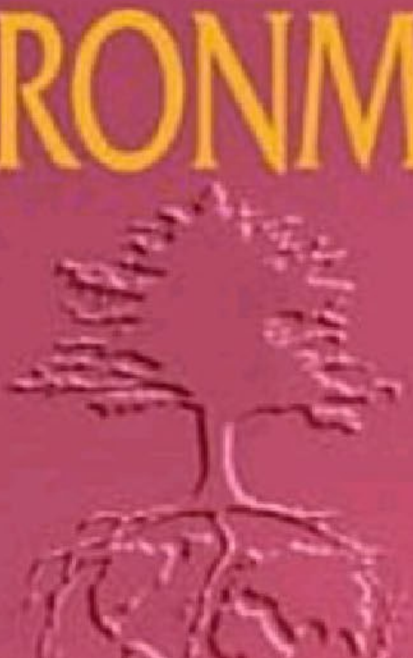
ENCYCLOPEDIA OF SOILS IN THE ENVIRONMENT



EDITED BY
DANIEL HILLEL
CYNTHIA ROSENZWEIG
DAVID POWLSON
KATE SCOW
MICHAIL SINGER
DONALD SPARKS

VOLUME ONE

ENCYCLOPEDIA OF SOILS IN THE ENVIRONMENT

FOUR-VOLUME SET

by Daniel Hillel (Editor-in-Chief)

Hardcover: 2200 pages
Publisher: Academic Press; 1 edition (November 8, 2004)
Language: English
ISBN-10: 0123485304
ISBN-13: 978-0123485304

Book Description

More than ever before, a compelling need exists for an encyclopedic resource about soil the rich mix of mineral particles, organic matter, gases, and soluble compounds that foster both plant and animal growth. Civilization depends more on the soil as human populations continue to grow and increasing demands are placed upon available resources.
The Encyclopedia of Soils in the Environmentis a comprehensive and integrated consideration of a topic of vital importance to human societies in the past, present, and future.

This important work encompasses the present knowledge of the world's variegated soils, their origins, properties, classification, and roles in the biosphere. A team of outstanding, international contributors has written over 250 entries that cover a broad range of issues facing today's soil scientists, ecologists, and environmental scientists.
This four-volume set features thorough articles that survey specific aspects of soil biology, ecology, chemistry and physics. Rounding out the encyclopedia's excellent coverage, contributions cover cross-disciplinary subjects, such as the history of soil utilization for agricultural and engineering purposes and soils in relation to the remediation of pollution and the mitigation of global climate change.

This comprehensive, yet accessible source is a valuable addition to the library of scientists, researchers, students, and policy makers involved in soil science, ecology, and environmental science.

Also available online via ScienceDirect featuring extensive browsing, searching, and internal cross-referencing between articles in the work, plus dynamic linking to journal articles and abstract databases, making navigation flexible and easy. For more information, pricing options and availability visit www.info.sciencedirect.com.

* A distinguished international group of editors and contributors
* Well-organized encyclopedic format providing concise, readable entries, easy searches, and thorough cross-references
* Abundant visual resources — photographs, figures, tables, and graphs — in every entry
* Complete up-to-date coverage of many important topics — essential information for scientists, students and professionals alike

EDITOR-IN-CHIEF

Daniel Hillel
Columbia University
New York, NY
USA

EDITORS

Jerry L Hatfield
National Soil Tilth Laboratory
Ames, IA
USA

David S Powlson
Rothamsted Research
Harpenden
UK

Cynthia Rosenzweig
NASA Goddard Institute for Space Studies
New York, NY
USA

Kate M Scow
University of California
Davis, CA
USA

Michael J Singer
University of California
Davis, CA
USA

Donald L Sparks
University of Delaware
Newark, DE
USA

FOREWORD

The *Encyclopedia of Soils in the Environment* is a vitally important scientific publication and an equally important contribution to global public policy. The *Encyclopedia* brings together a remarkable range of cutting-edge scientific knowledge on all aspects of soil science, as well as the links of soils and soil science to environmental management, food production, biodiversity, climate change, and many other areas of significant concern. Even more than that, the *Encyclopedia* will immediately become an indispensable resource for policy makers, analysts, and students who are focusing on one of the greatest challenges of the 21st century. With 6.3 billion people, our planet is already straining to feed the world's population, and is failing to do so reliably in many parts of the world. The numbers of chronically poor in the world have been stuck at some 800 million in recent years, despite long-standing international goals and commitments to reduce that number by several hundred million. Yet the challenge of food production will intensify in coming decades, as the human population is projected to rise to around 9 billion by mid-century, with the increased population concentrated in parts of the world already suffering from widespread chronic under-nourishment.

Unless the best science is brought to these problems, the situation is likely to deteriorate sharply. Food production systems are already under stress, for reasons often related directly to soils management. In Africa, crop yields are disastrously low and falling in many places due to the rampant depletion of soil nutrients. This situation needs urgent reversal, through increasing use of agro-forestry techniques (e.g. inter-cropping cereals with leguminous nitrogen-fixing trees) and increasing the efficient applications of chemical fertilizers. In other impoverished, as well as rich, parts of the planet, decades of intensive agriculture under irrigation have led to salinization, water-logging, eutrophication of major water bodies, dangerous declines of biodiversity and other forms of environmental degradation. These enormous strains are coupled with the continuing pressures of tropical deforestation and the lack of new promising regions for expanding crop cultivation to meet the needs of growing populations. Finally, there looms the prospect of anthropogenic climate change. Global warming and associated complex and poorly understood shifts in precipitation extremes and other climate variables all threaten the world's natural ecosystems and food production systems in profound yet still imperfectly understood ways. The risks of gradual or abrupt climate change are coupled with the risks of drastic perturbations to regional and global food supplies.

The *Encyclopedia* offers state-of-the-art contributions on each of these challenges, as well as links to entries on the fundamental biophysical processes that underpin the relevant phenomena. The world-scale and world-class collaboration that stands behind this unique project signifies its importance for the world community. It is an honor and privilege for me to introduce this path-breaking endeavor.

Jeffrey D Sachs
Director
The Earth Institute at Columbia University
Quetelet Professor of Sustainable Development
Columbia University, New York, USA

PREFACE

The term ‘soil’ refers to the weathered and fragmented outer layer of our planet’s land surfaces. Formed initially through the physical disintegration and chemical alteration of rocks and minerals by physical and biogeochemical processes, soil is influenced by the activity and accumulated residues of a myriad of diverse forms of life. As it occurs in different geologic and climatic domains, soil is an exceedingly variegated body with a wide range of attributes.

Considering the height of the atmosphere, the thickness of the earth’s rock mantle, and the depth of the ocean, one observes that soil is an amazingly thin body – typically not much more than one meter thick and often less than that. Yet it is the crucible of terrestrial life, within which biological productivity is generated and sustained. It acts like a composite living entity, a home to a community of innumerable microscopic and macroscopic plants and animals. A mere fistful of soil typically contains billions of microorganisms, which perform vital interactive biochemical functions. Another intrinsic attribute of the soil is its sponge-like porosity and its enormous internal surface area. That same fistful of soil may actually consist of several hectares of active surface, upon which physicochemical processes take place continuously.

Realizing humanity’s utter dependence on the soil, ancient peoples, who lived in greater intimacy with nature than many of us today, actually revered the soil. It was not only their source of livelihood, but also the material from which they built their homes and that they learned to shape, heat, and fuse into household vessels and writing tablets (ceramic, made of clayey soil, being the first synthetic material in the history of technology). In the Bible, the name assigned to the first human was Adam, derived from ‘adama,’ meaning soil. The name given to that first earthling’s mate was Hava (Eve, in transliteration), meaning ‘living’ or ‘life-giving.’ Together, therefore, Adam and Eve signified quite literally ‘Soil and Life.’

The same powerful metaphor is echoed in the Latin name for the human species – Homo, derived from humus, the material of the soil. Hence, the adjective ‘human’ also implies ‘of the soil.’ Other ancient cultures evoked equally powerful associations. To the Greeks, the earth was a manifestation of Gaea, the maternal goddess who, impregnated by Uranus (god of the sky), gave birth to all the gods of the Greek pantheon.

Our civilization depends on the soil more crucially than ever, because our numbers have grown while available soil resources have diminished and deteriorated. Paradoxically, however, even as our dependence on the soil has increased, most of us have become physically and emotionally detached from it. Many of the people in the so-called ‘developed’ countries spend their lives in the artificial environment of a city, insulated from direct exposure to nature, and some children may now assume as a matter of course that food originates in supermarkets.

Detachment has bred ignorance, and out of ignorance has come the delusion that our civilization has risen above nature and has set itself free of its constraints. Agriculture and food security, erosion and salination, degradation of natural ecosystems, depletion and pollution of surface waters and aquifers, and decimation of biodiversity – all of these processes, which involve the soil directly or indirectly – have become abstractions to many people. The very language we use betrays disdain for that common material underfoot, often referred to as ‘dirt.’ Some fastidious parents prohibit their children from playing in the mud and hurry to wash their ‘soiled’ hands when the children nonetheless obey an innate instinct to do so. Thus soil is devalued and treated as unclean though it is the terrestrial realm's principal medium of purification, wherein wastes are decomposed and nature's productivity is continually rejuvenated.

Scientists who observe soil closely see it in effect as a seething foundry in which matter and energy are in constant flux. Radiant energy from the sun streams onto the field and cascades through the soil and the plants growing in it. Heat is exchanged, water percolates through the soil's intricate passages, plant roots extract water and transmit it to their leaves, which transpire it back to the atmosphere. Leaves absorb carbon dioxide from the air and synthesize it with soil-derived water to form the primary compounds of life. Oxygen emitted by the leaves makes the air breathable for animals, which consume and in turn fertilize plants.

Soil is thus a self-regulating bio-physio-chemical factory, processing its own materials, water, and solar energy. It also determines the fate of rainfall and snowfall reaching the ground surface – whether the water thus received will flow over the land as runoff, or seep downward to the subterranean reservoir called groundwater, which in turn maintains the steady flow of springs and streams. With its finite capacity to absorb and store moisture, and to release it gradually, the soil regulates all of these phenomena. Without the soil as a buffer, rain falling over the continents would run off entirely, producing violent floods rather than sustained river flow.

Soil naturally acts as a living filter, in which pathogens and toxins that might otherwise accumulate to foul the terrestrial environment are rendered harmless. Since time immemorial, humans and other animals have been dying of all manner of disease and have then been buried in the soil, yet no major disease is transmitted by it. The term *antibiotic* was coined by soil microbiologists who, as a consequence of their studies of soil bacteria and actinomycetes, discovered streptomycin (an important cure for tuberculosis and other infections). Ion exchange, a useful process of water purification, also was discovered by soil scientists studying the passage of solutes through beds of clay.

However unique in form and function, soil is not an isolated body. It is, rather, a central link in the larger chain of interconnected domains and processes comprising the terrestrial environment. The soil interacts both with the overlying atmosphere and the underlying strata, as well as with surface and underground bodies of water. Especially important is the interrelation between the soil and the climate. In addition to its function of regulating the cycle of water, it also regulates energy exchange and surface temperature.

When virgin land is cleared of vegetation and turned into a cultivated field, the native biomass above the ground is often burned and the organic matter within the soil tends to decompose. These processes release carbon dioxide into the atmosphere, thus contributing to the earth's greenhouse effect and to global warming. On the other hand, the opposite act of reforestation and soil enrichment with organic matter, such as can be achieved by means of conservation management, may serve to absorb carbon dioxide from the atmosphere. To an extent, the soil's capacity to store carbon can thus help to mitigate the greenhouse effect.

Thousands of years are required for nature to create life-giving soil out of sterile bedrock. In only a few decades, however, unknowing or uncaring humans can destroy that wondrous work of nature. In various circumstances, mismanaged soils may be subject to erosion (the sediments of which tend to clog streambeds, estuaries, lakes, and coastal waters), to leaching of nutrients with attendant loss of fertility and eutrophication of water bodies, to waterlogging and impaired aeration, or to an excessive accumulation of salts that may cause a once-productive soil to become entirely sterile. Such processes of soil degradation, sometimes called 'desertification,' already affect large areas of land.

We cannot manage effectively and sustainably that which we do not know and thoroughly understand. That is why the tasks of developing and disseminating sound knowledge of the soil and its complex processes have assumed growing urgency and importance. The global environmental crisis has created a compelling need for a concentrated, concise, and definitive source of information – accessible to students, scientists, practitioners, and the general public – about the soil in all its manifestations – in nature and in relation to the life of humans.

Daniel Hillel
Editor-in-Chief
May 2004

INTRODUCTION

The *Encyclopedia of Soils in the Environment* contains nearly 300 articles, written by the world's leading authorities. Pedologists, biologists, ecologists, earth scientists, hydrologists, climatologists, geographers, and representatives from many other disciplines have contributed to this work. Each of the articles separately, and all of them in sequence and combination, serve to summarize and encapsulate our present knowledge of the world's variegated soils, their natural functions, and their importance to humans.

Concise articles surveying specific aspects of soils (soil genesis, soil chemistry and mineralogy, soil physics and hydrology, and soil biology) are complemented by articles covering transdisciplinary aspects, such as the role of soils in ecology, the history of soil utilization for agricultural and engineering purposes, the development of soil science as a discipline, and the potential or actual contributions of soils to the generation, as well as to the mitigation, of pollution and of global climate change.

This comprehensive reference encompasses both the fundamental and the applied aspects of soil science, interfacing in general with the physical sciences and life sciences and more specifically with the earth sciences and environmental sciences.

The *Encyclopedia of Soils in the Environment* manifests the expanding scope of modern soil science, from its early sectarian focus on the utilitarian attributes of soils in agriculture and engineering, to a wider and much more inclusive view of the soil as a central link in the continuous chain of processes constituting the dynamic environment as a whole. Thus it both details and integrates a set of topics that have always been of vital importance to human societies and that are certain to be even more so in the future.

Daniel Hillel
Editor-in-Chief
May 2004

CONTENTS

Contents are given as follows: CHAPTER NAME Author(s) Page number

VOLUME 1

A

ACID RAIN AND SOIL ACIDIFICATION L Blake 1

ACIDITY N Bolan, D Curtin and D Adriano 11

AERATION D E Rolston 17

AGGREGATION

Microbial Aspects S D Frey 22

Physical Aspects J R Nimmo 28

AGROFORESTRY P K R Nair 35

AIR PHASE see AERATION; DIFFUSION

ALBEDO see ENERGY BALANCE; RADIATION BALANCE

ALLOPHANE AND IMOGOLITE see AMORPHOUS MATERIALS

ALLUVIUM AND ALLUVIAL SOILS J L Boettinger 45

ALUMINUM SPECIATION D R Parker 50

AMMONIA D E Kissel and M L Cabrera 56

AMORPHOUS MATERIALS J Harsh 64

ANAEROBIC SOILS P W Inglett, K R Reddy and R Corstanje 72

ANION EXCHANGE see CATION EXCHANGE

APPLICATIONS OF SOILS DATA P J Lawrence 78

ARCHAEA J E T McLain 88

ARCHEOLOGY IN RELATION TO SOILS J A Homburg 95

B

BACTERIA

Plant Growth-Promoting Y Bashan and L E de-Bashan 103

Soil L J Halverson 115

BACTERIOPHAGE M Radosevich, K E Williamson and K E Wommack 122

BIOCONTROL OF SOIL-BORNE PLANT DISEASES C E Pankhurst and J M Lynch 129

BIODIVERSITY D H Wall 136

BUFFERING CAPACITY B R James 142

BULK DENSITY see POROSITY AND PORE-SIZE DISTRIBUTION

C

CALCIUM AND MAGNESIUM IN SOILS N Bolan, P Loganathan and S Saggar 149

CAPILLARITY D Or and M Tuller 155

CARBON CYCLE IN SOILS

Dynamics and Management C W Rice 164

Formation and Decomposition C A Cambardella 170

CARBON EMISSIONS AND SEQUESTRATION K Paustian 175

CATION EXCHANGE L M McDonald, V P Evangelou and M A Chappell 180

CHEMICAL EQUILIBRIA A P Schwab 189

CHEMICAL SPECIATION MODELS see SURFACE COMPLEXATION MODELING

CHERNOZEMS see GRASSLAND SOILS

CHILDS, ERNEST CARR E G Youngs 195

CIVILIZATION, ROLE OF SOILS D Hillel 199

CLASSIFICATION OF LAND USE see LAND-USE CLASSIFICATION

CLASSIFICATION OF SOILS R W Arnold 204

CLASSIFICATION SYSTEMS

Australian R W Fitzpatrick 211

FAO F O Nachtergaele 216

Russian, Background and Principles M Gerasimova 223

Russian, Evolution and Examples D Konyushkov 227

USA D J Brown 235

CLAY MINERALS D G Schulze 246

CLIMATE CHANGE IMPACTS P Bullock 254

CLIMATE MODELS, ROLE OF SOIL P Smith 262

COLD-REGION SOILS C-L Ping 268

COLLOID-FACILITATED SORPTION AND TRANSPORT R Kretzschmar 276

COMPACTION J J H van den Akker and B Soane 285

COMPOST T L Richard 294

CONDITIONERS R E Sojka, J A Entry and W J Orts 301

CONSERVATION see EROSION: Water-Induced; Wind-Induced; SUSTAINABLE SOIL AND LAND MANAGEMENT; TERRACES AND TERRACING

CONSERVATION TILLAGE M R Carter 306

COVER CROPS L Edwards and J Burney 311

CROP ROTATIONS C A Francis 318

CROP WATER REQUIREMENTS L S Pereira and I Alves 322

CROP-RESIDUE MANAGEMENT D C Reicosky and A R Wilts 334

CRUSTS

Biological J Belnap 339

Structural R L Baumhardt and R C Schwartz 347

CULTIVATION AND TILLAGE M R Carter and E McKyes 356

D

DARCY’S LAW D Swartzendruber 363

DEGRADATION C J Ritsema, G W J van Lynden, V G Jetten and S M de Jong 370

DENITRIFICATION D A Martens 378

DESERTIFICATION D Hillel and C Rosenzweig 382

DIFFUSION T Addiscott and P Leeds-Harrison 389

DISINFESTATION A Gamliel and J Katan 394

DISPERSION see FLOCCULATION AND DISPERSION

DISSOLUTION PROCESSES, KINETICS K G Scheckel and C A Impellitteri 400

DRAINAGE, SURFACE AND SUBSURFACE N R Fausey 409

DRYLAND FARMING G A Peterson 414

E

EARTHWORMS see FAUNA

EDAPHOLOGY A L Ulery 419

ELECTRON PARAMAGNETIC RESONANCE see ELECTRON-SPIN RESONANCE SPECTROSCOPY

ELECTRON-SPIN RESONANCE SPECTROSCOPY N Senesi and G S Senesi 426

ELECTROSTATIC DOUBLE-LAYER see CATION EXCHANGE

ENERGY BALANCE M Fuchs 438

ENVIRONMENTAL MONITORING P J Loveland and P H Bellamy 441

ENZYMES IN SOILS R P Dick and E Kandeler 448

EROSION

- Irrigation-Induced G A Lehrsch, D L Bjorneberg and R E Sojka 456
- Water-Induced J E Gilley 463
- Wind-Induced T M Zobeck and R S Van Pelt 470

ESSENTIAL ELEMENTS E A Kirkby 478

EUTROPHICATION A J Gold and J T Sims 486

EVAPORATION OF WATER FROM BARE SOIL C W Boast and F W Simmons 494

EVAPOTRANSPIRATION G Stanhill 502

F

FACTORS OF SOIL FORMATION

- Biota A H Jahren 507
- Climate O C Spaargaren and J A Deckers 512
- Human Impacts J Sandor, C L Burras and M Thompson 520
- Parent Material K R Olson 532
- Time E F Kelly and C M Yonker 536

FAUNA T Winsome 539

VOLUME 2

FERTIGATION U Kafkafi and S Kant 1

FERTILITY J L Havlin 10

FERTILIZERS AND FERTILIZATION H W Scherer 20

FIELD CAPACITY see WATER CYCLE

FLOCCULATION AND DISPERSION I Shainberg and G J Levy 27

FLUORESCENCE SPECTROSCOPY N Senesi and V D’Orazio 35

FOLIAR APPLICATIONS OF NUTRIENTS M Tagliavini and M Toselli 53

FOOD-WEB INTERACTIONS P C de Ruiter and J C Moore 59

FORENSIC APPLICATIONS W F Rowe 67

FOREST SOILS J R Boyle 73

FOURIER TRANSFORM INFRARED SPECTROSCOPY D Peak 80

FRACTAL ANALYSIS Y Pachepsky and J W Crawford 85

FREEZING AND THAWING

Cycles B Sharratt 98

Processes G N Flerchinger, G A Lehrsch and D K McCool 104

FUNGI K Ritz 110

G

GEOGRAPHICAL INFORMATION SYSTEMS J Bo¨hner, T Selige and R Ko¨the 121

GERMINATION AND SEEDLING ESTABLISHMENT A Hadas 130

GLOBAL WARMING see CARBON EMISSIONS AND SEQUESTRATION; CLIMATE CHANGE IMPACTS;

GREENHOUSE GAS EMISSIONS

GRASSLAND SOILS J A Mason and C W Zanner 138

GREEN MANURING see COVER CROPS

GREENHOUSE GAS EMISSIONS K A Smith 145

GROUNDWATER AND AQUIFERS Y Bachmat 153

GROUNDWATER POLLUTION see POLLUTION: Groundwater

H

HEAT AND MOISTURE TRANSPORT R Horton and A Globus 169

HEAT CAPACITY see THERMAL PROPERTIES AND PROCESSES

HEAT FLOW see THERMAL PROPERTIES AND PROCESSES

HEAVY METALS D C Adriano, N S Bolan, J Vangronsveld and W W Wenzel 175

HILGARD, EUGENE WOLDEMAR R Amundson 182

HOOGHOUDT, SYMEN BAREND P A C Raats and R R van der Ploeg 188

HUMIFICATION T C Balser 195

HYDRAULIC PROPERTIES, TEMPERATURE EFFECTS S A Grant 207

HYDRIC SOILS G W Hurt 212

HYDROCARBONS P Kostecki, R Morrison and J Dragun 217

HYDRODYNAMIC DISPERSION see SOLUTE TRANSPORT

HYDRODYNAMICS IN SOILS T P A Ferre´ and A W Warrick 227

HYSTERESIS J H Dane and R J Lenhard 231

I

IMMISCIBLE FLUIDS R J Lenhard, J H Dane and M Oostrom 239

INCEPTISOLS A Palmer 248

INDUSTRIAL POLLUTION see POLLUTION: Industrial

INFILTRATION T P A Ferre´ and A W Warrick 254

INFRARED SPECTROSCOPY see FOURIER TRANSFORM INFRARED SPECTROSCOPY

IRON NUTRITION K Mengel and H Kosegarten 260

IRRIGATION

Environmental Effects S Topcu and C Kirda 267

Methods D L Bjorneberg and R E Sojka 273

ISOTOPES IN SOIL AND PLANT INVESTIGATIONS K Reichardt and O O S Bacchi 280

ISOTROPY AND ANISOTROPY T-C J Yeh, P Wierenga, R Khaleel and R J Glass 285

J

JENNY, HANS R Amundson 293

K

KELLOGG, CHARLES J D Helms 301

KINETIC MODELS P M Jardine 307

KIRKHAM, DON D R Nielsen and R R van der Ploeg 315

L

LAMINAR AND TURBULENT FLOW see HYDRODYNAMICS IN SOILS

LANDFILLS see WASTE DISPOSAL ON LAND: Municipal

LAND-USE CLASSIFICATION J A LaGro Jr 321

LAWES, JOHN BENNET AND GILBERT, JOSEPH HENRY A E Johnston 328

LEACHING PROCESSES B E Clothier and S Green 336

LIEBIG, JUSTUS VON R R van der Ploeg, W Bo¨hm and M B Kirkham 343

LIMING E J Kamprath and T J Smyth 350

LIPMAN, JACOB G. J C F Tedrow 358

LOESS A J Busacca and M R Sweeney 364

LOWDERMILK, WALTER CLAY J D Helms 373

LYSIMETRY T A Howell 379

M

MACRONUTRIENTS C W Wood, J F Adams and B H Wood 387

MACROPORES AND MACROPORE FLOW, KINEMATIC WAVE APPROACH P F Germann 393

MAGNESIUM IN SOILS see CALCIUM AND MAGNESIUM IN SOILS

MANURE MANAGEMENT J T Sims and R O Maguire 402

MARBUT, CURTIS FLETCHER J P Tandarich 410

MATRIC POTENTIAL see HYDRODYNAMICS IN SOILS; WATER POTENTIAL; WATER RETENTION AND CHARACTERISTIC CURVE

MEDITERRANEAN SOILS J Torrent 418

METAL OXIDES A C Scheinost 428

METALS AND METALLOIDS, TRANSFORMATION BY MICROORGANISMS S M Glasauer, T J Beveridge, E P Burford, F A Harper and G M Gadd 438

METALS, HEAVY see HEAVY METALS

MICROBIAL PROCESSES

Environmental Factors P G Hartel 448

CommunityAnaly sis C H Nakatsu 455

Kinetics N S Panikov 463

MICRONUTRIENTS L M Shuman 479

MINERAL–ORGANIC–MICROBIAL INTERACTIONS P M Huang, M C Wang and M K Wang 486

MINERALS, PRIMARY P M Huang and M K Wang 500

MINERALS, SECONDARY see CLAY MINERALS

MINIMUM TILLAGE see CONSERVATION TILLAGE

MISCIBLE DISPLACEMENT see SOLUTE TRANSPORT

MORPHOLOGY P R Owens and E M Rutledge 511

MULCHES C L Acharya, K M Hati and K K Bandyopadhyay 521

MYCORRHIZAL FUNGI L M Egerton-Warburton, J I Querejeta, M F Allen and S L Finkelman 533

VOLUME 3

N

NEMATODES D A Neher and T O Powers 1

NEUTRON SCATTERING M J Fayer and G W Gee 6

NITROGEN IN SOILS

Cycle M S Coyne and W W Frye 13

Nitrates D S Powlson and T M Addiscott 21

Nitrification J I Prosser 31

Plant Uptake A Hodge 39

Symbiotic Fixation J I Sprent 46

NITROGEN FERTILIZERS see FERTILIZERS AND FERTILIZATION

NUCLEAR WASTE DISPOSAL G W Gee, P D Meyer and A L Ward 56

NUTRIENT AVAILABILITY N K Fageria and V C Baligar 63

NUTRIENT MANAGEMENT G D Binford 71

O

ORGANIC FARMING C A Francis 77

ORGANIC MATTER

Principles and Processes M Schnitzer 85

Genesis and Formation K M Haider and G Guggenberger 93

Interactions with Metals N Senesi and E Loffredo 101

ORGANIC RESIDUES, DECOMPOSITION A J Franzluebbers 112

ORGANIC SOILS D L Mokma 118

OVERLAND FLOW T S Steenhuis, L Agnew, P Ge´rard-Marchant and M T Walter 130

OXIDATION–REDUCTION OF CONTAMINANTS C J Matocha 133

P

PADDY SOILS C Witt and S M Haefele 141

PARENT MATERIAL see PEDOLOGY: Basic Principles; FACTORS OF SOIL FORMATION: Parent Material

PEDOLOGY

Basic Principles M J Singer 151

Dynamic F C Ugolini 156

PEDOMETRICS I O A Odeh and A B McBratney 166

PENMAN, HOWARD LATIMER J L Monteith 176

PENMAN–MONTEITH EQUATION R Allen 180

PERCOLATION see HYDRODYNAMICS IN SOILS

PERMAFROST see POLAR SOILS

PERMEABILITY see HYDRODYNAMICS IN SOILS

PERSISTENT ORGANIC POLLUTANTS (POPS) see POLLUTANTS: Persistent Organic (POPs)

PESTICIDES R H Bromilow 188

PETROLEUM see HYDROCARBONS

pH N Bolan and K Kandaswamy 196

PHOSPHORUS IN SOILS

Overview J T Sims and P A Vadas 202

Biological Interactions M D Mullen 210

PHYTOTOXIC SUBSTANCES IN SOILS M Qadir, S Schubert and D Steffens 216

PLANT–SOIL–WATER RELATIONS R A Feddes and J C van Dam 222

PLANT–WATER RELATIONS C Gimenez, M Gallardo and R B Thompson 231

POISEUILLE’S LAW see HYDRODYNAMICS IN SOILS

POLAR SOILS J C F Tedrow 239

POLLUTANTS

Biodegradation P B Hatzinger and J W Kelsey 250

Effects on Microorganisms M E Fuller 258

Persistent Organic (POPs) D Johnson 264

POLLUTION

Groundwater H Rubin 271

Industrial S P McGrath 282

POLYMERS AND MICROORGANISMS M C Rillig 287

POORLY CRYSTALLINE ALLUMINOSILICATES see AMORPHOUS MATERIALS

POROSITY AND PORE-SIZE DISTRIBUTION J R Nimmo 295

POTASSIUM IN SOILS P M Huang, J M Zhou, J C Xie and M K Wang 303

PRECIPITATION, WATERSHED ANALYSIS J V Bonta 314

PRECIPITATION–DISSOLUTION PROCESSES W P Robarge 322

PRECISION AGRICULTURE see SITE-SPECIFIC SOIL MANAGEMENT

PREFERENTIAL FLOW see UNSTABLE FLOW; MACROPORES AND MACROPORE FLOW, KINEMATIC WAVE APPROACH

PRODUCTIVITY D L Karlen 330

PROFILE see MORPHOLOGY

PROTOZOA W Foissner 336

Q

QUALITY OF SOIL B J Wienhold, G E Varvel and J W Doran 349

R

RADIATION BALANCE J L Hatfield, T J Sauer and J H Prueger 355

RADIONUCLIDES see ISOTOPES IN SOIL AND PLANT INVESTIGATIONS

RAINFED FARMING see DRYLAND FARMING

RANGE MANAGEMENT G L Anderson 360

RECYCLING OF ORGANIC WASTES see POLLUTANTS: Biodegradation

REDISTRIBUTION see WATER CYCLE

REDOX POTENTIAL R D DeLaune and K R Reddy 366

REDOX REACTIONS, KINETICS P S Nico and S Fendorf 372

REMEDIATION OF POLLUTED SOILS E Lombi and R E Hamon 379

REMOTE SENSING

- Organic Matter D K Morris, C J Johannsen, S M Brouder and G C Steinhardt 385
- Soil Moisture T J Jackson 392

RHIZOSPHERE A C Kennedy and L Z de Luna 399

RICHARDS, LORENZO A. W R Gardner 407

ROOT ARCHITECTURE AND GROWTH L E Jackson 411

ROOT EXUDATES AND MICROORGANISMS B-J Koo, D C Adriano, N S Bolan and C D Barton 421

S

SALINATION PROCESSES I Shainberg and G J Levy 429

SALINITY

Management D Hillel 435

Physical Effects D Russo 442

SALT BALANCE OF SOILS see SALINATION PROCESSES

SALT-AFFECTED SOILS, RECLAMATION R Keren 454

SAND DUNES H Tsoar 462

SATURATED AND UNSATURATED FLOW see HYDRODYNAMICS IN SOILS; VADOSE ZONE: Hydrologic Processes

SCALING

Physical Properties and Processes G Sposito 472

Transport Processes R P Ewing 477

SEPTIC SYSTEMS R L Lavigne 485

SHIFTING CULTIVATION R Lal 488

SITE-SPECIFIC SOIL MANAGEMENT C J Johannsen and P G Carter 497

SLASH AND BURN AGRICULTURE see SHIFTING CULTIVATION

SLUDGE see WASTE DISPOSAL ON LAND: Liquid; Municipal

SODIC SOILS G J Levy and I Shainberg 504

SOIL–PLANT–ATMOSPHERE CONTINUUM J M Norman and M C Anderson 513

SOLUTE TRANSPORT M C Sukop and E Perfect 521

SORPTION

Metals D L Sparks 532

Organic Chemicals B Xing and J J Pignatello 537

Oxyanions C P Schulthess, H Wijnja and W Yang 548

SORPTION–DESORPTION, KINETICS D L Sparks 556

SPATIAL PATTERNS J H Go¨rres and J A Amador 562

VOLUME 4

SPATIAL VARIATION, SOIL PROPERTIES R Webster 1

SPECIFIC SURFACE AREA K D Pennell 13

STATISTICS IN SOIL SCIENCE R Webster 19

STERILIZATION see DISINFESTATION

STOCHASTIC ANALYSIS OF SOIL PROCESSES D Russo 29

STRESS–STRAIN AND SOIL STRENGTH S K Upadhyaya 38

STRUCTURE V A Snyder and M A Va´zquez 54

SUBSOILING R L Raper 69

SULFUR IN SOILS

Overview M A Tabatabai 76

Biological Transformations S D Siciliano and J J Germida 85

Nutrition M A Tabatabai 91

SURFACE COMPLEXATION MODELING S Goldberg 97

SUSTAINABLE SOIL AND LAND MANAGEMENT J L Berc 108

SWELLING AND SHRINKING D Smiles and P A C Raats 115

T

TEMPERATE REGION SOILS E A Nater 125

TEMPERATURE REGIME see THERMAL PROPERTIES AND PROCESSES

TENSIOMETRY T K Tokunaga 131

TERMITES see FAUNA

TERRA ROSSA see MEDITERRANEAN SOILS

TERRACES AND TERRACING G R Foster 135

TESTING OF SOILS A P Mallarino 143

TEXTURE G W Gee 149

THERMAL PROPERTIES AND PROCESSES D Hillel 156

THERMODYNAMICS OF SOIL WATER P H Groenevelt 163

TILLAGE see CONSERVATION TILLAGE; CULTIVATION AND TILLAGE; ZONE TILLAGE

TILTH D L Karlen 168

TIME-DOMAIN REFLECTOMETRY G C Topp and T P A Ferre´ 174

TROPICAL SOILS

Arid and Semiarid H C Monger, J J Martinez-Rios and S A Khresat 182

Humid Tropical S W Buol 187

U

UNSTABLE FLOW T S Steenhuis, J-Y Parlange, Y-J Kim, D A DiCarlo, J S Selker, P A Nektarios, D A Barry and F Stagnitti 197

URBAN SOILS J L Morel, C Schwartz, L Florentin and C de Kimpe 202

V

VADOSE ZONE

Hydrologic Processes J W Hopmans and M Th van Genuchten 209

Microbial Ecology P A Holden and N Fierer 216

VIRUSES see BACTERIOPHAGE

VOLCANIC SOILS G Uehara 225

W

WAKSMAN, SELMAN A. H B Woodruff 233

WASTE DISPOSAL ON LAND

Liquid C P Gerba 238

Municipal D A C Manning 247

WATER AVAILABILITY see PLANT–SOIL–WATER RELATIONS

WATER CONTENT AND POTENTIAL, MEASUREMENT G S Campbell and C S Campbell 253

WATER CYCLE D K Cassel and B B Thapa 258

WATER EROSION see EROSION: Water-Induced

WATER HARVESTING D Hillel 264

WATER MANAGEMENT see CROP WATER REQUIREMENTS

WATER POTENTIAL D Or, M Tuller and J M Wraith 270

WATER REQUIREMENTS see CROP WATER REQUIREMENTS

WATER RETENTION AND CHARACTERISTIC CURVE M Tuller and D Or 278

WATER TABLE see GROUNDWATER AND AQUIFERS

WATER, PROPERTIES D Hillel 290

WATER-REPELLENT SOILS J Letey 301

WATERSHED MANAGEMENT M D Tomer 306

WATER-USE EFFICIENCY M B Kirkham 315

WEED MANAGEMENT D D Buhler 323

WETLANDS, NATURALLY OCCURRING E K Hartig 328

WIDTSOE, JOHN A. AND GARDNER, WILLARD G S Campbell and W H Gardner 335

WIND EROSION see EROSION: Wind-Induced

WINDBREAKS AND SHELTERBELTS E S Takle 340

WOMEN IN SOIL SCIENCE (USA) M J Levin 345

WORLD SOIL MAP H Eswaran and P F Reich 352

Z

ZERO-CHARGE POINTS J Chorover 367

ZONE TILLAGE J L Hatfield and A T Jeffries 373

Table of Contents, Volume 1

A **1**
Acid Rain and Soil Acidification **1**
Introduction 1
Acid Rain 1
Atmospheric Transport, Secondary Chemistry, and Acid Rain 2
Dry oxidation of sulfur and nitrogen oxides 2
Wet oxidation of sulfur 2
Acid Soil 3
Sources of Acids in Soils 4
Internal Acid Production in Soils 4
Carbonic and organic acid production 4
Nitrogen accumulation and transformations 5
Assimilation of Nutrients by Vegetation 5
Deposition of Nitrogen Species and Nitrogen Transformations 5
Addition of Dissolved Strong Acids Through Acid Deposition 5
Observations of Changes in Soil Acidification due to Acid Deposition 5
Proton Sinks and Buffer Ranges 6
Carbonate and Silicate Buffers 7
Cation Exchange Buffer 7
Aluminum and Iron Buffer 8
Proton Adsorption Reactions 9
Acid Sources and Sinks: Input-Output Relationships and the Calculation of Proton Balances 9
Acidification, Ecosystem Stability and Global Change 9
Further Reading 11
Acidity **11**
Introduction 11
Processes Generating Acidity in Soils 12
Acid Drainage 12
Acid Precipitation 12
Elemental Cycling 13
Fertilizer Reactions 14
Measurement and Effects of Soil Acidity 15
Amelioration of Soil Acidity 16
Further Reading 17
Aeration **17**
Introduction 17
Soil-Air Composition 17
Gas Exchange Mechanisms 18
Diffusion 18
Convection 18
Gas Reactions 19
Respiration 19
Oxidation-Reduction Processes 19
Production and Consumption of Other Gases 19
Aeration Requirements 20
Plants 20
Remediation of Contaminated Soils 20
Summary 20
Further Reading 21
Aggregation **22**
Microbial Aspects 22
Introduction 22
Microorganisms and Aggregate Formation 22

Soil Aggregates as Habitats for Microbiota 25
Soil Structure and Microorganism Interactions..... 26
Protection of Organic Matter Conferred by Soil Aggregates 26
Soil as a Spatially Continuous Medium 27
List of Technical Nomenclature 28
Further Reading 28
Physical Aspects..... 28
Forces on Soil Particles..... 29
Aggregate Physical Properties and their Measurement 31
Basic Properties..... 31
Size and Stability 32
Fundamentals 32
Commonly used methods 32
Miscellaneous methods 33
Issues of general importance..... 33
Representation and Interpretation..... 33
Summary 34
Further Reading 34
Agroforestry 35
Introduction..... 35
Emergence of Agroforestry as a Land-Use Approach 35
Agroforestry Systems 36
Examples of Common Agroforestry Systems 38
Improved Fallow 38
Alley Cropping 38
Homegarden, Shaded-Perennial, and Multistrata Systems 39
Silvopastoral Systems 40
Agroforestry Applications for Soil Conservation and Reclamation 41
Other Agroforestry Applications 41
Agroforestry and Ecosystem Services..... 42
Soil Productivity and Protection..... 42
Carbon Sequestration 42
Agroforestry Research..... 43
Future Directions 43
Acknowledgments..... 43
List of Technical Nomenclature 44
Further Reading..... 44
Alluvium and Alluvial Soils 45
Introduction..... 45
Alluvium 45
Composition 45
Texture 45
Landforms 46
Alluvial Soils..... 47
Recent Alluvial Soils..... 47
Older Alluvial Soils 47
Classification 48
Human Use of Alluvial Soils 48
List of Technical Nomenclature 49
Further Reading..... 49
Aluminum Speciation 50
Introduction..... 50
Inorganic Aluminum Species in Soil Solution 50
Organic Complexes of Aluminum 51
Determining Aluminum Speciation in Experimental Samples..... 53
Computational Methods 53

    - Fractionation Based on Size, Charge, and Reactivity 53
    - Speciation Using More Direct Analytical Methods 54
  - Further Reading 56
- **Ammonia 56**
  - Introduction 56
  - Chemical Reactions of Ammoniacal N 57
    - Ammonia-Ammonium Equilibrium 57
    - Retention of Ammonia by Soils 57
  - Retention and Loss of Surface-Applied Ammoniacal N in Soils 59
    - Reactions of Urea in Soils 59
    - Environmental Conditions and Ammonia Formation 60
  - Urea concentration 60
  - Soil pH and H+ buffering capacity 61
  - Factors affecting the rate of urea hydrolysis 61
  - Soil urease concentration 61
  - Urea concentration 62
  - Soil temperature 62
  - Soil water content 62
  - Soil pH 62
    - Reactions of NH4+-N Applied to Calcareous Soils 62
  - List of Technical Nomenclature 63
  - Further Reading 63
- **Amorphous Materials 64**
  - Introduction 64
  - Occurrence in Soils 64
  - Identification 65
  - Structure and Charge 67
  - Chemisorption and Ligand Exchange 68
  - Reaction with Soil Organic Matter 70
  - Aluminum Solubility 70
  - Physical Properties 71
  - Further Reading 71
- **Anaerobic Soils 72**
  - Physical Characteristics 72
  - Biological Characteristics 73
  - Chemical Characteristics 73
  - Morphological Characteristics of Wetland Soils 74
  - Biogeochemical Characteristics 74
    - Microbial Respiration 75
  - Agronomic, Ecologic, and Environmental Significance 77
  - Further Reading 78
- **Applications of Soils Data 78**
  - Introduction 78
  - Development of Interpretations in the USA 78
  - Interpretations Beyond Agriculture 80
  - Today’s Interpretive Categories 81
    - National Inventory Groupings 81
    - Land-Use Planning 81
    - Farmland 81
    - Rangeland 81
    - Forest Land 81
    - Windbreaks 81
    - Recreation 81
    - Wildlife Habitat 83
    - Construction Materials 83
    - Building Sites 83

Waste Disposal ... 83
Water Management ... 83
Uses of Soils Data and Interpretations ... 83
Federal and State Programs ... 83
State and Local Planning ... 84
Natural Resource Management ... 84
Land Appraisal and Assessment ... 84
Engineering and Construction ... 86
Hazardous Waste, Brownfields, and Remediation ... 86
Research and Analyses ... 86
Conclusion ... 88
Further Reading ... 88
Archaea ... 88
Introduction ... 88
Archaeal Evolution ... 89
The Archaeal Cell ... 89
Three Archaeal Groups ... 90
Extreme Halophiles ... 90
Methanogens ... 91
Hyperthermophiles ... 92
Psychrophiles ... 93
Thermoplasma ... 93
Biotechnological Use of Extremoenzymes ... 93
Further Reading ... 94
Archeology in Relation to Soils ... 95
Introduction ... 95
Pedostratigraphy and the Archeological Record ... 95
Archeological Soil Chemistry ... 97
Soil as a Resource ... 98
Further Reading ... 101
B ... 103
Bacteria ... 103
Plant Growth-Promoting ... 103
Introduction ... 103
Plant Growth-Promoting Bacteria in Agriculture and the Environment ... 103
Endophytic PGPB ... 108
Biocontrol of Phytopathogens ... 108
Mechanisms Employed by Biocontrol-PGPB to Control Phytopathogens ... 109
Production of Antibiotics ... 109
Production of Siderophores ... 110
Production of Small Molecules ... 110
Production of Enzymes ... 110
Competition and Displacement of Pathogens ... 111
Modification of Plant Metabolism ... 112
Induced and Acquired Systemic Resistance ... 112
Modification of Plant Ethylene Levels ... 112
Prospects for Improving PGPB by Genetic Manipulation ... 112
PGPB Inoculants ... 112
Conclusions and Prospects ... 113
Further Reading ... 114
Soil ... 115
Introduction ... 115
Bacterial Diversity ... 115
The Microbe as the Scale of Study ... 116
Two Basic Survival Strategies ... 117

Avoidance ... 117
Stress Tolerance ... 118
Water Availability ... 118
Extracellular Polysaccharide Production ... 118
Biofilms ... 119
Competition for Resources ... 119
Cooperation Among Bacteria ... 120
Siderophores ... 120
Quorum-Sensing ... 120
Death ... 121
Summary ... 121
Further Reading ... 122
Bacteriophage ... 122
Introduction ... 122
Why Soil Bacteriophage? ... 122
Ecologic Role of Viruses in Microbial Communities ... 123
Bacteriophage Contain Genetic Elements that Can Increase Host Fitness: Bacteriophage Conversion ... 123
The Role of Phage in Horizontal Gene Transfer and Bacterial Evolution ... 123
Evidence for Natural Transduction in Microbial Communities ... 124
Cultivation-Independent Investigation of Soil Viral Communities ... 125
Microscopy of Soil Viruses ... 125
Genomic Approaches to Viral Diversity ... 127
Summary ... 128
List of Technical Nomenclature ... 128
Further Reading ... 128
Biocontrol of Soil-Borne Plant Diseases ... 129
Introduction ... 129
Enhancement of Biocontrol by Agronomic Practices ... 130
Crop Rotations ... 130
Organic Amendments ... 130
Biocontrol with Introduced Microbial Biocontrol Agents ... 131
Selection, Evaluation, and Delivery of Microbial Biocontrol Agents ... 131
Mode of Action of Biocontrol Agents ... 133
Competition ... 133
Antibiosis ... 133
Parasitism ... 133
Induced systemic resistance ... 134
Plant growth-promoting rhizobacteria ... 134
Future Prospects ... 134
Conclusions ... 135
Further Reading ... 135
Biodiversity ... 136
Introduction ... 136
Diversity and Abundance ... 136
Distribution ... 138
Soil Biodiversity and Ecosystem Functioning ... 139
Global Change and Impacts on Soil Biodiversity ... 140
Sustaining Soil Biodiversity ... 141
Further Reading ... 141
Buffering Capacity ... 142
Introduction ... 142
The Samovar Analogy for Buffering in Soils ... 142
Buffering of Soil Acidity ... 143
Dissolution and Precipitation of CaCO3 and Al(OH)3 ... 144

Cation Exchange Reactions and pH Buffering........144
pH-Dependent Surface Charge and pH Buffering........145
Buffering of Soil Oxidation-Reduction Status........146
Buffering of Ion Activities via Dissolution-Precipitation, Ion Exchange, and Ligand Complexation........146
Further Reading........147

C........149

Calcium and Magnesium in Soils........149
Introduction........149
Input to Soils........149
Reactions in Soils........151
Modeling the Dissolution Reactions of Calcium and Magnesium Compounds in Soils........151
Plant Requirements and Deficiency Symptoms........152
Plant Uptake........153
Interactions with Other Nutrients........153
Animal Requirements........153
Further Reading........154

Capillarity........155
Introduction........155
Liquid Properties........155
Surface Tension........155
Contact Angle........155
Curved Surfaces and Capillarity........156
The Capillary Rise Model........157
Capillarity in Soils........158
Capillarity in Angular Pores........158
Dynamic Aspects of Capillarity........159
Dynamics of Capillary Rise........159
Dynamic Contact Angle........160
Heterogeneous Surfaces and Microscale Hysteresis........161
Contact Angle on Chemically Heterogeneous and Rough Surfaces........161
Hysteresis........163
Further Reading........163

Carbon Cycle in SoilsDynamics and Management........164
Dynamics and Management........164
Introduction........164
Fluxes........166
Soil-Forming Factors........166
Functions........166
Regulation of Soil Organic Carbon Dynamics........167
Composition of Soil Organic Carbon........168
Management........168
Further Reading........170
Formation and Decomposition........170
Introduction........170
Sources of Carbon........170
Aerobic Decomposition........171
Anaerobic Decomposition........172
Factors Controlling Decomposition........173
Soil Organic Matter Formation and Turnover........174
Further Reading........175

Carbon Emissions and Sequestration........175
Introduction........175
Soils and Carbon-Cycling........175
Controls on Emissions and Sequestration........176

Land Use and Land-Use Change 177
Management Practices for Soil C Sequestration 178
Regional and Country Estimates of C-Sequestration Potential 179
List of Technical Nomenclature 179
Further Reading 180
Cation Exchange 180
Introduction 180
Cation Exchange Capacity 181
Qualitative Description of Cation Exchange 182
Properties of Ions 182
Properties of Surfaces 183
Properties of Solvents 183
Quantitative Description of Cation Exchange 184
Homovalent Binary Exchange 184
Heterovalent Binary Exchange 185
The Future of Cation Exchange 187
List of Technical Nomenclature 188
Further Reading 188
Chemical Equilibria 189
Introduction 189
Chemical Equilibrium 189
Kinetic Definition 189
Free-Energy Definition 189
Applicability of Chemical Equilibrium Principles to Soils 190
Limitations to Equilibrium in Soils 190
Soil Systems in Which Equilibrium Concepts May Apply 191
Reasonable Expectations from Application of Equilibrium 191
Predict endpoints of reactions 191
Determine ionic activities in soil solution 191
Calculating final conditions for systems that attain equilibrium rapidly 191
Experimental Approach to Equilibrium in Soils 191
System Variables 191
Reactions and Associated Equilibrium Constants 191
Equilibrium Modeling of Soil Systems 192
Geochemical models 192
Model inputs 192
Application of Chemical Equilibrium to Soils: Examples 192
Iron Oxides in Reduced Systems 192
Manganese(III) Phosphate 193
Calcite in Soils: Equilibrium versus Nonequilibrium 193
The Outlook for Studying Chemical Equilibrium in Soils 193
List of Technical Nomenclature 194
Further Reading 194
Childs, Ernest Carr 195
Further Reading 198
Civilization, Role of Soils 199
Human Management of the Soil 199
Historical Attitudes Toward the Soil 199
Human Origins 200
The Paleolithic Transformation 201
The Agricultural Transformation 201
Soil Husbandry and Ceramics 202
From Rainfed to Irrigated Farming 203
Silt and Salt in Ancient Mesopotamia 203
The Sustainability of Egyptian Agriculture 203
Further Reading 204

Classification of Soils 204
Principle of Purpose 205
Principle of Domain 205
Principle of Identity 206
Principle of Differentiation 206
Principle of Prioritization 207
Principle of Diagnostics 209
Principle of Membership 209
Principle of Certainty 210
Summary 210
List of Technical Nomenclature 210
Further Reading 210
Classification Systems 211
Australian 211
Introduction 211
History of General-Purpose National Soil Classifications 211
The New National Classification 212
Soil Classifications Used on Australian State and Regional Scales 213
The Basis for State and Regional Soil-Classification Systems 213
State and Regional Examples 213
Special-Purpose Soil-Classification Systems 214
Viticultural Soils 214
Engineering Applications: Laying of Telecommunication Optical Fiber Cables 214
Future Directions 216
Further Reading 216
FAO 216
Introduction 216
The Development of the FAO Legend 1960-—1981 216
Set 1 218
Set 2 218
Set 3 218
Set 4 218
Set 5 219
Set 6 219
Set 7 219
Set 8 219
The FAO Revised Legend 1981-1990 219
The World Reference Base for Soil Resources 1981-1998 220
Summary 222
Further Reading 222
Russian, Background and Principles 223
Introduction 223
Principles, Categories, and Classes in Major Classification Systems Developed in Russia 223
Early Systems 223
NonOfficial Systems in the Second Half of the Twentieth Century 224
Official Systems 224
Classification and diagnostics of soils of the USSR - 1977 225
Russian soil classification system - 1997 225
Conclusion 227
Further Reading 227
Russian, Evolution and Examples 227
Introduction 227
Pre-Dokuchaev Period 228
Agroproductive Soil Groups, Folk Soil Nomenclature, and First Scientific

Soil Classifications 228
Dokuchaev's Period 229
Genetic Approach to Soil Classification 229
Post-Dokuchaev Period 230
Diversification of Classification Decisions 230
Official Soil Classification and Unofficial (Authors') Systems 232
The Problem of Basic Substantive-Genetic-Soil Classification and the new Russian Classification 232
Further Reading 235
USA 235
Introduction 235
Soil Classification Criteria 236
Soil Composition 236
Genesis and Soil Taxonomy 237
Climate Zones in Soil Taxonomy 238
Structure and Nomenclature 240
How it Works 240
Example Profile 242
Closing Thoughts 245
List of Technical Nomenclature 245
Further Reading 245
Clay Minerals 246
Introduction 246
Major Element Composition of Clay Minerals 246
Basic Structural Concepts 246
Tetrahedra and Octahedra 246
Representing Crystal Structures 247
Tetrahedral and Octahedral Sheets 247
Tetrahedral Sheet 247
Octahedral Sheet 247
Phyllosilicate Minerals Common In Soil Clays 248
1:1-Type Minerals 248
1:1 Layer structure 248
Kaolinite 248
Halloysite 248
2:1-Type Minerals 249
2:1 Layer structure 249
Pyrophyllite 249
Micas 249
Vermiculites 251
Smectites 251
Chlorites 251
Hydroxy-interlayered vermiculite and smectite 251
Interstratification in phyllosilicates 252
Palygorskite and sepiolite 252
Other Minerals that Occur in Soil Clays 252
Zeolites 253
Allophane and Imogolite 253
Aluminum Hydroxide Minerals 253
Iron Oxide Minerals 253
Manganese Oxide Minerals 253
Titanium Oxide Minerals 253
Carbonates, Sulfates, and Soluble Salts 253
List of Technical Nomenclature 254
Further Reading 254
Climate Change Impacts 254

Introduction 254
Climate Change Predictions 254
Timescale for Change 255
Climate Change Impacts on Soil Water and Soil Temperature 255
Soil Water 255
Soil Temperature 257
Changes in Soil-Forming Processes and Properties 257
Soil Organic Matter 257
Soil Structure 258
Soil Fauna and Soil Flora 258
Acidification and Nutrient Status 258
Changes in Land-Degradation Processes 259
Soil Erosion 259
Desertification 259
Salinization 260
Climate Change Impacts on Soil Functions 260
Further Reading 262
Climate Models, Role of Soil 262
Introduction 262
Approaches to Modeling SOC Dynamics for the Study of Climate Change 262
Process-Based, Multicompartment Models of SOC Dynamics 263
Cohort Models Describing Decomposition as a Continuum 263
Food-Web Models 263
Factors Affecting the Turnover of Soil Organic Carbon in Models 263
Model Performance 264
Models Used to Study the Impacts of Climate Change 264
Climate Models with Coupled Description of Biospheric Carbon Feedbacks 264
List of Technical Nomenclature 267
Further Reading 267
Cold-Region Soils 268
Introduction 268
Definition of Cold Soils 268
Characteristics and Genesis of Cold-Region Soils 269
Physical Properties 269
Boreal soils 269
Arctic soils 269
Alpine and high arctic soils 269
Morphological Properties 270
Biological and Chemical Properties 271
Alpine and Plateau Regions 272
Vegetation and Fire 273
Land Use of Cold-Region Soils 273
Engineering 274
Agriculture Development 274
Further Reading 275
Colloid-Facilitated Sorption and Transport 276
Introduction 276
Nature and Stability of Soil Colloids 276
Colloid Transport in Porous Media 279
Factors Influencing Colloid Release 280
Factors Influencing Colloid Deposition 281
Colloid-Facilitated Transport of Contaminants 283
Further Reading 284
Compaction 285
Introduction 285
Factors and Processes Affecting Distribution and Intensity of Compaction 285

Compaction under Running Gear 285
Loading Characteristics of Individual Wheels or Caterpillar Tracks 285
Field Traffic Intensity and Distribution 286
Soil Compactibility 287
Effects on Soil Physical and Mechanical Properties 287
Bulk Density, Porosity, and Packing State 287
Hydraulic Properties 287
Aeration Characteristics 288
Strength Characteristics 288
Compaction in Crop Production 288
Effects on Germination and Establishment 289
Effects on Root Growth and Distribution 289
Effects on Plant Growth and Yield 289
Interactive Crop Responses to Compaction and Fertilizer Application 289
Crop Responses to Subsoil Compaction 289
Modeling of Crop Responses to Soil Compaction 290
Effects on Environmental Components 290
Techniques for the Reduction of Compaction 292
Amelioration of Compacted Soils 293
Further Reading 293
Compost 294
Introduction 294
Compost Production 294
Compost Application and Incorporation 295
Compost-Soil Interactions 296
Physical Effects 296
Chemical Effects 297
Biological Effects 299
List of Technical Nomenclature 300
Further Reading 300
Conditioners 301
Introduction 301
Early Use of Mineral and Organic Materials 301
Use of Waste Materials as Conditioners 302
Advent of Synthetic Conditioners 302
Hydrogels and Super Water-Absorbent Polymers 303
Recent Advances Using Polyacrylamide 304
Further Reading 305
Conservation Tillage 306
Introduction 306
Evolution of Conservation Tillage Systems 306
Defining Conservation Tillage 307
Variants of Conservation Tillage 307
Common Types of Conservation Tillage Practices 308
Constraints to the Adoption of Conservation Tillage Practices 308
Climate and Soil 308
High Levels of Crop Residue 308
Mixed Cropping Systems 309
Beneficial Influence of Conservation Tillage 309
Improving Soil Quality and Health 309
Enhancing Soil Organic Matter Storage 309
Reducing Soil Erosion and Environmental Risk 310
Tillage Costs 310
Further Reading 310
Cover Crops 311
Introduction 311

Agronomy ...312
Species Selection...312
Seeding and Husbandry...313
Role in Soil Quality Improvement ...314
Role in Soil Erosion Control...315
Nature of Soil Erosion ...315
Runoff and Erosion Control...316
Perspective...316
Further Reading...317
Crop Rotations ...318
Introduction...318
Soil Fertility and Crop Yields...319
Cultural Management ...320
Pest Management with Rotations...320
Economic Diversity ...321
Future Perspectives...321
Further Reading...322
Crop Water Requirements...322
Introduction...322
Crop Evapotranspiration...323
Crop Coefficients...324
The Single-Crop Coefficient Approach...325
Kc Adjustment for Climate...330
Kc Adjustment for Nonpristine Conditions...330
The Dual-Crop Coefficient Approach ...330
Kc for Nonpristine and Unknown Conditions...332
Irrigation Water Requirements...332
Summary ...332
List of Technical Nomenclature ...333
Further Reading...334
Crop-Residue Management...334
Introduction...334
Conservation, Carbon Cycle, Soil Organic Matter and Carbon Sinks ...334
Crop Residues and Nutrient Cycling...335
Factors Controlling Residue Decomposition and Soil Quality ...336
Crop Residues: Social and Environmental Benefits ...336
Crop Residues, Research and Global Change...337
Summary ...337
Further reading ...338
Crusts...339
Biological ...339
Introduction ...339
Species Composition and Growth Forms...339
Ecological Roles...341
Carbon Fixation ...341
Nitrogen Fixation...341
Dust Trapping ...343
Effects on Vascular Plants...343
Nutrient Levels in Vascular Plants...343
Soil Hydrology and Stabilization...343
Effects of Disturbance ...344
Species Composition...344
Water Erosion...344
Wind Erosion ...344
Nutrient Cycles ...345
Albedo...345

Recovery from Disturbance 345
Natural Recovery Rates 345
Assisted Recovery 345
Conclusion 345
Further Reading 347
Structural 347
Introduction 347
Crust Types and Morphology 348
Factors Governing Crust Formation 349
Rainstorm Characteristics 349
Soil Properties 351
Hydrologic Impact of Crusts 351
Infiltration - Developed Crusts 352
Infiltration - Developing Crusts 352
Numerical Infiltration Models for Developing Crusts 353
Agronomic Importance of Crusts 354
Further Reading 355
Cultivation and Tillage 356
Introduction 356
Goals of Soil Tillage 356
Main Types of Tillage Tools 357
Moldboard Plow 357
Applications 357
Disadvantages and limitations 357
Chisel Plow 358
Applications 358
Disk Harrow 358
Toothed Cultivator 359
Subsoiler 359
No-tillage 359
Basic Tillage Operations 360
Types of Tillage Systems 360
List of Technical Nomenclature 361
Further Reading 361
D 363
Darcy’s Law 363
Introduction 363
Water-Saturated Soil Conditions 363
Validity of Darcy’s Equation 364
Solving Saturated-Flow Problems 366
Water in Unsaturated Soils 366
Solving Unsaturated-Flow Problems 368
List of Technical Nomenclature 368
Further Reading 369
Degradation 370
Introduction 370
Factors and Processes Affecting Degradation of Soils 370
Types of Soil Degradation 371
Assessment of Degradation 371
Approaches 371
Potentials and Limitations 373
Degree and Impact of Degradation 374
Degree of Degradation 374
Rate of Soil Degradation 375
Impact of Degradation 375
Preventing and Combating Degradation 375

Further Reading ... 377

Denitrification ... 378

Introduction ... 378

Definitions and Pathways ... 378

Organisms and Substrates ... 378

Denitrification ... 378

Nitrifier Denitrification ... 379

Environmental Factors ... 380

Organic C Availability ... 380

Controls on O2 Availability ... 380

Nitrate Supply ... 381

Temperature ... 381

pH ... 381

Conclusion ... 381

Further Reading ... 382

Desertification ... 382

Introduction ... 382

The Role of Drought ... 383

Primary Production and Carrying Capacity ... 384

Soil Degradation and Rehabilitation ... 385

Potentialities and Problems of Irrigation ... 385

Social Factors ... 386

Monitoring Desertification ... 386

The Role of Climate Change ... 387

Overview ... 387

Further Reading ... 388

Diffusion ... 389

Introduction ... 389

How Diffusion Occurs ... 389

Fick’s Laws ... 390

Initial and Boundary Conditions ... 390

Components of the Diffusion Coefficient in Soil ... 390

Diffusion and Geometry ... 391

Some Useful Analytical Solutions for Various Geometries ... 391

Diffusion from a Semiinfinite Medium ... 391

Diffusion in a Cylinder ... 392

Diffusion in a Sphere ... 392

Simple Model for Various Geometries ... 392

Applications in Soil Science ... 393

Plant Nutrition ... 393

Kinetics of Potassium Release From Clays ... 393

Gaseous Diffusion ... 393

Leaching ... 393

List of Technical Nomenclature ... 393

Further Reading ... 394

Disinfestation ... 394

Introduction ... 394

Principles of Soil Disinfestation ... 395

Methods of Soil Disinfestation ... 395

Physical ... 395

Chemical ... 396

Methyl bromide ... 396

Metham sodium ... 396

Dazomet (3,5-dimethyltetrahydro-1,3,5,(2H)-thiodiazinothione) ... 396

Other fumigants and mixtures ... 396

Application ... 396

Soil Solarization 397
Principles of soil solarization...397
Soil heating: simulation models for the prediction of soil temperatures..........398
Combining Disinfestation Methods 398
Benefits and Limitations..399
The MBr Crisis and Its Implications 399
Further Reading...399
Dissolution Processes, Kinetics..400
Introduction...400
Rates and Limits of Element-Cycling...401
Kinetics of Proton- and Ligand-Promoted Mineral Dissolution 401
Metal-Promoted Mineral Dissolution..403
Reductive Dissolution of Minerals..404
Dissolution of Metal Surface Precipitates 404
The Good and Bad of Dissolution Kinetics 407
List of Technical Nomenclature 408
Further Reading...409
Drainage, Surface and Subsurface...409
Introduction...409
Hydrology and Drainage...410
Accelerated Drainage 410
Surface Drainage Principles and Practices...410
Subsurface Drainage Principles and Practices 411
Recent Innovations in Drainage Practice...412
Further Reading...413
Dryland Farming...414
Introduction...414
Definition...414
World Scope of Dryland Farming..414
Major Dryland Farming Issues..415
Dryland Farming Management Techniques...415
Fallowing 415
Water Capture Management..416
Water Retention Management 416
Efficient Water Use by Crop Plants..417
Conclusion..417
Further Reading...417
E 419
Edaphology...419
Introduction...419
Edaphic Properties 419
Edaphic Factors Affecting Water and Air 420
Edaphic Factors Affecting Plant Nutrients..421
Edaphic Factors Affecting Root Growth 423
Research 424
List of Technical Nomenclature 424
Further Reading...425
Electron-Spin Resonance Spectroscopy..426
Introduction...426
Basic Principles 426
The Resonance Phenomenon 426
Spectral Parameters..426
Instrumentation and Methodology 427
The Instrument and the Experiment...427
Sample Preparation 428
Sensitivity and Resolution 429

Determination and Interpretation of Spectral Parameters........................429
Whole Soil..429
Organic Free Radicals in Soil Humic Substances 430
ESR Spectra and Parameters..430
Interpretation of ESR Parameters 431
Factors Affecting ESR Parameters of Humic Substances 431
Structural Implications..431
Interactions of Humic Free Radicals with Organic Chemicals and Metal Ions 432
Paramagnetic Transition Metal Ions in Soil Constituents 432
ESR Spectra and ESR Parameters...432
Ferric iron...433
Divalent copper..434
Vanadyl ion 435
Divalent manganese 435
Spin-Derivatization Studies...435
Spin-Labeling and Spin-Trapping...435
Metal Spin Probes..435
Complexation chemistry of humic substances...435
Ion exchange on layer silicates...436
Chemisorption on mineral surfaces 436
Thermodynamic constants from ESR parameters...436
List of Technical Nomenclature 436
Further Reading..437
Energy Balance 438
Introduction 438
The Energy Input From the Sun 438
Energy Balance of a Planet Without Atmosphere..438
Energy Balance of the Earth’s Surface...438
Anthropogenic Change of the Energy Balance..441
Further Reading...441
Environmental Monitoring...441
Introduction 441
Why Monitor Soils?..442
The Starting Point 443
Soil-Monitoring Networks..444
Protocols...444
Monitoring Sites 444
Soil Sampling..446
Quality Control 447
Summary 447
Further Reading..447
Enzymes in Soils...448
Introduction 448
Spatial Distribution of Enzymes in Soils 449
Extracellular Enzyme Stabilization in Soil Matrix.....................................449
Microscale Distribution 449
Macroscale Distribution..450
Methods of Studying Enzyme Activities in Soils 450
Enzyme Kinetics...451
Soil-Quality Indicators and Technologies 452
Sensitivity to Soil Management and Ecosystem Stress 452
Ecological Dose Value 454
Detoxification of Polluted Soils..454
Summary 455
Further Reading...455

Erosion 456
Irrigation-Induced 456
Introduction 456
Significance 457
Erosion Under Sprinkler Irrigation 458
Processes Causing Soil Loss 458
Practices Controlling Soil Loss 459
Irrigation practices 459
Soil and crop management practices 459
Erosion with Surface Irrigation 460
Processes Causing Soil Loss 460
Practices Controlling Soil Loss 460
Irrigation practices 460
Runoff management practices 461
Soil and crop-management practices 462
Summary 462
Prospects for Future Control 463
Further Reading 463
Water-Induced 463
Introduction 463
Impacts of Erosion 464
Types of Erosion 464
Rill Erosion 464
Interrill Erosion 464
Gully Erosion 464
Ephemeral Gully Erosion 465
Stream Channel Erosion 465
Factors Affecting Erosion 465
Rainfall Characteristics 465
Soil Factors 465
Topography 465
Climate 466
Land Use 466
Cropland areas 466
Rangelands and pastures 466
Forest areas 466
Construction sites 467
Erosion Control Measures 467
Erosion Control Measures on Agricultural Areas 467
Contouring 467
Strip cropping 467
Conservation tillage 467
Terraces 467
Buffer strips 468
Grassed waterways 468
Polyacrylamide 469
Erosion Control Measures on Forest Areas 469
Erosion Control Measures on Construction Sites 469
Erosion Control Measures in Stream Channels 469
Further Reading 469
Wind-Induced 470
Introduction 470
Wind Profile 470
Modes of Transport 470
Soil Surface Conditions 471
Vegetation 473

Wind Erosion Modeling 473
Wind Erosion Sampling 474
Use of Anthropogenic Radioisotope Tracers 475
Effects of Wind Erosion 475
Wind Erosion Control 476
List of Technical Nomenclature 477
Further Reading 477

Essential Elements 478

Introduction 478
Classification of Essential Elements 478
Nutrient Deficiency Symptoms 478
Nitrogen 479
Sulfur 480
Phosphorus 480
Potassium 481
Calcium 481
Magnesium 482
Iron 482
Manganese 483
Zinc 483
Copper 483
Molybdenum 484
Boron 484
Chlorine 485
Nickel 485
Further Reading 485

Eutrophication 486

Introduction 486
Aquatic Eutrophication 486
Causes 486
Limiting Nutrients 486
Consequences of Aquatic Eutrophication 489
Stratification and Oxygen Depletion 489
Changes in Species Composition 489
Sea-Grass Destruction 490
Degradation of Corals 491
Modeling Eutrophication 491
Phosphorus-Limited Freshwater Systems 491
Nitrogen-Limited Coastal Marine Systems 491
Terrestrial Eutrophication 491
Watershed Export and Models 492
Controlling Eutrophication: Future Trends 493
List of Technical Nomenclature 494
Further Reading 494

Evaporation of Water from Bare Soil 494

Introduction 494
Evaporation Rate and Cumulative Evaporation 494
Processes Involved in Evaporation, and Latent Heat of Evaporation 495
Net Radiation 495
Energy Balance 495
Stages of Evaporation 496
Evaporation under Conditions of Unlimited Potential Evaporation 497
Evaporation Below the Soil Surface 497
Methods for Measuring Evaporation 498
Weigh the Loss of Water Mass in the Soil 498
Detect the Arrival of Water Vapor Within the Air 499

Quantify the Upward Transport of Water Vapor in the Air........................499
Measure Enough Elements in the Energy Balance to Calculate LeE 499
Compare the Soil of Interest to a Completely Dry Soil 500
List of Technical Nomenclature 501
Further Reading..502
Evapotranspiration 502
Introduction...502
Development of the Concept 502
Methods of Measurement 503
Magnitude of Evapotranspiration...503
Control of Evapotranspiration 504
Climate Change and Evapotranspiration 505
Further Reading..506
F 507
Factors of Soil Formation..507
Biota..507
Introduction 507
The Biosequence 507
The Biotic Factor 508
Vegetation as a Soil-Forming Factor...508
Grassland-Forest Biosequence 510
Conifer Forest-Deciduous Forest Biosequence...511
List of Technical Nomenclature..512
Further Reading 512
Climate..512
Introduction 512
Soils in Tropical Climates..514
Soils in Dry Climates 515
Soils in Warm Temperate Climates...516
Soils in Cool Temperate Climates 517
Soils in the Polar Climates 518
Soils as Climatological Archive 518
Summary...519
Further Reading 519
Human Impacts...520
Introduction 520
Scales and Scope of Human Impacts on Soil Formation.........................520
Evaluating Human Impacts on Soil Formation 522
Changes in Soil Properties and Processes Resulting From Human Activities 523
Agriculture...523
Cities and Industry 525
Mining 527
War 529
Climate Change and Soil Formation 529
Significance and Future of Human Impacts on Soil Formation 530
List of Technical Nomenclature..531
Further Reading 531
Parent Material 532
Introduction 532
Defining Soil Parent Material and Role in Soil Genesis 532
Soil-Forming Factors..532
Weathering..533
Effects of Parent Material on Soil Properties..534
Influence of Parent Materials in Soil Genesis...534
Parent Material and Soil Classification...535

Summary 535
List of Technical Nomenclature 535
Further Reading 535
Time 536
Introduction 536
Key Soil Properties that Vary with Time 536
Studies of Time and Soil Formation 538
List of Technical Nomenclature 538
Further Reading 539
Fauna 539
Introduction 539
Characterization of Soil Macrofauna 539
Saprovores and Microbivores 539
Herbivores and Granivores 541
Predators and Omnivores 541
The Role of Soil Macrofauna in Ecosystem Function 542
Decomposition and Nutrient Cycling 542
Soil Modification 544
Participation in Belowground and Aboveground Food Webs 548
Further Reading 548

A

ACID RAIN AND SOIL ACIDIFICATION

L Blake, Rothamsted Research, Harpenden, UK

Introduction

During the last three decades of the twentieth century, several studies in northern, central, and western Europe and in North America demonstrated, almost invariably, increasing acidity in soils beneath natural and seminatural ecosystems. Several ecosystem and soil processes may decrease soil pH, particularly changes in land use or soil management; thus the interpretation of acidifying sources can be difficult and complex. However, long-term studies of changes in deposition and soil chemistry have shown that inputs of atmospherically derived acidity cause soils to acidify rapidly, depending on the intensity of the acid input and the acid-buffering capacity of the soil. This acidification process is characterized by specific changes in soil and soil-solution chemistry.

Acid Rain

Rain is naturally acidic, usually in the pH range 5.0–5.5, mainly because of dissolved atmospheric carbon dioxide (CO_2) forming carbonic acid (H_2CO_3):

$$CO_2 \rightarrow CO_2 \cdot H_2O \rightarrow HCO_3^- + H^+ \qquad [1]$$

This 2.5×10^{-5} H^+ represents only a minute quantity of acid and is probably beneficial, even essential, to such processes as the weathering of soil from rocks and the release of insoluble nutrients from soil minerals. Only when the concentration of protons is increased by factors of 10–100 over 'natural' pH values by the presence of stronger acids do the detrimental effects of 'acid rain' occur.

The precursors of stronger acids in the atmosphere are pollutants attributable to (1) the combustion of fossil fuels and (2) the smelting of sulfide ores. However, these sources can be significantly augmented by ammonia, resulting from the management practices of intensive agriculture, particularly from livestock wastes.

There are essentially two kinds of pollutants of concern, gaseous and particulate, each of which can be subdivided by their modes of formation: primary pollutants are produced directly from industrial and domestic activities, and secondary pollutants are created in the atmosphere by chemical processes acting on primary pollutants. These distinctions enable four groups of pollutant to be identified:

1. Gaseous, primary:
   Sulfur dioxide (SO_2);
   Nitric oxide (NO) and nitrogen dioxide (NO_2); (collectively designated NO_x);
   Nitrous oxide (N_2O);
   Ammonia (NH_3);
   Carbon dioxide (CO_2);
   Hydrocarbons;
2. Gaseous, secondary:
   NO_2 from oxidation of NO;
   Ozone (O_3) and other photochemical oxidants formed in the lower atmosphere by the action of sunlight on mixtures of NO_x and hydrocarbons;
   Nitric acid (HNO_3) formed from the oxidation of NO_x;
3. Particulate, primary: fuel ash and metallic particles, from smelting and heavy industry;
4. Particulate, secondary: the reaction products of sulfuric acid and nitric acid with other atmospheric constituents, notably NH_3 (NH_4HSO_4, NH_4NO_3, etc.). H_2SO_4 and HNO_3 formed by the oxidation of SO_2 and NO_x, respectively.

The primary source of oxidized sulfur from human activity is the burning of coal. Much of the world's coal used for energy purposes contains 2% S, half of which is present as pyrite (Fe_2S) and the remainder is organic. SO_2 is produced readily on burning:

$$4FeS_2 + 11O_2 \rightarrow 2Fe_2O_3 + 8SO_2 \qquad [2]$$

NO and NO_2 enter the atmosphere from natural and pollutant sources, such as the burning of fossil fuels in motor vehicles or stationary furnaces. The formation of NO from the reaction of N_2 and O_2 occurs at high temperatures:

$$N_2 + O_2 \rightarrow 2NO \qquad [3]$$

Once NO has been formed, rapid cooling of exhaust gases prevents further reaction and traps the oxide in the atmosphere. NO is also formed naturally in the atmosphere through reaction of O_2 and N_2 caused by lightning.

N_2O is formed mainly from biological agents and is an essential intermediate in the N cycle in soils; for example, in the process of denitrification:

$$NO_3 \rightarrow NO_2 \rightarrow NO \rightarrow N_2O \rightarrow N_2(g) \qquad [4]$$

This process is often observed in organic-rich sediments and flooded soils and is facilitated by microorganisms growing anaerobically. Significant emissions of nitrous oxide also arise in aerobic or partially aerobic soils from the rapid oxidation of organic matter or ammonium fertilizers by the process of nitrification:

$$\begin{array}{l} NH_4^+ \rightarrow NH_2OH \rightarrow (NOH) \rightarrow NO_2 \rightarrow NO_3^- \\ \qquad\qquad \downarrow \\ \qquad\qquad N_2O \end{array} \qquad [5]$$

A very important source of acidification in ecosystems is the production of protons from NH_4^+ ions by the nitrification process:

$$NH_4^+ + 2O_2 \rightarrow NO_3 + H_2O + 2H^+ \qquad [6]$$

Impacts from atmospheric NH_3 and ammonium have been shown to be especially important in forest ecosystems, particularly in the Netherlands, where almost 50% of the acidic deposition at the ‘edges’ of Dutch forests is considered to be derived from volatilized ammonia. The ammonium ion is a conspicuous constituent of areas suffering impacts from acidic deposition due to the rapid nitrification (oxidation) of ammonium salts; sulfate, nitrate, and bicarbonate are typical counterions:

$$(NH_4)_2SO_4 + 4O_2 \rightarrow 2NO_3^- + 4H^+ + 2H_2O + SO_3^{2-} \qquad [7]$$

Atmospheric Transport, Secondary Chemistry, and Acid Rain

The concentrations of primary pollutants and secondary pollutants determine the nature and pattern of atmospheric deposition over a region. These concentrations are related to the chemical processes that occur simultaneously with transport and dispersion through the atmosphere. Emitted primary pollutants such as SO_2 and NO_x, which have not already been deposited directly by dry deposition, are oxidized to secondary pollutants and converted to acidic, water-soluble aerosols. For example, SO_2 is oxidized to H_2SO_4 and NO_2 to HNO_3.

Atmospheric cycling of these pollutants is a complex process (Figure 1). Atmospheric oxidation is mediated by chains of free-radical reactions. The oxidation proceeds in a system of very dilute reactants using an external energy source provided by solar radiation. Within the troposphere, the most important reactions – gas-phase reactions – are those initiated by the hydroxyl radical, OH. This is known to react with many pollutants and it controls the tropospheric concentrations of many of them. The OH radical is often regarded as the natural cleaning agent of the atmosphere. In turn, trace gases influence atmospheric concentrations of OH. The two other key oxidizing species in the production of atmospheric acidity are ozone (O_3) and hydrogen peroxide (H_2O_2).

Dry oxidation of sulfur and nitrogen oxides In the absence of clouds, the major route for the production of H_2SO_4 from SO_2 is via reaction with OH radicals in the gas phase:

$$OH + SO_2 \rightarrow HSO_3 \qquad [8]$$

$$HSO_3 + O_2 \rightarrow HO_2 + SO_3 \qquad [9]$$

$$SO_3 + H_2O \rightarrow H_2SO_4 \qquad [10]$$

In daylight, the major gas-phase route for the production of HNO_3 is via reaction of NO_2 with OH:

$$OH + NO_2 \rightarrow HNO_3 \qquad [11]$$

At night a further route is available via the formation of the nitrate radical, which is photolytically unstable in sunlight:

$$NO_2 + O_3 \rightarrow NO_3 + O_2 \qquad [12]$$

$$NO_2 + NO_3 \rightarrow N_2O_5 \qquad [13]$$

$$N_2O_5 + H_2O \rightarrow 2HNO_3 \qquad [14]$$

Wet oxidation of sulfur SO_2 dissolves in atmospheric water droplets to form the bisulfite ion (HSO_3^-):

$$SO_2 \rightarrow SO_2 \cdot H_2O \rightarrow HSO_3^- + H^+ \qquad [15]$$

Oxidation of HSO_3^- yields a sulfate ion (SO_4^{2-}) and a further proton:

$$HSO_3^- + \text{Oxidant} \rightarrow SO_4^{2-} + H^+ \qquad [16]$$

The major oxidants for the production of dissociated SO_4^{2-} and H^+ ions are H_2O_2 and O_3. The H^+ ions are deposited as wet-deposited acidity.

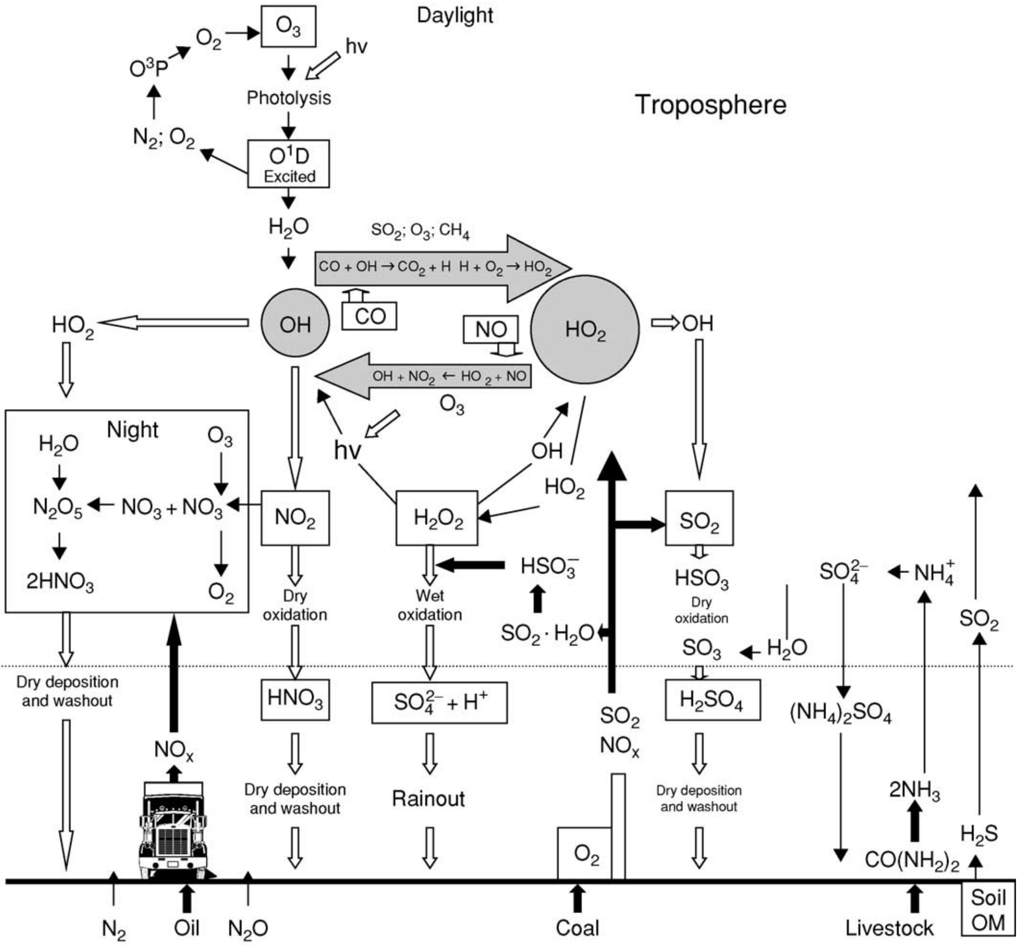


Figure 1 Summary of the processes governing acid deposition. The recycling of the hydroxyl radical is outlined in gray. The dotted line represents the boundary layer. OM, organic matter; hV, UV protons from solar radiation; O^3P, ground state atomic oxygen.

Deposition of both 'wet' and 'dry' products of S and N reactions occurs. Dry deposition is the direct transfer of gases and particulates from the atmosphere. The primary gaseous species of S, SO_2, may be deposited directly from the pollution source, as may the N products from the dry oxidation stages of the secondary reactions, NO_2 and HNO_3 (most N_2O and NO emissions are oxidized). Gaseous species also directly deposited are NH_3, HCl, and O_3.

Dry deposition of particles and gases occurs by sedimentation and surface adsorption; it has greatest significance close to pollutant areas. Dry-deposited species themselves react with surface moisture (in soil or on plant foliage) to produce acidity by the dissociation of H^+ ions and mobile conjugate anions, although some H^+ tends to be neutralized by NH_3, resulting in the formation of NH_4^+ ions.

Wet deposition – the precipitation of scavenged pollutants from the atmosphere – is the dominant removal mechanism remote from the pollutant source. Wet deposition involves the removal of primary SO_2 by reaction with atmospheric water and oxidation with H_2O_2 or O_3 and the removal of secondary NO_2, as HNO_3, by reaction with the hydroxyl radical. Wet deposition is characterized by two processes: (1) rainout – whereby atmospheric species are associated with cloud phenomena; (2) washout – where species are removed by falling precipitation. Figure 2 shows the long-term trend in the deposition of acidity (H^+) and the precursors of acidity in southern England since 1853.

Acid Soil

Soil acidification is defined as a decrease in acid-neutralizing capacity (ANC) or an increase in base-neutralizing capacity (BNC), resulting in an increase in acid strength as represented by a decrease in soil pH:

$$\text{Soil acidity}(+\Delta\text{BNC}) = -(\text{soil alkalinity}) = -\Delta\text{ANC} \quad [17]$$

Soil acidification processes in aerated soils are a consequence of: (1) production of various acids in the

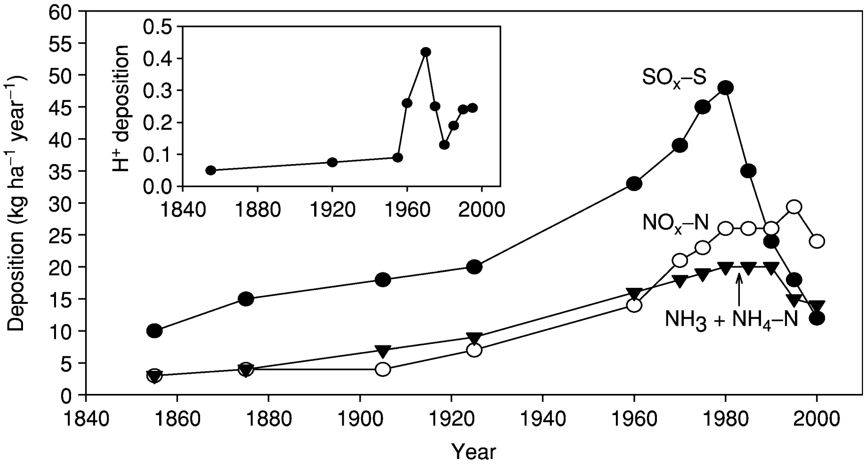


Figure 2 Wet-deposited acidity (H$^+$; the truly acid rain) and the estimated total deposition (wet plus dry deposition) of the precursors of acidity at Rothamsted Research in southern England, UK. Wet deposition has been measured at Rothamsted since 1853. Dry deposition of NO_3^- and NH_4^+ measured since 1990, and of SO_4^{2-} since 1980, calculated in other years from the ratio of wet/dry since 1990.

soil, (2) the effect of ion uptake by biota, (3) nitrogen inputs and transformations, and (4) the addition of dissolved strong acids (acid deposition).

Sources of Acids in Soils

Several ecosystem and soil processes decrease soil pH through internal acid production. These processes can also be directly or indirectly influenced by atmospheric deposition.

Substances which react in soils as proton donors (acidity) are as follows:

1. In the solid phase:
   - Cations forming weak hydroxides (mainly Al, Fe, and Mn, either as exchangeable cations or bound within clay minerals or associated with organic matter – M$_a$ cations). For Al:

$$Al^{3+} + H_2O \leftrightarrow AlOH^{2+} + H^+,$$
$$Al^{3+} + 2H_2O \leftrightarrow Al(OH)^{2+} + 2H^+ \qquad [18]$$

   - Mineralization of organically bound N followed by nitrification:

$$R{-}C{-}NH_2 \rightarrow NH_3 + H^+ \rightarrow NH_4^+$$
$$+ 2O_2 \rightarrow NO_3 + H_2O + 2H^+ \qquad [19]$$

   - Undissociated acidic groups on (1) clay minerals (pH-dependent charge) and (2) organic matter (R-OOH);
   - Aluminum hydroxysulfates and sulfate sorbed on aluminum hydroxides;
   - Exchangeable NH_4^+ (nitrification or assimilation of the ammonium cation);
   - Organically bound S (organic S → H_2SO_4);
2. In solution:
   - $CO_2 \cdot H_2O \rightarrow HCO_3^- + H^+$(carbonic acid); [20]
   - NH_4^+: ($NH_4^+ \leftrightarrow NH_3 + H^+$; $NH_4^+ \rightarrow$ organic N + H$^+$; assimilation); [21]
   - cations forming weak acids (as for the first item in the solid-phase list);
   - Organic acids (R-COOH):

$$CO_2 + R{-}CH_2OH \rightarrow RCOO^- + H_2O + H^+ \qquad [22]$$

Internal Acid Production in Soils

In order to isolate the specific impact that acid deposition has on soil pH, the effects of internal acid production in soils need to be assessed.

Carbonic and organic acid production The production of weak acids such as $CO_2 \cdot H_2O$ or dissolved organic acids by plants and by decomposition of organic substances does not result in soil acidification, because the weathering of silicates releases alkali and earth alkali (M$_b$) cations into solution. However, ANC decreases if M$_b$ cations are removed from the soil by drainage. Thus, dissolution of Na, Ca, and Mg from soil minerals by $CO_2 \cdot H_2O$, followed by leaching with HCO_3^-, is the dominant soil-acidification process in nature (pH >5).

Stronger organic acids replace CO$_2$ as the acidifying agent in acidic soils of pH 4–5 (eqn [22]) and are the dominant acidifying agent in such acid soils under natural conditions. Low-molecular-weight species of humic acids as well as aliphatic acids are produced in the A horizon of soils and are transported downward toward the B horizon. Humic acids contain phenolic and carboxylic groups, the latter readily binding metal ions such as Al and Fe. The metal ions link the organic molecules, producing larger molecules with lower solubility, and release protons to the soil solution of the B horizon. The production of organic acids results largely from microbial activity in the A horizon; chemical precipitation reactions take place in the B horizon. As a result, this wholly natural process of acidification is restricted to the A and B horizons of soils beneath typically podzolizing plant communities such as heather or spruce stands. Podzolization cannot explain the acidification of soils below the B horizon of freely draining soils.

Nitrogen accumulation and transformations The accumulation of organic N in soils is potentially a large proton source for acidity because of nitrification following mineralization. Changing conditions of organic matter decomposition and humus formation strongly influence this powerful proton source. Thus, changes to ecosystems which influence the turnover on organic matter can result in considerable internal proton production. These are largely the effect of anthropogenic changes in ecosystem management such as increased rates of agricultural or forest production, and also to atmospheric N enrichment. The capacity for proton production by this process is only limited by the amount of organic N accumulated in the mineral soil during ecosystem development, but is almost entirely restricted to the surface soil horizons (rooting zone) where this accumulation occurs.

Assimilation of Nutrients by Vegetation

In ecosystems where efficient N cycling occurs, N species are unimportant in terms of soil acidification. In such ecosystems, however, plants take up more basic components (M_b cations) than acidifying components. To maintain electroneutrality, plants excrete protons as counterions to cation uptake. As a consequence, ecosystems that increase in biomass acidify the soil. For natural ecosystems in steady state, with no export of cations in biomass, mineralization equals assimilation and the soil does not acidify by assimilation of nutrient cations. However, the export of biomass (e.g., timber) results in the removal of bases stored in the vegetation, leading to a decline in ANC and further soil acidification.

Deposition of Nitrogen Species and Nitrogen Transformations

Atmospheric inputs of N species influence soil acidification either (1) directly through the deposition of acid nitrogen compounds or (2) indirectly through biological N transfers within the ecosystem. N transfers cause proton fluxes if there is a net input or output of NH_4^+ or NO_3^- from the soil. These fluxes result from deposition-derived NH_4^+ or NO_3^- and uptake by vegetation (which causes an equivalent release of H^+ or HCO_3^- from the roots, respectively) or nitrification of NH_4^+ (eqn [19]) followed by leaching of NO_3^-. For soils with high inputs of NH_4^+ from the atmosphere, this can be the dominant form of acidity.

Addition of Dissolved Strong Acids Through Acid Deposition

HNO_3 and H_2SO_4 are the major components of directly derived atmospheric acidity. After their addition ANC is decreased rapidly when basic cations (M_b) released from soil minerals by strong acid weathering are leached with sulfate or nitrate in drainage. The relative contribution of HNO_3 and H_2SO_4 to acidification can be assessed from comparison of the input/output (deposition/seepage) balance of SO_4-S/Cl and NO_3-N/Cl (chloride is not bound in soil, so output equals input). ANC is also decreased if H_2SO_4 is retained in the soil, by sulfate adsorption or by precipitation of sulfates of Fe and Al.

Observations of Changes in Soil Acidification due to Acid Deposition

To assess the causes and effects of acid deposition on ecosystems involves measuring changes determined by soil measurement over a period of time, or making measurements on soils that have been sampled chronologically and stored. Relatively little information on the analysis of long-term data is available, but numerous measurements based on lysimeter studies have been made since the 1970s.

There are four prime sources that show the results of reexamining soils in polluted regions and the long-term (more than 30 years) changes involved:

1. Eastern USA: A reexamination of soils in the Adirondack Mountains, New York, after a lapse of 50 years, revealed that the pH of organic horizons initially at pH 4.0 (Al buffer) remained unchanged, but the pH of originally less acid horizons had declined by 0.3–0.5. Acidic deposition and natural leaching were considered to be the major contributors to these pH changes;
2. Sweden: Long-term pH changes in southern Sweden were documented by repeating studies from 1927 on the Tonnersjoheden experimental area. Two trends were detected: (a) the pH of the humus layer (O horizon) and others decreased as stands of Norway spruce increased in age; and (b) there was a significant pH decrease in all soil horizons, including the deepest C horizon, where pH was independent of stand age. The decreases ranged between pH 0.31 and 0.65. A combination of acid rain and biological acidification was the cause of the changes;

3. Scotland: Soils were resampled in 1987 from 15 sites in Alltcailleach Forest, first sampled in 1949–1950. The surface horizons at 12 sites had decreased by 0.07 units to pH 1.28, while decreases of 0.16–0.54 units were found in the deeper mineral horizons. All of these decreases were also associated with decreased base saturation and markedly increased amounts of extractable Al;

4. England: The most detailed study of long-term (1876–1991) changes in soil chemistry in response to acidic deposition used archived soil samples taken from woodland and grassland at Rothamsted Research in southern England. The changes were assessed in light of the measured increases in acid deposition at this location, as shown in Figure 2, and are used here to illustrate the long-term effects of acid deposition and soil acidification.

Proton Sinks and Buffer Ranges

The major changes in the chemistry of soils subject to acidification result from proton-buffering processes. Substances which react in soils as proton acceptors (basicity) are responsible for soil ANC. In the solid phase these are: (1) carbonates, (2) silicates, and (3) alkali and earth alkali (M_b) cations which are bound or exchangeable on weakly acidic groups of soil minerals and organic matter.

In the solution phase, OH^- and $R\text{-}COO^-$ are usually at too low a concentration to be of importance. HCO_3^- is the dominant base, but only at pH >5:

$$HCO_3^- + H^+ \rightarrow CO_2 + H_2O \qquad [23]$$

The soil phases accepting protons release M_a and M_b cations through dissolution reactions, which change with increasing acid strength. This leads to a sequence of buffer reactions as pH decreases. In addition to dissolution reactions, proton adsorption also takes place on negatively charged surfaces associated with clay minerals and organic matter. These processes vary according to the composition of the buffer substances and the reaction products. Figure 3 shows how proton sinks have influenced soil pH in silty clay loam soils at Rothamsted and illustrates two important aspects of soil acidification. First, the pH value does not change in a simple, constant manner with the quantity of H^+ input. This is due to stronger

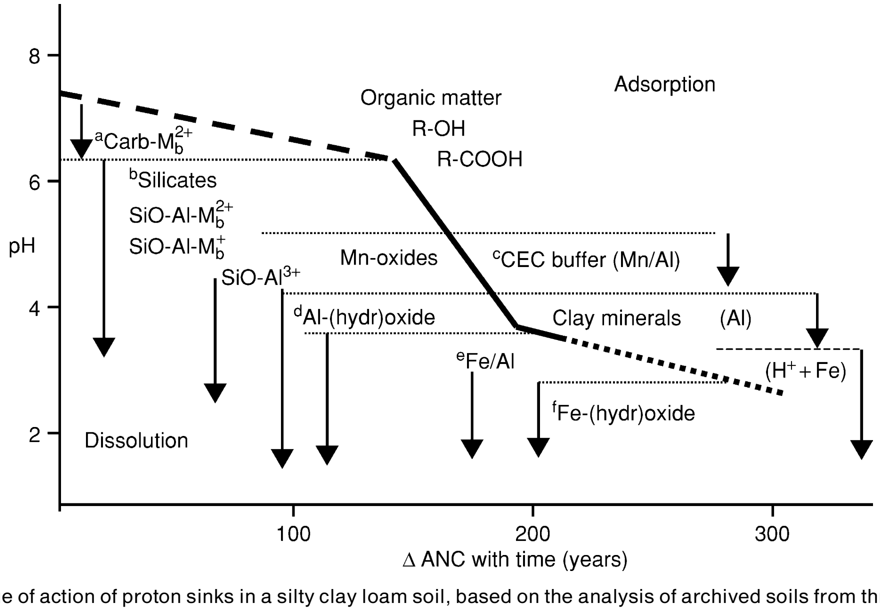


Figure 3 pH range of action of proton sinks in a silty clay loam soil, based on the analysis of archived soils from three long-term (more than 100 years) field experiments receiving acidifying inputs at Rothamsted, UK: thick dashed line, woodland heavily limed in 1881 subject to acid deposition; thick continuous line, unlimed woodland subject to acid deposition; thick dotted line, grassland strongly acidified by annual applications of ammonium sulfate (144 kg N ha^{-1} per year; since 1856). Reaction products: a, carbonate buffer: $Ca(HCO_3)_2$ in solution, leaching of Ca and basicity; b, silicate buffer: Basic cations (M_b) in solution. At pH <5, Al species in solution (monomeric Al-hydroxy cations, e.g., $Al(OH)^{2+}$, Al^{3+}, polymeric Al-hydroxy cations: $n(Al(OH)_x^{3-x})^+$; Mn^{2+} from oxides. Overlaps with cation exchange capacity (CEC), Al and Fe buffers; c, cation exchange (CEC) buffer: nonexchangeable polymeric Al-hydroxy cations (blockage of permanent charge – reduction of CEC). Exchangeable Mn^{2+}, monomeric Al-hydroxy cations, and Al^{3+} reduce base saturation by displacement of M_b cations. At pH <4.5, amphoteric Al-hydroxy cations, (blocks and/or changes permanent charge to pH dependent charge – reduction of CEC). Formation of Al-hydroxy sulfates at pH >4.2; d, aluminum buffer: Al^{3+} in solution (from interlayer Al, Al-hydroxy sulfates, and Al hydroxides). H^+ adsorption – permanent charge to pH-dependent charge – reduction of CEC; e, Al/Fe buffer: exchangeable H^+ and Fe (displacement of Al); f, iron buffer: exchangeable H^+ and Fe (mineral disruption). ANC, acid-neutralizing capacity.

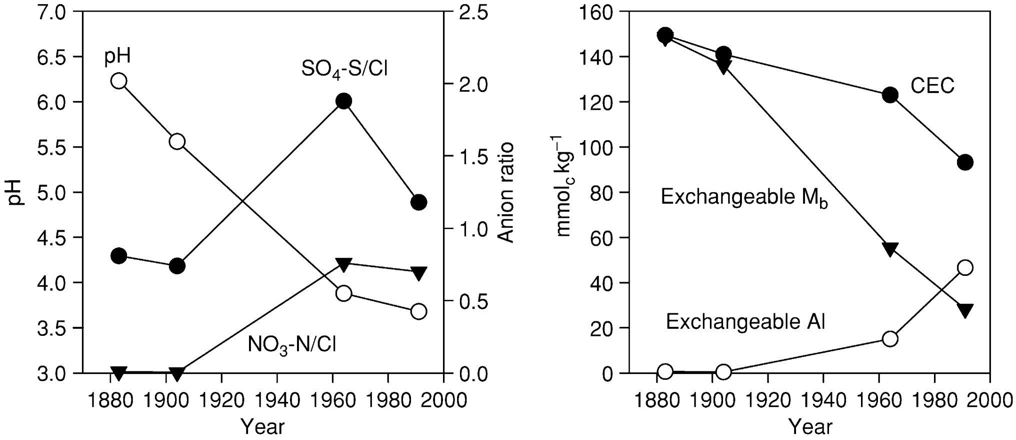


Figure 4 Long-term (1883–1991) changes in soil chemistry in archived soil samples (0–23 cm depth) taken from woodland (Geescroft wilderness) subject to acid deposition at Rothamsted, UK. CEC, cation exchange capacity.

buffering in certain pH ranges: by carbonates at near-neutral pH values, and by hydroxy Al and Fe compounds and clay minerals at pH 4.2–3.2. Second, the rapid addition of strong mineral acids from acid deposition depresses the pH much more rapidly than slower addition from most internal sources. This is particularly characterized by a relatively rapid decline of the CEC buffer. Major long-term changes in soil chemistry through soil acidification beneath woodland at Rothamsted are shown in Figure 4. A schematic representation of changes through the soil profile is shown in Figure 5.

Carbonate and Silicate Buffers

The dissolution of Ca and Mg carbonates is driven by the protonation of CO_3^{2-} at pH values greater than 6.2; the reaction products are soluble bicarbonates which are leached through the soil:

$$CaCO_3 + H_2CO_3 \leftrightarrow Ca^{2+} + 2HCO_3^- \qquad [24]$$

As soon as $CaCO_3$ is used up, the M_b cations leached with bicarbonate originate from silicate weathering.

The weathering of primary silicates is the dominant buffer reaction in soils at pH <6. For relatively abundant SiO–Al–M groups, there is a two-step dissolution process. At pH values less than 6.5, SiO–Al–M_b groups become progressively protonated, liberating M^+ and M^{2+}, while SiO–Al groups remain intact. The products of acid weathering of silicates are clay minerals with permanent negative charge and mobile cations. Thus, stable ANC present in silicates is transferred to labile-mobilizable ANC as exchangeable base cations adsorbed on clay minerals. When pH has decreased to approximately pH 5, the SiO–Al become progressively protonated, releasing Al species into solution. The mobilized reaction products are: (1) monomeric Al-hydroxy cations and Al^{3+}, which are adsorbed on to exchange surfaces; and (2) nonexchangeable polymeric Al-hydroxy cations, which lead to a reduction in CEC. Mn oxides (if present) are also readily protonated at pH <5 to such an extent that the acidification front is characterized by a high proportion of exchangeable Mn^{2+}.

Cation Exchange Buffer

As acidity increases at pH values of less than 5, Al^{3+} in solution and on exchangeable surfaces increases. Al^{3+} saturation on exchange surfaces can reach high values at low Al^{3+} concentrations in solution, because of highly selective bonding of Al^{3+} on exchange surfaces. As a result, M_b cations bound to the exchange sites are leached together with the anion of the acid-generating proton and the output of anions comes close to the input. This results in a progressive decline in base saturation. Where the dominating anion leached is SO_4^{2-}, the main source of acidity is H_2SO_4 from acidic deposition.

In most cases, the continued weathering of silicates results in substantially more Al^{3+} than M_b cations and is a major source of acidity, because Al^{3+} and hydrolyzed Al ions are intermediaries of protons (BNC; eqn [18]). As a result, acidity moves down the soil profile with the Al species in the leachate. The CEC partially buffers this potential acidity through adsorption on to exchange surfaces. If the acid loading is H_2SO_4, the formation of Al-hydroxy sulfates occurs at the lower pH range of the CEC buffer:

$$Al(OH)_3 + H_2SO_4 \leftrightarrow AlOHSO_4 + 2H_2O \qquad [25]$$

This gives the soil the additional property to react as a sink for H_2SO_4. At the same time the CEC is strongly reduced by the covering of clay surfaces

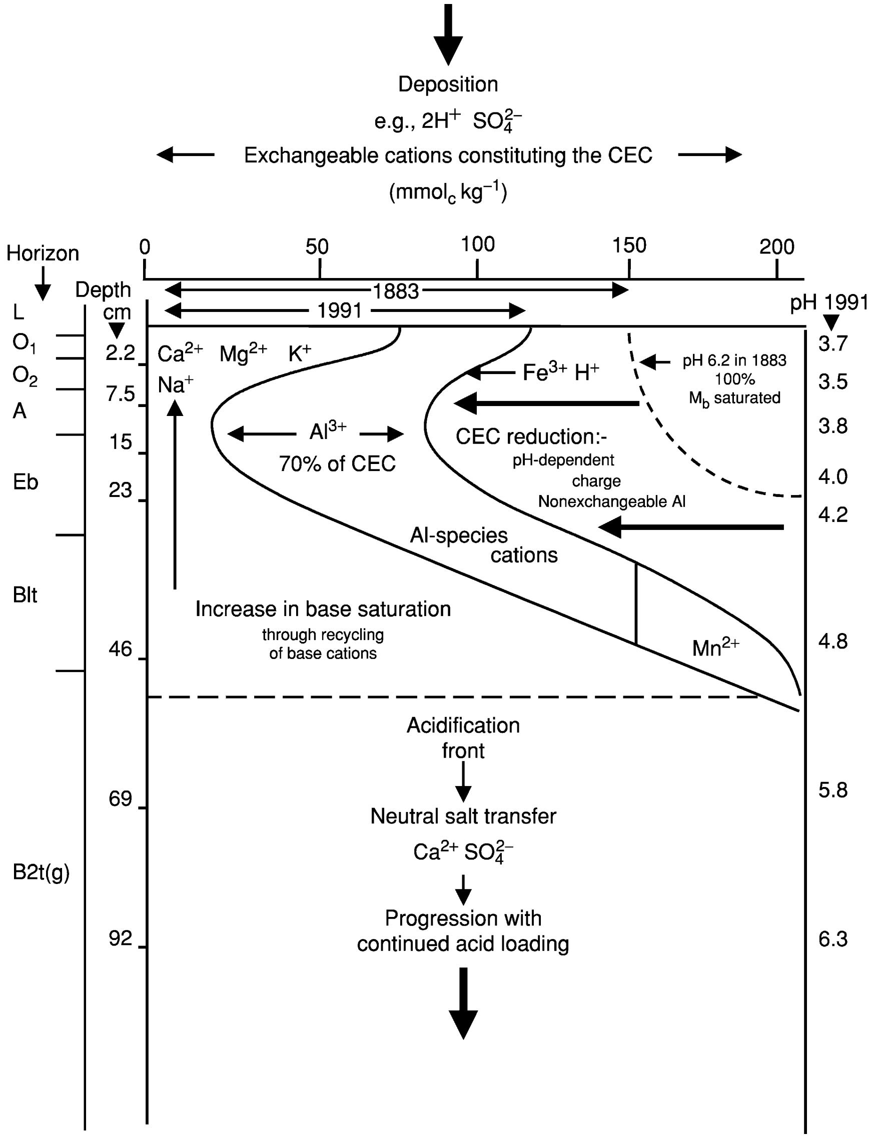


Figure 5 A schematic model of soil acidification beneath a typical deciduous woodland (Geescroft wilderness) receiving long-term acid deposition. The model is based on the analysis of archived soils (1883–1991) taken in incrementing depths to a total depth of 92 cm. CEC, cation exchange capacity.

with amphoteric-Al hydroxy cations, blocking adsorption sites and 'changing' permanent charge into pH-dependent charge.

The degree of saturation of the CEC with Al species indicates the remaining capacity of the solid phase for proton binding. Acid mineral subsoils with Al-saturated exchange surfaces (less than 5% M_b) characterize soils subject to long-term acid deposition.

Aluminum and Iron Buffer

At pH <4.2, Al-hydroxy compounds and interlayer Al become increasingly weathered to an extent that Al^{3+} becomes the dominant cation in the soil solution and exchange surfaces become saturated with Al^{3+}. Al-hydroxy cations and sulfates accumulated in the CEC buffer range decrease through dissolution:

$$AlOHSO_4 + H^+ \rightarrow Al^{3+} + SO_4^{2-} + H_2O \qquad [26]$$

Leached (output) SO_4^{2-}, H^+, and Al^{3+} levels show a tendency to exceed SO_4^{2-} and H^+ inputs where acidic deposition is dominated by H_2SO_4.

At pH <3.8, Fe (hydr)oxides become increasingly protonated. This buffer mechanism runs parallel with the Al buffer to pH 3.2 and is accompanied by a progressive shift in the exchangeable cation composition from Al^{3+} to Fe^{3+} and H^+ ions, with Al^{3+} being desorbed.

Proton Adsorption Reactions

The adsorption of H^+ between pH 5 and 4 is through negatively charged surfaces on organic matter (R-COO^-) and (hydr)oxides (CEC buffer). Such storage is reversible through exchange with M_b cations. At pH >4, only adsorption sites that exert a high preference for protons contribute significantly to H^+ adsorption on to clay minerals. The direct protonation of adsorption sites on clay minerals contributes to ANC most at pH <4 and permanent charge at mineral surfaces becomes progressively pH-dependent. Isomorphous substitution may also be restricted through mineral disruption and the inclusion of M_a cations. These processes reduce the CEC.

The permanent effect of dissolution and adsorption reactions on soil clay may be reflected by X-ray diffraction analysis in a reduction in expansion of swelling mineral (where present) upon solvation. This can be accounted for by the introduction of interlayer Al species.

Acid Sources and Sinks: Input–Output Relationships and the Calculation of Proton Balances

In order to quantify the impact of acidic deposition on changes in soil pH in relation to other soil acidification processes, all important H^+ fluxes need to be assessed. Protons, acid precursors, and bases entering or transferred within the soil (inputs) leave the system as solutes or in biomass (outputs). Protons can be produced or consumed within the system by the buffering processes described above. From the principle of electroneutrality, for a given soil compartment, the storage of cations ($M_a = M_b = M^+$) equals the storage of anions (A^-) expressed in moles ion charge (equivalents):

$$M^+ + H^+ = A^- \qquad [27]$$

The change in protons (H^+) can therefore be calculated from the change in storage fluxes of cations and anions within soil compartments:

$$H^+ = A^- - M^+ \qquad [28]$$

On this basis the total proton load (TPL) can be estimated by the determination of fluxes within the individual compartments (H^+ sources and sinks) as follows:

$$\begin{gathered}
\text{Input}-\text{output} \qquad\qquad \text{Internal proton} \\
\text{ion flux (weathering)} \qquad\qquad \text{production} \\
H^+ = (H^+_{in} - H^+_{out}) + [(M^+_{in} - M^+_{out}) - (A^-_{in} - A^-_{out})] \\
\text{Protons} \qquad\qquad \text{Cations} \qquad\qquad \text{Anions} \\
+ [(M^+_{up} - A^-_{up}) + (HCO^-_{3\ in} - HCO^-_{3\ out}) \\
\text{Ion uptake} \qquad\qquad \text{Dissociation} \\
+ (R{-}COO^-_{in} + RCOO^-_{out})] \\
\text{Protolysis} \\
+[(NH^+_{4\ in} - NO^-_{3\ in})(NH^+_{4\ out} - NO^-_{3\ out})] = TPL \\
\text{N-transformations: deposition} + \text{organic N}
\end{gathered} \qquad [29]$$

The weakness of this approach is the difficulty involved in determining the rates of these processes. Cation and anion weathering and organic acid dissociation can be estimated by the differences between the sum of measured cations and anions in soil solution or drainage using standard lysimeters or suction-cup lysimeters. If the amount of biomass and its cation and anion content are known, the cation excess can be calculated; the proton production due to the development and export of biomass is equal to the cation excess. The kind of N nutrition cannot be measured directly; its effect on the soil proton flux is included in the production and consumption of NH_4^+ and NO_3^- in the input–output balance of the soil; this assesses the whole effect of the N cycle, which includes N mineralization as well as N uptake.

Table 1 shows examples of proton budgets determined for four of the 'classic' ecosystem studies of acidification with the necessary calculation steps. Note that proton sources should balance proton sinks. The isolation of internal and external proton sources highlights the profound effect that acid deposition has had on the ecosystems at Solling (Germany) and Rothamsted (England) and the much lesser effect of internal proton production at Cedar River (USA). The predominant source of acidity at these locations is also identified by the anion composition (SO_4-S:Cl and NO_3-N:Cl ratios) in the soil solution or leachate, providing clear evidence of the contribution of acid deposition to recent soil acidification.

Acidification, Ecosystem Stability and Global Change

The influence of acid deposition on soil acidity can be evaluated by looking for the characteristic effects that it has. The transfer of soil horizons beneath the rooting zone of nonpodzolized soils into the Al-buffer range only occurs through long-term acid deposition because

Table 1 Processes which have to be considered for the estimation of proton budgets, with four examples from different locations. Values are in kilomoles of charge per hectare per year

	Solling, Germany[g]		*Geescroft wilderness,*	
	Spruce	*Beech*	*UK*[h] *deciduous woodland*	*Cedar river, USA Douglas fir*[i]
Source				
Wet deposition of H^+[a]	↓	↓	0.2	↓
$SO_2 + NO_2$ deposition[a]	3.7	1.7	2.5	0.3
N transformations[b]	1.1	0.8	2.8	0.1
Base cation uptake[c]	2.1	1.5	0.9	0.5
Carbonic acid and/or organic acid dissociation[d]	0.6	0.5	0.1	0.4
Anion weathering[e]	0.2	0.1	0	0.3
Total	7.7	4.6	6.5	1.6
Depth of soil profile (M)	1.0	1.0	0.46	0.45
pH of soil solution/output	4.0	4.2	4.0	6.7
SO_4-S/Cl deposition	2.0	1.6	1.3[f]	–
output	3.0	2.0	0.9	0.7
NO_3-N/Cl deposition	0.40	0.33	1.6	–
output	0.35	0.05	0.3	0.04
Sink				
Leaching of H^+	0.4	0.5	1.1	0.0
N transformations	0.0	0.5	0	0.1
Anion accumulation	0.6	0.3	0	0.1
Cation weathering	6.1	3.0	5.9	1.3
Total	7.1	4.3	7.0	1.5

[a] Wet-deposited acidity and dry deposition of acid precursors such as SO_2 and NO_x.

[b] N transformation refers to the input–output balance of NH_4^+ and NO_3^-. There is no proton production if $(NH_4^+{}_{in}–NH_4^+{}_{out}) + (NO_3^-{}_{out}– (NO_3^-{}_{in}) > 0$.

[c] Cation and anion accumulation refer to the accumulation in biomass and humus of cations and anions other than NH_4^+ and NO_3^-.

[d] Estimated from the difference between the sum of measured cations and the sum of measured anions in soil solution or drainage water.

[e] The weathering of individual cations and anions is estimated from the following mass balance: $W = O - I = \Delta B$, where O is the output in drainage water; I is the input through deposition; ΔB is accumulation in biomass and humus.

[f] Reflects local decline in S deposition since 1980 (see also Figure 4).

[g] Source: van Breeman N, Driscoll CT, and Mulder J (1984) Acidic deposition and internal proton sources in acidification of soils and waters. *Nature* 307: 599–604.

[h] Source: Blake L, Goulding KWT, Mott CJB, and Johnston AE (1999) Changes in soil chemistry accompanying soil acidification over more than 100 years under woodland and grass at Rothamsted Experimental Station, UK. *European Journal of Science* 50: 401–412.

[i] Source: van Miegroet H (1986). *Role of N Status and N Transformations in H⁺ Budget, Cation Loss and S Retention Mechanisms in Adjacent Douglas-Fir and Red Alder Forests.* PhD Thesis. Washington, DC: University of Washington.

nitrification of organic N is restricted to the rooting zone. Once the rooted soil is acidified, the release of M_a cations compounds the buildup of BNC and the acidifying front passes into the deeper horizons together with leached M_b cations as sulfates and/or nitrates. Under continued acid deposition loading, this process continues through the soil profile beneath the rooting zone with soluble Al-species, sulfate, and/or nitrate accumulating in the soil solution and seepage waters.

Since 1980 the deposition of S from the atmosphere has decreased throughout Europe and North America as the result of pollution abatement strategies based on the quantitative determinations of critical loads of acidifying pollutants. For many natural ecosystems to maintain soils at their present pH, this approach requires deposition values to be restricted to preindustrial emissions of less than 5 kg ha^{-1} per year each of total S and N. The deposition of acidifying N-species in many locations still considerably exceeds this value. Deacidification can only occur if acid deposition tends to zero and the release of M_b cations from silicates and organic matter can replenish base saturation. However, if acidification has been allowed to proceed to a point where the CEC is nearly empty of nutrient cations (Al buffer), the time for the soil system to repair itself naturally to a state that is ecologically acceptable will be in the order of centuries.

The acidification of ecosystems has a considerable influence on soil biological and chemical processes. Al mobilization and release to surface waters are of most concern, because Al is highly toxic to terrestrial plants, and to humans and animals drinking the water. Soil acidification also causes solubilization of heavy metals in soils, both those naturally contained in minerals and any that entered the soil from pollution arising from waste disposal. As with Al, metals such as Pb, Cu, Zn, or Cd can be toxic to plants or, if they enter the food chain, to humans and animals.

See also: **Acidity**; **Buffering Capacity**; **Liming**; **pH**; **Sulfur in Soils:** Overview

Further Reading

Berden M, Nilsen SI, Rosen K, and Tyler G (1987) *Soil Acidification: Extent, Causes and Consequences*. National Swedish Environment Protection Board Report 3292. Solna, Sweden: National Swedish Environmental Protection Board Information Section.

Blake L, Goulding KWT, Mott CJB, and Johnston AE (1999) Changes in soil chemistry accompanying acidification over more than 100 years under woodland and grass at Rothamsted Experimental Station, UK. *European Journal of Soil Science* 50: 401–412.

Billet MF, Parker Javis F, Fitzpatrick EA, and Cresser MS (1990) Forest soil chemical changes between 1949/50 and 1987. *Journal of Soil Science* 41: 133–145.

Bredemeier M (1989) Nature and potential of ecosystem – internal acidification processes in relation to acid deposition. In: Longhurst JWS (ed.) *Acid Deposition*, pp. 197–212. London, UK: British Library Technical Communications.

Hallbacken L and Tamm CO (1986) Changes in soil acidity from 1927 to 1982–84 in a forest area of southwest Sweden. *Scandinavian Journal of Forest Research* 1: 219–232.

Helyar KR and Porter WM (1989) Soil acidification, its measurement and the processes involved. In: Robson AD (ed.) *Soil Acidity and Plant Growth*, pp. 61–101. Sydney, Australia: Academic Press.

Johnson DW (1987) A discussion of changes in soil acidity due to natural processes and acid deposition. In: Hutchinson TC and Meema K (eds) *Effects of Acidic Deposition on Forests, Wetlands and Agricultural Ecosystems*, pp. 333–346. New York: Springer-Verlag.

Kennedy IR (1992) *Acid Soil and Acid Rain*. New York: Research Studies Press, LTD/John Wiley.

Last FT and Watling R (eds) (1991) Acidic deposition its nature and impacts. *Royal Society of Edinburgh Proceedings. Section B Biological Sciences* 97: 17–34, 81–116, 155–168, 169–191.

Reuss JO and Johnson DW (1986) Acid deposition and the acidification of soils and waters. *Ecological Studies Analysis and Synthesis*, vol. 59. New York: Springer-Verlag.

Reuss JO and Walthall PM (1989) Soil reaction and acidic deposition. In: Norton SA, Lindberg SE, and Page AL (eds) *Acidic Precipitation*, vol. 4, pp. 1–31. *Soils, Aquatic Processes and Lake Acidification*. New York: Springer-Verlag.

Ulrich B (1983) Effects of acid deposition. In: Beilke S and Elshout AJ (eds) *Acid Deposition*, pp. 31–41. Dordrecht, The Netherlands: Reidel.

Ulrich B (1983) A concept of forest ecosystem stability and of acid Deposition as a driving force for destabilization. In: Ulrich B and Pankrath J (eds) *Effects of Accumulation of Air Pollutants in Forest Ecosystems*, pp. 1–29. Dordrecht, The Netherlands: Reidel.

Ulrich B and Sumner ME (eds) (1991) *Soil Acidity*. Berlin, Germany: Springer-Verlag.

van Breeman N, Driscoll CT, and Mulder J (1984) Acidic deposition and internal proton sources in acidification of soils and waters. *Nature* 307: 599–604.

ACIDITY

N S Bolan, Massey University, Palmerston North, New Zealand
D Curtin, New Zealand Institute for Crop and Food Research, Christchurch, New Zealand
D C Adriano, University of Georgia, Aiken, SC, USA

Introduction

Soil acidification is a natural process that can either be accelerated by certain plants and human activities or slowed down by careful management practices. Industrial and mining activities lead to soil acidification due to acid produced from pyrite oxidation and from acid precipitation caused by the emission of sulfur (S) and nitrogen (N) gases. In managed ecosystems, soil acidification is mainly caused by the release of protons (H^+) during the transformation and cycling of carbon (C), N, and S, and fertilizer reactions. Soil acidification caused by these processes can have adverse impacts where soils are unable to buffer against further pH decrease. For example, in parts of North America and Europe, soil acidification caused by acid precipitation has resulted in forest decline and, in some parts of Australia, continuous legume cultivation and inappropriate use of N fertilizer have generated sufficient soil acidity that cereal crop cultivation has had to be abandoned due to aluminum (Al) and manganese (Mn) toxicity.

Historically, liming is the most common management practice used to neutralize soil acidity. Most plants grow well in the pH range 5.5–6.5, and the usual objective of liming programs is to maintain pH in this range. Liming enhances the physical, chemical, and biological properties of soil through its direct effect in ameliorating soil acidity and through its indirect effects in mobilizing plant nutrients, immobilizing toxic heavy metals, and improving soil physical conditions. In variable-charge soils, liming can be used as a management tool to manipulate the surface charge, thereby controlling the reactions of nutrient ions and heavy metals. Liming provides optimum conditions for a number of biological processes, including N_2 fixation and mineralization of N, phosphorus (P), and S.

Processes Generating Acidity in Soils

Processes generating acidity in soils can be broadly grouped into two categories: (1) those occurring in natural ecosystems through industrial activities, and (2) those occurring in managed ecosystems because of farming activities. The various reactions involved in these processes are given in Table 1.

The two important acid-generating processes resulting from industrial activities in natural ecosystems are acid drainage following pyrite oxidation and deposition of acidity in precipitation. While the first process occurs at a local level, the second process can lead to acidification far away from the source of acid production and hence an international effort is needed to combat it.

Acid Drainage

Acid drainage has both anthropogenic and natural origins, including mining of coal and sulfide-containing metal ores, land disturbances (e.g., rice cultivation), and industrial activities (mineral processing, manufacturing or recycling of batteries, electronic equipment, wood pulp, tanneries and textile manufacturing, and food processing).

Pyrite is commonly associated with coal and metal ores, as well as mine deltas, wetlands, and rice fields. Exposure of pyrite to the atmosphere leads to its oxidation and the production of extremely acidic drainage water. Pyrite oxidation includes biological and electrochemical (abiotic) reactions (Table 1). Abiotic oxidation of pyrite is pH-sensitive and it is extremely slow in very acidic conditions. *Thiobacillus ferrooxidans* and *T. thiooxidans* are mainly responsible for the oxidation of pyrite, especially in acid soils.

Acid Precipitation

Carbon dioxide combines with water to form a dilute solution of carbonic acid (H_2CO_3) with an equilibrium pH of approximately 5.6. For this reason, acid precipitation is arbitrarily defined as precipitation with a pH value of less than 5.6. Natural sources of acid

Table 1 Proton generation and consumption processes in the soil–plant system, including those associated with acid precipitation, pyrite oxidation, and the C, N, and S biogeochemical cycles

Process	*Reaction equation*	*H^+ ($mol_c mol^{-1}$)*	*Eqn no.*
Acid precipitation			
Oxidation of sulfur dioxide	$2SO_2 + O_2 \rightarrow 2SO_3$	0	1.1
Hydrolysis of sulfur trioxide	$SO_3 + H_2O \rightarrow H_2SO_4 \rightarrow SO_4^{2-} + 2H^+$	+2	1.2
Photochemical oxidation of nitric oxide	$O_3 + NO \rightarrow N_2O + O_2$	0	1.3
Hydrolysis of nitrogen dioxide	$2NO_2 + H_2O \rightarrow HNO_3 + HNO_2 \rightarrow NO_3 + H^+$	+1	1.4
Pyrite oxidation			
Pyrite oxidation by oxygen	$2FeS_2 + 7O_2 + 2H_2O \rightarrow 2Fe^{2+} + 4SO_4^{2-} + 4H^+$	+2	2.1
Ferrous iron oxidation	$4Fe^{2+} + O_2 + 4H^+ \rightarrow 4Fe^{3+} + 2H_2O$	−1	2.2
Ferric iron precipitation	$Fe^{3+} + 3H_2O \rightarrow Fe(OH)_3 + 3H^+$	+3	2.3
Pyrite oxidation by ferric iron	$FeS_2 + 14Fe^{3+} + H_2O \rightarrow 15Fe^{2+} + 2SO_4^{2-} + 16H^+$		2.4
Carbon cycle			
Dissolution of carbon dioxide	$CO_2 + H_2O \rightarrow H_2CO_3 \rightarrow H^+ + HCO_3^-$	+1	3.1
Synthesis of organic acid	Organic C $\rightarrow$ RCOOH $\rightarrow$ $RCOO^- + H^+$	+1	3.2
Nitrogen cycle			
N fixation	$2N_2 + 2H_2O + 4R{\cdot}OH \rightarrow 4R\text{-}NH_2 + 3O_2$	0	3.3
Mineralization of organic N	$R\text{-}NH_2 + H^+ + H_2O \rightarrow R\text{-}OH + NH_4^+$	−1	3.4
Urea hydrolysis	$(NH_2)_2CO + 3H_2O \rightarrow 2NH_4^+ + 2OH^- + CO_2$	−1	3.5
Ammonium assimilation	$NH_4^+ + R{\cdot}OH \rightarrow R\text{-}NH_2 + H_2O + H^+$	+1	3.6
Ammonia volatilization	$NH_4^+ + OH^- \rightarrow NH_3\uparrow + H_2\mathit{O}$	+1	3.7
Nitrification	$NH_4^+ + 2O_2 \rightarrow NO_3^- + H_2O + 2H^+$	+2	3.8
Nitrate assimilation	$NO_3^- + 8H^+ + 8e^- \rightarrow NH_3 + 2H_2O + OH^-$	−1	3.9
Denitrification	$4NO_3^- + 4H^+ \rightarrow 2N_2 + 5O_2 + 2H_2O$	−1	3.10
Sulfur cycle			
Mineralization of organic S	2Organic S $+ 3O_2 + 2H_2O \rightarrow 2SO_4^{2-} + 4H^+$	+2	3.11
Assimilation of sulfate	$SO_4^{2-} + 8H^+ + 8e^- \rightarrow SH_2 + 2H_2O + 2OH^-$	−2	3.12
Oxidation of S^0	$2S^0 + 2H_2O + 3O_2 \rightarrow 2SO_4^{2-} + 4H^+$	+2	3.13

precipitation include geologic weathering, volcanic eruptions, anaerobic decomposition of organic matter, air-borne sea-salt sprays, and production of N oxides (NO_x) during lightning storms. The increased acid precipitation burden in recent decades has been attributed to anthropogenic sources that include combustion of fossil fuels (especially coal and oil), certain industrial processes (especially smelting of ores), exhausts from internal combustion engines, and N fertilization of agricultural and forest lands.

Widespread occurrence of acid precipitation (i.e., both wet and dry deposition) results in large part from industrial emissions of oxides of S (SO_x) and N (NO_x). These compounds are transformed in the atmosphere to sulfuric and nitric acids (**Table 1**), which can be transported over great distances before deposition on vegetation, soils, surface waters, and building structures. The average annual ratio of sulfuric to nitric acids in North America is approximately 2:1, but nitric acid is becoming progressively important relative to sulfuric acid because of the installation of flue gas desulfurization (FGD) systems in coal-fired power stations. The major impact of acid precipitation has been on freshwater ecosystems, while its impact on terrestrial ecosystems is more controversial. Forest health may be one casualty of acid precipitation, but increasing tropospheric ozone may also be a factor in forest decline.

Elemental Cycling

The most significant proton (H^+)-generating processes in soil are associated with the cycling of C, N, and S (**Table 1**) and these processes can be grouped into two main categories: plant-induced (i.e., uptake and assimilation) and soil-induced (i.e., transformation) processes.

In the case of the C cycle, the metabolism of photosynthates results in the synthesis of organic acids and, at the cytoplasmic pH of the plants (pH ≈7.2–7.4), some of the carboxyl groups of these acids dissociate to produce H^+ ions. Cytoplasmic pH regulation is generally achieved by transport of excess H^+ ions out of the cytoplasm. Some terrestrial plant species counteract the change in cytoplasm pH by excreting H^+ ions into the soil solution and, at the same time, taking in a nutrient basic cation to balance the charge.

Plants take up N in three main forms – as an anion (nitrate, NO_3^-), as a cation (ammonium, NH_4^+), or as a neutral N_2 molecule (N_2 fixation). Depending upon the form of N taken up and the mechanism of assimilation in the plant, excessive cation or anion uptake may occur. To maintain charge balance during the uptake process, H^+, OH^-, or bicarbonate (HCO_3^-) ions must pass out of the root into the surrounding soil. While the uptake of NH_4^+ and N_2 fixation result in a net release of H^+ ions, uptake of NO_3^- can result in a net release of OH^- ions. In the case of N_2 fixation, acidity is generated even when no ionic species of N are taken up by the plant. This is because basic cations are imported into the legume in exchange for H^+ ions generated during C assimilation into carboxylic acids. The amount of H^+ ion released during N_2 fixation is a function of C assimilation and therefore it depends mainly on the nature and amounts of amino acids and organic acids synthesized by the plant. When ammonium (NH_4^+) assimilation occurs in the root, deprotonation of NH_4^+ to R-NH_2 releases one mole of H^+ per mole of NH_4^+ (eqn [3.6]; **Table 1**). When plants take up NO_3^-, the NO_3^- ion is first reduced to NH_4^+, which is subsequently assimilated into amino acids. This NO_3^- reduction produces one mole of OH^- ion for every mole of NO_3^- reduced to NH_4^+ (eqn [3.9]; **Table 1**).

Sulfate (SO_4^{2-}) is assimilated into sulfur-containing amino acids (cysteine, cystine, and methionine) in the form of sulfhydryl (–SH) groups. This reduction process produces two net moles of OH^- for each mole of S assimilated (eqn [3.12]). On decomposition of sulfhydryl-containing amino acids, two moles of H^+ ions are generated for each mole of –SH oxidized to sulfate. Since plants require roughly 10 times less S than N, assimilation of SO_4^{2-} will have only a minor effect on H^+ balance in plants.

As microorganisms decompose soil organic matter, they respire CO_2, which upon hydrolysis forms H_2CO_3 (eqn [3.1]), eventually dissociating to H^+ and HCO_3^- ions (eqn [3.2]). The continuous production of CO_2 through soil and root respiration increases the concentration of CO_2 in the soil air spaces. Soil microorganisms also produce organic acids when they decompose plant litter that is rich in organic compounds but low in charge-balancing basic cations (eqn [3.2]). In general, conifer litter tends to produce more organic acids than does leaf fall from deciduous woodlands.

Nitrification and ammonia (NH_3) volatilization result in the release of H^+ ions. While the ammonification process (conversion of organic forms of N to NH_4^+-N) results in the release of OH^- ions, the subsequent oxidation of NH_4^+ to NO_3^- (nitrification) results in the release of H^+ ions. Combined ammonification (eqn [3.4]) and nitrification (eqn [3.8]) of organic N compounds, including urea, in theory generates one net mole of H^+ for every mole of N transformed. In alkaline media NH_4^+ ions dissociate into gaseous NH_3, which is subject to volatilization (eqn [3.7]), resulting in a net decrease in pH due to the consumption of OH^- ions (or release of H^+ ions) as NH_4^+ is converted to NH_3.

In aerobic soils, H^+ ions are produced during the mineralization and subsequent oxidation of S in soil organic matter (eqn [3.11]). As soil bacteria and fungi decompose plant litter and soil organic matter rich in C but low in S, soil solution SO_4^{2-} may be immobilized. In this case eqn [3.11] (Table 1) is reversed and it becomes an H^+-consuming reaction, because SO_4^{2-} is assimilated into microbial biomass. Under anaerobic conditions, some bacteria have the capacity to use SO_4^{2-} as a terminal electron acceptor, resulting in H^+ consumption as SO_4^{2-} is reduced along a chain of intermediate compounds to hydrogen sulfide (H_2S). The H_2S reacts with metal ions to precipitate as metal sulfides, which is an H^+-consuming process. However, when these metal sulfides are reoxidized, they generate H^+, acidifying the soil (acid drainage).

Fertilizer Reactions

Fertilizer application in ecosystems managed for agricultural production is a major contributor to soil acidification. The acidifying effects of fertilizers commonly used in agricultural production are presented in Table 2.

Application of N fertilizers such as urea and ammonium sulfate to soils produces H^+ by two processes: nitrification (eqn [3.8]) and NO_3^- leaching. Part of the H^+ produced is neutralized by OH^- released by plants during the subsequent uptake of the NO_3^- ions (eqn [3.9]). The depletion of basic cations (Ca, K, Mg, and Na) during the leaching of NO_3^- ions (i.e., as ion pair) and product removal (i.e., silage) accelerates the acidification process. In the case of ammonium sulfate, the concomitant leaching of SO_4^{2-} and NO_3^- ions causes greater depletion of basic cations. With urea, the initial conversion of amide N (R-NH_2) to NH_4^+ (ammonification) releases OH^- (eqn [3.5]), which neutralizes part of the H^+ produced during the subsequent oxidation of NH_4^+ to NO_3^-, explaining why urea-based N fertilizers are less acidifying than the NH_4^+-based fertilizers (Table 2).

Superphosphates, the most common phosphate fertilizers, contain monocalcium phosphate (MCP) as the principal P component. Dissolution of MCP in soils results in the formation of dicalcium phosphate with a release of phosphoric acid, which subsequently dissociates into phosphate anions ($H_2PO_4^-$) and H^+. Part of H^+ is subsequently neutralized by the OH^- released during the adsorption of the $H_2PO_4^-$ by soil particles. Compared with weakly adsorbed SO_4^{2-} and NO_3^- ions, $H_2PO_4^-$-induced leaching of basic cations is unlikely to occur, since the $H_2PO_4^-$ ions are strongly adsorbed by most soils.

In legume-based pasture and crop-production systems, P fertilizers are added to promote N_2 fixation by the legumes. Regardless of P fertilizer source, application of P to legume-based systems promotes N_2 fixation, thereby indirectly causing soil acidification.

Table 2 Acidity production by various fertilizers and its effect on pH of two soils with contrasting buffer capacities

			Number of years required to reduce soil pH by 1 unit[c]	
Fertilizer	*Acidity equivalence*[a]	*Acidity produced (kmol H^+ ha^{-1})*[b]	*Tokomaru silt loam*	*Egmont clay loam*
Ammonium sulfate	110	2.60	8	26
Ammonium chloride	93			
Ammonium nitrate	60			
Diammonium phosphate	74	2.06	10	33
Monoammonium phosphate	55			
Urea	79	0.86	25	78
Potassium nitrate	−23			
Calcium nitrate	−50			
Sodium nitrate	−29			
Nitrogen fixation	70–250	–	–	–
Single superphosphate	8	0.48	45	140
Triple superphosphate	15	0.50	43	135
North Carolina phosphate rock	−50			
Calcium sulfate	−57			
Potassium sulfate	−64			
Elemental sulfur (S^0)	310	1.55	14	43

[a] Acidity equivalence is the number of parts by weight of pure lime (calcium carbonate) required to neutralize the acidity caused by 100 parts of the fertilizer. Negative values indicate the liming value (kilograms $CaCO_3$/100 kg) of the fertilizer.

[b] Ammonium sulfate, diammonium phosphate, and urea added at the rate of 25 kg N ha^{-1} $year^{-1}$; single and triple superphosphate at the rate of 30 kg P ha^{-1} $year^{-1}$; and S^0 at the rate of 30 kg S ha^{-1} $year^{-1}$.

[c] pH-buffering capacity (kilomoles H^+ per hectare) of 21.7 and 67.5 for the Tokomaru and the Egmont soil, respectively.

The amount of acidity produced indirectly by N_2 fixation depends mainly on the extent of NO_3^- leaching and it can be greater than that produced directly by the dissolution of MCP in superphosphate fertilizer granules (Table 2).

Elemental sulfur (S^0) is used as an acidifying agent, as a slow-release S fertilizer or, in a finely divided form, as a fungicide. When S^0 is added to soils, it is oxidized to sulfuric acid, which dissociates into SO_4^{2-} and H^+ ions (eqn [3.13]).

Measurement and Effects of Soil Acidity

The acidity of a soil is assessed from the activity of H^+ ions in the soil solution. Since there is a huge range of $[H^+]$ in soil solution, a logarithmic scale (known as pH scale; $pH = -\log [H^+]$) is used to quantify acidity. In soils, acidification is indicated by a decrease in either pH or anion neutralizing capacity. For a given input of acidity, the extent to which pH decreases depends mainly on the pH buffering capacity of the soil. Various soil constituents, such as organic matter, Fe and Al oxides, and $CaCO_3$ (in calcareous soils) contribute to the pH buffering of soils at different pH values. Depending on organic matter content, texture, and the nature of the clay mineralogy, the amount of acidity needed to reduce pH of topsoil (noncalcareous) by 1 unit is normally in the 1- to 8-cmol (H^+) kg^{-1} range. These values translate to buffer capacities of approximately 10–70 kmol (H^+) ha^{-1} for the top 7.5 cm soil layer. The amounts of acidity produced by selected fertilizers and the number of years required to reduce soil pH by 1 unit are shown in Table 2 for two soils with contrasting pH buffer capacities. It is clear that annual application of ammonium sulfate or ammonium phosphates could acidify soil within a short time (10 years), particularly when soil pH-buffering capacity is low.

Acidification affects the transformation and biogeochemical cycling of both nutrients and heavy metals through its effect on the physical, chemical, and biological characteristics of soils. Soil pH can be viewed as the master variable of all the driving factors, because it can affect the surface charge and subsequent adsorption of solutes by variable-charge soil components. In addition to its effect on the sorption of metal cations and anions in soils, it also influences metal speciation, complexation of metals with organic matter, precipitation/dissolution reactions, redox reactions, mobility and leaching, dispersion of colloids and, ultimately, the bioavailability of trace metals.

Soil pH affects the surface charge through the supply of H^+ for adsorption on to the metal oxides and the dissociation of the functional groups in the soil organic matter. A decrease in pH decreases the net negative charge (often referred to as cation exchange capacity or CEC) and increases the net positive charge (often referred to as anion exchange capacity or AEC).

Because acidity governs the type, number, and activity of microorganisms in soil, it influences the rate of organic matter mineralization and availability of elements such as N, S, and P. Because of suboptimal microbial activity in highly acid conditions, organic matter accumulates, giving rise to a storehouse of nutrients that can be exploited by liming. Likewise, nitrification is markedly reduced below pH 6, whereas the ammonification reaction is relatively insensitive to acidity, resulting in the accumulation of NH_4^+ in acidic soil where nitrification is inhibited. Thus, for certain crops unable to use NH_4^+, acidification can result in restricted uptake of N and it may promote NH_4^+ toxicity.

Acidity has a deleterious effect on the symbiotic relationship between *Rhizobia* and legumes, and nodulation and N_2 fixation are generally poor below pH 6. Physiological reasons for this include: (1) inhibition of rhizobial infection of legume roots, decreasing nodule formation; (2) inhibition of nitrogenase enzyme activity in the nodule due to modification of the nitrogenase iron protein; (3) decrease in bacterial membrane potential and the inhibition of the leghemoglobin; (4) decrease in the supply of photosynthate to the rhizobium due to the poor supply of major nutrients, such as P; and (5) poor supply of Mo and Ca, which are essential for N_2 fixation.

A decrease in soil pH will increase the concentrations of Fe and Al in soil solution and this may lead to adsorption/precipitation of phosphate. In variable-charge soils, a decrease in pH increases the AEC, increasing the retention of phosphate. However, at very low pH, solubilization of P compounds may result in an increase in the concentration of P in soil solution. Under acid conditions, weathering liberates K from micaceous and feldspar minerals, enabling it to enter the soluble and exchangeable pools; but, in variable-charge soils, increasing acidity decreases CEC, reducing the ability of the soil to retain K, and this may cause K to be more prone to leaching.

In many soils, organic matter is the main source of S and, since mineralization of organic matter may be slowed by acidity, the release of S for plant uptake may decline when soil is acidified. Furthermore, in highly weathered acid soils, SO_4^{2-} may be adsorbed by sesquioxide surfaces, precipitated as Al-OH-SO_4-type minerals, such as alunite and basulminite, and/or held as an exchangeable anion on positively charged sites on sesquioxides.

One of the major consequences of acidification is a decline in basic cations such as Ca^{2+} and Mg^{2+}, leading to potential deficiency of these cations for plant growth. Furthermore, at low pH, the bioavailability of Ca may be restricted due to antagonistic effects of soluble Al. With increasing soil acidification, smaller amounts of Mg^{2+} remain in exchangeable form due to reduction in negative charge. Since Mg^{2+} is a poor competitor with Al^{3+} and Ca^{2+} for the exchange sites, it tends to accumulate in the solution phase and is therefore more prone to leaching.

Acidification affects the transformation of metal ions through: (1) changes in surface charge in variable-charge soils; (2) changes in metal speciation; and (3) shifts in the oxidation–reduction reactions of metals. In general, the solubility and phytoavailability of metals such as Cu and Zn are inversely related to soil pH. Since Cu is largely complexed with humic substances, the greater solubility of humic compounds at high pH may result in more soluble Cu when pH is raised, though this organically complexed Cu may not have low bioavailability. One of the major consequences of soil acidification is the increase in concentrations of soluble Al and Mn, both of which are highly toxic to plant growth. A primary aim of liming soils for agricultural production is to decrease the concentration of these elements. While Mn toxicity is directly related to the metabolic requirements of plants, the effect of Al toxicity appears to be largely manifested as malformation and malfunction of the root system, a syndrome which is exacerbated by low levels of solution Ca in acid soils.

Under acid conditions, boron (B) occurs in soil solution as the uncharged H_3BO_3 molecule and thus it is readily available in acid soil. Unlike other anions, Mo is highly insoluble in low pH conditions and it may become limiting when soil is acidified.

Cadmium (Cd) has been identified as a potential human health hazard, which enters the food chain through plant uptake. Adsorption of Cd^{2+} by soil decreases with decreasing pH for the following reasons: (1) in variable-charge soils, a decrease in pH results in a decrease in cation adsorption and (2) a decrease in soil pH is likely to result in an increase in free Cd^{2+} at the expense of the more strongly adsorbed hydroxy-cadmium species. In general, Cd uptake by plants increases with decreasing pH and consequently it is recommended that soil pH be maintained at pH 6.5 or greater in land receiving Cd-rich biosolids.

The effect of soil acidity on the adsorption of metalloids such as arsenic (As) and selenium (Se) is controlled by two interacting factors – pH-related changes in negative potential in the plane of adsorption and pH-induced changes in amount of negatively charged ionic species present in soil solution. While the first factor results in a decrease in As(V) adsorption when pH is raised, the latter factor is likely to cause the opposite effect. Thus, the effect of pH on As(V) adsorption depends on the nature of the mineral surface. In soils with low oxide content, increasing the pH will have little effect on As(V) adsorption, while in highly oxidic soils, adsorption will decrease with increasing pH.

Amelioration of Soil Acidity

Three approaches can be taken to minimize the rate and impacts of soil acidification: (1) reduce the amount of H^+ ions generated, (2) minimize the extent to which the H^+ and OH^- ions generating processes are uncoupled, and (3) neutralize the acidity produced. In managed ecosystems, the rate of acid generation can be altered by applying nutrients in forms that produce less acidity, selecting plant species that accumulate less cation, and reducing the loss of C, N, and S from soil.

Traditionally liming has been the most common practice used to alleviate soil acidification. A range of liming materials is available, including calcite ($CaCO_3$), burnt lime (CaO), slaked lime ($Ca(OH)_2$), dolomite ($CaMg(CO_3)_2$), and slag ($CaSiO_3$). Recently, the liming potential of other Ca-containing compounds has been evaluated. Some of these materials include phosphate rocks (PRs), FGD gypsum, fluidized bed boiler ash, fly ash, and lime-stabilized organic composts. The amount of liming material required to rectify soil acidity depends on the neutralizing value of the liming material and pH-buffering capacity of the soil.

Unlike soluble P fertilizers, PRs can have a liming value, because they contain some free $CaCO_3$, which itself can act as a liming agent and, secondly, the dissolution process of the P mineral component (i.e., apatite) consumes H^+, thereby reducing the soil acidity. While the free $CaCO_3$ in PRs dissolves reasonably fast, providing a small, immediate liming effect, the apatite generally dissolves at a slower (but variable) rate and has liming value over a longer period of time.

Alkaline-stabilized biosolids are increasingly being used as agricultural lime substitutes and soil amendments. Alkaline stabilization of biosolid utilizes a combination of high pH, heat, and composting to kill pathogens and stabilize organic matter. A range of alkaline materials are used for this purpose, including cement-kiln dust, lime-kiln dust, lime, limestone, alkaline coal fly ash, FGD, other coal-burning ashes, and wood ash. To minimize metal mobility and bioavailability in biosolid-amended soils, the US Environmental Protection Agency (USEPA) recommends the application of alkaline-stabilized biosolids and other liming agents to increase the soil pH to 6.5 or more.

The primary purpose of liming arable soils is to overcome the chemical problems associated with soil acidity that include high concentrations of acid ions (H^+ and Al^{3+}) and toxic elements (Mn^{2+}), and low concentrations of basic cations (Ca and Mg) and other nutrient ions such as Mo and P. The hydrolysis of the basic cations in lime produces OH^- ions, which neutralize the H^+ ions, thereby decreasing the activity and bioavailability of Al and Mn. Liming also increases the solubility of Mo and P, thereby increasing their availability. Lime provides the basic nutrient cations (Ca^{2+} and Mg^{2+}) and also reduces the solubility of heavy metals, thereby minimizing their bioavailability and mobility in soils.

Further Reading

Adams F (1984) *Soil Acidity and Liming*. Madison, WI: Soil Science Society of America Publishing.

Evangelou VP (1995) *Pyrite Oxidation and its Control.* New York, NY: CRC Press.

Haynes RJ (1984) Lime and phosphate in the soil–plant system. *Advances in Agronomy* 37: 249–467.

Longhurst JWS (1991) *Acid Deposition: Origin, Impacts and Abatement Strategies*. New York, NY: Springer-Verlag.

Marschner H (1995) *Mineral Nutrition of Higher Plants*, 2nd edn. London, UK: Academic Press.

Rengel Z (ed.) (2003) *Handbook of Soil Acidity*. New York, NY: Marcel Dekker.

Robson AD (ed.) (1989) *Soil Acidity and Plant Growth.* New York, NY: Academic Press.

Ulrich B and Sumner ME (eds) (1991) *Soil Acidity.* New York, NY: Springer-Verlag.

Wright RJ, Baligar VC, and Murrmann RP (eds) (1990) *Plant–Soil Interactions at Low pH.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

AERATION

D E Rolston, University of California–Davis, Davis, CA, USA

Introduction

In a general sense, aeration is the interchange of various gases between the atmosphere and the Earth and the various reactions that either consume or produce gases in the soil. The interchange results from concentration gradients established within soil by respiration of microorganisms and plant roots, by production of gases associated with biological reactions such as fermentation, nitrification, and denitrification, reduction–oxidation reactions of soil chemicals, and by soil incorporation of materials such as fumigants, anhydrous ammonia, pesticides, and various volatile organic chemicals from toxic waste sites. The two major gases associated with aeration are oxygen (O_2) and carbon dioxide (CO_2), where O_2 moves from the atmosphere to soil and is consumed, and CO_2 is produced in soil and moves from the soil to the atmosphere. Figure 1 indicates the general direction of the flow of gases within soil profiles. Soil aeration has been reviewed extensively over the years.

Soil-Air Composition

The amount of air or soil-air content is directly related to the bulk density of the soil and the amount of water in the soil profile. The bulk density of natural soil varies from approximately $1.0\,\mathrm{Mg\,m^{-3}}$ to $1.7–1.8\,\mathrm{Mg\,m^{-3}}$. Thus, the relative amount of void or pore space in the soil varies between approximately 30 and 60%. The soil pores or voids can be filled with either air or water. Therefore the soil-air content or air-filled porosity can vary between approximately 30 and 60%.

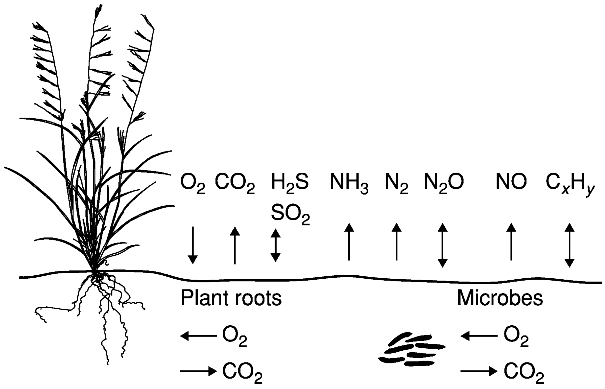


Figure 1 Schematic indicating the general flow directions of important gases within soil.

The composition of soil air depends on the relative magnitude of both the sources and the sinks of the various gas components, the interchange between soil air and atmospheric air, and the partitioning of the gases between the gaseous, liquid, and solid (mineral and organic matter) phases of the soil. If a soil were completely 'aerated' the concentrations of the gases in the soil air would be similar to that in the atmosphere. Oxygen concentrations in the soil air will be somewhat below that in the atmosphere (approximately 20% by volume), since O_2 is consumed in soil by plant root and microbial respiration and through chemical reactions. Under some conditions, O_2 concentrations can fall to zero and the soil becomes anaerobic (anoxic). It is now widely accepted that under some conditions soil profiles do not have to be either fully aerated or fully anaerobic but may be partially aerobic and partially anaerobic. Anaerobic pockets or 'hot spots' may exist within the soil due to pockets of very high O_2 consumption such as around incorporated carbon materials and/or due to very slow diffusion to regions of O_2 consumption. For example, the interior of large aggregates may be anaerobic for these reasons.

CO_2 concentrations in the soil air can be as high as 10 times more than in the atmosphere (0.036% by volume). Since nitrogen gas (N_2) is more abundant than other gases in the atmosphere (approx. 78%) and there are generally no sources or sinks for N_2 in the soil (except N_2 absorbed during nitrogen fixation or produced during denitrification), the concentration of N_2 in the soil air will be similar to that in the atmosphere, varying only slightly depending on the production and consumption of other soil gases. The soil air will also contain varying amounts of nitric oxide and nitrous oxide (from nitrification and denitrification); methane, hydrogen sulfide, and ethylene (from anaerobic processes); water vapor; and trace amounts of inert gases such as argon (Figure 1). Human activities also result in the accidental or intentional introduction of gases in the soil profile such as fumigants, anhydrous ammonia, pesticides, and various volatile organic chemicals that exist partially in the vapor phase.

Gas Exchange Mechanisms

Diffusion

Diffusion is considered to be the principal mechanism in the exchange of gases between the soil and the atmosphere. The diffusion velocities of gas mixtures in soil are related to each other in a complex manner dependent upon the mole fraction of each gas, the molar fluxes of each gas, and the binary diffusion coefficient of each gas pair. General equations for steady transport of a multicomponent mixture of gases have been developed based on gas kinetic theory. If gravity effects are ignored or diffusion occurs only horizontally, the well-known Stefan–Maxwell equations provide the theoretical framework for diffusion of gases in soils. Ficks law is generally applicable for only a few special cases. One of these cases is for the diffusion of a trace gas in a binary mixture, meaning that the mole fraction of the tracer gas is small. Since the binary diffusion coefficients of N_2 in air and O_2 in air are very similar and CO_2 may be considered to exist in trace amounts, the diffusion of the two major gases associated with aeration may come under this case. Diffusion of some of the other gases existing in soil may not meet this criterion, however.

Assuming that the special case conditions are met, Ficks law is given by:

$$\frac{M_g}{At} = f_{g,d} = -D_p \frac{dC_g}{dx} \qquad [1]$$

where M_g is the amount of gas diffusing (kilograms of gas), A is the cross-sectional area of the soil (square meters of soil), t is time (seconds), $f_{g,d}$ is the gas flux density (kilograms of gas per square meter of soil per second) due to molecular diffusion, C_g is concentration in the gaseous phase (kilograms of gas per cubic meter of soil air), x is distance (meters of soil), and D_p is the soil-gas diffusion coefficient (cubic meters of soil air per meter of soil per second).

The soil-gas diffusion coefficient is the main variable controlling the degree of soil aeration. It is highest when the soil is dry and approaches zero as the soil becomes very wet or saturated. The D_p has been related both empirically and theoretically to the soil-air content by a number of authors. Since discrepancies between measured soil-gas diffusion coefficients stemming from eqn [1] and those calculated from the many empirical equations occur, it is often desirable to measure the soil-gas diffusion coefficient for particular situations. Laboratory methods using soil cores and field methods for measuring soil-gas diffusion coefficients have been developed.

Convection

In addition to molecular diffusion processes, soil gases may also exchange with the atmosphere through convection (advection). Convective flow of gases means that the whole air parcel is moving through soil pores due to a pressure difference (gradient) between the soil and the atmosphere. Pressure gradients may develop due to barometric pressure changes in the atmosphere; wind blowing across the soil surface or against a hill or other landscape feature; infiltration of rainfall or irrigation water into the soil, soil-water redistribution, and evaporation; temperature differences across the upper part of the soil profile; and density differences due to high concentrations of gases that have densities much different from air.

Convective gas flow is described by a form of Darcy's Law:

$$f_{\mathrm{g,c}} = -\frac{K_{\mathrm{a}}}{\mu}\left[\frac{\mathrm{d}P}{\mathrm{d}x}\right] \qquad [2]$$

where $f_{\mathrm{g,c}}$ is the flux density of gas due to convection (cubic meters of gas per square meter of soil per second), K_{a} is the air permeability (cubic meters of gas per meter of soil), μ is the viscosity of the gas mixture (pascal-seconds), P is pressure (pascals), and x is distance (meters of soil).

Air permeability is strongly influenced by soil-water content. Air permeability is a maximum in dry soil and decreases as the soil becomes wet, until it reaches zero at saturation. This is caused by progressive blockage of the soil pores by water. Several field and laboratory methods for measuring K_{a} have been developed, involving either steady-state or non-steady-state flow, though steady-state measurements of gas permeability are preferred. Most field methods are based on the same principles as the laboratory methods, i.e., they involve measuring the flow rate of air through a soil column under known pressure differences across the column.

Convective processes occur rapidly and sporadically. Thus, it is very difficult to observe, measure, and predict the gas exchange that occurs by convection. During infiltration into soil, there is ample evidence that air pressure increases ahead of the water wetting front, and convection of gases occurs. It is often assumed that diffusion is the main gas exchange process overall, because it is operating continuously, whereas convection occurs episodically. Of the convective processes, rainfall and irrigation may contribute the most to aeration of soils, with estimates that rainfall may account for 7–9% of total aeration. Adequately quantifying these processes and being able to predict and model convective flow processes is a goal yet to be fully attained.

Gas Reactions

Respiration

O_2 is continuously consumed and CO_2 produced by plant roots and by soil microorganisms. Even plants such as rice that grow best in water-submerged soil transport O_2 from the leaves to the roots for respiration. The rates of O_2 consumption and CO_2 production are directly related to the rate of plant and microbial growth, which in turn is related to several environmental factors, including air and soil temperature, substrate availability, and soil moisture. For aerobic conditions, the amount of CO_2 produced and O_2 consumed tends to be about equal. For anaerobic conditions, the CO_2 production will tend to be larger than the O_2 consumption because other reactions are occurring.

The concentrations of O_2 and CO_2 that occur in the soil pore space vary widely, especially for CO_2, and depend on the rate of consumption and production and upon the rate that the soil is able to exchange these gases between the soil and the atmosphere through diffusion and convection. The diurnal and annual variability in the soil-gas concentrations are generally much greater for clayey soils than for sandy soils owing to the ability of sandy soils to transmit gases at a higher rate (larger soil-gas diffusion coefficients and air permeabilities) and maintain more constant concentrations.

Oxidation–Reduction Processes

Oxidation and reduction are connected with the transfer of electrons from soil organic matter (or organic contaminants) to oxidized inorganic compounds catalyzed by enzymes produced by soil microorganisms. For well-aerated conditions (aerobic), O_2 is the electron acceptor. When O_2 becomes limiting (anaerobic), other substances will accept electrons or be reduced. Examples of compounds that can be reduced under anaerobic conditions are nitrate (denitrification), manganic manganese, ferric iron, sulfate, and perchlorate (a natural and anthropogenic contaminant). The reduction of nitrate and sulfate results in N_2 (and N_2O) and hydrogen sulfide gases, respectively. Another gas produced from reduction processes is methane (CH_4). Both CH_4 and N_2O are strong greenhouse gases and contribute to global warming. In waterlogged or very wet soils and sediments that are unable to transmit O_2 sufficiently fast through the profile, O_2 is the first to disappear, followed by nitrate and then sequentially by the reduction of manganese, iron, and sulfate. Perchlorate is reduced at about the same point as nitrate. Several toxic organic chemicals also undergo redox reactions that greatly affect the kind and toxicity of the reaction products.

Production and Consumption of Other Gases

There are a few gases produced in soils that are not necessarily associated with redox reactions. When fertilizer materials such as urea and ammonium salts are applied to the soil, reactions can occur, produce ammonia gas (NH_3), particularly in soils with a pH of more than about 7.5 or 8, that fairly large emissions of NH_3 can occur, for instance, if urea is applied to the soil surface, since urea hydrolysis results in a large increase in pH near the fertilizer granules, with NH_3 being emitted. The deeper the material is placed, the less NH_3 will be emitted. Ammonia may also be produced in soil from the incorporation of animal waste.

Under aerobic conditions, ammonium-based materials either from fertilizer or mineralization of organic materials will be oxidized to nitrite and then to nitrate by microbial processes (nitrification). During nitrification, some N_2O and nitric oxide (NO) can be produced and emitted to the atmosphere. Nitric oxide enters into the tropospheric ozone cycle and can contribute to very small particle (less than 10 μm) generation in the atmosphere.

Besides O_2, gases including hydrocarbons, N_2O, NO, CH_4, and some sulfur gases may move from the atmosphere into the soil and be consumed by biological processes. Thus, the soil may act as a significant sink for atmospheric gases under some circumstances.

Aeration Requirements

Plants

The plant response to inadequate aeration or lack of O_2 is highly dependent upon the plant species, stage of growth, and upon several soil and environmental conditions such as temperature, water relations, and occurrence of toxic by-products of anaerobic conditions. The response of plants to inadequate aeration may be due to either direct or indirect effects. The direct effect is because of the lack of oxygen for root respiration. The indirect effect is due to changes in redox conditions that affect nutrient and water availability, soil pH, buildup of toxic concentrations of metabolites and metals, and the viability of pests and diseases.

The direct effect of lack of O_2 or slow diffusion of O_2 to plant roots has been characterized by a measurement called the oxygen diffusion rate (ODR). The ODR is measured by placing a cylindrical platinum electrode into the soil. Oxygen is reduced at the electrode surface and creates a current that can be measured. Diffusion of O_2 to the water-covered electrode is meant to mimic the diffusion through the water film around roots. Many studies have attempted to relate the ODR to plant response. ODR values smaller than approximately 0.2 μg cm^{-2} min^{-1} indicate potentially poor aeration for many plant species. Values of more than approximately 0.4 μg cm^{-2} min^{-1} are indicative of relatively good aeration for growth of most plants.

The plant symptoms of poor aeration are poor root growth or death, negative geotropism of roots, depressed shoot growth, wilting, leaf senescence, abortion of flowers, and termination of shoot apex growth. Poor aeration may also result in decreased transpiration; accumulation of ethylene, ethanol, and other metabolites; and decreased nutrient uptake. Reduced conditions in soil may also result in increased solubility of some chemicals that are toxic to plants.

Remediation of Contaminated Soils

With the large use of chemicals in modern society, both inorganic and organic chemicals, either intentionally or accidentally, end up in soil as contaminants with varying degrees of toxicity. Large amounts of petroleum products end up contaminating soil, both as point and nonpoint sources of pollution. The hydrocarbons in crude petroleum include alkanes, cycloalkanes, aromatics, polycyclic aromatics, asphaltines, and resins. In addition, chlorinated solvents, pesticides, detergents, metals, and other kinds of chemicals may pollute soil. Microbial processes in soil can degrade many of the organic compounds, but the biodegradability of various compounds is greatly influenced by their physical state and toxicity, as well as soil environmental conditions, including soil water content, soil pH, temperature, levels of inorganic nutrients, levels of electron acceptors, and aeration. Organic contaminants provide a source of carbon to microorganisms or they provide electrons, which the organisms can use to obtain energy. Metal contaminants may undergo redox reactions affecting their solubility and toxicity, which are also dependent upon the aeration status of the soil.

In most cases, biodegradation of organic contaminants depends on the activities of aerobic organisms. Thus, the presence of an adequate supply of O_2 in the soil is essential for biodegradation or bioremediation to occur. In general, it takes 2–3 kg of O_2 to degrade 1 kg of petroleum hydrocarbon. On the other hand, some compounds like the highly chlorinated chemicals are not easily degraded and may be broken down more effectively under anaerobic conditions. Some chemicals, perchlorate for instance, are only broken down under anaerobic conditions. Table 1 gives a list of some chemicals, indicating whether they are degradable by aerobic or anaerobic processes. For more information on biodegradation and bioremediation of contaminated soils, *see* **Pollutants:** Biodegradation in this encyclopedia.

Summary

Aeration is the interchange of various gases between the atmosphere and soil and the various reactions that

Table 1 Organic chemicals and their biodegradability

Chemical class	*Examples*	*Biodegradability*
Aromatic hydrocarbons	Benzene, toluene	Aerobic and anaerobic
Ketones and esters	Acetone, methylethyl ketone	Aerobic and anaerobic
Petroleum hydrocarbons	Fuel oil	Aerobic
Chlorinated solvents	Trichloroethylene, perchloroethylene	Aerobic (methanotrophs), anaerobic (reductive dechlorination)
Polyaromatic hydrocarbons	Anthracene, benzo[a]pyrene, creosote	Aerobic
Polychlorinated biphenyls	Arochlors	?
Organic cyanides		Aerobic

Adapted from Baker KH and Herson DS (1994) *Bioremediation*. New York: McGraw-Hill. © 1994 with permission.

either consume or produce gases in the soil. The composition of soil air depends on the relative magnitude of both the sources and sinks of the various gas components, the interchange between soil air and atmospheric air, and the partitioning of the gases between the gaseous, liquid, and solid (mineral and organic matter) phases of the soil. The two major gases associated with aeration are O_2 and CO_2, where O_2 moves from the atmosphere to soil and is consumed by plant roots and microorganisms and CO_2 moves from the soil, where it is produced by plant and microbial respiration, to the atmosphere. In addition to root and microbial respiration, O_2 may also be consumed by reaction with metals and other compounds. The two transport mechanisms that result in aeration of soils are molecular diffusion and convection. Although convection can result in significant transport under certain situations, such as infiltration of rainfall or irrigation water into soil, diffusion is considered to be the dominant mechanism of exchange over the long term. The plant symptoms of poor aeration are poor root growth or death, negative geotropism of roots, depressed shoot growth, wilting, leaf senescence, abortion of flowers, and termination of shoot apex growth. Aeration is also important for the soil's ability to degrade pollutants. In most cases, biodegradation of organic contaminants depends on the activities of aerobic organisms. Thus, the presence of an adequate supply of O_2 in the soil is essential for biodegradation or bioremediation to occur. On the other hand, some chemicals will only degrade under anaerobic conditions. Knowledge of a chemical's biodegradability and whether the chemical will degrade under aerobic or anaerobic conditions plays a major role in design of effective bioremediation schemes.

See also: **Anaerobic Soils**; **Carbon Emissions and Sequestration**; **Diffusion**; **Greenhouse Gas Emissions**; **Hydrocarbons**; **Oxidation–Reduction of Contaminants**; **Pollutants:** Biodegradation; **Remediation of Polluted Soils**; **Vadose Zone:** Hydrologic Processes

Further Reading

Baker KH and Herson DS (1994) *Bioremediation*. New York: McGraw-Hill.

Dane JH and Topp GC (eds) (2002) *Methods of Soil Analysis*, part 4, *Physical Methods*. Soil Science Society of America Book Series 5. Madison, WI: Soil Science Society of America.

Eweis JB, Ergas SJ, Chang DPY, and Schroeder ED (1998) *Bioremediation Principles*. Boston: McGraw-Hill WCB.

Gerstl Z, Chen Y, Mingelgrin U, and Yaron B (eds) (1989) *Toxic Organic Chemicals in Porous Media*. New York: Springer-Verlag.

Glinski J and Steniewski W (1985) *Soil Aeration and its Role for Plants*. Boca Raton, FL: CRC Press.

Hillel D (1998) *Environmental Soil Physics*. San Diego, CA: Academic Press.

Jury WA, Gardner WR, and Gardner WH (1991) *Soil Physics*. New York: John Wiley.

Scott HD (2000) *Soil Physics, Agricultural and Environmental Applications*. Ames, IA: Iowa State University Press.

AGGREGATION

Contents
Microbial Aspects
Physical Aspects

Microbial Aspects

S D Frey, University of New Hampshire, Durham, NC, USA

Introduction

Aggregates are the basic unit of soil structure, consisting of primary particles (sand, silt, and clay), organic materials in various stages of decay, and living organisms all bound together in clusters ranging in size from less than 2 μm to greater than 2 mm. The size, arrangement, and stability of aggregates is of primary importance in determining many soil physical, chemical, and biological properties, including soil water and air relations, microbial activity, the turnover of soil organic matter, and the release of plant-available nutrients. Good soil structure depends on the presence of aggregates that remain stable when disturbed by wetting–drying and freezing–thawing cycles and that contain a range of pore sizes. There should be sufficient small pores (0.2–30 μm) to retain water for plant and microbial growth, yet enough large pores ($>30\,\mu$m) to promote rapid oxygen diffusion, water infiltration, and drainage.

The activities of soil organisms are, to a large degree, responsible for the formation and stabilization of soil aggregates. During the breakdown of organic matter, decomposer organisms produce 'microbial glues' that bind soil particles together into aggregates of various sizes and stabilities, depending on the type and persistence of the binding agent. At the same time, soil structure regulates the movement and growth of soil organisms by affecting the size and degree of continuity of the soil pores within which soil organisms live. Thus there are strong interactions and feedbacks between soil organisms and soil structure. Research related to soil aggregation has significantly increased in the past decade as it has become increasingly clear that integrating soil structure into microbiological and ecologic studies is necessary to understand fully how soils function and how they can be managed sustainably.

Microorganisms and Aggregate Formation

The improvement of soil physical conditions brought about by the addition and incorporation of organic materials has been appreciated for centuries. However, organic matter additions have little effect on aggregation, unless microorganisms are present and actively decomposing the added materials. Fungi physically entangle soil particles in their hyphal networks, and both fungi and bacteria produce extracellular polysaccharides and other by-products of growth that cement soil particles together.

Studies with pure and mixed cultures of soil microorganisms indicate that a wide range of bacteria, actinomycetes, and fungi influence the formation and stabilization of aggregates in the presence of a suitable carbon source. The important role that microorganisms play in aggregate formation has been substantiated in a number of ways. Some of the earliest studies confirmed that aggregation could be brought about by adding microbial cells or the products of microbial biosynthesis to soil. For example, soil aggregation can be increased by adding bacterial or fungal extracellular polysaccharides that have been isolated from pure cultures or from soil. Another approach has been to observe changes in aggregation after soil has been treated with sodium periodate, a chemical that selectively oxidizes polysaccharides. Periodate treatment has been reported in numerous studies to reduce the quantity of stable aggregates, suggesting that microbial extracellular polysaccharides play a key role in aggregate stabilization.

When soil is sieved (e.g., $<250\,\mu$m) to remove large aggregates, amended with organic matter, and incubated in the laboratory under optimal moisture and temperature conditions, a significant number of new aggregates form rapidly, reaching a maximum within days to a few weeks. These newly formed aggregates often remain stable for a considerable period of time, even while the amount of microbial biomass declines (Figure 1), providing further support that by-products that remain after microbial growth subsides act as stabilizing agents. If a bacteriocide or fungicide

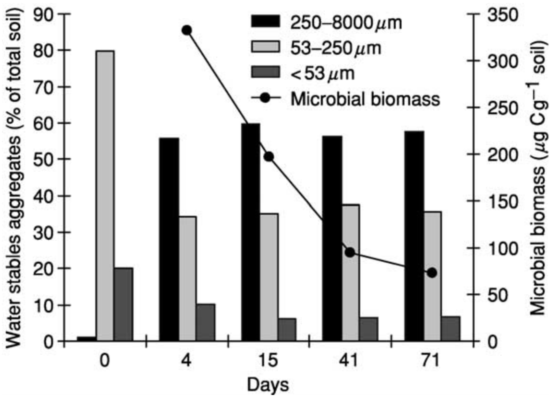


Figure 1 Percentage of water-stable aggregates following amendment of microaggregates (53–250 μm) with starch and incubation for 71 days. The closed circles represent the amount of microbial biomass in the macroaggregates (250–8000 μm) that formed during the incubation. Data from Guggenberger G, Elliott ET, Frey SD, Six J, and Paustian K (1999) Microbial contributions to the aggregation of a cultivated grassland soil amended with starch. *Soil Biology and Biochemistry* 31: 407–419.

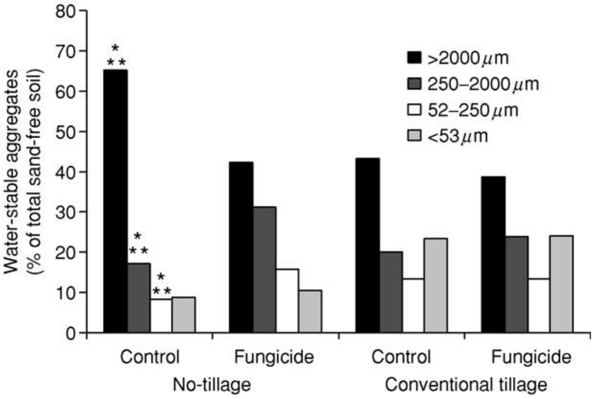


Figure 2 Effects of a fungicide on the percentage of water-stable aggregates in surface soils (0–5 cm) of no-tillage and conventional tillage plots at the long-term tillage comparison experiment in Horseshoe Bend, Georgia. Asterisks indicate significant differences ($P < 0.05$) among aggregate-size classes within a tillage treatment (**) and between tillage treatments within an aggregate-size class (*). Data from Beare MH, Hu S, Coleman DC, and Hendrix PF (1997) Influences of mycelial fungi on soil aggregation and organic matter storage in conventional and no-tillage soils. *Applied Soil Ecology* 5: 211–219.

is applied to the soil to inhibit bacterial or fungal growth, aggregate formation is suppressed. When already well-aggregated field soils are treated with a fungicide to inhibit fungal activity and growth, the level of aggregation is significantly reduced, especially in soils where fungi contribute significantly to the total microbial biomass, such as in no-tillage agroecosystems (Figure 2).

Aggregates are subjected to a variety of forces such as freeze–thaw and wetting–drying cycles that can lead to aggregate disintegration. Rapid wetting of dry aggregates, which occurs frequently at the soil surface during rain or irrigation events, can be particularly damaging. As water enters the soil pores, air becomes entrapped in the interior of the aggregate, leading to a pressure buildup which can cause the aggregate to fall apart (i.e., slake). Stable aggregates resist slaking and are thus termed 'water-stable.' Microbial exudates may influence aggregate stability by enhancing the ability of aggregates to withstand the pressure caused by slaking and/or by increasing the water repellency of the aggregates such that water entry into dry aggregates is significantly reduced or at least sufficiently slowed to allow entrapped air to escape.

Stable aggregates are not a random arrangement of soil particles, but are formed by a complex process involving a variety of binding mechanisms interacting across a range of spatial scales. According to the aggregate hierarchy concept, aggregate formation and stabilization occur in four distinct stages, with each stage dependent on a different type of organic binding agent and resulting in aggregates of increasing size (Table 1). At the lowest level of organization, individual clay plates flocculate into stable particles <2 μm in diameter in response to van der Waal's forces and hydrogen bonding. The presence of persistent binding agents consisting of inorganic and organic coatings on the clay surfaces may enhance attraction between clay plates and stabilize the particles against dispersion. These flocculated particles attach to and encrust living bacterial cells, fungal hyphae, and microbial debris (dead cells and extracellular products), forming stable particles of 2–20 μm in diameter. Electron

Table 1 Aggregate size classes and their primary binding agents

Aggregate class	*Particle size (μm)*	*Primary binding agents*
Macroaggregates	>250	Plant roots and fungal hyphae
Microaggregates	20–250	Plant and microbial cells and by-products (e.g., polysaccharides) encrusted with clay particles
Silt-sized particles	2–20	Bacterial and fungal debris encrusted with clay particles
Flocculated clay	<2	Amorphous aluminosilicates, oxides, and organic polymers on clay surfaces; electrostatic bonding

Adapted from Tisdall JM and Oades JM (1982) Organic matter and water-stable aggregates in soils. *Journal of Soil Science* 33: 141–163.

micrographs of soils clearly show that bacterial cells and fungal hyphae produce and are surrounded by an extracellular polysaccharide coating to which clay particles firmly attach (Figure 3a, b).

Microaggregates, particles 20–250 μm in diameter, are highly stable owing to their small size and to the persistent nature of the binding agents that hold them together. They resist destruction when subjected to rapid wetting or the disturbance caused by cultivation. This aggregate size class is critically important for the protection of organic matter sequestered inside.

Microaggregates are enmeshed and bound together into macroaggregates (greater than 250 μm) by a network of plant roots and fungal hyphae (Figure 3c, d). This largest class of aggregates is often separated into small (250–2000 μm) and large (greater than 2000 μm) macroaggregate fractions. The binding agents responsible for macroaggregate formation (roots and hyphae) are considered transient because they are readily decomposed by microorganisms. The temporary nature of macroaggregate binding makes this aggregate class especially susceptible to disturbance; therefore, the proportion of macroaggregates found in a particular soil is highly dependent on how that soil is managed. Soils that are regularly plowed generally have significantly fewer macroaggregates than uncultivated soils or cultivated soils where reduced tillage practices are employed.

The aggregate hierarchy model has been found generally to apply in temperate region soils where soil organic matter is a main binding agent. Aggregate hierarchy may be less evident in soils dominated by 1:1-type clays, and iron and aluminum oxides, where mineral interactions are the dominant stabilizing force. If aggregate hierarchy does exist, the 'porosity exclusion principle' proposed by Dexter should apply. Namely, large aggregates will have a greater total porosity compared with small aggregates, because they will contain

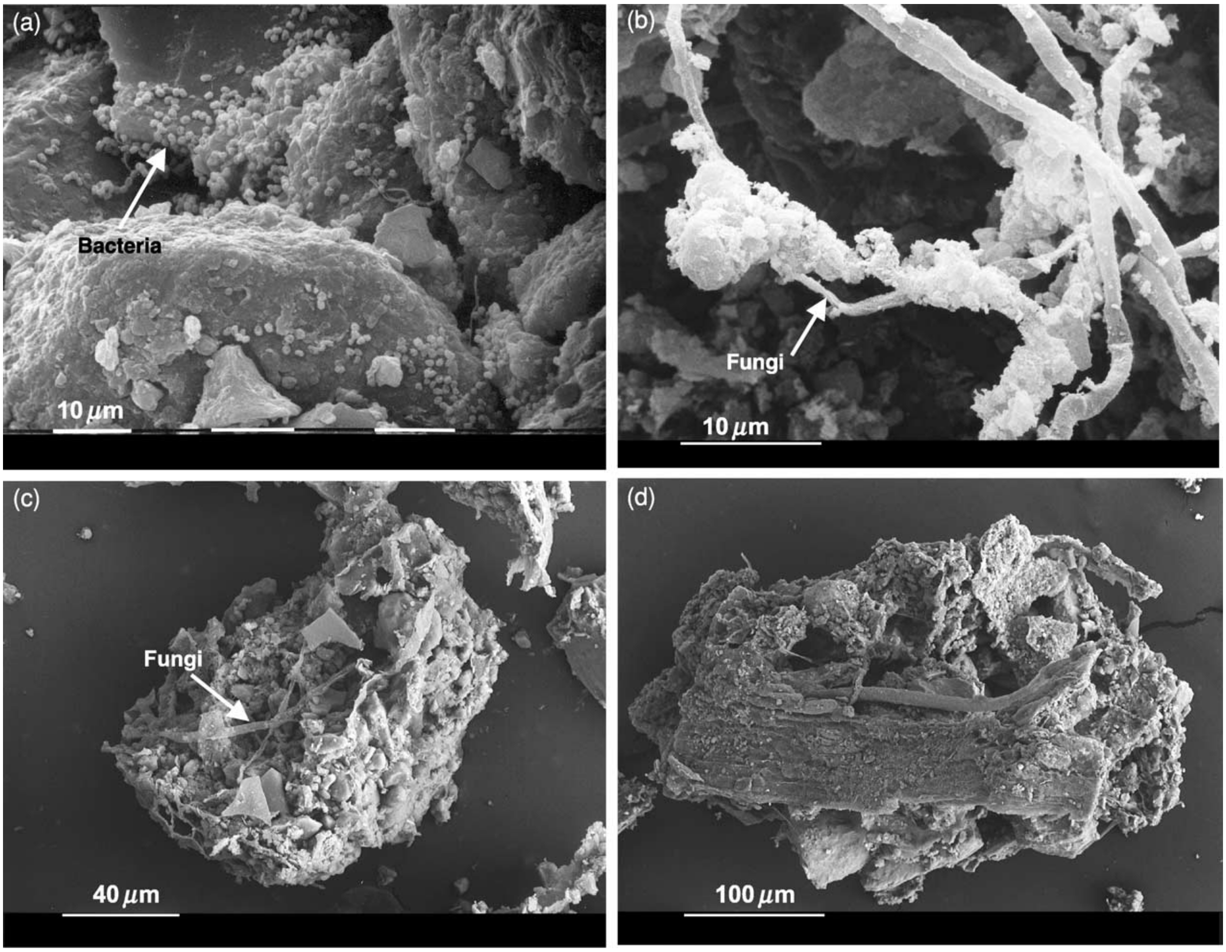


Figure 3 Scanning electron micrographs of (a) a colony of bacteria adhered to particle surfaces; (b) fungal hyphae encrusted with clay particles; (c) a microaggregate between 53 and 250 μm in size; and (d) particulate organic matter and soil particles bound together by roots, fungal hyphae, and microbial exudates into a macroaggregate (>250 μm). Courtesy of V.V.S.R. Gupta, CSIRO Land and Water, Glen Osmond, SA, Australia.

pores both within and between the smaller aggregates that comprise them.

Soil Aggregates as Habitats for Microbiota

The shape and arrangement of soil mineral and organic particles are such that pores of various shapes and sizes are created during the aggregate formation process. In a well-aggregated soil, up to 60% of the total soil volume will be comprised of pores that are either air- or water-filled depending on the moisture conditions. These pores may be open and connected to adjoining pores or closed and isolated from the surrounding soil. Thus soil at a microscopic scale is not dissimilar to a large cave system, with its complex maze of underground tunnels and chambers.

There are four categories of pores: micropores ranging in size from less than 0.2 to 10 μm and found inside microaggregates; pores of between 10 and 100 μm, located between microaggregates but within macroaggregates; pores between macroaggregates; and macropores created by roots and earthworms or by abiotic processes such as cracking or the shrinking and swelling of certain clay minerals when exposed to drying–wetting cycles. The size distribution and degree of continuity of soil pores depend to a large degree on soil texture. In clay soils, nearly 50% of the pores, which are often poorly interconnected, are less than 0.2 μm in size. Pores of 6–30 μm are most abundant in sandy soils and less than 20% are smaller than 0.2 μm. Less than 10% of pores in all soil types are greater than 150 μm.

Pores of different shapes, sizes, and degree of continuity provide a mosaic of microbial habitats with very different physical, chemical, and biological characteristics, resulting in an uneven distribution of soil organisms. Since soil organisms themselves vary in size, structural heterogeneity determines where a particular organism can reside, the degree to which its movement is restricted, and its interactions with other organisms. In this regard, it is the diameter of pore necks, or pore openings, rather than the enlarged section of pores that determines the location of soil organisms.

Every aggregate is a microcosm containing a highly variable microbial community of hundreds to thousands of different species of bacteria, actinomycetes, fungi, protozoa, and algae. The numbers and types of organisms vary from aggregate to aggregate and even between pores within a given aggregate. Mycelial fungi, unlike the other microbial groups, do not require a water film for growth and movement, and can therefore extend their hyphae across air-filled macropores, connecting aggregates and binding them together. Bacteria and protozoa, however, require water for motility and are thus largely restricted from movement between aggregates, since there is typically a discontinuous water film in the pore network, except when the soil becomes saturated following a precipitation or irrigation event.

Soil bacteria, which typically average 0.2–1.0 μm in size, can occupy both large and small pores; however, more than 80% of bacteria are thought to reside preferentially in small pores. The maximum diameter of pores most frequently colonized by bacteria is estimated to range from 2.5 to 9 μm for fine- and coarse-textured soils, respectively. There is a positive correlation between bacterial biomass and pores with a mean diameter of 1.2 μm. Few bacteria have been observed to reside in pores less than 0.8 μm in diameter, which means that 20–50% of the total soil pore volume, depending on soil texture and the pore-size distribution, cannot be accessed and utilized by the microbial community. Electron microscopy has revealed that bacteria often occur as isolated cells or small colonies (less than 10 cells) associated with decaying organic matter; however, larger colonies of several hundred cells have been observed on the surface of aggregates isolated from a clayey pasture soil and in soils under native vegetation (Figure 3a). Bacterial cells are often embedded in mucilage, a sticky substance of bacterial origin to which clay particles attach. Clay encapsulation and residence in small pores may provide bacteria with protection against desiccation, predation, bacteriophage attack, digestion during travel through an earthworm gut, and the deleterious effects of introduced gases such as ethylene bromide, a soil fumigant.

Fungi, protozoa, and algae are mainly found in pores larger than 5 μm. Fungi are commonly observed on aggregate surfaces (Figure 3c) and typically do not enter small microaggregates (less than 30 μm). Like bacteria, fungal hyphae are often sheathed in extracellular mucilage, which not only serves as protection against predation and desiccation, but also is a gluing agent in the soil aggregation process (Figure 3b). Mycelial fungi develop extensive hyphal networks and, as they grow through the soil and over aggregate surfaces, they bind soil particles and microaggregates (53–250 μm) together, thereby playing an important role in the formation and stabilization of macroaggregates (greater than 250 μm).

The relative abundance and distribution of bacteria and fungi vary across aggregate size classes within a given soil and between soils differing in clay content. Macroaggregates have been observed to contain higher total microbial biomass and higher fungal biomass in particular than microaggregates. Aggregates isolated from sandy soils tend to have a more even distribution of microorganisms than those from clayey soils. That is, bacterial cells and fungal hyphae occur on both the surfaces and inside aggregates from sandy soils, while microbes are largely concentrated on the surface of aggregates in clayey soils, with few microbes being present inside. This is probably due to the differential pore-size distribution of soils with different textures and especially the preponderance of pores less than 0.2 μm in diameter in clay soils.

The heterogeneous nature of the soil-pore network plays a fundamental role in determining microbial abundance, activity, and community composition by affecting the relative proportion of air- versus water-filled pores, which in turn regulates water and nutrient availability, gas diffusion, and biotic interactions such as competition and predation. Microbial activity, measured as respiratory output (i.e., CO_2 evolution), is maximized when approximately 60% of the total soil-pore space is water-filled. As soil moisture declines below this level, pores become poorly interconnected, water circulation becomes restricted, and dissolved nutrients which are carried by the soil solution become less available for microbial utilization. Soil drying leads to a reduction of microbial biomass, particularly in the larger pores, where organisms are subjected to more frequent alterations between desiccation and wetting.

At the other extreme, when most or all of the pores are filled with water, oxygen becomes limiting, since diffusion rates are significantly greater in air than through water. Gas diffusion into micropores is particularly slow, since small pores often retain water even under dry conditions. Restricted oxygen diffusion into micropores combined with biological oxygen consumption during the decomposition of organic matter can lead to the rapid development and persistence of anaerobic conditions. Thus survival of bacteria residing in small pores depends on their ability to carry out anaerobic respiration (e.g., denitrification), replacing oxygen with an alternative electron acceptor (e.g., nitrate, NO_3^-). More than 85% of the potential denitrifying activity of whole soil has been attributed to the microaggregate fraction, where most micropores are located.

Soil Structure and Microorganism Interactions

Ecologic interactions between soil organisms, such as competition and predation, regulate the flow of nutrients in ecosystems. For example, the release of plant-available nutrients such as nitrogen is stimulated when bacterial cells are consumed by protozoa. It is estimated that as much as 30% of the inorganic nitrogen released into the soil solution from decomposing organic matter is due to protozoan predation of bacteria; and more nitrogen is taken up by plants in soils containing protozoa compared with those without protozoa. The release of carbon from soils as CO_2 is also often enhanced in the presence of protozoa.

Soil heterogeneity indirectly influences nutrient-cycling dynamics by restricting organism movement and thereby modifying the interactions between organisms. For example, small pores influence trophic relationships and nutrient mineralization by providing refuges and protection for smaller organisms, particularly bacteria, against attack from larger predators (e.g., protozoa) that are typically unable to enter smaller pores. The location of bacteria within the pore network is a key factor in their survival and activity. Bacterial populations are consistently high in small pores, but highly variable in large pores, where they are vulnerable to being consumed. This may explain, in part, why introduced bacteria (e.g., *Rhizobium* and biocontrol organisms) often exhibit poor survival relative to indigenous bacteria. When they are introduced in such a way as to be transported by water movement into small, protected pores, their ability to persist is enhanced. This example stresses the importance of integrating structural aspects into soil microbiological studies.

Protozoa can consume 2000–12 000 bacteria per protozoan cell division and, since bacteria often occur as individual cells or small colonies, predation is greatest under conditions in which protozoa can readily move between and access a large number of pores. Thus predation rates are high if protozoan numbers are high and they are present in a large number of pores; whereas predation is low if protozoa are restricted to a few large pores. Under typical soil-moisture conditions, protozoan movement is restricted, since they require a continuous water film for migration between pores. Only at high soil-moisture contents (more than 60% of soil water-holding capacity) are protozoa free to move from pore to pore and perhaps from aggregate to aggregate. This explains, in part, why nutrient release is stimulated following a rain or irrigation event.

Protection of Organic Matter Conferred by Soil Aggregates

Bacteria and fungi are the primary decomposers in terrestrial ecosystems and most organic materials entering the soil must eventually pass through the microbial equivalent of the ‘eye of the needle’ on their way to becoming soil organic matter. Soil microorganisms also recycle the products that they themselves produce during microbial growth and biosynthesis, including cell debris from dead and decaying cells and substances exuded into the soil environment by living microorganisms (e.g., extracellular polysaccharides). They are also responsible for the further processing and turnover of various soil organic matter pools. These activities lead to the loss of carbon from the soil as carbon dioxide (CO_2), which ultimately diffuses out of the soil, contributing significantly to atmospheric CO_2 levels.

Soil structure is a dominant control on microbial decomposition processes and thus indirectly influences the amount of carbon and other nutrients released from decomposing plant material and soil organic matter. During decomposition, particulate organic materials, colonized by bacteria and fungi, become encrusted with microbial exudates and soil particles, initiating the aggregate formation process (Figure 3). In turn, newly formed and stabilized aggregates protect organic matter from further decomposition by controlling microbial access and activity.

That aggregates protect organic matter from decomposition is indirectly supported by experiments where CO_2 and nutrient release is monitored during the incubation of disturbed versus undisturbed aggregates. When macroaggregates are crushed, for example, there is a flush of microbial activity and the net release of CO_2 and nitrogen (Table 2), suggesting that previously protected organic materials are made more accessible for microbial attack when aggregates are disturbed. Aggregates collected from soils that were previously relatively undisturbed (e.g., soils under native vegetation or no-tillage agricultural management) generally exhibit the greatest response to crushing. Air drying of soils followed by rapid rewetting and incubation also results in a greater release of CO_2 compared with soils kept continuously moist, due partially to aggregate disruption caused by slaking.

More than 25% of the carbon stored in arable soils worldwide has been lost due to cultivation over the past century. The greatest impact of cultivation on aggregation is typically the loss of macroaggregates greater than 250 μm in size. Macroaggregate disruption and the subsequent release and decomposition of organic material once protected within the aggregate structure is one important mechanism by which carbon is lost from soils. There is growing interest in determining to what extent this trend can be reversed. By implementing management practices (e.g., no-tillage, cover crops, perennial vegetation) that promote the formation and stabilization of soil aggregates, it is thought that soils can sequester carbon and thereby mitigate, to some degree, the rapid accumulation of CO_2 in the atmosphere.

Soil as a Spatially Continuous Medium

Much of what is known about the relationships between soil structure, microorganisms, and microbial-mediated processes is based on studies where soil has been broken down into aggregates of different sizes. The isolated size fractions represent those aggregates that are resistant to the method of disruption employed and as such are arbitrary structures. Intact soil is a continuum of soil particles, pore spaces, organic

Table 2 Effect of crushing on the mineralization of carbon and nitrogen from macroaggregates

		Carbon mineralized (mg kg^{-1} soil)		*Nitrogen mineralized (mg kg^{-1} soil)*	
Collection site	*Incubation time (days)*	*Intact*	*Crushed*	*Intact*	*Crushed*
Sidney, NE, USA[a]					
Native grassland	20	1085	1129	68	94
Cultivated field	20	414	493	17	38
Saskatchewan, Alberta, Canada[b]					
Native prairie	14	508	581	45	64
Cultivated field	14	311	326	23	28
Horseshoe Bend, GA, USA[c]					
No-tillage	20	2186	2804	207	305
Conventional tillage	20	1361	1402	150	162

[a] Aggregates (300–2000 μm) were isolated from soil samples collected to a depth of 20 cm and incubated intact or crushed for 20 days (Elliott ET (1986) Aggregate structure and C, N and P in native and cultivated soils. *Soil Science Society of America Journal* 50: 627–633).

[b] Soil samples were collected to a depth of 15 cm at 10 points across each site. Macroaggregates (250–8000 μm) were incubated intact or crushed for 14 days (Gupta VVSR and Germida JJ (1988) Distribution of microbial biomass and its activity in different soil aggregate size classes as affected by cultivation. *Soil Biology & Biochemistry* 20: 777–786, with permission).

[c] Intact soil cores (0–5 cm) were separated into aggregate size fractions and the three macroaggregate size classes (>2000, 250–2000, and 106–250 μm) were incubated intact or crushed for 20 days. Data for the fraction >2000 μm are shown here and expressed on a sand-free aggregate basis (Beare MH, Cabrera ML, Hendrix PF, and Coleman DC (1994) Aggregate-protected and unprotected organic matter pools in conventional- and no-tillage soils. *Soil Science Society of America Journal* 58: 787–795).

materials, and organisms rather than a collection of discrete aggregates. Traditional aggregate-isolation techniques remove aggregates and their associated microbial communities from their spatial context and fail to capture the heterogeneity and connectivity of the pore network within which soil organisms live. Methods which combine soil thin-sectioning with image analysis, geostatistical tools, and mathematical models are now available to describe and quantify the spatial distribution of microbial cells in relation to soil particles and pore spaces. Such nondestructive approaches should provide a more complete understanding of soil microbial communities and the ecosystem processes they mediate.

List of Technical Nomenclature

CO_2	Carbon dioxide
NO_3^-	Nitrate

See also: **Cultivation and Tillage**; **Factors of Soil Formation:** Biota; **Structure**; **Tilth**

Further Reading

Degens BP (1997) Macro-aggregation of soils by biological bonding and binding mechanisms and the factors affecting these: a review. *Australian Journal of Soil Research* 35: 431–459.

Dexter AR (1988) Advances in characterization of soil structure. *Soil Tillage Research* 11: 199–238.

Elliott ET (1986) Aggregate structure and C, N and P in native and cultivated soils. *Soil Science Society of America Journal* 50: 627–633.

Elliott ET and Coleman DC (1988) Let the soil work for us. *Ecological Bulletins* 39: 23–32.

Gupta VVSR and Germida JJ (1988) Distribution of microbial biomass and its activity in different soil aggregate size classes as affected by cultivation. *Soil Biology & Biochemistry* 20: 777–786.

Lynch JM and Bragg E (1985) Microorganisms and soil aggregate stability. *Advances in Soil Science* 2: 133–171.

Miller RM and Jastrow JD (1990) Hierarchy of root and mycorrhizal fungal interactions with soil aggregation. *Soil Biology & Biochemistry* 22: 579–584.

Ranjard L and Richaume A (2001) Quantitative and qualitative microscale distribution of bacteria in soil. *Research in Microbiology* 152: 707–716.

Tisdall JM (1994) Possible role of soil microorganisms in aggregation in soils. *Plant and Soil* 159: 115–121.

Tisdall JM and Oades JM (1982) Organic matter and water-stable aggregates in soils. *Journal of Soil Science* 33: 141–163.

van Veen JA and Kuikman PJ (1990) Soil structural aspects of decomposition of organic matter by micro-organisms. *Biogeochemistry* 11: 213–233.

Vargas R and Hattori T (1986) Protozoan predation of bacterial cells in soil aggregates. *FEMS Microbiology Ecology* 38: 233–242.

Young IM and Ritz K (1998) Can there be a contemporary ecological dimension to soil biology without a habitat? *Soil Biology & Biochemistry* 30: 1229–1232.

Physical Aspects

J R Nimmo, US Geological Survey, Menlo Park, CA, USA

Published by Elsevier Ltd.

A soil aggregate is “a group of primary soil particles that cohere to each other more strongly than to other surrounding particles.” Soil aggregates form through the combined action of aggregation and fragmentation processes. That is, attractive and disruptive forces act on the particles in the soil to cause greater cohesion among some individual particles and groups of particles than others. Most soils break up naturally into some form of aggregate, as shown in Figure 1. Important physical aspects of aggregates include their size, density, stability, structure, and effect on the transport of fluids, solutes, particles, and heat.

The analysis of soil aggregation is important in a variety of applications. Aggregation is a major influence on the growth and effectiveness of roots. Aggregate stability and size information may be used to evaluate or predict the effects of various agricultural techniques such as tillage or addition of organic matter. Aggregate analysis is often used in experiments where various tillage methods are applied and then evaluated by examining the stable aggregates that result. Because of their direct relation to cohesive forces, aggregate size and stability are important to the understanding of soil erosion and surface sealing. Analysis of dry aggregates is logically related to wind-erosion effects, while wet analysis may be more appropriate to evaluate or predict erosion due to rainfall impact and runoff. The stability of wet aggregates can be related to surface-seal development and field infiltration, as water-stable cohesion among particles may lead to restriction of water entry and formation of surface seals. Through these erosion and sealing effects, as well as the relation between aggregation

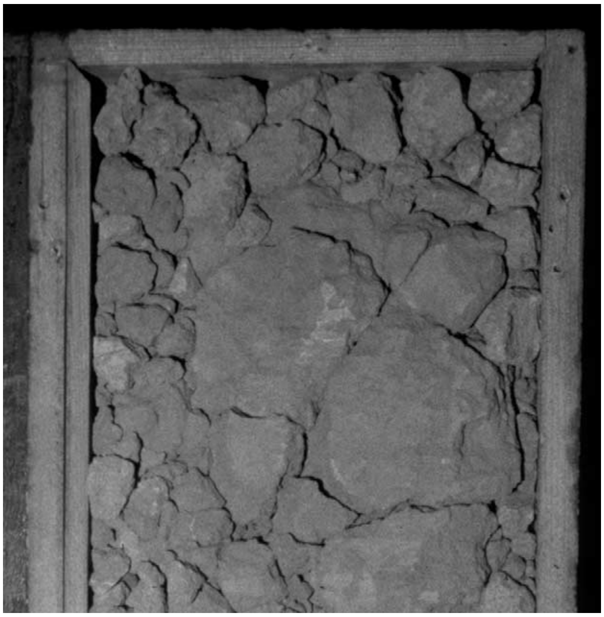


Figure 1 Soil with aggregates partially separated, in a tray. (From the Historic Russian Soil Collection of the St. Petersburg Academy of Forestry, provided by Jennifer Harden.)

and structural features such as macropores, aggregate analysis may help in the understanding of most aspects of soil water behavior, including runoff, infiltration, and redistribution, as well as soil aeration. Increasingly, aggregate properties are used in models that predict soil hydraulic properties, including water retention and unsaturated hydraulic conductivity.

Closely related terms include 'ped,' 'clod,' and 'crumb.' A ped is an aggregated unit representative of the innate structural classification of the soil. It has a characteristic shape related to structural designations such as prismatic, columnar, and blocky. The term 'clod' applies to an aggregate separated from the bulk soil by artificial means such as digging or plowing. A crumb is a small aggregate, typically less than 5 mm in diameter.

Forces on Soil Particles

The strength of interparticle cohesion depends on a variety of soil physical, chemical, and biological influences, some of the most important being air–water surface tension, intermolecular attractive forces between water and solids, cementation by precipitated solutes, entanglement by roots and fungal hyphae, and various chemical phenomena. The forces of soil cohesion depend strongly on water content and other conditions.

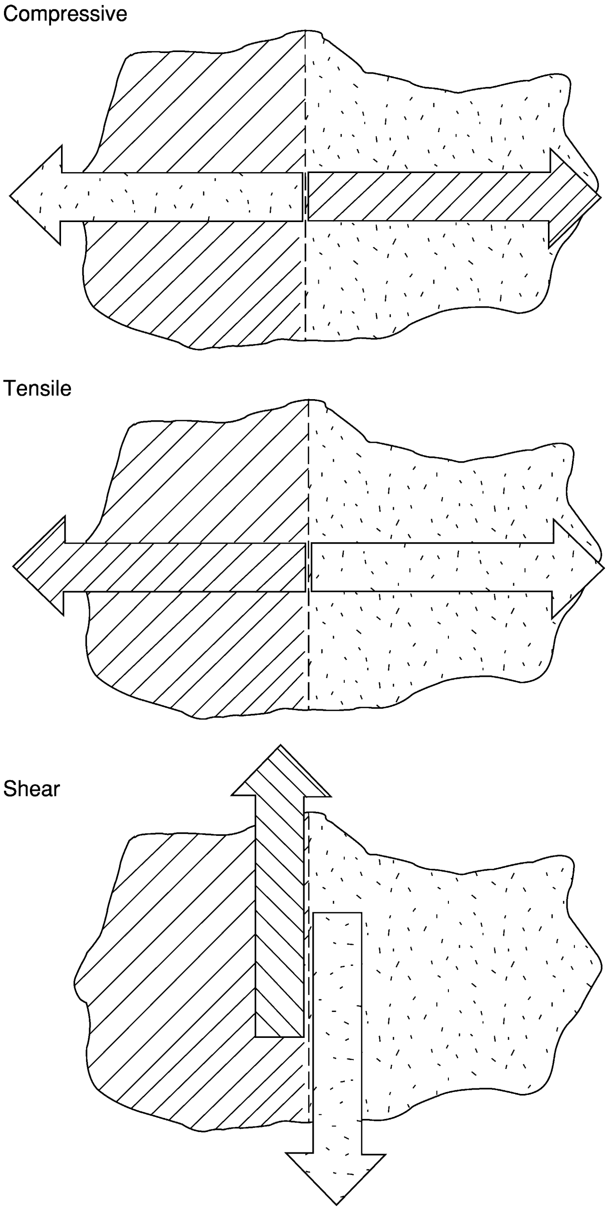


Figure 2 Types of stresses on an aggregate. Stresses are defined with respect to a selected plane, shown as the dashed line in this cross-sectional diagram. Each arrow indicates the force acting on the portion of the aggregate with the same type of shading, at the selected plane.

Fundamentally, the forces important in aggregation can be considered as stresses, that is, force per unit area acting on a given cross-sectional plane within the aggregate. These categorize as compressive, tensile, and shear stresses (Figure 2). Compressive stresses act to push soil particles closer together, as for example by the weight of soil above a given horizontal plane. Tensile stresses pull apart; for example, forces from soil shrinkage. Shear stresses act along a plane parallel to the direction of force, as in an aggregate at the edge of a zone of compaction. Tensile and shear stresses tend to disrupt aggregates; compressive stresses tend to consolidate aggregates, except that, when uneven across a plane, they lead to shear stresses that disrupt.

Several types of forces act to hold soil particles together. Water in the soil does this through surface tension, as well as through the attractiveness of water molecules for soil solids and for each other. Dissolved ions are important, especially in terms of the electrical double layer. The tendency of soil particles to have a negative surface charge means that water close to them is rich in positive ions, which in turn attracts other particles, in a process of flocculation. Because clay particles are especially sensitive to flocculating influences, higher clay content of a soil generally makes for more aggregation. Chemicals that precipitate or otherwise turn into cementing agents also play an obvious role. Various other chemicals, especially certain organic compounds such as polysaccharides, attract soil particles. Some organic materials just happen to be attractive; others, like roots and fungal hyphae, adhere to soil as part of their natural function. Because aggregation strongly affects air, water, roots within the soil, and other factors affecting plant growth, one may assume that plants have evolved so as to generate decay products that affect aggregation in ways favorable to the plant. A major part of this influence is to promote aggregation. Bacteria and other microorganisms contribute similarly to aggregation. To increase aggregation artificially by adding organic material to the soil is not straightforward. Not just the type and quantity of organic compounds but also their microscale distribution is critically important. Typically particles within an aggregate may be held together by a sort of glue made up of water, clay, and organic materials (Figure 3).

There is a similar variety of mechanisms that pull soil particles apart, either directly or by a decrease in attractive force. Some of the most common are associated with the addition of water. The breakup of aggregates that results from this, especially from sudden immersion, is called slaking. Increased water content can dissolve cementing precipitates and can decrease flocculation, while the resultant dilution weakens the effects of electrical double layers. As water infiltrates an aggregate, the expansion of trapped air, as well as the release of adsorbed air from newly wetted surfaces, can generate substantial disruptive force. Other disruptive mechanisms include the expansion of water upon freezing, impacts of rain or falling objects, and vibrations – either natural or artificial such as ultrasound or jostling on a sieve. Mechanisms associated mainly with compressive force disrupt by the generation of shear stresses; examples include foot or wheel traffic, which is always to some degree uneven across the land surface, and gravity acting on an uneven mass distribution of soil or on an aggregate unevenly supported from below.

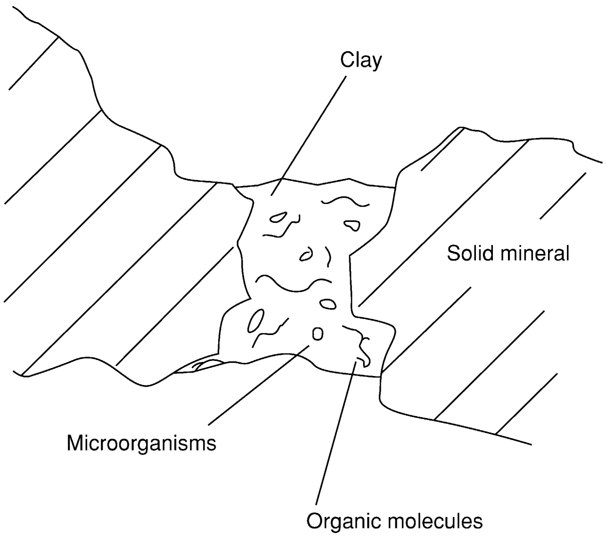


Figure 3 The microscopic region between solid grains within an aggregate.

Aggregates are less stable as they get bigger. This generalization applies within a given soil and should not be confused with the idea that soils forming larger aggregates have greater aggregate stability. One reason large aggregates are less stable is simply that interparticle forces vary, and the bigger the aggregate, the greater likelihood that it contains a plane-like region of low tensile strength where it breaks in response to stress. Similarly, the bigger the aggregate, the greater likelihood that it contains an expanding root or other agent that breaks it apart. Some disruptive stresses increase as the size of an aggregate increases, for example its weight increases with size. Forces due to shrinking and swelling are cumulative so that their net effect within the volume of a single aggregate is greater for a larger aggregate. These effects can be generalized by noting that attractive forces (cementing, intermolecular attraction, etc.) are predominantly short-range, whereas disruptive forces (which are mechanically transmitted through the soil fabric) are predominantly longer-range. Thus, as long as the same basic soil material is considered, the balance of forces within an aggregate increasingly favors disruption as larger aggregates are considered. This fundamental linkage between aggregate size and aggregate stability is a crucial factor in virtually any assessment of soil aggregation.

Whether a given sample of soil tends toward relatively large or relatively small aggregates depends chiefly on the interparticle attractive forces. This is because these forces vary more from soil to soil than do disruptive effects such as gravity and surface traffic, and therefore they dominate the issue of how the attractive and repulsive forces balance out. The soil's characteristic response to shear stresses – the extent to which it undergoes plastic in contrast to brittle deformation – is also important. Both interparticle attraction and plasticity depend strongly on soil texture and the composition of the soil–plant–water system. The aggregates of a fine-textured soil with much organic matter are likely to be larger than those of a sandy soil. The net interparticle attractive force in a sand can easily be so small that the characteristic aggregate size it indicates is smaller than the individual soil particles, so aggregates do not occur.

Aggregate Physical Properties and their Measurement

Basic Properties

Some physical properties of aggregates can be determined directly, especially where it is possible physically to isolate individual aggregates. Aggregate size and shape can be determined optically, by comparison with a ruled grid or by analysis of digital images. Because aggregates have irregular shape, their size cannot be indicated by a single linear dimension. Thus a choice must be made whether to indicate size by greatest dimension, average dimension, diameter of an equivalent sphere, or some other measure. The volume and bulk density of an aggregate can be measured by the clod method, using the aggregate's weight in air and in a liquid of known density, after coating it to prevent liquid intrusion. Alternatively, to reduce or eliminate the need for coating, a fine granular material of known bulk density may be used instead of a liquid. The strength of an aggregate can be measured, at least operationally, by breaking it with a known mechanical force applied by impact or by gradual increase in magnitude.

In most applications, attributes such as size, density, and strength are important as characterizations of the bulk soil, rather than of particular aggregates. Then, if the measurements are performed on individual aggregates, they must be made on enough aggregates to enable the determination of representative property values by statistical techniques. Alternatively, there is a wide range of methods that can be applied on aggregates in bulk.

Another important aspect is the internal structure of an aggregate. Soil within an aggregate may be more homogeneous than within a greater volume of soil, but, like any body of soil, it is not perfectly homogeneous. At one extreme, it might have a monolithic character not readily subdivided into units larger than individual particles. Alternatively, an aggregate may comprise smaller aggregates held by greater forces within themselves than between each other. In this way, each subdivision of aggregates may comprise a smaller subdivision of aggregates, down to a limit as the subdivided aggregates approach the size of particles. This sort of structure has led some scientists to propose fractal models for the structure of individual aggregates. Discussions below, on aggregate density dependence on aggregate size, and on mathematical representations of aggregate size distribution, explore this issue further.

Aggregate density and how it correlates with aggregate size can provide evidence for or against hypothesized forms of structure and data for predicting or correlating with other soil properties. Figure 4 shows an example of such data for soils of three different textures. For two of the soils, the smaller aggregates have greater density, which more closely approximates the particle density of soil minerals and indicates a tighter, more compact, and probably more stable structure. The material labeled 'sand' or 'quartz sand with pebbles' in the original figure has an aggregate density that has little size dependence and differs little from the particle density of pure quartz, approximately 2.65 g cm^{-3}. This indicates that this sand essentially has no aggregates, except

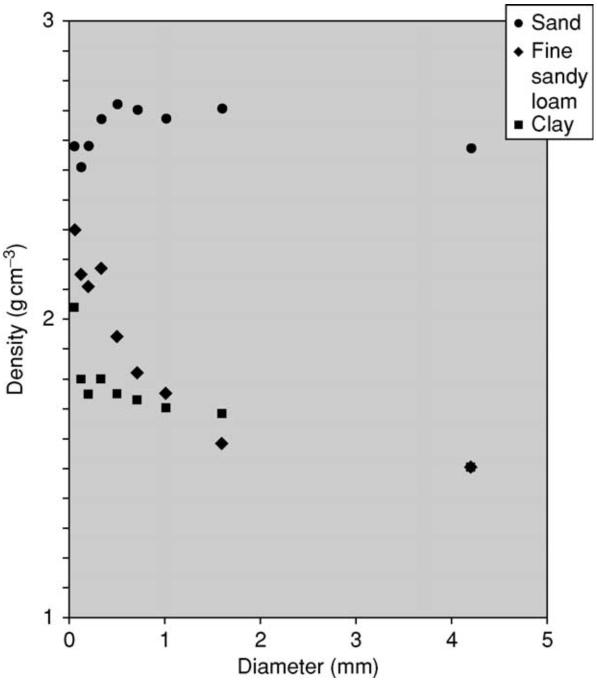


Figure 4 Measured aggregate density as a function of mean aggregate size for soils of three textures. (Source: Chepil WS (1950) Methods of estimating apparent density of discrete soil grains and aggregates. *Soil Science* 70: 351–362.)

possibly for some small aggregates (approx. 0.1 mm in diameter) of the smallest particles.

Size and Stability

Fundamentals Measuring physical aspects of aggregation of a coherent volume of soil containing a representative number of aggregates is the most common way to characterize the size distribution or stability of the soil's aggregates, but it is complicated by the interrelationships of aggregate properties. Especially important is that aggregate size and net cohesive force are conceptually inseparable. Measured sizes depend on the disruptive force applied to separate the aggregates, so force and size cannot be measured independently; all techniques characterizing these for a bulk soil involve some combination of both. For example, size determination by sieving cannot be done without the disruptive force of collisions between the aggregates and the sieve. Methods tend to be called stability methods or size methods depending on which of these gets more emphasis. This difference may be in the technique itself, for example a size method relying on a specified disruptive force or a stability method relying on the effect of force on a given size of aggregates. Alternatively, it may be in the interpretation of its results, for example whether a measured size distribution is presented as a distribution function or as a single index (such as an average aggregate size) that is considered to indicate stability.

Ideally, the size and strength of aggregates would be defined on a fundamental physical basis, in the way that hydraulic conductivity can be defined in terms of flux and potential gradients. In that case, any given measurement technique would provide an approximation to the defined ideal. Improved methods would produce results that are increasingly close approximations of the ideal. For aggregates, the definition would have to encompass both size and force, but the difficulty of quantifying the force prevents use of such a definition. Research on disruptive and cohesive forces may eventually solve this problem. Present characterizations, however, have to rely on operational definitions that endorse the result of a particular procedure with a particular apparatus, so the method cannot be separated from the definition.

In choosing a method for obtaining aggregate size and stability information, either stability or size distribution, and either wet or dry aggregates can be focused upon. The needs of the application should guide this choice. Erosion applications, for example, usually relate more directly to stability, while hydraulic and gas transport properties may relate more directly to the size distribution. The choice of wet or dry aggregates for measuring may depend on which condition most resembles the field situation, or on such considerations as reproducibility or consistency with other measurements.

Commonly used methods The forces applied to fragment or separate aggregates of the main bulk of soil are fundamentally artificial, though they may in some ways resemble forces in the natural setting. In the laboratory it is impossible to exert disruptive forces that exactly oppose the microscopic forces of cohesion, so practical methods rely on the variable and poorly known forces in a process that usually involves some combination of sieving, grinding, or vibration. Some methods make use of other phenomena that break aggregates apart, especially the forces involved when water or another liquid is introduced in relatively dry soil. In specifying the procedures that effectively define the aggregate characteristics, there are three realms of variables: the disrupting force or energy applied, the distribution of aggregates and particles, and the conditions of testing.

Some methods afford ways of quantifying some aspect of the applied force or energy. For example, the rupture-threshold approach considers one aggregate at a time, squeezing the aggregate between parallel plates while measuring both the applied force and the linear displacement. The drop-shatter method considers a known mass of nominally undisturbed soil, which is dropped from a known height on to a hard surface. The difference in potential energy associated with the distance of fall serves as an index of the energy applied to break apart the aggregated soil. The translation of this energy into energy that promotes sample rupture is imperfect, though it may be a useful index of the applied disruptive energy.

Some stability methods produce their own stability index as a result of the specified procedures and data analysis techniques, for example the fraction of soil weight that comprises stable aggregates. Stability interpretations may also be derived from aggregate-size distributions, usually by mathematically converting the tabular data or parameterized distribution formula to an average or other simple index. In this way the mathematical representation of size distribution not only serves as a convenience, but also provides the link, where needed, between size distribution and stability. In relating aggregate size to stability, the basic idea is that bigger aggregates imply greater stability. The most widely used index for this purpose is the mean weight diameter, defined as the sum of the weighted mean diameters of all size classes, the weighting factor of each class being its proportion of the total sample weight. Ideally this would be determined from integration of the cumulative abundance of aggregates as a function of diameter. The geometric mean diameter can also serve as an aggregate size index, though in recent literature it appears less than the mean weight diameter.

Miscellaneous methods Many techniques involve deliberate wetting or immersion of the sample. Wet compared with dry measurements on aggregates effectively measure different physical properties of the soil. It is not only the degree of wetness that is important, but also the means by which water has been applied. Wetting the soil in a vacuum, for example, reduces the disruptive forces associated with trapped air and thus produces larger aggregates.

Fast wetting with no vacuum involves immersion of air-dried aggregates in water for a period of time before beginning the mechanical sieving process. This type of wetting of dry aggregates produces disintegration and slaking, which may be undesirable. High-vacuum fast wetting involves de-airing aggregates in a vacuum chamber under high vacuum, then instantaneously wetting the aggregates in the chamber. It generally produces minimal disruption. Slow aerosol wetting, in which samples on screens are wet by vapor from below, produces little disintegration. Stabilities are higher and more reproducible with this type of wetting, in contrast to vacuum wetting. Wetting by slow wicking with or without a vacuum allows aggregates to draw moisture in from moist filter paper. As an alternative to water in initial wetting or other procedures, organic solvents such as methanol may reduce aggregate disintegration by slaking and may better preserve aggregate structure in drying.

A method derived from the older, 'high-energy moisture characteristic' method is based on differences in the water retention curves for fast-wetted and slow-wetted aggregates from replicate soil samples. Soil water retention is measured for thin beds of sieved aggregates on sintered glass. The wettest portion of the retention curve is considered, since the matric pressure range relatively close to zero is most affected by interaggregate pores. The key principle is that the less stable the aggregates, the more vulnerable they will be to disintegration during fast wetting, so the greater will be the difference between retention curves for the fast-wetted and slow-wetted samples. The method produces an index of aggregate stability on a dimensionless zero-to-one scale that is easily compared among different soils but is not directly comparable with other stability indices.

Ultrasonic dispersion can supply the disruptive force to associate with aggregate stability. The energy level that achieves a plateau in the quantity of aggregates remaining intact serves as an index of stability.

Stability is sometimes considered operationally in terms of the fraction of sample weight remaining after a prescribed sieving operation. Other methods measure the energy needed to break aggregates by crushing with parallel plates, as described above in connection with quantification of the energy of rupture. The results for a significant number of aggregates need to be reduced to a statistical representation indicative of the properties of the bulk sample. The energy required per increase in aggregate surface area can serve this purpose, as can a distribution function that indicates the probability of failure for a given applied rupture energy.

Issues of general importance The size distribution and stability of aggregates depend on numerous secondary factors, apart from the soil type. An obvious consideration as it is for spatial variability, expected to be substantial like most other soil properties. Aggregate stability can increase with storage time of the sample; it can also increase with increasing salt content of the water and is likely to decrease with temperature.

Soils with concretions (assemblages of primary particles that cannot be broken apart by the disaggregation processes of the chosen method) must be analyzed with respect to the application. In some cases the concretions may be treated as indivisible particles, because they are stable under normal cultivation practices; or in other cases as stable aggregates, because they usually have porosity, internal surface area, and substantial exchange capacity.

Some soils, especially from humid regions, may be nearly 100% stable in terms of the fraction remaining after prescribed sieving. Greater disruptive force, achievable by increasing the duration and amplitude of sieving, or by a more disruptive wetting technique, enables the detection of differences among highly stable soils, though with the drawback of precluding comparison with results obtained by more standard procedures on less-stable soils. The use of multiple methods increases the likelihood that comparisons will be possible among diverse soils.

Representation and Interpretation

To represent size distributions, the fraction of material at particular values of effective aggregate diameter can be graphed directly or cumulatively, as in Figure 5. For convenience in representation or for further mathematical development, these data can be fitted to a specific mathematical form.

Of various mathematical functions that have been used to fit the aggregate data, the lognormal distribution is one of the most useful and reasonably fits data from a variety of soils. Being a normal (Gaussian) distribution on a log scale, this distribution is skewed toward the small-diameter end of the range covered. It also has appropriate tapering-off of abundance at

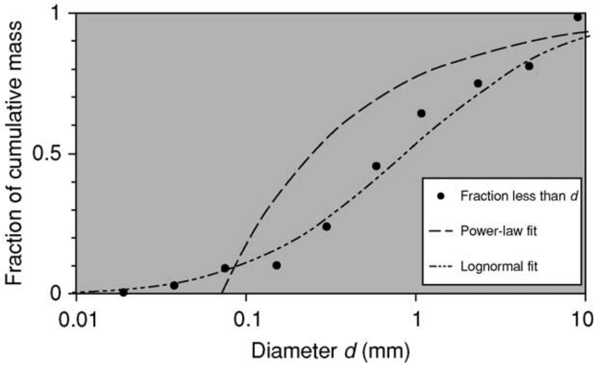


Figure 5 Measured and fitted aggregate size distribution for Sharpsburg silty clay loam. (Source: Wittmuss HD and Mazurak AP (1958) Physical and chemical properties of soil aggregates in a brunizem soil. *Soil Science Society of America Proceedings* 22: 1–5.)

both the small- and large-diameter extremes. The lognormal representation has also been used in some of the recent hydraulic property models that are based on aggregate properties.

Fractal interpretations have been applied to both aggregate stability and size distribution. A fractal characterization is valid if each subunit of the system is structurally identical (at a reduced scale) to the whole system. This idea is attractive for aggregates, since they are not made of primary soil particles on a fully equal basis. Larger aggregates may be thought of as being made of smaller aggregates that are more strongly bound internally than to each other. One attribute of fractal representation is that the cumulative number-size distribution can be represented as a power law; the cumulative abundance of objects greater than a given size is proportional to that size raised to some exponent. Like the lognormal model, a fractal model with an appropriate fractal dimension has a distribution skewed toward the small diameters. By fractal theory the power-law exponent is directly related to the mass fractal dimension, so this dimension may be known once the value of the exponent is established. Geometrically, the fractal dimension depends on the shape of the objects and the extent of fragmentation. Tests with the density, shape, and relative diameter variables represented fractally do not always show the degree of consistency over different scales that the most straightforward fractal models would predict. Natural aggregates may tend toward a monolithic internal structure or otherwise to deviate from true fractal character. Another shortcoming of fractal models is that they are fundamentally unrealistic at extremes of the range. Even so, fractal models remain useful for relating disruptive force to aggregate size, and in general for the modeling of relationships between mechanical properties and other soil properties and conditions.

Summary

The concept of an aggregate arises simply because some particles in the soil adhere more strongly than others. Physical aspects of aggregation are fundamental to the character, function, and behavior of soil. They give insight, with the possibility of much-needed quantitative insight, into soil structure.

The characterization of physical aspects of aggregation requires crucial tradeoffs. Little can be learned quantitatively about aggregates without operational definitions and criteria. Thus the specifics of measurement techniques are more closely bound to the consideration of aggregate physical properties than to many other aspects of soil. The choice between widely used, informally standardized methods and more novel ones often involves a substantial tradeoff between the need for consistency and the ultimate appropriateness of the method. The more standardized methods facilitate comparability, but the quantitative indices they generate may not give the sort of aggregate characterization most pertinent to the application at hand. Reliance on operational definitions also makes it awkward to incorporate ongoing scientific advances in conceptualizations and techniques. Research that leads to standardization of the specified force and a fundamental physical definition would help in allowing aggregate-measurement technology to advance without loss of comparability. The difficulty of this undertaking parallels the difficulty of bringing the general concept of soil structure into an objective, quantitative realm, but the potential benefits of even partial success justify much effort.

See also: **Aggregation:** Microbial Aspects; **Flocculation and Dispersion**; **Fractal Analysis**; **Structure**; **Swelling and Shrinking**

Further Reading

Ashman MR and Puri G (2002) *Essential Soil Science*. Oxford, UK: Blackwell.

Currie JA (1966) The volume and porosity of soil crumbs. *Journal of Soil Science* 17(1): 24–35.

Dickson JB, Rasiah V, and Groenevelt PH (1991) Comparison of four prewetting techniques in wet aggregate stability determination. *Journal of Soil Science* 71: 67–72.

Holden NM (1995) Temporal variation in ped shape in an old pasture soil. *Catena* 24: 1–11.

Kemper WD and Chepil WS (1965) Size distribution of aggregates. In: Black CA (ed.) *Methods of Soil Analysis*, part 1, *Agronomy*, pp. 499–510. Madison, WI: American Society of Agronomy.

Kosugi K and Hopmans JW (1998) Scaling water retention curves for soils with lognormal pore-size distribution. *Soil Science Society of America Journal* 62: 1496–1505.

Nimmo JR and Perkins KS (2002) Aggregate stability and size distribution. In: Dane JH and Topp GC (eds) *Methods of Soil Analysis*, part 4, *Physical Methods*, pp. 317–328. Soil Science Society of America Book Series No. 5. Madison, WI: Soil Science Society of America.

Oades JM and Waters AG (1991) Aggregate hierarchy in soils. *Australian Journal of Soil Research* 29: 815–828.

Rieu M and Sposito G (1991a) Fractal fragmentation, soil porosity, and soil water properties. I. Theory. *Soil Science Society of America Journal* 55: 1233–1238.

Rieu M and Sposito G (1991b) Fractal fragmentation, soil porosity, and soil water properties. II. Applications. *Soil Science Society of America Journal* 55: 1239–1244.

AGROFORESTRY

P K R Nair, University of Florida, Gainesville, FL, USA

Introduction

Agroforestry, a loosely defined term, involves the deliberate growing of trees and shrubs with crops and/or animals in interacting combinations for a variety of objectives. Such farming practices have been used throughout the world for a long time; but scientific attention was focused on them, and thus they attained prominence as a land-use practice, only since the late 1970s. Since then, substantial progress has been made in the science and practice of agroforestry. Today, acting as an interface between agriculture and forestry, agroforestry is considered to be a promising and sustainable approach to land use. The objective here is to present the essential features and forms of agroforestry and review the problems and prospects of its future development.

Emergence of Agroforestry as a Land-Use Approach

Why and how did this old form of land use attain prominence lately? During the mid-to-late 1900s, agriculture and forestry in the industrialized nations were transformed into commercial enterprises with emphasis on the production of single commodities. Certainly, it paid rich dividends. This apparently successful model was embraced by many newly independent nations of the developing world that were faced with the problem of feeding their millions. Several food-production technologies were developed according to this model and tried in the tropics. Some of them resulted in substantial increases in agricultural production, especially through the so-called Green Revolution of the 1970s. The traditional, mixed production systems of raising food crops, trees, and animals together, as well as exploiting a multiple range of products from natural woodlots, did not fit into the single-commodity paradigm, and were discouraged.

Serious doubts began to be expressed, however, about the relevance of single-commodity strategies and policies promoting them. In particular, there was concern that the basic needs of the poorest farmers, especially those in rural areas, were neither being considered nor adequately addressed. Soon it became clear that many of the technologies that contributed to the Green Revolution were not affordable to the poor farmer and that most tropical soils were unable to withstand the impact of high-input technology. The disastrous consequences of increasing rates of deforestation in the world's tropical regions also became a matter of serious concern. It was recognized that a major cause of deforestation was the search for more land to provide food and fuelwood for rapidly increasing populations.

The search for appropriate strategies to address these problems – strategies that would be socially acceptable, would enhance the sustainability of the production base, and would meet the need for production of multiple outputs – led to studies of age-old practices based on combinations involving trees, crops, and livestock on the same land unit. The inherent advantages of traditional land-use practices involving trees, such as sustained yield, environmental conservation, and multiple outputs, were quickly recognized. Agroforestry thus began to come of age in the late 1970s; the trend was institutionalized with the establishment of the International Council (Centre, since 1991) for Research in Agroforestry (ICRAF) in 1977, in Nairobi, Kenya. Since then, ICRAF has been the world leader in tropical agroforestry research.

In the temperate region, agroforestry has had a slower adoption rare than in the tropics. During the

1970s and 1980s, agroforestry as a concept had very little support in the industrialized countries. But the situation changed gradually and agroforestry gained better acceptance in the 1990s. For example, faced with the environmental consequences of agricultural and forestry practices that focused on the economic bottom line, the American public is now demanding greater environmental accountability for land-use practices and the application of more ecologically and socially friendly management approaches. Agroforestry fits well into that mold.

Agroforestry Systems

An agroforestry system normally involves two or more species of plants (or plants and animals), at least one of which is a woody perennial. The system has two or more outputs and a production cycle of more than 1 year, and both its ecology and economics are more complex than a monocultural system of agriculture or forestry. The essence of agroforestry can be expressed in four key 'I' words: intentional, intensive, interactive, and integrated. The term 'intentional' implies that systems are intentionally designed and managed as a whole unit, and 'intensive' means that the systems are intensively managed for productive and protective benefits. The biological and physical interactions among the system's components (tree, crop, and animal) are implied in the term 'interactive,' and 'integrative' refers to the structural and functional combinations of the components as an integrated management unit. It is often emphasized that all agroforestry systems are characterized by three basic sets of attributes: productivity (production of preferred commodities as well as productivity of the land's resources), sustainability (conservation of the production potential of the resource base), and adoptability (acceptance of the practice by the farming community or other targeted clientele).

In addition to agroforestry, several other 'forestry' terms became prominent in the 1980s, reflecting the increasing global interest in tree-planting activities. These include community forestry, social forestry, and farm forestry. Although these terms have also not been defined precisely, they emphasize people's participation in tree-planting activities. Social forestry refers to using trees and/or tree planting specifically to pursue social objectives, usually betterment of the poor, through delivery of the benefits of trees and/or tree planting to local people. Community forestry, a form of social forestry, refers to tree-planting activities undertaken by a community on communal lands or the so-called common lands; it is based on local people's direct participation in the process, either by growing trees themselves, or by processing the tree products. Farm forestry, a term used mainly in Asia, refers to tree planting on farms. The major distinction between agroforestry and these other systems is that, while agroforestry emphasizes the interactive association between woody perennials and agricultural crops and/or animals for multiple products and services, the other terms refer to tree planting, often as woodlots. In practice, however, all these labels directly or indirectly refer to growing and using trees to provide food, fuelwood, fodder, medicines, building materials, and cash income. Therefore, in common land-use parlance, these different terms are often used as synonyms, and sometimes even out of context.

A large number of traditional agroforestry systems have been recognized from different parts of the world. A major characteristic of these systems is their location-specificity. Each is a specific local example of the association or combination of the components, characterized by the plant species and their arrangement and management, and environmental and socioeconomic factors. In spite of the large variations among them, broad similarities can be recognized among the systems. For example, all these systems are composed of three basic groups of components or constituents: woody perennials (trees), herbaceous and other agricultural species (crops), and/or livestock. Agrisilvicultural systems involve trees and crops; silvopastoral systems involve trees and pasture/animals, and agrosilvopastoral systems involve all three major groups of components (crops, trees, and pasture/animals). Another way of describing the systems is based on the temporal arrangements of the components. Thus, 'simultaneous system' is used when the trees and crops exist together on the same unit of land during the same period of time; similarly, when the trees and crops are separated in time (such as when one alternates with the other), it is a 'sequential system.'

Because of the indigenous and traditional nature of agroforestry systems, there is some ambiguity in the terms used to describe them. All these systems consist, however, of a few specific patterns of component arrangements in space and time; these patterns are sometimes called agroforestry practices. In other words, although there are several hundred agroforestry systems, they all consist of a relatively few agroforestry practices. The same or similar such practices are found in various systems in different situations; the common practices and their brief descriptions are given in Table 1. Both the systems and the practices are known by similar names, but the systems are related to the specific locality or the region where they exist, or other descriptive characteristics that are specific to it. For example, growing

Table 1 Major agroforestry practices in the tropics and the temperate regions

Agroforestry practice	*Brief description*
Tropical agroforestry	
Alley cropping (hedgerow intercropping)	Fast-growing, preferably leguminous woody species grown in crop production fields; woody species are periodically pruned at low height (<1.0 m) to reduce shading of crops; prunings are applied as mulch into alleys as a source of organic matter and nutrients, or used as animal fodder
Taungya	Growing agricultural crops during the early stages of establishment of forestry (timber) plantations
Homegardens	Intimate multistorey combinations of a large number of various trees and crops in homesteads; livestock may or may not be present
Improved fallow	Fast-growing, preferably leguminous, woody species planted and left to grow during the fallow phase of shifting cultivation; the woody species cause site improvement and may yield economic products
Trees in soil conservation and reclamation	Trees on terraces and raisers with or without grass strips; use of trees for reclamation of saline, acidic, or otherwise degraded lands
Multipurpose trees (MPTs) on farms and rangelands	Fruit trees and other MPTs scattered haphazardly or according to some systematic planting arrangements in crop or animal production fields; trees provide fruits, fuelwood, fodder, and timber
Pasture under plantations (a form of silvopasture)	Cattle grazing on pasture under widely spaced rows of plantation species
Shaded perennial-crop systems	Integrated multistorey mixtures of tree crops such as coconut, cacao, coffee, and rubber with other tree crops, shade trees, and/or herbaceous crops
Protein banks (a form of silvopasture)	Production of protein-rich tree fodder on farms/rangelands for cut-and-carry fodder production
Shelterbelts and windbreaks	Use of trees to protect fields from wind damage, sea encroachment, and floods
Temperate-zone agroforestry	
Alley cropping	Trees planted in single or grouped rows with herbaceous (agricultural or horticultural) crops in the wide alleys between the tree rows
Forest farming	Utilizing forested areas for producing specialty crops that are sold for medicinal, ornamental, or culinary uses
Riparian buffer strips	Strips of perennial vegetation (trees/shrubs/grass) planted between croplands/pastures and streams, lakes, wetlands, and ponds
Silvopasture	Combining trees with forage (pasture or hay) and livestock production
Windbreaks	Row trees around farms and fields, planted and managed as part of crop or livestock operation to protect crops, animals, and soil from wind hazards

Source: Nair PKR (1993) *An Introduction to Agroforestry*. Dordrecht, The Netherlands: Kluwer; and Garrett HE, Rietveld WJ, and Fisher RF (eds) (2000) *North American Agroforestry: An Integrated Science and Practice*. Madison, WI: American Society of Agronomy.

coffee (*Coffea* spp.) under the shade of the timber tree, *Cordia alliodora*, in Costa Rica and growing cacao (*Theobroma cacao*) under the shade of coconuts (*Cocos nucifera*) in Kerala, India, are two examples of the same (or similar) practice, but they represent different systems. In spite of these differences, in common usage, the words 'systems' and 'practices' are used synonymously in agroforestry just as in other land-use forms. To make matters worse semantically, the same 'base' word is retained even after a traditional system or practice may have been improved through scientific intervention. For example, there are traditional fallows, enriched fallows, and improved fallows, and traditional intercropping and hedgerow intercropping. Another anomaly in the use of these terms is the variants or 'forms' of the same practice in different contexts. For example, the term 'alley cropping' as used in the tropics is different from temperate-zone alley cropping (Table 1), although conceptually they are similar.

Within the tropics, the complexity of agroforestry systems is greatest in the lowland humid and subhumid tropics, where climatic conditions generally favor rapid growth of a large number of plant species. Homegardens, shaded perennial systems (or plantation-crop combinations), and multilayer tree gardens are common in such regions with high human populations, and less intensive systems such as taungya and shifting cultivation are common in areas with less population density. In the semiarid tropics, the nature of agroforestry systems is also influenced by population pressure: homegardens and multilayer tree gardens are found in the relatively wetter areas; windbreaks and shelterbelts, and multipurpose trees on croplands are found in the drier regions. In the highland tropics (with favorable rainfall regimes), sloping lands and rolling topography make soil erosion an issue of major concern; consequently, soil conservation is one of the main objectives of agroforestry in these regions. Shaded perennial systems, use of woody perennials in soil conservation, improved fallows, and silvopastoral systems are the major forms of agroforestry in these tropical highlands. Several other specific systems also exist in the tropics: for example, apiculture with trees, aquaculture involving trees and shrubs, and woodlots of multipurpose trees.

Alley cropping, forest farming, riparian buffer strips, silvopasture, and windbreaks are the five major agroforestry practices recognized in North America (Table 1). Other temperate-zone agroforestry systems include ancient tree-based agriculture involving a large number of multipurpose trees such as chestnuts (*Castanea* spp.), oaks (*Quercus* spp.), carob (*Ceratonia siliqua*), olive (*Olea europa*), and figs (*Ficus* spp.) in the Mediterranean region. The 'dehesa' system of land use, involving grazing under oak trees with strong linkages to recurrent cereal cropping in rangelands, is also a very old system in this region.

The nature, complexity, and objectives of agroforestry systems vary considerably between the tropics and the temperate region. The climatic conditions in many parts of the tropics favor longer production cycles within a year and a large diversity of species, facilitating the existence of more numerous and diverse agroforestry systems in the tropics and subtropics than in the industrialized countries. Besides, socioeconomic factors such as human population pressure, availability of labor, land-holding size, land tenure, and proximity to markets have a major influence on the nature and form of agroforestry systems. As a consequence, considerable variations are found among systems existing in similar agroclimatic conditions. In general, small family farms, subsistence food crops, and emphasis on the role of trees in improving soil quality of agricultural lands are characteristic of tropical agroforestry systems. On the other hand, the driving force of agroforestry in the industrialized nations is environmental protection. The emphasis on monocultural production systems of agriculture and forestry in these countries has led to reduced biodiversity and loss of forest resources and wildlife habitat, increased erosion, nonpoint-source pollution of groundwater and rivers, greenhouse gas emissions, and social changes such as deterioration of family farms.

In summary, the primary objectives of tropical agroforestry systems are to exploit the role of trees on farms to provide products (such as fuelwood, poles, timber, animal fodder, food, fruits, and medicines) and ecosystem services (such as nutrient input and cycling, and soil erosion control). In the temperate zone, environmental amelioration and enhanced economic returns from tree-production systems are the key motivations.

Examples of Common Agroforestry Systems

Innumerable examples of agroforestry systems that exist in different parts of the world have been documented at various levels of detail. This section gives some examples of systems that have received research and development attention during the 1980s and 1990s.

Improved Fallow

The term implies the deliberate planting of species, usually legumes, with the primary purpose of improving soil fertility, mainly through nitrogen fixation and nutrient cycling, in a crop–fallow rotation. Growing herbaceous green-manure species in rotation with economic crops, usually called green manuring, is an age-old agricultural practice. The use of woody species as short-rotation fallows in the context of reduced fallow length in shifting cultivation cycles is a relatively new practice, and that is what is implied by the term 'improved fallow' in the context of agroforestry. The term 'managed fallow' is sometimes used to refer to such improved fallows in order to distinguish them from natural fallows characterized by colonization by natural vegetation, as in traditional shifting cultivation. Research on improved fallow has increased substantially during the 1990s, spear-headed by ICRAF, primarily in Africa. The main woody species used are leguminous, belonging to the genera *Sesbania*, *Tephrosia*, *Crotalaria*, *Mimosa*, and *Cajanus*.

Alley Cropping

In the tropics, alley cropping was developed in the 1970s and 1980s in an effort to find alternatives to long-fallow shifting cultivation. In this practice, arable crops are grown between hedgerows of planted shrubs and trees, preferably leguminous species that are periodically pruned to prevent shading to crops (Figure 1). The biomass, sometimes called prunings, obtained by pruning the trees and shrubs, is a source of mulch and green manure. Besides, leguminous woody species add nitrogen to the system through biological nitrogen fixation. The main motivation and rationale for alley cropping were the perceived soil-improving potential of fast-growing, leguminous woody species and their ability to withstand repeated cutting. It was hypothesized that by integrating such woody species into food crop production systems in simultaneous combinations, some of the benefits of natural fallows of shifting cultivation could be realized. Following that, many studies on alley cropping were undertaken in various parts of the tropics. Indeed, alley cropping was the

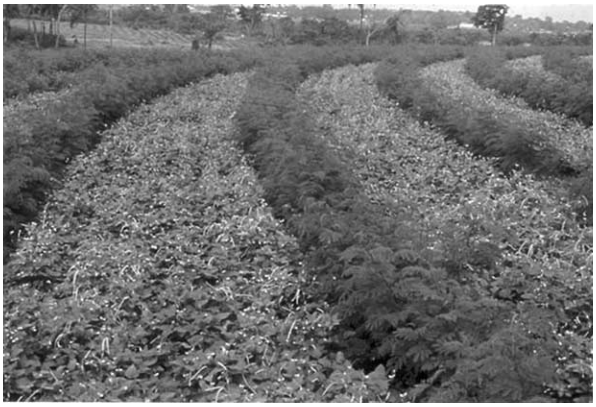


Figure 1 Hedgerow intercropping, commonly called alley cropping, is a much-researched tropical agroforestry technology. A large number of tree/shrub species have been tried as hedgerow species in association with a number of agricultural crops in different situations. This picture shows an alley-cropping field with *Leucaena leucocephala* in contour hedgerows, with cow pea (*Vigna unguiculata*) as the crop in the alleys, in Ibadan, Nigeria. Courtesy of B.T. Kang.

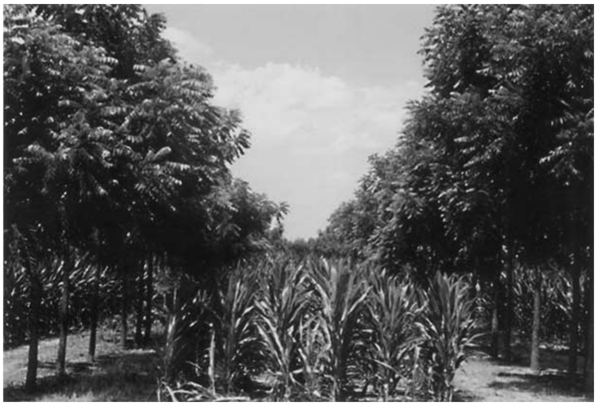


Figure 2 In the temperate region, alley cropping refers to growing agricultural (herbaceous) crops in the wide alleys between rows of trees; each tree row bordering the alley could be single-, double-, or multiple-row thick. Black walnut, *Juglans nigra*, is the most widely used tree for alley cropping in North America. The photo shows an alley-cropping plot of black walnut and maize (*Zea mays*) in Indiana, USA. Courtesy of S. Jose.

most widely researched topic in tropical agroforestry during the 1980s and much of the 1990s. In interpreting the results of these studies, some experts have used the data to defend alley cropping, others to denigrate it. A key issue concerning alley cropping is its ecological adaptability. The provision of nutrients through decomposing mulch, a basic feature of alley cropping, depends on the quantity of the mulch as well as on its quality and time of application. If the ecological conditions do not favor the production of sufficient quantities of mulch (as is the case in the dry tropics), there is no perceptible advantage in using alley cropping.

In temperate regions, alley cropping refers to growing herbaceous (agricultural) crops in the wide alleys between rows of trees that are planted in single or grouped rows (Table 1). Black walnut (*Juglans nigra*), a valuable timber- and nut-producing tree, is the species most widely used in alley cropping in North America (Figure 2). The common spacing adopted is 12.5 m between tree rows and 3 m between trees within a row (270 trees ha^{-1}). The driving force behind most temperate-zone alley cropping is economic benefits from the intercrop. Soil conservation in gently sloping lands could be an additional advantage. Quite a large amount of research has been conducted on various aspects of this system, especially in the eastern and Mid-western USA and Canada. Other similar systems in the temperate regions include the traditional intercropping practice used in the establishment of fruit and nut trees, including olive (*Olea* spp.) and grapes (*Vitis* spp.) in Europe, with pecan (*Carya illinoensis*) trees in south-eastern USA, and with paulownia (*Paulownia* spp.) trees in China.

Homegarden, Shaded-Perennial, and Multistrata Systems

These terms are used to describe a set of intimate plant associations in the highly populated parts of tropics. The word 'homegarden' is used rather loosely to denote diverse practices from growing vegetables in backyards, to complex multistrata systems (Figure 3). Homegardens can be found in almost all tropical and subtropical ecozones where subsistence land-use systems predominate. The choice of plant species in homegardens is determined by several factors: in addition to environmental and socioeconomic factors, dietary habits of people and market demands of the locality are important. The nature of species is remarkably similar among different homegardens in various places: fruit trees and other food-producing trees are the dominant components in most. The combination of crops with different production cycles provides an uninterrupted supply of food products throughout the year. These products are usually used for home consumption, with very little reaching the markets.

Growing shade-tolerant commercial species under the canopy of shade trees, commonly called shaded-perennial systems or plantation crop combinations, is another common agroforestry practice that is structurally similar to the homegarden. Traditionally, most of the so-called tropical plantation crops such as oil-palm (*Elaeis guineensis*), rubber (*Hevea brasiliensis*), coconut, cacao, cashew (*Anacardium occidentale*), tea (*Camellia sinensis*), and black pepper (*Piper

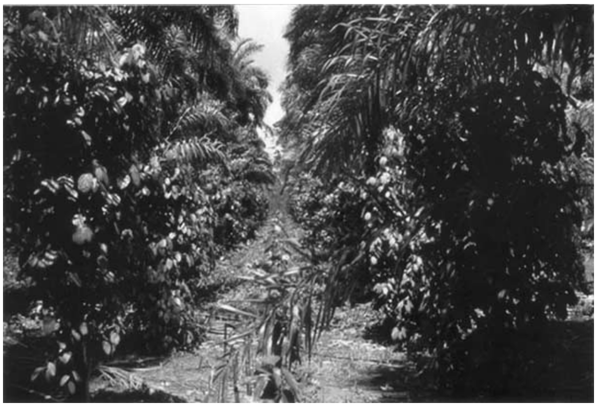


Figure 3 Multistrata systems involving commercial crops such as cacao (*Theobroma cacao*) and coconut (*Cocos nucifera*) are common in many parts of the tropics. The photo shows such a system including young cacao plants in between rows of peach palm (*Bactris gasipaees*) in association with black pepper (*Piper nigrum*) in south-eastern Bahia, Brazil.

nigrum) were developed as monocultural production enterprises. Contrary to popular belief, however, substantial areas under these crops are grown by smallholder farmers who cultivate them in association with a number of other crops. Crop combinations with coconut and coffee and cacao under shade trees are two systems on which considerable research has been done in South Asia and Central America respectively (Figure 4). These plantation crop mixtures ensure efficient use of available resources such as water, nutrients, and sunlight, as well as production of different products throughout the year. Moreover, predominance of perennial rather than annual crops in the system results in a relatively high ratio of nutrients stored in the vegetation to those stored in the soil; this ensures an effective nutrient cycle and relatively small hazard for leaching and erosion.

Silvopastoral Systems

This term is used to denote animal production systems in association with trees; these represent a major form of agroforestry in both tropical and temperate environments. Both fodder-producing and commercial timber species of trees are used in these systems. The systems in many temperate regions such as Australia and New Zealand, southern Europe, and south and south-eastern USA involve mostly pasture and animals under commercial tree plantations (Figure 5). In these cases, where the trees do not have any fodder value, the main objective is to enhance the economic return from the enterprise during the early growth of plantation when there is no economic return from timber sales. Tropical silvopastoral systems include both grazing systems of pasture under commercial tree stands as in the temperate regions and growing and managing fodder-producing trees on farmlands or as community woodlots. In the latter case, trees and shrubs

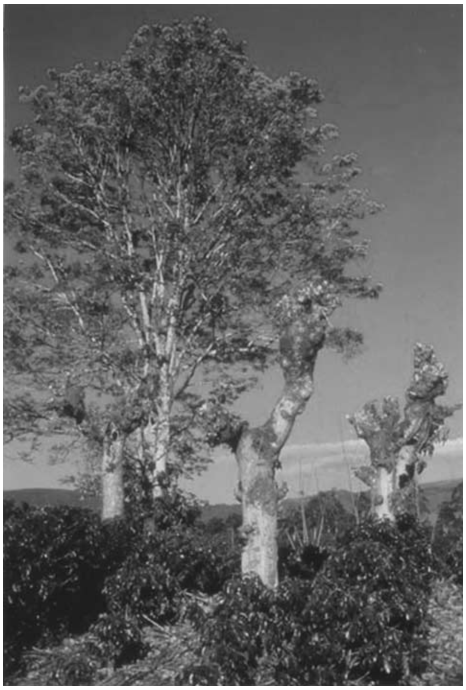


Figure 4 Shaded-perennial crop systems involve growing shade-tolerant commercial species such as coffee (*Coffea* spp.) and cacao (*Theobroma cacao*) under overstorey shade trees. The photo shows coffee under *Erythrina poeppigiana*, a leguminous tree in Costa Rica. The tree is pollarded (pruned) two to three times a year to regulate shade received by the coffee bush. Pollarded tree trunks are in the foreground, while a nonpollarded tree is in the rear. Courtesy of R.G. Muschler.

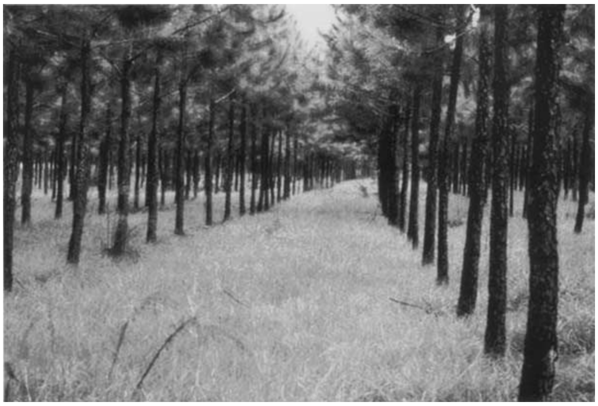


Figure 5 Various types of silvopastoral systems exist in different parts of the world. This photo, from Florida, USA, shows a stand of widely spaced rows of slash pine (*Pinus elliottii*) with understorey growth of forage grass; the system allows cattle grazing for many years after plantation establishment.

that are known for their excellent fodder value are grown in intensive or extensive systems. Intensive systems include cut-and-carry systems or protein banks (Table 1) where the trees and shrubs are grown in block configurations or along field and plot boundaries or other designated places, and the foliage is lopped periodically and carried for stall-feeding of animals. Sometimes such fodder species are also grown in alley-cropping combinations. The fodder trees grown on field and plot boundaries may be used as live-fence posts, which are then managed as in cut-and-carry systems. Because of the importance of livestock in both subsistence and commercial production systems, tree-fodder and silvopastoral systems are a major area of research in agroforestry.

Agroforestry Applications for Soil Conservation and Reclamation

Direct or supplementary use of trees and shrubs to control soil erosion is a widespread agroforestry practice in both tropical and temperate regions. Direct use encompasses increased soil cover by trees and shrubs, permeable hedgerow barriers, natural terrace formation by soil accumulation up-slope of hedgerows, and increased soil resistance to erosion by maintenance of organic matter. In supplementary use, the trees and shrubs are not the primary means of checking erosion, but they support other soil conservation structures through the stabilizing effect of the tree root system, while yielding valuable products such as fruits, animal fodder, and firewood. Use of trees and shrubs as a windbreak is a major form of agroforestry practice for soil conservation in drier and wind-prone areas. Windbreaks are narrow strips of trees, shrubs, and/or grasses planted to protect fields, homes, canals, and other areas from the wind and blowing sand. Shelterbelts are a type of windbreak consisting of long, multiple rows of trees and shrubs, usually along the seacoast, to protect agricultural fields from inundation by tidal waves. Riparian buffer strips formed by planting strips of perennial vegetation (of trees, shrubs, and grasses) in between crop fields or pastures and water bodies such as streams, ponds, lakes, and wetlands, are used increasingly in North America to rehabilitate degraded waterways that drain such heavily fertilized lands. Investigations on multiple benefits of stream rehabilitation and bioremediation using multispecies riparian buffer strips in the Midwestern USA have shown the substantial buffering capacity of such vegetated riparian buffers to remedy nonpoint-source contamination of waterways.

Other Agroforestry Applications

In addition to the common types of practices and systems described above, there are many others that have been relatively little studied. These include extensive silvopastoral systems, fuelwood lots, scattered multipurpose trees on farmlands, and tree planting for reclamation of wastelands and problem soils. For example, establishing woodlots of multipurpose (fodder- and fuelwood-yielding), salt-tolerant trees is widely practiced for reclaiming extensive tracts of salt-affected soils in north-western India. The parkland system is an extensive tree-intercropping system in sub-Saharan Africa that originated from the traditional shifting cultivation. During clearance of forest for agricultural production, farmers conserve trees that grow on crop fields for shading, fodder, fruits, and medicines. The trees that are most common include *Faidherbia albida*, *Parkia biglobosa*, and *Vitelleria paradoxa*. Other agroforestry systems involving a whole host of trees and production of a variety of nontimber forest products such as fruits, food, tannins, medicines, resins, and honey are present in several tropical and some temperate regions. The role of agroforestry in the buffer zones around protected parks and bioreserves is another aspect that is often mentioned, but not studied in deserving seriousness.

Forest farming is another practice, recognized especially in temperate-zone agroforestry (Table 1). It refers to growing specialty crops and products in forests and other wooded areas on a regular annual basis. Most notable among them in North America include cultivation of ginseng (*Panax* spp.) for medicinal purposes and the pharmaceutical industry, ferns and other ornamental species, mushrooms and other specialty food products, and pine straw for mulching. The concept of forest farming is not unique, however, to the temperate regions. As mentioned earlier, agroforestry is also used as an approach to exploiting a large number of nontimber forest products around the tropics.

Several indigenous agroforestry systems involve a multitude of such lesser-known species, especially woody species. They have come to be known as 'multipurpose trees' or 'multipurpose trees and shrubs.' The exploitation of these species and the agroforestry practices involving them have wide implications in food security and environmental protection, as well as conservation and use of genetic resources to meet current and future needs. The main difficulty in basing large-scale development plans on these practices is that the available qualitative descriptions by themselves often do not provide the necessary technical basis for preparing sound projects that involve large-scale investments.

Agroforestry and Ecosystem Services

Soil Productivity and Protection

One of the main conceptual foundations of tropical agroforestry is that trees and other vegetation improve the soil beneath them. Observations of interactions in natural ecosystems and subsequent scientific studies have identified a number of facts that support this concept. Research results during the past two decades show that three main tree-mediated processes determine the extent and rate of soil improvement in agroforestry systems. These are: (1) increased N input through biological nitrogen fixation by nitrogen-fixing trees; (2) enhanced availability of nutrients resulting from production and decomposition of substantial quantities of tree biomass; and (3) greater uptake and utilization of nutrients from deeper layers of soils by deep-rooted trees.

Nitrogen-fixing trees (NFTs) are a valuable resource in agroforestry systems. Some of the widely held assumptions about their benefits could, however, be wrong or incomplete. Because of methodological difficulties in quantifying N_2 fixation, especially in older trees, our understanding of the extent of N_2 fixation, and, therefore, the benefit that is actually realized by using NFTs in agroforestry systems is unsatisfactory. Furthermore, it is not clearly understood how much of the N_2 that is fixed by an NFT is actually utilized or potentially made available to an associated crop during its growth cycle, and how much goes into the soil's N store for eventual use by subsequent crops.

Biomass decomposition patterns vary greatly among agroforestry tree species. Several biomass (litter)-quality parameters, based on the chemical composition of plant tissues, have been developed to interpret these patterns: ratios of C to N, polyphenols to N, lignin to N, and (polyphenols + lignin) to N. Using this information, we can develop management strategies to manipulate the decomposition of plant biomass in agroforestry systems, thereby regulating the rates of nutrient release in the short term, and improving soil fertility, via improved soil organic-matter status, in the long term.

Compared to our knowledge of aboveground biomass additions in agroforestry systems, much less is known about the dynamics of belowground biomass. Experimental evidence shows that deep roots of trees can take up subsoil nitrate from beyond the rooting depth of crops. There seems to be little potential, however, for tree roots to take up immobile nutrients such as P from below or beyond the root zone of interplanted crops. Furthermore, the role of trees in the uptake of nutrients other than N and P is little studied in agroforestry systems.

The other major avenue of soil improvement through agroforestry is through soil conservation. When properly designed and managed, agroforestry techniques can contribute to ecosystem protection and restoration functions by reducing water and wind erosion and enhancing soil productivity.

Carbon Sequestration

Considerable interest has been evinced lately in the scientific community about the carbon sequestration potential of agroforestry. Agroforestry systems such as fuelwood plantations, shelterbelt/windbreak systems and woodlots may have the potential to sequester C, or offset fossil fuel emissions by substituting sustainably produced fuelwood and fodder. Based on a preliminary assessment of national and global terrestrial C sinks, two primary beneficial attributes of agroforestry systems have been identified: (1) direct near-term C storage (decades to centuries) in trees and soils; and (2) the potential to offset immediate greenhouse gas emissions associated with deforestation and subsequent shifting cultivation. A projection of carbon stocks for smallholder agroforestry systems indicated C sequestration rates ranging from 1.5 to 3.5 Mg C ha^{-1} $year^{-1}$ and a tripling of C stocks in a 20-year period, to 70 Mg C ha^{-1}. According to one estimate, median carbon storage by tropical agroforestry practices is around 9, 21, and 50 Mg C ha^{-1} in semiarid, subhumid, and humid ecozones, respectively. The total carbon emission from global deforestation at the currently estimated rate of 17 million ha per year is 1.6 Pg. Assuming that 1 ha of agroforestry could save 5 ha from deforestation and that agroforestry systems could be established in up to 2 million ha in the low-latitude (tropical) regions annually, a significant portion of carbon emission caused by deforestation could be reduced by establishing agroforestry systems. These are rough estimates however. C sequestered in agroforestry systems varies with a number of site- and system-specific characteristics, including climate, soil type, tree-planting densities, and tree management. The key point is that agroforestry could have a major role to play in the global terrestrial C budget; this is an area that needs to be researched in more depth.

Other ecosystem benefits such as water-quality improvement through agroforestry have received little attention in tropical agroforestry research, compared with the level of interest and research in this area in the temperate zone.

Agroforestry Research

The establishment of ICRAF in 1977 is considered to be the starting point of agroforestry research on a global scale. Since it happened in the wake of frustrations arising from the Green Revolution's failure to benefit poor farmers and escalating land-management problems such as tropical deforestation, fuelwood shortage, and soil degradation, the expectation was that investments in agroforestry research would contribute substantially to addressing these problems. Research results in agroforestry were therefore expected to be of a problem-solving and application-oriented nature.

Although the scientific gains in agroforestry during recent decades have been impressive, doubts and concerns have been expressed that the original expectations of agroforestry research have not been fulfilled. Some of the reasons for these concerns can be related to the prevailing myths about the science of agroforestry, and misconceptions about the practice of conducting agroforestry research. A major myth centers around the popular perception that agroforestry involves 'miracle' tree species and their 'magical' ability to improve soil productivity in agroforestry systems. Our understanding of the main tree-mediated processes that determine the extent and rate of nutrient cycling and soil improvement in agroforestry systems is far from satisfactory, so some of the widely held assumptions about their benefits could be wrong or incomplete.

Misconceptions also abound in agroforestry research. A major one is about research methodologies. In agroforestry research, be they of biophysical or socioeconomic nature, we rely on methodologies that have been developed in specific disciplines, often for conditions that are simpler than (or different from) those of agroforestry. Such methodologies may not encompass the special conditions and requirements of agroforestry. Agroforestry systems are often a puzzle to economists too. Since the basic-needs approach of classical political economics has been replaced by the market-superiority premise, economic efficiencies of farm enterprises are calculated based on their profit generation. Although tropical agroforestry systems that are primarily of subsistence nature fulfill the basic needs of farm families, they rank very low in the value premises and theoretical assumptions that underline neoclassical analysis. Yet, agroforestry systems have flourished for a long time. Thus, agroforestry systems pose challenges to ecologists and economists alike. Many such misconceptions about research in agroforestry are, however, not unique to agroforestry, but are characteristic of most such application-oriented research on low-input, integrated land-use systems to support the evasive and ill-defined goals of rural development.

Future Directions

The gains and developments of more than two decades of agroforestry research and development are certainly impressive. In particular, the second half of the 1990s has witnessed a large number of top-quality research and specialized publications. Undoubtedly agroforestry is now on a firm scientific footing. Today agroforestry is no longer a mysterious enigma that defies science and scientific principles, as it was perceived two decades ago. Agroforestry is well on its way to becoming a specialized science at a level similar to those of crop science and forestry science.

These hard-earned scientific gains need to be put to practice for solving the problems we had set out to address when we started research. In order to accomplish that, a two-pronged approach seems to be the best strategy. First, process-oriented research should be extended to hitherto underresearched aspects of agroforestry, and second, a rethinking is needed on overall research and development efforts at all levels (local, regional, and global). Relatively underresearched aspects include the 'agro' (i.e., crop) component of agroforestry. All crops and their cultivars/varieties that are used in agroforestry research today are those that have been developed for sole-crop conditions, and they may not be the best for mixed stands, as in agroforestry systems. Domestication and improvement of indigenous and underexploited trees is an area that can yield immediate dividends. Furthermore, research endeavors have so far been limited mostly to plot and field levels; they need to be scaled up to landscape, watershed, and ecosystem levels. Another research priority is policy and market issues pertaining to valuation of nontimber forest products. Enhancing the adoption potential of agroforestry technologies through improved extension techniques and methodologies for impact assessment are two other areas that need immediate research attention in the near future.

Agroforestry systems, be they practiced in the tropics or temperate regions, provide many basic needs and ecosystem services, and thus contribute to many regional developmental goals. These underexploited systems have the potential to develop into a set of major land-use options in the twenty-first century.

Acknowledgment

This chapter is adapted from the author's work: Nair PKR (2001) Agroforestry. In: Tolba MK (ed.) *Our Fragile World: Challenges and Opportunities for*

Sustainable Development. Forerunner to *The Encyclopedia of Life Support Systems*, vol. I, pp. 375–393. Paris, France: UNESCO.

List of Technical Nomenclature

Agroforestry Purposeful growing of trees, crops, and sometimes animals in interacting combinations for a variety of objectives. Agrisilviculture = trees + crops; silvopasture = trees + pasture/animals; agrosilvopasture = trees + crops + animals/pasture

Alley cropping Growing crops in the interspaces or alleys between planted trees or shrubs. In the tropics, alley cropping usually takes the form of hedgerow intercropping, where crops are grown in the alleys between regularly pruned hedgerows of planted shrubs or trees

Fallow Land resting from cropping, which may be grazed or left unused, often colonized by natural vegetation. An improved fallow refers to deliberate planting of fast-growing species for rapid replenishment of soil fertility

Homegarden A subsistence farming system consisting of integrated mixtures of multipurpose trees and shrubs in association with crops and sometimes livestock around homes, the whole unit managed intensively by family labor

ICRAF International Centre for Research in Agroforestry, now World Agroforestry Centre (Nairobi, Kenya)

Multipurpose tree (and shrub) A tree/shrub that is grown for multiple products and/or services

Multistoried or multistrata system An arrangement of plants forming distinct layers from the lower (usually herbaceous) layer to the uppermost tree canopy

See also: **Windbreaks and Shelterbelts**

Further Reading

Agroforestry Systems. The Netherlands: Kluwer Academic. Available online at: www.kluweronline.nl.

Buck LE, Lassoie JP, and Fernandes ECM (eds) (1999) *Agroforestry in Sustainable Agricultural Systems*. Boca Raton, FL: Lewis/CRC.

Buresh RJ and Cooper PJM (eds) (1999) The science and practice of short-term improved fallows. *Agroforestry Systems* 47: special issue.

CABI International (1998) *The Forestry Compendium*. Wallingford, Oxon, UK: CABI. CD-ROM.

Garrett HE, Rietveld WJ, and Fisher RF (eds) (2000) *North American Agroforestry: An Integrated Science and Practice*. Madison, WI: American Society of Agronomy.

Leakey RRB, Temu AB, Melynk M, and Vantomme P (eds) (1996) *Domestication and Commercialization of Non-Timber Forest Products in Agroforestry Systems*. Rome, Italy: FAO.

Nair PKR (ed.) (1989) *Agroforestry Systems in the Tropics*. Dordrecht, The Netherlands: Kluwer.

Nair PKR (1993) *An Introduction to Agroforestry*. Dordrecht, The Netherlands: Kluwer.

Nair PKR and Latt CR (eds) (1998) *Directions in Tropical Agroforestry Research*. Dordrecht, The Netherlands: Kluwer.

Nair PKR, Rao MR, and Buck LE (2004) *New Vistas in Agroforestry: A Compendium for the First World Congress of Agroforestry 2004*. Dordrecht, The Netherlands: Kluwer.

The Overstory. Available online at: www.agroforester.com/overstory/ovs.html.

Young A (1997) *Agroforestry for Soil Management*, 2nd edn. Wallingford, Oxon, UK: CAB International.

Air Phase *See* **Aeration**; **Diffusion**

Albedo *See* **Energy Balance**; **Radiation Balance**

Allophane and Imogolite *See* **Amorphous Materials**

ALLUVIUM AND ALLUVIAL SOILS

J L Boettinger, Utah State University, Logan, UT, USA

Introduction

Alluvial soils are some of the world's most useful and productive soil resources. Alluvium, the parent material of alluvial soils, is the sediment deposited by fluvial systems such as rivers and streams. Alluvium includes a wide variety of compositions and textures, depending on the source of geologic materials and the depositional environment. Alluvium occurs in all climate regimes and underlies geomorphic surfaces ranging in age from zero to millions of years. Therefore, alluvial soils are also extremely diverse.

Alluvium

Alluvium is ultimately dervied from the weathering and erosion of bedrock such as basalt or granite, or other unconsolidated sedimentary deposits, including colluvium (gravity-deposited), loess or eolian sand (wind-deposited), or alluvium. Important characteristics of alluvium, which ultimately influence the properties of alluvial soils, include composition, texture, and landform.

Composition

The chemical and mineralogical composition of alluvium depends on the lithology, or type of rock materials, from which the sediments are derived: alluvium derived from calcareous sedimentary rocks such as limestone and dolomite will be calcareous; alluvium derived from gypsum-bearing shales will be gypsiferous; alluvium derived from siliceous crystalline igneous rocks will be rich in quartz and feldspar, and lack carbonates and gypsum; alluvium derived from several lithologic sources will have a mixed composition; and so on. For example, the composition of alluvium in the vast San Joaquin Valley of central California differs from east to west. The eastern side of the valley comprises alluvium derived from granitic rocks of the Sierra Nevada mountain range and therefore is rich in quartz and feldspar, with minor amounts of mafic minerals such as hornblende and biotite. The western side of the valley comprises alluvium from the dominantly marine sedimentary rocks of the southern Coastal Ranges. The west-side alluvium contains varying amounts of sulfates, carbonates, and a diversity of silicate minerals, reflecting the range in composition of the marine shales, siltstones, sandstones, and limestones.

Texture

The texture of alluvium, or relative distribution of particle size, depends primarily on the energy of the fluvial environment in which the sediments were deposited. High-energy fluvial environments such as channels of braided streams can carry and therefore ultimately deposit relatively large particles. For example, glacial outwash deposited by high-discharge streams carrying glacial meltwater is typically cobbly to gravelly. Lower-energy fluvial environments deposit finer-textured alluvium. For example, sediments deposited by overbank flooding are typically clays, silts, and fine sands. An example of gravelly channel deposits and silty overbank flood and distal fan deposits is shown in Figure 1.

Different components of the stream channel vary in their ability to carry and deposit sediment, and the stream channel migrates across the landscape in response to changing discharge and bed loads. As a result, alluvium is typically stratified, i.e., composed of alternating layers of different textures (Figure 2b). Increasingly finer textured deposits, or a fining-upward sequence, often overlie coarser-textured alluvium.

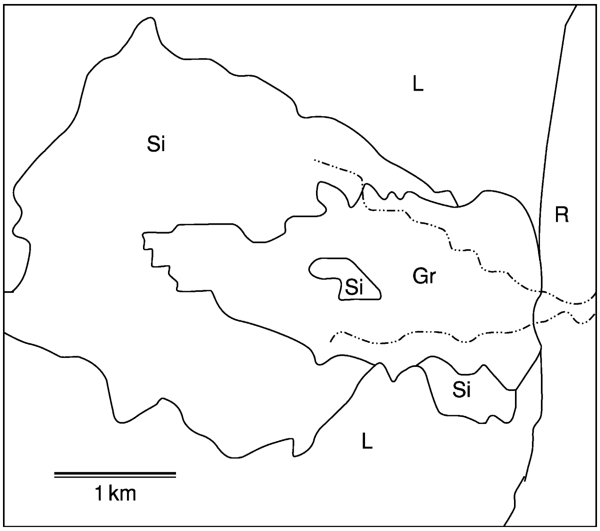


Figure 1 Schematic representation of the Holocene Green Canyon alluvial fan cut into and deposited on lacustrine deposits (L) of Pleistocene Lake Bonneville, Cache Valley, northern Utah, USA. Streams draining rocky limestone and dolomite uplands (R) deposited gravelly channel alluvium (Gr) and silty overbank and distal fan alluvium (Si). Holocene soils are the Green Canyon series (loamy–skeletal Typic Haploxerolls) formed in gravelly alluvium, and the Millville series (coarse–silty Typic Haploxerolls) formed in silty alluvium. (Information derived from the Soil Survey of Cache Valley Area, Utah.)

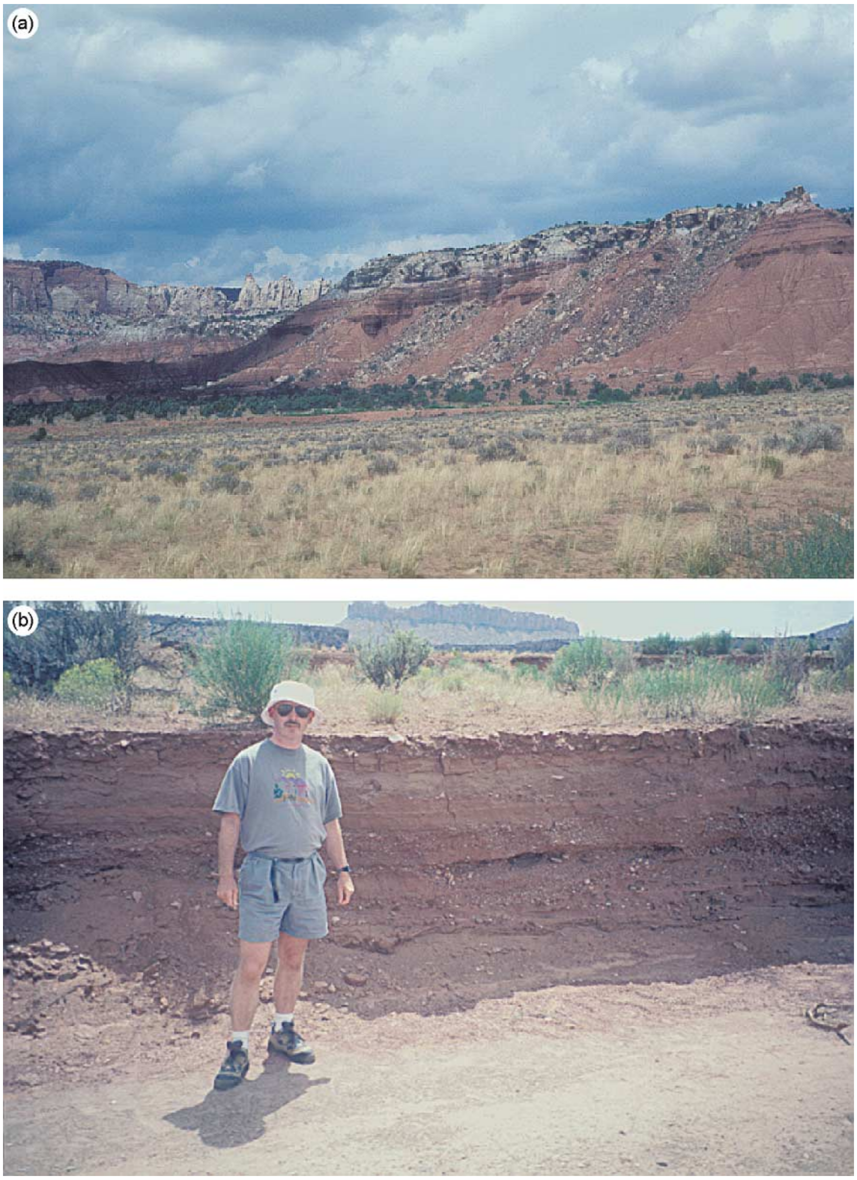


Figure 2 (a) Alluvium from the Mesozoic sedimentary rocks of the Circle Cliffs area of southern Utah, USA, deposited in a small, low-gradient alluvial fan that slopes away from the sediment source to the right of the frame; (b) recent alluvial soil (Ustic Torrifluvent) in the Circle Cliffs, USA, showing typical stratification of alluvial sediments. The man is approx. 1.7 m tall.

Alluvial deposits that fine upward reflect the migration of the stream channel away from its initial position on the landscape. The stratigraphy of alluvial deposits can provide clues to the history of landscape evolution.

The composition of the geologic source material also influences the texture of alluvium. Clearly, alluvium derived from sandstone is typically sandy, whereas alluvium derived from loess is silty. The physical weathering characteristics of the rock source will also influence texture of the alluvium. For example, alluvium derived from granitic rocks is typically sandy to fine gravelly – the size range of the individual mineral grains – because the rock fabric typically shatters into individual grains.

Landforms

Alluvium occurs in deposits that compose a variety of different landforms. Recent alluvium occurs on nearly level floodplains adjacent to streams and rivers. Older alluvium occurs on alluvial terraces, which are generally higher than associated floodplains and are not subject to frequent flooding. Higher terraces subject to upland erosion are often dissected. Alluvial terraces are also known as stream or river terraces. Pleistocene glacial outwash occurs on outwash terraces.

Streams draining mountainous terrain that discharge into wide valleys or basins, such as in the Basin and Range Province of western North America, deposit alluvial fans. Components of the alluvial fans include floodplains, fan terraces, and dissected fan remnants. Alluvial fans emanating from adjacent streams can coalesce and form bajadas. Finer-textured distal fan deposits can form alluvial plains or basins on the valley floor, and overbank deposits can form interfan basins between low-gradient alluvial fans. Alluvial fans are generally sloped away from the sediment source (Figure 2a), with steeper slopes near the source and lower gradient slopes far from the source.

Alluvial Soils

The morphological, physical, chemical, and mineralogical properties of alluvial soils depend greatly on the characteristics of the alluvial parent material in which the soils formed, especially when the soils are young. As alluvial soils develop with time, the other soil-forming factors influence the resulting soil properties.

Recent Alluvial Soils

Recent alluvial soils are often highly stratified, containing layers of alluvium that were deposited successively and/or in fining-upward sequences (Figure 2b). Soils on active floodplains receive deposits of new alluvium with each flooding episode. The amount of alluvium deposited during each event will vary. Small amounts of material deposited on the soil can be barely perceptible and incorporated into the underlying surface horizon rapidly, the rate of which depends on the climate and biota. Larger amounts of new alluvium can completely bury underlying soils.

Because of periodic disturbance by flooding, soils on recent floodplains often develop only A or O horizons, resulting from the near-surface deposition and decomposition of plant material. Subsequent deposition of new alluvium and reinitiation of landform stability and soil formation results in soils containing one or more buried A or O horizons.

Recent alluvial soils typically can have somewhat elevated concentrations of organic carbon at depth. New alluvium is often derived from the eroded A or O horizons of upland and/or upstream soils. In addition, soils with buried A or O horizons clearly demonstrate an irregular decrease in organic carbon with increasing depth (Figure 3).

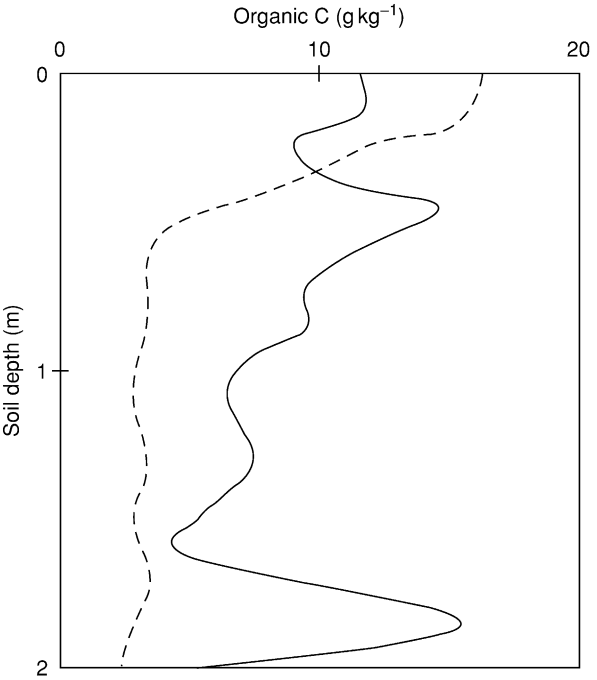


Figure 3 Idealized representation of a recent alluvial soil with an irregular decrease in organic carbon with depth (solid line) compared with an older alluvial soil with a regular decrease in organic carbon (dashed line). The recent alluvial soil probably has a buried A horizon (Ab) at approx. 1.8 m. Organic carbon distributions are typical of alluvial soils in the semiarid, temperate climate of southeastern Utah, USA.

Climate and associated biota further influence the properties of recent alluvial soils. These soils in humid climates generally support dense vegetation, thus developing A and/or O horizons more rapidly than in arid, semiarid, or subhumid climates. Highly soluble minerals such as gypsum and salts, if present in the parent alluvium, will be rapidly dissolved and leached in humid climates, whereas these constituents are often retained in arid climates.

Older Alluvial Soils

In general, older alluvial soils develop when they are no longer subject to periodic flooding events. Surfaces are more stable and thus able to support a stable vegetation cover. Organic carbon in the subsoil is eventually decomposed and the soil develops a regular distribution of organic carbon with increasing depth (Figure 3).

Climate, which influences vegetation and associated biota, further modifies the properties of alluvial soils. In cold climates with permafrost, cryoturbation can disrupt stratification of alluvial layers; in warmer climates, the available precipitation influences the resulting soil properties; in humid climates, alluvial soils are commonly leached. Depending on the mineralogical composition and texture of the parent material, B horizons develop in the subsoil and accumulate constituents such as silicate clay, free iron oxides, and metal humus complexes. In subhumid, semiarid, and arid climates, alluvial soils are incompletely leached. Depending on alluvium composition and texture, B horizons can accumulate carbonates (nearly ubiquitous), gypsum, soluble salts, etc., with or without silicate clay.

Old alluvial soils have often been subject to changes in climate during their development. This is particularly the case in areas that have existed as alluvial valleys or basins for hundreds of thousands to millions of years, such as the valleys and basins of the Basin and Range physiographic province of western North America. Very old alluvial soils can show the imprinting of several climate regimes, e.g., well developed, clay-rich B horizons that have been engulfed by calcium carbonate accumulations such that clay skins are no longer visible in the field.

Alluvial deposits, landforms, and the associated soils can be subject to active fluvial processes at one time or another. Therefore all or a portion of an alluvial soil profile is subject to erosion. If only a portion of the soil is removed by erosion, the soil has been truncated. Often, soils are eroded down to the more resistant clay- or carbonate-rich B horizon; older, truncated alluvial soils can have a B horizon exposed at the soil surface. Recent alluvial deposits may bury a truncated soil. As soil development begins again, soil-forming processes influence the recent overlying alluvium and the underlying older alluvial soil if shallow enough, resulting in a soil with a welded profile.

Alluvial soils often retain at least a partial record of the history of alluvial deposition. As soils develop on stable or truncated surfaces, they also record the passage of time since deposition, landform stability, and/or truncation. Therefore, alluvial soils are useful for soil-geomorphic studies. Soil development on alluvial surfaces has been used to correlate surfaces of similar ages, decipher paleoclimate, and estimate activity of faults that cut alluvial surfaces.

Classification

The diversity of alluvial soils results in a complex array of potential soil classifications. Recent alluvial soils that lack significant development of surface or subsurface diagnostic features are classified into the order of Entisols (US Department of Agriculture Soil Taxonomy System) or Fluvisols (Food and Agriculture Organization World Reference Base for Soil Resources).

In the Soil Taxonomy System, the formative element 'fluv' is used to connote the alluvial origin and stratified nature of recent alluvial soils. Better-drained Entisols that have high organic carbon levels (more than 0.2%) at depth (125 cm) or an irregular decrease in organic carbon with depth (see Figure 3) are in the suborder of Fluvents. Fluvents are further subdivided into 'great groups' by soil moisture regimes (e.g., Udifluvents, Torrifluvents, etc.). Alluvial Entisols that are saturated for prolonged periods during a normal year are Fluvaquents. In arid climates, the organic carbon level in the soil is often low; thus, many youthful alluvial soils in arid climates are classified as Torriorthents.

With greater landform stability and soil development, alluvial soils can develop into a myriad of different soils and thus can occur in all of the other soil orders of the Soil Taxonomy System. Alluvial soils with mollic epipedons and high base status throughout are Mollisols. Alluvial soils in arid climates that develop subsurface diagnostic horizons (e.g., argillic, calcic, gypsic) are Aridisols. Other alluvial soils with ochric or umbric epipedons are Inceptisols, Alfisols, Ultisols, Oxisols, and Spodosols, depending on the diagnostic subsurface horizons and properties present and the degree of leaching. Alluvial soils subject to cryoturbation are Gelisols; alluvial soils that accumulate deep organic soil materials are Histosols. Vertisols and Andisols are also possible, depending on the mineralogical composition, texture, and the degree of soil development.

Human Use of Alluvial Soils

For millennia, humans have used alluvial soils, especially those that are young and less developed, for the production of food. New alluvium rich in organic matter and nutrients provided fertile soils for agriculture.

Alluvial soils have been the sites of the earliest agriculture. This is illustrated by the concentration of early civilization advancement along river corridors in the Mediterranean Basin. For example, early agricultural societies developed along the Tigres and Euphrates rivers in the Middle East, the Nile in Egypt, and the Ebro in the Iberian Peninsula.

Indigenous cultures throughout the world still rely on the productivity and ease of cultivation of recent alluvial soils. Periodic flooding results in the rejuvenation of soil fertility by depositing organic-matter rich sediment on soil surfaces. Zuni and other indigenous peoples of the arid and semiarid southwestern US rely also on the periodic deposition of sand, which acts as mulch to increase infiltration rates and decrease soil moisture loss via evaporation.

Even today, alluvial soils underlie the most productive agricultural regions of the world. For example, the Central Valley of California supports a vibrant, multimillion-dollar agricultural industry. However, many of the soil and landscape properties that make alluvial soils attractive for agriculture are also attractive for urban development. Alluvial soils often have low slopes and occur in wide valleys or plains and are easy to excavate. Therefore, thousands of hectares of productive agricultural soils are being lost each year to urbanization and industrialization, making the preservation of prime agricultural land a world-wide concern.

List of Technical Nomenclature

Alluvial	Relating to alluvium
Alluvial fans	Fan-shaped landforms deposited by streams usually emanating from mountains and emerging on to lower gradient valleys
Alluvium	Sediment deposited by fluvial systems such as rivers and streams
Bajadas	Continuous skirt of sediment adjacent to a mountain range formed by coalescent alluvial fans
Bed loads	Debris moved along the bottom of the stream
Braided streams	Stream that flows in a network of dividing and rejoining channels because flow is diverted around sediment deposited by the stream
Buried soil	Older, developed soil overlain with more recently deposited sediment
Channel	Deepest part of a stream where the main flow of water occurs
Discharge	Rate of flow (volume per unit time)
Dissected	Incised by erosion
Distal fan deposits	Sediments deposited at the far end of the alluvial fan where the stream runs out of water and energy
Floodplains	Landform adjacent to the stream constructed by materials deposited by present flooding activities, and still covered with water during floods
Glacial outwash	Glacier-derived sediments moved by high-discharge steams carrying glacial meltwater
Landscape evolution	Development of deposits and landforms through time
Overbank flooding	Flooding where water rises and flow over the tops of the banks of the stream
Paleoclimate	Past climate(s)
Stratified	Composed of alternating layers
Terrace	Relatively flat to gently sloped surface generally bounded by short, steep slopes
Truncated soil	Soil from which the upper part was removed by erosion

Further Reading

Birkeland PW (1999) *Soils and Geomorphology*, 2nd edn. New York: Oxford University Press.

Busacca AJ, Singer MJ, and Verosub KL (1989) *Late Cenozoic Stratigraphy of the Feather and Yuba Rivers Area, California, with a Section on Soil Development in Mixed Alluvium at Honcut Creek*. In: US Geological Survey Bulletin 1590-G, pp. G1–132. Washington, DC: Government Publishing Office.

Eghbal MK and Southard RJ (1993) Stratigraphy and genesis of Durorthids and Haplargids on dissected alluvial fans, western Mojave Desert. *Soil Science Society of America Journal* 59: 151–174.

Gile LH and Grossman RB (1979) *The Desert Project Soil Monograph: Soils and the Landscapes of a Desert Region Astride the Rio Grande Valley, Near Las Cruces, New Mexico*. Washington. DC: USDA Soil Conservation Service.

Gile LH and Hawley JW (1966) Periodic sedimentation and soil formation on an alluvial-fan piedmont in southern New Mexico. *Soil Science Society of America Journal* 30: 261–268.

Harden JW (1987) *Soils Developed in Granitic Alluvium Near Merced, California*. In: US Geological Survey Bulletin 1590-G, pp. A1–65. Washington, DC: Government Printing Office.

Hillel D (1991) *Out of the Earth: Civilization and the Life of the Soil*. New York: Free Press.

Norton JB (2003) Hillslope soils and organic matter dynamics within a Native American agroecosystem on the Colorado Plateau. *Soil Science Society of America Journal* 67: 225–234.

ALUMINUM SPECIATION

D R Parker, University of California, Riverside, CA, USA

Introduction

Aluminum (Al) is the third most abundant element in the Earth's crust, and thus plays a pivotal role in soil science, geochemistry, ecology, and numerous related disciplines. Although it has only one stable oxidation number in nature (+III), the chemistry of Al is extremely complex, and has been studied intensively for more than a century. Although only sparingly soluble at environmental pH values, Al is demonstrably toxic to a variety of organisms, including higher plants, soil microorganisms, phytoplankton, fish, and other aquatic animals.

During the last two decades, considerable emphasis has been placed on the aqueous speciation of Al, specifically its tendency to form stable complexes with an assortment of inorganic and organic ligands. Aluminum is classified as an 'A-type' or 'hard-sphere' metal, and thus has considerable affinity for oxygen-containing ligands (including carboxyl groups on organic molecules) and fluoride (F^-). Understanding the speciation of Al is essential to explaining and predicting its toxicity, as only certain aqueous species seem to elicit any toxic response. An accurate description of Al speciation is also necessary for modeling its geochemical behavior. The solubility of Al-bearing minerals, as an example, is described thermodynamically in terms of the activity of the free, trivalent metal ion, and is independent of the 'side reactions' involving other ligands.

Inorganic Aluminum Species in Soil Solution

The most notable feature of the chemistry of Al(+III) in water is its tendency to hydrolyze. In simple, dilute solutions of low pH, the dissolved Al is primarily the hexaaquo ion, $Al(H_2O)_6^{3+}$. As pH is increased, however, Al reacts with water according to the general scheme:

$$xAl^{3+} + yH_2O \rightleftarrows Al_x(OH)_y^{(3x-y)+} + yH^+ \qquad [1]$$

with overall formation constants of the form:

$$\beta_{x,y} = \frac{\left(Al_x(OH)_y^{(3x-y)+}\right)\left(H^+\right)^y}{\left(Al^{3+}\right)^x} \qquad [2]$$

where the waters of hydration have been omitted for simplicity, and parentheses indicate ionic activities. The four mononuclear hydrolysis products (i.e., the coefficient $x = 1$) are reasonably well-understood and agreed-upon, and formation constants (log β) are provided in Table 1.

Polynuclear hydroxy complexes of Al (where $x > 1$) can be readily demonstrated in the laboratory, most often by adding a strong base (e.g., NaOH) to a solution of a soluble salt such as $AlCl_3$. As long as the ratio of added OH^- to total Al is <3, stable and soluble polynuclear complexes are formed that can persist for months or even years. The exact structure of these complexes has been the subject of fierce debate, but the most convincing evidence is for a tridecameric, 'Keggin' structure formulated as $AlO_4Al_{12}(OH)_{24}(H_2O)_{12}^{7+}$, and often denoted simply as 'Al_{13}.' This tridecamer consists of a highly symmetrical tetrahedrally coordinated Al centralized in a cage-like structure composed of 12 octahedrally coordinated Al atoms. At present, however, there is little evidence that this polynuclear Al complex exists outside the laboratory, and its existence in soil solution is especially uncertain.

The third hydrolytic reaction of Al involves its precipitation as oxyhydroxide minerals such as gibbsite and bayerite. The solubility of these minerals shows an extremely strong dependence on solution pH, as predicted by the stoichiometry of reaction:

$$Al^{3+} + 3H_2O \rightleftarrows Al(OH)_{3(s)} + 3H^+ \qquad [3]$$

The mineral gibbsite is often assumed to control the solubility of Al in soils, and measured solubility

Table 1 Aluminum hydrolysis and complexation constants at infinite dilution and 25°C (waters of hydration are omitted for simplicity)

Reaction	*log β*
$Al^{3+} + H_2O \rightleftarrows Al(OH)^{2+} + H^+$	−5.0
$Al^{3+} + 2H_2O \rightleftarrows Al(OH)_2^+ + 2H^+$	−10.1
$Al^{3+} + 3H_2O \rightleftarrows Al(OH)_3^0 + 3H^+$	−16.8
$Al^{3+} + 4H_2O \rightleftarrows Al(OH)_4^- + 4H^+$	−23.0
$Al^{3+} + SO_4^{2-} \rightleftarrows AlSO_4^+$	3.5
$Al^{3+} + 2SO_4^{2-} \rightleftarrows Al(SO_4)_2^-$	5.0
$Al^{3+} + F^- \rightleftarrows AlF^{2+}$	7.0
$Al^{3+} + 2F^- \rightleftarrows AlF_2^+$	12.7
$Al^{3+} + 3F^- \rightleftarrows AlF_3^0$	16.8
$Al^{3+} + 4F^- \rightleftarrows AlF_4^-$	19.4
$Al^{3+} + 5F^- \rightleftarrows AlF_5^{2-}$	20.6
$Al^{3+} + 6F^- \rightleftarrows AlF_6^{3-}$	20.6
$Al^{3+} + H_2PO_4^- \rightleftarrows AlH_2PO_4^+$	~3
$Al^{3+} + HPO_4^{2-} \rightleftarrows AlHPO_4^-$	~7

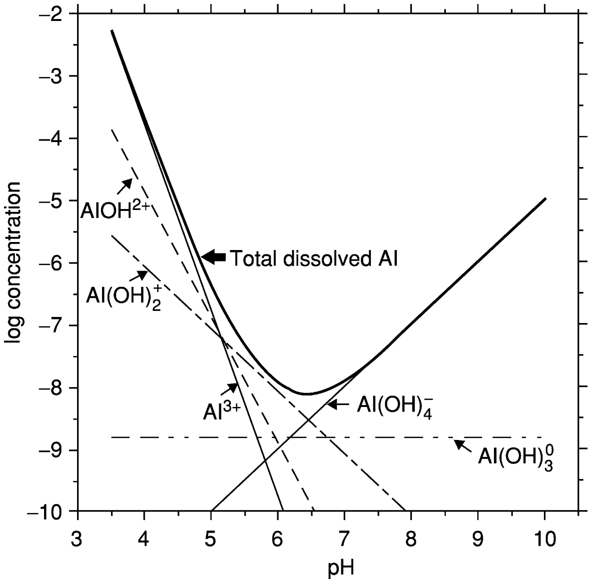


Figure 1 Distribution of mononuclear Al species in a simple system contains no ligands other than the hydroxyl ion that arises from hydrolysis. The total Al in solution is limited, especially at near-neutral pH values, by sparingly soluble oxy-hydroxide minerals such as gibbsite, modeled here as $Al^{3+} + 3H_2O \rightleftarrows Al(OH)_{3\,(s)} + 3H^+$, with log $\beta = -8.0$.

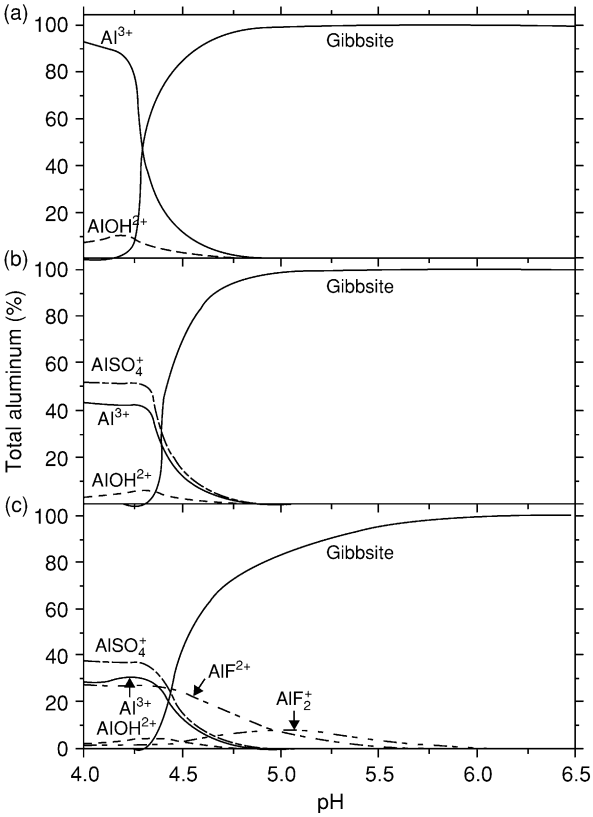


Figure 2 Distribution of Al as a function of pH for a simple solution initially containing 50 μmol l^{-1} total Al. Species comprising ≤1% of the total Al are omitted for clarity. (a) Ionic medium is 1 mmol l^{-1} $Ca(NO_3)_2$; (b) ionic medium is 1 mmol l^{-1} $CaSO_4$; (c) ionic medium is 1 mmol l^{-1} $CaSO_4$, and includes 15 μmol l^{-1} total F in solution. In all three cases, the overall solubility of Al is limited above pH 4.5. Speciation was calculated using the formation constants in Table 1 and the assumed gibbsite solubility depicted in Figure 1.

products have varied significantly. However, a log β of -8.0 for the foregoing reaction would be a reasonable value, and the resulting solubility of Al as a function of pH is depicted in Figure 1. Note the logarithmic nature of the y-axis; at pH ~6.5, the expected concentration of dissolved Al is only about 10^{-8} mol l^{-1}. In acidic soils with pH <5.0, the free Al^{3+} ion dominates, and toxicity is often best correlated with its concentration (or activity). Interestingly, Al solubility also increases at alkaline pH: the solubility of Al at pH 8.7 is about what it is at pH 5.0 (Figure 1). Yet, no toxic effects of Al have been demonstrated at alkaline pH, presumably because of the dominance of an anionic species, $Al(OH)_4^-$.

Other inorganic ligands that are important to the aqueous speciation of Al in soils include sulfate, fluoride, and perhaps phosphate; formation constants for the relevant complexes are given in Table 1. Interactions between Al and anions such as chloride, nitrate, and (bi)carbonate are too weak to be of any general significance. In light of the stability of the Al-O-Si bonds in aluminosilicate minerals, one might expect significant complexation of Al by aqueous Si, but this does not seem to be the case, presumably due in part to the low solubility of silicic acid in water.

In Figure 2, the distribution of Al as a function of pH in acidic soils (e.g., 4.0–6.5) is presented using a linear rather than logarithmic y-axis. In a simple matrix of 1 mmol l^{-1} $Ca(NO_3)_2$, the only significant solution species are Al^{3+} and $AlOH^{2+}$; precipitation of gibbsite becomes dominant at about pH 4.5 (Figure 2a). If the nitrate is replaced by sulfate, then a codominant complex – $AlSO_4^+$ – prevails until gibbsite's insolubility again dominates (Figure 2b). When a small amount of the stronger ligand F is also included, the precipitation of gibbsite is partially suppressed, and somewhat more Al is maintained in the solution phase out to a pH of about 5.5 (Figure 2c). Note the complexity of the Al speciation in Figure 3 over the pH range 4.0–5.0, even in the absence of any organic ligands – a complexity with significant implications for direct experimental measurement of Al speciation (see below).

Organic Complexes of Aluminum

All soil solutions naturally contain dissolved organic carbon (DOC), although the quantities vary

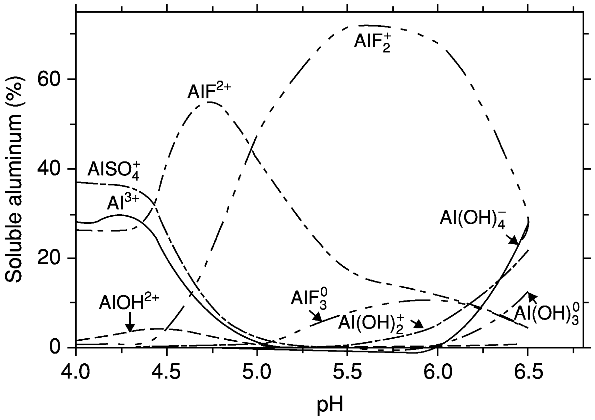


Figure 3 Distribution of Al species in Figure 2c, but expressed here as a percentage of the Al that remains in solution after gibbsite precipitates. Species comprising ≤1% of the total dissolved Al are omitted for clarity. The absolute concentration of these species is quite low above pH ~5.0, and the distribution of the numerous complexes present shows a marked dependence on pH.

enormously, and may be vanishingly small in some cases. Because of the affinity of carboxyl groups (and, to a lesser degree, acidic hydroxyls and amines) for Al, DOC is potentially a potent chelator of Al. The most ubiquitous form of DOC in soils is 'fulvic acid' (FA), a term used to describe a group of relatively high-molecular-weight compounds that are water-soluble at near-neutral pH values. One of the stable end-products of microbial processing of soil carbon, the FAs have highly variable and complex macromolecular structures, and contain a number of functional groups of varying affinity for Al. Attempts to model predictively the binding between Al and FA have ranged from rather simple to quite elaborate conceptual and mathematical representations. For the former, conditional stability constants (i.e., those that are specific to a given pH and ionic strength) for the formation of a '1:1' complex (here, the '1' for the ligand refers to a single binding site on the macromolecule) have generally been estimated to be in range of 10^3–10^6. Thus, the FAs may have a binding affinity approaching that of F^-, but may also be as weak a chelator of Al as is SO_4 (Table 1). The range in reported binding affinities is clearly quite wide, and generalizations about the overall importance of DOC are therefore difficult to make. In certain soils, the FAs may dominate the speciation of Al, while in others they be of little or no significance.

The other notable class of organic ligands for Al are the organic acid anions such as citrate, oxalate, malate, succinate, and propionate, among others. These are common by-products of both plant and microbial metabolism, and are sometimes present at measurable levels in soil solution. They are particularly likely to be abundant in the immediate vicinity of the root surface (i.e., the 'rhizosphere'), because roots often exude organic acid anions and because they are heavily colonized by microflora. The appearance of these ligands may be extremely transient, however, as they are readily metabolized as a carbon and energy source by heterotrophic microorganisms.

Root exudation of organic acid anions has recently been implicated as an important mechanism of the resistance to Al toxicity exhibited by certain plant taxa, consistent with the notion that complexed and chelated forms of Al are relatively harmless, and that it is the free Al^{3+} ion that elicits the toxic response. The stability constants for binding of Al by these ligands are not always precisely known, as different researchers have utilized quite disparate reaction schemes to fit similar potentiometric titration data (which are often highly problematic due to the uncontrolled formation of polynuclear hydroxy species, discussed above). But it seems clear that the tricarboxylate citrate has the highest binding affinity for Al, with 1:1 stability constant of about 10^8. Oxalate and perhaps malate are also sufficiently strong chelators of Al to reduce its toxicity in the rhizosphere, but the other common dicarboxylic acids are probably not under most conditions; monocarboxylic acids such as lactate and acetate are almost certainly of no general significance to Al speciation in soils.

Phenolic groups are important components of humic substances, and various phenolic acids, such as salicylic acid, caffeic acid, and gallic acid, have at times been implicated in the solubilization and mobility of Al in soils. It has been tempting to assign significance to these compounds because their reported log β values for Al binding are often quite high (>10). However, it is important to recognize that these binding constants refer to the fully deprotonated ligand. With the phenolic acids, the ring hydroxyl does not deprotonate except at extremely alkaline pH values; thus the conditional constants that describe their binding affinity in the acidic pH range of relevance are orders of magnitude lower. In comparison to the aliphatic organic acids described above, the phenolics are probably of minor significance to the speciation of Al in most soils.

The speciation of Al is of particular significance to its ecotoxicological effects. With both terrestrial and aquatic organisms, toxicity seems to correlate best with the concentration of free Al^{3+}, an affirmation of the 'free ion model' that is observed with a number of other trace metals. In general, none of the other inorganic or organic mononuclear species have been shown to contribute to the toxic response, including the aluminate anion that prevails at alkaline pH. The Al_{13} polynuclear complex has been shown to be uniquely and acutely toxic to both plants and algae reared in laboratory solutions, but the environmental significance of this finding remains unknown.

Determining Aluminum Speciation in Experimental Samples

Computational Methods

If all of the Al-complexing ligands present in a soil-solution sample could be accurately quantified, and if all of the pertinent binding constants were exactly known, then the speciation of Al could be readily computed using any one of a number of ‘off-the-shelf’ chemical equilibrium models (GEOCHEM-PC, MINEQL+, MINTEQA2, etc.). However, several obstacles to this procedure usually persist. Not all of the ligands can be readily identified and quantified at the levels that typically prevail in soil solution. Fluoride is problematic in that few analytical methods exist that are accurate in the needed micromolar range, and F has some affinity for metals other than Al and perhaps for colloidal forms of Al. Polynuclear forms of Al, if present, cannot be predicted using the thermodynamically based ion-association models that underlie all chemical speciation programs. Other problems may include difficulties in the accurate separation of dissolved Al from microcolloidal forms that might pass a submicron filter prior to analysis for total Al.

The highly heterogeneous nature of DOC is particularly problematic, and makes it difficult or impossible routinely to quantify and accurately model the Al-binding characteristics of organic ligands such as FA. The interactions between FA and Al have been modeled by treating the individual binding sites as a hypothetical soluble ligand, and binding stoichiometries include simple AlL complexes, as well those of the form AlL_n and Al_nL. Sometimes a single metal–ligand complex is employed, but often two or more ligands or binding stoichiometries are invoked to simulate the variation in binding affinity among the many sites on natural FA. The most elaborate computational models take into account electrostatic effects and make use of polyelectrolyte theory. To date, none of these approaches has shown any general predictive capability, and the models have usually required site- or even sample-specific empirical recalibration.

These limitations have led researchers to develop and employ a host of empirical methods for the speciation of Al, some of which rely partly on the computational approach described above. It should be noted, however, that chemical equilibrium modeling is often the method of choice for ‘clean’ solutions prepared from known components in the laboratory, and is invaluable as a heuristic tool for probing the complexities of aqueous Al chemistry in hypothetical solutions such as those depicted in Figures 1–3.

Fractionation Based on Size, Charge, and Reactivity

Much of the purported ‘speciation’ of Al in the literature would not meet a strict definition of the term, as the chemical entities in solution were not directly probed and quantified. More commonly, the dissolved Al is analytically allocated to several operationally defined categories based on physical separation due to differences in size and/or charge, or based on chemical reactivity with an Al-complexing agent.

Because much of the natural DOC in soil solutions or extracts consists of high-molecular-weight compounds, the DOC-complexed Al should be separable from the smaller species using dialysis or ultrafiltration membranes. Nominal pore diameters as small as 1 nm are available, and the method has been employed with some success. Care must be taken to avoid adsorptive losses on to the membrane, as well as contamination with reactive metals and ligands. The largest difficulty with such methods is achieving an unambiguous cutoff with respect to molecular size. Transport efficiency across the membrane can become very low as the molecular diameter approaches the nominal pore size, and there may be a range of molecular weights present with natural DOC.

A number of chromogenic Al-complexing reagents have been employed over the years to fractionate dissolved Al, including ferron, 8-hydroxyquinoline, chrome azurol S (CAS), and pyrocatechol videt (PCV), among others. When bound to Al, these ligands form a distinctly colored complex that can be quantified using ultraviolet-visible spectrometry (UV-Vis) or, occasionally, fluorimetry. Most typically, these are timed analyses in that the reading must be exactly made after a fixed period of reaction (e.g., 60 s), and this has been achieved in some cases by automating the method using flow-injection analysis (FIA). Alternatively, the reaction is similarly stopped by extraction of the Al complex into an organic solvent. Less commonly, the formation of the colored complex is monitored continuously, and the absorbance versus time data are then fitted to a kinetic model. In all cases, the method depends critically on the assumption that certain Al species are measured quantitatively during the brief reaction with the analytical ligand, and that other, less labile species are completely excluded from the analysis.

When examined critically, this assumption has seldom been found to be completely valid. The mononuclear hydrolysis products are, through proton exchange, in rapid equilibrium with the free Al^{3+} ion, so all should be quantified together. The sulfate complexes are weak, and seem to be readily ‘stripped’ by the Al-complexing reagent. With stronger Al complexes, however, such as those with F or citrate, the Al may or may not be exchanged on to the analytical reagent and, if it is, the rate may be both sluggish and unknown. The reactivity of Al–DOC complexes (e.g., the FAs) would be even more difficult to predict.

In a number of cases, this ‘rapidly reacting’ Al has been assumed to represent the sum of all of the inorganic, mononuclear Al species. Regardless of the validity of this assumption, it is important to note that, in order to construct a detailed picture of the Al speciation (for example, to know the concentration of free Al^{3+}), it is necessary to model computationally the distribution of these species as described previously. The measured, reactive Al is used as the ‘total’ Al concentration, and the concentrations of F and SO_4 must also be determined and used as model inputs. Thus, many of the empirical fractionation procedures in use are actually combination or ‘hybrid’ methods, in that part of the ‘speciation’ is done empirically, and part is done computationally.

Cation exchange resins of both strong- and weak-acid character have been used by a number of researchers to fractionate Al, often in combination with other methods. For example, a column of strong-acid resin may be used to separate inorganic and organic forms of mononuclear Al. The underlying assumption is that the inorganic complexes are positively charged and/or sufficiently labile to be retained by the resin, while organic complexes are neutral or negatively charged, and are sufficiently strong so as not to dissociate while passing through the column. Thus the inorganic species are retained, while the organic complexes are collected in the eluent where they are quantified by analyzing for total Al. Figure 4 presents a typical fractionation scheme wherein total mononuclear Al is quantified by one of the timed spectrophotometric methods, the organic subfraction by ion exchange, and the inorganic species by difference. The method also allows differentiation of any very nonreactive Al such as small colloidal solids, and of dissolved forms that are particularly refractory (e.g., large polynuclear complexes, unusually strong organic chelates).

Note that all of these empirical approaches contain some critical underlying assumptions. Foremost is that, during the analysis or separation of Al, the equilibrium distribution of species is not perturbed. Of course this cannot be strictly true as the separation of some of the Al into different phases, or its selective complexation by a reagent ligand, must inevitably shift the speciation. However, if kinetic hindrances sufficiently retard this reequilibration, then the method may not be severely compromised. In addition, many of the methods cause shifts in pH and ionic strength as compared to the original sample due to, for example, the inclusion of concentrated buffers in many of the timed spectrophotometric methods. It must be assumed that these do not cause major shifts in the distribution of Al species, although the pH-sensitivity that is well-illustrated in Figures 1–3 should instill extra caution in this regard.

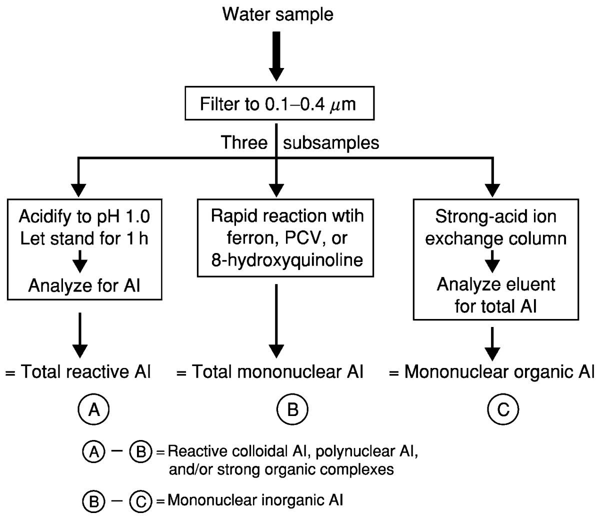

Water sample
Filter to 0.1–0.4 μm
Three subsamples
Acidify to pH 1.0
Let stand for 1 h
Analyze for Al
Rapid reaction wtih ferron, PCV, or 8-hydroxyquinoline
Strong-acid ion exchange column
Analyze eluent for total Al
= Total reactive Al
= Total mononuclear Al
= Mononuclear organic Al
A
B
C
A − B = Reactive colloidal Al, polynuclear Al, and/or strong organic complexes
B − C = Mononuclear inorganic Al

Figure 4 Schematic diagram of a typical fractionation procedure where ‘dissolved’ Al is allocated into five operationally defined chemical fractions. PCV, pyrocatechol violet.

Speciation Using More Direct Analytical Methods

Aluminum has a single natural isotope, ^{27}Al, with a small quadrupole moment and high resonance frequency, making it an ideal candidate for analysis using nuclear magnetic resonance (NMR). This is an appealing tool for aqueous speciation, because samples can be probed nondestructively at ambient temperatures, and no reagents or resins need be added to the sample matrix; original chemical conditions are thus readily preserved. Indeed, ^{27}Al-NMR has been widely used to study the hydrolysis behavior of Al and, increasingly, its complexation and chelation by various ligands in laboratory-synthesized solutions. An NMR spectrum reflects differences in coordination chemistry in two ways, via chemical shifts and via changes in line width. The former are most profound when the coordination number changes, for example from octahedral to tetrahedral coordination. Line broadening results from asymmetry in the ligand field which, due to ^{27}Al’s quadrupole moment, is a general problem that often limits the resolution of the technique.

To date, the greatest limitation to the application of ^{27}Al-NMR to ‘real-world’ speciation has been its lack of sensitivity: total Al concentrations of

$\sim 10^{-4}$ mol l^{-1}, or, more typically, $\geq 10^{-3}$ mol l^{-1}, have been required using conventional instrumentation. Line broadening is partially to blame for this poor sensitivity, but the presence of Al as an ubiquitous impurity in NMR probes and glass tubes is also responsible. Recently, utilization of specially constructed probes, coils, and sample tubes using low-Al materials has enhanced sensitivity such that concentrations approaching 10^{-6} mol l^{-1} can be addressed. It remains to be seen what the utility of such enhanced-sensitivity ^{27}Al-NMR might be for complex environmental samples containing multiple ligands. In acidic samples, virtually all of the Al is in octahedral coordination, and line-broadening can make it difficult or impossible to resolve the overlapping resonances arising from multiple Al species.

Various forms of liquid chromatography (e.g., IEC, HPLC, FPLC, SEC: see Table 2 for definitions) have been utilized for some 20 years in attempts to separate analytically and quantify various complexes of Al. The number of published methods is very large, and the last decade has seen a tremendous surge in interest in these techniques. The earlier papers tended to use postcolumn reaction with one of the aforementioned chromogenic reagents, with subsequent detection using UV-Vis or fluorimetry. More recently, there has been increased emphasis in 'hyphenated' techniques, where the chromatographic separation is coupled with an atomic spectroscopy instrument such as ICP-MS, ICP-AES, or ET-AAS (see Table 2 for definitions). These element-specific methods of detection have greatly improved sensitivity; detection limits are now $\sim 10^{-7}$ mol l^{-1} and even lower. The chromatographic separations have been based on differences in chemical properties such as charge, size (e.g., SEC), and hydrophobicity (in the case of organic complexes using RPC).

Table 2 Abbreviations used in the description of aluminum speciation methods

Abbreviation	*Definition*
CAS	Chrome azurol S
Ferron	8-Hydroxy-7-iodo-5 quinoline-sulfonic acid
FIA	Flow-injection analysis
FPLC	Fast protein liquid chromatography
HPLC	High-performance liquid chromatography
ICP-MS	Inductively coupled plasma mass spectrometry
ICP-AES	Inductively coupled plasma atomic emission spectrometry
ET-AAS	Electrothermal atomic absorption spectrometry (graphite furnace)
IEC	Ion-exchange chromatography
LC-ES-MS	Liquid chromatography electrospray mass spectrometry
morin	2,3,4,5,7-Pentahydroxy-flavone
NMR	Nuclear magnetic resonance
PCV	Pyrocatechol violet
RPC	Reversed-phase chromatography
SEC	Size-exclusion chromatography
UV-Vis	Ultraviolet-visible spectrometry

To date, most of the published studies have focused on synthetic laboratory solutions and biological fluids (e.g., human serum). Among the organic ligands, the emphasis as been on mammalian proteins and the aliphatic organic acids; comparatively little work has been done on environmental samples containing heterogeneous DOC. Thus, the utility of these chromatographic methods for soil solutions remains unknown, although the emergence of techniques such LC-ES-MS will undoubtedly help us to understand better the structure of Al-complexing humic substances. As with the fractionation methods described previously, a general caution applies to these chromatographic techniques: because of the eluents used, there can be marked changes in pH, ionic strength, or other chemical properties during the analysis. In order for the results to reflect the *in situ* speciation of the sample, assurance must be gained that the complexes of interest do not shift due to these changes. Such stability during analysis might arise thermodynamically, but more likely would be a result of kinetic stability that requires careful evaluation and verification.

Finally, mention can be made of two indirect approaches to the speciation of soil solution Al. The first utilizes an F-selective electrode to measure directly the activity of the free F$^-$ ion in the sample. A separate measurement is made after addition of a metal-complexing buffer to liberate all of the F, thus generating a value for total F in solution. These two values are combined with the measured prevailing pH and, utilizing the equilibrium relations shown in Table 1, the free Al^{3+} concentration (or activity) can be calculated. A similar approach involves the addition of a trace quantity of the reagent morin to the sample, which forms a fluorescent complex with Al that can be quantified at very low concentrations. Again, by knowing the concentrations of both total morin and the Al–morin complex, the free Al^{3+} concentration can be 'backed out' computationally. There are several limitations to these methods. With the F-electrode method, there must be sufficient free F$^-$ in the sample to obtain accurate readings. When the pH exceeds $\sim$5 and the hydrolysis products are more prevalent, small errors in measured F$^-$ can lead to relatively large errors in computed speciation. With the morin method, it is necessary that the quantity of morin added is small enough to cause only minimal shifts in the original speciation, requiring some foreknowledge of the prevailing Al chemistry; fluorescence of natural DOC in the sample can also cause interferences. Finally, the information gained from these two methods is limited to an estimate of the concentrations of free Al^{3+} ion and the mononuclear hydrolysis products (which are linked by pH); little is learned about the relative importance of the various Al complexes that cause total Al and free Al^{3+} to differ widely in many samples. Neither method has enjoyed widespread utilization at this time.

Further Reading

Bertsch PM and Parker DR (1996) Aqueous polynuclear aluminium species. In: Sposito G (ed.) *The Environmental Chemistry of Aluminum*, 2nd edn, pp. 117–168. Boca Raton, FL: CRC Press.

Bi S-P, Yang Y-D, Zhang F-P, Wang X-L, and Zou G-W (2001) Analytical methodologies for aluminium speciation in environmental and biological samples – a review. *Fresenius Journal of Analytical Chemistry* 370: 984–996.

Bloom PR and Erich MS (1996) The quantitation of aqueous aluminium. In: Sposito G (ed.) *The Environmental Chemistry of Aluminum*, 2nd edn, pp. 1–38. Boca Raton, FL: CRC Press.

Campbell PGC (1995) Interactions between trace metals and aquatic organisms: a critique of the free-ion activity model. In: Tessier A and Turner DR (eds) *Metal Speciation and Bioavailability in Aquatic Systems*, pp. 45–102. New York: John Wiley.

Gensemer RW and Playle RC (1999) The bioavailability and toxicity of aluminum in aquatic environments. *Critical Reviews in Environmental Science and Technology* 29: 315–450.

Kinraide TB (1991) Identity of the rhizotoxic aluminum species. *Plant and Soil* 134: 167–178.

LaZerte BD, van Loon G, and Anderson B (1997) Aluminum in water. In: Yokel RA and Golub MS (eds) *Research Issues in Aluminum Toxicity*, pp. 17–45. Washington, DC: Taylor & Francis.

Ryan PR, Delhaize E, and Jones DL (2001) Function and mechanism of organic anion exudation from plant roots. *Annual Review of Plant Physiology and Plant Molecular Biology* 52: 527–560.

Stumm W and Morgan JJ (1996) *Aquatic Chemistry*, 3rd edn. New York: John Wiley.

Tipping E (1998) Humic ion binding model VI: an improved description of the interactions of protons and metal ions with humic substances. *Aquatic Geochemistry* 4: 3–48.

AMMONIA

D E Kissel and M L Cabrera, University of Georgia, Athens, GA, USA

Introduction

The primary forms of nitrogen (N) taken up from soil by plants are ammonium (NH_4^+) and nitrate (NO_3^-). Soils used for the production of food and fiber often store large quantities of N in soil organic matter; for example, a highly fertile prairie soil in the central USA may typically contain 10 000 kg N per hectare in the top meter of soil, with most in organic compounds in the soil's organic matter and not available to plants. A small proportion of this N, around 1%, undergoes biological reactions each year to form ammoniacal nitrogen, i.e., either ammonia (NH_3) or NH_4^+. A natural equilibrium exists between NH_3 and NH_4^+ that depends largely on pH and temperature. The physical properties of these two forms of ammoniacal N are totally different. Both forms are water-soluble, but whereas NH_4^+ can be easily metabolized by plants and other organisms, NH_3 is highly toxic to living organisms. Because the pH of most soils is in the range of 5–8, most ammoniacal N exists in the soil as NH_4^+. However, due to the volatility and toxicity of NH_3, its concentration in soil may become sufficiently high above pH 7 to be toxic to organisms or to be lost to the atmosphere.

Over the past 50 years, there has been an approximately linear increase in the world production of industrially fixed N fertilizers, from around 10 million metric tons in 1960 to just over 80 million metric tons by 2000. Over 75% of this N is in the form of ammoniacal N, or in forms of N such as urea (about 40%) that react in the soil to produce ammoniacal N. Because N fertilizers are often added to soils in concentrated zones by surface or band applications, subsequent chemical reactions are often more pronounced than fertilizers that are uniformly distributed throughout a larger soil volume. The resulting chemical and physical reactions are of considerable practical importance because N fertilizer is sometimes lost into the atmosphere. Nitrogen applications with less loss result in more efficient and economical use by crops, and they minimize off-site losses to the ecosystem.

Chemical Reactions of Ammoniacal N

Ammonia–Ammonium Equilibrium

Ammonium in the soil solution can be considered as a weak acid in equilibrium with its dissociation products NH_3 and H^+ as follows:

$$NH_4^+ = NH_3 + H^+ \qquad pKa = 9.25 \qquad [1]$$

The pKa value indicates the pH at which half of the ammoniacal N is NH_4^+ and half is NH_3. Together with pH, the pKa can be used to estimate the proportion of ammoniacal N as NH_3 using the equation:

$$pH = pKa - \log[NH_3/NH_4^+] \qquad [2]$$

The value of the pKa at 25°C is 9.25; however, this dissociation is temperature-sensitive. The effect of temperature on the proportion of NH_3 in the mixture is substantial. For example, at a pH of 8, the proportion of ammoniacal N as NH_3 is about 2% at 10°C, and 13% at 40°C (Figure 1). Another physical constant that affects the behavior of ammoniacal N is the Henry's constant that describes the equilibrium of partitioning of NH_3 between a solution and the gaseous phase in contact with the solution. It too is affected by temperature, with an approximately threefold increase in the proportion of NH_3 that partitions into the air when the temperature is raised from 10 to 40°C. Taken together, higher temperatures enhance ammonia loss from ammoniacal N due to its effect on both of these physical constants.

Retention of Ammonia by Soils

In some agricultural production areas, anhydrous NH_3 is applied directly to soils as a source of N fertilizer for crop production. It is typically released into the soil from application points attached to tools that are pulled through the soil, resulting in line (banded) applications. The applied NH_3 reacts first with soil nearest the line of release, and then diffuses outward, resulting in cylindrical NH_3 retention zones just below the soil surface. The properties of the NH_3 retention zones are shown in Figure 2. Because NH_3 is a base when dissolved in water, it raises soil pH following its application to soil. The measured value of pH will vary with distance from the injection point or line, and it will also depend on the amount of NH_3 added and the H^+-buffering properties of the soil. As noted in Figure 2, the soil pH at the center of the retention zone is typically about 9 by 1 day after application. The outward expansion of the retention zone is complete by 1 day following application. The spatial distribution of nitrate concentrations at 1, 2, and 4 weeks following application is also shown in Figure 2. It indicates that nitrification proceeds most rapidly at an intermediate radial distance from the retention zone center, due to more favorable conditions for nitrification. Nitrate concentrations are lowest in the retention zone center where high concentrations of NH_3 are toxic to nitrifying organisms, thereby slowing the rate of nitrification. With more time and continued nitrification, soil pH decreases, and nitrification rates increase in the retention zone center.

The concentration of NH_3 retained in the center of a retention zone by 1 day after application can be predicted from the soil's titratable acidity to pH 9. Titratable acidity to pH 9 is a measure of the quantity of H^+ in soil that will potentially react with the applied NH_3 at the retention zone center, as well as the soil's pH-dependent negative charge that will result from raising soil pH to 9 that will in turn adsorb NH_4^+. This reaction can be written as:

$$NH_3 + H - Soil = NH_4 - Soil \qquad [3]$$

'H – Soil' in this case represents titratable acidity that is reactive with a base but not cation-exchangeable with an unbuffered salt. Because titratable acidity to pH 9 is a good predictor of ammoniacal N concentration in the center of the retention zone, it follows that initial soil pH should also affect the ammoniacal N concentration in the center of the retention zone. As the pH of a soil becomes more acid, the reduction in the soil's effective cation exchange capacity is balanced by an equivalent increase in the soils unionized pH-dependent charge that is measured as titratable acidity. If NH_3 is applied to two soils equal in all respects except for their pH, the differences in the ammoniacal N concentrations in the center of their retention zones will be equal to the difference

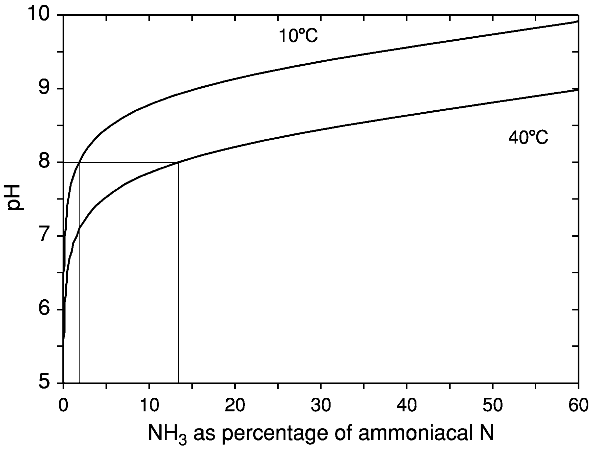


Figure 1 The percentage ammonia (NH_3) of a dilute ammoniacal N solution as affected by solution pH and temperature.

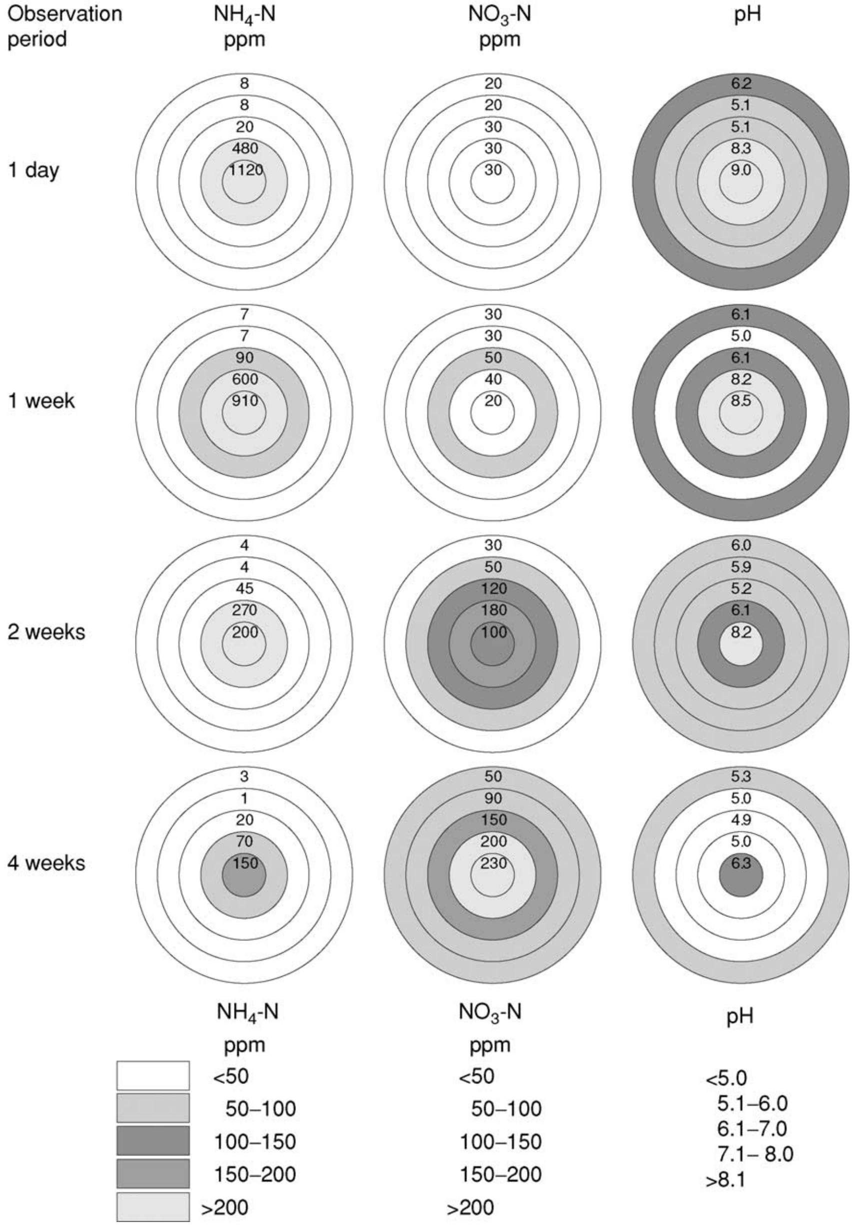


Figure 2 Distribution of NH_4^+, NO_3^-, and pH in soil with distance of 0–1.5, 1.5–3, 3–5, 5–7, and 7–9 cm from the injection line. Ammonia applied at 116 kg N ha^{-1} at 25 cm spacing. (Reproduced with permission from Nommik H and Vahtras K (1982) Retention and fixation of ammonium and ammonia in soils. *Agronomy* 22: 123–171.)

in titratable acidity to pH 9, being greater, of course, in the soil of lower pH. Likewise, as noted above, two soils with identical pH but with different unionized pH-dependent charge (titratable acidity) to pH 9.0 will also differ in NH_3 retention, with greater retention in the soil with greater titratable acidity.

Because the application of NH_3 varies greatly in practice by factors such as the spacing between adjacent retention zones, as well as the application rate per hectare, it follows that these too must be considered when estimating the properties of an NH_3 retention zone. The factors that will finally determine the properties of an NH_3 retention zone will be the soil's capacity of unionized pH-dependent charge per volume of soil (determined by its pH buffering capacity, its pH, and its bulk density), plus the retention zone spacing and the amount of NH_3 applied per hectare. A computer simulation model has been developed that describes the distribution of ammoniacal N and pH, as well as ammonia loss based on soil bulk density and its titratable acidity to pH 9, depth and spacing of application bands, and the amount of ammonia applied per hectare.

Retention and Loss of Surface-Applied Ammoniacal N in Soils

Nitrogen fertilizer is sometimes uniformly spread on the soil surface and not incorporated into the soil by tillage. This application method is often employed with established crops for which tillage will be detrimental, such as the topdressing of small grain, forages, or crops produced under no-tillage culture. In these cases, the chemical, physical, and biological reactions that control the formation and adsorption of NH_3 at the soil surface are of special interest. These reactions determine if NH_3 losses are significant, and the subsequent availability of the N fertilizer to crops. Although other biological reactions may be important in affecting the availability of the N fertilizer to crops, only those affecting the loss and retention of NH_3 will be considered here. Two different cases of N fertilizer reactions will be considered: (1) those involving urea fertilizers applied to all soils, and (2) those involving the reaction of NH_4^+ fertilizers with calcareous soils. Each will be considered separately.

Reactions of Urea in Soils

When added to soil, urea fertilizer granules will typically dissolve in the soil water after a short time (minutes to hours), depending on the soil water content and temperature. Even applications at the soil surface will dissolve, typically within a few days, from water near the soil surface and from dews. The dissolved urea then reacts (catalyzed by soil urease enzyme) to form NH_4^+ and HCO_3^-, as described by eqn [4]. Concurrently, bicarbonate produced in eqn [4] forms carbon dioxide in eqn [5]. Below pH 8.2, eqn [5] will continue to consume all of the bicarbonate, provided that CO_2 diffuses freely from the soil, which would be the case for surface applications of urea. In the case that eqns [4] and [5] go to completion, two H^+ are consumed and two NH_4^+ ions released for each urea hydrolyzed (or one H^+ consumed for each NH_4^+ released).

$$CO(NH_2)_2 + 2H_2O + H^+ = 2NH_4^+ + HCO_3^- \qquad [4]$$

$$HCO_3^- + H^+ = CO_2 + H_2O \qquad [5]$$

The ratio of one H^+ consumed for each NH_4^+ released results in an effective cation exchange site being formed and occupied by an NH_4^+ ion, as shown in eqn [6]. The H^+ on the right side of eqn [6] is consumed in eqns [4] or [5]. The term '$H^+ - $Soil' in eqn [6] refers to the pH-dependent charge in the soil, which can be measured as titratable acidity.

$$H^+ - \text{Soil} + NH_4^+ = NH_4 - \text{Soil} + H^+ \qquad [6]$$

Some of the soil cations are not adsorbed on the cation exchange sites. These cations (typically Ca^{2+}, Mg^{2+}, K^+, NH_4^+, and Na^+) exist in the soil solution and exchange freely with the same cation species on the exchange sites. Since the equilibria described by eqn [1] only apply to the ions in the soil solution, the cation exchange equilibria, shown in eqn [7] for the case of Ca^{2+}, affects the amount of ammoniacal N in the soil solution, and therefore the proportion that will exist in solution as NH_3.

$$1/2Ca^{2+} + NH_4 - \text{Soil} = NH_4^+ + Ca_{1/2} - \text{Soil} \qquad [7]$$

The consumption of H^+ causes soil pH near the dissolved urea to rise. The amount of pH rise depends on the soil's H^+ buffering capacity, which will be discussed below. If the soil pH rises above 7, a significant amount of NH_3 can form, which depends primarily on the soil pH, temperature, and the concentration of NH_4^+ in the soil solution, as described by the equilibrium in eqns [7] and [1]. Based on equilibrium in eqn [1], any movement of NH_3 away from the reaction site will release H^+, thereby lowering pH.

Eqns [1], [4], [5], [6], and [7] are the predominant ones below a soil pH of 8.2. The process of nitrification, in which NH_4^+ is oxidized to NO_3^-, will add two H^+ to the soil for each NH_4^+ oxidized. This process will not be discussed any further here. Its effect on NH_3 volatilization is not very pronounced in the first week following application, since there is often a lag time of a few days in the buildup of a nitrifier population following N application.

Above pH 8.2, eqns [8] and [9] become important, and CO_2 formation from reaction [5] stops, as shown graphically in **Figure 3**. The predominant species is HCO_3^- (99.2% HCO_3 and 0.8% CO_3^{2-}) around pH 8.2. Any HCO_3^- from eqn [4] remains in the soil solution or is changed to CO_3^{2-} as pH is raised further, as shown by eqn [8]. If there are sufficient Ca^{2+} ions in the soil solution to exceed the solubility product of $CaCO_3$, then solid $CaCO_3$ is formed in the soil, as shown in eqn [9].

$$HCO_3^- = CO_3^{2-} + H^+ \qquad [8]$$

$$Ca^{2+} + CO_3^{2-} = CaCO_{3(\text{solid})} \qquad [9]$$

The formation of solid $CaCO_3$ removes CO_3^{2-} from the soil solution, causing a new equilibrium to be established in eqn [8], releasing H^+ that slows the rise in pH. This effect is shown in **Figure 4**. In this study, soils were titrated with either NH_4OH or urea + urease enzyme to raise soil pH. Up to pH 8.2, soil pH was raised identically by NH_4OH and urea that had completely hydrolyzed. Above pH 8.2, urea was

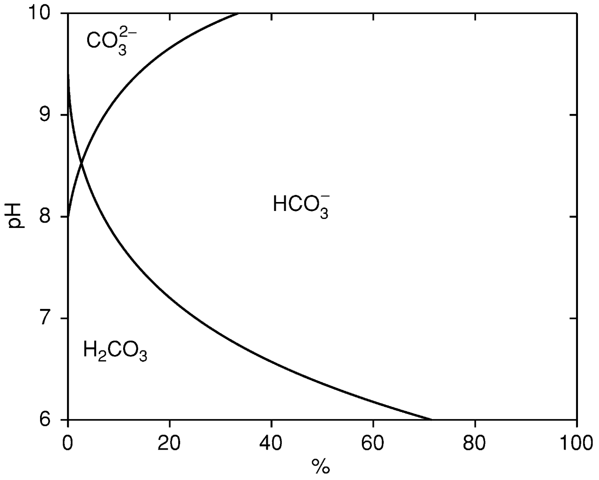


Figure 3 The distribution of inorganic C species in a dilute solution as affected by pH. (Reproduced with permission from Koelliker JK and Kissel DE (1988) Chemical equlibria affecting ammonia volatilization. In: Bock BR and Kissel DE (eds) *Ammonia Volatilization from Urea Fertilizers*, pp. 37–52. Bulletin Y-206. National Fertilizer Development Center. Muscle Shoals, AL: Tennessee Valley Authority.)

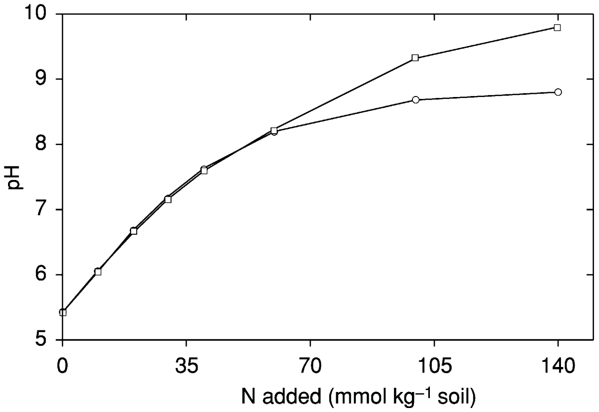


Figure 4 The change in soil pH of Kahola silt loam as affected by the addition of NH_4OH (squares) or urea (circles) allowed to hydrolyze completely. (Reproduced with permission from Kissel DE, Cabrera ML, and Ferguson RB (1988) Reactions of ammonia and urea hydrolysis reaction products with soil. *Soil Science Society of America Journal* 52: 1793–1796.)

less effective than NH_4OH in raising soil pH. The lower effectiveness of urea was due in part to the urea-C not forming CO_2 but remaining as HCO_3^- and CO_3^{2-} in the soil solution or by precipitating as $CaCO_3$.

Environmental Conditions and Ammonia Formation

When urea is surface-applied, the formation of NH_3 at or near the soil surface may allow some NH_3 volatilization, and when urea is banded near germinating seeds or seedlings, some NH_3 toxicity may result, causing plant damage or loss. The severity of both depends primarily on the concentration of NH_3 formed. The concentration of NH_3 depends on the amount and method of urea application, soil properties, and soil environmental conditions for several days after application. Specifically, the most important factors affecting NH_3 concentrations are:

1. The concentration of urea in soil immediately following application directly affects the rise in soil pH. If urea is more concentrated, eqns [4–6] take place in a smaller mass of soil, consuming H^+ from a smaller volume, which will raise soil pH to higher levels. More concentrated urea applications will also result in more concentrated levels of ammoniacal N.
2. The soil pH for a 3- to 5-day period following urea application affects NH_3 concentrations because pH determines the proportion of the two forms of ammoniacal N, as described in eqn [1].
3. The rate of urea hydrolysis affects NH_3 concentration because the reactants and reaction products all tend to become less concentrated with time due to diffusion from the site of the reactions. Therefore, faster urea hydrolysis rates will increase the concentration of NH_3 at the site of the reactions.

Each of these factors will be discussed further in the following sections.

Urea concentration The concentration of urea in soil depends largely on two factors: (1) the amount applied per hectare; and (2) the method of application. Methods such as band application or surface application cause higher urea concentrations than urea that is incorporated and mixed with soil by tillage. Rainfall or irrigation immediately after surface application can also move urea into soil and cause it to disperse and to be less concentrated. Molecular diffusion of urea in soil water also disperses urea and makes it more dilute, allowing it to interact with more urease. Hydrogen ions from the soil are consumed when urea is hydrolyzed, so it follows that the H^+ will be consumed from less soil where urea is more concentrated, provided there is enough urease enzyme and adequate temperature and water to sustain a fast reaction. Rapid hydrolysis and consumption of H^+ from a small volume of soil will cause the soil pH at the site of application to increase further and cause more NH_3 to form. When urea hydrolysis is complete, the rise in pH will be directly proportional to the initial urea concentration, and it will be highly soil-dependent, as shown in Table 1 for two soils from Kansas. In this comparison, the pH was measured after all urea had hydrolyzed due to the addition of urease enzyme. For comparison, the soil

Table 1 Soil pH of two Kansas soils 13 h after adding urea and urease

Urea-N concentration (mg N kg^{-1})	*Haynie (soil pH)*	*Kahola (soil pH)*
0	5.74	5.42
140	7.17	6.05
280	8.13	6.67

Source: Kissel DE, Cabrera ML, and Ferguson RB (1988) Reactions of ammonia and urea hydrolysis reaction products with soil. *Soil Science Society of America Journal* 52: 1793–1796.

pH after the hydrolysis of 280 mg urea-N kg^{-1} was 6.67 for the silt loam soil (Kahola) and 8.13 for the sandy loam (Haynie). The differences in the rise in pH were due to differences in soil H$^+$ buffering capacities of the two soils, which can be measured from a pH titration of the soils, to be discussed further in the next section.

Soil pH and H$^+$ buffering capacity The pH rise that occurs when a given concentration of urea hydrolyzes depends largely on the H$^+$ buffering capacity of the soil, which is a measure of the quantity of H$^+$ that will be adsorbed or released by the soil per unit pH change. The H$^+$ buffering capacity arises primarily from soil organic matter and clays, and is sometimes referred to as pH-dependent charge in soils. The type of clay mineral greatly affects buffering. Kaolinite clays have very little H$^+$ buffering capacity, whereas 2:1 layer clays, especially hydroxy-interlayered vermiculites, have significant amounts. Sandy soils are often low in both clay and organic matter, and therefore have less H$^+$ buffering capacity. The effect of soil buffering on soil pH change for a uniformly mixed sample is shown in Figure 5. The effect of H$^+$ buffering capacity on loss from surface-applied urea will depend in part on how deep and how quickly the applied urea and urea reaction products diffuse into the soil. Urea and its reaction products move by molecular diffusion to depths of approximately 25 mm within 24 h after the urea dissolves at the soil surface. With movement this deep into the soil, it is apparent that differences in H$^+$ buffering capacity will affect the rate and amount of loss, as shown in Figure 5. In this study, soil mixes were prepared that differed in their H$^+$ buffering capacities, but had the same initial pH and cation exchange capacities. Urea fertilizer was then applied to the surface of each soil mix and NH_3 loss and surface pH were measured for 19 days. As would be expected, the NH_3 loss was greatest from the soil mix that had the least H$^+$ buffering capacity (soil mix 1). Because this soil mix had the fewest H$^+$ ions, it allowed the soil surface pH to be higher than the other soil mixes 3–4 days after urea was surface-applied. The higher pH allowed a greater proportion of the ammoniacal N to be in the gaseous NH_3 form, which resulted in more loss. In contrast, the soil mix with the most H$^+$ buffering capacity (soil mix 3) had the lowest surface soil pH, and the least NH_3 loss. In summary, it can be generalized that soils with more H$^+$ buffering capacity will have a lower pH increase following surface application of urea and will therefore lose less NH_3.

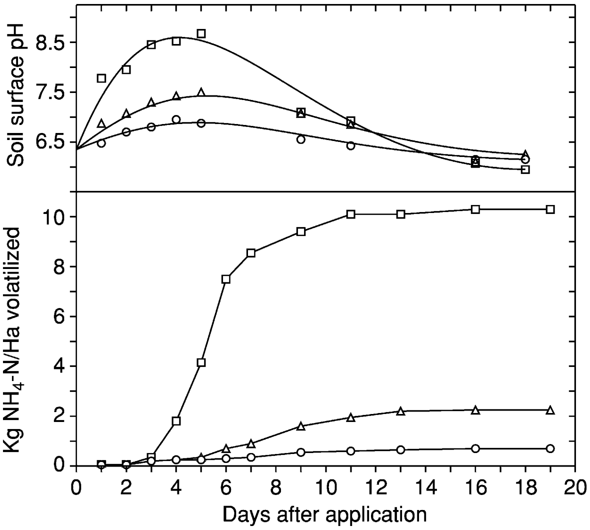


Figure 5 Soil surface pH and total NH_3 loss with time after surface applications of urea to soil mixes with different H$^+$ buffering capacity. Squares, soil mix 1; triangles, soil mix 2; circles, soil mix 3. (Reproduced with permission from Ferguson RB, Kissel DE, Koelliker JK, and Basel W (1984) Ammonia volatilization from surface-applied urea: effect of hydrogen ion buffering capacity. *Soil Science Society of America Journal* 48: 578–582.)

Factors affecting the rate of urea hydrolysis The factors that are most important in affecting the rate of urea hydrolysis are the concentration of urease enzyme in soil, urea concentration, soil temperature, soil water content, and soil pH. Each will be discussed separately.

Soil urease concentration The number of active urease molecules in soils cannot be measured directly. However, urease activity (the rate of urea hydrolysis) can be measured under standard conditions of temperature, water content, pH, and urea concentration. The measured value of a soil's urease activity under the standard conditions (including a defined urea concentration) then represents the number of urease molecules. Several investigators have attempted to find the soil property that correlates best with urease activity in cultivated soils. There was general agreement among these studies that soil organic C was a relatively good predictor of soil urease activity. In the

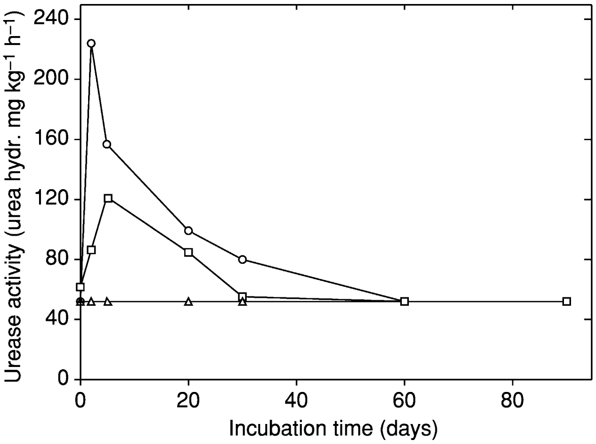


Figure 6 The effect of incubation time on the urease activity of unamended soil (triangles) and soil amended with 5000 mg of corn residue (squares) or glucose (circles) per kg of soil. (Reproduced with permission from Zantua MI and Bremner JM (1976) Production and persistence of urease activity in soils. *Soil Biology and Biochemistry* 8: 369–374.)

absence of a measurement of a soil's urease activity, it can be estimated from the organic C content of cultivated soils. However, this estimation is complicated in soils that contain a significant amount of decomposing residues, as shown in Figure 6. In this study, soil samples were treated with glucose or crop residues that enhanced microbial activity, which increased the urease activity of the soil. The persistence of the urease produced depended on the soil type and the amendments added, but eventually the level of urease activity returned to that found in unamended soil. Apparently, this is due to a fixed number of protective sites for urease molecules, which varies with soil type.

In the practice of surface application of urea, the urea hydrolysis rates observed during the first 5–20 h after surface application of urea are sometimes slower than subsequent hydrolysis rates. These observations are not due to an increase in urease activity. Studies have shown that urea uniformly mixed with soil results in linear increases in NH_4^+ over the first 6 or 7 h after application, indicating no change in urease activity. A more appropriate explanation for the lag in hydrolysis rates from surface application of urea is that time is required for urea to diffuse down into the soil and to interact with more urease molecules as time progresses. This increased contact of urea with urease as time progresses explains the lag in hydrolysis rates over the first couple of days. Of course, NH_3 loss lags even more because of the time required to increase soil pH and levels of ammoniacal N.

Urea concentration When urea fertilizer granules dissolve in soil, the concentration of urea in the soil solution varies from a nearly saturated solution near the granule to very low concentrations at the outer reaches of the area affected. In laboratory studies of urea mixed uniformly with soil, the effects of urea concentration across the full range of concentrations have been shown to be well described by the sum of two Michaelis–Menten reactions, but with some slowing of hydrolysis rates at concentrations near saturation, apparently due to inhibition of the enzyme.

Soil temperature The effect of soil temperature on urea hydrolysis has been shown to follow a Q_{10} of 2, i.e., a doubling of reaction rate with each 10°C rise in temperature. For example, a rise in soil temperature from 5 to 25°C will cause the rate of urea hydrolysis to be approximately four times faster. Urea hydrolysis is a relatively fast process, even at cool temperatures that would typically occur when urea fertilizers are applied to cool-season crops. For example, for a typical silt loam surface soil with about 2% organic matter, a water potential of −0.5 MPa or wetter, and a neutral pH, one would expect urea hydrolysis at 27°C to be 90% complete by 3 days after application, whereas at 2°C, urea hydrolysis would still be 90% complete by 9 days after application.

Soil water content Urea hydrolysis proceeds at optimum rates when soil water contents are in the range that is readily available to plants. In the water content range from wilting point (−1.5 MPa) to air-dry, the rate slows greatly, and essentially stops as the soil approaches air dryness.

Soil pH The effect of soil pH on urease activity has been described in the scientific literature primarily as a single effect, regardless of the urea concentration. The optimum pH has been reported to vary from around 7 to as high as 8.8–9.0. More recent research indicates that the optimum pH is different for the two reactions described briefly in the section on urea concentration.

Reactions of NH_4^+-N Applied to Calcareous Soils

Based on experimental evidence with surface-applied urea fertilizers, NH_3 loss is not very significant from most soils unless the pH at the soil surface reaches a value of 6.5 or higher (Figure 5). Because of this, the loss of NH_3 from most NH_4^+-N fertilizers applied to the surface of neutral to acid pH soils is very low or insignificant. This may also be partly true because the most common NH_4^+-N sources such as NH_4NO_3, $(NH_4)_2SO_4$, and NH_4Cl are also slightly acid, which may decrease the pH of the soil surface even further. An exception is $(NH_4)_2HPO_4$, which has a pH of about 8.5.

When NH_4^+-N sources are applied to calcareous soils, NH_3 loss can occur, not only because the pH of calcareous soils is above 7 (typically in the range of 7.5–8.2), but also because some of the NH_4^+ sources will react with $CaCO_3$ to form reaction products that increase soil pH to even higher values, as shown in Figure 7. Ammonium sources that form relatively insoluble calcium reaction products with the anion of the NH_4^+ fertilizer will have higher pH for the first 5–20 h after application than those sources that do not. For the sources shown in Figure 7, the solubilities were 0.0016, 0.20, and 134 g 100 ml^{-1} H_2O for CaF_2, $CaSO_4$, and $Ca(NO_3)_2$, respectively. Losses of NH_3 by 24 h after surface application were 60, 40, and 8% respectively for NH_4F, $(NH_4)_2SO_4$, and NH_4NO_3 respectively. These losses are consistent with the relative soil pHs for the three sources for the first day following application. Although pH following NH_4NO_3 application was by far the lowest, around pH 7, losses were still significant, due to the high concentration of NH_4^+ in the soil solution near the soil surface. The reactions responsible for the high losses from these sources are the following:

$$CaCO_{3(solid)} = Ca^{2+} + CO_3^{2-} \qquad [10]$$

Carbonate ions dissolved from the solid phase at pH of around 8 will immediately revert to HCO_3^-, as shown by eqn [6] and Figure 3. As Ca^{2+} is removed from solution by precipitation with the anion of the NH_4^+ N source (for those sources with a relatively insoluble reaction product), more $CaCO_3$ will be dissolved and the CO_3^{2-} will react via eqn [8] to remove H^+ from the soil, which will tend to raise pH, as will loss of CO_2 because eqn [5] will reestablish equilibrium. Based on these reactions, $CaCO_3$ will be consumed in an amount chemically equivalent to the NH_3 lost from the soil. This consumption of $CaCO_3$ can also be understood from an H^+ balance, since the fertilizers are applied to the calcareous soils as NH_4^+, and volatilize as NH_3 with a difference of one H^+ added to the soil for each NH_3 molecule volatilized.

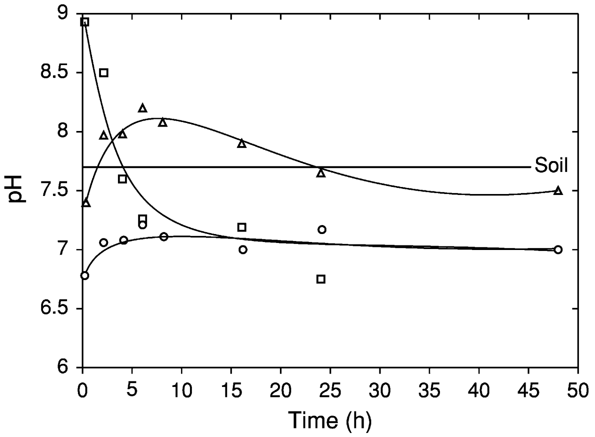


Figure 7 Change in soil pH with time of the surface 6.4 mm of Houston Black clay soil at 33°C for three ammonium compounds surface applied at 550 kg N ha^{-1}. Squares, NH_4F; triangles, $(NH_4)_2SO_4$; circles, NH_4NO_3. (Reproduced with permission from Fenn LB and Kissel DE (1973) Ammonia volatilization from surface applications of ammonium compounds on calcareous soils. I. General theory. *Soil Science Society of America Proceedings* 37: 855–859.)

Since $CaCO_3$ will be consumed from the calcareous soil due to these NH_3 loss reactions, and since the reactions occur within a time period of hours, the $CaCO_3$ must be of sufficiently small particle size to dissolve quickly to sustain the loss reactions. A concentration of about 10% by weight of clay size $CaCO_3$ sustains the reaction at rates determined by other factors, such as temperature and rate of N application.

List of Technical Nomenclature

(MPa)	Water potential
(mg kg^{-1})	Concentration
(multiplicative reaction rate (10°C)$^{-1}$: Q_{10})	Temperature quotient

See also: **Nitrogen in Soils:** Cycle

Further Reading

Bates RG and Pinching GD (1949) Acidic dissociation constant of ammonium ion at 0° to 50°C, and the base strength of ammonia. *Journal of Research of the National Bureau of Standards* 42: 419–430.

Fenn LB and Kissel DE (1973) Ammonia volatilization from surface applications of ammonium compounds on calcareous soils. I. General theory. *Soil Science Society of America Proceedings* 37: 855–859.

Ferguson RB, Kissel DE, Koelliker JK, and Basel W (1984) Ammonia volatilization from surface-applied urea: effect of hydrogen ion buffering capacity. *Soil Science Society of America Journal* 48: 578–582.

Izaurralde RC, Kissel DE, and Cabrera ML (1987) Titratable acidity to estimate ammonia retention. *Soil Science Society of America Journal* 51: 1050–1054.

Izaurralde RC, Kissel DE, and Cabrera ML (1990) Simulation model of banded ammonia in soils. *Soil Science Society of America Journal* 54: 917–922.

Kissel DE and Cabrera ML (1988) Factors affecting urea hydrolysis. In: Bock BR and Kissel DE (eds) *Ammonia Volatilization from Urea Fertilizers*, pp. 53–66. Bulletin Y-206, National Fertilizer Development Center. Muscle Shoals, AL: Tennessee Valley Authority.

Kissel DE, Cabrera ML, and Ferguson RB (1988) Reactions of ammonia and urea hydrolysis reaction products with soil. *Soil Science Society of America Journal* 52: 1793–1796.

Koelliker JK and Kissel DE (1988) Chemical equilibria affecting ammonia volatilization. In: Bock BR and Kissel DE (eds) *Ammonia Volatilization from Urea Fertilizers*, pp. 37–52. Bulletin Y-206. National Fertilizer Development Center. Muscle Shoals, AL: Tennessee Valley Authority.

Moyo CC, Kissel DE, and Cabrera ML (1989) Temperature effects on soil urease activity. *Soil Biology and Biochemistry* 21: 935–938.

Nommik H and Vahtras K (1982) *Retention and Fixation of Ammonium and Ammonia in Soils.* Agronomy. 22, pp. 123–171. Madison, WI: American Society of Agronomy.

Rachhpal-Singh and Nye PH (1984) The effect of soil pH and high urea concentrations on urease activity in soil. *Journal of Soil Science* 35: 519–527.

Zantua MI and Bremner JM (1976) Production and persistence of urease activity in soils. *Soil Biology and Biochemistry* 8: 369–374.

AMORPHOUS MATERIALS

J Harsh, Washington State University, Pullman, WA, USA

Introduction

Poorly crystalline aluminosilicates (‘amorphous materials’) in soils consist primarily of the mineral imogolite and allophane, a term that refers to all remaining short range-ordered (SRO) aluminosilicate clays. Although long thought to derive only from volcanic ash parent material, these materials are ubiquitous in soils under a variety of vegetation, parent materials, and climate. They are characterized not only by their poor crystallinity, but by their small particle size, imparting high specific surface areas, and by their variable and permanent surface charge properties. These characteristics make them important ion exchange materials even when present at relatively low concentration. SRO materials also play roles in Al solubility, organic matter stabilization, and soil shrink–swell properties. As a result, they affect a number of environmental processes, including solute transport, mineral weathering, formation of humic materials, and soil stability.

Occurrence in Soils

Allophane and imogolite result from the rapid precipitation of soluble Al and Si that are released either by labile parent materials, such as volcanic ash, or by intense weathering of any parent material ranging from sandstone to granite. They are metastable products that form in favor of more stable crystalline minerals such as kaolinite, because SRO materials have lower surface tension and, thus, nucleate more rapidly in aqueous solutions. Eventually they are replaced by the more crystalline aluminosilicate clay minerals and, as Si is weathered and leached from the soil, by Fe and Al (hydr)oxides. Allophane and imogolite may begin to form immediately in an ash deposit but increase in concentration over tens of thousands of years before declining in favor of the more crystalline minerals. Therefore, their metastable nature does not diminish their importance in soils.

Poorly crystalline aluminosilicates can form in any environment where weathering leads to sufficient Al and Si in solution. The association of allophane and imogolite with soils of volcanic origin arises from the fact that ash, tephra, and other pyroclastic materials contain amorphous volcanic ash that can rapidly release Al and Si. Allophane formation is favored by Si concentrations between 0.1 and 4 $\text{mmol}\,l^{-1}$. The type of allophane that forms will depend on the Si-to-Al ratio of the solution, the pH, and soluble organic matter. Soils derived from volcanic debris and dominated by SRO materials (or by Al–humus complexes) often fall into the Andisol soil order under the US classification system. It would be a mistake to assume that allophane and imogolite are limited to soils derived from volcanic materials and formed in humid environments. They have been found in soils that have formed from gneiss, sandstone, igneous and sedimentary rock, and loess. Allophane has been found in soils of six of the 12 orders: Entisols, Inceptisols, Spodosols, Alfisols, Aridisols, and Ultisols, as well as in humic, xeric, and arid moisture regimes, including the deserts of Iceland. The nearly ubiquitous nature of SRO materials underscores the fact that it is the presence of sufficient Al and Si in solution that determines their formation, not any particular environment or parent material.

The environment and parent material largely determine the nature of the materials formed. The Al/Si ratio in solution largely governs the type of allophane that results. Allophanes range from Al-rich, which have an Al/Si molar ratio of approximately 2, to

Si-rich, which have Al/Si of approximately 1. At high soluble Si concentration in the soil solution, there is a tendency for halloysite, a 1:1 phyllosilicate clay mineral, to form. Halloysite is favored at soluble Si concentrations greater than $10^{-3.45}$ mol l^{-1}, whereas Al-rich allophane and imogolite are favored at lower concentrations. When the parent material contains volcanic glass, an equilibrium concentration of $10^{-2.7}$ mol l^{-1} is expected, and the formation of these minerals will depend upon the rate at which Si is leached from the system. Illustrative of this climatic influence is a climatosequence of a 170-ky soil in Hawaii where 'noncrystalline' materials (primarily allophane and imogolite) increase and halloysite decreases with increasing rainfall (Figure 1). Increasing rainfall lowers readily soluble Si, favoring allophane and imogolite.

Identification

The poorly crystalline nature of allophane and imogolite defies their easy identification in soils by X-ray diffraction, the method of choice for the phyllosilicate clay minerals and many oxides. Instead, a combination of methods is used to identify and characterize these materials, including electron microscopy, selective chemical dissolution, Fourier transform infrared (FTIR) spectroscopy, and differential thermal analysis. Synthetic allophane and imogolite or purified samples from soils can be further characterized by nuclear magnetic resonance (NMR) and X-ray absorption spectroscopy (XAS). Electron microscopy is useful for defining the morphology of individual particles and aggregates of SRO materials, whereas the other methods give information on the structural and chemical bonding properties.

Allophane and imogolite are not truly noncrystalline or amorphous materials and do exhibit X-ray diffraction patterns, albeit with diffuse and often weak diffraction peaks. These patterns can be used to identify imogolite and distinguish it from Al-rich allophane (protoimogolite) (Figure 2), something not readily accomplished with FTIR. All allophanes, regardless of Al/Si ratio, display the two broad peaks

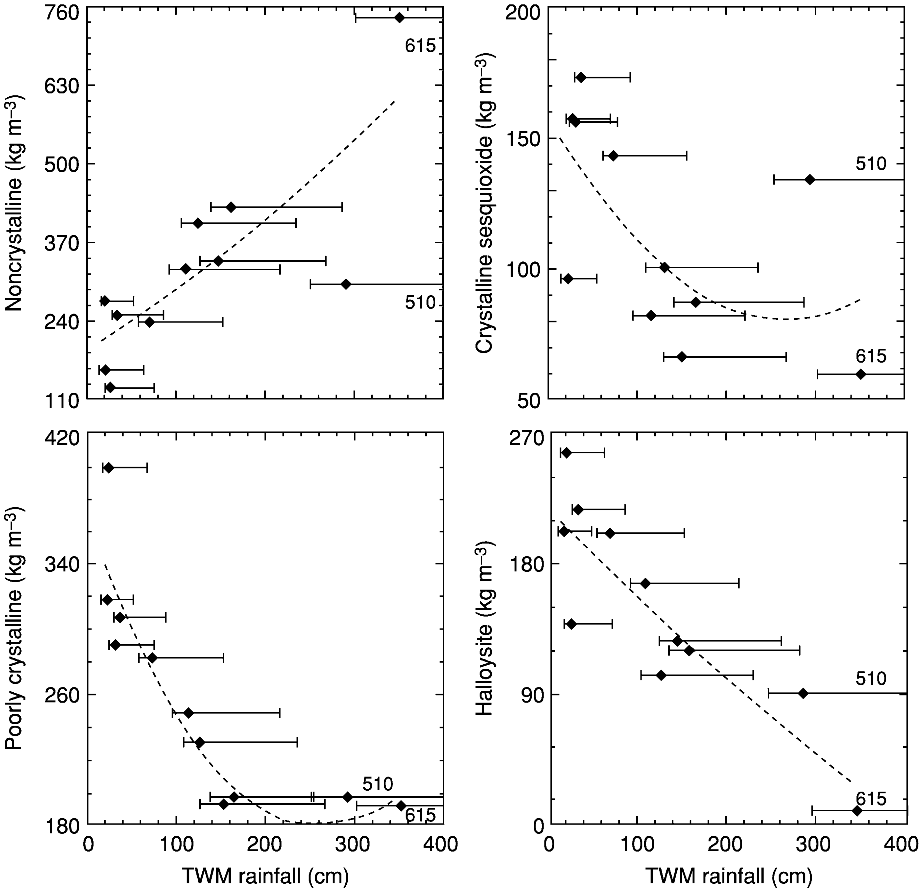


Figure 1 Soil components extracted from a climatosequence of soils in Hawaii. The 'noncrystalline' components are acid oxalate-extracted aluminosilicates – primarily allophane and imogolite. TWM, total weighted mean. (Source: Chadwick OA, Olson CG, Hendricks DM, Kelly EF, and Gavenda RT (1994) Quantifying climatic effects on mineral weathering and neoformation in Hawaii. *Transactions of the 15th World Congress of Soil Science,* International Soil Science Society and the Mexican Society of Soil Science.)

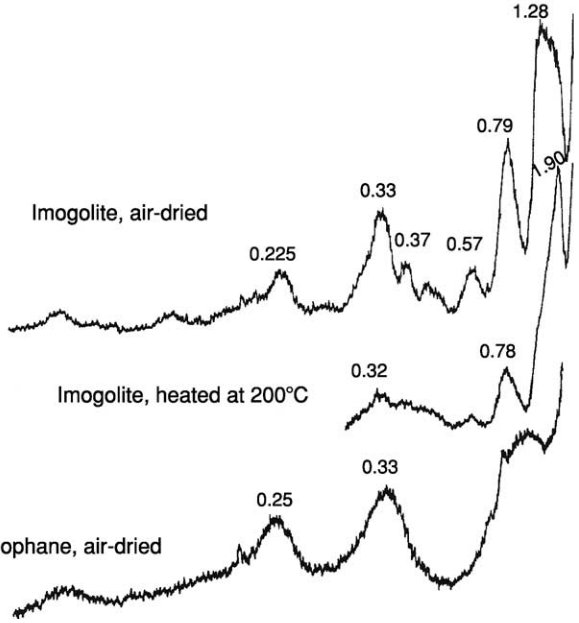


Figure 2 X-ray diffraction patterns of naturally occurring imogolite and allophane. (Reproduced from Wilson MJ (1987) *A Handbook of Determinative Methods in Clay Mineralogy*. Glasgow, UK: Blackie & Son.)

centered around 0.33 and 0.25 nm. The imogolite peaks result both from the long-range order along the longitudinal direction of the imogolite fiber and from the close packing structure of fiber bundles.

The clearest way to distinguish between allophane and imogolite is with electron microscopy. Transmission electron micrographs reveal the fibrous bundles of imogolite (Figure 3) that are readily identifiable even in complex soil mixtures. The imogolite fibers have cylindrical morphology, consisting of a gibbsite-like sheet of aluminum hydroxide octahedra pulled into a cylinder by individual silica tetrahedra. A two-dimensional (2D) ball-and-stick model in Figure 4 represents a cross-section of the imogolite tube. Allophane appears amorphous under the electron microscope, appearing gel-like (Figure 5) or as aggregates of small spherules. The structure, composition, and morphology of allophane depend largely on its Al/Si ratio.

Estimates of the quantity of SRO materials can be made by selective dissolution of these materials from soil. Extraction with acid ammonium oxalate removes Al and Si from poorly crystalline aluminosilicates and iron oxides (i.e., ferrihydrite) and Al from complexes with organic matter. Iron oxide- and organic-associated Al and Si can be removed with dithionite-citrate bicarbonate (DCB) and sodium pyrophosphate, respectively, before extraction with oxalate to quantify the amount coming from sources other than allophane and imogolite.

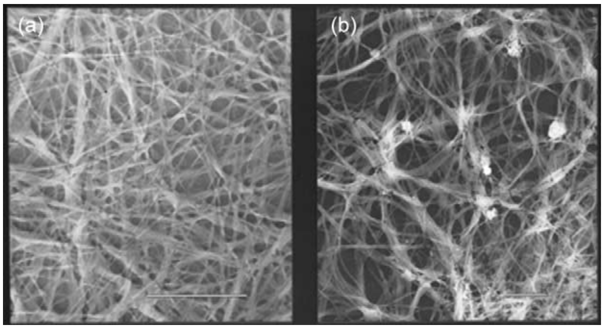


Figure 3 Transmission electron micrographs and electron diffraction patterns of (a) synthetic imogolite and (b) natural imogolite.

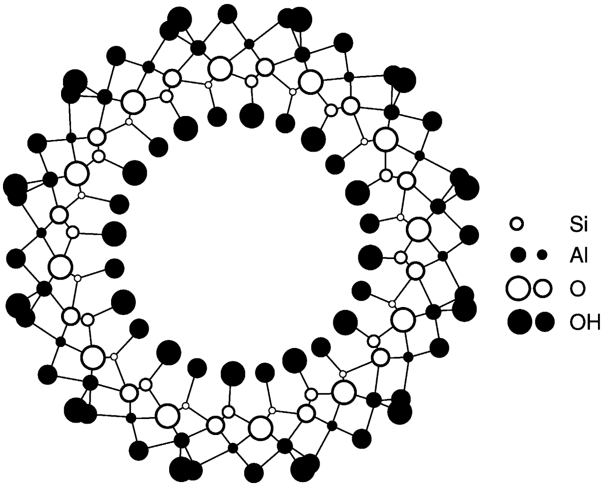


Figure 4 Crystallographic representation of the cross-section of an imogolite tube. (Reproduced from Cradwick PDG, (Farmer VC, Russell JD, Masson CR, Wada K, and Yoshinaga N (1972) Imogolite, a hydrated aluminum silicate of tubular structure. *Nature Physical Science* 240:187–189.)

Neither DCB nor pyrophosphate dissolves a significant amount of aluminosilicate material. Once the amount of Al and Si coming solely from SRO materials has been calculated, the percentage of allophane and imogolite in the soil can be estimated from:

$$\%\ \text{allophane in soil} = \text{Si}/(-0.067\,\text{Al}/\text{Si} + 0.27) \qquad [1]$$

where Si is the percentage Si from SRO aluminosilicates. This formula is based on an average Al/Si ratio for allophanes found in soils.

Imogolite can be distinguished from allophane and its concentration in soil estimated by both differential thermal analysis (DTA) and FTIR. Imogolite gives rise to a unique endothermic peak at 400°C due to structural water loss (Figure 6). The area of this peak is proportional to the quantity of imogolite in the soil. An FTIR band at 348 cm^{-1} is indicative of

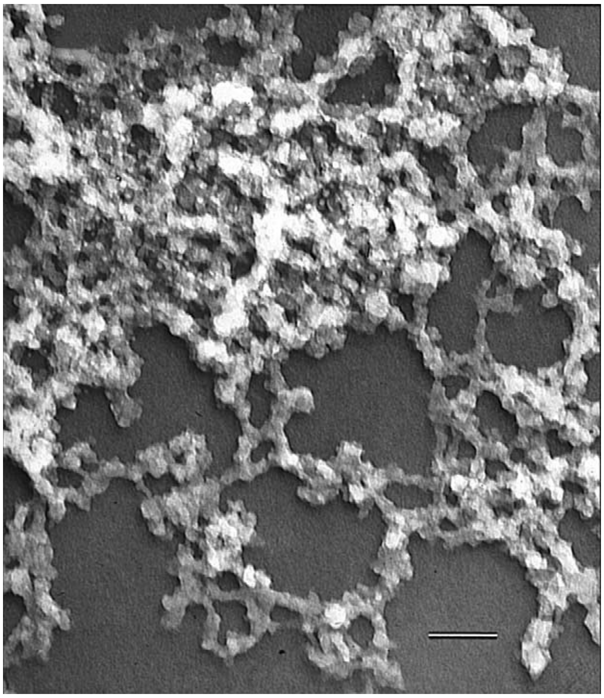


Figure 5 Transmission electron micrographs and diffraction pattern of synthetic allophane with a 1.12 Al/Si ratio.

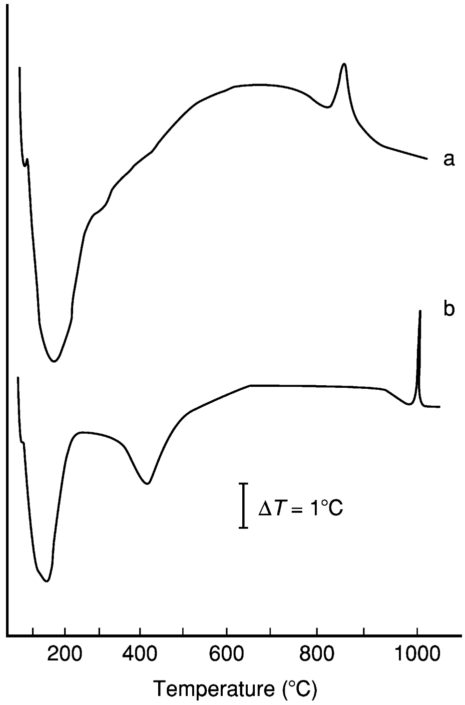


Figure 6 Differential thermal analysis curves of (a) Al-rich allophane and (b) imogolite.

the structure of both imogolite and protoimogolite allophane. This peak can be normalized to provide the quantity of 'imogolite-like structures' in a given soil.

Structure and Charge

Infrared and NMR spectra have been instrumental in helping to infer the structure of both imogolite and allophane. The IR band at 940 cm^{-1} is assigned to the unshared OH groups of the orthosilicate groups in imogolite and protoimogolite allophane. For all other allophanes, the Si-O stretching band shifts from 975 to 1020 cm^{-1} as the Al/Si ratio decreases (Figure 7). This indicates that polymerization of the silicate groups increases with increasing Si concentration. The −79 ppm NMR shift is also indicative of Si coordinated to three Al-O groups and one OH (Figure 8). The negative shift in the NMR spectrum as Al/Si decreases supports increased O sharing among Si atoms.

Polymerization is accompanied by an increase in tetrahedrally coordinated Al, as indicated by ^{27}Al-NMR and XAS spectra, inferring that the Si-rich allophanes have a feldspathoid-like structure. The feldspathoids are network silicates that, like zeolites, are characterized by high structural negative charge. Such charge also increases in allophanes as Al/Si decreases.

In Al-rich allophane and imogolite, the surface charge arises almost exclusively from variably charged aluminol (Al-OH) and silanol (Si-OH) groups. Chemically, the surface charging reactions can be represented as protonation–deprotonation reactions on these groups:

$$\text{Al-O}^- + \text{H}^+ \leftrightarrow \text{Al-OH} + \text{H}^+ \leftrightarrow \text{Al-OH}_2^+ \qquad [2]$$

$$\text{Si-O}^- + \text{H}^+ \leftrightarrow \text{Si-OH} + \text{H}^+ \leftrightarrow \text{Si-OH}^+ \qquad [3]$$

Aluminol groups in gibbsite have no net charge around pH 9, whereas silanol groups in quartz have a net neutral charge near pH 2. As a result, it is assumed that the acidic silanol groups contribute negative charge to the SRO aluminosilicates, and aluminol groups primarily bear positive charge, because, typically, the pH range of soils falls between 3 and 9.

Figure 9 shows the variable-charge behavior of an Al-rich allophane with the same Al/Si ratio as imogolite (i.e., protoimogolite). The data points represent the amount of Na and Cl adsorption determined as a function of pH and ionic strength. The magnitude of the adsorption increases with increasing ionic strength, whereas the sign of the charge depends on the pH. The pH at which Na and Cl adsorption is equal (i.e., no net surface charge) is known as the point of zero net charge (PZNC). The dotted line in Figure 9 represents the surface charge determined by adsorption of H$^+$ and OH$^-$ (σ_H). The pH at which σ_H is zero is known as the point of zero proton charge (PZNPC). The PZNC and PZNPC are equivalent

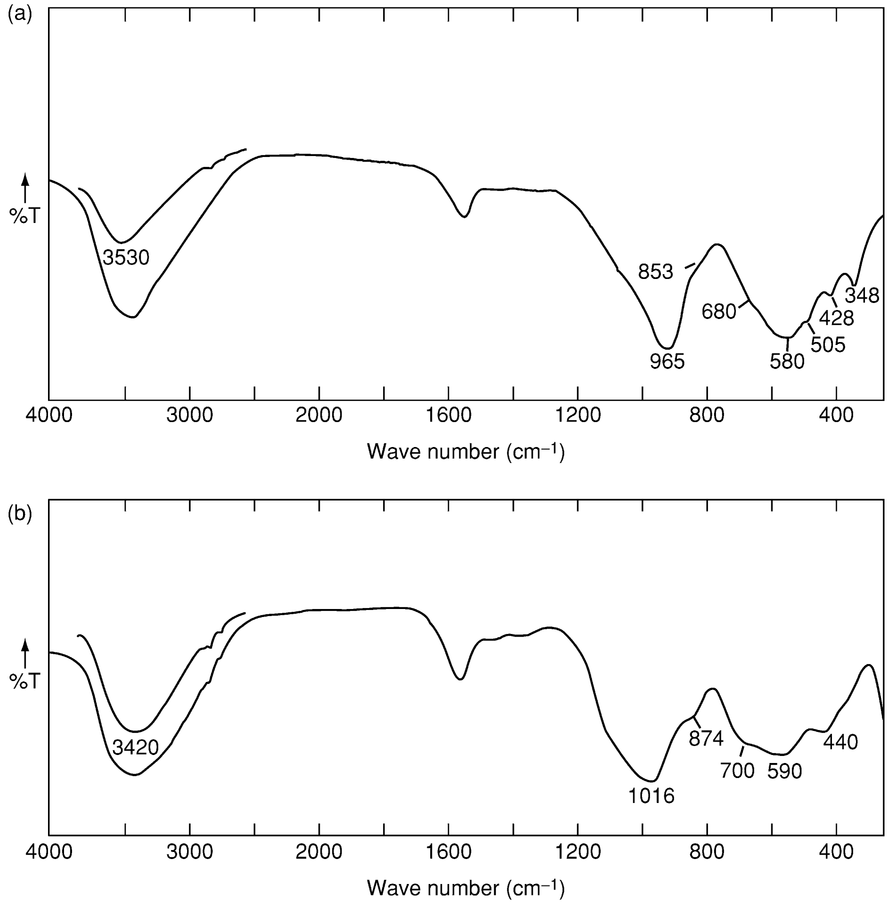


Figure 7 Infrared spectra of (a) Al-rich (protoimogolite) and (b) Si-rich allophanes. The *y*-axis is %T, which represents percentage of transmittance.

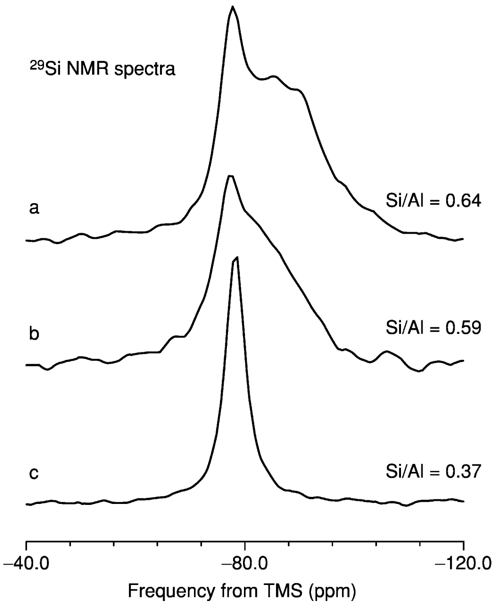


Figure 8 The ^{29}Si-NMR spectra of three allophanes with decreasing Si/Al molar ratio: A, 0.64; B, 0.59; C, 0.37.

when there is no permanent structural charge in the allophane. The fact that these two are nearly equivalent in Figure 9 also shows that Al is not substituting for Si in tetrahedral sites, giving rise to structural charge.

It is evident from Figure 9 that protoimogolite allophane is positively charged in all but calcareous and other alkaline soils. The amount of charge is significant, rivaling smectites in magnitude of charge in both acid and highly alkaline soils. Indeed, allophane and imogolite may be among the most important anion exchangers in soils. The use of soils containing allophane as filters for anions such as arsenate, iodide, and technetium has been considered. Silica-rich allophanes, on the other hand, bear significant negative structural charge and will function as important cation exchangers.

Chemisorption and Ligand Exchange

In addition to the nonspecific cation and anion exchange reactions that occur on positively and negatively charged sites, respectively, metals and ligands

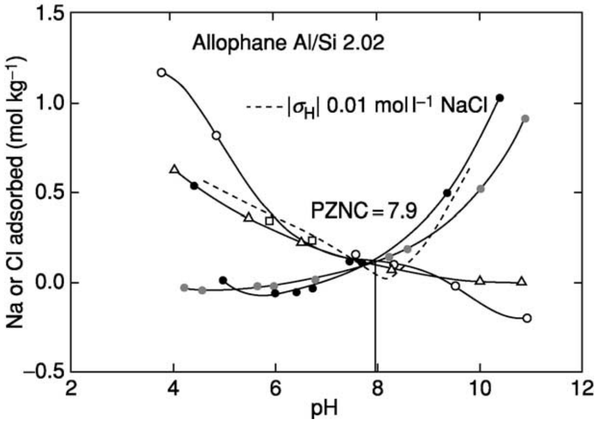


Figure 9 Ion adsorption and proton charge (σ_H) of synthetic Al-rich allophane in NaCl. (Reproduced with permission from Su C, Harsh JB, and Bertsch PM (1992) Sodium and chloride sorption by imogolite and allophanes. *Clays Clay Mineral* 40: 280–286.)

can chemisorb to the reactive aluminol and silanol groups. The nonspecific interactions on charged surfaces are characterized by the fact that the ions can be easily exchanged with other ions of like charge from the soil solution. The adsorbed ions remain hydrated (outer-sphere complexes), are kinetically labile, and equivalently balance the surface charge. Chemisorbed ions, on the other hand, bond directly to the aluminol and silanol groups (inner-sphere complexes), are not easily or rapidly exchanged, and do not require an oppositely charged surface group to adsorb. These interactions have the ability to govern the transport, stability, and bioavailability of solutes including toxic metals, essential plant nutrients, and soil organic matter.

Chemisorption of heavy metals such as Zn, Cu, Pb, Co, and Cd generally involves exchange with a proton from a surface silanol or aluminol group. An inner-sphere complex is formed as the metal bonds directly with the oxygen of the surface group. For example, on an aluminol group on allophane:

$$\text{Al-OH} + \text{M}^{2+}(\text{aq}) \leftrightarrow \text{Al-OM}^{+} + \text{H}^{+}(\text{aq}) \qquad [4]$$

represents the chemisorption of a divalent metal cation (M^{2+}) where one proton is exchanged. The result, in this example, is a more positively charged surface, a reduction in the mobility of the metal cation, and a decrease in soil pH. Given the high specific surface areas of SRO aluminosilicates and the relatively high number of reactive surface groups, allophane and imogolite can play an important role in metal behavior in soils, even when these materials are present in low concentration.

The cation exchange behavior of metals depends primarily on their charge and ionic radius. The selectivity for a negatively charged site tends to increase with both charge and size as long as the cation remains hydrated. In the case of the chemisorbed metals, however, the selectivity depends on the nature of the reaction with the aluminol or silanol group. The selectivity series for metals on aluminol groups follows the order:

$$\text{Cu}^{2+} > \text{Pb}^{2+} > \text{Zn}^{2+} > \text{Ni}^{2+} > \text{Co}^{2+}$$
$$> \text{Cd}^{2+} > \text{Mg}^{2+} > \text{Sr}^{2+}$$

This series represents the position of the adsorption edge when the amount of metal adsorbed is plotted against pH. It implies that the metals such as Cu and Pb will be chemisorbed to aluminol groups even in acid soils, whereas the alkaline metals like Mg and Sr are not chemisorbed in slightly alkaline to acid soils, but only undergo cation exchange. Chemisorption to silanol groups is similar.

Just as allophane and imogolite can be important anion exchangers in soil, they are also capable of chemisorbing anions and neutral solutes through ligand exchange:

$$\text{Al-OH}_2^{+} + \text{A}^{-} \leftrightarrow \text{Al-A}^{-} + \text{H}_2\text{O} \qquad [5]$$

In the above reaction, an anion (A^{-}) exchanges with a water molecule, increasing the negative charge on the surface. The anion bonds directly with the surface Al or Si ion, forming an inner-sphere complex. Several solutes are known to chemisorb to allophane and imogolite through ligand exchange, including phosphate, fluoride, selenite, and organic acids.

The sorption of phosphate to SRO materials has a major effect on the availability of phosphate to crops grown in allophanic soils. Although it was long assumed that this reaction was governed strictly by ligand exchange, it is more likely that phosphate reacts with allophane to form insoluble Al-phosphate phases similar to variscite ($\text{AlPO}_4 \cdot 2\text{H}_2\text{O}$). The high retention of phosphate may limit fertility of both agricultural and forested soils. Solutions to this potential problem have included not only high application of inorganic fertilizers, but addition of silica and organic matter to compete with phosphate for reactive sites.

Fluoride is strongly adsorbed to both silanol and aluminol groups and results in the release of OH^{-} by the reaction:

$$\text{Si-OH} + \text{F}^{-} \leftrightarrow \text{Si-F} + \text{OH}^{-} \qquad [6]$$

This reaction has been used as a test for the presence of SRO aluminosilicates in soil. The high concentration of reactive groups can easily result in a pH >10 when a few grams of an allophanic soil is reacted with 0.1 mol l^{-1} NaF, initially at neutral pH. Other SRO materials, such as ferrihydrite, and Al-humic complexes also contribute to this reaction. As in the reaction with phosphate, insoluble metal fluorides can replace allophane.

Reaction with Soil Organic Matter

Soils that are derived from volcanic ash or contain high amounts of SRO materials often fall into the Andisol order. The term 'andisol' originated in Japan and refers to 'dark soils.' These nearly black soils resulted from relatively high quantities of organic matter in their surface horizons. Organic matter contents as high as 20% can be found in Andisols that occur in warm and humid climates, where degradation should be rapid. The association between allophanic soils and high organic matter has long been known, but is not well understood.

The high amount of organic matter in allophanic soils has been attributed to several possible factors. One is a high concentration of Al-organic complexes, which inhibit microbial activity. These complexes not only slow organic matter decomposition, but may preclude the formation of SRO aluminosilicates. In this case, it is the Al complexes themselves that impart allophanic properties to the soil. Chemisorption of humic materials to the high-surface-area SRO materials is assumed to protect them from degradation through steric factors or by limiting microbe accessibility via occlusion. It is also possible that reaction between allophane and organic matter alters the humification process, resulting in more recalcitrant, polymethylene-type molecules.

Aluminum Solubility

The activity of Al in soils is an important consideration from an environmental standpoint, because Al is toxic to various plant species, especially agricultural crops, and to aquatic organisms. Aluminum activity increases 1000-fold for every unit decrease in pH when Al solubility is controlled by solid phases such as gibbsite, kaolinite, and imogolite. Among the soils for which a solid-phase dissolution mechanism is likely to control Al activity in the soil solution are those containing SRO aluminosilicates. The B horizons of Spodosols, in particular, near equilibrium with an imogolite-like phase as determined in both field and laboratory studies. As mentioned above, the SRO aluminosilicates are metastable with respect to more crystalline minerals such as kaolinite; however, they are persistent in soils and approach equilibrium solubility fairly rapidly in soils.

The dissolution reaction for imogolite can be written as follows:

$$0.5\mathrm{Al_2SiO_3(OH)_4(s)} + 3\mathrm{H^+(aq)} \leftrightarrow \mathrm{Al^{3+}(aq)} + 0.5\mathrm{Si(OH)_4(aq)} + 1.5\mathrm{H_2O(l)} \quad [7]$$

and the log solubility product (log K_{sp}) is:

$$\log K_{sp} = \log(\mathrm{Al^{3+}(aq)}) + 0.5\log(\mathrm{Si(OH)_4(aq)}) + 3\mathrm{pH} \quad [8]$$

where parentheses denote activities.

Aluminum solubility can then be obtained by rearranging:

$$\log(\mathrm{Al^{3+}}) = \log K_{sp} - 0.5\log(\mathrm{Si(OH)_4(aq)}) - 3\mathrm{pH} \quad [9]$$

Thus, the activity of Al^{3+} in a soil solution controlled by imogolite depends negatively on $\mathrm{Si(OH)_4(aq)}$ activity and pH.

A stability diagram based on some of the published values for imogolite and protoimogolite allophane in relation to more crystalline minerals is shown in Figure 10. The two imogolite lines are from two different investigations. The more stable of the two synthetic imogolites crosses the halloysite stability line at an aqueous silica activity near 10$^{-3.5}$, consistent with the formation of halloysite in soils with higher values. However, the formation of halloysite in preference to imogolite may be kinetic, requiring significant supersaturation. Although the solubility products of these minerals are not yet well defined, Figure 10 still shows the metastability of SRO

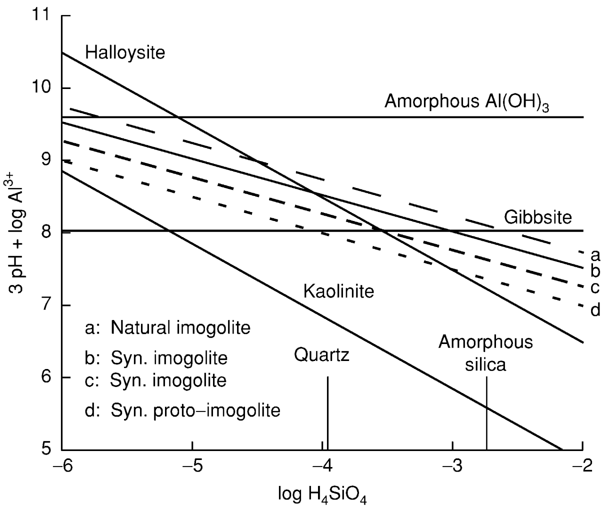


Figure 10 Stability diagram for selected Al hydrous oxide and aluminosilicate solid phases. Syn, synthetic.

materials with respect to kaolinite at nearly all soluble silica activities: the relative stability of gibbsite being higher at low ($Si(OH)_4(aq)$) and of halloysite being higher ($Si(OH)_4(aq)$).

Physical Properties

Soils high in SRO materials generally have very low bulk densities, typically less than 0.85 $Mg\,m^{-3}$. Indeed, this is a defining characteristic of Andisols or soils classified as having 'andic' properties. The low bulk densities result from high microporosity brought about by particle–particle associations and from the strong association between allophanes and soil organic matter, discussed above. Allophane particles form linear associations that flocculate into microporous aggregates (Figure 2) when near the point of zero charge or the point where the particles have zero mobility in an electric field. Individual allophane and imogolite particles have densities of 2.7–2.8 $Mg\,m^{-3}$ and 2.6–2.75 $Mg\,m^{-3}$, respectively, which are higher than for crystalline clay minerals.

The surfaces of allophane and imogolite can be assumed to be largely accessible to the soil solution. Specific surface areas determined by a polar molecule such as ethylene glycol monoethyl ether are around $10^6\,m^2\,kg^{-1}$, rivaling the specific surface of smectites. Specific surface areas of allophanes determined with N_2 are significantly lower, ranging from 0.3 to $0.7 \times 10^6\,m^2\,g^{-1}$.

The physical characteristics of SRO materials and associated organic matter have a profound effect on soils containing these materials. In addition to the low bulk densities, such soils have unusually high water retention at both 33 kPa ('field capacity') and 1500 kPa ('permanent wilting point') suction. Saturated water retention has been reported as high as 1.8 $kg\,kg^{-1}$. This can be attributed to the high microporosity and specific surface and leads to favorable conditions for plant growth, particularly in forested soils.

Meso- and macroporosity are also high in andic soils as a result of granular structure brought about by the aggregation of particles and organic matter. Consequently, saturated hydraulic conductivity is generally high in allophanic soils.

Both physical and chemical alteration of these soils can change hydraulic conductivity and other physical properties. When soils containing allophane and imogolite are air-dried, there is an irreversible increase in void volume, which results in higher saturated hydraulic conductivity and lower water retention. It is evidently difficult to rehydrate all surfaces and micropores after drying. In the field, mechanical disturbance that enhances drying or disrupts through shearing can lead to significant lowering of soil plasticity, whereas moist andic soils are considered highly plastic. Finally, adjusting soil pH to values higher or lower than the near-neutral point of zero charge of Al-rich allophanes disturbs particle–particle associations and can reduce hydraulic conductivity as much as 75% or more.

In general, the physical properties of soils containing allophanic materials are conducive to soil productivity. High water retention and hydraulic conductivity not only make water available for plant growth, but result in soils resistant to erosion. Under the wetting and drying conditions common in cropping systems, these soils tend to resist compaction by machinery and mechanical disruption by rainfall. This environment is also conducive to the physical aspects of plant growth, including seedling emergence and root penetration.

See also: **Clay Minerals**

Further Reading

Dahlgren RA (1994) Quantification of allophane and imogolite. In: Amonette J and Zelazny LW (eds) *Quantitative Methods in Soil Mineralogy*, pp. 430–451. Madison, WI: Soil Science Society of America.

Farmer VC and Russell JD (1990) The structure and genesis of allophanes and imogolite: their distribution in non-volcanic soils. In: De Boodt MF, Hayes MHB, and Herbillon A (eds) *Soil Colloids and Their Association in Aggregates*, pp. 165–178. New York: Plenum Press.

Harsh J, Chorover J, and Nizeyimana E (2002) Allophane and imogolite in soil. In: Dixon JB and Schulze DE (eds) *Soil Mineralogy with Environmental Applications*. Madison, WI: Soil Science Society of America.

Huang PM (1995) The role of short-range ordered mineral colloids in abiotic transformations of organic components in the environment. In: Huang PM, Berthelin J, Bollag J-M, McGill WB, and Page AL (eds) *Environmental Impact of Soil Component Interactions*, vol. 1, pp. 135–167. Boca Raton, FL: CRC Press.

Maeda T, Takenaka H, and Warkentin BP (1977) Physical properties of allophane soils. *Advances in Agronomy* 29: 229–269.

Shoji S, Nanzo M, and Dahlgren RA (1993) Volcanic ash soils: genesis, properties and utilization. *Developments in Soil Science* 21: 145–187.

Su C and Harsh JB (1994) Gibbs free energies of formation at 298 K for imogolite and gibbsite from solubility measurements. *Geochimica Cosmochimica Acta* 58: 1667–1677.

Wada K (1989) Allophane and imogolite. In: Dixon JB and Weed SB (eds) *Minerals in Soil Environments*, 2nd edn, pp. 1051–1087. SSSA Book Series. Madison, WI: Soil Science Society of America.

ANAEROBIC SOILS

P W Inglett, K R Reddy, and R Corstanje, University of Florida, Gainesville, FL, USA

Anaerobic soils occur in areas where oxygen consumption by soil biota exceeds the diffusion of oxygen into the soil profile. This condition is also termed 'soil anaerobiosis' and results in a predominantly oxygen-free environment in the soil profile. Anaerobic soils occur in a number of environments in the landscape, including: wetlands; paddy soils; organic soils; poorly drained, heavy textured soils; areas with a high water table; soils amended with heavy rates of organic materials such as animal wastes, biosolids, and composts; and soils treated with ammoniacal fertilizers.

In upland environments, anaerobic soil conditions may be temporary and may not last more than a few days; while, in wetland environments, soil anaerobic conditions last for several months. Thus anaerobic soils include: all types of wetland soils (swamps, marshes, floodplains, coastal wetlands, and bottomland hardwood forests), hydric soils, paddy soils, organic soils, and any other waterlogged or flooded soils. Because much of our knowledge of anaerobic soils has been gained through research in wetlands and rice paddies, the discussion in this paper will largely focus on the morphological and biogeochemical features of these types of soils.

Wetland soils are widely distributed throughout the world and can be found in all climates, ranging from the tropics to tundra, with the exception of Antarctica. Approximately 6% of the Earth's land surface, which equals approximately 800 million hectares (approx. 2 billion acres) is covered by wetlands. The USA alone contains approximately 12% of the world's wetlands, or approximately 111 million hectares (274 million acres). In any given landscape, wetlands are located in areas with a low elevation and a high water table. Wetlands can be broadly defined as marshes, swamps, bogs, and similar areas. These areas are poorly drained and retain water during rainy periods. Thus, the physical, chemical, and biological characteristics of anaerobic soils are important in determining the properties and functioning of wetlands.

The creation of anaerobic soil conditions is predicated in the situation where demand exceeds the supply of oxygen. Once a soil becomes saturated, the supply of oxygen is immediately reduced owing to the displacement of oxygen contained in the available pore space. Following consumption of the relatively small amount of available oxygen in the pore water, oxygen can only be supplied to respiring organisms through the process of diffusion from the nearest aerobic zone. This process is comparatively slow under saturated soil conditions as oxygen diffusion in water is approximately 10 000 times slower than through air. Under these conditions, even moderate rates of soil or root respiration can quickly deplete available oxygen and result in anaerobic soil conditions.

Depending on hydrologic conditions, wetland soils can be present: (1) flooded, with defined water depth above the soil surface; (2) under saturated soil conditions, with no excess floodwater; and (3) when the water table lies below the soil surface at a certain depth, depending on soil characteristics. Under the first two conditions, wetland soils can be classified as hydric soils, while the third group can mimic the characteristics of both wet- and upland soils, depending on soil type and hydrologic conditions. Soil taxonomy classifies soils with these characteristics into a suborder 'aquic,' which implies that soil pores are filled with water (from soil surface to a depth of 2 m), and many of the oxidized compounds are enzymatically reduced, with end products of these reductive processes accumulating in the soil. Soil taxonomists classify aquic soils according to soil color and not the accumulation of reduced products. Gray colors or low chroma (2) are used generally as indicators of soil anaerobiosis.

Physical Characteristics

Soil volume primarily comprises solid matter, water, and air. When soils are flooded, most of their pore volume is occupied by water. Upland mineral soils generally consist of about 50% by volume of solids, 25% of water, and 25% of air. In wet mineral soils, approximately 50% of the soil volume is solids, while the remaining 50% is occupied by water. In wetland organic soils, a large proportion (up to 80%) of soil volume is occupied by water, with soil organic matter and mineral matter occupying less than 20%.

Generally, reduced compounds are not found in upland soils. Gaseous exchange is not restricted because of continuation of air spaces in upland soils, and oxygen dominates the respiratory and chemical environment. Gaseous composition of soil pores is approximately 10–21% O_2, 0.03–1% CO_2, and trace amounts of N_2O and NH_3. In wetland soils, there is less oxygen, because soil pores are filled with water. In the absence of oxygen, reduced

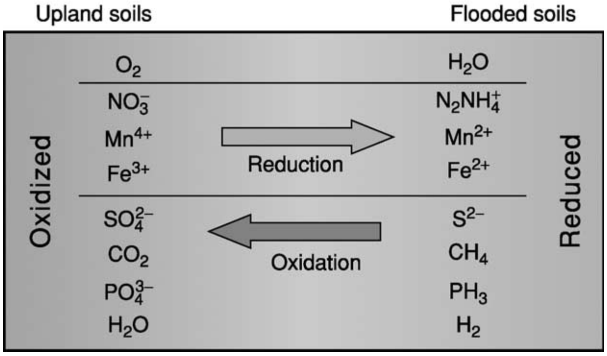


Figure 1 Various inorganic oxidized and reduced compounds in upland and wetland soils.

chemical forms predominate and are regulated by associated biogeochemical processes (Figure 1). In recently flooded soils, N_2O can be present as a result of denitrification of nitrate nitrogen (NO_3-N). In moderately reduced soils, H_2S can be observed, followed by CH_4 under more reducing conditions. In highly reduced soils, C_2H_4 and PH_3 (phosphine) can be observed. However, the presence and accumulation of these gases depend on respective oxidants available for reduction.

Biological Characteristics

Saturated soil conditions support microbial populations adapted to anaerobic environments. Aerobic microbial populations are restricted to zones where O_2 is available. Most of the aerobic organisms become quiescent or die, and new inhabitants, largely facultative (organisms which can function under both aerobic and anaerobic environments) and obligate anaerobic bacteria, take over.

Fungi, which are active in upland environment, are inhibited in the anaerobic wetland soil environment. This is primarily due to absence of O_2 and alteration in soil pH (acid to neutral) under anaerobic conditions. Similarly, microbial biomass decreases under saturated soil conditions. This decrease in microbial activity is primarily due to the shift from aerobic to anaerobic respiration. Thus, under wetland soil conditions, rates of many microbially mediated reactions decline, and some reactions may be eliminated and replaced by new ones.

Saturated soil conditions can support the growth of microphytic communities, including a variety of planktonic, epiphytic, and benthic algae at the soil–floodwater interface. The species composition varies with physicochemical conditions within the wetland. Many of the species in these microbial assemblages have the capacity to carry out photosynthesis. Diel fluctuations in dissolved O_2 produced as a consequence of photosynthesis often increase the O_2 levels in the floodwater beyond saturation levels during daytime and to low levels during nighttime. These large fluctuations in available oxygen have special significance in wetlands for regulating biogeochemical cycles of nutrients.

Wetland or anaerobic soil conditions also support the presence of hydrophytic vegetation, or plants that are adapted to the reducing wetland environment. These plants have unique characteristics to adapt to oxygen-deficient conditions, including physiological adaptations (such as capability to respire anaerobically), anatomical adaptations (such as development of intercellular air spaces), and morphological adaptations (such as water roots and adventitious roots). With these adaptations, hydrophytic plants are able to survive under reducing conditions considered toxic to other macrophytes. In many cases, adapted plant communities become the dominant source of organic matter in wetland systems.

Chemical Characteristics

When oxygen availability becomes limited, bacteria must utilize other compounds as electron acceptors to maintain their metabolism. These compounds, many of which are nutrient elements, can exist in both dissolved and solid phases, and include oxidized forms of elements such as N, Fe, Mn, and S. As they are utilized during respiration, these elements gain electrons and thus become chemically reduced. The result of microbial metabolism therefore is the conversion of oxidized elements to the corresponding reduced form under anaerobic conditions. When wetland soils are drained, many of the reduced compounds are oxidized either by chemical or biochemical reactions. Therefore, in upland and/or drained soils, oxidized forms of chemical species dominate the system, while reduced forms dominate the wet soil system (Figure 1).

Reduction–oxidation, or redox potential (E_h), reflects the intensity of reduction *or* a measure of electron (e^-) activity analogous to pH (which measures H^+ activity). Depending on soil characteristics, E_h generally decreases with time and approaches a steady value after flooding. Redox potential is the most common parameter used to measure degree of soil wetness or intensity of soil anaerobic conditions. The range of E_h values observed in wetland soils is from $+700$ to -300 mV (Figure 2). Negative values represent high electron activity and intense anaerobic conditions typical of permanently waterlogged soils. Positive values represent low electron activity and aerobic to moderately anaerobic conditions typical of wetlands in transition zone. Soils with

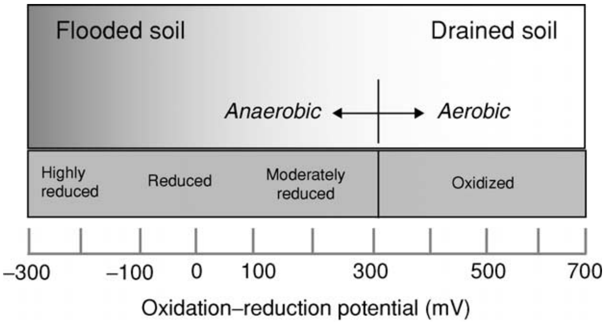


Figure 2 Schematic showing relationship between oxidation–reduction potential and oxidized and reduced conditions in wetland soils.

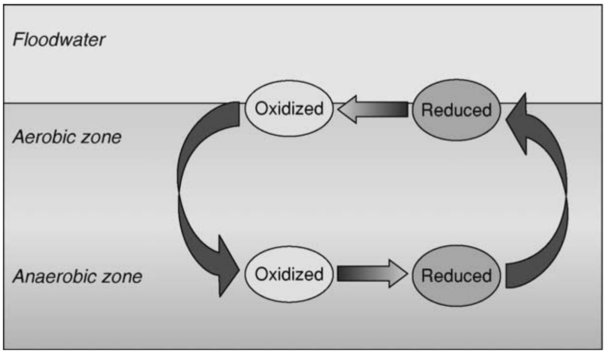


Figure 3 Diffusional patterns of reduced and oxidized compounds in response to anaerobic gradients in flooded soils.

$E_h > 300$ mV are considered aerobic or upland. Under these conditions, oxygen is used as the dominant electron acceptor. Soils with $E_h < 300$ mV are considered anaerobic or wetland.

The chemical nature of the reduction process also affects soil pH, electrical conductivity, cation exchange capacity, and sorption and desorption processes. In general, saturated soil conditions result in an increase in pH, electrical conductivity, and ionic strength, but a decrease in soil redox potential. The pH of most soils tends to approach the neutral point under flooded conditions, with acid soils increasing and alkaline soils decreasing in pH. Increase in pH of acid soils depends on the activities of oxidants (such as NO_3^-, Fe^{3+}, Mn^{2+}, and SO_4^{2-}) and proton consumption during reduction of these oxidants under flooded conditions. In alkaline soils, pH is controlled (and generally lowered) by the accumulation of dissolved CO_2 and organic acids.

Accumulation of reduced compounds in the anaerobic soil layer results in the establishment of concentration gradients across the aerobic–anaerobic interface. The concentration of reduced compounds is usually higher in the anaerobic layer, which results in upward diffusion into aerobic soil or floodwater, where they are oxidized (Figure 3). Similarly, some of the dissolved, oxidized compounds diffuse downward, i.e., from floodwater or aerobic soil layer into underlying anaerobic soil layer, where they will be reduced. The exchange rates between soil and overlying water determine whether the wetlands soils or sediments are functioning as a sink or source for nutrients. The rate of exchange of dissolved species depends upon: (1) concentration of dissolved species in soil pore water; (2) soil type and other related physicochemical properties (pH, cation exchange capacity, organic matter content, and bulk density); (3) concentration of dissolved species in the floodwater; and (4) kinetics of related biogeochemical processes in soil and floodwater.

Morphological Characteristics of Wetland Soils

Wetland protection now requires identification of the boundaries between uplands and wetlands. Criteria based on hydrology, vegetation, and soils are individually or together used to determine these boundaries. Among these three components, soils assessment is particularly critical because, while vegetation and hydrology are temporally affected by climatic fluctuations, soils are the most stable and respond only to long-term inundation. The term ‘hydric soils’ is now commonly used in jurisdictional language synonymous with wetland soils. Hydric soils are defined as soils formed under conditions of saturation, flooding, or ponding long enough during the growing season to develop anaerobic conditions in the upper part of the soil profile.

Saturated soils develop several unique morphological characteristics as a result of several oxidation–reduction reactions. These features are now used as soil indicators to evaluate independently wetland boundaries. Some of the key hydric soil indicators are formed by the accumulation or loss of iron and manganese, hydrogen sulfide, or accumulation of organic matter. In many cases, soil color is used to assess both the accumulation of organic matter (dark horizons) and the reduction of iron species (formation of gray or gley colors).

Biogeochemical Characteristics

Microbial communities in anaerobic soils play a key role in regulating a number of essential biogeochemical cycles such as carbon, nitrogen, phosphorus, and sulfur. Organic matter released by the primary producers is degraded by microbial communities, releasing nutrients back into the environment. Degradation of organic materials allows heterotrophic microbial groups to obtain energy and nutrients

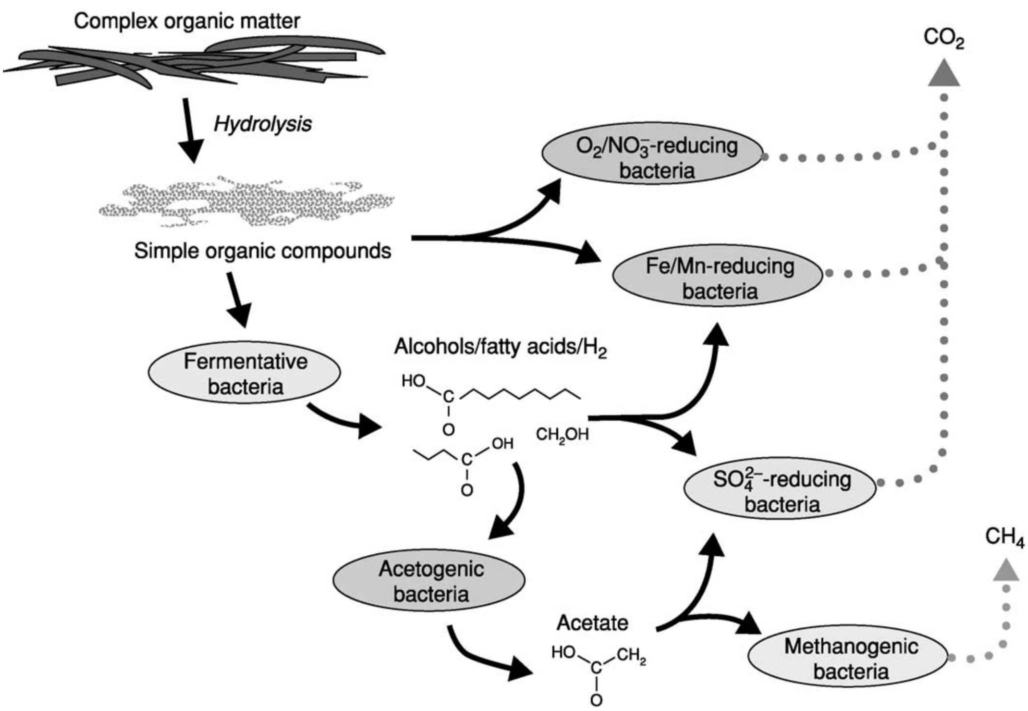


Figure 4 Pathways of organic matter decomposition in wetland soils.

for growth. Enzymatic hydrolysis of organic matter produces several monomers, including glucose, xylose, fatty acids, and amino acids. These simple reduced compounds then serve as substrates for microbial metabolism as chemical energy is released during their oxidation.

Anaerobic decomposition of organic matter is more complex and less energetically favorable than aerobic decomposition. It is mediated by a complex group of physiologically different microorganisms which participate in the decomposition pathways (Figure 4). Often, the end product of one metabolism is substrate for the next until the decomposition is complete. Initially, fermenting bacteria mediate the extracellular hydrolysis of high-molecular-weight polymers (i.e., proteins, polysaccharides, lipids, and nucleic acids) and ferment their respective monomers (e.g., amino acids, sugars, fatty acids, and nucleotides) to CO_2, H_2, acetate, propionate, butyrate, and other fatty acids and alcohols.

Microbial Respiration

Oxygen, if present in wetland soils, provides an electron acceptor (oxidant) for supporting microbial oxidation of reduced carbon compounds (respiration). Thermodynamically, oxygen is the most preferred electron acceptor by microorganisms (Table 1), and as a result, aerobic microorganisms maintain a competitive advantage while oxygen is present. Oxygen is used preferentially because it receives electrons from the reductant material (organic matter) more readily than do other oxidants. The greater energy yield during the aerobic process is due to: (1) complete oxidation of carbon atoms in the organic substrates to CO_2, and (2) the oxygen redox couple having a relatively high, positive reduction potential. This leads to a large net difference in electrical potentials between electron donor (organic substrate) and terminal electron acceptor (oxygen).

Once oxygen is depleted in the soil, bacteria must respire anaerobically. Under these conditions, bacteria capable of utilizing the electron acceptor with the next-highest thermodynamic potential dominate. Thus, the microbial use of electron acceptors proceeds in a sequential manner dependent on their electron affinity, energy yield, and related enzyme systems in the bacteria. In this manner, the order of electron acceptor use following oxygen depletion is: NO_3^-, Mn^{4+}, Fe^{3+}, SO_4^{2-} and finally CO_2 (Figure 5, Table 2). Anaerobic respiration is typically reflected in vertical profiles of these electron acceptors, with soil and/or sediment depth, as the most favorable electron acceptors are utilized first. However, the exclusion of less-favorable respiration pathways is not complete, resulting in considerable overlap between the different forms of organic carbon mineralization. The rate at which electron acceptors are consumed in soil systems depends on their concentration, the availability of organic compounds, and the activity of the microbial population involved in the reductive processes.

Table 1 Microbial groups involved in various redox reactions in wetland soils

Redox potential (mV)	*Electron acceptor*	*Decomposition end products*	*Microbial groups*
Aerobic			
>300	O_2	CO_2, H_2O	Aerobic fungi and bacteria
Fermenting			
less than -100 to $+300$	Organics	Organic acids, CO_2, H_2, alcohols, amino acids	Fermenting bacteria
Facultative anaerobic			
100–300	NO_3^-	N_2O, N_2, CO_2, H_2O	Denitrifying bacteria
	Mn^{4+}	Mn^{2+}, CO_2, H_2O	Mn(IV) reducers
	Fe^{3+}	Fe^{2+}, CO_2, H_2O	Fe(III) reducers
Obligate anaerobic			
less than -100	SO_4^{2-}	HS^-, CO_2, H_2O	Sulfate reducers
	CO_2 and acetate	CH_4, CO_2, H_2O	Methanogens
	Organic acids	Acetate, CO_2, H_2	H_2-producing bacteria

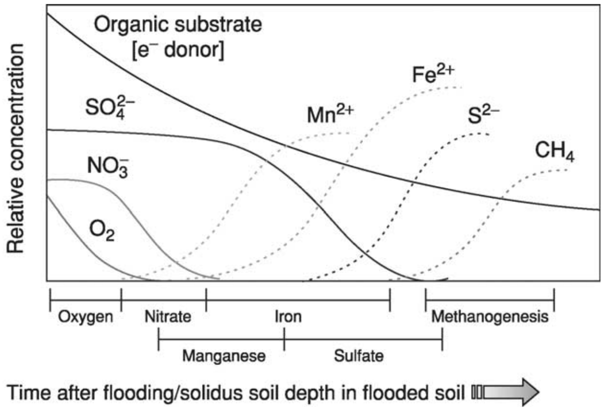


Figure 5 Sequential reduction of oxidants (oxidized compounds) and accumulation of reductants (reduced compounds) in wetland soils.

Nitrate reduction can occur in wetlands according to two major pathways:

- Dissimilatory nitrate reduction to ammonia (DNRA):

$$NO_3^- \rightarrow NO_2^- \rightarrow NH_4^+$$

- Denitrification:

$$NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$$

Dissimilatory reduction of NO_3^- is performed by a variety of facultative anaerobic bacteria. During this process, NO_3^- is first converted to nitrite NO_2^-, which may be further reduced to NH_4^+. Reduced NH_4^+ produced through dissimilatory NO_3^- reduction results in high NH_4^+ levels characteristic of wetland soils and sediments. Denitrifiers are heterotrophic bacteria (most of them facultative anaerobic) that couple the oxidation of organic substrates to the reduction of NO_3^- to either N_2O or N_2. This reaction occurs in moderately reduced conditions in the absence of oxygen and is one of the dominant mechanisms for removal of nitrogen from aquatic systems.

The oxidation and reduction of iron in many soils is possibly one of the main components of soil formation. Its relatively ubiquitous presence in soils and sediments makes this respiratory pathway a major contributor of organic-matter mineralization. A large variety of microorganisms are capable of reducing iron, including fungi. When Fe(III) is reduced, Fe(II) is the reduced end product. For Mn, Mn(II) is generally accepted as the end product of Mn(IV) reduction; however, Mn(III) may also be encountered as an intermediate species.

In the general Fe(III) and Mn(IV) reduction model, complex organic matter is hydrolyzed to smaller components (i.e., sugars, amino acids, fatty acids). The sugars and amino acids are metabolized by fermentative microorganisms, which may reduce a small amount of Fe(III) or Mn(IV) in the process. The majority of the primary products from this first stage of the metabolism of sugars and amino acids are short-chain fatty acids and possibly hydrogen. This hydrogen can then be oxidized by Fe(III) and Mn(IV) reducers (e.g., *Pseudomonas* sp.), while other fermentation products are oxidized through Fe(III) or Mn(IV) reduction by species such as *Shewanella putrefaciens*. Alternatively, *Thiobacillus thiooxidans* or *T. ferrooxidans* can reduce Fe(III) or Mn(IV), with elemental sulfur S^0 as the electron donor.

Sulfate reducers are obligate anaerobes that couple oxidation of organic substrates to CO_2 with the reduction of terminal electron acceptor SO_4^{2-} to sulfides ($-S^{2-}$). Gram-negative bacteria such as *Desulfobacterium*, *Desulfobulbus*, and *Desulfotomaculum* are the most common types of sulfate-reducing bacteria in freshwater sediments. Sulfate-reducing bacteria cannot

Table 2 Summary reactions for microbial respiration pathways in wetland soils

Electron acceptor	*Reaction coupled to glucose oxidation*	ΔG_l^o *(kJ mol^{-1})*
O_2	$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O$	−2879
NO_3^-	$5C_6H_{12}O_6 + 24NO_3^- + 24H^+ \rightarrow 30CO_2 + 12N_2 + 42H_2O$	−2713
MnO_2	$C_6H_{12}O_6 + 12MnO_2 = 6CO_2 + 12M_n^{2+} + 18H_2O$	−1916
$Fe(OH)_3$	$C_6H_{12}O_6 + 24Fe(OH)_3 + 48H^+ = 6CO_2 + 24Fe^{2+} + 66H_2O$	−418
SO_4^{2-}	$C_6H_{12}O_6 + 4H_2O \rightarrow 2CH_3COO^- + 2HCO_3^- + 4H_2 + 4H^+$	−207
	$CH_3COO^- + SO_4^{2-} + 3H^+ = 2CO_2 + H_2S + 2H_2O$	−63
CO_2	$C_6H_{12}O_6 + 4H_2O \rightarrow 2CH_3COO^- + 2HCO_3^- + 4H_2 + 4H^+$	−207
	$2CH_3COO^- + 2H_2O \rightarrow 2CH_4 + 2HCO_3^-$	−31

synthesize enzymes to hydrolyze polymers such as polysaccharides. Also, many groups of sulfate-reducing bacteria cannot use monomers such as monosaccharides (e.g., glucose) as substrates for energy, and thus sulfate reducers are dependent on fermenting bacteria to produce simple organic compounds (e.g., acetate, propionate).

Sulfate reducers are widely studied groups of microorganisms with special significance in coastal wetland ecosystems due to the high concentration of sulfate in seawater. Sulfate reduction can occur over a wide range of pH, temperature, and salinity. One product of sulfate reduction, hydrogen sulfide, is extremely toxic to aerobic organisms, because it reacts with the heavy metal groups of the cytochrome system. Hydrogen sulfide is very reactive with metals and usually results in the precipitation of metallic sulfides (e.g., FeS).

The terminal step in the anaerobic degradation of organic macromolecules, in the absence of all other electron acceptors, is the conversion of acetate and H_2/CO_2 to methane. This is an intricate process involving a net of interactions, possibly encompassing the largest set of microbial dependencies. Methanogens are obligate anaerobes that grow autotrophically (they use CO_2 as C source and as electron acceptor) and heterotrophically (they use organic substrates as energy source).

Like sulfate reducers, methanogens cannot directly utilize large-molecular-weight polymers; so methanogens must depend on at least three groups of microbes, including hydrolytic, fermentative, and H_2-producing acetogenic bacteria. Methanogens are typically found in the archaeal families of Methanobacteriaceae, Methanomicrobiaceae, Methanosaeteaceae, and Methanosarcinaceae.

Fermentation pathways vary depending on the original substrate, and quantity and presence of alternate electron acceptors. Denitrification, and Fe(III) and Mn(IV) reduction may utilize any of these fermentation products as the final step in respiration. Acetate and H_2, and other small organic acids are utilized directly by sulfate-reducing bacteria, while methanogens can only use acetate and H_2. Acetogenic bacteria cleave organic acids and alcohols into acetate, H_2, and CO_2. This conversion is only possible in the presence of sulfate-reducing bacteria or methanogens that consume H_2, resulting in low hydrogen concentrations, ensuring that acetogenesis is thermodynamically favorable.

Agronomic, Ecologic, and Environmental Significance

Anaerobic soils occupy an important niche in the biosphere, and their importance in wetlands and paddy soils is widely recognized by scientists, environmental managers, and policy-makers. Agronomically, anaerobic soils commonly known as paddy soils are widely used throughout the world for rice production. Anaerobic soils in wetlands are primary drivers of natural ecosystem function, as many of the biogeochemical processes have important feedback to ecosystem productivity and function.

Anaerobic soils are primary nutrient sources to plants grown in the paddy soils or wetlands. The decomposition process described here results in production of bioavailable nitrogen and phosphorus, which supports the productivity of plants. Furthermore, the extent of Fe(III) and/or Mn(IV) reduction can strongly influence the distribution of toxic trace metals and availability of P.

Environmentally, anaerobic soils may have both positive and negative attributes. One negative aspect is that wetlands are one of the primary sources of methane, a potent greenhouse gas. Approximately 25% of methane emitted to the atmosphere is derived from wetlands. Alternatively, anaerobic soils in wetlands also function as sinks, sources, transformers of nutrients and contaminants, and their role in improving water quality is widely recognized. This function of anaerobic soils has resulted in developing low-cost constructed wetland technology for water treatment. At present several thousands of such wetlands are in operation throughout the world.

See also: **Carbon Cycle in Soils:** Dynamics and Management; **Hydric Soils**; **Microbial Processes:** Environmental Factors; **Nitrogen in Soils:** Cycle; **Organic Soils**; **Paddy Soils**; **Sulfur in Soils:** Overview; **Wetlands, Naturally Occurring**

Further Reading

Cowardin LM, Carter V, Golet FC, and LaRoe ET (1979) *Classification of Wetlands and Deepwater Habitats of the United States*. Fish and Wildlife Service, US Department of Interior. Washington, DC: US Government Printing Office.

Federal Register, Notice (1994) *Changes in Hydric Soils of the United States*. Federal Register, vol. 59. Washington, DC: US Government Printing Office.

Federal Register, Notice (1995) *Hydric Soils of the United States*. Federal Register, vol. 60. Washington, DC: US Government Printing Office.

Holmer M and Storkholm P (2001) Sulphate reduction and sulphur cycling in lake sediments: a review. *Freshwater Biology* 46: 431–455.

Lovley DR (1991) Dissimilatory Fe(III) and Mn(IV) reduction. *Microbiological Reviews* 55(2): 259–287.

National Technical Committee for Hydric Soils (1991) *Hydric Soils of the United States*. USDA–SCS Miscellaneous Publication 1491. Washington DC: USDA–SCS.

Ponnamperuma FN (1972) The chemistry of submerged soils. *Advances in Agronomy* 24: 29–96.

Reddy KR and D'Angelo EM (1994) Soil processes regulating water quality in wetlands. In: Mitsch W (ed.) *Global Wetlands – Old World and New*, pp. 309–324. New York: Elsevier.

Reddy KR, D'Angelo EM, and Harris WG (2000) Biogeochemistry of wetlands. In: Sumner ME (ed.) *Handbook of Soil Science*, pp. G89–G119. Boca Raton, FL: CRC Press.

Schwertmann U (1993) Relations between iron oxides, soil color, and soil formation. In: Bigham JW and Crolkosz EJ (eds) *Soil Color*, pp. 51–70. Madison, WI: Soil Science Society of America.

US Department of Agriculture, Natural Resource Conservation Service (1998) *Field Indicators of Hydric Soils in the United States*. version 4.0. Ft. Worth, TX: USDA–NRCS.

Vepraskas MJ (1992) *Redoximorphic Features for Identifying Aquic Conditions*. Technical Bulletin 301, North Carolina Agricultural Research Service. Raleigh, NC: North Carolina State University.

Anion Exchange *See* **Cation Exchange**

APPLICATIONS OF SOILS DATA

P J Lawrence, USDA Natural Resources Conservation Service, Washington, DC, USA

Introduction

Soil survey, classification, and interpretation efforts are well established in countries around the world, and play significant roles in land use and natural resource decision-making. Scientific inquiry into soils began in the nineteenth century, with centers in Western Europe, Russia, and the USA. The early US pioneers in soil science and survey emphasized the soil genesis concepts developed in Russia in the 1870s and worked to validate and apply these landmark principles. Soils interpretations bring information on tangible and measur- able soil properties together with desired uses to make predictions about site suitability or limitations. Applications of soils data and survey interpretations have steadily broadened in range and precision, from the earliest interpretations that identified potential for salt accumulation in soils to today's sophisticated uses for precision agriculture and evaluating nonpoint-source pollution risks. This article examines the development and uses of soils interpretations and data in the USA.

Development of Interpretations in the USA

Interpretations were emphasized to varying degrees throughout the history of the US soil survey. Milton Whitney, the first Chief of the Soil Survey Division, US Department of Agriculture (USDA), started the systematic survey of soils in four areas of the country in 1899. These early surveys concentrated on measurable soil characteristics that correlated with the properties Whitney thought important: soil texture, soil moisture, soil temperature, and concentration of soluble salts in soils.

The most accurate and economically valuable interpretation early soil surveyors could make was to identify the potential for 'alkali' soils – reflecting the presence of soluble salts in soil and water. In 1901 and 1903, interpretations from work in the Imperial Valley of Southern California accurately identified large areas of heavy-textured impervious soils of high salt content that would pose substantial limitations for drainage and reclamation. This report coincided with a large-scale irrigation project promoted by an influential company. A subsequent survey for the Modesto-Turlock area of the San Joaquin Valley in California (1909) also reported 'alkali' presence. Predictably, such information was not always popular with landowners, developers, and others who perceived that the information diminished land values and development opportunities.

Whitney's successor, Curtis Marbut, generally deemphasized the federal role in interpretations, but worked assiduously at establishing the concepts of soil genesis first introduced in Russia by Dokuchaiev, developing a soil classification system, and establishing the scientific reputation of the published soil survey. State governments, through cooperating soil surveyors and other agricultural scientists, took the lead in preparing and publishing interpretations of completed surveys. A common theme was to identify soils particularly suited to certain crops, for example, apples or tobacco. Other federal agencies, such as the Bureau of Reclamation and the Bureau of Public Roads, established cooperative studies with the Bureau of Soils in acquiring soils information for their particular uses.

Despite Marbut's view on interpretations, some effort was made to develop soil erosion interpretations. In the late 1920s, as Hugh Hammond Bennett was escalating his campaign for research on erosion and soil conservation measures, H.E. Middleton of the Bureau of Chemistry and Soils was making substantial progress toward understanding the complex processes related to erodibility. Bennett identified areas where the combination of geology and agricultural practices combined to produce serious soil erosion and he took this discovery to both the public and politicians (Figure 1). Subsequently, Congress authorized a series of soil erosion experiment stations where interdisciplinary teams of researchers measured erosion conditions under different crops, soils, rotations, and a variety of structural and management practices. These pioneering studies added to interpretations and led to national-level soil erodibility data – the origins of the erodibility data that support current conservation planning tools such as the Universal (and the Revised) Soil Loss Equation.

When Charles E. Kellogg became Chief of the Soil Survey Division in 1935, the period of scant federal interest in soil interpretation ended. Kellogg brought new vigor to developing soil interpretations, which coincided with Bennett's success in influencing Congress to create the Soil Conservation Service (SCS). (The Soil Conservation Service was renamed the Natural Resources Conservation Service (NRCS) in 1994.) Kellogg felt strongly that pedologists had a

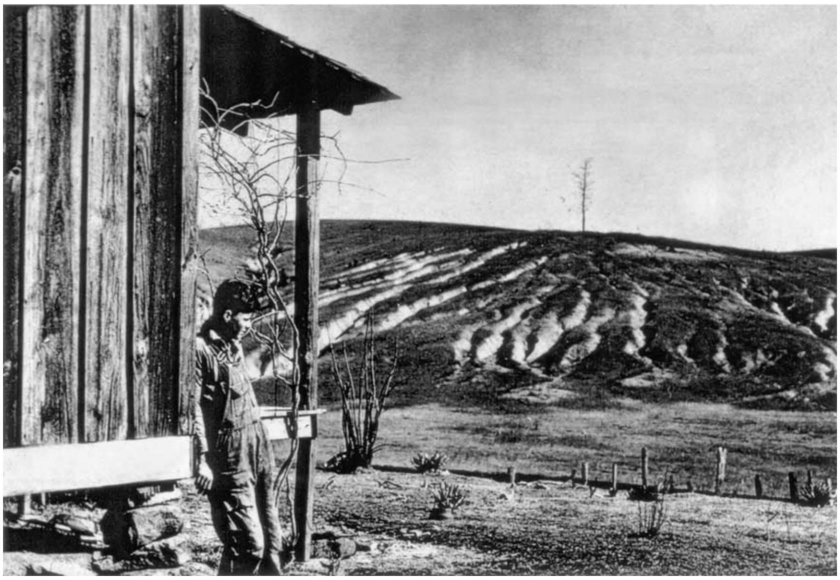


Figure 1 The combination of severe drought, geology, and agricultural practices contributed to severe erosion problems in the US Dust Bowl days. (Photo: Location unknown, 1930s). Courtesy of the Natural Resources Conservation Service USDA.

responsibility to summarize and make public the results of their work. As a student at Michigan State University, he had worked on the 1922 Michigan Land Economic Survey, a landmark in multidisciplinary research and inventorying for land classification. Kellogg also established a policy that soil surveys would include productivity ratings. Earlier attempts to develop natural or inherent productivity ratings had met with criticism as interpreters wrestled with how to rate soils that were productive under certain management but nonarable otherwise. Under Kellogg, productivity as an interpretation was to be relative to the technology and management applied. Kellogg's re-energized emphasis on interpretations was timely, corresponding with the emergence of new agricultural technologies that when matched to refined differentiation of soil types could generate substantial productivity increases.

The 1940s found the USDA in the position of having two soil survey efforts: the federal component of the National Cooperative Soil Survey (NCSS), which was aimed at producing a national inventory of soils, and the more localized farm planning surveys of soils that were conducted by SCS for use in assisting farmers and ranchers to apply soil-conserving practices. While the general soil classification framework was similar, scale, legends, and interpretations differed considerably. In 1945, USDA made clear that its requirements for legend preparation, field reviews, and soil correlation applied only to the NCSS. The SCS soil maps produced for farm planning would not be integrated into the NCSS national portrait of soil resources. Later, however, the two survey efforts were united. In 1952, the Federal NCSS component (the Soil Survey Division) was integrated into SCS; mapping was accelerated, and interpretations benefited from the expertise of a large cadre of conservationists with training in range science, agronomy, forestry, biology, engineering, and other disciplines. The merger also linked the soil survey to a major user group – the agricultural landowners with whom the SCS conservationists worked directly (Figure 2).

During this period, a new soil classification system was being developed under the leadership of Guy D. Smith. The new system, *Soil Taxonomy*, retained the deep-rooted concept of the soil series, but made classification more quantitative, and provided a more accurate, clearly defined basis for making interpretations.

Interpretations Beyond Agriculture

Before the 1950s, the primary applications of soil surveys were farming, ranching, and forestry, although some states recognized applications for highway planning as early as the late 1920s. In the post-World War II environment, however, nonagricultural interpretations flourished. Rapid urbanization brought forth example after example of problems stemming from development on poorly suited sites. Soil scientists made the point that soils information could be used to avoid some problems, but more information was needed about the response of soils to developmental uses. In the 1950s, soil scientists, engineers, and others worked to develop interpretations of soils for building sites, sewage-disposal systems, highways, pipelines, and recreation. By the late 1950s, some states and counties had begun to integrate soil-survey data and interpretations into their land-use planning functions. Fairfax County, Virginia, a rapidly urbanizing suburb of Washington, DC, is thought to be the first county in the USA to hire a full-time soil scientist.

A special symposium at the 1965 annual meeting of the Soil Science Society of America highlighted the rapid evolution of soil interpretations in the postwar period, especially those related to land-use planning. The resulting publication, *Soil Surveys and Land Use Planning*, represented widely ranging disciplines – soil science, civil engineering, architecture, and city and regional planning, among others. The symposium revealed the cost-effectiveness of integrating soil-survey interpretations into planning efforts. By passing the Soil Information Assistance for Community Planning and Resource Development Act of 1966, Congress clarified the legal basis for conducting soil surveys on nonfarm

Figure 2 Hugh Hammond Bennett, 1946. In 1946, over a decade after the establishment of the Soil Conservation Service (SCS), Hugh Hammond Bennett (left) visits a farm in Coon Valley, Wisconsin. Coon Valley was the first SCS field demonstration of the benefits of conservation systems to agricultural productivity and sustainability. Photograph by M.F. Schweers. Courtesy of the Natural Resources Conservation Service USDA.

areas in order to assist states and other geopolitical units in planning and resource development.

The decades of Kellogg's leadership were active for interpretations. The land capability classification system was refined and updated to help identify appropriate land uses and needed conservation practices for agricultural sustainability (Figure 3). The classification system, however, did not adequately address range or forestland capability, thus prompting the first forest soil-site correlations in the 1960s. The work on range and forestland quickly became firmly integrated into the soil survey. Soon biologists were developing criteria and ratings for soils related to food and cover for wildlife. The soil survey was well along in its transition from a purely agricultural productivity perspective to a much broader view.

Today's Interpretive Categories

Today, soil-survey information helps to answer a wide range of soil-related questions. The National Soil Survey Manual of the Natural Resources Conservation Service (NRCS) identifies 12 standard interpretive categories for soil-survey data:

National Inventory Groupings

echnical soil groupings present soil features or attributes of specific interest on a national scale. Such groupings have been developed as criteria for national-level programs, for example, prime farmlands, unique farmlands, hydric soils, and highly erodible lands.

Land-Use Planning

Interpretations provide information that allows evaluation of the potential environmental and economic effects of proposed or competing land uses. Interpretive maps at different scales and with different taxonomic levels are used as appropriate to the planning area. For example, regional planning may employ maps on entire associations of soil series and at higher taxonomic levels.

Farmland

Farmland interpretations traditionally place soils in management groups, identifying the soil properties important to crop production, and conservation needs, among other aspects of agriculture. Interpretations may be made to determine areas suitable for specialty crops, match crops with appropriate soils, delineate fields for greater soil homogeneity, and identify needs for specific management practices.

Productivity interpretations are one of the most important farmland interpretations and are described in terms of the output of product per unit land area (or as carrying capacity or liveweight gain for pastureland) under defined management. Interpretations may also present potential productivity ratings based on best practices or generalized soil productivity based on a number of different crops in the survey area.

Resiliency interpretations provide information on the ability of a soil to rebound from depletion or degradation. Resiliency ratings are important in evaluating alternative management systems and long-term effects of management systems on soils.

Rangeland

Soil-range site correlation identifies the suitability of the soil to produce various kinds, proportions, and amounts of plants, which is important in developing management alternatives to maintain site productivity. Range site descriptions provide information on the landscape, climate, soil, and vegetation factors, and the typical species (plant and animal) and their dynamics. The site interpretation describes the potential importance of the site for each of its uses and the feasibility of reclaiming degraded or depleted areas of the site.

Forest Land

Soil-forest site correlation describes the suitability of the soil to produce wood products and provides information on erosion and windthrow hazards, equipment limitations, seedling mortality, plant competition, and recommended trees for reforestation. Estimated productivity of the common trees may be given for each soil in the survey area if it is an important component of the overall survey. Surveys provide information on understory vegetation, potential problems associated with typical forestry operations, and effects of potential hazards such as burning, soil-borne pests, and diseases.

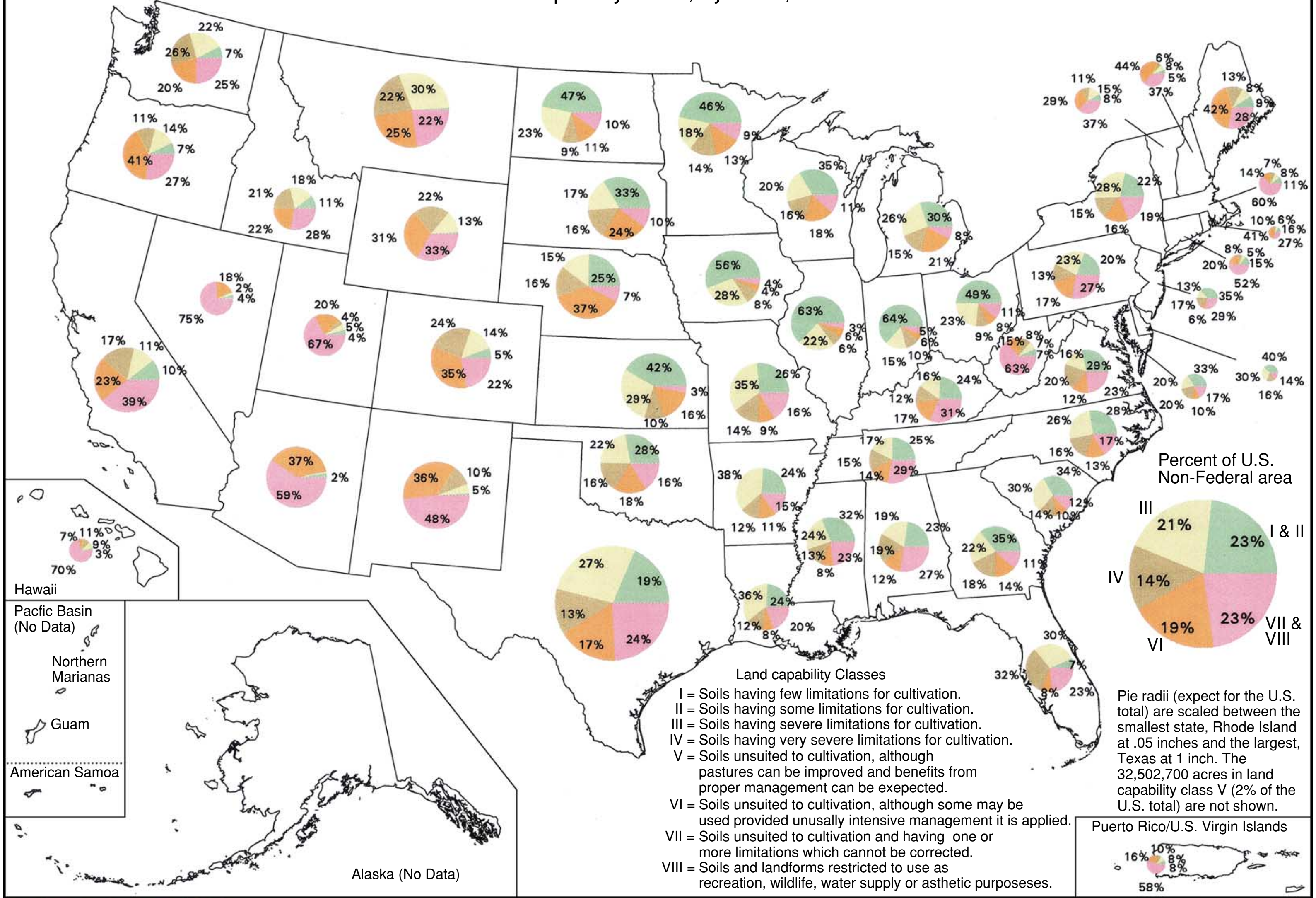
Windbreaks

Windbreaks of suitable species are used to protect soil resources, conserve moisture and energy, provide wildlife habitat, and protect homes, among other purposes. Correlation of soil properties and adaptable windbreak species helps in the selection of appropriate species to achieve the intended objective.

Recreation

Interpretations are made for a variety of recreational purposes such as golf courses, picnic sites, playgrounds, hiking paths, ski areas, snowmobile trails, and campsites. Restrictive soil-interpretative properties such as slope and texture of surface horizons tend

Land Capability Class, by State, 1997



Land capability Classes

I = Soils having few limitations for cultivation.
II = Soils having some limitations for cultivation.
III = Soils having severe limitations for cultivation.
IV = Soils having very severe limitations for cultivation.
V = Soils unsuited to cultivation, although pastures can be improved and benefits from proper management can be exepected.
VI = Soils unsuited to cultivation, although some may be used provided unusally intensive management it is applied.
VII = Soils unsuited to cultivation and having one or more limitations which cannot be corrected.
VIII = Soils and landforms restricted to use as recreation, wildlife, water supply or asthetic purposeses.

Pie radii (expect for the U.S. total) are scaled between the smallest state, Rhode Island at .05 inches and the largest, Texas at 1 inch. The 32,502,700 acres in land capability class V (2% of the U.S. total) are not shown.

U.S. Department of Agriculture
Natural Resources Conservation Service
Resource Assessment Division
Washington DC, December 2001

Map ID: m6175
For proper interprtation, see Explanation of Analysis for this map at our web site. Search for "USDASOTL" to locate our map index.

Data Source : 1997 Nation Reseources Inventory
Revised December 2000

to form the basis for ratings. Other factors such as location, accessibility, and infrastructure also tend to be important; however, they are not evaluated in map units for the survey.

Wildlife Habitat

Interpretations for wildlife habitat describe the suitability for different vegetation groups or habitat elements (e.g., hardwood trees, shallow-water areas), and a rating for types of wildlife supported (for example species adapted to open spaces, woodland, or wetlands). Current land use and existing vegetation and wildlife populations are not considered because they are subject to change; they nevertheless may express a significant influence on the potential for an area to support wildlife.

Construction Materials

Interpretations describe the suitability of the soil as a construction material and locations for obtaining materials such as gravel, sand, or low-shrink–swell potential soils. Survey information may also rate soils as potential sources of materials for other purposes, such as organic or mineral soil material that may be used as soil amendments or in horticultural uses. Soils can be rated as possible sources of these materials, but the quality of the site generally cannot be defined.

Building Sites

Interpretations describe the suitability of sites for small-building construction, as well as for road, street, and utility installation, among other construction purposes. Building-site interpretations provide information that can be used to compare suitability of alternative sites, but onsite evaluation is necessary for site selection.

Waste Disposal

Interpretations rate soils for their capacity to handle waste in a relatively small area, such as a septic-tank absorption field, or as distributed at low rates over a larger area. Typically, the evaluation begins with determining how disposal systems have performed on specific kinds of soil in the area, either from experience or related research. Soil scientists and specialists in other disciplines determine what properties are critical and how to appraise the effects of the properties.

Water Management

Interpretations for water management focus on the construction of small to medium impoundments, waterway control, installation of drainage and irrigation systems, and control of surface runoff. Detailed onsite evaluations are required to design engineered projects; however, interpretations assist in the evaluation of alternative sites.

Uses of Soils Data and Interpretations

While soils information is sometimes used alone, it is also used as one layer of information in integrated systems that consider other natural resources, demographics, climate, and ecological and environmental factors. Soil-survey data are used in models that deal with regional planning, erosion prediction, crop yields, and even global change. The following examines a few important or unique ways in which soil survey interpretations have been and are used today.

Federal and State Programs

While all USDA conservation programs and activities rely to some degree on soils information and interpretations, certain efforts and programs have more prominent links. The Farmland Protection Policy Act (FPPA; PL 97–98, 7 USC 4201), administered by USDA, was enacted by Congress in 1981 after finding that federally funded projects were playing a role in the loss of important agricultural lands. The FPPA mandates federal entities to conduct a land evaluation and site assessment to determine the potential for proposed projects to result in conversion of important agricultural lands to nonagricultural uses – soils information is the basis for identifying the presence of those important agricultural lands. Soils information is also a basic element in USDA farm program eligibility determinations. The Food Security Act of 1985 (1985 Farm Bill, PL 99–198, 7 USC 1631)

Figure 3 Land capability class, by state, 1997. The land capability classification system refined by Kellogg is still used today. This map allows a general comparison of land capability among states. Courtesy of the Natural Resources Conservation Service USDA.

included landmark provisions linking eligibility for USDA farm programs to compliance with the highly erodible land (HEL) and wetlands requirements in the legislation (Figure 4). Soils information forms the basis for determining the presence of HEL or wetlands. Soils information is also used to determine if lands are eligible for farmland protection programs. The presence of prime and important soils is a requirement for participation in USDA's Farm and Ranch Lands Protection Program, which works in conjunction with state, local, or nongovernmental organization programs to purchase conservation easements to protect important agricultural lands.

State and Local Planning

Interpretations have formed the foundation for land-use planning ordinances across the USA (e.g., floodplain setbacks; slope protection; agricultural zoning). One of the earliest uses of the soil-survey information for planning was in Fairfax, VA in the 1950s. So useful were the survey and interpretations that the county retained a soil scientist to provide consultation to planners. Within 4 years the county had a soil scientist on permanent staff to assess: (1) septic tank disposal systems; (2) rezoning cases; (3) floodplain extent; (4) determining best trees/shrubs for planting; (5) type of and depth to bedrock; (6) slope stabilization and soil slippage problems; and (7) new school sites. Today, soil survey data are used to develop suitability models that evaluate soils for their capacity for specific uses (for example, as water recharge or environmentally sensitive areas), and thus support local land-use zoning decisions.

Natural Resource Management

Early interpretations focused on the capabilities of land for agricultural production including productivity ratings and soil suitability for specific crops. One of the earliest agronomic interpretations was identifying 'tobacco soils' in the southeast for expansion of production to other suitable areas. Over time, interpretations have become much broader, covering a wide array of environmental effects.

With the emergence of precision farming, soils information is increasingly being used in geographic information systems (GIS) to develop variable-rate application plans tailored to specific field conditions. While the resolution of data needed in precision agriculture may exceed that generally provided by soil surveys, the development of improved dissemination and expert systems will likely benefit some kinds of precision agriculture.

Soil-landscape information provided by soils interpretations is an important tool for managing forestland, rangeland, and wildlife. Range and forest site correlations are being enhanced to distinguish sites by their ability to produce a characteristic natural plant and animal community. This ecological site data for forestland and rangeland will be available through the NRCS Ecological Site Information System (ESIS) and will provide data on the ecological site, plant and animal composition, history, and condition of the community over time. The National Park Service uses soil-survey information to support vegetation management in the park system. These soil–ecological site interpretations may also be used to identify forest ecosystems with high potential for containing specialty forest products such as truffles or ginseng.

Interpretations can have a large role in the conservation of wildlife populations. Soils maps help botanists identify areas with high probabilities of having rare or endemic species – often species that are endangered or threatened as they are restricted to unusual soils (e.g., soils produced from substrates like gypsum, marine clays, or from the interactions of substrates and dynamic geomorphic processes). Some such searches have led to the discovery of additional populations. This method has proved successful for locating new populations of green pitcher plant (*Sarracenia oreophilia*), Mohr's Barbara button (*Marshallia mohri*), and geocarpon (*Geocarpon minimum*).

Land Appraisal and Assessment

Soil interpretations historically have had significant influence in property sales. The most vociferous opposition to early soil surveys arose in response to interpretations that identified limitations on land potentials, primarily of an agricultural nature. While early uses of soils information included land valuation for tax assessment, the same information has been used to reverse assessments. In one case, a community won a reduced tax assessment as they convincingly demonstrated for the court that land values were lower than assessed because of the soils' natural unsuitability to support building foundations.

Soils data and interpretations can be used to evaluate potential risks and insurance needs. Insurance underwriters commonly use soil-survey information and interpretations that identify high flood potentials, shrink–swell soils unsuitable for construction, and other important characteristics upon which development and insurance decisions depend. On the St Regis Mohawk reservation in New York, soil composition was used to identify homes that should be tested for radon levels. Twenty-five buildings on soil with a high correlation to radon emissions were targeted for testing; of these, four were identified as exceeding the standard.

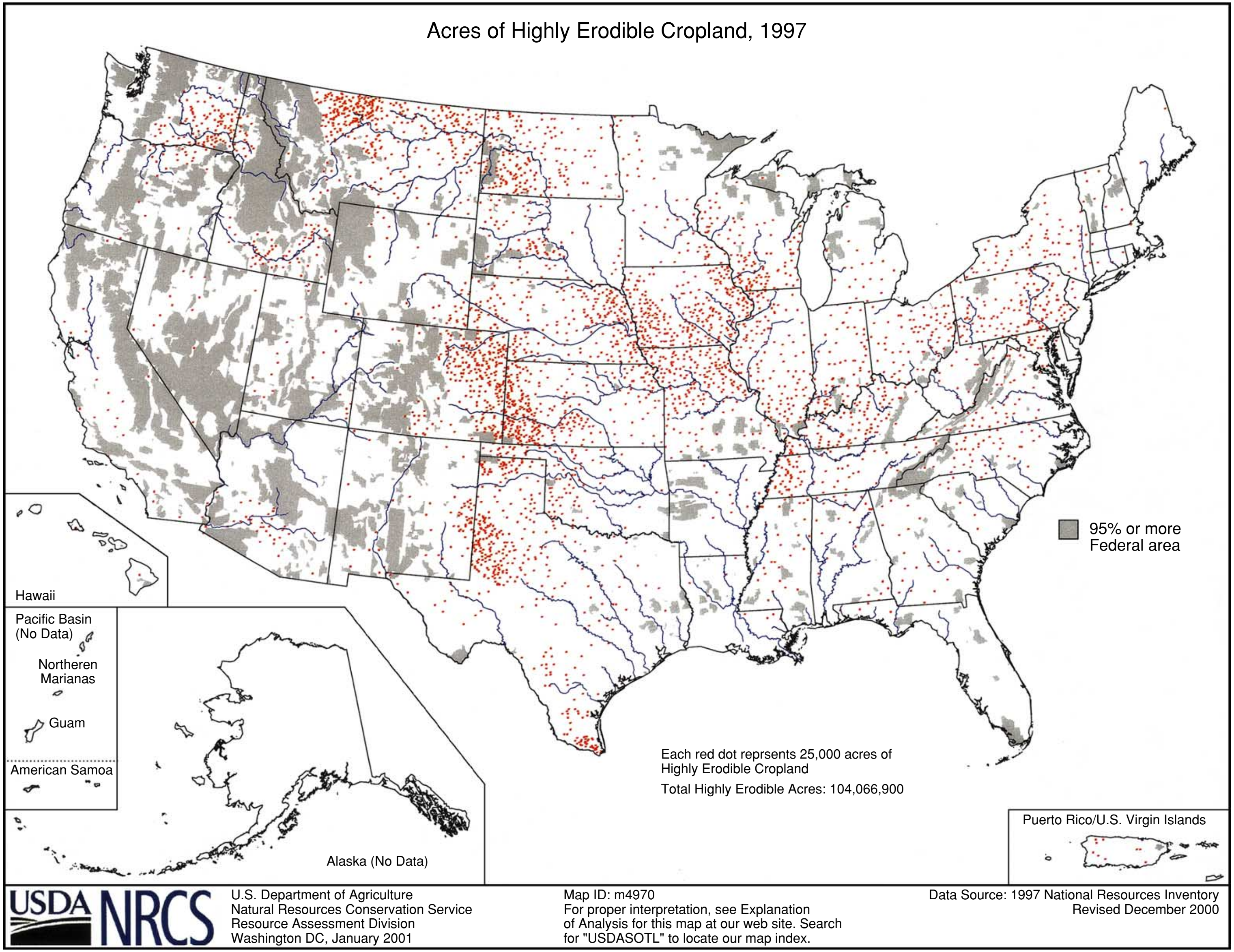

Acres of Highly Erodible Cropland, 1997
Hawaii
Pacific Basin
(No Data)
Northeren
Marianas
Guam
American Samoa
Alaska (No Data)
▨ 95% or more
Federal area
Each red dot reprsents 25,000 acres of
Highly Erodible Cropland
Total Highly Erodible Acres: 104,066,900
Puerto Rico/U.S. Virgin Islands
USDA NRCS
U.S. Department of Agriculture
Natural Resources Conservation Service
Resource Assessment Division
Washington DC, January 2001
Map ID: m4970
For proper interpretation, see Explanation
of Analysis for this map at our web site. Search
for "USDASOTL" to locate our map index.
Data Source: 1997 National Resources Inventory
Revised December 2000

Figure 4 Acres of highly erodible cropland in the USA, 1997. A focus on the conservation needs of highly erodible land led to the origins of the Soil Conservation Service (SCS) in 1935 and a half-century later it was also a focal point of the 1985 Food Security Act. This map shows the total number and distribution of highly erodible acres as defined by the 1985 Act. This presentation is useful for national-scale evaluation of areas potentially needing conservation treatment, however, a site-specific evaluation is needed to establish the erodibility index for a specific land area. Courtesy of the Natural Resources Conservation Service USDA.

Engineering and Construction

Soils maps and interpretations are used widely in siting development such as highways, buildings, and other construction. Selecting sites with soils with sufficient load-bearing quality, not subject to flooding, and suitable for on-site sewage disposal result in substantial costs-savings, and reduce the potential for unforeseen failures. Soils data used by officials in one Illinois county helped decision-makers select a highway site with the fewest limitations, which resulted in savings of several thousand dollars per acre because of reduced excavation and construction costs.

Interpretations can also be used in estimating costs. The cost of laying buried pipelines, for example, is affected by a variety of soil factors. Soils interpretations can inform decisions on special material needs (e.g., pipe composition and rigidity), mitigation required (e.g., extremely wet or differentially draining soils creating needs for protective coatings or cathodic protection), and costs of excavation (e.g., short depth to bedrock or presence of hardpan can increase excavation costs).

Hazardous Waste, Brownfields, and Remediation

Evaluation for toxic substances, such as heavy metals, is an emerging area for soil surveys, particularly in urban or developed settings. The survey for LaTourette Park on Staten Island, New York, includes an evaluation of heavy metals in soils in addition to the more traditional interpretations for playgrounds and picnic areas. A soil survey completed before the Plattsburgh, New York, Air Force Base closure in 1995 is now a resource for redevelopment. The survey and interpretations are providing the information needed to undertake environmental-impact statements, planning, and remediation for reclamation of contaminated sites (e.g., dumps, fuel spills, etc.).

Research and Analyses

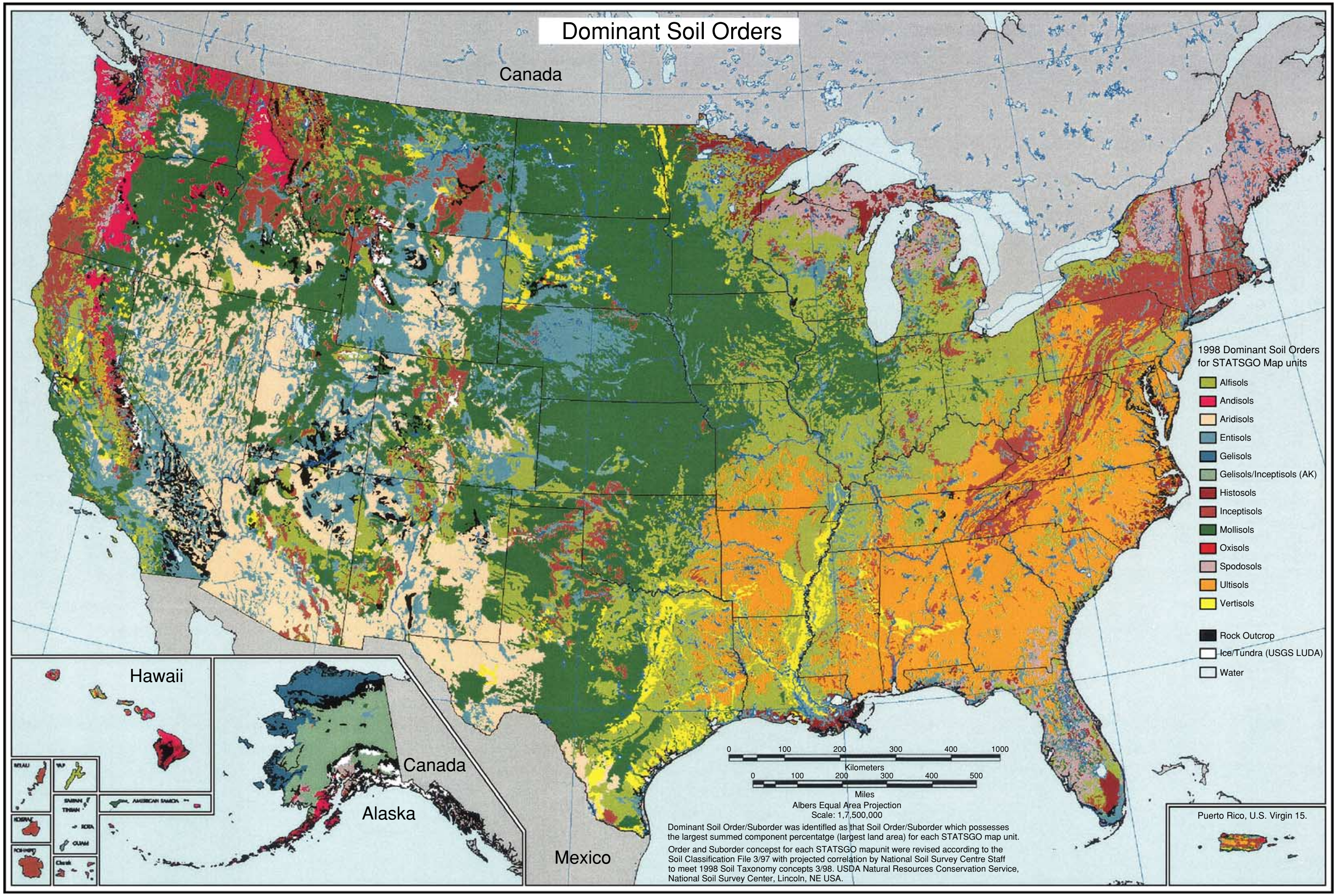
Soil survey data and interpretations are a substantial resource for research efforts. Interpretations have been vital to the development of improved management practices that optimize productivity. Interpretations are also used to identify soils with desired characteristics for specific experiments (e.g., soils of varying acidity) and to evaluate site suitability and limitations for agricultural research. This use of interpretations was first seen in 1911 when the state of Alabama requested a detailed survey and interpretation of its agricultural experiment stations. During World War II, interpretations helped to identify areas of high potential for producing goods in short supply. Interpretations identified areas suitable for the production of guayule as a substitute for rubber; oil-producing alternatives such as castor bean; and fiber crops such as American hemp.

Interpretations have proven useful in historic and archeological research. Soil maps and interpretations provide a resource to identify features associated with early settlements (e.g., access to rivers or tributaries, productive soils) as well as physiographic changes over time. Comparison of older surveys with current aerial photos can identify shifts that help to target potential sites. Some uses have included verifying property boundaries, which were commonly associated with stream channels. Early survey maps also included farmstead boundaries, rural cemeteries, and other settlement features (e.g., churches, towns, city buildings, etc.) – now of historic interest.

The array of analytical uses of soils data continues to expand. Detailed soil survey data are available through the Soil Survey Geographic (SSURGO) database to support GIS uses (mapping scales from 1:12 000 to 1:63 360). Field-mapping methods using national standards are used to construct SSURGO soil maps and SSURGO digitizing duplicates the original soil-survey maps. This level of mapping is designed for use by landowners, townships, and county natural-resource planning and management. State Soil Geographic (STATSGO) database maps present smaller-scale generalized soil-survey data designed for broad planning and management uses covering state, regional, and multistate areas (Figure 5) (mapping scales of 1:250 000, with the exception of Alaska, which is 1:1 000 000).

Soils data are an essential component of natural-resource modeling. For example, a soils data layer is a fundamental source of information for modelers simulating nonpoint-source pollution potentials from agricultural areas. Some productivity models, such as the Environmental Policy Integrated Cimate model (EPIC), integrate soil, climatic, economic, management, and other variables to simulate impacts of cropping systems on the environment and productivity. EPIC uses up to 11 variables to describe the soil and up to 20 variables to describe physical and chemical characteristics of each identifiable soil layer in the profile.

Today, soil survey data are being used to evaluate potential for carbon sequestration and other soil conditions that relate to global change. The role that soils play in mediating the effects of agriculture and forestry on the global atmospheric composition of greenhouse gases is a major component of the USDA Global Change research and development program.



Albers Equal Area Projection
Scale: 1,7,500,000

Dominant Soil Order/Suborder was identifled as that Soil Order/Suborder which possesses the largest summed component percentatge (largest land area) for each STATSGO map unit.

Order and Suborder concepst for each STATSGO mapunit were revised according to the Soil Classification File 3/97 with projected correlation by National Soil Survey Centre Staff to meet 1998 Soil Taxonomy concepts 3/98. USDA Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE USA.

Figure 5 Dominant soil orders in the USA. This polygon map shows the dominant soil order present in each State Soil Geographic (STATSGO) database map unit. Dominant soil orders represent the largest land area for each map unit. This type of geographical information systems analysis allows users to compare the general distribution of soil orders. STATSGO-level mapping is designed to be used for broad planning and management uses covering regional, state, and multistate areas. Courtesy of the Natural Resources Conservation Service USDA.

Conclusion

Applications of soils data and survey interpretations have steadily broadened in range and precision in response to economic, environmental, social, and political influences. Today, interpretations are becoming increasingly dynamic with the application of information technology to support decision-making. As soils data are integrated into analytical and expert systems, access to digitized soils data is essential. Advances in digital orthophotography (digital imagery that has been rectified to remove distortions resulting from topography and the camera angle, thus equalizing distances represented on the image), maps and digital data for computer manipulation and retrieval enhance the delivery efficiency of soil-survey information. Detailed resource information on specific land areas can now be provided quickly and interactively to help landowners, communities, and others in land-use decision-making.

Further Reading

Bartelli LJ, Klingebiel AA, Baird JV, and Heddleson MR (eds) (1966) *Soil Surveys and Land Use Planning*. Madison, Wisconsin: Soil Science Society of America and American Society of Agronomy.

Gardner DR (1998) *The National Cooperative Soil Survey of the United States*. Historical Notes Number 7. Washington, DC: Natural Resources Conservation Service.

Helms D, Effland A, and Durana P (eds) (2001) *Profiles in the History of the U.S. Soil Survey*. Ames, IA: Iowa State Press.

Jenny H (1961) *E. W. Hilgard and The Birth Of Modern Soil Science*. Pisa, Italy: Collana Della Rivista 'Agrochimica.'

Kellogg CE (1943) *The Soils That Support Us*. New York: Macmillan.

Klingebiel AA (1991) Development of soil survey interpretations. *Soil Survey Horizons* 32(3): 53–65.

Olson GW (1981) *Soils and the Environment, A Guide to Soil Surveys and Their Applications*. New York: Chapman & Hall.

Simonson RW (1987) Historical aspects of soil survey and soil classification. *Soil Survey Horizons*. Madison, Wisconsin: Soil Science Society of America.

Smith GD (1983) Historical development of soil taxonomy – background. In: Wilding LP, Smeck NE, and Hall GF (eds) *Pedogenesis and Soil Taxonomy*, pp. 23–49. New York: Elsevier.

US Department of Agriculture, Natural Resources Conservation Service, Soil Survey Staff (1999) *Soil Taxonomy, A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 2nd edn. Agriculture Handbook No. 436. Washington, DC: US Department of Agriculture.

US Department of Agriculture, Natural Resources Conservation Service, Soil Survey Division Staff (1993) *Soil Survey Manual*. USDA Handbook No. 18. Washington, DC: US Department of Agriculture.

ARCHAEA

J E T McLain, USDA Agricultural Research Service, Tucson, AZ, USA

Published by Elsevier Ltd.

Introduction

In 1977, Carl Woese and his colleagues announced the discovery of 'a new form of life.' The immediate reaction of biologists was largely skeptical. Prior to this, the existing paradigm was that all organisms, except viruses, could be assigned to one of two primary groups, prokaryotes and eukaryotes. The rRNA work of Woese and others confirmed the existence of the archaeal domain, and currently recognized biotic diversity now consists of three groups, two of which are exclusively microbial (Archaea and Bacteria), while the third (Eukarya) contains both microbial life (as unicellular protists) and multicellular organisms.

Although they are metabolically diverse, a property common to the majority of archaeal organisms identified to date is the ability to exist in extreme habitats, including environments of high salt, high temperature, low pH, and acute anoxia. Three general ecological categories represent the overall patterns of archaeal adaptations to extreme environments: thermophilic (heat-loving), methanogenic (methane-producing), and halophilic (salt-loving). Most Archaea belong to at least one of these categories, and a number belong to two. This does not mean, however, that Archaea are limited to extreme environments. Archaea also thrive in freshwater sediments, temperate soils, and other less extreme conditions, confirming that microbes of this domain are ubiquitous.

The harsh environments in which many Archaea flourish have intrigued scientists interested in strategies of coping with life at the extremes. The recent sequencing of the entire genomes of several Archaea have provided a wealth of knowledge, including the fact that some archaeal genes, including those encoding major metabolic pathway enzymes, are similar to those of Bacteria, while others, such as those for RNA polymerase subunits, are more similar to eukaryal genes, while still others appear to be Archaea-specific.

Archaeal Evolution

Phylogenetic relationships derived on the basis of 16S and 23S rRNA indicate that the domain Archaea consists of three major kingdoms, the Crenarchaeota, Euryarchaeota, and Korarchaeota (Figure 1). One of the two main branches of this phylogenetic tree contains, in large part, the thermophilic sulfur-dependent Archaea, while the other branch contains the methanogens, the extreme halophiles, and a few thermophilic organisms.

Current phylogenetic evidence deduced from comparison of 16S rRNA sequences suggests that Archaea have evolved more slowly than either Bacteria or the Eukaryotes. This is especially true of hyperthermophilic Archaea. It is not known why Archaea are the slowest-evolving of the three domains, but it may be related to their habitation of extreme environments. For example, organisms living in hyperthermal environments must maintain those genes that specify phenotypic characteristics critical to life at high temperatures, as evolutionary alteration of gene expression could impair organism survival.

Because thermophilic Archaea may have evolved very slowly, it has been proposed that these organisms are likely to have been among the earliest life forms on Earth. The phenotypic properties of thermophiles, including the ability to withstand high temperatures and the use of anaerobic chemoorganotrophic (the use of organic chemicals as electron donors) or chemolithotrophic (the use of organic chemicals as electron donors) metabolism, agree well with the phenotype of primitive organisms one would predict, given the geochemical conditions present on the Earth three billion years ago or more. If indeed life first occurred in the form of thermophiles in boiling hot springs deep on the ocean floor, this may explain the mystery surrounding the importance of phosphate in information storage and energy transfer in living cells. In other sea and freshwater environments, phosphate is present in very low concentrations and is often the limiting nutrient for organisms that live there, but water exiting hydrothermal systems often percolates through phosphate-rich minerals, assuring a rich phosphate supply to organisms inhabiting this environment. If thermophiles were present as an early life form, then life on Earth may have evolved in a phosphate-rich environment, thus developing a necessity for this mineral for cellular function.

The Archaeal Cell

Archaeal cells display a wide variety of morphological types. Some, similar to species of the bacterial domain, are strictly rod-shaped or spherical. Others

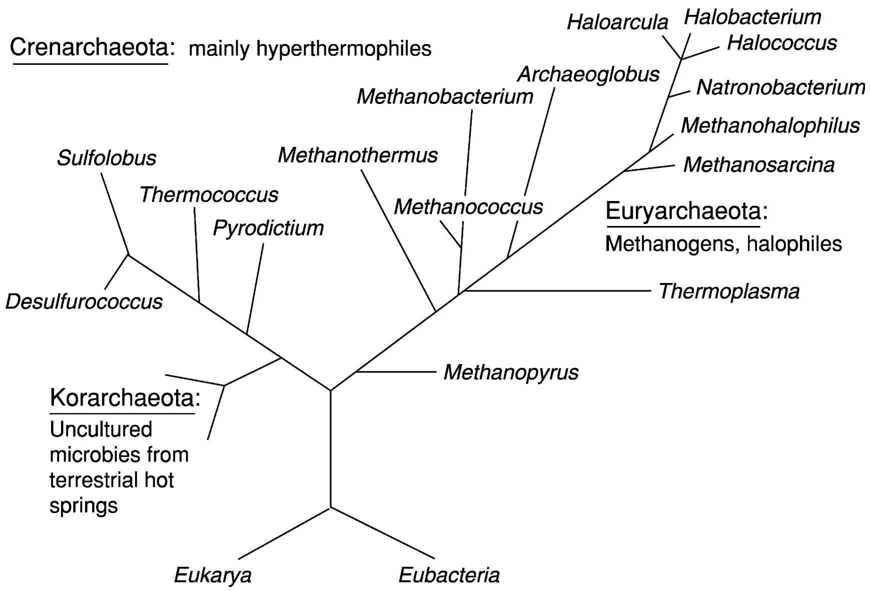


Figure 1 The phylogenetic tree of the Archaea. (Reproduced with permission from Howland JL (2000)) *The Surprising Archaea: Discovering Another Domain of Life*. New York: Oxford University Press.

Table 1 Some cellular features of the Archaea

Feature of cell	*Most similar to domain*
Morphology	Some unique
Cell wall with pseudopeptidoglycan	Resembles bacteria
Genome consisting of single circular piece of DNA	Bacteria
DNA polymerase, DNA helicase, DNA ligase	Bacteria
Protein trafficking for sugars and inorganic ions	Bacteria
Multisubunit RNA polymerase	Eukarya
Ether-linked lipids	Unique
Promoter site for transcription	Eukarya
Lack of nuclear membrane	Eukarya
Flagellar composition and assembly	Unique
Chaperonin heat-shock proteins	Bacteria

are disks, spirals, or filaments, or exhibit amebal irregularity, with variable protuberances. Still others have a mineral-like geometry, with shapes similar to cubes or triangles. Other irregular cells are bumpy spheres (e.g., *Thermococcus*) or are vaguely rod-shaped, but of highly variable diameter (e.g., *Pyrodictium*).

Archaeal cells possess many characteristics similar to those of the Bacteria and Eukarya, and others that are unique to the archaeal domain (Table 1). With one exception, all Archaea contain a cell wall that, like the cell wall of bacteria, functions to prevent osmotic lysis and to define cell shape. The ability of Archaea to adapt to extreme environments is assisted by the possession of unique cell wall types, which vary from those containing molecules composed of pseudopeptidoglycan, closely resembling bacterial peptidoglycan, to cell walls completely lacking a polysaccharide component. A common wall type is a paracrystalline surface layer (S-layer), consisting of protein or glycoprotein, generally of hexagonal symmetry. Although S-layers are common to all groups of Archaea, the biochemical makeup of S-layers among species is very diverse and, in some cases, this layer is too supple to contribute to the stability of the cell. In these cases, the function of the S-layer is unknown, but it has been proposed that the space between the cytoplasmic membrane and the outer surface of the S-layer may fulfill the role of a periplasmic space reminiscent of Gram-negative bacteria.

Archaeal RNA polymerases are complex, consisting of up to 14 subunits (compared to 4 in the bacterium *Escherichia coli*). Like Bacteria, Archaea possess a single RNA polymerase, but the archaeal RNA polymerase resembles those of eukaryotes in multisubunit complexity and sequence homology. In addition, archaeal RNA polymerases are unable to initiate transcription *in vitro*, a feature also seen in eukaryotes where general transcription factors are required for initiation.

Three Archaeal Groups

Extreme Halophiles

Extremely halophilic Archaea inhabit environments with salt concentrations high enough to kill most organisms. They occur in hypersaline bodies of water such as the Dead Sea or the Great Salt Lake, saline soils, and have also been isolated from dry deposits where salt is mined commercially. Halobacteria can also be found on salted fish, salted hides, bacon, and sausage, and these microorganisms can often be attributed to the spoilage of these foods.

The formation of saline waters throughout the world can result from the geologic separation of seawater from the open ocean and concentration of salts by evaporation, or may result from the dissolution of salts from rocks into bodies of water. The ionic composition of hypersaline water bodies varies widely, from high Na^+ ($\sim 100\,g\,l^{-1}$) and Cl^- ($\sim 200\,g\,l^{-1}$) in the Great Salt Lake, to high Cl^- ($>200\,g\,l^{-1}$) in the Dead Sea, to high SO_4^{2-} ($>200\,g\,l^{-1}$) in Hot Lake in Washington state, USA. The varied chemical and physical properties of saline habitats result in colonization by a widely diverse group of archaeal prokaryotes, which possess some common features that allow them to thrive in these high-solute environments.

Research has shown that extreme halophiles not only tolerate salty conditions, but they also require high concentrations of salt. A generally accepted definition of an extreme halophile is that the organism requires at least $1.5\,mol\,l^{-1}$ NaCl for growth, but most species require $2–4\,mol\,l^{-1}$ for optimal growth. Virtually all extreme halophiles can grow at the limit of saturation for NaCl ($5.5\,mol\,l^{-1}$), although some species grow only very slowly at this salinity. The plasma membrane of halophiles does not exclude salts, but can select for certain ions so that internal concentrations of ions can be controlled and cellular functioning can be maintained. Cells of *Halobacterium*, for example, pump large amounts of potassium from the environment into the cell such that the concentration of K^+ inside the cell is higher than the Na^+ concentration outside. Thus, the total ionic strength remains the same on both sides of the plasma membrane but potassium, required for many cellular functions, is the prevailing cation on the inside. In this manner, *Halobacterium* employs an inorganic ion as its compatible solute and remains in positive water balance

The cell wall of *Halobacterium* is stabilized by sodium ions and in low-Na^+ environments the cell wall breaks down, resulting in cell lysis. The halobacterial cell wall is composed of a glycoprotein with an exceptionally high content of the acidic amino acids aspartate and glutamate, and the sodium ions shield the negative charges contributed by the carboxyl groups of these amino acids. When sodium is diluted, the negatively charged parts of the proteins actively repel each other, leading to cell lysis. Thus, the cellular components of halophiles exposed to the external environment require high sodium for stability, whereas internal components require high potassium. This requirement for specific cations in such high amounts is a feature unique to halophilic bacteria. Another unique quality of Halobacteria are cytoplasmic proteins with very low levels of hydrophobic amino acids, perhaps representing an evolutionary adaptation to the highly ionic cytoplasm of extreme halophiles. In environments of high ionic strength, polar proteins would tend to remain in solution, whereas nonpolar molecules would tend to cluster and perhaps lose activity.

Some extreme halophiles possess a unique light-mediated synthesis of adenosine triphosphate (ATP) that does not involve chlorophyll pigments. When light is available and their ability to obtain energy through respiration is compromised, *Halobacterium salinarum* and certain other extreme halophiles synthesize and insert a protein, bacteriorhodopsin, into their membranes. Conjugated to this is a molecule of retinal, a carotenoid-like molecule that can absorb light and catalyze the transfer of protons across the cytoplasmic membrane. Light-mediated ATP production in *H. salinarum* has been shown to support slow growth of this organism anaerobically in the absence of organic energy sources and under conditions in which other energy-generating reactions do not occur.

Methanogens

Two features common to all methanogenic Archaea are the inability to tolerate oxygen or reactive oxygen species and the ability to produce methane gas. Methane production in soils characteristically occurs under anaerobic, highly reducing conditions in the absence of nitrate, sulfate, or ferric iron, including the mud of swamps and marshes, the beds of fresh and marine bodies of water, and mud originating from sewage plants and rice paddies. In some instances, methanogens live in small anoxic pockets in soils in an otherwise oxygen-rich area. Such regions are often formed by the action of microorganisms that locally consume all of the available oxygen, and several studies have confirmed that methanogens can be present throughout a macroscopically oxic soil and that methane-producing activity can respond rapidly to the establishment of appropriate anoxic conditions.

Methanogens also flourish in the digestive systems of ruminants. Production of methane gas worldwide from cattle, goats, sheep, and camels is greater than methane production from paddy fields and swamps. The rumen utilizes a mixed population of microorganisms, including Eubacteria, methanogenic Archaea, and anaerobic protozoa to carry out the digestion of cellulose and other polymeric sugars. Methanogens are responsible for regulating the overall fermentation in the rumen by removing hydrogen gas during methane production. This action keeps the hydrogen concentration in the rumen low, encouraging the activity of hydrogen-producing species and altering their metabolism towards higher-yielding pathways. These pathways result in the synthesis of more microbial cells, increasing the available protein to the ruminant.

Methane is the final product of a complex community of organisms breaking organic materials down into the simple methanogenic substrates and thus, methanogens complete the last step in the anaerobic decomposition of organic matter. Because they are dependent on other organisms for provision of substrate and the establishment of reducing conditions, methanogens can only function as members of microbial communities. Methanogenesis is frequently rate-limited by the activities of the other members of the community, and particularly by how rapidly hydrogen or acetate is made available. At least 10 substrates can be converted to methane by methanogens (Table 2) and three main classes of metabolic reactions can be used to create energy for ATP synthesis. The first class utilizes CO_2-type substrates (Table 2, lines 1–5), while the second class of reaction involves reduction of the methyl group of methyl-containing compounds to methane (Table 2, lines

Table 2 Reactions for methanogenesis by methanogenic bacteria

Substrates		*Products*
$4H_2 + CO_2$	→	$CH_4 + 2H_2O$
$4H_2 + HCO_3^- + H^+$	→	$CH_4 + 3H_2O$
4 Formate + $4H^+$	→	$CH_4 + 3CO_2 + 2H_2O$
4 (2-Proponal) + CO_2	→	CH_4 + 4 acetone + $2H_2O$
2 Ethanol + CO_2	→	CH_4 + 2 acetate + $2H^+$
4 Methanol	→	$3CH_4 + CO_2 + 2H_2O$
4 Methanol	→	$3CH_4 + HCO_3^- + H^+ + H_2O$
4 Methylamine + $2H_2O$	→	$3CH_4 + CO_2 + 4NH_4^+$
Methanol + H_2	→	$CH_4 + H_2O$
Acetate + H^+	→	$CH_4 + CO_2$
Acetate + H_2O	→	$CH_4 + HCO_3^-$

6–9). The third class of methanogenic reactions is acetotrophic, the cleavage of acetate to CH_4 plus CO_2 (Table 2, lines 10–11). The conversion of acetate to methane appears to be a very significant ecological process, especially in sewage digesters and in freshwater anoxic environments where competition for acetate between sulfate-reducing bacteria and methanogenic bacteria is not extensive. The reduction of CO_2 to CH_4 is generally H_2-dependent, but formate, carbon monoxide, and even elemental iron can serve as electron donors for methanogenesis. A few methanogens can utilize simple organic compounds as electron supplies for CO_2 reduction.

Methanogenesis employs a unique collection of cofactors in its reactions. Methanogens often contain high concentrations of these compounds, whose presence can be used to determine the prevalence of methanogenic Archaea. One cofactor commonly used to identify methanogens is coenzyme 5-deazaflavin F420, an electron carrier involved in several reactions in the methanogenic pathway which also has a role in DNA photorepair. Another cofactor, coenzyme F430, contains nickel, which all methanogens require for growth. A third cofactor unique to methanogenesis, coenzyme M, acts as a methyl-carrying coenzyme in the last step of the methanogenic pathway and is thus involved in the final reduction of a methyl group to form methane.

Hyperthermophiles

The archaeal kingdom Crenarchaeota consists of a great diversity of organisms. Some form a stable part of the soil microbial community in boreal environments, while others are defined by their extremely thermophilic nature. This latter group, the hyperthermophiles, contains organisms that are the most heat-loving of all known prokaryotes. Several hyperthermophiles are capable of growth at temperatures above the normal boiling point of water, and all have temperature optima above 80°C.

Many hyperthermophilic Archaea have been isolated from geothermally heated soils or waters containing elemental sulfur and sulfides, and most hyperthermophilic species metabolize sulfur in some way. In fact, the first hyperthermophile discovered, *Sulfolobus*, grows in sulfur-rich hot acid springs. Sulfur-rich environments are found throughout the world, and extensive studies of hyperthermophilic Archaea have been made in Yellowstone National Park (USA), where the highest concentration of sulfur-rich thermal features in the world has been attracting researchers since the first scientific study of the region in the late nineteenth century.

Hyperthermophilic Archaea have also been found in artificial thermal habitats such as the boiling outflow of geothermal power plants. In addition, a phylogenetically distinct set of hyperthermophilic Archaea has been isolated from submarine volcanic habitats, where the pressure of even a few meters of seawater can raise the boiling point of water sufficiently to select for organisms capable of growth above 100°C. *Pyrolobus fumarii*, a submarine organism, has a growth optimum at 106°C but can grow at 113°C, and can even survive autoclaving at 121°C for 1 h.

Depending on the surrounding geology, geothermally heated environments may be either slightly alkaline to mildly acidic (pH 5–8) or extremely acidic, with pH values below 1 not uncommon. Such extreme acidity does not deter the hyperthermophiles: *Picrophilus oshimae* has a pH optimum for growth of 0.7 and can grow significantly at pH values approaching zero; however, the majority of thermophilic Archaea inhabit neutral or mildly acidic habitats.

With a few exceptions, hyperthermophiles are obligate anaerobes. Their sulfur requirement is based on the need for an electron acceptor to carry out anaerobic respiration or an electron donor for chemolithotrophic metabolism. Organisms of the genera *Thermococcus* and *Thermoproteus* oxidize a variety of organic compounds (e.g., small peptides, glucose, starch) anaerobically in the presence of S^0 as an electron acceptor. *Sulfolobus* is an obligate aerobe capable of oxidizing organic compounds, H_2S, or S^0 to H_2SO_4 and fixing CO_2 as a carbon source. Many hyperthermophilic Archaea can grow chemolithotrophically with H_2 as an energy source. *Acidianus*, a facultative aerobe resembling *Sulfolobus*, grows anaerobically using S^0 as an electron acceptor and H_2 as an electron donor, forming H_2S as the reduced product, and *Pyrodictium* can be cultured under strict anaerobic conditions in a mineral-salts medium supplemented with H_2 and S^0 at temperatures up to 110°C.

A unique property shared by many hyperthermophiles such as *Sulfolobus* and *Acidianus* is an unusually low guanine-cytosine (GC) base ratio. The DNA of *Sulfolobus* is ~38% GC, whereas that of *Acidianus* is ~31%. These low GC base ratios are intriguing: in a testtube, DNA of 30–40% GC content would melt almost instantly at 90°C. Research into how DNA of these organisms is prevented from melting is ongoing; however, it is hypothesized that the DNA may be protected, in part, by high cytoplasmic solute concentrations (the melting temperature of DNA increases as the solute concentration increases). Cytoplasmic concentrations of solutes such as cyclic 2,3-diphosphoglycerate in cells of *Methanopyrus fervidus* and other thermophilic methanogens are closely correlated with the temperature at which the organisms are grown in laboratory incubations. In addition, it is thought that specific DNA binding proteins of hyperthermophiles somehow prevent DNA from melting, perhaps by folding the DNA into a conformation consistent with thermal stability. *Pyrodictium* cells grown at 110°C produce 80% of their protein biomass as a single protein that functions as a molecular chaperonin, stabilizing other cellular proteins by refolding them as they begin to denature near the upper temperature limits of growth. At 100°C (near the optimum for growth of *Pyrodictium*), very little of this chaperonin protein is made, suggesting that only at very extreme temperatures do the otherwise thermally stable proteins of this organism begin to denature. Additional protein adaptations of hyperthermophiles that allow these organisms to withstand extreme temperatures include sequence modifications, addition of salt bridges, increased hydrophobic interactions, additional ion pairing and hydrogen bonding, improved core packing, and shortening of loops. These strategies, used to differing extents by different thermophilic proteins, not only confer higher thermal stability but also enhance rigidity and resistance to chemical denaturation.

Within the last several years, molecular techniques have uncovered a unique lineage of the kingdom Crenarchaeota that is phylogenetically distinct from the hyperthermophiles. Nonthermophilic Crenarchaeota, which have been identified in marine picoplankton, freshwater sediments, soils, and in continental shelf anoxic sediments, have been shown to account for as much as 2% of microbial rRNA in soils analyzed. Phylogenetic analyses suggest that the nonthermophilic Crenarchaeota may have a common ancestor with the hyperthermophiles, but as yet the ecological significance of the nonthermophiles remains unknown.

Psychrophiles

Archaea have also been detected in ecosystems with characteristics in direct contrast to hyperthermophilic environments. Psychrophilic (cold-loving) Archaea account for over a third of the prokaryotic biomass in coastal Antarctic surface waters, and the hypersaline lakes of the Vestfold Hills lake system in Eastern Antarctica have been the subjects of a number of studies on microbial distribution. One of these lakes is Deep Lake, with a salinity of $320\,\mathrm{g\,l^{-1}}$ and temperatures between −14 and −18°C. The biodiversity of Deep Lake is low, and is dominated by Archaea of the family Halobacteriaceae. To date, only three psychrophilic archaeal strains, all free-living and members of the subdomain Eukaryarchaeota, are available in pure culture. *Methanococcoides burtonii* (T_{min} −2.5°C) and *Methanogenium frigidum* (T_{min} −10°C) were originally isolated from the bottom of Ace Lake, Antarctica, where the methane-saturated waters remain between 1 and 2°C.

Thermoplasma

Thermoplasma acidophilum is a prokaryote that does not possess a cell wall and in this respect resembles the mycoplasma. Phylogenetically, however, *Thermoplasma* is a member of the Archaea. *Thermoplasma* is an acidophilic, aerobic, thermophilic chemoorganotroph, and with one exception, all strains of *Thermoplasma* have been obtained from self-heating coal refuse piles, which contain coal fragments, pyrite, and other organic materials extracted from coal. When this refuse is dumped into piles in coal-mining operations, it tends to self-heat by spontaneous combustion and creates conditions conducive to growth of *Thermoplasma*, which apparently metabolizes leached organic compounds. A chemically unique cell membrane allows *Thermoplasma* to survive the osmotic stresses of life without a cell wall and to withstand the dual environmental extremes of low pH and high temperature. This membrane contains a unique lipopolysaccharide that, together with other molecules, renders the *Thermoplasma* stable to hot acidic conditions.

Biotechnological Use of Extremoenzymes

Biotechnologically useful enzymes represent the main focus of industrial interest in the Archaea, as a result of the abilities of these microbes to function at the temperature, salinity, and pH limits of life. Heat-tolerant enzymes are currently the most investigated of all extremoenzymes because performing biotechnologically related processes at higher temperatures is often advantageous for many reasons. In chemical reactions involving organic solvents, decreased viscosity and increased diffusion at elevated temperatures result in higher reaction rates. In addition, performing reactions at higher temperatures reduces the possibility of complications resulting from contamination. One thermophilic compound of particular interest is DNA polymerase, an enzyme that is responsible for the elongation of the primer strand of a growing DNA molecule and is thus central to the polymerase chain reaction for DNA amplification. DNA polymerases from various hyperthermoarchaea (including *Pwo* from *Pyrococcus* woesei and *Pfu* from *P. furiosus*) are showing biotechnological promise, based on their stringent proofreading abilities and suitability for the amplification of longer DNA fragments. These hyperthermophilic DNA polymerases possess error rates that are five- to 10-fold lower than that of the widely used thermobacterial *Taq* polymerase from *Thermus aquaticus*.

Applied uses exist or have been proposed for a variety of other archaeally derived materials. The extremely stable lipids of archaeal membranes may represent a novel drug delivery system because of their enhanced stability under temperature extremes. Archaeal components such as the S-layer glycoprotein have drawn interest for their use as possible vaccine carriers and other nanotechnological potentials, and it has been shown that much higher immune responses in mice are shown to protein antigens encapsulated in archaeosomes than in conventional liposomes. A thermostable ligase for the ligase chain reaction (an amplification method that involves the ligation of two sets of adjacent oligonucleotides) would be of obvious benefit because the ligation must be carried out near the melting temperature of the DNA, and the ligase enzymes must be stable during the dissociation step that follows. Currently, a ligase from *T. aquaticus* is used, but a more stable equivalent may be available from hyperthermophilic Archaea. Haloarchaeal polymers have been considered as a raw material for biodegradable plastics. Hydrolases from hyperthermophiles could be used in the food-processing industry to hydrolyze fats at high temperatures, reducing bacterial contamination problems. Addition of polymer-hydrolyzing extremoenzymes such as beta-glycanases from psychrophiles to detergents would allow for efficient washing in cold water. The food industry could exploit pectinases that act at lower temperatures in the processing of fruit juices or cheeses.

Often, the mere presence of archaeal communities carries considerable potential economic value. Methanogenic Archaea have proven to be quite valuable in their capacity as clean and inexpensive energy sources, and acidophilic Archaea have been identified at several acid mine drainage sites where their mineral-sulfide oxidizing abilities play an important role in the geochemical sulfur cycle.

The days of Archaea being considered as just ‘odd bacteria’ adapted to living in extreme environments appear to be over. In the past few years, information on the isolation, characterization, description, and applications of Archaea has mushroomed. Researchers will continue to study the adaptations that allow Archaea to grow at the extremes and to search for new species that will extend the boundaries that limit life of Earth. There is no doubt that many novel features remain to be discovered about these microorganisms, and their continued study will have a major impact on science in the decades to come.

See also: **Bacteria:** Soil

Further Reading

Bernander R (1998) Archaea and the cell cycle. *Molecular Microbiology* 29: 955–961.

Eichler J (2001) Biotechnological uses of archaeal extremozymes. *Biotechnology Advances* 19: 261–278.

Howland JL (2000) *The Surprising Archaea: Discovering Another Domain of Life*. New York: Oxford University Press.

Jarrell KF, Bayley DP, Correia JD, and Thomas NA (1999) Recent excitement about the Archaea (Archaebacteria). *BioScience* 49: 530–542.

Kates M, Kushner DJ, and Matheson AT (eds) (1993) *The Biochemistry of Archaea (Archaebacteria)*. New Comprehensive Biochemistry Series, vol. 26. Amsterdam, Netherlands: Elsevier Science.

Madigan MM, Martinko J, and Parker J (2000) *Brock Biology of Microorganisms*. Upper Saddle River, NJ: Prentice-Hall.

Oren A (2002) Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiology Ecology* 39: 1–7.

Reysenbach A-L, Voytek M, and Mancinelli R (eds) (2001) *Thermophiles: Biodiversity, Ecology, and Evolution*. New York: Kluwer Academic Plenum.

Vreeland RH and Hochstein LI (eds) (1993) *The Biology of Halophilic Bacteria*. Boca Raton, FL: CRC Press.

Woese CR and Wolfe RS (eds) (1985) *The Bacteria: A Treatise on Structure and Function,* vol. VIII. *Archaebacteria*. Orlando, FL: Academic Press.

ARCHEOLOGY IN RELATION TO SOILS

J A Homburg, Statistical Research Inc., Tucson, AZ, USA

Introduction

Archeology, soil science, and geology are intimately related disciplines because they are all historical sciences. Records of each of these scientific disciplines take time to develop, and indicators of each are encoded in the sedimentary and soil matrix in which the archeological record is embedded. Physical and chemical soil properties are commonly used to make interpretations of the archeological record, including how the soil was formed, how it has been altered, and how it is preserved. Soils are also used to draw archeological inferences about human behavior in the past. Examples include identifying chemical signatures of different human activities that may leave no physical traces, assessing the long-term effects of cultivation on soil productivity, and reconstructing how earthen material was used to make pottery and build many types of cultural features (e.g., adobe houses, earth ovens, clay-lined fire hearths, roasting pits, postmolds, etc.). An understanding of soil variability within a site is important in deciding where archeological test pits should be placed and how deep they should be excavated.

As an interdisciplinary science, archeology is unrivaled in terms of the number of fields on which it draws, except perhaps by forensic science. Today, archeological projects often include one or more soil scientists who work with experts in geomorphology, geochemistry, ethnobotany, pollen, phytoliths, paleontology, paleoclimatology, dating methods (e.g., radiocarbon, tree-ring, archeomagnetic, optical luminescence, and obsidian hydration dating), and many other disciplines. Soil science has major applications in archeological research. Archeology can also contribute significantly to pedology, especially with the use of cross-dating techniques, whereby temporally diagnostic artifacts such as pottery and projectile points from archeological deposits are used to determine the age of soil horizon development.

Pedostratigraphy and the Archeological Record

A basic understanding of soil science, especially pedology (soil morphology and pedogenesis) and soil chemistry, is essential for making meaningful interpretations of archeological context and site-formation processes that account for the contemporary archeological record. This record, as expressed in surficial, buried, and stratified cultural deposits, is an imperfect and biased record of past human activities. Many postoccupational disturbance processes (e.g., animal burrowing, displacement by root growth, and infilling of soil cracks caused by freeze–thaw and wetting–drying cycles) modify the archeological record by translocating and mixing artifacts. The degree of mixing can be assessed by observing and documenting soil profiles, focusing on the extent of krotovinas (that is, infilled animal burrows) and the level of soil horizonation. Preservation of materials in the archeological record is also highly variable. Durable materials such as stone and ceramic artifacts are usually preserved for very long periods, but perishable items made of wood and leather are rarely preserved except in dry caves and anaerobic wetland soils.

Calcareous materials such as bone and shell may or may not be preserved in the archeological record, depending on the climate, soil pH, and how long they were exposed to weathering before being buried. Bone, which consists mainly of the calcium phosphate mineral hydroxyapatite, is best preserved at pH 7.88 and it becomes increasingly soluble above and below this level. It is well known that bone is poorly preserved in acidic soils, but it is also poorly preserved in moderately to very strongly alkaline soils. It is interesting to note that human bone at the Port Hudson Confederate Cemetery in Louisiana was almost completely decomposed in acidic soils in approximately 120 years, with a soil pH of approximately 5.5. Calcareous remains such as shellfish in coastal shell middens have been added in such large quantities in some cases that soil pH has altered sufficiently to preserve bone.

Site-formation processes, including soil formation, if properly interpreted, are useful for helping to explain the archeological record. Archeological context refers to the relationship between cultural and natural deposits, as well as to the relationship between artifacts and cultural features within archeological sites. An understanding of soils and how they develop is crucial for explaining why archeological deposits are located or concentrated in particular places and for assessing the stratigraphic integrity of cultural deposits. This is because the interaction of many soil-forming factors, including climate, topography, vegetation, and parent material, were also important considerations to humans in deciding where

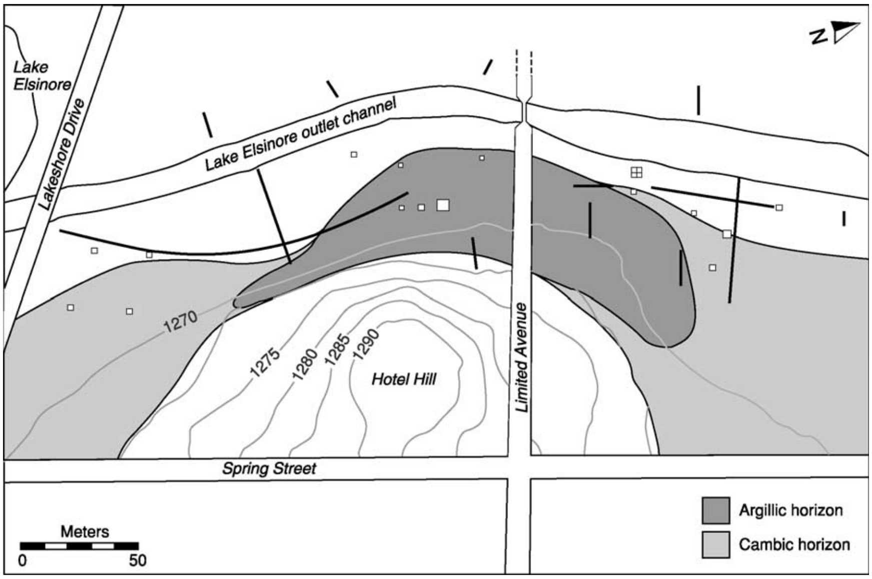


Figure 1 Map of the distribution of cambic and argillic horizons at the Elsinore Site. (Adapted from Grenda DR (1997) *Continuity and Change: 8500 Years of Lacustrine Adaptation on the Shores of Lake Elsinore*, Fig. 26. Technical Series No. 59. Tucson, AZ: Statistical Research/The University of Arizona Press, with permission.)

habitations and other site types were established. Soil maps are commonly used by archeologists in developing models to predict where sites of particular ages and functions are located, and how land-use patterns changed through time.

Soil development is helpful in assessing geomorphic stability, or the extent to which landforms have been preserved, buried, or modified. Soils showing the most development are usually the oldest in a given study area, and these soils are places where older archeological deposits may be preserved. Artifacts are commonly concentrated in well-developed A horizons, because these are surface horizons that served as occupation surfaces where artifacts and refuse were discarded and incorporated into the soil. Older archeological deposits commonly overlie or are found within diagnostic subsurface horizons such as argillic and calcic horizons. An example of this relationship has been evidenced from backhoe trenches and testpits at the Elsinore site, an 8500-year-old site located near the outlet channel of Lake Elsinore in southern California. Excavations have revealed that the richest and oldest cultural deposits are in and above sediments where an argillic horizon has developed. The youngest cultural remains, by contrast, are on the fringe of the site, where a weakly developed soil horizon, a cambic horizon, has formed (Figure 1).

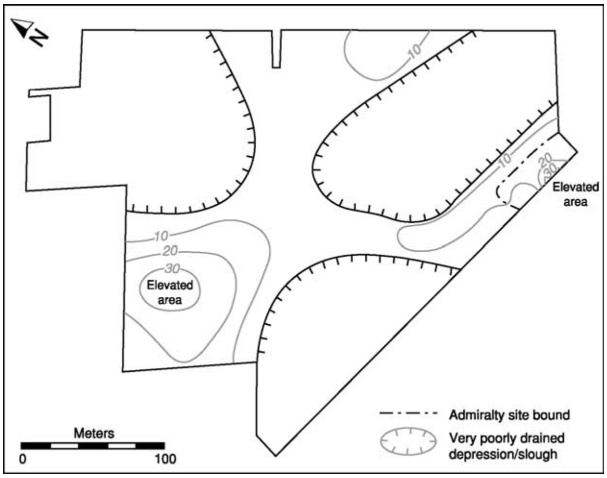


Figure 2 Map of the Channel Gateway project area in Los Angeles, showing E horizon thickness, depression, and the Admiralty Way site. (Adapted from Altschul JH, Homburg JA, and Ciolek-Torrello RS (1992) *Life in the Ballona: Archaeological Investigations at the Admiralty Site and the Channel Gateway Site*, Fig. 71. Technical Series No. 33. Tucson, AZ: Statistical Research/The University of Arizona Press, with permission.)

Soil properties are useful for identifying places where past human activities have been concentrated and explaining why this is so. At the Admiralty site, a coastal shell midden in the Ballona Wetlands of Los Angeles, an E horizon has been identified in subsurface tests. Soils in this wetland setting, although seasonally moist or saturated, are dry enough most of the year for organic matter, clay, and bases to be translocated downward to form an eluvial horizon where these materials have been depleted. Thickness of a grayish eluvial horizon, the E horizon, marks the best-drained areas on and adjacent to the site (Figure 2). On- and off-site subsurface tests show that the site location coincides with an area with the thickest E horizons. This use of soil morphology has helped to explain why the site is located where it is in the wetlands.

Soil morphology has also been used to study the Louisiana State University Campus Mounds, a 5000-year-old earthen-mound site built by Native Americans during a time that archeologists refer to as the Middle Archaic Period. Three soil cores were collected from each of the two 5-m-tall mounds, to determine their age and construction history. Three radiocarbon samples of humates in the A horizon buried under one of the mounds and in the lower mound fill have yielded dates of approximately 5000 years BP, which makes this mound site one of the earliest in North America. Although similar in appearance from the outside, the soil stratigraphy varies greatly between the two mounds, which suggests different construction histories (Figure 3). Many thin, dark-colored bands of A horizon material have been documented in Mound A. If these A horizons have formed *in situ*, it implies different construction stages, with enough time between stages for natural A horizons to form as organic matter was added to the soil from decomposing vegetation. The boundaries between the dark-brown bands are abrupt to very abrupt in Mound A, which contrasts sharply with the clear-to-gradual horizon boundaries in the off-mound natural soils. Consequently, the bands have been interpreted as basket-loaded A horizon material that has been placed there during mound construction, not *in situ* A horizons marking construction stages. Brief interruptions during the construction process have been identified, however, one of the indicators being a reddish oxidized zone that apparently marks a fire hearth, which may have served a ritual purpose during mound construction, and a laminated zone, resulting from slope-wash deposition during a storm. The fill of Mound B was quite different, consisting mainly of dark yellowish-brown subsurface soil derived from the argillic and fragic soil (a Btx horizon) from off-mound areas adjacent to the mounds. Thin grayish bands are also found in Mound B, and these are interpreted as a mix of material from the E horizons that overlie the Btx horizon off-mound and from infilled vertical cracks (or interprisms) of the fragipan. Soil chemistry supports the interpretation that Mound A was built mainly from near-surface soil and Mound B from subsurface material. The Mound A fill has a significantly higher pH, organic matter content, and phosphorus and calcium concentrations. Interestingly, a fragipan, a loamy soil horizon with a high bulk density that can restrict root and water penetration, has been found on the lower slopes of both mounds. The fragipan has formed *in situ*, thus indicating that this diagnostic subsurface horizon can develop within 5000 years. This is a good example of how archeology can help ascertain how long certain soil features take to develop.

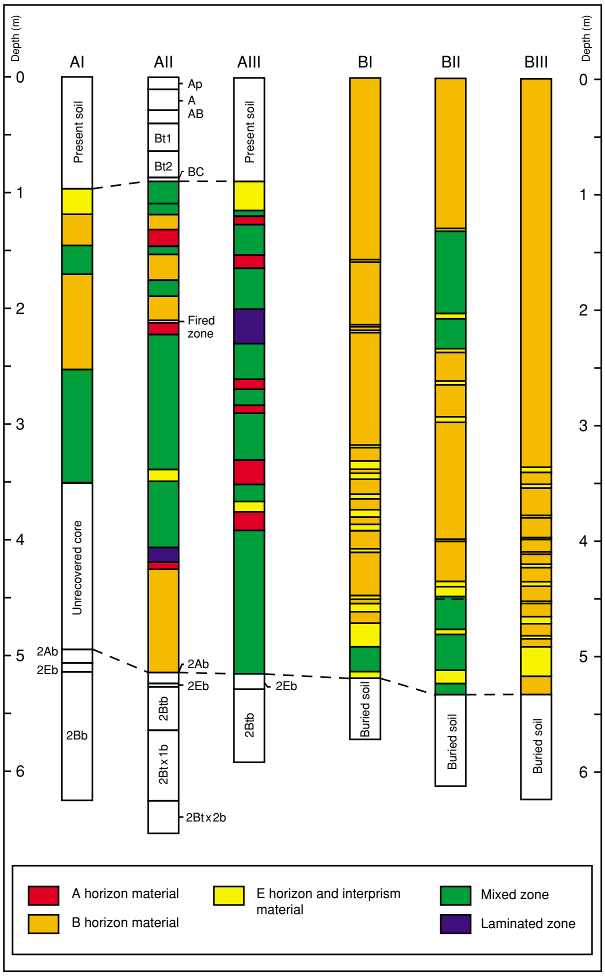


Figure 3 Stratigraphy of cores at the Campus Mounds of Louisiana State University. (Adapted from Homburg JA (1988) Archaeological investigations at the LSU Campus Mounds. *Louisiana Archaeology* 15: 31–204, Fig. 18, with permission.)

Archeological Soil Chemistry

Soil chemistry is being utilized increasingly in archeological interpretations. It is sometimes used for paleoenvironmental reconstruction, such as the use of carbon isotopes to reconstruct vegetation, such as prairie versus forest vegetation, in places where the boundary has shifted. More commonly, however, soil chemistry is used to reconstruct past human activities, identify activity areas within sites, and define site boundaries and stratigraphic relationships. Detecting traces of past human activities depends largely on the degree and kind of impact that humans had on soils. Two types of surface horizons (or epipedons) that are clearly the result of human activities are anthropic horizons, which are characterized by elevated levels of phosphorus that result from fertilizer additions, and plaggens, which result from many years of manure additions. Other signatures of human activity are much more subtle and subject to interpretive difficulties. This is because humans impact soils in such a wide variety of ways and with differing intensities, and at all scales of time and space.

Elements commonly enriched in soil due to human activity include carbon, nitrogen, phosphorus, and calcium, and, to a lesser degree, potassium, magnesium, sulfur, copper, and zinc. The most common addition to soil in preindustrial societies that is often easy to recognize today is organic matter from plant and animal residues composed mainly of carbon, nitrogen, phosphorus, sulfur, and humus. This type of human effect typically results in more organic matter in the surface soil and a thickened A horizon that has a lower chroma value than unmodified soils; for example, the dark, fertile soils of the Amazon Basin, known as Terra Preta. Differentiating organic matter that is anthropogenic from natural soils often relies on other traces of human activity such as artifacts, remains of plant foods (e.g., pollen, phytoliths, charcoal, and ash), and chemicals such as phosphates and fatty acids. Traces of these materials can be identified in thin sections using soil micromorphology.

Because of its high stability and immobility in many soils, phosphorus has served as a key element for identifying areas of human activity. Phosphorus is concentrated in surface horizons owing to many kinds of human activity, especially disposal of rubbish and waste products from humans and livestock. Similarly, aluminum, calcium, manganese, and iron levels may be elevated in anthropogenic soils, because they are often bound in phosphate compounds. Calcium and calcium carbonate levels have been used to identify places where animal bone and limestone were processed. Potassium is enriched where animal remains decompose and where wood ash is added to the soil as a result of burning vegetation. Nitrogen, which is abundant in animal and human waste, may also become elevated in soils, but, because it rapidly decomposes, traces of this element rarely persist for long periods in the archeological record. Various heavy metals have also been useful for archeological studies, especially in reconstructing activities associated with metal-working and processing in cultures that used iron, copper, and other metals.

Since the 1990s, researchers have used a wide range of analytical techniques to characterize the elemental chemistry of anthropogenic soils relative to natural, unaltered soil. Two of the most sophisticated methods include inductively coupled plasma–atomic emission spectroscopy (ICP/AES) and inductively coupled plasma–mass spectroscopy (ICP/MS). Soil samples are typically collected using grid- or transect-sampling methods to cover both archeological sites and off-site areas. In some cases interpretations of ancient activity areas have been improved by studying activity areas in modern or recently abandoned pre-industrial societies. This research is a type of ethnoarcheology, which involves studying existing cultures and how they use space for different activities to bolster interpretations of the archeological past. Examples of this research include studies of modern farming villages in the Mayan area and of pastoral societies in Africa such as the Masai. It is important to note that research on human impacts on soils in archeological sites is relevant to interpretations of impacts on soils today. The archeological record provides a long-term perspective on human impacts that is available from no other source.

Soil as a Resource

Soil has always been a basic resource for sustaining human populations. This section focuses on the agricultural use of soils, including archeological studies of the long-term effects of cultivation on soil productivity. Soils served as the planting medium for ancient agricultural systems, and also provided material for building earthen architectural structures and for making pottery.

Studies of ancient agricultural soils can contribute significantly to research on agricultural sustainability and soil quality in the context of modern and ancient farming systems. Ancient agricultural soils in arid and semiarid regions are particularly useful for these kinds of studies because: (1) soil-formation processes (e.g., weathering, leaching, and illuviation) proceeds slowly, so soil changes caused by ancient cultivation practices tend to persist and be detectable for 1000 years or longer; (2) many ancient fields have not been cultivated since they were abandoned, so historic farming practices such as plowing and artificial-fertilizer applications have not masked or erased soil properties that reflect ancient farming practices; and (3) the presence or absence of agricultural features for planting crops, such as rock alignments, rock piles, and terraces, provide important clues for discerning, sampling, and comparing cultivated and uncultivated soils.

Since the 1960s, several studies have been conducted in the American Southwest to assess the long-term, anthropogenic effects of cultivation. This work has been conducted in agricultural fields that are among the oldest identifiable fields in the New World. Ancient agricultural fields in the Sonoran desert are commonly associated with soil horizons (e.g., argillic and petrocalcic horizons and duripans) or shallow bedrock that strongly impedes or blocks water infiltration, thereby holding moisture in the rooting zone for long periods after rain. Soil studies in Arizona and New Mexico indicate that the consequences of cultivation in terms of soil productivity and agricultural sustainability are highly variable due to many interacting factors (e.g., climate, topography, soil type, types of crops and native vegetation, agricultural technology, and the duration and intensity of cultivation). Cultivation effects are not easily predicted, so soil testing is essential for making informed evaluations of soil productivity and anthropogenic effects. Some investigations have found that ancient farming practices had long-lasting degradational effects on soils, whereas others have found that cultivation had little effect on soil quality or even that it was improved.

Degradational changes associated with organic matter depletion, compaction, erosion, and enrichment with salts are typically less severe for land where minimal or no-tillage farming has been practiced (as in the prehistoric American Southwest) than for intensively tilled or irrigated fields, but production levels can be reduced substantially nevertheless. Cultivation, especially of highly consumptive crops such as maize, a crop that was cultivated extensively in the New World, can rapidly deplete already low stores of N. Studies of Puebloan fields, at Mesa Verde and near Flagstaff and Santa Fe, and terraced Mimbres fields in southwest New Mexico have found that phosophorus reductions were so severe that agricultural soils became unproductive and were abandoned. By contrast, studies of rock mulch field systems in Arizona have found that soil productivity was not seriously degraded by cultivation. Rock mulch systems are ubiquitous in cobbly landscapes cultivated by the Hohokam, Anasazi, and other prehistoric cultures of the American Southwest. Similar rock mulch systems have also been documented in Israel, Italy, Peru, Argentina, New Zealand, China, the Canary Islands, and other places that have a soil moisture deficit during the growing season.

The Mimbres study has found that the primary anthropogenic soil changes were degradational, and that the effects persisted for more than 800 years after fields were abandoned. Rock alignments were built to function as dams to slow runoff, increase infiltration into agricultural soils, and thicken naturally thin A horizons in the terrace soils. Compared with uncultivated soils, cultivated terrace soils had lighter colors, thicker A horizons with blockier structures and higher bulk density, lower levels of organic carbon, nitrogen, and total and available phosphorus, and higher pH and manganese levels. Accompanying these soil changes were dramatic differences in vegetation patterns, with terrace soils having little to no grass cover compared with adjacent control. Chapalote maize grown in the terrace soils under controlled greenhouse conditions are significantly lower in weight than those grown in control soils. Different fertilizer treatments indicate that nitrogen is the most limiting nutrient for plant growth.

Studies of Classic Period (*ca.* AD 1150–1450) rock mulch field systems in the Gila River valley of southeast Arizona and in the Lower Verde River valley of central Arizona have found that agricultural soils were not degraded and that soil fertility was often improved. The lack of deleterious changes may reflect either direct cultivation effects or postcultivation vegetation associations with agricultural rock features, short-term use of fields, replenishment of nutrients by naturally deposited organic debris, cultivation of crops with low nutrient requirements, or recovery after abandonment. Use of gravel and cobble mulches on planting surfaces reduces soil erosion by wind and water, increases soil temperature to extend the growing season, increases water infiltration, and reduces evaporative water loss from the soil. Rock-pile soils in the Lower Verde have gravimetric water contents that average approximately 30% higher than adjacent controls (Figure 4). The Gila and Verde River studies have found that cultivated soils tend to have similar to significantly elevated levels of organic carbon (Figure 5), nitrogen, and total and available phophorus levels compared with uncultivated soils. Cultivated Gila soils had pH levels that were reduced from approximately 8.1–8.4 to 7.7–8.0, levels that would have been beneficial for crop production due to increased plant availability of many essential nutrients. Bulk density tests indicate that cultivation did not cause long-term soil compaction, which is a common degradational effect in many agricultural systems.

Determining what crops were or might have been cultivated in ancient fields can be very difficult, and

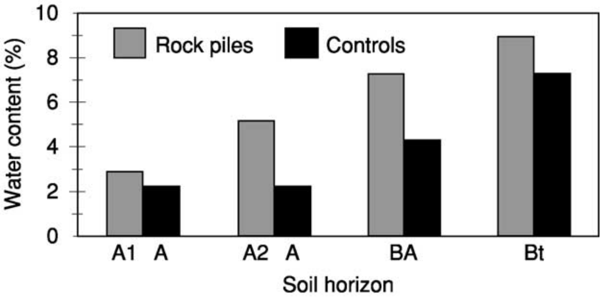


Figure 4 Gravimetric water content (grams of water per gram of soil, expressed as a percentage) of rock mulch soils in the Horseshoe Basin compared with adjacent controls. Reproduced from Homburg JA and Sandor JA (1997) Tucson, AZ: SRI Press.

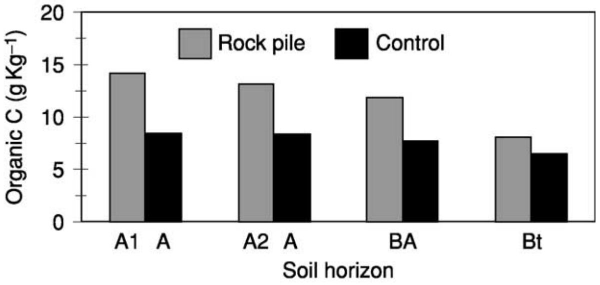


Figure 5 Organic C content of rock mulch soils in the Horseshoe Basin compared with adjacent soils. Reproduced from Homburg JA and Sandor JA (1997) Tucson, AZ: SRI Press.

soil data alone are insufficient to solve this problem. Where possible, evaluations of cultivation effects on soils need to make reference to the crops that were cultivated. Traces of plant remains (e.g., pollen, phytoliths, and carbonized plant remains) from soils in fields or in habitation features such as houses, agricultural processing areas, storage facilities, and waste deposits are used to reconstruct the range of crops that were cultivated. Soils can supply important clues about potential crops, however, especially when evaluated in the context of modern soil and native plant associations relative to prehistoric agricultural features. For example, agave thrives in cobbly, calcareous, droughty soils, even soils with low fertility in rugged terrain that support little other vegetation. Overall, soil nutrient levels in the Gila gridded-field complex are sufficient to have supported maize agriculture, but the thin soils, high temperatures, low rainfall, and low runoff throughout most landscape positions of the field suggest that crops such as agave or other drought-tolerant plants were the probable focus of agricultural production.

Studies of traditional Native American agriculture may supplement archeological studies of prehistoric agricultural systems in the Southwest. The rationale of these kinds of studies is that knowledge about prehistoric agriculture is inherently based on inference and reconstruction from sites abandoned for centuries. Historic and modern Native American agriculture may provide valuable sources of baseline information on identification of agricultural fields, placement of fields with respect to geomorphic and soil variables, crop and soil management, agricultural productivity, and the effects of agriculture on the environment. The relatively subtle nature of prehistoric agriculture, combined with subsequent overprinting by natural processes and cultural effects, underscores the need for such baseline studies. Although there are limitations in extrapolating from the past to the present, there are definite threads of continuity in agricultural strategies between prehistoric cultures and Native American groups that still practice similar forms of traditional agriculture.

One example is a study of the runoff agricultural system of the Zuni, a Puebloan tribe in west-central New Mexico. The Zuni and other Native American groups of the American Southwest have successfully cultivated maize and other crops for more than two millennia without using formal irrigation or artificial fertilizers. Zuni fields are among the oldest more or less continuously cultivated areas in the USA. Traditional Zuni agriculture is based on a runoff farming system that involves capturing storm water flow and organic-rich sediment from small watersheds and directing it on to agricultural fields. The Zuni and other Native American groups of the semiarid Southwest US have successfully cultivated maize and other crops for over three millennia without using artificial fertilizers. The Zuni soil study concludes that tillage in recent decades has altered some soil properties but there is no indication that agricultural soils have been degraded. Paired cultivated soils are 7.6% higher in bulk density on average and they have greater massive structure and reduced granularity. Compaction at this level, especially given the friability of the soils, suggests that root elongation is not restricted and there may even be advantages in moisture retention for this desert environment. No consistent differences in organic carbon, nitrogen, and available and total phosphorus have been identified in the cultivated, abandoned, and uncultivated soils, thus supporting the perception of Zuni farmers that long-term cultivation has not caused a decline in agricultural productivity.

List of Technical Nomenclature

Anthropogenic Relating to, or influenced by, human activity. Anthropogenic changes in soils can be degradational (e.g., accelerated erosion, organic matter and nutrient losses, salinization, and pollution) or improvements (e.g., elevated organic matter and nutrient status, and reduced erosion, salt, and pollution levels)

Argillic horizon A subsurface diagnostic horizon marked by significant accumulation of clay translocated from above. In general, a clay increase of at least 1.2 times that of an overlying horizon, and that is either 15 cm or more thick or at least one-tenth the thickness of all overlying horizons, is needed to qualify as an argillic horizon

Cambic horizon A subsurface diagnostic horizon marked by a pedogenic change from the soil parent material, such as reddening by oxidation, development of soil structure, or accumulation of illuvial clay at levels that do not meet the definition of an argillic horizon

Geomorphic stability The degree or extent that a landform or landscape has been preserved or altered. Highly stable land surfaces have undergone little change for long periods of time, so associated soils are expected to be at or near equilibrium with major soil-forming factors, such as climate and vegetation cover. By contrast, unstable geomorphic surfaces are modified by erosion or buried by sediment, so the surface may change too fast for soil development to reach equilibrium with the environmental conditions

Illuviation A soil process whereby material (e.g., clay, iron oxides, calcium carbonate, or organic matter) accumulates in a soil due to translocation from overlying horizons. These materials are leached from above and translocated downward by wetting fronts in the soil

Midden A German term widely used in archeology that refers to anthropogenic accumulations of trash deposited either intentionally or unintentionally to form layers, mounds, or infilled pits. Midden soils are enriched with cultural materials, such as ash and carbonized plant debris, microfossil traces (pollen and phytoliths from cultivated and wild plant foods), butchered animal bones, shellfish, manure, and stone, ceramic, and metal artifacts are commonly found in middens, as well as human burials in some cultures. Midden soils are typically dark in color and have elevated organic matter and phosphorus levels relative to natural soils unaffected by human activities

Paleosol An ancient soil. Types of paleosols include: (1) buried soils – ones covered by sediment; (2) exhumed soils – ones formerly buried but now exposed by erosion; and (3) relict soils – ones formed under the influence of preexisting landscape or climatic regime that was never buried. This term is widely used in archeology and Quaternary geosciences, but there is no consensus as to a minimum age requirement or specific differences between soil formed in preexisting versus present landscapes and climates

Pedoarcheology A study that involves the application of soil science, especially studies in soil morphology and genesis, to answer archeological research questions, problems, or hypotheses

Pedostratigraphy Layering in soil material caused by soil formation processes. Soil horizons form through time in physically and chemically weathered sedimentary or residual geologic material, with differences in soil layers reflecting the integrated effects of climate, organisms, topography, and parent material. Pedostratigraphic units may or may not correlate to natural depositional layers, and they may result in layers that are well-preserved or mixed by burrowing soil fauna. Other kinds of stratigraphic layers include lithostratigraphic, chronostratigraphic, biostratigraphic, and ethno- or archeostratigraphic units

Soil micromorphology A study involving analysis of undisturbed soil at a microscopic scale. In archeological investigations, other kinds of earthen materials (e.g., unconsolidated sediments and cultural materials, like ceramics, adobe, bricks, and mortars) can be examined using micromorphology. Thin sections are 25–30-μm slices of soil or other earthen materials that are examined through a microscope using plane and/or crossed polarized light, or by other imaging techniques, such as ultramicroscopy (e.g., scanning or transmission electron microscopy). Thin sections are obtained by: (1) collecting and drying structurally intact samples; (2) impregnating them under vacuum with a solvent mixed with an unsaturated polyester or epoxy resin, with or without a stain or fluorescent dye; and (3) grinding, polishing, and mounting the thin section on glass slide, with or without a cover slip, depending on the kind of analysis

See also: **Applications of Soils Data**; **Civilization, Role of Soils**; **Factors of Soil Formation:** Biota; Climate; Human Impacts; Parent Material; Time; **Forensic Applications**; **Geographical Information Systems**

Further Reading

Brown AG (1997) Interpreting floodplain sediments and soils. In: Brothwell D, Market G, Dincauze D, and Renouf P (eds) *Alluvial Geoarchaeology: Floodplain Archaeology and Environmental Change*, pp. 63–103. Cambridge, UK: Cambridge University Press.

Collins ME, Carter BJ, Gladfelter BG, and Southard RJ (eds) (1995) *Pedological Perspectives in Achaeological Research*, SSSA Special Publication No. 44. Madison, WI: Soil Science Society of America.

Courty MA, Goldberg P, and MacPhail R (1989) *Soils and Micromorphology in Archaeology*. Cambridge, UK: Cambridge University Press.

Hill CL, Rapp GR, and Rapp G Jr (1998) Sediments and soils and the creation of the archaeological record. In: *Geoarchaeology: The Earth-Science Approach to Archaeological Interpretation*, pp. 18–49. New Haven, CN: Yale University Press.

Holliday VT (ed.) (1992) *Soils in Archaeology: Landscape Evolution and Human Occupation*. Washington, DC: Smithsonian Institution Press.

Holliday VT (1997) Stratigraphy, soils, and geochronology of Paleoindian sites. *Paleoindian Geoarchaeology of the Southern High Plains*, pp. 50–148. Austin, TX: University of Texas Press.

Holliday VT (2004) *Soils and Archaeological Research*. Oxford, UK: Oxford University Press.

Homburg JA (1988) Archaeological investigations at the LSU Campus Mounds. *Louisiana Archaeology* 15: 31–204.

Homburg JA and Sandor JA (1997) An agronomic study of two Classic Period agricultural fields in the Horseshoe Basin. In: Homburg JA and Ciolek-Torrello R (eds) *Vanishing River: Landscapes and Lives of the Lower Verde Valley: The Lower Verde Archaeological Project*, vol. 2, *Agricultural, Subsistence, and Environmental Studies*, pp. 127–148. CD-ROM. Tucson, AZ: SRI Press/The University of Arizona Press.

Limbrey S (1975) *Soil Science and Archaeology*. New York: Academic Press.

Mandel RD and Bettis EA III (2001) Use and analysis of soils by archaeologists: a North American perspective. In: Goldberg P, Holliday VT, and Ferring CR (eds) *Earth Sciences and Archaeology*, pp. 173–204. New York: Kluwer Academic/Plenum.

Sandor JA, Gersper PL, and Hawley JW (1986) Soils at prehistoric agricultural terracing sites in New Mexico. *Soil Science Society of America Journal* 50: 166–180.

Schiffer MB (1987) *Formation Processes of the Archaeological Record*. Albuquerque, NM: University of New Mexico Press.

Waters MR (1992) The geoarchaeological interpretation of sediments, soils, and stratigraphy. In: *Principles of Geoarchaeology: A North American Perspective*, pp. 88–113. Tucson, AZ: The University of Arizona Press.

Wood RW and Johnson DL (1978) A survey of disturbance processes in archaeological site formation. In: Schiffer MB (ed.) *Advances in Archaeological Method and Theory*, vol. 1, pp. 315–381. New York: Academic Press.

B

BACTERIA

Contents
Plant Growth-Promoting
Soil

Plant Growth-Promoting

Y Bashan and L E de-Bashan, Center for Biological Research of the Northwest (CIB), La Paz, Mexico

Introduction

Plant growth-promoting bacteria (PGPB) are defined as free-living soil, rhizosphere, rhizoplane, and phylosphere bacteria that, under some conditions, are beneficial for plants (Figure 1). Most of the activities of PGPB have been studied in the rhizosphere, and to lesser extent on the leaf surface; endophytic PGPB that reside inside the plant have also been found.

PGPB promote plant growth in two different ways: (1) They directly affect the metabolism of the plants by providing substances that are usually in short supply. These bacteria are capable of fixing atmospheric nitrogen, of solubilizing phosphorus and iron, and of producing plant hormones, such as auxins, gibberelins, cytokinins, and ethylene. Additionally, they improve a plant's tolerance to stresses, such as drought, high salinity, metal toxicity, and pesticide load. One or more of these mechanisms may contribute to the increases obtained in plant growth and development that are higher than normal for plants grown under standard cultivation conditions. However, these bacteria do not enhance the genetic capacity of the plant, as genetic material is not transferred. (2) A second group of PGPB, referred to as biocontrol-PGPB, indirectly promote plant growth by preventing the deleterious effects of phytopathogenic microorganisms (bacteria, fungi, and viruses). They produce substances that harm or inhibit other microbes, but not plants, by limiting the availability of iron to pathogens or by altering the metabolism of the host plant to increase its resistance to pathogen infection. Biocontrol-PGPB can also possess traits similar to PGPB; they may fix nitrogen or produce phytohormones, for example.

Plant Growth-Promoting Bacteria in Agriculture and the Environment

Many soil and especially rhizosphere bacteria can stimulate plant growth in the absence of a major pathogen by directly affecting plant metabolism. These bacteria belong to diverse genera, including *Acetobacter*, *Achromobacter*, *Anabaena*, *Arthrobacter*, *Azoarcos*, *Azospirillum*, *Azotobacter*, *Bacillus*, *Burkholderia*, *Clostridium*, *Enterobacter*, *Flavobacterium*, *Frankia*, *Hydrogenophaga*, *Kluyvera*, *Microcoleus*, *Phyllobacterium*, *Pseudomonas*, *Serratia*, *Staphylococcus*, *Streptomyces*, and *Vibrio* and including the legume-symbiotic genus *Rhizobium*.

Treatment of plants with agriculturally beneficial bacteria can be traced back for centuries. Inoculation of legumes with symbiotic *Rhizobium* has been practiced for almost 100 years and has had a major impact worldwide on crop yields. In Eastern Europe in the 1930s and 1940s, large-scale inoculation with associative nonsymbiotic bacteria such as *Azotobacter* and *Bacillus* failed. Two major breakthroughs in PGPB research that were largely responsible for the renewed interest in the field occurred in the late 1970s: in Brazil, the late Dr J Döbereiner and coworkers rediscovered that *Azospirillum* is capable of enhancing nonlegume plant growth, and in the USA, the work of JW Kloepper and MN Schroth and coworkers showed that biocontrol agents such as *Pseudomonas fluorescens* and *P. putida* can act as pesticides to control soilborne diseases.

The best-known among the nonsymbiotic PGPB are bacteria of the genus *Azospirillum*. These bacteria

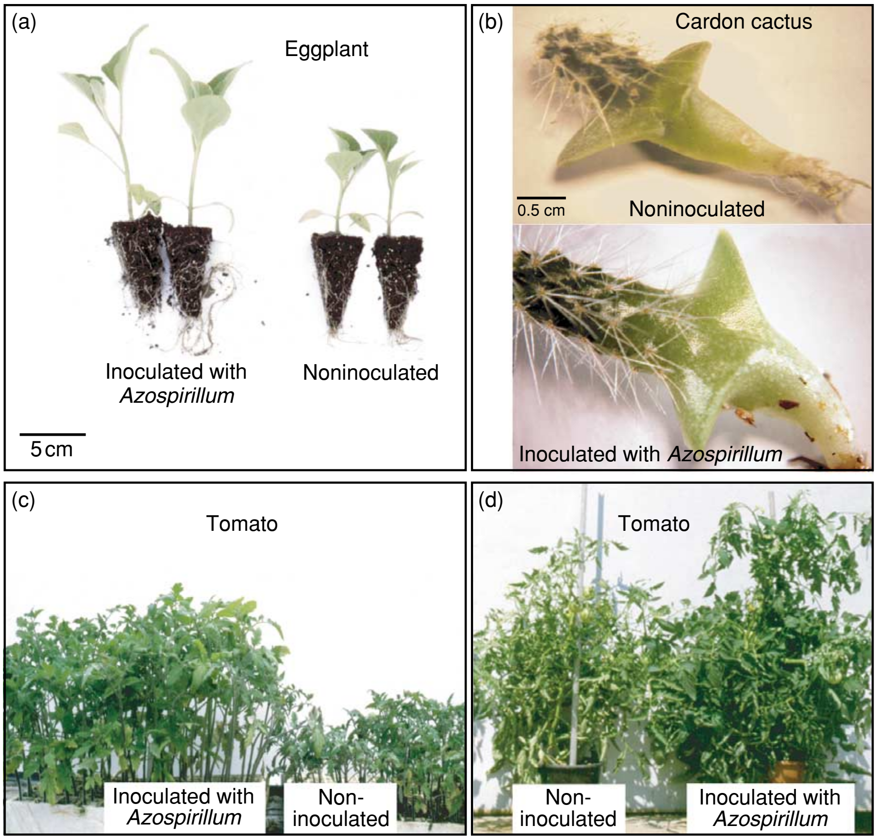

(a)
Eggplant
Inoculated with *Azospirillum*
Noninoculated
5 cm
(b)
Cardon cactus
0.5 cm
Noninoculated
Inoculated with *Azospirillum*
(c)
Tomato
Inoculated with *Azospirillum*
Non-inoculated
(d)
Tomato
Non-inoculated
Inoculated with *Azospirillum*

Figure 1 Enhanced seedling and plant growth after inoculation with *Azospirillum brasilense*. (a) Eggplant seedlings; (b) giant cardon cactus seedlings; (c) a nursery of tomatoes; (d) mature tomato plants. © Y Bashan.

enhance plant growth using a number of different mechanisms (Figure 2). Some are similar to those used by other PGPB (Table 1). Worldwide efforts to characterize this genus extensively have resulted in the availability of several commercial inoculants used for growth promotion of corn, wheat, rice, vegetables, and turf grass like Azogreen^{MR} (*Azospirillum lipoferum* on maize in France) and BioYield (*Paenobacillus macerans* and *Bacillus amyloliquefaciens* for biocontrol of tomato and pepper diseases in the USA). Although some strains of *Azospirillum* have an affinity for certain crops, the major advantage of this genus is that it is not plant-specific as it can enhance the growth of numerous plant species. Many field studies have shown that inoculation with *Azospirillum* increases crop yields by 5–30%; however, 30–40% of inoculations are unsuccessful. This inconsistency in yield stimulated experimentation with mixed inoculants, i.e., the combination of *Azospirillum* with other PGPB (Table 2). Enhanced plant growth following co-inoculation is due to the synergistic effect of both bacteria and *Azospirillum* functioning as a 'helper' bacterium to enhance the performance of other PGPB. Mixed inoculations have a higher success rate. It seems that in co-inoculated plants, nutrition is more balanced and the adsorption of nitrogen, phosphorus, and other mineral nutrients is significantly improved, yielding a better crop.

Although the PGPB described herein are associated with the plant, they are not symbiotic. Therefore, secure attachment of the bacteria to the root is essential for a long-term association for three main reasons: (1) If the bacteria are not attached to the root epidermal cells, plant growth substances excreted by the bacteria diffuse into the rhizosphere and are consumed by nutritionally versatile microorganisms before reaching the plant. (2) Without a secure attachment, water may wash the bacteria away from the rhizoplane to perish in the surrounding nutrient-deficient soil; many PGPB survive poorly in bulk soil. (3) Potential association sites for bacteria on roots, unoccupied by a PGPB, are vulnerable to colonization by aggressive, possibly nonbeneficial, root microbes. Thus, many PGPB have developed ways to remain attached to the roots, either temporarily or permanently. For example, *Azospirillum* has developed two modes of attachment (Figure 3). The first is a short-term attachment within hours after contact (after the bacteria migrate towards the roots by chemotaxis and aerotaxis, or the root reaches the

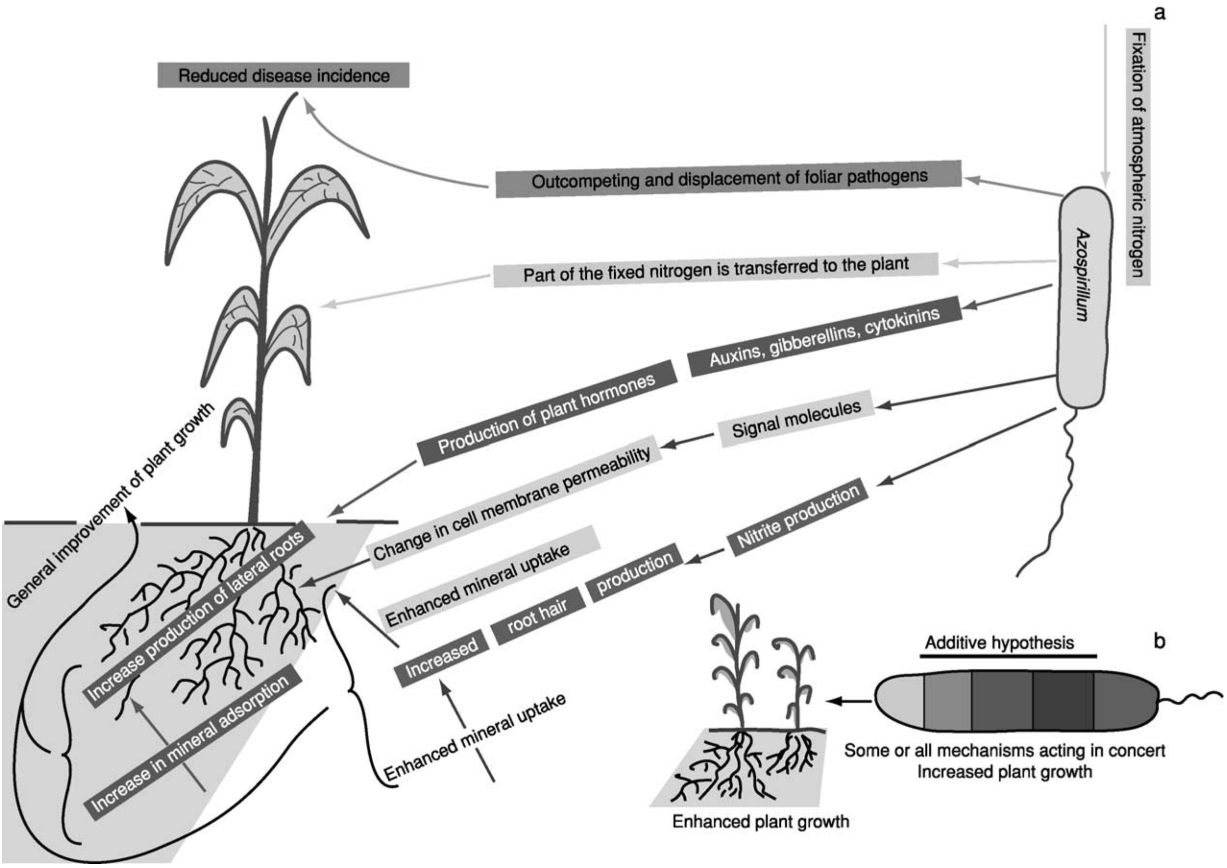

a
Reduced disease incidence
Fixation of atmospheric nitrogen
Outcompeting and displacement of foliar pathogens
Part of the fixed nitrogen is transferred to the plant
Azospirillum
Auxins, gibberellins, cytokinins
Production of plant hormones
Signal molecules
Change in cell membrane permeability
Nitrite production
General improvement of plant growth
Increase production of lateral roots
Enhanced mineral uptake
Increased root hair production
Increase in mineral adsorption
Enhanced mineral uptake
Additive hypothesis
b
Some or all mechanisms acting in concert
Increased plant growth
Enhanced plant growth

Figure 2 Mode of action of *Azospirillum* in promoting plant growth.

site of an applied inoculant). It involves hydrophobic interactions and lectin recognition between the bacteria and the plant cell wall. The second involves elaboration of a network of polysaccharide/protein fibrillar material, which anchors the bacteria permanently to the root surface. Eventually the bacterial cells multiply and form small aggregates (Figure 4) that provide an ecological advantage over the single cell state with respect to competition for nutrients that leak from the root. Similarly, the nitrogen-fixing cyanobacterium *Microcoleus chthonoplastes* enhances the production of thick mucigel layers on the roots of associated plants in which the bacteria are protected from the rhizosphere and from excessive oxygen which inhibits nitrogen fixation (Figure 5).

Because appropriate plants for colonization are not always available, Gram-negative PGPB (nonspore-forming bacteria) have developed mechanisms that allow them to survive in the absence of a host. Cells can form cysts and flocs (large, visible aggregates) that protect them from desiccation, produce melanin blocking ultraviolet irradiation, and reduce cell metabolism to the minimum required for survival. Furthermore, in times of plenty, many PGPB store large amounts of poly-β-hydroxybutyrate, that can sustain them for prolonged periods of nutrient scarcity.

In addition to their usefulness as a crop inoculant, the potential benefits of PGPB were extended to environmental applications in recent years. For example, *Azospirillum* species can enhance bioremediation of wastewater by microalgae by increasing microalgal proliferation and metabolism, allowing the microalgae to clean the water better than when used alone.

Mangrove ecosystems enhance fisheries along tropical coasts because they serve as breeding, refuge, and feeding grounds for many marine animals in the tropics during their younger and more vulnerable life stages. *Azospirillum* and cyanobacteria species may improve mangrove reforestation by increasing the rate of survival and development of seedlings in an otherwise unfavorable environment. Inoculation with several PGPB of the genera *Vibrio*, *Bacillus*, and *Azospirillum* improves domestication of the wild oilseed

Table 1 Mechanisms employed by plant growth-promoting bacteria

Mechanisms	*Effect on plant growth*	*Examples of bacterial species*
Root-associated nitrogen fixation	Increase nitrogen content and biomass	*Azospirillum, Acetobacter, Azotobacter,* Cyanobacteria, *Herbaspirillum*
Production of plant hormones (auxin, giberellin, cytokinin)	Stimulate root branching, increase shoot and root biomass, and induce the reproductive cycle	*Azospirillum*
Phosphate solubilization	Increase biomass and P content	*Bacillus lichiniformis, Vibrio*
Inhibition of plant ethylene synthesis	Increase root length	*Pseudomonas putida*
Sulfur oxidation	Increase biomass and foliar nutrient content	Undefined
Production of signal molecules and enhanced proton extrusion	Change in plant metabolism related to mineral absorption	*Azospirillum, Achromobacter*
Increase root permeability	Increase biomass and nutrient uptake	*Azospirillum*
Enhance general mineral uptake	Increase biomass and nutrient uptake	*Azospirillum*
Increase nitrite production	Increase formation of lateral roots	*Azospirillum*
Increase nitrate accumulation	Increase biomass and nitrate content	*Azospirillum*
Reduce heavy-metal toxicity	Protection against nickel toxicity	*Kluyvera*
Increase legume nodules or size	Increase biomass, N content, and reproductive yield	*Azospirillum*
Increase alder root nodules or size	Increase biomass and N content	*Frankia*
Increase frequency of infection by endomycorrhizal fungi	Increase biomass	*Pseudomonas*
Increase number of ectomycorrhizal root tips	Increase biomass	*Pseudomonas*
Increase temporary 'rain root' production in cacti	Improve survival of seedlings during drought	*Azospirillum*
Increase resistance to adverse conditions (drought, salinity, compost toxicity)	Improve survival of seedlings and increase biomass	*Azospirillum*
Additive hypothesis	Increase biomass as a result of several small-magnitude mechanisms working in concert or at the same time	*Azospirillum*

Table 2 Mixed inoculation of *Azospirillum* with other microorganisms and plants – a sampler

Mixed inoculation with species	*Plant*	*Effect on plants*
Rhizobium	Soybean, French bean, lentil, chickpea, alfalfa, bean, peanuts	Increase in nodule stimulation and function, total number and weight of nodules, epidermal cell differentiation in root hairs, straw and grain yield, root surface area, yield
Bacillus polymyxa	Sorghum	Increase in N and P uptake, grain and dry matter
Agrobacterium radiobacter or *Arthrobacter mysoreus*	Barley	Increase in grain yield, N_2 fixation, N accumulation in plant
Azotobacter chroococcum or *Streptomyces mutabilis*	Wheat, sugarcane	Increase in plant growth, indole acetic acid, P, Mg, N, and total soluble sugars in shoots, soil N content
Mycorrhizal fungi	Sorghum, wheat, jute, strawberry	Increase in dry weight of roots and shoots, grain and straw yield, mycorrhizal infection, P content, and uptake of N, Zn, Cu, and Fe

plant *Salicornia*, normally grown in mangrove ecosystems, which could be used in a seawater-irrigated agriculture system.

Treatment of cacti with *Azospirillum* enhances seedling establishment and survival in eroded desert areas. Re-vegetation of eroded and disturbed desert areas, aided by PGPB and vesicular-arbuscular mycorrhizal (VAM) fungi invigoration of desert plants responsible for soil stabilization, prevents soil erosion and promotes abatement of dust pollution (Table 3 and Figure 6). Finally, plants inoculated with *Kluyvera* have reduced nickel toxicity and can therefore grow in and rehabilitate nickel-contaminated wastelands.

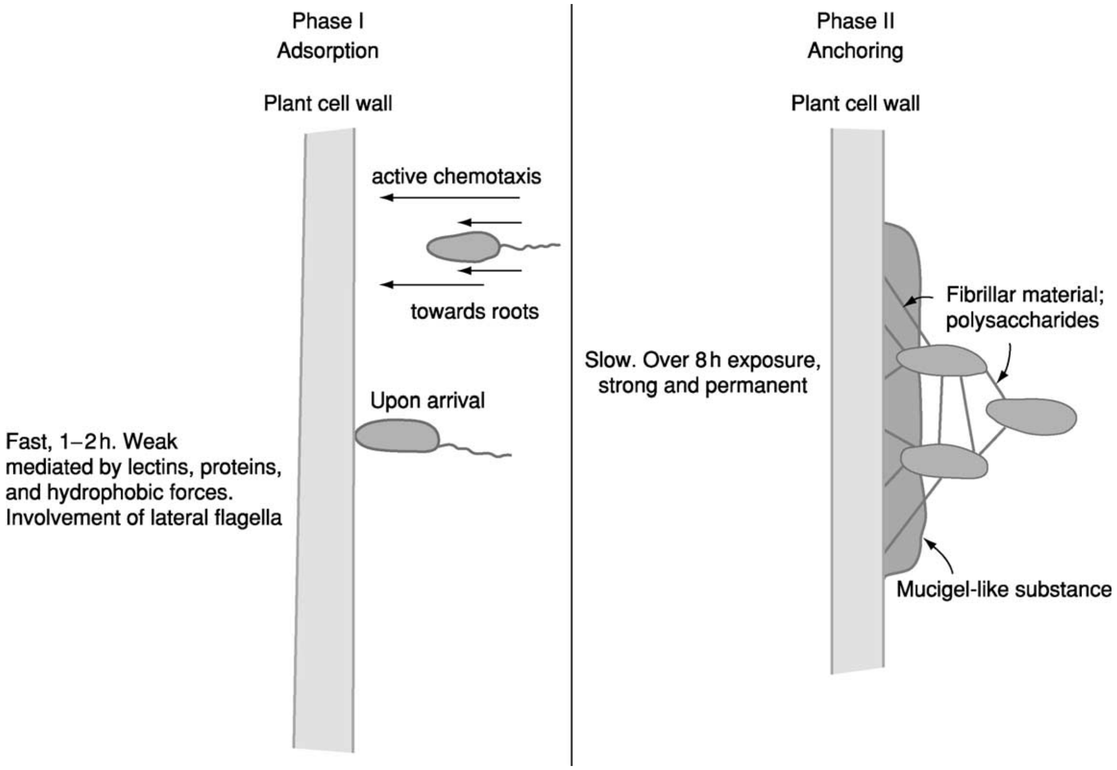


Figure 3 Mechanisms of attachment of *Azospirillum* to roots.

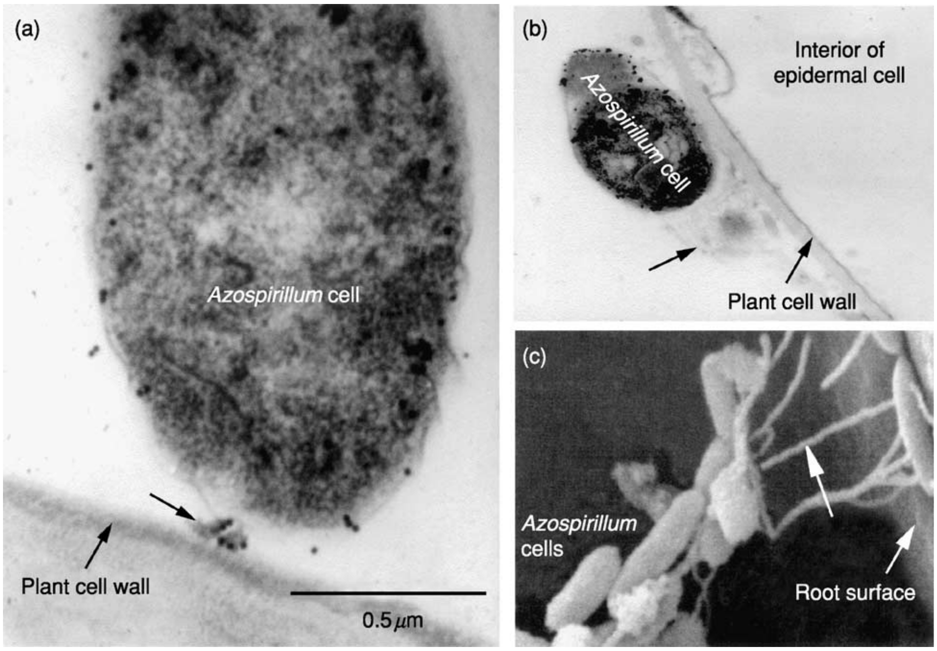


Figure 4 Transmission (a, b) and scanning (c) electron microscopy of attachment of *Azospirillum brasilense* to the root surface of wheat and cotton by fibrillar material. (a) polar attachment of bacterium to wheat plant cell wall by short fibrils (arrow); (b) nonpolar attachment of bacterium to wheat cell wall by unidentified electron-dense material (arrow); (c) permanent attachment of bacteria to cotton root surfaces through formation of long fibrils (arrow). © Y Bashan.

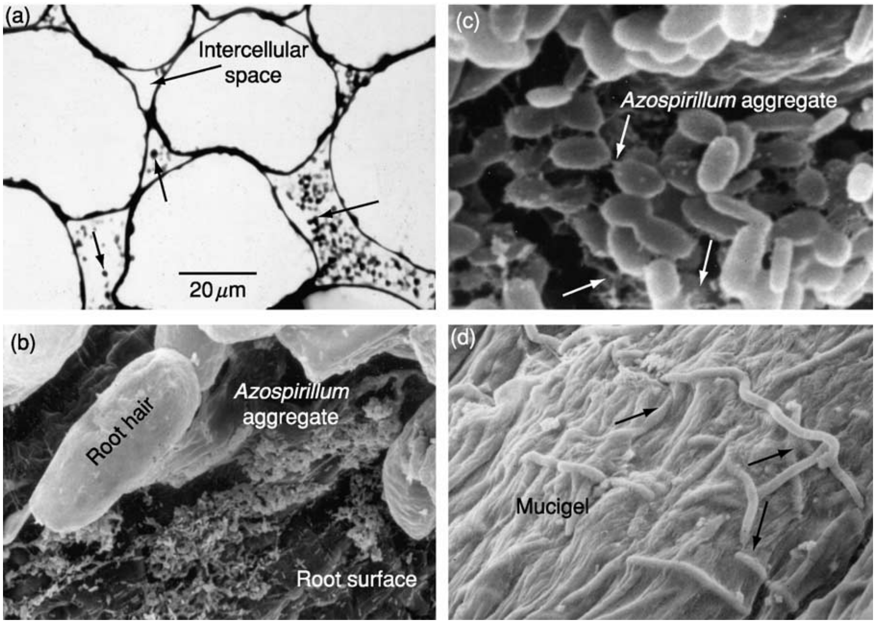


Figure 5 Light and scanning electron microscopy of interior and surface root colonization by plant growth-promoting bacteria (PGPB). (a) Light micrograph of thick cross-section of wheat roots showing the localization of *Azospirillum brasilense* within the roots. Bacteria (arrows) are located in the intercellular spaces of inner layers of cortical cells in the root elongation zone. (b, c) Root surface colonization of the elongation zone of wheat roots by inoculated *A. brasilense*. Note aggregation type of colonization. (c) is enlargement of small section of (b). Note fibril connections (arrows) between cells within the aggregate. (d) Production of mucigel layer on black mangrove roots in response to inoculation with the filamentous diazotroph cyanobacterium *Microcoleus chthonoplastes* in which the PGPB (arrows) is embedded. © Y Bashan.

Table 3 Plant growth-promoting bacteria usefulness in the environment

Bacterial species	*What it does*
Azospirillum	Helps microalgae *Chlorella* spp. to clean wastewater
Azospirillum + cyanobacteria *Microcoleus*	Improve reforestation of mangrove plants
Azospirillum + mycorrhizal fungi	Enhance seedling establishment and promote cactus growth to reduce soil erosion and dust pollution
Azospirillum, Bacillus, and *Vibrio*	Increase growth of wild oilseed plants destined for domestication
Kluyvera	Reduce nickel toxicity in polluted soil, which allows plant growth
Bacillus, Pseudomonas, and *Frankia*	Improve germination and increase growth of forest trees

Endophytic PGPB

Many bacterial species can live harmlessly as endophytes within plant tissues. They can reside latently or actively, and can colonize the plant within an organ or in the vascular system. Although most are saprophytes, some species are considered PGPB and biocontrol-PGPB because they increase plant growth and resistance to phytopathogens. For example, endophytic *Acetobacter diazotrophycus* can fix all the nitrogen required for cultivation of sugarcane and can promote pineapple growth. Some endophytic bacteria can promote growth of forest trees, from temperate pines to tree-shaped cacti. Seed endophytic PGPB from desert plants can weather rocks, unbind minerals essential for plant growth, and allow cactus seedlings to establish in barren areas as primary colonizers. These microbial activities are producing soil in bare rock areas, and consequently allow other plant species to grow. Endophytes, such as *Pseudomonas fluorescens* and *Bacillus* spp., can serve as biocontrol-PGPB controlling the soil pathogens *Fusarium* in cotton and *Rhizoctonia solani* and *Sclerotium rolfsii*.

Many endophytic bacteria invade plant tissues using mechanisms similar to pathogens, i.e., using hydrolytic enzymes, or natural (e.g., stomata) or artificial (wound) openings; however, their population density is generally lower than pathogens. Most are not recognized by the plants as potential pathogens. Endophytic bacteria rely heavily on nutrients supplied by the host plant; thus, variables affecting plant nutrition also affect the endophytic communities.

Biocontrol of Phytopathogens

Phytopathogenic microbes have an immense impact on agricultural productivity, greatly reducing crop yields and sometimes causing total crop loss. Usually,

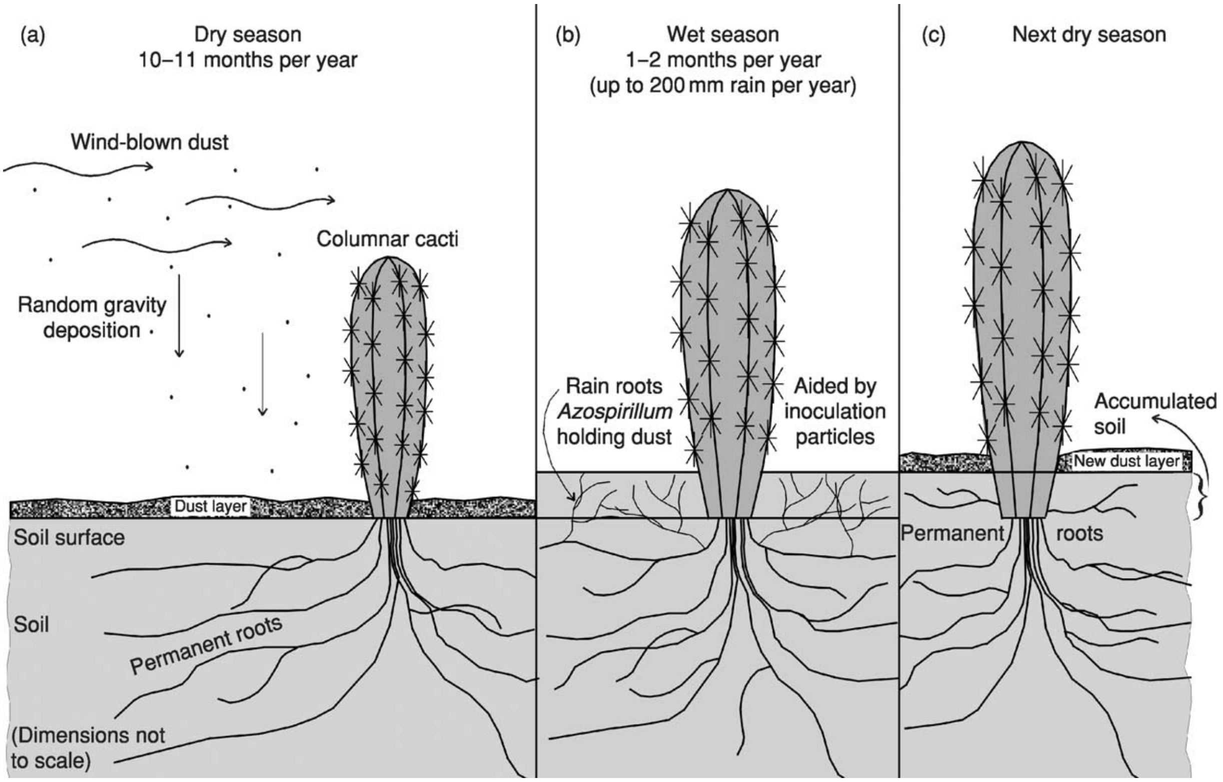


Figure 6 Mechanism of dust abatement and soil accumulation by cacti inoculated with *Azospirillum brasilense*.

growers manage phytopathogens by employing chemical pesticides and, to a lesser extent, expensive steam sterilization and 'soil solarization.' The main drawback of the chemical management strategy is that the target plants often remain infected but nonsymptomatic for prolonged periods, thus, untreated. Small environmental shifts can produce uncontrollable epidemics. Additionally, pesticides are expensive, hazardous, affecting human and animal health when they accumulate in the plants and soil, and eliminate beneficial soil and biocontrol organisms. A better strategy to avert the development of epidemics is to treat the pathogen when its levels in the field are low, to prevent further increases over the growing season. Effective options include employing the pathogen's natural enemies as biological control agents or developing transgenic plants that are resistant to the pathogen. Both strategies are considered less destructive or more 'environmentally friendly' than chemical treatments. Several biocontrol-PGPB are commercially available.

Mechanisms Employed by Biocontrol-PGPB to Control Phytopathogens

The mechanisms employed by biocontrol-PGPB to deter phytopathogens can be chemical, environmental (outcompetition and displacement of pathogens), or metabolic (induction of acquired or induced systemic resistance and modification of hormonal levels in plants) (Table 4). A large array of microbial substances is involved in the suppression of pathogenic growth and subsequent reduction in damage to plants. These substances include antibiotics such as Agrocin 84, Agrocin 434, 2,4-diacetylphloroglucinol, herbicolin, phenazine, oomycin, pyoluteorin, and pyrrolnitrin, siderophores, small molecules such as hydrogen cyanide (HCN), and hydrolytic enzymes such as chitinase, laminarinase, β-1,3-glucanase, protease, and lipase. Most studies of biocontrol mechanisms were conducted under laboratory and greenhouse conditions; however, ultimately the efficacy of biocontrol-PGPB must be tested under field conditions. Thus, the importance of the below mechanisms in controlling pathogens should be considered conditional.

Production of Antibiotics

Antibiotic production by biocontrol-PGPB is perhaps the most powerful mechanism against phytopathogens. Many different types of antibiotics are produced and have been shown to be effective under laboratory conditions, although not necessarily under field conditions. Because the genes involved in the production of some antibiotics are known, it is possible to enhance antibiotic activity, hence enhance suppression

Table 4 Mechanisms employed by biocontrol plant growth-promoting bacteria against phytopathogens

Mechanisms	*Effect on plant growth*
Competition for Fe^{3+} ions through siderophore production	Reduced disease incidence and severity and increase biomass of plants
Antibiotic production	
Production of small toxic molecules	
Production of hydrolytic enzymes	
Competition for nutrients, colonization sites, and displacement of pathogens	
Induced and acquired systemic resistance	Make plant more resistant to infection by pathogens and increase biomass of plants
Change ethylene levels in plants	Reduce noxious effect of excess ethylene production in plants
Suppression of deleterious rhizobacteria	Increase biomass of plants

of phytopathogens, at least theoretically. Regardless of the progress under laboratory conditions, only one antibiotic-producing bacterium, *Agrobacterium radiobacter*, which produces the antibiotic Agrocin 84, is commercially available. This biocontrol-PGPB (which was later genetically modified to prevent the target pathogen from easily acquiring resistance) currently controls the pathogen *A. tumefaciens*, the causative agent of crown gall of stone fruit trees.

Production of Siderophores

Iron, an element essential for microbial growth, is mostly unavailable because it is mainly present in soil in a hard-to-solubilize mineral form (Fe^{3+}). To sequester iron from the environment, numerous soil microorganisms secrete low-molecular-weight, iron-binding molecules, called siderophores, which have a high capacity for binding Fe^{3+}. The now-soluble, bound iron is transported back to the microbial cell and is available for growth.

Siderophores produced by biocontrol-PGPB have a higher affinity for iron than the siderophores produced by fungal pathogens, allowing the former microbes to scavenge most of the available iron, and thereby prevent proliferation of fungal pathogens. Depletion of iron from the rhizosphere does not affect plant growth as plants can thrive on less iron than can microorganisms. Moreover, some plants can bind and release iron from bacterial iron–siderophore complexes, and use the iron for growth. Thus, the plant benefits in two ways: from the suppression of pathogens and from enhanced iron nutrition, resulting in increased plant growth (Figure 7).

Examples of the involvement of siderophores in disease suppression are many. A mutant strain of *P. putida* that overproduces siderophores was more effective than the wild bacterium in controlling the pathogenic fungus *Fusarium oxysporum* in tomato. Many wild strains that lose siderophore activity also lose biological control activity. The extent of disease suppression as a consequence of bacterial siderophore production is affected by several factors, including the specific pathogen, the species of biocontrol-PGPB, the soil type, the crop, and the affinity of the siderophore for iron. Thus, disease suppression under controlled laboratory conditions is only an indication of the efficacy of the biocontrol agent in the field.

Production of Small Molecules

Some biocontrol-PGPB produce a wide range of low-molecular-weight metabolites with antifungal potential. The best known is hydrogen cyanide (HCN), to which the producing bacterium, usually a pseudomonad, is resistant. HCN produced by bacteria can inhibit the black root rot pathogens of tobacco, *Thielabiopsis basicola*. In soil, a biocontrol pseudomonad was capable of using seed exudates of sugar beet to produce substances inhibitory to the pathogen *Pythium ultimum*, even though the pathogen was not inhibited when the two organisms were grown together in culture medium.

Production of Enzymes

Hydrolytic enzymes produced by some biocontrol-PGPB can lyse fungal cell walls, but not plant cell walls, and thereby prevent phytopathogens from proliferating to the extent that the plant is endangered. For example, *Pseudomonas stutzeri* produces extracellular chitinase and laminarinase that lyse the pathogen *Fusarium solani*. Similarly, *Burkholderia cepacia* produces β-1,3-glucanase and reduces disease caused by the fungi *Rhizoctonia solani, Scelrotium rolfsii*, and *Phytium ultimum*. Another strategy used by biocontrol-PGPBs to reduce disease severity in plants is the hydrolysis of fungal products that are harmful to the plant. *Cladosporium werneckii* and *B. cepacia* can hydrolyze fusaric acid (produced by the fungus *Fusarium*) that causes severe damage to plants.

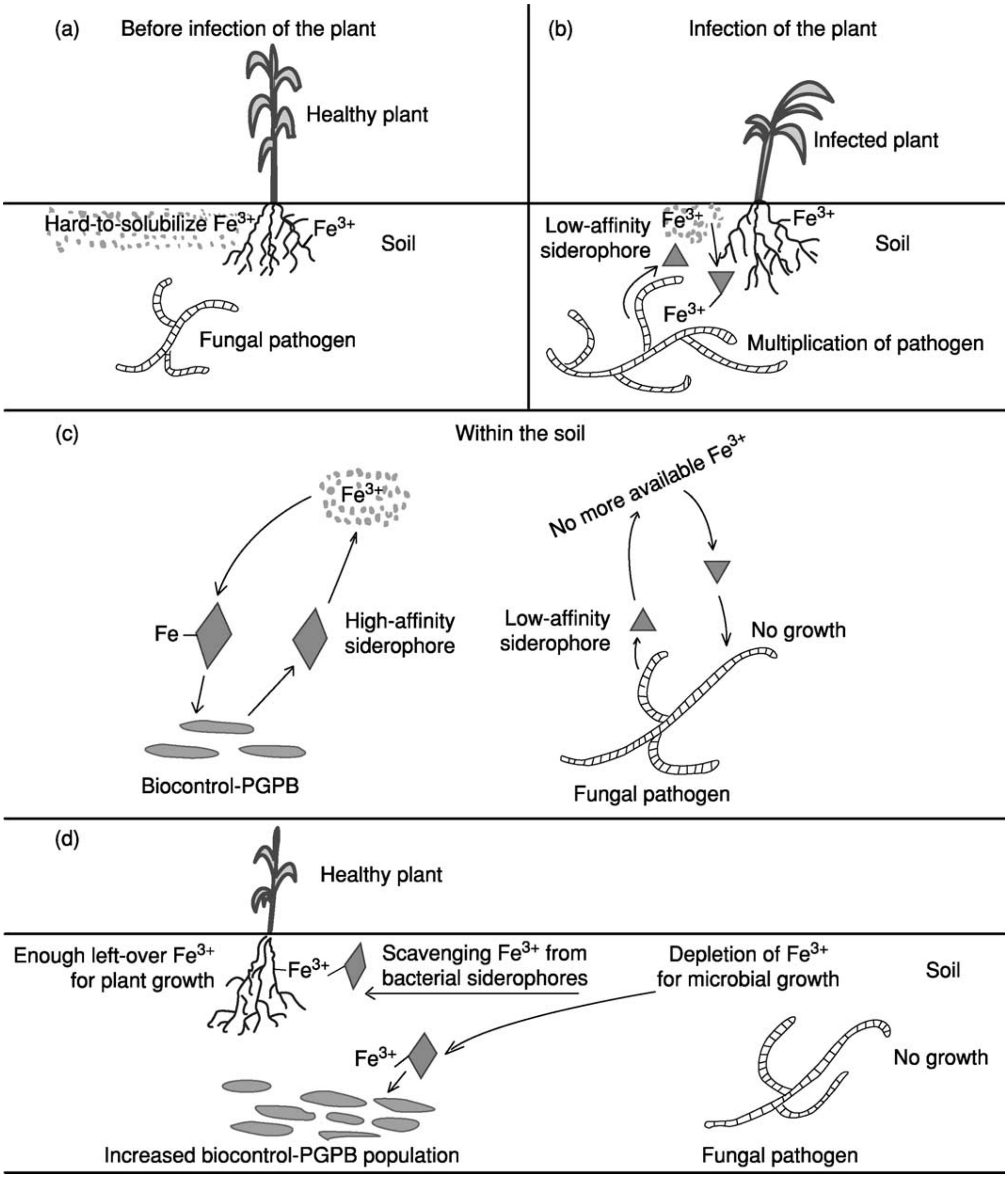


Figure 7 Biological control of fungal pathogens by biocontrol plant growth-promoting bacteria (PGPB) using siderophores.

Competition and Displacement of Pathogens

Competition for nutrients and suitable niches among pathogens and biocontrol-PGPB is another mechanism of biocontrol of some plant diseases. For example, high inoculum levels of a saprophytic *Pseudomonas syringae* protected pears against *Botrytis cinerea* (gray mold) and *Penicillium expansum* (blue mold).

On leaves there are a limited number of sites where a pathogen can attack the plant. Bacteria capable of multiplying on the leaf surface to form a large population can compete successfully with pathogens for these sites and often reduce disease. These agents can be saprophytic strains, PBPB, or nonvirulent strains of the pathogen. For example, the PGPB *A. brasilense* was able to displace the causal agent of bacterial speck disease of tomato, *P. syringae* pv. *tomato*, on tomato leaves, and consequently decreased disease development. Similarly, when a nonpathogenic strain of *P. syringae* pv. *tomato* was co-inoculated on to leaves with a pathogenic strain, disease incidence was significantly reduced. An ice-nucleation-deficient mutant of *P. syringae* displaced pathogenic *P. syringae*, and protected tomato and soybean against early frost induced by the pathogen.

Modification of Plant Metabolism

Induced and Acquired Systemic Resistance

Plants can be protected against pathogens for long periods and across a broad spectrum of disease-causing microbes by making them more resistant to infection. Exposure to pathogens, nonpathogens, PGPB, and microbial metabolites stimulates a plant's natural self-defense mechanisms before a pathogenic infection can be established, effectively 'immunizing' the plant against fungal, viral, and bacterial infections. Protection occurs by accumulation of compounds such as salicylic acid, which plays a central protective role in acquired systemic resistance, or by enhancement of the oxidative enzymes of the plant. While acquired systemic resistance is induced upon pathogen infection, induced systemic resistance can be stimulated by other agents, such as PGPB inoculants. The feasibility of protecting plants by induced systemic resistance has been demonstrated for several plant diseases. Plants inoculated with the biocontrol-PGPB, *P. putida* and *Serratia marcescens* were protected against the cucumber pathogen *P. syringae* pv. *lachrymans*.

Modification of Plant Ethylene Levels

In response to stress or pathogenic attack, plants commonly synthesize higher than normal amounts of the hormone ethylene. Ethylene stimulates senescence and leaf and fruit abscission, inhibits plant growth, and triggers cell death near infection sites. A biocontrol-PGPB that can lower plant ethylene levels after infection might be beneficial for the plant. Some PGPB synthesize the enzyme ACC deaminase, which has no known role in bacterial metabolism, but can lower the plant's ethylene levels and thereby stimulate plant growth when the plant is inoculated with the PGPB. This effect was demonstrated with the PGPB *P. putida* and with a strain of *A. brasilense* (which naturally lacks the enzyme) genetically manipulated to carry the gene for ACC deaminase.

Prospects for Improving PGPB by Genetic Manipulation

As our understanding of the mechanisms used by PGPB advances, it becomes feasible to enhance their capacity to stimulate plant growth by modifying promising traits. The activity and utility of a biocontrol-PGPB may be enhanced by supplanting it with genes responsible for the biosynthesis of antibiotics, extending the range of pathogens against which a single biocontrol-PGPB can be used, or by genetically manipulating the bacterium to increase production of the antibiotic. Similarly, it is possible to extend the range of iron–siderophore complexes that a single strain can use, allowing a biocontrol-PGPB strain to use siderophores synthesized by other soil microorganisms, hence giving it a competitive advantage. Because many of the enzymes that hydrolyze fungal cell walls are encoded by a single gene, it would be relatively easy to isolate these genes, transfer them to other biocontrol-PGPB, and thus construct new biocontrol-PGPB armed with antibiotics, siderophores, and hydrolytic enzymes. When developing an attenuated pathogenic strain to displace a pathogen in the environment, it is not only necessary to delete the virulence genes from the bacteria but also to insert copper-resistant genes in the chromosome, since copper is a major bactericide used in agriculture. It is possible to isolate bacterial genes for ACC deaminase and transfer them to biocontrol-PGPB that employ other plant growth-promoting mechanisms, allowing them to modulate ethylene levels in the host plant, and reduce disease severity.

Potential growth-promoting traits can be transferred from any bacteria to a PGPB. For example, the transfer of the gene for acid phosphatase from the saprophytic soil bacterium *Morganella morganii* to *B. cepacia* and *Azospirillum* strains would create a biocontrol-PGPB and a nitrogen-fixing bacterium with phosphate solubilization activity and therefore enhanced phosphate uptake.

Regardless of the type of genetic insertion aimed at improving the PGPB, as a general rule, wild strains (nontransformed) are likely to persist in the environment longer than their transformed relatives. However, a transformed PGPB with a short survival capacity (but long enough to last a growing season) is a bonus for commercial suppliers, who can then provide fresh inoculant to the grower on a regular basis. In addition, a short-lived genetically engineered PGPB may be more acceptable from an environmental standpoint.

PGPB Inoculants

A 'carrier' is a vehicle for delivery of live PGPB from the factory to the field. Without a suitable formulation, many promising PGPB will never reach the marketplace. A universal carrier or formulation is presently unavailable (Table 5). A good carrier has one essential characteristic: the capacity to deliver the right number of viable cells in good physiological condition at the right time (Tables 6 and 7). Peat is the most common carrier for rhizobia and many PGPB (Figure 8); however, more advanced formulations based on polymers such as alginate and liquid formulations in water and oils are constantly being

Table 5 Inoculant type carrier for plant growth-promoting bacteria

Category	*Materials*
Soils	Peat, coal, clays, and inorganic soil
Plant waste materials	Composts, farmyard manure, soybean meal, soybean and peanut oils, wheat bran, sugar industry waste, agricultural waste material, spent mushroom composts, and plant debris
Inert materials	Vermiculite, perlite, ground rock phosphate, calcium sulfate, polyacrylamide, polysaccharide-like alginate, and carraginan
Plain lyophilized microbial cultures	Culture media and cryoprotectants

Table 6 Characteristics of inoculant for plant growth-promoting bacteria (PGPB)

Characteristic	*Degree of importance*
Deliver right number of viable cells in good physiological condition at the right time	****
Bacteria released from the inoculant can inoculate the plant efficiently	****
Inexpensive raw material	****
Provide slow release of bacteria for long periods (or short periods in case of some biocontrol-PGPB)	****
Uniform, with consistent quality	***
Biodegradable by soil microorganisms	***
Contains large and uniform bacterial population	***
Ease of handling by the farmer	***
Nontoxic in nature, causes no ecological pollution (like air dispersion or entering the ground water)	**
Relatively small in volume and in nonrefrigerated conditions	**
Ease to manipulate its chemical properties in relation to the biological requirement of the PGPB	**
Applied with standard agrochemical field machinery	**
Sufficient shelf-life (1–2 years at room temperature)	**
Dry and synthetic	*
Easy quality control by the industry	*
Nearly sterile or easily sterilized	*
High water-holding capacity (for wet inoculants)	*
Suitable for many bacterial species or strains	*
Easily manufactured and mixed by existing industry	*
Allows the addition of nutrients and has an easily adjustable pH	*

Note: No single carrier can have all these qualities, but a good one should have as many as possible. Degrees of importance: *important to ****essential.

Table 7 Formulations of inoculants for plant growth-promoting bacteria

Dispersal form	*Main characteristics*	*Popularity of use*
Powders	Used as seed coating before planting. Sizes vary from 0.075 to 0.25 mm	Most common
Slurries	Powder-type inoculant suspended in liquid (usually water). Suspension drips into the furrow or seeds are dipped just prior to sowing	Less popular
Granular	Applied directly to the furrow together with the seeds. Size ranges from 0.35 to 1.18 mm	Popular
Liquids	Seeds are dipped into inoculant before sowing, or an applicator evenly sprays liquid inoculant on the seeds. After drying, the seeds are sown	Popular

evaluated. Since peat has reached its maximum development potential and still presents bacterial delivery difficulties, it appears that the future lies in the production of synthetic inoculant carriers in forms such as macro- and microbeads or powders on seed coatings (Figure 9).

Conclusions and Prospects

Application of PGPB in the field (apart from rhizobia) has yielded satisfactory results in controlled experiments, although results are less promising under agricultural conditions. Compared with chemical applications, their presence in the agricultural market is small, and nonexistent for environmental applications, where they are only experimental. However, the public, and therefore, the agrochemical industry, are now more sympathetic to the concept of PGPB inoculants. The notion prevailing today is that PGPB inoculants will complement the chemicals already on the market. It is relatively easy to isolate a new biocontrol-PGPB or to find a bacterium that will increase root development. Yet, identification of the best bacterium for the task is still difficult as the

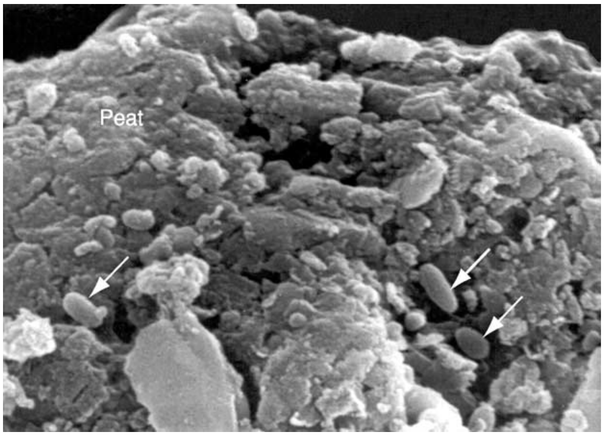


Figure 8 Scanning electron microscopy of peat inoculant (commonly used for *Rhizobium*) having a population of *Azospirillum brasilense* (arrows). © Y Bashan.

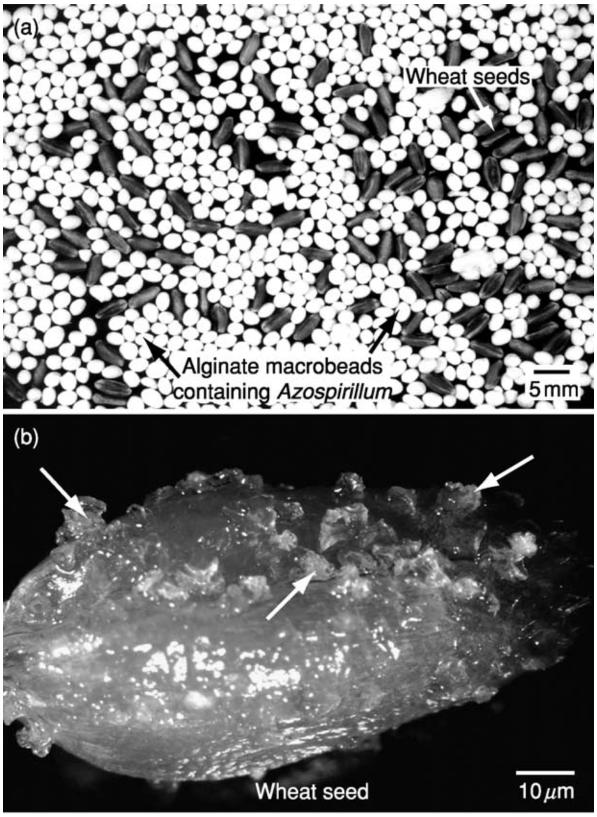


Figure 9 Synthetic inoculant carriers. (a) Macrobead inoculant made of alginate and *Azospirillum brasilense* mixed with wheat seeds before sowing. (b) Microbeads of alginate and *A. brasilense* (arrows) attached to the surface of a wheat seed. © Y Bashan.

characteristics necessary for such a PGPB are still poorly understood. Little is known about the features required for rhizo-competence and for survival and function in the environment after application.

While it is feasible to extract traits from PGPB controlled by a single gene and to use these genes to create transgenic plants that obviate the use of PGPB, PGPB use multiple mechanisms to promote plant growth, or mechanisms such as nitrogen-fixation that are impossible, as yet, to transfer. Thus, transfer of a mechanism encoded by a single gene from PGPB to plants may not provide significant benefits, although the engineering of plants with traits of PGPB has a significant presence in PGPB research.

Realistically, chemical fertilizers and pesticides will continue to dominate the marketplace in the near future. PGPB inoculants will only gradually and modestly displace chemical solutions to agricultural and environmental problems.

See also: **Mineral–Organic–Microbial Interactions**; **Mycorrhizal Fungi**; **Nitrogen in Soils:** Symbiotic Fixation

Further Reading

Bashan Y (1998) Inoculants of plant growth-promoting bacteria for use in agriculture. *Biotechnology Advances* 16: 729–770.

Bashan Y and Holguin G (2002) Plant growth-promoting bacteria: a potential tool for arid mangrove reforestation. *Trees Structure and Function* 16: 159–166.

Bashan Y, Bashan LE, and Moreno M (2001) Environmental applications of plant growth-promoting bacteria of the genus *Azospirillum*. In: De Boer SH (ed.) *Plant Pathogenic Bacteria*, pp. 68–74. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bashan Y, Holguin G, and de-Bashan LE (2004) *Azospirillum*–plant relationships: agricultural, physiological, molecular and environmental advances (1997–2003). *Canadian Journal of Microbiology* (in press).

Benizri E, Baudoin E, and Guckert A (2001) Root colonization by inoculated plant growth-promoting rhizobacteria. *Biocontrol Science and Technology* 11: 557–574.

Chanway CP (1997) Inoculation of tree roots with plant growth promoting soil bacteria: an emerging technology for reforestation. *Forest Science* 43: 99–112.

Cook RJ (1993) Making greater use of introduced microorganisms for biological control of plant pathogens. *Annual Review of Phytopathology* 31: 53–80.

Dobbelaere S, Vanderleyden J, and Okon Y (2003) Plant growth-promoting effects of diazotrophs in the rhizosphere. *Critical Reviews in Plant Sciences* 22: 107–149.

Glick BR (1995) The enhancement of plant growth by free-living bacteria. *Canadian Journal of Microbiology* 41: 109–117.

Glick BR (2003) Phytoremediation: synergistic use of plants and bacteria to clean up the environment. *Biotechnology Advances* 21: 383–393.

Glick BR, Patten CL, Holguin G, and Penrose DM (1999) *Biochemical and Genetic Mechanisms Used by Plant Growth Promoting Bacteria*. London: Imperial College Press.

Hallmann J, Quadt-Hallmann A, Mahaffee WF, and Kloepper JW (1997) Bacterial endophytes in agricultural crops. *Canadian Journal of Microbiology* 43: 895–914.

Jetiyanon K and Kloepper JW (2002) Mixtures of plant growth-promoting rhizobacteria for induction of systemic resistance against multiple plant diseases. *Biological Control* 24: 285–291.

Pankhurst CE and Lynch JM (1995) The role of soil microbiology in sustainable intensive agriculture. *Advances in Plant Pathology* 11: 229–247.

Sturz AV, Christie B, and Nowak J (2000) Bacterial endophytes: potential role in developing sustainable systems of crop production. *Critical Reviews in Plant Science* 19: 1–30.

Sutton JC and Peng G (1993) Manipulation and vectoring of biocontrol organisms to manage foliage and fruit diseases in cropping systems. *Annual Reviews of Phytopathology* 31: 473–493.

van Loon LC, Bakker PAHM, and Pieterse CMJ (1998) Systemic resistance induced by rhizosphere bacteria. *Annual Reviews of Phytopathology* 36: 453–483.

Vessey JK (2003) Plant growth-promoting rhizobacteria as biofertilizers. *Plant and Soil* 255: 571–586.

Zehnder GW, Yao C, Murphy JF *et al.* (1999) Microbe-induced resistance against pathogens and herbivores: evidence of effectiveness in agriculture. In: Agrawal AA (ed.) *Induced Plant Defenses Against Pathogens and Herbivores: Biochemistry, Ecology and Agriculture*, pp. 335–355. St. Paul, MN: APS Press.

Soil

L J Halverson, Iowa State University, Ames, IA, USA

Introduction

At densities ranging from 10^7 to 10^{11} per gram of soil, bacteria are clearly common residents of soil. The ability of bacteria to thrive in soil is due, in part, to their unequaled metabolic versatility and phenotypic plasticity, which allows them to colonize the vastly different habitats soils are comprised of. Soil is a remarkably diverse, complex, heterogeneous habitat comprised of solid, liquid, and gaseous phases that vary in both space and time, and whose complexity is influenced by the development, movement, and metabolic activity of plant roots. Furthermore, environmental conditions such as, for example, soil temperature, moisture, pH, and water availability, can fluctuate rapidly or slowly, creating stresses that bacteria have to cope with. By necessity soil bacteria need to utilize a range of colonization and survival strategies.

Given the complexity of the soil habitat, the questions arise as to how soil bacteria are able successfully to colonize a particular habitat, and what spectrum of strategies they employ for growth and survival. To understand this requires developing an appreciation of what their habitat is like, what the landscape and environment is like, and how the residents, the neighbors, and the community behave, and how their combined behaviors influence growth and survival strategies. If the life history of each individual bacterium could be monitored we would know what adaptive strategies were employed, when they were employed, and the context in which they were employed. Instead, conceptual models are developed based on our knowledge of microbial physiology and ecology and of soils themselves to understand better the spectrum of strategies bacteria utilize to survive. One such model describes bacterial fitness as the ability to grow and/or survive in soil in a particular habitat under a given environmental regime. Fitness is therefore context-dependent, since it will depend on the habitat and the environment conditions in which it is being assessed (Figure 1). The ability of bacteria to deliberately alter their habitat in response to specific environmental signals to alleviate the effect of environmental stressors on cellular physiology or to increase their ability to successfully colonize a particular habitat is particularly noteworthy.

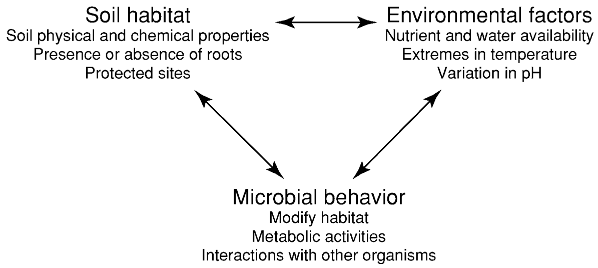


Figure 1 Model illustrating the interrelated attributes influencing bacterial fitness in soil.

Bacterial Diversity

There are two types of microbial classification systems. Phenetic classification is where relationships among organisms are based on a maximum correlation of attributes when all properties are more-or-less weighed equally. In contrast, phylogenetic classification systems are based on evolutionary relatedness, which is usually based on the similarity of nucleic acids of highly conserved macromolecules such as ribosomal ribonucleic acids (rRNAs). From comparative analyses of the small subunit of rRNA, three phylogenetically distinct cellular lineages have been identified, two of which, called the Bacteria and Archaea, are prokaryotic in cell structure (Figure 2). Representatives of the Bacteria and Archaea domains are found in soil. Proteobacteria is comprised of five

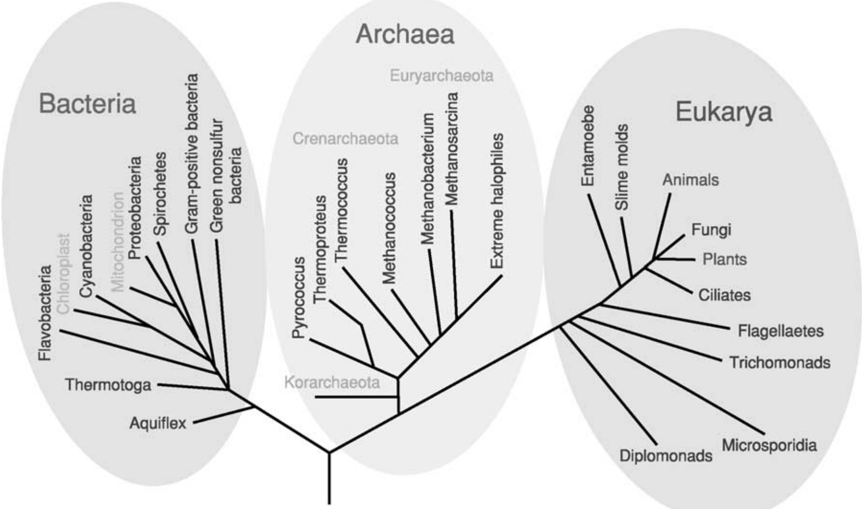


Figure 2 The three-domain classification of life based on comparative sequencing of small-subunit ribosomal ribonucleic acid molecules. This classification system separates prokaryotic organisms into two domains: Bacteria and Archaea. The third domain, Eukarya, contains all organisms composed of eukaryotic cells. Branch points and branch lengths do not reflect actual evolutionary position or distance. Figure based on data obtained from the Ribosomal Database Project II.

Table 1 Examples of physiological and ecologic classifications of soil bacteria

Category	*Classification*	*Description*
Nutritional	Heterotroph	Able to use organic compounds as energy and carbon source
	Phototroph	Able to use light as its energy source
	Autotroph	Able to use carbon dioxide as a sole carbon source
	Chemolithotroph	Able to use an inorganic substrate as an electron donor
	Saprophyte	Degrades nonliving organic material
Functional	Nitrifiers	Oxidation of ammonia to nitrite and nitrate
	Denitrifiers	Reduction of nitrate to dinitrogen gas
	Nitrogen-fixers	Reduction of dinitrogen gas to ammonia
	Sulfate-reducers	Reduction of sulfate to hydrogen sulfide
	Sulfate-oxidizers	Oxidation of elemental sulfur to sulfate
	Decomposers	Breakdown of complex compounds into simpler compounds
Interactions	Commensalism	One organism benefits while the other is neither harmed nor benefits
	Parasitism	One organism benefits while the other is harmed
	Mutualism	Both organisms benefit

subgroups and contains most of the Gram-negative bacteria found in soil. The ability to extract nucleic acids directly from soil has improved our understanding of the diversity of both cultivated and uncultivated bacteria, such as those in the division *Acidobacterium*, present in soil.

In addition to diversity at the organismal level (number of different taxonomic groups), soil microbiologists are also interested in genetic diversity (variation in genes caused by horizontal gene transfer and mutation). Currently we understand little about how genetic diversity is translated into organismal diversity and less about how genetic and organismal diversity influence ecologic and evolutionary processes. The metabolic activities and functions of soil bacteria are of especially strong interest, since they are vital to many ecologic processes.

Consequently, soil bacteria are frequently described by their physiological and ecologic properties rather than, or in addition to, their phylogenetic relationships (Table 1). This question remains as to how this vast organismal and functional diversity arise and how fitness traits, stress adaptation, and environmental parameters in soil influence this diversity.

The Microbe as the Scale of Study

It is important not to neglect the element of scale as we try to understand bacterial behavior. While macro- and mesoscale distributions of microbes indicate the average abundance of microorganisms through volumes of soil, it is the patterns of distribution and metabolic activities at the scale of the microbe (the microscale) that will control many microbial behaviors. It is also the scale at which adaptive traits of individual cells will influence both short- and long-term survival strategies, which ultimately affect the outcomes of a population (diversification) or community. From the perspective of a bacterium, a volume of soil or a root several centimeters away can represent the distance needed to travel across a small state in the USA. The question arises as to how biological, chemical, and physical attributes that define a habitat and can vary along distances of less than 1 μm to 100 μm or more affect bacterial behavior. There is a limited understanding of how bacteria perceive their immediate environment or change to it on a scale that is most relevant to them.

Increasingly bioreporters utilizing reporter-gene technology are being used to provide information on the physiological activity of an individual microbe, usually at the level of transcription of a target gene that is responsive to a particular stimulus (Figure 3). Moreover, this information may help us gain an understanding of how bacteria perceive their habitat and the physical, chemical, and biological properties of that habitat. Reporter-gene technologies are also being developed that permit the identification of environmental conditions, such as bioavailability of a nutrient or pollutant, or patterns in expression of target genes even if the target compound is ephemeral or the gene is transiently expressed. The future of bioreporters in microbial ecology is great, since they provide a means to obtain a higher resolution of the bacterial habitat that is needed for understanding many bacterial processes.

Two Basic Survival Strategies

There are at least two strategies for bacteria to use to survive the large and often rapid fluctuations in environmental conditions that occur in a soil. One is a strategy of avoidance that requires the ability to seek and exploit sites that are protected from stresses. Alternatively, a strategy of tolerance requires the ability to tolerate environmental stresses (nutrient deprivation, low water availability, extremes in temperature). There is most likely a spectrum of strategies employed, from those which employ solely a tolerance strategy to those that employ both strategies to various extents. If a bacterium is to avoid a stress it must have active mechanisms by which to facilitate movement to sites that are amenable to growth and survival and, once there, it must be able to adapt to those specific conditions or to modify the habitat so that it is more suitable for survival.

Avoidance

To avoid a stress, bacteria need actively to position themselves to a more hospitable site. Avoidance can happen passively when bacteria are physically repositioned in the soil profile by root movement or through some other physical disturbance, but this is not an effective strategy for surviving stressful times. If bacteria grow or survive at particular sites in soils then the ability to move actively to such sites is an important arsenal in their survival strategies. Bacterial motility is likely to play a role in positioning them in a niche that

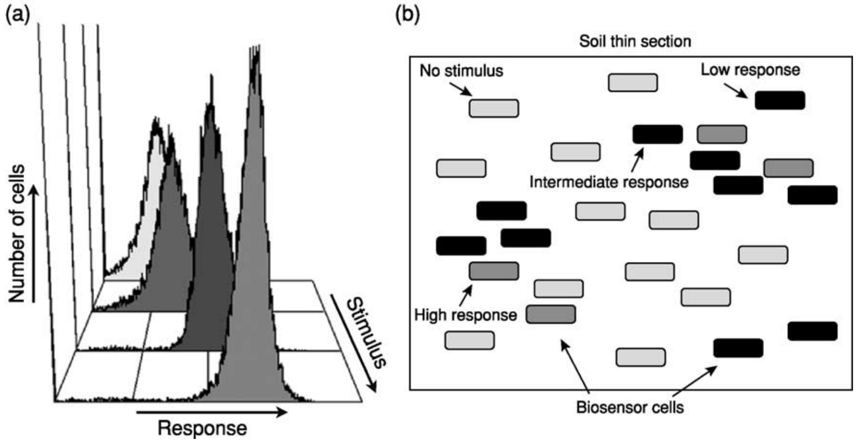


Figure 3 Use of living bacterial biosensors to describe properties of soil microhabitats: (a) response of a population of biosensor cells to a stimulus; (b) *in situ* observation of biosensor cells. For many bioreporters, reporter gene activity is proportional to the magnitude of the stimulus and is expressed as the response over the entire population (a) or on a cell-to-cell basis (b). As indicated by the different degrees of shading, there can be variation in a population response to a particular magnitude of a stimulus (a) or in the biosensors' ability to perceive a stimulus of a particular magnitude (b). Variation in stimulus exposure in a soil is inferred from the biosensors' behavior.

is abundant in desirable nutrients, low in toxic or undesirable compounds, or provides protective sites. Although flagellar motility requires a sufficiently thick water film, which is more common in wetter soils, it is possible that twitching or surface motility may play a role in positioning the cell in a favorable site, particularly when water films are too thin for flagellar motility, which more commonly occurs in unsaturated soils. It is uncertain what distances bacteria could move using this form of motility, but it would be shorter than that covered by flagellar motility.

Once a bacterium reaches a favorable site, the ability to resist removal may be advantageous. The strength of the selective pressure depends on the strength of the removal pressure, such as water percolation following rain, tillage, or movement of a root through soil. Once associated with a surface, bacteria can adhere to it by specific attachment structures such as cellulose fibrils, membrane proteins, pili, or extracellular polysaccharides (EPS). It has been hypothesized that bacteria employ specific adhesins in environments where there are strong physical forces. Due to the improbability of continually strong shear forces in soil, bacteria probably do not use adhesion extensively, except when they occupy sites that offer some protection, or are nutrient-rich, or on roots, because root movement could provide sufficiently strong physical forces to cause their removal.

Stress Tolerance

Bacteria in soil are exposed to a variety of environmental conditions that are going to affect them such as, for example, fluctuations in temperature, water availability, and nutrients, the presence of toxic environmental pollutants, or toxic metabolic wastes produced by themselves or by their neighbors. Survival in a continually changing environment requires a wide range of fast, adaptive responses, some of which do not require transcriptional activation of genes whose products facilitate coping with a given stress, since these responses take too long to complete. For example, modification of fatty acid composition in pseudomonads provides an extremely rapid response to maintain membrane fluidity when they are exposed to stresses such as desiccation or organic solvents that perturb membrane properties. Transcriptional activation of genes whose products facilitate coping with stress provide more long-term solutions. It is not feasible to review bacterial responses to every stress they will potentially encounter in soil. Consequently, reduced water availability is highlighted, and this is likely to be a dominant factor influencing bacteria in many soils; the mechanisms of adaptation to this stress are likely to be similar to cold or heat stress, which are two common stresses bacteria frequently experience in surface soils, since they all alter physical properties of bacterial membranes.

Water Availability

If solutes are abundant in free soil-water, they could become sufficiently concentrated upon drying to stress the resident bacteria. Osmotic stress is only one component of the total water stress that a bacterium may encounter in soil as it dries. The total soil-water potential is a quantitative term reflecting water availability. It is comprised primarily of the sum of the osmotic potential, which is due to the interaction of water with solutes, and the matric potential, which is due to the interaction of water with soil constituents that increases as the soil dries. From a bacterial perspective, the primary difference between these two stresses is that with an osmotic stress bacteria are bathed in water (albeit water with diminished activity), whereas with a matric stress bacteria become desiccated by removal of water from its environment, and the availability of the water that is remaining is reduced due to its sorptive interactions with soil constituents. There is ample evidence demonstrating that the stress imposed by a low matric potential has a stronger influence on a bacterial cell than an equivalent osmotic potential. This is due in part to the greater cellular dehydration that occurs under a matric than an osmotic stress. Dehydration produces deoxyribonucleic acid (DNA) damage, denatures proteins, and causes membranes to become less fluid and potentially damaged. Resistance to dehydration is crucial for microbial life in soil habitats. These include various strategies to counter the life-threatening, lipid-solidifying effects of desiccation, synthesizing solutes compatible with cellular physiology to create an intracellular water potential that is in equilibrium with the external environment, protecting and repairing DNA, and producing EPS.

Extracellular Polysaccharide Production

There is considerable evidence that soil bacteria are capable of producing EPS, and increased EPS production by soil bacteria has been shown to improve soil aggregation and aggregate stability, and water retention properties. These polysaccharides are likely to have a dominant role in the ecology of these organisms, since their presence defines the environment immediately surrounding the bacteria. As such, they probably mediate the interactions of the bacteria with the surface, other microorganisms, and the physical and chemical environment. The EPS matrix that surrounds them can have ion-exchange capabilities, which can concentrate nutrients from dilute sources in the vicinity of the cell or provide protection from predators and shield cells from the action of lytic enzymes, antibiotics, and other inhibitory compounds. EPS production may be particularly important during desiccating conditions, because many EPS are hydroscopic and their production results in a more hydrated microenvironment in the immediate vicinity of the cell relative to the bulk environment. The greater water content in the immediate vicinity of cell membranes may reduce the physical damage to membrane properties that might occur otherwise. This is an excellent example of how habitat modification can benefit the resident bacterium. If individual cells are capable of producing EPS then the cooperative production of an EPS layer surrounding surface-dwelling bacteria to form a microcolony is likely to provide them with numerous benefits.

Biofilms

Many studies have revealed aggregates of bacteria in bulk soil or on root surfaces. These bacteria are closely packed and covered with an amorphous material that is presumably EPS. This association of aggregates of bacteria in a matrix is analogous to that of biofilms, which are defined as assemblages of bacteria on surfaces encased in EPS of their own making. The concept of biofilms in soil is essentially unused within the soil science and microbiology research community. One reason for this omission may be that biofilms are generally viewed as occurring in aquatic systems and experimentally examined in this context. Another reason is that biofilm growth is generally considered to be continuous (such as on the surface of submerged rocks in a stream) and not patchy like microbial growth in soils. However, by definition, bacteria in soil grow as assemblages on either biotic or abiotic surfaces and are enmeshed in a hydrated matrix of microbially synthesized, extracellular polymeric substances.

Biofilms are not simply organism-containing slime layers on surfaces, but instead represent communities with a high level of organization. A defining feature of a biofilm is the microcolony, which can exhibit a complex, three-dimensional architecture (Figure 4), and cells within the microcolony exhibit gene-expression patterns specific to their position in the microcolony. Biofilms are essentially a community of bacteria that live in a specialized habitat of their own making. Many species have shown distinct developmental steps in biofilm formation, including: (1) attachment to a surface, (2) formation of a microcolony, and (3) maturation of microcolonies into an EPS-encased structure. This implies that biofilm formation may require complex coordinated communication and behaviors between multiple bacterial cells and species. It is presently unclear whether biofilms or microcolonies in soil form the complex architecture that has been observed in aquatic systems. However, given the close physical proximity and density of bacteria in soil, and the fact that they frequently grow on surfaces, it is likely that they will assume some of the coordinated behaviors and interactions typically associated with biofilms in aquatic systems.

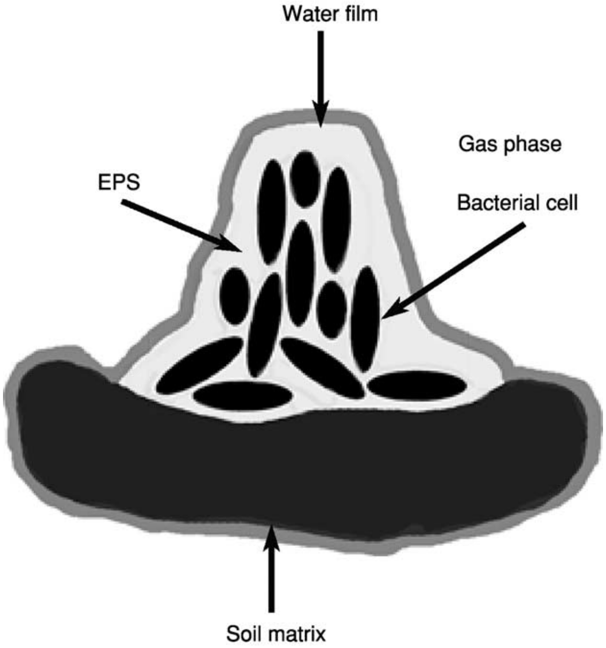


Figure 4 Ultrastructure of a bacterial aggregate or biofilm in an unsaturated soil. Cells are arranged in a structured fashion and are enmeshed in an extracellular polysaccharide (EPS) layer. Soil-water content controls water-film thickness surrounding the biofilm. Nutrients are provided by the underlying matrix that cells are adhered to, from those dissolved in soil water, or dead community members.

Competition for Resources

Bacteria of different species often occur together in soil and, if these species occupy the same microsites, the survival of any one depends on its ability to compete successfully for shared resources or to coexist by utilizing different resources. These resources can include the same physical site or be nutritionally based, and competition for nutrients is a driving factor in microbial ecology of bacteria in soil, particularly in the rhizosphere. Thus a broad nutrient utilization profile would increase the potential for an organism to survive in the presence of others and the ability to utilize an abundant resource should be highly advantageous as long as other resources are not limiting. Many soil bacteria, and in particular, rhizosphere (the root surface and the region immediately surrounding a root) colonists have broad nutrient utilization profiles. Likewise, the ability of soil bacteria to tolerate prolonged nutrient deprivation would also be highly advantageous. If microbes must compete for limited resources, production of antimicrobial metabolites, such as an antibiotic, could improve their competitive ability. Many soil bacteria are capable of producing antibiotics that target dissimilar organisms or bacteriocins that target more closely related strains of the same species, and their production would conceivably provide a competitive advantage.

The rhizosphere constitutes an ecologic niche where nutrients are more readily available compared with the bulk soil. The ability to compete effectively for these nutrients is likely to be an important factor contributing to the ability of a bacterium to become a rhizosphere colonist. The deposition of organic material from the roots (rhizodeposition) enhances microbial growth, drives the structuring of microbial communities, and controls their metabolic activities in the rhizosphere. The interactions between the plant and the surrounding bacteria in the bulk soil select for the establishment of certain populations (rhizobacteria). While rhizodeposition strongly influences the size and activity of microbial populations at the soil–plant interface, the activity of these populations in turn affects plant health and thus influences both the quantity and quality of rhizodeposition. The potential for either an exudation response to bacteria or a response by bacteria to exudation suggests a certain degree of coevolution between plants and rhizobacteria.

Cooperation Among Bacteria

The ability to modify a habitat may be augmented by cooperative interactions among bacteria and these interactions may occur among both homogeneous and heterogeneous populations. The occurrence of aggregated populations of bacteria in soil or in association with plant roots provides an opportunity to maximize opportunities for interactions among community members. There have been several recent descriptions illustrating sophisticated cooperation among community members, even though at times they may appear not to be cooperative attributes.

Siderophores

Although iron is abundant, the extreme insolubility of ferric hydroxide limits the amount of free iron available in aerobic soil environments, since the iron exists in the form of insoluble oxides in neutral or alkaline pH soils. Many bacteria use siderophores and corresponding membrane receptors to acquire this essential nutrient. Siderophores, which are low-molecular-weight molecules produced under iron-limiting conditions, bind the ferric iron, the membrane receptor recognizes a particular siderophore–iron complex, and the iron is then transported into the cell. A particular bacterial species or strain generally produces one or more siderophores and the corresponding membrane receptors for transporting each siderophore–iron complex into the cell. The bacterium releases the siderophore into the environment and then the siderophore needs to chelate ferric iron and diffuse back to the cell for the cell to benefit from the energy spent to synthesize the siderophore. Yet many soil bacterial species have the capacity to utilize siderophores produced by other bacterial strains or species and possibly fungi. This cross-feeding clearly provides advantages to community members, since there would be a greater potential return on the metabolic investment of siderophore production by taking up ferric–siderophore complexes that it did not synthesize rather than depending solely on diffusion of the siderophore it produced back into the cell. This cooperativity is not surprising given that the entire microbial community is probably experiencing some level of iron deficiency, although there may be significant microsite heterogeneity in iron availability. This also illustrates that bacterial survival may not depend solely on the efforts of the individual but also, in part, on the efforts of the community.

Quorum-Sensing

Bacteria can communicate with one another using chemical signaling molecules as words. Specifically, they release, detect, and respond to the accumulation of molecules called autoinducers. Detection of autoinducers allows bacteria to distinguish between low and high cell population density, and to control gene expression in response to changes in cell number (Figure 5). In reality, this is not necessarily a function of high cell density, but is likely to be related more directly to high signal concentrations, which are likely to be achieved at lower population densities in soil due to their low water volumes that result in relatively high autoinducer concentrations. This process, termed ‘quorum-sensing,’ allows a population of bacteria to coordinately control gene expression of the entire community and allows the bacteria to behave as multicellular organisms and to reap the benefits that would be unattainable to them as individuals. Recent studies show that highly specific as well as universal quorum-sensing signaling systems exist which enable bacteria to communicate within and between species, and that more frequently this behavior is associated with bacteria that have been isolated from plant roots than from bulk soil. Many microbial

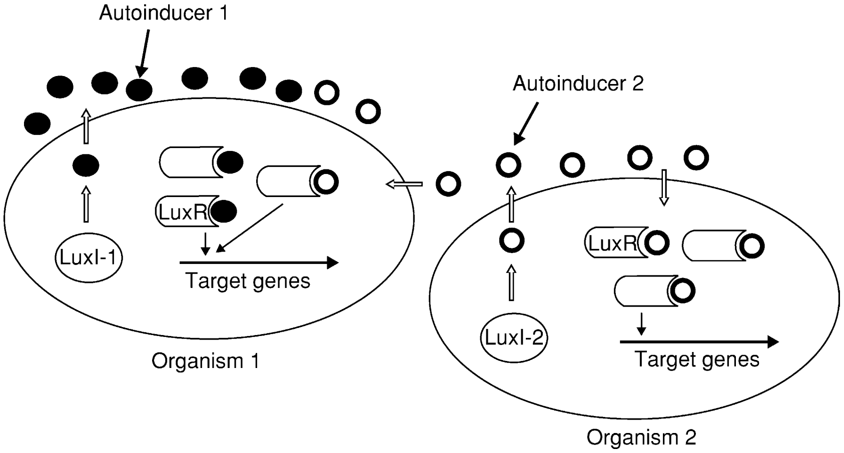


Figure 5 The LuxI/LuxR quorum-sensing signal transduction system. Two regulatory proteins control quorum-sensing. The LuxI-like proteins are the autoinducer synthases (*N*-acyl-homoserine lactones), which can freely diffuse through the cell membrane and accumulate in high density. At a particular intracellular concentration of the autoinducer, it will bind to the LuxR family of proteins that can then interact with target gene promoters to activate transcription. It is possible for a particular LuxR-family protein to bind to various autoinducers (produced by different species) that differ slightly chemically, although at different efficiencies, and still be able to activate transcription. The potential for cross-talk between two bacteria is demonstrated, because transcriptional activation of target genes in organism 1 can be mediated by autoinducer 1 and autoinducer 2 interacting with LuxR.

behaviors are regulated by quorum-sensing, including, for example, expression of nodulation genes in the *Rhizobium*–legume symbioses, virulence traits of bacterial pathogens of plants, conjugal transfer of plasmids between bacteria, antibiotic production, and biofilm formation. Bacterial biosensors using fluorescence-based reporter gene products have been particularly useful for visualizing cell–cell communication between different bacteria colonizing a plant rhizosphere. These studies illustrate the power and future of reporter-gene technology in microbial ecology research, and future studies, utilizing a spectrum of different-colored fluorescent proteins as bioreporters, will provide a more detailed picture of microbial behavior, habitat conditions, and interactions among community members.

Many bacteria have also developed mechanisms to interfere with bacterial quorum sensing, such as secreting enzymes that destroy autoinducers or producing autoinducer antagonists. Furthermore, some plants secrete substances that mimic or interfere with *N*-acyl-homoserine lactone autoinducer signal activities and affect population density-dependent behaviors in associated bacteria. These autoinducer signal-mimic compounds could prove to be important in determining the outcome of interactions between higher plants and a diversity of pathogenic, symbiotic, and saprophytic bacteria. Furthermore, some bacteria are capable of recognizing autoinducers produced by different bacterial species, indicating that interspecies quorum-sensing may play an important synergistic or competitive role in the dynamic behaviors of soil bacterial communities (Figure 5).

Death

One of the least understood aspects of bacterial life processes in soil is death. Clearly, the lack of sufficient types or amounts of nutrients, exposure to severe environmental stresses, or virus infections could lead to the death of individual cells, populations, or entire communities. The simplest examples of programmed cell death in soil bacteria are lysis of the mother cell during sporulation of bacilli and vegetative cells during myxobacterial fruiting-body formation. In these cases nutrient deprivation triggers developmental responses initiating spore formation, because spores are better suited than vegetative cells for surviving difficult times in soil. Furthermore, there is increasing evidence suggesting that bacteria have mechanisms to kill defective or damaged cells. From this perspective, programmed death of damaged cells may be beneficial to soil bacterial communities by releasing nutrients from dying cells to be used by neighbors, to provide a structural role during complex developmental cycles, such as during fruiting-body or biofilm formation, or to prevent spread of viral infections.

Summary

The ability of bacteria to survive in soil with its inherent complexity and heterogeneity under conditions that are often severe is a testament to their metabolic versatility and phenotypic plasticity (adaptability). Strategies for growth and survival in soil vary from transcriptional regulation of single genes to cooperative interactions among community members, and most likely a combination of both.

Many bacteria are able to modify their local environment, such as by producing EPS, to facilitate growth, and their increased numbers facilitate further habitat modification. Continued multiplication results in the formation of a microcolony (biofilm) that may be homogeneous if it contains progeny from only one cell or heterogeneous if multiple species are incorporated into it. Cooperative and coordinated behaviors are required for the development of such a community of bacteria as well as for maintaining the vitality and health of the community. Each individual member of the community, through its own actions, based in part on the signals it receives from other community members, initiates a spectrum of adaptive strategies that are appropriate for the particular habitat they reside in and the specific environmental challenges they are experiencing to optimize chances for survival.

See also: **Archaea**; **Bacteriophage**; **Fungi**; **Microbial Processes:** Environmental Factors; Community Analysis; Kinetics

Further Reading

Ashelford KE, Day MJ, and Fry JC (2003) Elevated abundance of bacteriophage infecting bacteria in soil. *Applied Environmental Microbiology* 69: 285–289.

Foster RC, Rovira AD, and Cock TW (1983) *Ultrastructure of the Root–Soil Interface*. St. Paul, MN: American Phytopathological Society.

Killham K (1994) *Soil Ecology*. Cambridge, UK: Cambridge University Press.

Koch B, Worm J, Jensen LE, Højberg O, and Nybroe O (2001) Carbon limitation induces σ^s-dependent gene expression in *Pseudomonas fluorescens* in soil. *Applied Environmental Microbiology* 67: 3363–3370.

Leveau JHJ and Lindow SE (2002) Bioreporters in microbial ecology. *Current Opinion in Microbiology* 5: 259–265.

O'Toole G, Kaplan GB, and Kolter R (2000) Biofilm formation as microbial development. *Annual Review of Microbiology* 54: 49–79.

Roberson EB and Firestone MK (1992) Relationship between desiccation and exopolysaccharide production in a soil *Pseudomonas* sp. *Applied Environmental Microbiology* 58: 1284–1291.

Torsvik V and Øvreås L (2002) Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion in Microbiology* 5: 240–245.

Vogel J, Normand P, Thioulouse J, Nesme X, and Grundmann GL (2003) Relationship between spatial and genetic distance in *Agrobacterium* spp. in 1 cubic centimeter of soil. *Applied Environmental Microbiology* 69: 1482–1487.

Wisniewski-Dye F and Downie JA (2002) Quorum-sensing in *Rhizobium*. *Antonie van Leeuwenhoek* 81: 397–407.

BACTERIOPHAGE

M Radosevich, University of Tennessee, Knoxville, TN, USA
K E Williamson and K E Wommack, University of Delaware, Newark, DE, USA

Introduction

Despite the importance of bacteriophage as model systems in genetics, until recently little attention was given to natural populations of bacteriophage. The observation that viruses commonly outnumber bacteria by a factor of 10 in aquatic environments revealed our limited understanding of the larger role bacteriophage play in the evolution, population biology, and ecology of bacteria. Even fewer studies have been conducted in soil ecosystems, and the overwhelming majority of these have focused on culturing bacteriophage of specific host bacteria.

Soil is the most biologically diverse ecosystem on Earth, with estimates of the number of bacterial species on the order of 5000–10 000 g^{-1} of surface soil. If we assume that the virus-to-bacteria ratio (VBR) of 10:1 that appears to be relatively constant across aquatic environments extends to the soil ecosystem, then there may be as many as 50 000–100 000 different viral taxa in 1 g of soil. The uncharacterized genetic diversity of soil bacteriophage communities is by extension simply staggering. Recent studies in aquatic environments and in rhizosphere soil suggest that viral infection may play a very critical role in shaping the structure and function of microbial communities. The literature is replete with investigations of the fate and transport of model bacteriophage of potential human pathogens and other nonindigenous viruses in porous media. There are relatively few culture-independent and model studies in bacteriophage ecology in soils, but parallels can be drawn from the fascinating findings that are more regularly emerging in aquatic ecosystems, and which raise major research questions.

Why Soil Bacteriophage?

Perhaps the most compelling reason to improve our understanding of soil bacteriophage ecology is that infection of host bacterial communities by bacteriophages is a fundamental driver of host abundance, function, and diversity. Bacteriophage can affect host biology via three distinct mechanisms: (1) host mortality, (2) phage conversion, and (3) horizontal gene transfer by transduction. These three processes have serious implications on the natural ecologic function of microbial communities in native and managed ecosystems (e.g., nutrient cycling, control and/or abundance of plant pathogens, ecologic impact of genetically engineered organisms, and spread of antibiotic resistance genes) and bioremediation of contaminated soils (e.g., survivability of introduced microbial degraders and dissemination of catabolic genes). From a more applied perspective, a better understanding of bacteriophage–host interactions in natural ecosystems may provide an opportunity to develop bacteriophage as biocontrol agents of plant and animal pathogens.

Ecologic Role of Viruses in Microbial Communities

Bacteriophage Contain Genetic Elements that Can Increase Host Fitness: Bacteriophage Conversion

Through years of intensive study, it has been discovered that bacteriophage are an integral component in the survival and fitness of populations of pathogenic microorganisms. The wealth of evidence supporting this broad claim is too extensive to be provided here; however, a few examples are particularly illustrative of the possible role of viruses within soil microbial communities, as is the case for several pathogens, notably *Corynebacterium diphtheriae*, *Vibrio cholerae*, and enterohemorrhagic *Escherichia coli*. The critical element of the virulence of these pathogens is a toxic gene product carried on the genome of a temperate bacteriophage. Thus, only lysogenic strains of the pathogen gain the fitness-improving capability of invading and replicating within a host. It is possible that, like communities of pathogenic bacteria and their respective bacteriophage, intimate genetic relationships exist in which temperate bacteriophages confer advantageous phenotypes upon soil host populations. An example relationship between a bacteriophage and a phytopathogen is that of *Clavibacter toxicus*, the likely causative agent in ryegrass gumming disease. Strong circumstantial evidence indicates that bacteriophage infection of *C. toxicus* is important to the etiology of this disease. Acquisition of virulence determinants by disease-causing bacteria can be associated with transduction and lysogenic conversion and is a noteworthy reminder of the power of bacteriophage-mediated horizontal gene transfer. Additionally, bacteriophage and other genetic elements carry genes useful in the survival of host bacteria. Perhaps the best example is antibiotic and metal resistance. Given the collective example of pathogenic bacteria and the acknowledged role of bacteriophage in enhancing their survival and fitness, it seems appropriate to ask whether analogous systems exist between viruses and bacteria in soil.

The Role of Phage in Horizontal Gene Transfer and Bacterial Evolution

In the bacterial world, genetic recombination appears to be a relatively rare event in which a small amount of genetic material (one or a few alleles) is exchanged between two species. This occurs without the requirement that the two species be close genetic relatives. Because bacterial populations in active growth may exhibit short generation times and fast growth rates, even rare genetic recombination events can quickly alter (increase or decrease) the diversity of genotypes within a community. The mechanisms by which genetic exchange occurs in bacterial populations are well understood and have become valuable tools for microbial geneticists. However, these same processes that provide the technology selectively to engineer new bacterial phenotypes also provide the means by which genetic material can be introduced into the multitude of genetic backgrounds in natural environments. Concern about release of genetically engineered microorganisms (GEMs) continues to stimulate examination of genetic exchange processes. Whole-genome sequencing of bacteria has revealed that lateral transfer may also be an important factor in the evolution of prokaryotic genomes. Transduction (bacteriophage-mediated gene transfer) begins as a serendipitous event, in which the host deoxyribonucleic acid (DNA) (plasmid or chromosomal) is mistakenly packaged into the capsid during production of progeny bacteriophage particles. Two types of transduction are possible, generalized and specialized, according to the phenotype of the bacteriophage mediating the genetic transfer. Generalized transducing bacteriophage can carry, usually with equal frequency, any DNA contained within the host cell regardless of location, since integration of the bacteriophage genome into that of the host is entirely random. Both temperate and virulent bacteriophage can perform generalized transduction and these bacteriophage share common attributes with respect to bacteriophage genome replication and DNA packaging. Conversely, specialized transducing bacteriophage are solely temperate bacteriophage which integrate within the host chromosome and are capable of transferring only specific chromosomal genes located close to the bacteriophage integration site. Because of its indiscriminant nature, generalized transduction provides the means through which large amounts of genetic material, i.e., approximately equal to the size of the transducing bacteriophage genome, can be passed between two bacterial cells.

Although transduction has generally shown the lowest frequencies of gene transfer, it has been speculated that the bacteriophage-as-carrier aspect of this form of horizontal gene transfer (HGT) provides specific advantages in that transducing bacteriophage particles can be long-lived in the environment, provide for both protection of DNA and dispersal of genetic material, and eliminate the requirement for cell-to-cell contact. The realization that bacteriophage are abundant within many environments and data showing high transduction frequency in marine systems indicate that transduction may be an underappreciated means of HGT in bacterial populations.

Some of the best evidence for the involvement of HGT in bacterial evolution has come from detailed examination of complete bacterial genomes. Properties such as the codon usage bias and G + C content of a gene serve as signatures identifying those genes that have been introduced into a new genetic background. Using such criteria it has been determined that more than 12% of the *E. coli* genome is probably transferred from other bacteria at the rate of 31 kbp per million years. The mechanisms responsible for the movement of genes between genomes, however, do not occur slowly over the course of millions of years, but are sometimes random and sometimes directed events which occur with frequencies on the level of one per 10^4–10^6 generations. As a testament to the occurrence of horizontal transfer, elements which probably cause the transposition, such as transposons, plasmid transfer origins, or bacteriophage attachment sites (*att* sites) are often found bordering presumed transferred genes. The process of HGT is probably a common event for most groups of microorganisms, but establishment of new genes within a genome is subject to the more discriminating and deliberate force of natural selection. None the less, all gene transfer must begin with a singular event of transfer (i.e., conjugation, transformation, or transduction). The collective rates of these processes ultimately determine the appearance of new genotypes on which natural selection will ultimately act.

Evidence for Natural Transduction in Microbial Communities

Several studies have shown the potential for transduction by inoculating soil microcosms with bacteriophage and an appropriate host. The transduction of chloramphenicol and mercury resistance genes into *E. coli* has been shown, using specialized transducing derivatives of bacteriophage P1. Transfer occurs in both sterile and nonsterile soil, and when lysogenic bacteria and recipient host cells are added to soil. Transduced *E. coli* are lysogenic for phage P1 by hybridization with a DNA probe derived from the bacteriophage. Although *E. coli* is a nonindigenous host bacterium, there is potential for transduction to play an important role in gene transfer in soil.

While the potential for transduction is clearly present in soil communities, it should be stressed that culture-independent assessment of transduction has not been demonstrated in the soil ecosystem. However, we draw on several elegant studies in aquatic ecosystems as evidence illustrating the importance of transduction as a viable mechanism of HGT and driving force of host diversity in natural microbial populations. Bacteriophage DNA probes derived from several bacteriophage of *Pseudomonas aeruginosa* are used to detect the presence of prophage (lysogenized bacteria) in natural samples of lake water, sediments, and sewage. Hybridization studies have shown that 40% of *P. aeruginosa* in lake water samples contain sequences homologous to bacteriophage DNA, thus indicating that bacteriophage are widely distributed in the environment and may play a significant role in natural gene transfer by transduction or lysogenic conversion.

Streptomycin resistance has been transferred by a generalized transducing phage (F116) to *P. aeruginosa* in lake-water microcosms incubated *in situ*. Transduction frequencies range from 1.4×10^{-5} to 8.3×10^{-2} transductants per recipient and are comparable with laboratory-determined transduction frequencies. A model describing how selection and transduction affect the establishment of a new phenotype has been developed and validated in a bacterial population. In nonsterile aquatic microcosms in which transduction is prevented, the density of inoculated mock transductants (cells of a strain which possess the transductant phenotype but are not lysogenic for F116) decline over time. When transduction is permitted, the number of transductants increases suggesting that transduction could, in a diverse microbial population, maintain a phenotype that would otherwise be lost. The results of this and other studies clearly show the potential for genes from lysogens to be maintained as part of the community gene pool through transduction even if specific strains with similar phenotypes do not survive and proliferate. Some studies have shown as much as a 100-fold-higher transduction frequency in the presence of solid particle surfaces, suggesting that aggregation to soil particles could enhance contact rates between bacteriophage and host cells, thus increasing potential for infection.

If introducing bacteriophage DNA confers a selective advantage to the lysogen, such as antibiotic resistance, heavy metal resistance, ability to produce a toxin, or the ability to degrade a xenobiotic, then infection may result in enhanced host competitiveness. The long-term survival of a derivative of the temperate actinophage ØC31, called 'KC301,' which carries a thiostrepton resistance gene and confers resistance to the antibiotic to infected bacteria, has been investigated. KC301 was introduced as a lysogen of *Streptomyces lividans* TK24. The lysogen reached a population density slightly lower than the nonlysogenized counterpart. Addition of thiostrepton did not enhance survival or proliferation of KC301, but the antibiotic concentration was low and may not have provided a sufficient selective pressure to enhance the competitiveness of the lysogen. No comparable studies have been conducted in systems where xenobiotic degradation was used as the selective pressure for maintenance of lysogen-contaminated soils.

Cultivation-Independent Investigation of Soil Viral Communities

Microscopy of Soil Viruses

Bacteriophage can be found in any environment occupied by a suitable host; even in hosts that have been starved for long periods of time. Phage have been isolated from soil using specific enrichment with a variety of soil bacteria, including *Arthrobacter*, *Bacillus*, *Nocardia*, *Pseudomonas*, *Rhodococcus*, and *Streptomyces*, suggesting that phage are common inhabitants of soil. Early measurements of phage abundance in natural environments were quite low; however, these measurements were made using plaque formation in specific host bacteria and were not representative of the total bacteriophage populations. More recently, microscopic methods have been used to enumerate bacteriophage in environmental samples and the counts have increased by several orders of magnitude from previous determinations.

Major breakthroughs in our understanding of microbial ecology have come through technical improvements in direct enumeration of microorganisms. In the 1970s, application of epifluorescent light microscopy (EFM) to enumeration of bacteria revealed that natural abundance of this group in water and soil samples exceeds that obtained through cultivation-based techniques by 100- to 1000-fold. This realization has been pivotal in justifying the search for cultivation-independent approaches to the study of bacterial diversity such as the array of molecular genetic techniques based on small-subunit ribosomal ribonucleic acid (SSU rRNA). In addition, accurate estimation of bacterial abundance has led to dramatic improvements in understanding carbon and energy flow through natural ecosystems. The concept of the 'microbial loop,' in which a significant proportion of primary production, which would otherwise be lost to higher-order consumers, is incorporated into the bacterioplankton is supported by more recent knowledge of the high abundance of bacteria in many marine systems.

In comparison with the evolution of bacterial direct-counting approaches, the discovery of abundant viral populations within water samples was a dramatic, serendipitous event. Francisco Torrella and Richard Morita were the first directly to observe virus particles using transmission electron microscopy (TEM). In 0.2-μm filtered water samples, *c.* 10^4 virus-like particles (VLPs) per milliliter of seawater were observed; however, they felt natural abundance was probably higher, as prefiltration may have removed VLPs. A decade later, in a project originally designed to quantify the elemental composition of marine bacteria using TEM, extraordinarily high abundances of VLPs in water samples (*c.* 10^7–10^8 ml^{-1}) from a variety of marine environments were observed. In general, this and other early studies have demonstrated that viral abundance typically exceeds co-occurring host abundance by 10-fold; thus viruses are now widely acknowledged to be the most abundant members of marine microbial communities. The role of viruses and viral infection in influencing the composition, diversity, and productivity of marine microbial communities continues to be an open area of investigation.

TEM provided the earliest evidence of high viral abundance in water samples and continues to be the standard through which all succeeding EFM-based enumeration techniques are judged. The principal advantage of TEM-based enumeration is the provision of morphological data (e.g., capsid size and head–tail morphology) that confirms positive identification of a virus particle and enables additional characterization of a viral community. However, in comparison with epifluorescence techniques, TEM is difficult with samples that contain even a small proportion of particulate material and requires more time for sample preparation and observation. Several nucleic acid-binding fluorescent stains have been employed for direct EFM enumeration of viruses within water samples, including 4′-6′-diamidino-2-phenylindole (DAPI), YoPro, and the SYBR stains. Of these, the SYBR stains appear to be the most widely accepted due to ease of staining, exceptional brightness, and low nonspecific binding and fluorescence. Because EFM provides no morphological detail, counts using these techniques are reported as VLPs. Numerous comparative studies have shown that TEM virus counts are similar, but generally lower than counts of VLPs using EFM. The possible interference of particulates in TEM counting and nonviral, stain-positive particles in EFM are hypothesized to result in under- and overestimates of these two approaches, respectively. Nevertheless, the ease and relative precision of EFM techniques have resulted in widespread application of these approaches to viral enumeration.

In comparison with aquatic environments, there are only two reports of cultivation-independent observation or enumeration of indigenous viruses within soils. TEM has been used to show that there is an abundance of virus particles within British agricultural soils ($1.5 \times 10^7\,g^{-1}$). Corresponding bacterial direct counts, estimated by EFM and acridine orange staining, are *c*. 200-fold higher. However, control experiments indicate a 40-fold loss of bacteriophage particles to binding with the soil matrix. Adjusting for extraction efficiency, actual viral abundance is estimated to be 10-fold higher, yielding a of 0.04. This low VBR estimate for soil samples is among the lowest reported by direct counting.

EFM using SYBR Gold staining has been a robust means of estimating VLP abundance in 0.2-μm filtered extracts of Delaware agricultural soils. In an investigation of suitable methods for extraction of autochthonous viruses from soils, a grand mean of 4.3×10^8 VLP g^{-1} dry soil has been observed using EFM. Brightly staining particles are almost entirely encapsulated double-stranded DNA (dsDNA) because treatment with heat followed by DNase digestion completely eliminates VLPs (Figure 1). Treatment with DNase alone removes a statistically insignificant number of particles. While the DNase experiment demonstrates that SYBR Gold-positive particles contain dsDNA, conclusive evidence that VLPs seen using EFM are actually viruses requires TEM examination of soil extracts. Initial attempts to examine directly 0.2-μm filtered soil extracts using TEM have failed owing to excessive interference from particulate matter. Subsequently, viral extracts of soil are purified by CsCl gradient centrifugation prior to TEM. With this combination of techniques (extraction, 0.2-μm filtration, and CsCl gradient centrifugation), a morphologically diverse range of virus particles have been detected in soil samples (Figure 2). TEM examinations of both Matapeake and Evesboro soil samples indicate a viral abundance of $1.5 \times 10^8\,g^{-1}$ dry soil (grand mean). Both EFM and TEM direct counts of VLPs in Delaware soil samples exceed those reported from British soils by 10 and 28 times; however, the significantly greater abundance of viruses in Delaware soils is probably due to methodological differences in extraction of virus particles.

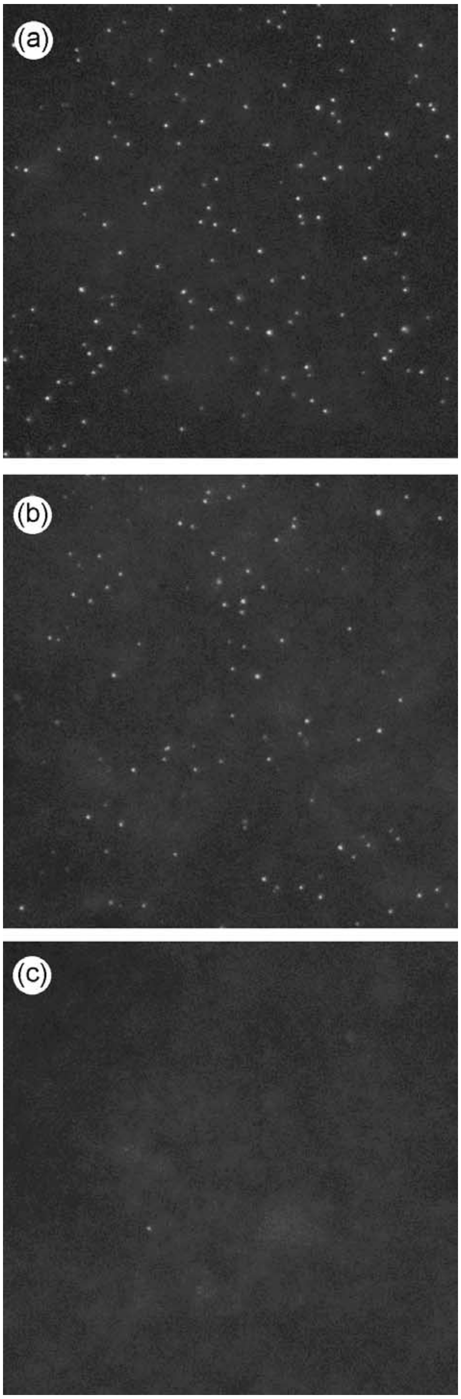


Figure 1 Epifluorescence micrographs of soil viruses stained with SYBR Gold: (a) untreated control; (b) pretreatment with DNase (1 h, 20°C); (c) pretreatment with heat (95°C, 20 min) then DNase.

TEM-based abundance is approximately five times lower than corresponding SYBR Gold epifluorescence counts for Delaware soils. The tendency of

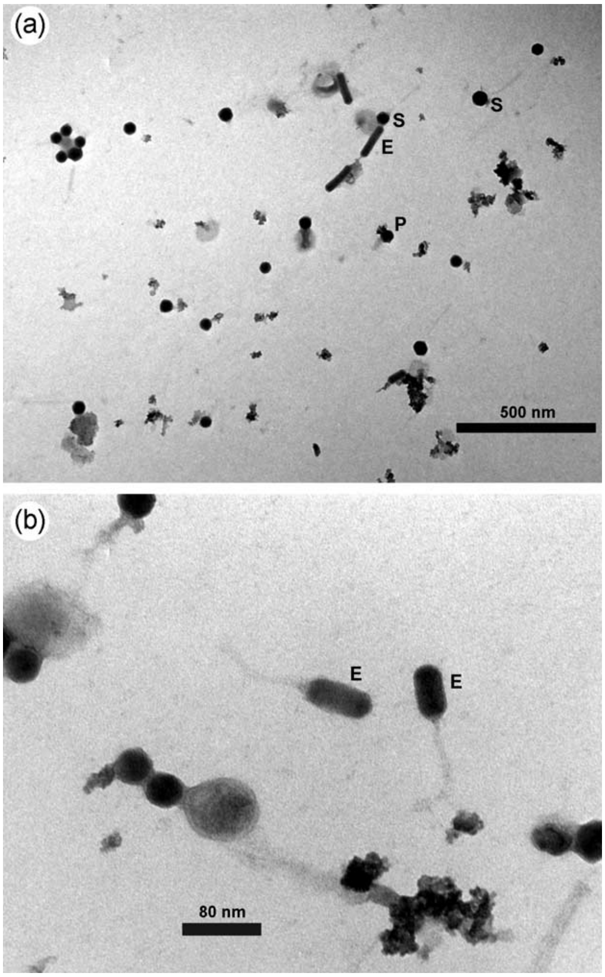


Figure 2 Transmission electron micrographs of virus-like particles from Matapeake soil: (a) ×50 000; (b) ×85 000. S, Syphoviridae; P, Podoviridae; E, elongate capsid phage.

EFM counts significantly to exceed TEM counts of virus particles is well-known for water and sediment samples and thus it is not surprising that this trend has been observed in Delaware agricultural soils. Approximately 65% of bacteriophage observed in Matapeake soil samples are tailed (49% short tails: Podoviridae; and 16% long tails: Myoviridae and Siphoviridae), and capsid diameters range from 27 to 114 nm, with a mean value of 49.6 nm ($N = 260$; from two sampling dates). Similar to aquatic viral communities in Chesapeake Bay, soil samples show a predominance of short-tailed bacteriophages as well as capsid diameters from 40 to 60 nm. Nearly 10% of soil viruses have elongated capsids (mostly morphotypes A3 (Myoviridae) and B3 (Siphoviridae)) (Figure 2). It is unlikely that the presence of elongate viruses has ever been documented for aquatic environments. Elongated morphotypes (A3, B3, and C3) comprise less than 2% of known phages, thus the observation of relatively abundant elongate bacteriophage is striking. Is it possible that these elongate viruses are specific to soils? And that such elongated capsids contain a relatively large (more than 200 kb) viral genome? This preliminary information is tantalizing evidence that soils contain morphologically diverse virus populations that differ dramatically from those occurring in better-studied, aquatic environments.

Genomic Approaches to Viral Diversity

While direct-counting approaches are an essential component to ecologic studies of microorganisms, true appreciation of extant diversity within microbial communities can only come through application of molecular genetic tools at the gene level. Approaches to studying prokaryotic diversity based on the sequence of the SSU (16S) rRNA have been critical to the formation of a universal taxonomy. In addition, techniques such as denaturing gradient gel electrophoresis (DGGE), terminal restriction fragment length polymorphism (TRFLP), and automated ribosomal intergenic spacer analysis (ARISA) applied to 16S rDNA polymerase chain reaction (PCR)-amplified gene products reveal a molecular fingerprint of prokaryotic diversity which can, in some cases, be linked to species composition. Application of molecular-fingerprinting tools to analysis of viral communities is somewhat more difficult in that universal genetic markers such as 16S rDNA do not exist for viruses. Nevertheless, certain subgroups of viruses, such as phycoviruses and cyanophages, have been examined by DGGE and TRFLP using genes (e.g., DNA polymerase, and a capsid gene) conserved among these viruses.

With the increasing availability of high-throughput DNA sequencing, it is now possible to explore, through shotgun sequencing, the genetic composition of whole microbial populations. Recently, such metagenomic approaches to the study of marine microorganisms have revealed entirely novel physiologies (i.e., phototrophy) within the pelagic ocean. Among all molecular genetic approaches used in microbial ecology, high-throughput shotgun sequencing of entire microbial communities holds the greatest promise for exploration of the vast genotypic diversity of prokaryotes. The conviction that metagenomics is the key to deeper understanding of the role of oceanic microbial communities in the global carbon cycle has led the US Department of Energy to fund a 3-year initiative to obtain the entire genetic sequence of microorganisms inhabiting the Sargasso Sea. Application of a metagenomics approach to characterization of marine viral communities demonstrates that the vast majority of viral genetic diversity is unknown. Early indications from an investigation of a Delaware agricultural soil indicate that soil

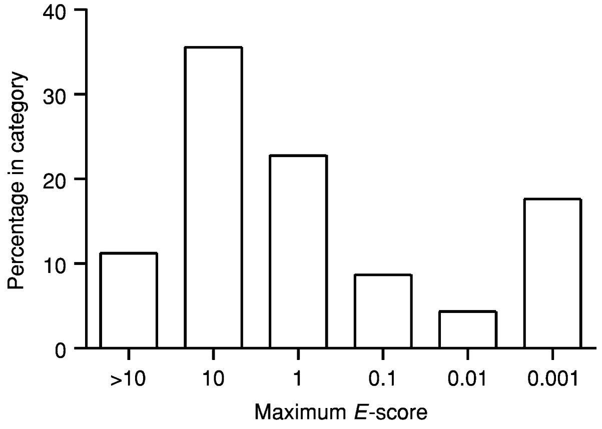


Figure 3 Distribution of BLASTX results for a soil virus metagenome library according to expectation value (*E*). Categories 10–0.1 are not significant and greater than 10 is considered 'no result.'

viruses may contain the greatest amount of unknown sequence (i.e., lacking significant homology to known sequence) of any microbial source investigated prior to 2003. Moreover, this soil-virus metagenomic library contains a larger number of sequences than any previously reported viral metagenome library. The library has been constructed from *c.* 10^9 soil VLPs using PCR-based approach for wholesale and unbiased amplification of viral DNA. After analysis of the metagenome library using PHRED/PHRAP, 9589 sequences were compared with the GenBank database using BLASTX and a minimum *E*-value of 10. Of these sequences 78% have no significant homology to known sequence at a permissive *E*-value of greater than 0.01 (Figure 3). Among sequences with significant hits ($E < 0.01$), only 33% align with known viral species, while most of those remaining align with bacterial sequences. Using a stricter criterion of $E < 0.001$, a viral metagenome library from California coastal seawater contains only 65% unknown sequence, with *c.* 34% of sequences aligning with viral sequences.

Given the prevalence of unknown sequence within metagenome libraries, it seems at first glance that this approach will offer little information concerning viral diversity. However, as more viral genomes and metagenomic sequence data become available, it is probable that significant homologies between unknown viral sequences will arise. In a way loosely similar to the discovery of new clades of bacteria through divergent SSU rRNA sequence, homologies between unknown viral sequences may lead to new genetic markers for exploration of viral evolution, taxonomy, and diversity.

Summary

In comparison with marine environments, the distribution, abundance, and ecology of soil viruses has not been thoroughly explored, yet the few initial studies available suggest that viruses are abundant members of soil ecosystems and probably play a significant role in modulating the structure, function, and productivity of soil communities. As seen in marine ecosystems, mortality of microorganisms and potentially higher organisms induced by viral infection has the potential significantly to affect the flow of energy and the biogeochemical cycling of nutrients through the soil food web. Viral transduction among indigenous host bacteria has not been definitively demonstrated in soil ecosystems but probably plays a significant role in the evolution of soil bacteria. Lack of progress in understanding these processes and the overall diversity of soil viruses has largely occurred owing to methodological barriers associated with the heterogeneous nature of the soil matrix and the general difficulties related to culture-independent surveys of biological diversity. As these barriers are broken, future research endeavors will undoubtedly discover the significance of viruses in the soil ecosystem.

List of Technical Nomenclature

EFM	Epifluoresence microscopy
HGT	Horizontal gene transfer. The transfer of genetic information from a donor cell to a recipient cell by transduction, transformation, or conjugation
TEM	Transmission electron microscopy
VLP	Virus-like particle

See also: **Biodiversity**; **Microbial Processes:** Environmental Factors; Community Analysis; Kinetics

Further Reading

Ackermann HW (1996) Frequency of morphological phage descriptions in 1995. *Archives of Virology* 141(2): 209–218.

Ashelford KE, Day MJ, and Fry JC (2003) Elevated abundance of bacteriophage infecting bacteria in soil. *Applied Environmental Microbiology* 69: 285–289.

Barksdale L and Arden SB (1974) Persisting bacteriophage infections, lysogeny, and phage conversions. *Annual Review of Microbiology* 28: 265–299.

Bergh O, Borsheim KY, Bratbak G, and Heldal M (1989) High abundance of viruses found in aquatic environments. *Nature* 340: 467–468.

Bratbak G, Heldal M, Naess A, and Roeggen T (1993) Viral impact on microbial communities. In: Guerrero R and Pedros-Alio C (eds) *Trends in Microbial Ecology*, pp. 299–302. Barcelona, Spain: Spanish Society for Microbiology.

Breitbart M, Salamon P, Andresen B *et al.* (2002) Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Science of the USA* 99(22): 14250–14255.

Danovaro R, Dell'anno A, Trucco A, Serresi M, and Vanucci S (2001) Determination of virus abundance in marine sediments. *Applied Environmental Microbiology* 67(3): 1384–1387.

Jin Y and Flurry M (2002) Fate and transport of viruses in porous media. In: Sparks DL (ed.) *Advances in Agronomy,* vol. 77, pp. 39–102. Amsterdam, the Netherlands: Academic Press.

Kepner RL Jr and Pratt JR (1994) Use of fluorochromes for direct enumeration of total bacteria in environmental samples: past and present. *Microbiological Reviews* 58(4): 603–615.

Kokjohn TA (1989) Transduction: mechanism and potential for gene transfer in the environment. In: Levy RV (ed.) *Gene Transfer in the Environment*, pp. 73–93. New York: McGraw-Hill.

Noble RT (2001) Enumeration of viruses. *Methods in Microbiology* 30: 43–50.

Paul JH and Kellogg CA (2000) Ecology of bacteriophages in nature. In: Hurst CJ (ed.) *Viral Ecology*, pp. 211–246. San Diego, CA: Academic Press.

Suttle CA (2000) Ecological, evolutionary, and geochemical consequences of viral infection of cyanobacteria and eukaryotic algae. In: Hurst CJ (ed.) *Viral Ecology*, pp. 248–296. San Diego, CA: Academic Press.

Williams ST, Mortimer AM, and Manchester L (1987) Ecology of soil bacteriophages. In: Goyl SM, Gerba CP, and Bitton G (eds) *Phage Ecology*, pp. 157–179. New York: John Wiley.

Woese CR (1987) Bacterial evolution. *Microbiological Reviews* 51(2): 221–271.

Wommack KE and Colwell RR (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews* 64: 69–114.

Zhong Y, Chen F, Wilhelm SW, Poorvin L, and Hodson RE (2002) Phylogenetic diversity of marine cyanophage isolates and natural virus communities revealed by sequences of viral capsid assembly protein g20. *Applied Environmental Microbiology* 68(4): 1576–1584.

BIOCONTROL OF SOIL-BORNE PLANT DISEASES

C E Pankhurst, CSIRO Land and Water, Townsville, QLD, Australia
J M Lynch, University of Surrey, Guildford, UK

Introduction

Soil-borne root diseases are one of the more intractable problems associated with achieving the sustainability of agriculture. Their occurrence is generally a sign of a biological imbalance within the soil ecosystem, where the natural enemies, predators, competitors, or antagonists of root parasites or disease-causing organisms are low in number and/or activity. This biological imbalance or loss of natural suppressiveness of the soil toward disease-causing organisms can, in many instances be associated with conventional agricultural management practices such as intensive soil cultivation, overuse of fertilizers and other agrichemicals, and continuous cropping. All of these practices tend to have a negative impact on the physical, chemical, and biological attributes of soil quality/health. In some cropping systems such as high-value vegetable and fruit crops, soil fumigation (or solarization) may be used to sanitize the soil and still permit growers to plant the same crop in the same field time after time. However, such practices, which work against the soil biota and debilitate its natural suppressive qualities, are not sustainable in the longer term.

Biological management or control of soil-borne root diseases has been a fruitful area of research for the past 40 years. There have been two basic research approaches. The first approach involves enhancement of the natural suppressiveness or biocontrol capability of microbial communities in the soil toward root disease organisms, through the use of alternative agricultural practices. These practices include the greater use of crop rotations and the use of organic amendments to stimulate soil microbial activity and provide a source of plant nutrients. The second approach involves the deliberate use of specific antagonist organisms for prevention and management of specific root diseases. These microbial biocontrol agents are usually single strains of bacteria or fungi which have been isolated from a soil and shown to have antagonistic properties toward the targeted root disease organism. They have been used in a number of ways, including direct addition to soil to reduce inoculum levels of pathogens in the soil and applied to seeds to protect against seed and root infections. The potential use of microbial biocontrol agents in agriculture has been hampered by numerous technical difficulties (inconsistent performance, formulation, and delivery) and complex and expensive regulatory protocols.

Enhancement of Biocontrol by Agronomic Practices

The value of enhancing the biocontrol of soil-borne plant diseases by agronomic practices has been the subject of major debate between advocates of so-called conventional agriculture and those who practice so-called alternative agriculture. Modern farming in developed and many developing countries has moved increasingly toward less dependency on naturally occurring biological controls and greater dependency on synthetic pesticides. In contrast, alternative agriculture gives emphasis to biological interactions and natural biological cycles, making them relevant rather than irrelevant to the farming system.

There are two agronomic practices frequently considered for pest control. These are (1) the greater use of crop rotations, and (2) the greater use of organic amendments and/or green-manure crops in the farming system.

Crop Rotations

The use of crop rotations in the farming system allows time for the soil microbiota to displace, weaken, or destroy the propagules of soil-borne pathogens of any one crop while another, usually unrelated crop is growing. In general, soil-borne plant pathogens multiply in the presence of their preferred host plant(s) and decline when the host plant is absent. Most of these pathogens could survive in the soil in the absence of the host plant if it were not for the combined action of competition, antibiosis, and predation/parasitism imposed by the associated soil microbiota. Whilst crop rotations do successfully reduce populations of some soil-borne pathogens, e.g., root disease of cereals caused by *Gaeumannomyces graminis* var. *tritici* (Take-all) and *Heterodera avenae* (cereal cyst nematode) is reduced by rotation with noncereal crops such as peas and lentils, there are instances, because of the broad host range of the pathogen, where this is less successful. For example, *Rhizoctonia solani*, the fungal root pathogen which causes ‘bare patch’ of cereals, has a wide host range and thus affects the growth of many plant species. It is also an efficient saprophyte, readily colonizing particulate soil organic matter. Thus, whilst there is evidence that inoculum levels of this pathogen decline following medic or pea crops, it is clearly more difficult to achieve effective control of this pathogen through crop rotations alone.

Yield decline of sugarcane in Queensland, Australia, is a good example where growth of a crop as a monoculture, coupled with intensive cultivation of the soil prior to crop establishment and mechanized harvesting, can lead to the buildup of root-pathogenic fungi (e.g., *Pachymetra chaunorhiza*) and nematodes (e.g., *Pratylenchus zeae*) which reduce sugarcane yields. The introduction of a rotation crop such as soybeans at the end of the cane-growing cycle is effective in reducing populations of these pathogens to low levels and increasing the yield of the following cane crop.

Despite the potential biological control benefits of crop rotations, there are often economic disincentives associated with this practice seen by farmers. These include the economics of growing more than one crop – the rotation crop may provide less economic return than the primary crop – and the time required to achieve effective disease reduction – it may be necessary to grow the rotation crop for one year or more. In many instances therefore, the benefits of crop rotation have given way to practices such as increased tillage, burning of crop residues, and soil fumigation, which are aimed at reducing pathogen levels in the soil. However, these practices mostly work at cross-purposes with the goal of making agriculture more sustainable.

Organic Amendments

Organic amendments, including animal manure, composts, mulches, green-manure crops, and municipal and industrial by-products, are being increasingly used in agricultural systems to recycle nutrients and energy as well as improve soil conditions for plant growth. Many organic amendments have also been shown to suppress soil-borne fungal pathogens and/or the disease they cause and several have been effectively used for control of plant-parasitic nematodes.

There are several mechanisms whereby organic amendments reduce populations of soil-borne plant pathogens. The most widely reported is the stimulation of the general microflora, including organisms that are suppressive toward the pathogen. Other mechanisms include the production of compounds toxic to pathogens (e.g., ammonia, nitrous acid, volatile fatty acids), after degradation of the amendments by soil microorganisms, and the induction of systemic resistance to pathogens. Production of compounds toxic to root pathogens is common when high nitrogen-containing organic amendments are added to soil. For example, the microsclerotia of *Verticillium dahliae*, which causes verticillium wilt of potato, are killed by ammonia and nitrous acid generated following the addition of bone meal, soymeal, poultry manure, or liquid swine manure to soil. Little is known about how organic amendments induce systemic resistance to root disease, and only a limited number of organic composts have been shown to cause this. For example, it is sufficient to expose only part of the root system of cucumber plants to a compost-amended horticultural potting mix suppressive to *Pythium* root rot in order to induce protection to the disease in the entire root system.

Enhancement of microbial populations suppressive toward soil-borne root pathogens has been demonstrated with a wide range of organic amendments in both horticulture and agriculture. Many reports show that *Pythium* and *Phytophthora* root rots are readily controlled by natural composts, whether applied as mulches to the soil surface, incorporated as soil amendments, or added as a component of potting mixes. Many types of microorganism appear to contribute toward the suppression of *Pythium* and *Phytophthora* spp. in these amended soils and potting mixes. Twenty percent of all bacterial strains recovered on 0.1-strength tryptic soy agar (TSA) from the rhizosphere of cucumber sown in a composted pine bark-amended potting mix induced biological control of *Pythium* damping-off when applied as seed treatments. Fluorescent pseudomonads, *Pantoea* and *Bacillus* spp. were the most effective in biocontrol and also the most abundant bacterial species present in the compost-amended potting mix.

Control of pathogens such as *R. solani* with organic amendments appears to be more variable than control of other pathogens such as *Pythium* spp. The basis for this difference is that *R. solani* is controlled by a much narrower spectrum of biocontrol agents and this microflora does not consistently colonize composts. This raises the important issue of the relationship between the quality (C:N ratio) and the degree of decomposition of the organic amendment and its efficacy in augmenting disease suppression. Generally speaking, fresh compost material such as green manures with a low C:N ratio can serve as a food source for both potential biocontrol agents and plant pathogens with high saprophytic ability. During this early phase in the decomposition process, the organic amendment may often increase root disease. However, in the next phase in the decomposition process, when the organic matter is fully colonized by microorganisms, the pathogen cannot effectively compete for resources and disease is suppressed. Later in the decomposition process, when the organic matter starts to become humified, the availability of resources to the biocontrol agents also becomes limiting, and this represents the phase when suppression begins to decline. Thus composted organic amendments with a high C:N ratio are more likely to be effective in pathogen control than fresh organic amendments with a low C:N ratio.

In horticulture the use of organic amendments is typically associated with the incorporation of green-manure crops or other readily available organic matter sources such as urban organic wastes. A comparison of organic and conventional farms in the central valley of California has shown that the severity of corky root of tomatoes, caused by *Pyrenochaeta lycopersici*, is less in organic farms than in conventional farms. The organically managed soils, which are fertilized by various composts and manures, have significantly higher microbial activity, which is positively correlated with lower corky root severity. In another case, municipal solid waste (fresh or composted), added 24 months previously to an arid Mediterranean soil, enhanced soil microbial activity in the soil, leading to improved biocontrol activity against *Pythium ultimum*.

Biocontrol with Introduced Microbial Biocontrol Agents

The deliberate introduction of microbial biocontrol agents into the soil to maintain the population of a targeted soil-borne plant pathogen at or below some economic threshold can be regarded as the more traditional or classic approach to biocontrol. This has been a productive area of research over the last 20 years and there are currently about 30 commercial products available worldwide for the biocontrol of soil-borne plant diseases. Some examples of commercial products of bacterial and fungal biocontrol agents used against soil-borne plant diseases are given in Table 1. The most common bacterial biocontrol agents are members of the genera *Bacillus*, *Burkholderia*, *Enterobacter*, *Pseudomonas*, and *Streptomyces*, and the most common fungal biocontrol agents are members of the genera *Gliocladium* and *Trichoderma*. Embodied in the classic approach to biocontrol are well-developed procedures for the selection, evaluation, and delivery of biocontrol agents to the soil, seed, or rhizosphere.

Selection, Evaluation, and Delivery of Microbial Biocontrol Agents

Several approaches have been used for the selection of microbial biocontrol agents, with no system apparently more successful than another. One approach has been to isolate potential microbial antagonists from the intended environment of use such as soils, seeds, or roots. This is based on the premise that any antagonist will be ecologically adapted to this environment and be able to survive and express activity when reapplied as a biocontrol agent. Another approach has been to isolate antagonists from soils suppressive to a particular pathogen. This approach has been

Table 1 Commercial biocontrol products for use against soil-borne plant diseases

Agent/product	*Target pathogens*	*Target crops*	*Application method*	*Manufacturer/distributor*
Bacteria				
Agrobacterium radiobacter				
Nogall	*A. tumefaciens*	Trees	Root dip	Bio-Care Technology, Australia
Bacillus subtilis				
Kodiak (HB, AT)	*Pythium/Rhizoctonia/ Fusarium*	Cotton/legumes	Seed treatment	Gustafson Inc., USA
Burkholderia cepacia				
Intercept	*Pythium/Rhizoctonia/ Fusarium*	Field crops/vegetables	Seed treatment/drench	Soil Technologies, USA
Pseudomonas aureofaciens				
Spotless	Turf diseases	Turf	Irrigation	Eco Soil Systems, USA
Streptomyces griseoviridis				
Mycostop	Broad	Field/vegetables/ornamental	Drench/spray	Kemira Agro OY, Finland
Fungi				
Coniothyrium minitans				
Contans	*Sclerotinia*	Field crops/vegetables	Spray	Prophyta, Germany
Fusarium oxysporum				
Biofox C	Pathogenic *Fusarium*	Vegetables/ornamental	Seed treatment/soil incorporation	SIAPA, Italy
Gliocladium virens				
SoilGard	*Pythium/Rhizoctonia*	Greenhouse crops	Granule incorporation	Grace-Sierra Co., MD, USA
Tricohoderma harzianum				
Root Shield	*Pythium/Rhizoctonia/ Fusarium*	Various	Granule incorporation/ drench	Bioworks Inc., USA

Complied from data in Whipps JM (1997) Developments in the biological control of soil-borne plant pathogens. *Advances in Botanical Research* 26: 1–135 and Stewart A (2001) Commerical biocontrol – reality or fantasy? *Australasian Plant Pathology* 30: 127–131.

used for the isolation of *Streptomyces griseoviridis*, a biocontrol agent for the control of *Fusarium* and *Pythium* spp. infections of ornamentals and vegetables. It is also used for the isolation of nonpathogenic *Fusarium oxysporum*, an effective biocontrol agent against *Fusarium* wilt of sweet potato and tomatoes. Alternatively, propagules or mycelia of pathogens have been placed in soils as baits from which antagonists have been isolated. In this case, antagonists have the potential to attack the pathogen and be adapted to the environment where the pathogen is active. This procedure has been used to obtain antagonists from sclerotia of several pathogens, including species of *Sclerotinia* and *Sclerotium*.

Having obtained a collection of antagonists, they are then screened for reproducible biocontrol activity. This usually involves a bioassay, which attempts to reproduce to some extent the conditions where biocontrol is required to act. As this often involves the screening of many hundreds of isolates, it usually involves some form of seedling bioassay carried out under controlled conditions. Successful isolates from the primary screens are then subjected to more rigorous evaluation and subsequent large-scale glasshouse or field trials.

Once a biocontrol agent has shown reproducible activity in a series of screening trials, methods for inoculum production, formulation, and application need to be considered in relation to the crop, disease, and environment of use. The production of suitable quantities of viable and active cells, spores, or biomass is the first step in this procedure. The most commonly used method is liquid fermentation utilizing a range of different and often inexpensive substrates such as molasses, brewer's yeast, and corn steep liquor. Liquid fermentations have the advantage in allowing control of nutrients, pH, temperature, and other environmental parameters, which can help optimize biomass production, spore quantity, and quality, and reduce contamination. An alternative method is solid substrate fermentation, involving a range of agricultural waste products, including lucerne powder, sugarcane bagasse, sawdust, various grains, bran, and peat.

Various methods of delivery of biocontrol agents have been developed. The method of choice for a particular agent is very much dependent on the nature of the biocontrol agent, the target pathogen, the target crop, and its method of cultivation. Methods include liquid formulations of the biocontrol agent, which are applied to the soil or growing medium as a drench, spray, or via irrigation. This is common with bacterial biocontrol agents. For example the biocontrol product Spotless, based on *Pseduomonas aureofaciens*, for control of turf diseases, is delivered via irrigation. Other methods include incorporation of the biocontrol agent into various kinds of granule formulations, which can be added to the soil at or before the time of planting, or applying the biocontrol agent to the seed before planting using specialized seed-coating or pelleting procedures.

Mode of Action of Biocontrol Agents

Several modes of action of microbial biocontrol agents have been identified, none of which are mutually exclusive. These can involve interactions between the antagonist and pathogen directly, either associated with roots or seeds, or free in the soil. Three direct modes of action are known: competition, where antagonist and pathogen compete for nutrients and/or space; antibiosis, where antagonists secrete metabolites harmful to pathogens; and parasitism, where the biocontrol agent infects the pathogen. In addition, indirect interactions are known where the plant itself responds to the presence of the antagonist, resulting in induced resistance or plant growth promotion. Often one antagonist may exhibit several modes of action simultaneously or sequentially. Also, in the case of natural biocontrol in some suppressive soils, several antagonists exhibiting a range of modes of action may act together to control disease.

Understanding the mode of action of biocontrol agents has been a fertile area of research in recent years. It holds the key for the improved selection and screening of new biocontrol agents as well as offering the possibility of improving biocontrol activity directly via genetic manipulation.

Competition Competition between antagonist and pathogens as a mechanism of biocontrol may occur at different levels. For example, competition for space or specific infection sites on roots or seeds has been proposed as a mechanism of biocontrol of pathogenic *F. oxysporum* by nonpathogenic strains of *F. oxysporum* and of pathogenic strains of *R. solani* by nonpathogenic *Rhizoctonia* spp. In both instances the pathogen is excluded by the more rapid and extensive colonization of the root surface by the biocontrol strain. Competition between microorganisms for carbon, nitrogen, and other nutrients in the rhizosphere is another well-researched mechanism of biocontrol. Competition for iron, mediated by production of iron-chelating siderophores, has been conclusively demonstrated as a mechanism of biocontrol by several species of bacteria in soils where iron is limiting. This is a widely recognized mechanism of biocontrol by fluorescent *Pseudomonas* spp., which produce a range of siderophores including pseudobactins and pyoverdines. The siderophores are thought to sequester the limited supply of iron that is available in the rhizosphere to a form that is unavailable to pathogenic fungi and other deleterious microorganisms, thereby restricting their growth.

Antibiosis Antibiosis is generally mediated by the production of low molecular weight antibiotics by the antagonist, which can inhibit the growth of the pathogen. Evidence for a role of antibiotics in biocontrol by both fungi and bacteria includes strong correlations between antibiotic production and biocontrol efficiency, demonstration that the purified antibiotic produced can mimic the effect of the biocontrol agent and demonstration that antibiotic-deficient mutants exert less biocontrol activity than wild-types. Antibiotic production by fungi exhibiting biocontrol activity has been most commonly reported for isolates of *Gliocladium* and *Trichoderma*. For example, production of gliotoxin by *G. virens* is implicated as the key factor in its biocontrol activity against *Pythium ultimum* and *R. solani*. There are large numbers of bacterial biocontrol agents that produce antibiotics and other secondary metabolites that appear to be important for the control of different fungal pathogens. One well-known example is the production of phenazine-carboxylic acid by *Pseudomonas fluorescens* strain 2-79, which has been demonstrated to be involved in the biocontrol activity of this strain against take-all of wheat caused by *Gaeumannomyces graminis* var. *tritici*.

An antibiotic produced by *P. fluorescens* (2,4-diacetylphloroglucinol) is especially active against the damping-off of sugar beet. In a recent study, its activity on *Pythium*-infected peas was comparable with that of *P. fluorescens* Q2-87, although the HCN-producer *P. fluorescens* CHAO and the competitive-excluder *P. fluorescens* SBW25 performed better.

Antibiosis also features in fungal biocontrol. For example, *Trichoderma harzianum* produces 6*n*-pentyl-2*H*-pyran-2-one and harzianopyridone, while *Gliocladium virens* (now *T. virens*) produces gliovirin. It should also be noted that *Trichoderma* spp. produce metabolites which stimulate plant growth directly.

Parasitism Parasitism of plant pathogens as a mechanism of biocontrol is usually associated with fungal biocontrol agents. Most evidence for this comes from field observations of infected fungal propagules such as spores or sclerotia. For example, oospores of *Phytophthora* and *Pythium* spp. are frequently found to be infected by *Olpidiopsis gracilis*, whilst sclerotia of *R. solani* are infected by the obligate sclerotial mycoparasite *Verticillium biguttatum*. The interaction between the mycoparasite and its host involves a sequence of processes encompassing location, contact, recognition, localized lysis, penetration, intracellular growth, and exit. Various chemical interactions are implicated in these processes, including involvement of lectins during the initial contact, and recognition between mycoparasite and the host fungus and a suite of different cell wall-degrading enzymes (e.g., β-1,3-glucanases, chitinases, proteinases, and lipases) during the penetration process. Other mechanisms of parasitism are associated with fungi such as *Verticillium chlamydosporium* and *Paecilomyces lilacinus*, which can infect the egg masses and cysts of the cereal cyst and root knot nematodes.

Induced systemic resistance Induced systemic resistance (ISR), also referred to as systemic acquired resistance, refers to the situation where the plant acquires resistance to infection from a pathogen following some initial triggering inoculation with a microorganism or some abiotic agent. Importantly, development of ISR is generally accompanied by the expression of a set of genes within the plant, including those that encode for pathogenesis-related proteins such as chitinases and β-1,3-glucanases. Most of the research on ISR has concerned foliar pathogens, and only relatively recently has the potential of this mechanism been recognized for biocontrol of soil-borne plant pathogens. Using split-root systems, both bacteria and fungi have been shown to induce resistance in several plants when applied to roots. For example, *Pseudomonas putida*, *Serratia marcescens*, and nonpathogenic isolates of *F. oxysporum* have been shown to induce systemic resistance to *F. oxysporum* f. sp. *cucumerinum* in cucumber.

The phenomenon of treating seeds, roots, or cuttings with inducing bacteria or fungi and achieving ISR to subsequent stem or foliar infection by a range of viral, bacterial, and fungal pathogens is also known, suggesting an important role for ISR in biocontrol in general.

Plant growth-promoting rhizobacteria Rhizobacteria which exert a beneficial effect on plant growth have been termed plant growth-promoting rhizobacteria (PGPR). The beneficial effect of PGPR may result from the direct biocontrol of root pathogens by any of the mechanisms described above (e.g., antibiosis), to indirect mechanisms including the direct promotion of growth via production of plant growth hormones, increased nutrient availability, stimulation of *Rhizobium* nodulation, and ISR. Growth promotion by PGPR may enable the plant to tolerate or escape the disease-causing pathogens, e.g., if the PGPR cause an enhanced seedling emergence rate, the time the plant is susceptible to preemergence damping-off pathogens such as *Pythium* spp. may be reduced.

Future Prospects

Soil-borne plant diseases are a particular problem with intensive agriculture. To a large degree, the problems they present can be associated with the cultural practices used to grow our crops including excessive tillage of the soil, continuous cropping, and reliance on synthetic inorganic fertilizers and pesticides. As a result, the health of our soils has declined, and with it most of the soils' natural capacity to suppress indigenous and exotic soil-borne plant diseases. There would appear to be two parallel drivers of change to overcome this situation. The first is based on improving soil health, by utilizing soil-management procedures that maintain soil community relationships, optimize soil organic matter levels, and encourage soil structural development. Here the objective is to capture the natural benefits of the soil biota in disease control through better crop management practices, including the improved use of crop rotations and improving the organic matter status of the soil through the use of natural and composted soil amendments. The second driver is based on the public's concern for maintenance of a clean and safe environment and hence interest in replacing chemical pesticides for disease control with biocontrol agents. However, it should be noted that, as with pesticide chemicals, the objective behind the use of biocontrol agents is the control rather than the management of root disease problems. Given that many root disease problems are caused by a suite of root pathogens (e.g., yield decline of sugarcane), some will argue that it may simply not be possible to control such problems with introduced biocontrol agents and that greater emphasis should be placed on developing management solutions to root disease problems rather than developing short-term control measures. However, the reality is that both approaches are equally valid in the quest for a more sustainable agriculture and that there will be many instances where one or both approaches will be more applicable.

Problems inherent in the quest for effective biocontrol agents of soil-borne diseases have been reviewed recently. The problems include inconsistency in performance, lack of broad-spectrum disease control activity, high costs associated with the registration of biocontrol products, and public concern over the potential risks of introducing exotic biocontrol agents into the environment. Inconsistency in the performance of biocontrol agents for soil-borne plant diseases is not surprising in view of the multitude of abiotic and biotic factors that can affect microbial survival plus the inherently large buffering capacity of the resident soil microbiota, which is more adapted to the soil conditions. A single introduced strain may flourish for a short period of time by virtue of its high initial inoculum level, and in some circumstances this longevity may be sufficient to give biocontrol (e.g., for damping-off diseases). However, the introduced strain will progressively become restricted to specific microenvironments to which it is more suited than the resident microbes and this will undermine its ability to elicit a biocontrol effect. Strategies to address this problem and also the lack of broad-spectrum control of most biocontrol agents include the use of mixtures of biocontrol agents. The use of combinations of *Pseudomonas* strains, *Phialophora graminicola*, and *Idriella bolleyi* to provide better control of cereal take-all has been proposed because of their ability to distribute themselves preferentially at different sites along the root. Multiple strains or mixtures of strains might also be expected to provide better breadth of activity by targeting multiple pathogens, and may also provide more stable and robust biocontrol across different sites and seasons. This may also prove more attractive to biotechnology companies who are reluctant to enter the biocontrol market because of the inconsistent disease control given by microbial products and the current limited size of the commercial market. The concept of combining induced resistance and treatment with a microbial biocontrol agent also warrants further investigation.

The registration of biocontrol products for soil-borne plant diseases has been hampered by them being classified as microbial pesticides. This generally means that the same time-consuming and extremely expensive efficacy–safety data package that is required for registration of a chemical pesticide is also required for registration of microbial biocontrol products. To change this situation in the foreseeable future, regulatory bodies need to be provided with more information in order to make valid judgments about the safety and environmental impact of microbial biocontrol agents. In addition distinctions may need to be made between those biocontrol agents whose mode of action is via competition, exclusion, or some other nonchemical-based mechanism and those where the primary mode of action is via the production of a toxic metabolite.

Concerns about the risks of introduced biocontrol agents to the environment have received considerable attention in recent years. Specific concerns range from the potential effect the introduced microorganism may have on nontargeted microorganisms, to the exchange and spread of genetic material between microorganisms, especially if the biocontrol agent has been genetically modified in any way. Conventional mutagenesis, protoplast fusion, and genetic modification have all been successfully used to enhance the biocontrol efficacy of biocontrol agents such as *Pseudomonas* spp., *Agrobacterium* spp., *Trichoderma* spp., *Gliocladium virens*, and saprophytic *Fusarium* spp. However, despite the fact that there has not been a single case of clear adverse effects to the environment due to the release of a genetically modified microorganism, there is a broad consensus that releases of 'novel' organisms must be preceded by careful examination of their threat to the environment. Regulations in many countries now require an analysis of environmental impact as part of an application for the registration and commercial development of not only genetically modified but also unmodified biocontrol agents.

Conclusions

Biocontrol of soil-borne plant diseases is continuing to develop strongly, utilizing approaches inherent in alternative agricultural practices where the emphasis lies in enhancing the natural biocontrol activities of microbial communities, and approaches where individual microbial biocontrol agents are introduced into the soil environment to control specific plant pathogens. Both approaches seek to harness the various forms of biocontrol that have been identified in the soil microbiota (competition, antibiosis, parasitism, induced plant resistance). With increased public awareness of the need to improve the health and safety of the environment, and to use more sustainable and environmentally friendly cultural practices in agriculture, biocontrol of soil-borne plant diseases is likely to become an increasingly important facet of food production.

Further Reading

Akhtar M and Malik A (2000) Roles of organic soil amendments and soil organisms in the biological control of plant-parasitic nematodes: a review. *Bioresource Technology* 74: 5–47.

Baker RR and Dunn PE (eds) (1990) *New Directions in Biological Control*. New York: Alan R. Liss.

Boland GJ and Kuykendall LD (eds) (1998) *Plant–Microbe Interactions and Biological Control*. New York: Marcel Dekker.

Cook RJ (1994) Introduction of soil organisms to control root diseases. In: Pankhurst CE, Doube BM, Gupta VVSR, and Grace PR (eds) *Soil Biota: Management in*

Sustainable Farming Systems, pp. 13–22. Melbourne: CSIRO Press.

Cook RJ and Baker KF (1983) *The Nature and Practice of Biological Control of Plant Pathogens*. St Paul, MN: American Phytopathological Society Press.

Dowling DN and O'Gara F (1994) Metabolites of *Pseudomonas* involved in the biocontrol of plant disease. *Trends in Biotechnology* 4: 239–249.

Hoitink HAJ and Boehm MJ (1999) Biocontrol within the context of soil microbial communities: a substrate-dependent phenomenon. *Annual Review of Phytopathology* 37: 427–446.

Hokkanen HMT and Lynch JM (eds) (1995) *Biological Control: Benefits and Risks*. Cambridge, UK: Cambridge University Press.

Hornby D (ed.) (1990) *Biological Control of Soil-borne Plant Pathogens*. Wallingford, UK: CAB International.

Kloepper JW (1993) Plant growth-promoting rhizobacteria as biological control agents. In: Metting FB Jr. (ed.) *Soil Microbial Ecology – Applications in Agricultural and Environmental Management*, pp. 255–274. New York: Marcel Dekker, Inc.

Lewis JA, Papavizas GC, and Lumsden RD (1991) A new formulation system for the application of biocontrol fungi to soil. *Biocontrol Science and Technology* 1: 59–69.

Lynch JM (1992) Environmental implications of the release of biocontrol agents. In: Tjamos EC, Papavizas GC, and Cook RJ (eds) *Biological Control of Plant Disease – Progress and Challenges for the Future*, pp. 389–397. New York: Plenum Press.

Rovira AD, Elliott LF, and Cook RJ (1990) The impact of cropping systems on rhizosphere organisms affecting plant health. In: Lynch JM (ed.) *The Rhizosphere*, pp. 389–436. Chichester, UK: Wiley-Interscience.

Stewart A (2001) Commerical biocontrol – reality or fantasy? *Australasian Plant Pathology* 30: 127–131.

Whipps JM (2001) Microbial interactions and biocontrol in the rhizosphere. *Journal of Experimental Botany* 52: 487–511.

BIODIVERSITY

D H Wall, Colorado State University, Fort Collins, CO, USA

Introduction

Soils teem with life. Representatives of almost every phylum of organism known aboveground also occur in soil. In less than a handful of soil, there can be billions of types of microbes and hundreds of species of microscopic invertebrates. The identities and natural histories of these microscopic flora and fauna and many of the larger, visible soil fauna are the least-known biota in terrestrial ecosystems. The major contributions of soil organisms to the maintenance of life on Earth are now being recognized by scientists and the public, opening a new frontier for exploration and a rising concern about the increasing degradation of the soil habitat. Soil biota are the primary drivers of numerous ecosystem processes. These activities provide a wealth of essential ecosystem services for humans, such as decaying organic matter, filtering water, stabilizing soil, generating and renewing soil fertility, providing nutrients for plant growth, modifying the hydrologic cycle (including mitigating floods and controlling erosion), and controlling pest and pathogens of plants and animals. Soil biodiversity is intimately linked to all ecosystems, terrestrial and aquatic, and to the atmosphere. Understanding how soil biodiversity will change with the rapid degradation of soils globally is important knowledge for policymakers and members of the public who need to develop and implement strategies for the future.

Diversity and Abundance

Soil biodiversity includes a plethora of life that is hidden from our everyday view: viruses; bacteria; actinomycetes; fungi; algae; the protozoans (single-celled eukaryotes); microscopic invertebrates such as rotifers, tardigrades (water bears), soil planaria, (flatworms), and nematodes (roundworms); the microarthropods, Acari (mites) and Collembola (springtails); and larger invertebrates, easily seen by the naked eye such as terrestrial gastropods (snails and slugs), isopods (pill bugs, sow bugs), Oligochaeta (enchytraeids and earthworms), spiders, scorpions, beetles, centipedes, millipedes, crustaceans, ants, and termites. Because of this diversity, the invertebrates are frequently grouped by size (body width), microfauna, mesofauna, and macrofauna (Table 1). Vertebrates also depend on the soil as a habitat, for example, moles, prairie dogs, meercats, wombats, small rodents, and some species of lizards, snakes, frogs, and even birds. Many aboveground invertebrates may be temporary inhabitants (such as nematodes parasitic on insects; immature flies in the Diptera such as tipulid crane flies and tephritid fruit flies; cicadas; and coleopteran scarab beetles) living in the soil for only one stage of their life cycle or occurring seasonally. Thus soil diversity is dynamic and tightly coupled to the aboveground biotic and climatic systems.

Table 1 Some invertebrates found in soil grouped by body width

Microfauna (<100 μm diameter)	*Mesofauna (100 μm to 2 mm diameter)*	*Macrofauna (>2 mm diameter)*
Nematoda (roundworms)	Acari (mites)	Opilones (daddy longlegs)
Protozoa	Collembola (springtails)	Isopoda (pill bugs, sow bugs)
Rotifera	Enchytraeidae (pot worms)	Chilopoda (centipedes)
Tardigrades	Isoptera (termites)	Diplopoda (millipedes)
	Formicoidea (ants)	Lumbricidae (earthworms)
		Coleptera (beetles)
		Arachnida (spiders and scorpions)
		Mollusca (snails and slugs)

Adapted from Swift MJ, Heal OW, and Anderson JM (1979) *Decomposition in Terrestrial Ecosystems*, Oxford, UK: Blackwell; Wall DH, Adams G, and Parsons AN (2001) Soil biodiversity. In: Chapin FS III, Sala OE, and Sannwald EH (eds) *Global Biodiversity in a Changing Environment*, pp. 48–49. New York: Springer-Verlag.

The diversity of species in soil appears so immense that knowledge of the numbers of species (species richness) in just two groups, the microarthropods and nematodes, could substantially increase global biodiversity estimates. Yet our knowledge of the numbers of soil species globally, even within any one group, is surprisingly poor. Data suggest 90–95% of the soil biota are unidentified. There is no one location where the total number of species or molecular types of microbes, invertebrates, or vertebrates has been assessed, although up to 1000 invertebrate species have been enumerated in a square meter of forest soil. Global estimates are based on many studies of varying numbers and sizes of soil samples from a variety of ecosystems and generally only to a shallow (0–20 cm) depth. This contributes to the underestimate of soil biodiversity given that mites and nematodes occur to 15-m depths in deserts and bacteria to greater depths. Global estimates of the total numbers of species identified so far from soils include 600 species of enchytraeids, 1500 species of soil protozoa, 5000 species of nematodes, and 20 000–30 000 species of mites (Figures 1 and 2). And, as in aboveground systems, many of the species identified in a soil sample are rare and have a limited distribution. Identification of species is complicated by the sheer numbers of organisms in the soil. Consider the billions of bacteria, meters of fungal hyphae, millions of protozoa, 10 million nematodes, 45 000 oligochaetes,

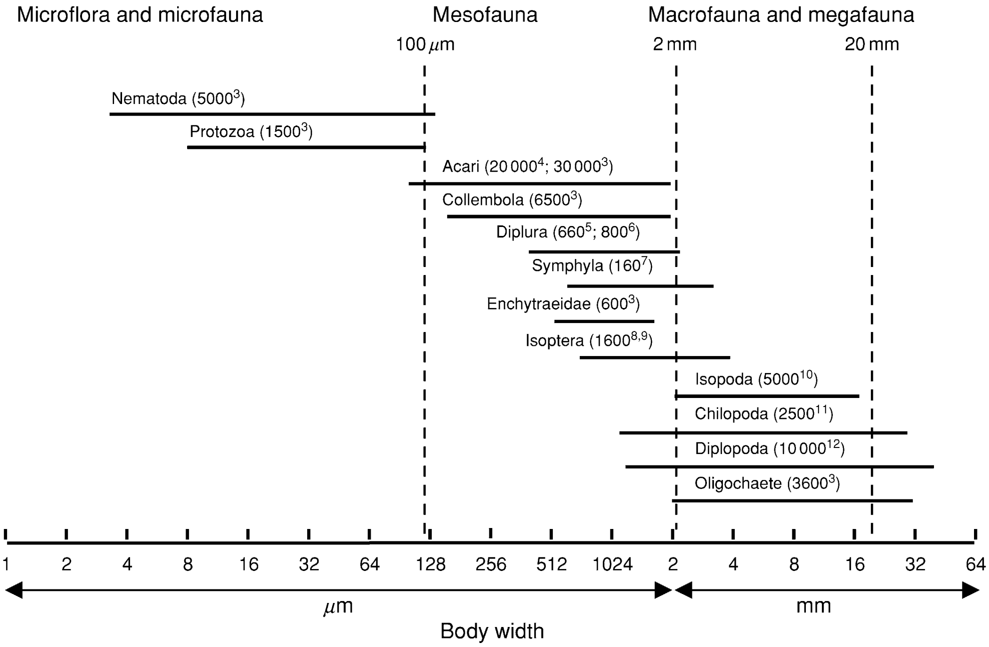


Figure 1 Invertebrate biodiversity represented by body width, with number of species described (as of 2003) for each group. (Source: Wall DH and Virginia RA (1997) The world beneath our feet: soil biodiversity and ecosystem functioning. In: Raven PH (ed.) *Nature and Human Society: The Quest for a Sustainable World*. Washington, DC: National Academy Press.)

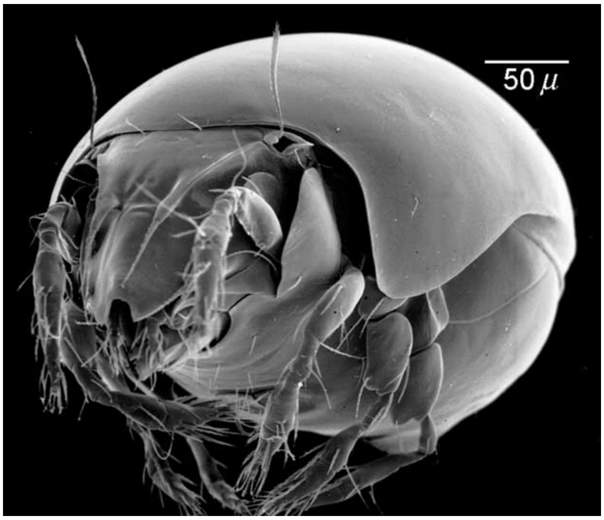


Figure 2 *Ceratozetes borealis*: a soil mite found in litter and soil in the Yukon and Alaska. (Courtesy of Valerie Behan-Pelletier.)

and up to a million arthropods that can inhabit a square meter of soil.

Even with the enormity of soil biodiversity, there is excitement about the progress and plans to overcome the difficulties of research on these and other species yet to be identified. Advancements in technology such as molecular technology, biochemical methods, and microscopic advancements (scanning electron microscope, image analysis) are enabling faster identification and/or characterization of organisms. Scientists, distributed globally with expertise in specific soil biotic groups, have accelerated our knowledge of the number of species, by sharing information, images and results via the internet, and creating web pages for the public. Today there are a very few taxonomists and systematists who specialize in the identification, natural history, and phylogenies of the soil biota. These specialists are also versed in the proper techniques for collecting, extracting, and identifying the various organisms, information that is needed for analysis of different ecosystems. As methods for the collection and extraction of soil fauna are primarily dependent on motility and size of the organism, there are no standard techniques for the collection or extraction of even a single group of fauna. This is comprehensible when one considers that collecting a handful of soil might suffice for extracting and identifying many species of microscopic nematodes, mites, protozoa, and microbes, but would not be adequate for earthworms. Technology, new educational tools, and training, and the excitement of conquering these challenges are attracting students to this new frontier of science.

Distribution

Soils, like air and water, are nonrenewable resources that have developed over geologic time. Both the type of soil and the evolutionary adaptation and dispersal mechanism of the organism to a particular location play a role in establishing the structure of a soil biotic community. Soil chemical and physical factors combine to make each soil a unique habitat. Chemical (e.g., pH, salinity, type and amount of organic matter and nutrients) and physical (e.g., texture (amount of sand, silt, and clay), pore size, moisture, temperature) variations, along with the composition of vegetation and geologic history all contribute to characterizing the diversity of species in soil. These components are all intertwined, and changes in any one factor can result in a new habitat for the biota. For example, an increase in organic matter generally has a ripple effect, increasing the soil water-holding capacity, which tends to create a darker soil and higher soil temperatures. Soil texture is also linked to microclimate. As clay content increases, soil fertility, soil moisture, and soil temperature can increase. Clay content is in turn largely dependent on the parent material (granite related to sandy soils versus limestone related to clayey soils) or landscape position (a floodplain is more fine-grained, a ridge top more coarse-grained). Thus, beneath our feet, there are thousands of soil habitats due to many combinations of factors such as pH, fertility, organic matter, soil moisture, climate, and geologic history.

Soil biota occupy remarkably different habitats within the soil: some species living only in decaying leaves and dead wood, or in the rich organic matter horizon, while others occur several centimeters to meters deeper in the mineral soil. For example, some species of earthworms occur in litter-organic horizons and others exist in deeper mineral soil. Species of organisms may be very different even at millimeter scales near a growing root than further away in bare ground. Increasing to larger scales, within a backyard or agricultural field, the diversity (species richness) might be similar. However, comparison of an agricultural field to a nearby natural area will generally show a very different community. At a global scale, primary factors affecting the biogeography of soil organisms are climate, the type (trees, grasses, forbs) and amount of vegetation (organic matter), and soil texture (composition of sand, silt, and clay). Earthworms are generally absent from dry ecosystems, illustrating an evolutionary adaptation to moist soils. Termite and ant diversity decreases with distance from the equator, but other invertebrates such as earthworms, enchytraeids, mites, and nematodes follow different patterns. As data are compiled on species and their natural history and combined with geographic information systems (GISs) maps of soil and vegetation, more reliable estimates of the global biogeography for soil microflora and fauna will become feasible.

Dispersal mechanisms of biota are a key to their distribution at local, regional, and global scales. Soil organisms are transported with soil: by humans transporting plants; or by cars, trucks, or other machinery; in rainwater, floods, seawater; carried externally and internally by animals such as birds, ungulates, and insects; in plant debris; and by wind. Protozoa, rotifers, tardigrades, and nematodes are aquatic biota, living in water films around soil particles, but as soil dries they enter a dormant state, anhydrobiosis (meaning life without water), reduce their body surface area, water content, and metabolism, and can live for years until they revive in water. While in anhydrobiosis, they can be wind-transported over hundreds of miles. Recent satellite photos from the year 2000 show large dust storms from Africa moving across the Atlantic ocean and transporting soil bacteria and fungi, including plant and animal pathogens into the North American soils. Collembola live in air pores of soil and in organic matter and were long thought to be confined to small geographic areas. However, they are now known to survive and move to other continents in seawater, because of a physiological advantage, their hydrophobic (unwettable) cuticles. This finding on a new means of dispersal will accelerate hypotheses on the global distribution patterns of Collembola. Other dispersal mechanisms are based in geologic time. For example, earthworm distribution in North America has been patterned largely by glaciation.

Soil Biodiversity and Ecosystem Functioning

Soil organisms moderate numerous ecosystem processes (e.g., decomposition, C and N transformation, hydrologic cycles) that are a major component of global cycles. Land-use change, climate change, atmospheric change, and an increase in invasive species are all components of 'global change.' Whether this global change results in the alteration of species composition within the soil community, and how this may impact ecosystem function, is an important question in the research of soil biology today. Little knowledge exists about the functional attributes of the majority of individual species, but there is considerable information on the consequence of larger species on soil processes. Macrofauna (earthworms and termites) species have a significant effect on soil structure and hydrology, and transfer and reconfigure organic matter, affecting nutrient mobility. The influence of these animal species on surface-soil plant litter can be measured at local and larger regional scales; but for smaller species scientists have used other approaches to determine their influence on ecosystem processes. Scientists lump species with similar morphologies and feeding sources into functional groups and measure the transformations of carbon and nitrogen through each functional group within the soil food web. A couple of examples illustrate this.

Perhaps some of the greatest diversity is involved in the decomposition (decay) component of the soil food web. In natural terrestrial ecosystems, the bulk of the world's dead plant material (leaves, stems, dead roots, grasses, also known as detritus) falls to the ground to be decomposed. The rate of decomposition in an ecosystem is affected by climate, the quantity and quality (type or chemical composition, generally C, N, and lignin) of the substrate or detritus (leaf versus wood versus animal, for example), and a 'detritivore' functional group in the food web that includes many species of generalist feeders from numerous phyla. If the substrate is more recalcitrant (has more lignin, higher C:N ratio, e.g., wood of trees) the decomposition will be primarily by fungi and fungal consumers. If the organic matter is more labile (lower lignin, lower C:N, e.g., some grasses) bacteria and their consumers are the major decomposers.

The breakdown of detritus is a succession of the species of invertebrates and fungi, actinomycetes, and bacteria, all with specific roles in breaking down complex organic substrates into the inorganic nutrients that are then available for plant growth. For example, some faunal species (e.g., arthropods) shred leaves into smaller particles; others channel through the soil, increasing the distribution of microbes and the organic matter while affecting soil porosity in the soil profile; other fauna ingest the pieces of organic matter, further exposing it to enzymes that aid the decay, and then pass it from their bodies in a different chemical form. Smaller microscopic animals, nematodes, mites, and Collembola prey on the decomposers, bacteria and fungi, maintaining them in an accelerated growth phase, speeding the transfer of nutrients to soil. As with other biotic food webs, predators are at the top of the soil food web. Interestingly, predators in soil food webs can be microflora (e.g., fungi that prey on nematodes) and microfauna (predaceous nematodes, or mites or tardigrades feeding on nematodes), as well as macrofauna. Determining predator–prey relationships of species contributes to biocontrol management of soil pathogens of humans and plants.

Global Change and Impacts on Soil Biodiversity

Land-use change is the major driver affecting soil biodiversity and future soil sustainability. Land-use changes (tillage, erosion, dams, change in plant species) affecting soil physical and chemical properties, soil structure, and the base of the soil food web (chemical composition of plants, amount of organic matter, oil, pollution, manure) have direct impacts on species composition. The conversion of a natural grassland or forest to a managed system for agriculture, pasture, urban, or industrial use, changes the determinants of soil biodiversity, the vegetation, soil structure, and microclimate. The disruption to the natural vegetation and the soil habitat with land-use change decouples the nutrients provided by the decomposition food web from plant uptake. The result is a loss in soil fertility provided by the original soil and its inhabitants. Additional fertilizer is required, and pesticides may be necessary, especially with more intensive agriculture, because the new soil food web generally has fewer predators, resulting in a change in or loss of biocontrol of plant pathogens. While tillage methods are beneficial in conserving carbon in soils and creating a food web that is more detritus-based, some lower level of herbicides and pesticides is generally used. There are examples from tropical wet and dry forests, grasslands, deserts, and other ecosystems showing that these land-use changes affect the total soil biota (macrofauna, microfauna, and microflora), generally reducing species diversity. Desertification resulting from land-use change has a considerable impact on soil processes, including soil carbon, soil structure, soil biota, and soil fertility.

Invasive species of plants and animals also influence changes in soils and biotic communities. The movement of plants, people, and machinery has not only increased the numbers of species invasions into soils globally, but has significantly increased impacts on soil biodiversity and some characteristic of the ecosystem (e.g., nutrient cycling, species richness and or abundance, plant factors, soil physical or chemical factors) and increased costs of eradication. An earthworm species introduced to New York has changed forest-floor litter quality and composition, soil chemistry, and water infiltration rates. European nations are examining ways to eradicate the Australian planarian flatworm, *Artioposthia triangulata*, a predator of earthworms. This invasive species has impacts aboveground (removing the food source for birds) and belowground (removing an animal species that influences organic matter transformation, soil hydrology, and structure). Invasive plant species with differing rooting depths and plant chemical composition can have repercussions for soil communities. Woody plant invasions into grasslands of the Great Plains of the USA have greater rooting depths (affecting soil carbon storage) and, at these depths, a more depauperate nematode community.

Increasing levels of CO_2 and climate change (temperature and rainfall patterns, and extreme, infrequent events such as droughts and rainfall) indirectly affect soil biodiversity through impacts on plant community composition and/or chemistry. Although data vary with ecosystem, modifications in the soil community and changes in belowground herbivory and decomposition pathways for soil food webs have been documented. Even in the Antarctic Dry Valleys where there are no visible plants, climate change has decreased soil nematode populations. Direct effects of elevated CO_2 are considered to be less important, because soils have high ambient levels of CO_2 owing to root and microbial respiration.

Soil food webs can be modeled to simulate the loss of species due to global change and to examine the effect, if any, on nutrient cycling and plant production. For example, in the short grass steppe of Colorado, USA, a model has been assembled using field data on the microbes and micro- and mesofauna. Other laboratory studies and results glean from the literature-generated information on the life cycles, generation time, biomass, and energetics for the dominant invertebrate species. The model accounted for transfers of carbon and nitrogen through the soil food web in 15 functional groups including plant symbionts (mycorrhizal fungi), herbivores (plant parasitic nematodes), decomposers (bacteria, saprophytic fungi), bacterial feeders (protozoa, nematodes), fungal feeders (nematodes, mites), omnivores, and predators (mites, nematodes). A global change scenario (elevated CO_2 and resulting change of plant species composition), which might indirectly result in the loss of biodiversity, was simulated for each of the 15 functional groups in the soil food web. For example, loss of bacteria and fungi led to extinctions of other groups in the soil food web. Removal of six of 15 groups impacted the abundance of other groups, and three of these groups (bacteria, saprophytic fungi, and root-feeding nematodes) caused up to a 10% change in two ecosystem processes, nitrogen mineralization, and plant primary production. This suggests that ecosystem functioning may be affected by the loss of soil biotic functional diversity, a result that needs further testing. Some scientists postulate that because of the high species diversity in soils, if species are lost, other species will perform the same function in the food web, i.e., there is a lot of redundancy of species in soils. Additional experimental results show that the effects of the change on soil community composition are idiosyncratic, or dependent on the ecosystem, vegetation, climate, and organisms examined. None of the studies so far have accounted for: (1) long-term loss of species; (2) infrequent climatic events; (3) highly variable changes in land use; (4) long-term effects of elevated CO_2 or increased nitrogen deposition, and other atmospheric changes; and (5) invasive species or combinations of these factors and soil disturbance.

Soil food web models may not reflect the complexity of individual species interactions and the underlying habitat but are critical for scientific analysis and forecasting for future scenarios of C and N cycling. The C and N cycling of the soil food web ties the soil to the aboveground ecosystem by relating the soil biota to soil carbon storage, nutrients available for plant growth and to atmospheric CO_2 and trace gas fluxes. Models, along with studies at species and functional group resolution, are currently being examined in experiments that manipulate biodiversity in field, laboratory (microcosms with generally less than 10 species), and in large soil experiments in plant growth facilities. Together these provide a theoretical and quantitative basis for integrating the soil biota into global ecosystem studies.

Sustaining Soil Biodiversity

Soil biodiversity is an important resource that provides ecosystem processes essential to the functioning of natural and global systems. Our understanding of the species, their interactions, and effect on processes occurring in the soil food web in natural systems are an important contribution to management of land, particularly agriculture. The link between aboveground and belowground diversity is strong, although occurring at different temporal scales for organisms, and changes affecting aboveground diversity and function are reflected in belowground ecosystems. An immediate effect is a decrease in the biological capacity of soils and a change in the regulation of interactions and processes. Knowledge on whether all or a few key taxa are important in this regulation of ecosystems processes is a high priority for planning for future sustainability.

Current priorities in the study of soil biodiversity include: (1) understanding the functions of rarely sampled fauna so that species estimates and global distributions are based on an accurate and current database; (2) determining which habitats are most vulnerable for soil biodiversity loss, e.g., where the 'hot spots' of biodiversity are, and which habitats and what time frames are most amenable to restoration; (3) synthesizing data and determining which invertebrate and microbial 'species' are key to ecosystem processes; (4) identifying gaps in knowledge of multispecies interactions and their influence on ecosystem functioning; (5) collaborating on long-term (more than 3 years) experiments to examine effects of global changes on aboveground–belowground biodiversity linkages and ecosystem functioning; and (6) determining, based on natural history information, which species are more likely to be invasive if introduced, and using this potentially to reduce spread and to identify threats to other species.

See also: **Archaea**; **Bacteria:** Plant Growth-Promoting; **Crusts:** Structural; **Fauna**; **Food–Web Interactions**; **Fungi**; **Microbial Processes:** Community Analysis; **Nematodes**; **Protozoa**

Further Reading

Coleman DC and Crossley DAJ (1996) *Fundamentals of Soil Ecology*. San Diego, CA: Academic Press.

Hendrix PF (ed.) (1995) *Earthworm Ecology and Biogeography in North America*. Boca Raton, FL: Lewis Publishers.

Hunt HW and Wall DH (2002) Modelling the effects of loss of soil biodiversity on ecosystem function. *Global Change Biology* 8: 33–50.

May RM (1988) How many species are there on earth? *Science* 241: 1441–1449.

Pace NR (2000) Microbial diversity and the biosphere. In: Raven PR and Williams T (eds) *Nature and Human Society: The Quest for a Sustainable World*, pp. 117–129. Washington, DC: National Academy of Sciences and National Research Council.

Schaefer M and Schauermann J (1990) The soil fauna of beech forests: comparison between a mull and a moder soil. *Pedobiologia* 34: 299–314.

Wall DH and Virginia RA (2000) The world beneath our feet: soil biodiversity and ecosystem functioning. In: Raven PR and Williams T (eds) *Nature and Human Society: The Quest for a Sustainable World*, pp. 225–241. Washington, DC: National Academy of Sciences and National Research Council.

Wall DH, Adams G, and Parsons AN (2001) Soil biodiversity. In: Chapin FS III, Sala OE, and Sannwald EH (eds) *Global Biodiversity in a Changing Environment: Scenarios for the 21st Century*, pp. 47–82. New York: Springer-Verlag.

Wall DH, Snelgrove PVR, and Covich AP (2001) Conservation priorities for soil and sediment invertebrates. In: Soule ME and Orians GH (eds) *Conservation Biology: Research Priorities for the Next Decade*, pp. 99–124. Washington, DC: Island Press.

BUFFERING CAPACITY

B R James, University of Maryland, College Park, MD, USA

Introduction

Buffering is the resistance of a system to change in response to a perturbation, and it is a key attribute of soils from the molecular level to the landscape scale. Chemical, physical, and biological processes may raise or lower solute concentrations (or activities) as intensive variables in soil solution, thereby temporarily disturbing dynamic equilibria or steady-state conditions between soil water and solid or gas phases of the soil. In response to such a perturbation, one or more processes may release solute to soil solution or remove it to restore wholly or partially the original concentration. The reservoir of such solutes in solution, in solid phases, or in the soil atmosphere for restoration of soil solution chemical composition is known as the 'capacity factor.' The measurable change in the capacity factor per unit change in the intensive variable is called the 'buffer index,' 'buffer intensity,' or, more commonly, 'buffer capacity.' Myriad chemical reactions (buffering mechanisms) govern such release or uptake of ions and molecules between soil solution and solid or gas phases, including cation and anion exchange, oxidation–reduction reactions, dissolution–precipitation processes, and metal–organic ligand complexation. Understanding and quantifying buffer capacities and buffering mechanisms at the colloid and molecular level of soils can aid in predicting the sensitivity and resilience of soil-water systems to anthropogenic and natural perturbations of ecosystems.

New ecological theory related to succession dynamics of disturbed ecosystems identifies that natural systems, including soils, are constantly and naturally recovering from regular and irregular disturbances of various severities. The processes and mechanisms of recovery determine the biodiversity and stability of ecosystems, and buffering in soils and natural waters is a key control of nutrient availability and pollutant bioavailability. Buffering reactions influence water quality and community regrowth, migration, and recruitment in disturbed patches on the landscape. The success of human efforts to restore disturbed ecosystems and soils is also determined in part by the scale and nature of soil-buffering processes, as related to natural ones that govern element transformations.

The Samovar Analogy for Buffering in Soils

The concepts of the intensive variable, capacity factor, buffer capacity, and buffering mechanisms for soils can be modeled and visualized by analogy with old-fashioned, Russian samovars used to make large volumes of tea (Figure 1). Each samovar comprises a large copper reservoir to boil water and a narrow glass tube on the outside of the samovar's tank to indicate how full the tank is. As boiling water is drawn out of the tank through the spigot, the indicator tube empties quickly and temporarily, but is refilled to the new tank level when the spigot is shut off (as indicated by the arrows between the tank and indicator tube in Figure 1). To decrease the volume of a samovar tank by a given fraction of its total capacity (e.g., from level 1 to level 2 in the indicator tube), a greater volume of water must be withdrawn from the larger samovar than from the smaller one.

Since the level of water in the indicator tube is proportional to how full each tank is, it is a parameter that is independent of the volume of the samovar. The height of the water column is analogous to an intensive variable in soil solution such as pH, a measure of how acidic an aqueous sample is, but not an indicator of its volume. The volume of the samovar represents the capacity factor (e.g., total soil acidity), and the

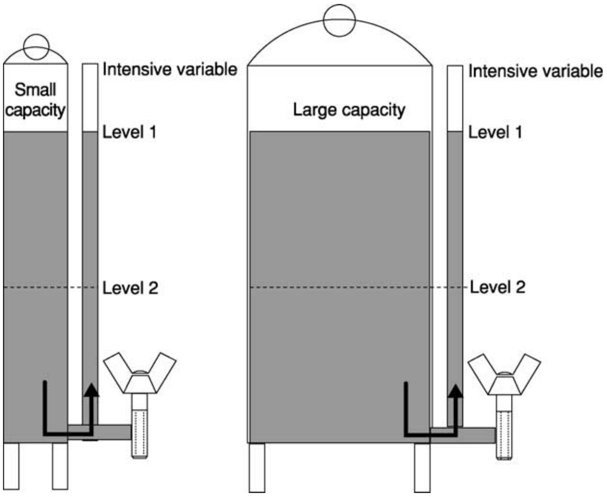


Figure 1 Samovar models representing small and large buffer capacities (tank sizes) in equilibrium with the same level of intensity variable (represented by the height of liquid in the narrow indicator tube outside the tank). (Adapted from Brady NC and Weil RR (2002) *The Nature and Properties of Soils*, 13th edn. Upper Saddle River, NJ: Prentice-Hall.)

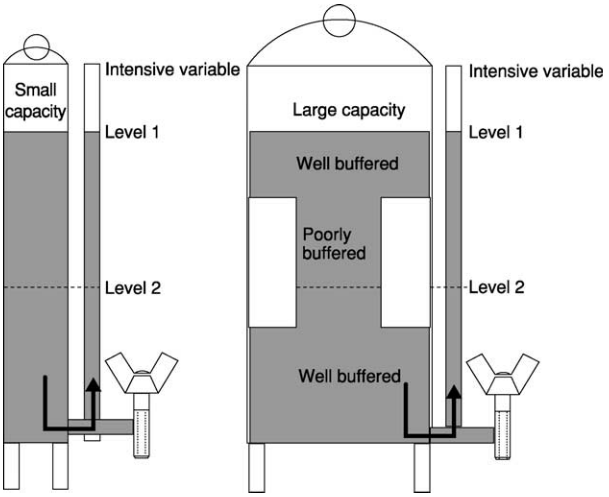


Figure 2 Samovar models representing small and large buffer capacities (tank sizes) in equilibrium with the same level of intensity variable (represented by the height of liquid in the narrow indicator tube outside the tank). The larger tank with the hourglass shape models a buffered system in which the buffer capacity (change in tank volume per unit change in the height of the liquid in the indicator tube) is not constant and varies during titration (drainage of the tank). (Adapted from Brady NC and Weil RR (2002) *The Nature and Properties of Soils*, 13th edn. Upper Saddle River, NJ: Prentice-Hall.)

change in the tank volume per unit change in the level of the indicator tube is the buffer capacity. The restoration of the water level in the indicator tube from the samovar after shutting off the spigot is due simply to the flow of water, i.e., the buffering mechanism in this model.

In this simple comparison of two samovars of different volumes, but with identical fractions of their capacities filled with water, as the boiling water is drained completely, the buffer capacity (change in tank volume per change in water level) remains constant. In complex soil colloid-water systems, however, buffer capacities do not usually remain constant as an intensive variable is changed, and the magnitude of the buffer capacity will change accordingly. This is modeled by a peculiar samovar with a narrow middle section and wide sections at the top and bottom (Figure 2). In this samovar model, the system would be well buffered at the beginning and end of draining the tank, but would be poorly buffered in the middle. In soils, measurements of intensive variables are often made routinely (e.g., pH, Ca concentration, partial pressure of O_2, and oxidation–reduction potential), and are used to indicate the energy state or 'ability of the system to do work' in affecting other coupled systems such as groundwater or plants rooted in the soil.

The size of the capacity factor and buffering mechanisms linking the intensive and capacity parameters are often not known with the same accuracy as are measurements of intensive variables in soils (i.e., the samovar tank is invisible). Therefore, investigations of the nature and scale of buffering processes in soils are needed to understand controls on the value of an intensive variable and what chemical and biological processes control its resistance to change.

The samovar analogy for soil buffering will be used to explain similarities and differences among several soil chemical processes responsible for buffering of soil acidity, oxidation–reduction status, and metal ion activities controlled by dissolution–precipitation and organic complexation by ligands. Soil acidity and oxidation–reduction status of soils are considered master variables that control the steady-state chemical composition of soils and the kinetics of approach to chemical equilibrium. As a result, predicting their resistance to change in response to disturbance of soil conditions is relevant to many environmental quality, plant nutrition, and human health impacts of soils. Dissolution and precipitation reactions are important in weathering, soil development, and leaching processes, as well as in controlling nutrient and pollutant solubility, speciation, bioavailability, and cellular absorption. Organic and inorganic ligand complexation, typically of cations, controls ion activity in equilibrium with soluble and insoluble forms of the element. Each of these processes is a coupling of an ion activity (intensive variable), a capacity factor (reserve of ion), and a buffering capacity and buffering mechanism that govern how the intensity and capacity factors interrelate in a soil system.

Buffering of Soil Acidity

The intensive variable for soil acidity is conceptually defined as the hydrogen ion activity of soil pore water or 'active acidity,' and is operationally defined by measuring pH in a soil-water or dilute salt suspension with a glass-reference electrode system, where pH is $-\log(H^+)$ and (H^+) is the hydrogen ion or proton activity in units of moles per liter. To neutralize the active acidity (e.g., raising pH from 5.0 to 7.0) in the soil solution of the top 15 cm of a hectare of loamy soil at field capacity moisture (−10 kPa water potential) containing 200 g montmorillonite and 30 g organic matter per kilogram of soil would require approximately 500 g of $CaCO_3$. This is just a handful of lime, compared with the 10 000 kg $CaCO_3$ ha^{-1} that might be needed to raise the whole soil pH to 7.0. The 20 000:1 ratio of lime needed to neutralize the capacity factor or 'reserve acidity' of the soil to that needed to neutralize the active acidity reflects the relative size of the metaphorical samovar tank and its indicator tube for soil acidity.

Soil properties that determine the buffer capacity for soil pH include initial pH, aluminosilicate clay content, and mineralogy; organic-matter content; and Al(III), Fe(III), and Mn(III,IV)(hydr)oxide contents. Soil pH buffer capacity is determined by adding increasing quantities of acid or base to a given mass of soil and measuring pH after allowing sufficient equilibration time for pH to stabilize. It is often designated:

$$\beta = \mathrm{d}C/\mathrm{dpH} \qquad [1]$$

where β is buffer index or capacity in units of moles of H^+ or OH^- added per kilogram of soil (dC) per unit pH change, dpH. The buffer capacity is quantified as the reciprocal of the slope of a linear relationship between measured pH plotted on the ordinate versus moles of H^+ or OH^- added per kilogram of soil on the abscissa.

Other, quicker indicator tests for reserve or total acidity as the capacity factor have been developed to estimate 'lime requirement' for acidic agricultural soils used to grow crops that are sensitive to acid soil conditions. Examples of these ways to estimate reserve acidity include chemical extractions of exchangeable and reactive Al^{3+}, measures of cation exchange capacity (CEC) and the fraction of exchange sites occupied by 'basic' cations (predominantly Ca^{2+}, Mg^{2+}, K^+, and Na^+), measurements of decreases in the pH of well-buffered solutions added to the soil sample, and equilibrations of a soil sample with a $BaCl_2$ solution buffered at pH 8.2, followed by titration to pH 5 to determine the quantity of acid that reacted with the buffer. Measures of free $CaCO_3$ content can be used to estimate the total alkalinity of a soil and its buffer capacity against acidification.

Diverse chemical reactions occurring on colloid surfaces of clay, oxides, and organic matter are responsible for pH buffering. They involve cation exchange reactions, dissolution and precipitation of sparingly soluble compounds, and surface charge changes in response to pH. In most soils, pH buffer capacity and buffering mechanisms are due to combinations of these reactions, and titrations and quick tests for total acidity do not distinguish between them. None the less, understanding the relative importance of these processes and their chemical basis allows predictions of the buffering behavior of diverse soils, and it allows extrapolation to unstudied soil systems and environments, particularly as related to ecological processes, soil contamination, and soil remediation.

Dissolution and Precipitation of $CaCO_3$ and $Al(OH)_3$

Soils containing free $CaCO_3$ have a pH between 7.0 and 9.5 (typically 8.0–8.5) as governed by the dissolution and precipitation of $CaCO_3$ in water containing partial pressures of CO_2 as high as 10 times that of atmospheric levels. Acidity produced by the hydration of dissolved CO_2 and by other acid-generating reactions reacts with $CaCO_3$ in accordance with:

$$CaCO_3 + H_3O^+ = Ca^{2+} + HCO_3^- + H_2O \qquad [2]$$

In the presence of high levels of soluble or exchangeable Mg^{2+} or Na^+, soluble $MgCO_3$ or Na_2CO_3 will form and allow the soil pH to rise to values higher than that controlled by $CaCO_3$–CO_2 equilibria (sometimes to pH >10). Addition of base or acid will be buffered by the lime in the soil, thereby maintaining its pH at approximately 8.2.

Under strongly acid soil conditions, Al^{3+} in soil solution and on cation exchange sites hydrolyzes and generates acidity, while the dissolution of $Al(OH)_3$ neutralizes it, leading to large β values at pHs <4.5:

$$Al^{3+} + 3H_2O = Al(OH)_3 + 3H^+ \qquad [3]$$

Soluble and exchangeable Al^{3+} is released from aluminosilicate clays upon weathering of Si-dominated tetrahedral sheets (releasing soluble monosilicic acid, $Si(OH)_4$), thereby exposing Al-dominated octahedral sheets to H^+ attack. As a result, the buffer capacity of strongly acid soils (pH <5) is dominated by Al chemistry.

Cation Exchange Reactions and pH Buffering

In the intermediate pH range between 4.5 and 6.5, cation exchange reactions on permanent aluminosilicate cation exchange sites, and on pH-dependent sites of organic matter and (hydr)oxides govern uptake and release of H^+ and OH^- as controls on the buffering-capacity size and mechanism for pH control. Permanent, negatively charged cation exchange sites on aluminosilicate clay minerals are due to isomorphous substitution of lower charge cations for Si^{4+} (e.g., Al^{3+}) in tetrahedral sheets and for Al^{3+} (e.g., Mg^{2+} and Fe^{2+}) in octahedral positions. Cation exchange sites created in this way during clay mineral formation are dominated by Ca^{2+} at and above pH 7 and become increasingly satisfied by H^+ and Al^{3+} at lower pH values. In this way, the so-called base saturation (an intensive variable reflecting the fullness of the acidity samovar) decreases as exchange sites are occupied by Al^{3+} and H^+.

Organic matter functional groups (principally carboxylic acids, represented by R–COOH, and phenolic acid groups, such as –OH on aromatic rings) are weak acids with association constants (pK_as) in the range of 2–7 for carboxylic acids and 7–10 for phenolic acid groups. These groups deprotonate at pHs >pK_a, and protonate at pHs <pK_a. When $pH = pK_a$, maximal pH buffering is observed. The complexity of humic and fulvic acid molecular aggregates and functional groups leads to overlapping pK_as that are not independent and which change as adjacent groups are protonated or deprotonated. In addition, conformational changes in the folding patterns of humic acid polymers as the suite of cations changes during titration result in changeable pK_as for the cation exchange sites responsible for pH buffering. As a result, titration curves of soil organic matter are often linear and do not reflect sharp inflection points characteristic of mono- or polyprotic simple organic acid titration curves. In addition to exchange reactions on organic acids between H^+, Al^{3+}, and Ca^{2+}, some precipitates, such as calcium oxalate, may exist in organic matter-rich horizons, such as in forest soils. These Ca–organic C precipitates may contribute to pH buffering via dissolution and precipitation reactions similar to those of $CaCO_3$, but at soil pHs <4.

pH-Dependent Surface Charge and pH Buffering

Oxide and hydroxide coatings of Fe(III), Mn(III,IV), and Al on sand, silt, and clay minerals are dominated by oxygen- and hydroxide-rich planes exposed to soil solution, and changes in pH protonate and deprotonate these functional groups, thereby creating variable or pH-dependent charge colloid surfaces that contribute to buffer capacity:

$$\begin{aligned} \mathrm{Me^{III}{-}OH_2^{+1}} + \mathrm{OH^-} &= \mathrm{Me^{III}{-}OH^0} + \mathrm{OH^-} \\ &= \mathrm{Me^{III}{-}O^{-1}} \qquad [4] \end{aligned}$$

where Me(III) represents a trivalent metal structurally associated with the mineral surface, and the charges on the surface become increasingly negative as pH is raised. Highly weathered soils that contain few aluminosilicate clay minerals or organic matter are often dominated by these variable-charge colloids, and they contribute to pH buffering through positive- and negative-charge creation.

The relative importance of $CaCO_3$ dissolution and precipitation, Al hydrolysis, cation exchange on permanently charged aluminosilicate minerals, organic-matter functional group protonation and deprotonation, and variation in negative and positive charges on variably charged oxide coatings depend on the relative abundance of these colloids in a given soil. In many soils, more than one of these soil chemical processes are responsible for pH buffering, and titrations of whole-soil materials reflect an integration of the several processes. Table 1 summarizes the relative magnitudes of the buffer capacities associated with these processes.

Table 1 Approximate buffer capacities of soil materials and horizons in the pH range 3.0–10.0

Soil constituent or horizon	*Buffering pH range*	*Buffering index (mmol kg^{-1} pH unit^{-1})*	*pH-buffering mechanism*
Aluminosilicate clays	3–10		
Smectite	3–10	178–333	Cation exchange; mineral dissolution/precipitation; Al hydrolysis
Vermiculite	3–10	333–444	Cation exchange; mineral dissolution/precipitation; Al hydrolysis
Illite	3–10	44–88	Cation exchange; mineral dissolution/precipitation; Al hydrolysis
Kaolinite	3–10	2–11	Cation exchange; mineral dissolution/precipitation; Al hydrolysis
Soil organic matter	5–8	360–444	Protonation–deprotonation of weak acid functional groups; conformational changes
Allophane and imogolite	3–10	44–111	Protonation–deprotonation of weak acid functional groups; conformational changes
Fe(III) and Al(III) (hydr)oxides	3–10	11–89	Protonation–deprotonation of weak acid functional groups; conformational changes
$CaCO_3$ and $MgCO_3$	>7	4444	Dissolution/precipitation
Forest floor organic horizons	3–7	180–360	Metal–ligand complexation/decomplexation; cation exchange; protonation/deprotonation; of weak acid functional groups
Mineral horizons of agricultural and forest soils	4–10	53	Cation exchange, mineral dissolution/precipitation, Al hydrolysis, protonation–deprotonation of weak acid functional groups, conformational changes, and metal–ligand complexation/decomplexation; principally associated with soil organic matter

Source: Bloom PR (2000) Soil pH and pH buffering. In: Sumner M (ed.) *Handbook of Soil Science*, pp. 333–352. Boca Raton, FL: CRC Press, with permission; McBride MB (1994) *Environmental Chemistry of Soils.* New York, Oxford University Press, with permission; James BR and Riha SJ (1986) pH buffering in forest soil organic horizons: relevance to acid precipitation. *Journal of Environmental Quality* 15: 229–234, with permission.

Buffering of Soil Oxidation–Reduction Status

The concept of 'electron activity' of soils refers to the thermodynamic tendency for electrons to be transferred from reductants (e.g., Fe(II)) to oxidants (e.g., O_2), and it is operationally defined by Pt electrode potentials versus a standard reference electrode. When the measured potential is corrected for the reference electrode potential relative to the hydrogen electrode ($E^0 = 0.0\,\mathrm{V}$), it is designated E_h. This voltage can be converted to pE, a parameter analogous to pH, by dividing E_h in volts by 0.059 (the Nernstian slope factor relating electrode voltage to electron activity). The E_h or pE is the intensity variable that is a measure of electron activity or 'electron pressure' in a soil system. Buffering of electron activity is called poise, and is analogous to proton buffering, except that the processes responsible for resistance to change in pE are due to electron transfer reactions, many of which are microbial and enzymatically catalyzed.

Heterotrophic microbial respiration in soils uses organic C as the electron source (reductant) and an array of electron acceptors as oxidants. Strict aerobes use dissolved O_2 as the required oxidant to derive energy from the oxidation of organic C, the most energy-efficient process for cellular respiration. In flooded soils in which all pores are filled with water, the diffusion rate for dissolved O_2 is 10 000 times slower than it is in air. As a result, microbial respiration may deplete available O_2 faster than it is replenished, thereby leading to the onset of anaerobic conditions, with a decrease in the measured pE. Under anaerobic conditions, facultative anaerobes and strict anaerobes use alternative electron acceptors, with decreasing metabolic efficiency in the order (depending on pH): NO_3^-, Mn(III,IV)(hydr)oxides, Fe(III)(hydr)oxides, SO_4^{2-}, CO_2, and H^+.

The quantities of each of these electron acceptors (the capacity factor for electron activity) in the soil will 'poise' the E_h at a given level as governed by the free energy of reaction associated with its reduction, as coupled to the oxidation of organic C. The expected pE and E_h values at which each of these couples would poise the soil are shown at pH 5 and 7 in Table 2. Since the molecular reaction mechanism for many reduction reactions involves the transfer of the electron with a proton to the oxidant (equivalent to the addition of a H atom), the overall reaction raises the soil pH in many cases. As a result, the higher the pH of a soil, the lower the E_h or pE at which a given reduction reaction is expected to occur and at which the soil system will be poised (i.e., at a lower (H^+), a greater (e^-), or 'electron pressure' is needed to effect the given reduction).

Table 2 Common reduction half-reactions that poise soils at designated pE values at pH 5 and 7; the log *K* value is the pE for the reaction at pH 0

Half-reaction		*Poising pE*	
(for 1 electron reduction)	*log* K	*pH 5*	*pH 7*
$1/4O_2 + e^- + H^+ = 1/2H_2O$	20.8	15.6	13.6
$1/5NO_3^- + e^- + 6/5H^+ = 1/10N_2 + 3/5H_2O$	21.1	14.3	11.9
$1/2Mn_3O_4 + e^- + 4H^+ = 3/2Mn^{2+} + 2H_2O$	30.7	16.7	8.7
$Fe(OH)_3 + e^- + 3H^+ = Fe^{2+} + 3H_2O$	15.8	4.8	−1.2
$1/8SO_4^{2-} + e^- + 5/4H^+ = 1/8H_2S + 1/2H_2O$	5.2	−1.0	−3.5
$1/8CO_2 + e^- + H^+ = 1/8CH_4 + 1/4H_2O$	2.9	−2.1	−4.1
$e^- + H^+ = 1/2H_2$	0	−5	−7
$1/4CO_2 + e^- + H^+ = 1/24C_6H_{12}O_6 + 1/4H_2O$	−0.21	−5.9	−7.9

Buffering of Ion Activities via Dissolution–Precipitation, Ion Exchange, and Ligand Complexation

The soil solution activities of ions other than H^+ and OH^- are intensity measurements that may be controlled by dissolution–precipitation, ion exchange, and ligand complexation reactions responsible for buffering in soils. Solubility, exchange, and complexation equilibria maintain ion activities in soil solutions as leaching, and plant and microbial uptake; and other biotic and abiotic processes occur and increase or decrease particular ion activities. The disturbance of these equilibria will induce restorative shifts in accordance with the LeChatelier principle, which states that the balance of products and reactants in a dynamic chemical equilibrium will shift in response to a perturbation in order to restore the original balance of reactants and products. For example, if sparingly soluble $PbCrO_4$ is present in a soil, Pb^{2+} and CrO_4^{2-} concentrations of approximately $5.3 \times 10^{-7}\,\mathrm{mol\,L^{-1}}$ will be maintained in soil solution as controlled by the solubility product, Ksp, of 2.8×10^{-13}. Any process that depletes the soluble Pb^{2+} or CrO_4^{2-} will induce the dissolution of a small amount of solid $PbCrO_4$ to restore the equilibrium activities of the Pb^{2+} and CrO_4^{2-}. Similarly, additions of these ions to the soil solution will induce precipitation of more $PbCrO_4$. In this way, the ion activities are buffered and maintained in soil solution, as shown in eqn [5]:

$$PbCrO_4 = Pb^{2+} + CrO_4^{2-} \qquad [5]$$

Similar reactions occur via cation and anion exchange on charged colloids to buffer ion activities, as discussed above for soil acidity buffering. The suite of exchangeable cations in soils is dominated by Ca^{2+}, Mg^{2+}, K^+, Na^{2+}, H^+, and Al^{3+}, depending on the pH of the soil, the mineralogy of the soil parent material, the aluminosilicate clay mineralogy, plant uptake, microbial processes, and human-induced changes in the soil conditions. Multiple equilibria are established in any soil, but all comprise a capacity factor, represented by the exchangeable cations, and an intensity factor. The capacity factor is sometimes called the 'quantity' term, and the quantity-to-intensity ratio (Q/I) is a measure of the nature of the cation or anion exchange buffering reactions responsible for maintaining ion activities in soil solution. Eqn [6] represents a Ca–K cation exchange reaction for such a Q/I relationship:

$$\text{Ex-Ca} + 2K^+(aq) = \text{Ex-}K_2 + Ca^{2+}(aq) \qquad [6]$$

in which Ex-Ca represents solid-phase, exchangeable Ca, $K^+(aq)$ is soluble K^+, Ex-K_2 is exchangeable K^+, and $Ca^{2+}(aq)$ is soluble Ca^{2+}. Any environmental conditions that change the activity ratio of K^+ to Ca^{2+} in soil solution (e.g., changing water content, preferential plant uptake of one cation over the other, precipitation of Ca^{2+}, or addition of Ca- or K-containing materials to the soil) will result in K-for-Ca or Ca-for-K exchange reactions that restore and buffer the ion activities in soil solution. The balance between the ratios of exchangeable Ca-to-K and soluble Ca-to-K in solution can be described by an equilibrium constant, a measure of the buffer capacity between exchangeable and soluble forms of the cations.

The activities of many ions in soil solution are also in equilibrium with complexed or chelated forms of the metals that are also soluble, but in forms in which the positive charge of the cation has been neutralized by negatively charged ligands. Any processes that increase or decrease the activity of the hexaquo form of the 'free' form of the cation, e.g., $Fe(H_2O)_6^{3+}$, that is in equilibrium with complexed forms, e.g., complexed with carboxylic acid groups, hydroxyl ions, or other Lewis bases, will induce a shift in the equilibrium to restore the balance of the complexed and free form of the ion. In this way, the 'free' form of the cation (its activity or intensity) is maintained through the equilibrium with the capacity factor, the complexed form. Similar to solubility products and cation exchange equilibria, a stability constant quantifies the relative thermodynamic stability of the complexed and free forms of a metal in a quantity–intensity relationship.

See also: **Acid Rain and Soil Acidification**; **Biodiversity**; **Environmental Monitoring**; **Eutrophication**; **Forest Soils**; **Microbial Processes:** Environmental Factors; **Nuclear Waste Disposal**; **Organic Soils**; **Pollutants:** Persistent Organic (POPs); **Pollution:** Groundwater; Industrial; **Remediation of Polluted Soils**; **Waste Disposal on Land:** Liquid; Municipal; **Wetlands, Naturally Occurring**

Further Reading

Bloom PR (2000) Soil pH and pH buffering. In: Sumner M (ed.) *Handbook of Soil Science*, pp. 333–352. Boca Raton, FL: CRC Press.

Brady NC and Weil RR (2002) *The Nature and Properties of Soils*, 13th edn. Upper Saddle River, NJ: Prentice-Hall.

James BR and Bartlett RJ (2002) Redox phenomena. In: Sumner M (ed.) *Handbook of Soil Science*, pp. 169–194. Boca Raton, FL: CRC Press.

James BR and Riha SJ (1986) pH buffering in forest soil organic horizons: relevance to acid precipitation. *Journal of Environmental Quality* 15: 229–234.

Lindsay WL (1979) *Chemical Equilibria in Soils*. New York: Wiley-Interscience.

Marschner H (1995) *Mineral Nutrition of Plants*, 2nd edn. London, UK: Academic Press.

McBride MB (1994) *Environmental Chemistry of Soils*. New York: Oxford University Press.

Reice SR (1994) Nonequilibrium determinants of biological community structure. *American Scientist* 82: 424–435.

Stumm W and Morgan JJ (1996) *Aquatic Chemistry*, 3rd edn. New York: Wiley-Interscience.

Bulk Density *See* **Porosity and Pore-Size Distribution**

C

CALCIUM AND MAGNESIUM IN SOILS

N S Bolan and P Loganathan, Massey University, Palmerston North, New Zealand
S Saggar, Landcare Research, Palmerston North, New Zealand

Introduction

Calcium (Ca) and magnesium (Mg) are the fifth and the eighth most abundant elements in the Earth's crust, respectively. While Ca is an important component of plant cell wall, Mg is the central component of chlorophyll. Both elements, regarded as secondary plant nutrients, undergo almost similar reactions in most soils and are not always considered explicitly in fertilizer-recommendation programs. One of the reasons for their exclusion in fertilizer programs is that Ca and Mg are added to soil as accessory elements in many fertilizers and liming materials. Increasing use of Ca-free ammonium phosphate fertilizers and reduced use of liming materials have resulted in increasing incidences of Ca deficiency in soils. Similarly, increasing incidences of Mg deficiency are attributed to the decreased use and reduced concentration of Mg in other major fertilizers (e.g., kainite). Further, accelerated soil acidification caused by modern agricultural practices has exacerbated Ca and Mg deficiency in soils. This article describes the inputs of Ca and Mg in soils, their plant and animal requirements, and modeling of the reactions of their compounds in soils.

Input to Soils

Calcium in soils is found mainly in minerals such as feldspar, calcite, dolomite, apatite, and hornblende. Calcium sulfate (gypsum) and calcium carbonate (calcite), which occur in arid and calcareous soils, respectively, control Ca concentration in these soils. Soils developed from calcite are generally alkaline in reaction. A high pH and the presence of Ca favor the formation of Ca humate complexes, which account for the dark color of these soils. The Ca content of soils depends on the type of parent materials and the extent of weathering. Although most soils contain 1.0–50 $\mathrm{g\,kg^{-1}}$ Ca, some of the calcareous soils contain more than 200 $\mathrm{g\,kg^{-1}}$ Ca.

Magnesium is a normal component of both igneous and sedimentary rocks, and of the soils developed from such rocks. Soils developed from basic rocks (diabase, basalts, limestone, and serpentine) generally contain high levels of Mg (2.7–28.6 $\mathrm{g\,kg^{-1}}$) and those developed on coastal sand and granite and sandstones contain low levels of Mg (0.1–3.4 $\mathrm{g\,kg^{-1}}$). In most soils, Mg is present in primary minerals such as biotite, serpentine, olivine, augite, and hornblende, and in secondary silicate clay minerals, chlorite, vermiculite, illite, and montmorillonite (Table 1), in organic matter as exchangeable cation, and also in soil solution. However, the majority of soil Mg is present in forms that are not readily available to the plant.

Calcium deficiency in soils can be overcome by adding Ca-containing compounds. Traditionally, superphosphates (single superphosphate, SSP; triple superphosphate, TSP) have been used as the major source of phosphorus, but they also supply Ca. The Ca in superphosphates is present in readily soluble gypsum ($CaSO_4 \cdot 2H_2O$ in SSP) and monocalcium phosphate ($Ca(H_2PO_4)_2$ in SSP and TSP) forms. The other two most commonly used Ca compounds are lime and gypsum. Lime is added mainly to overcome the problems associated with soil acidification; gypsum is used both as a sulfur (S) source and as an amendment to improve the physical conditions of soils.

A range of liming materials, which vary in their ability to neutralize the acidity, can supply Ca and Mg to soils. These include calcite ($CaCO_3$), burnt lime (CaO), slaked lime ($Ca(OH)_2$), dolomite ($CaMg(CO_3)_2$), and slag ($CaSiO_3$). The acid-neutralizing value of liming materials is expressed in terms of calcium carbonate equivalent (CCE), defined as the acid-neutralizing capacity of a liming material expressed as a weight percentage of pure $CaCO_3$. A neutralizing value of greater than 100 indicates greater efficiency of the material relative

Table 1 Calcium and magnesium minerals in soils

Mineral	*Chemical formula*	*Calcium content (g kg^{-1})*	*Magnesium content (g kg^{-1})*
Actinolite	$Ca(Mg,Fe)_3Si_4O_{12}$	40–70	100–160
Augite	$CaMg(SiO_3)_2$	90–120	45–100
Diopside	$CaMg(SiO_3)_2$	75–185	20–140
Hornblende	CaMg metasilicate	50–80	10–90
Gypsum	$CaSO_4 \cdot 2H_2O$	200–250	–
Calcite	$CaCO_3$	300–500	–
Fosterite	Mg_2SiO_4	–	320–350
Pyrope	$3MgO \cdot Al_2O_3 \cdot 3SiO_2$	–	60–130
Iolite	$H_2(Mg,Fe)_4Al_8Si_{10}O_{37}$	–	50–80
Enstatite	$MgSiO_3$	–	180–220
Serpentine	$H_4Mg_3Si_2O_9$	–	19–26
Talc	$H_2Mg_3Si_4O_{12}$	–	160–200
Phlogopite	$H_3Mg_3Al(SiO_4)_3$	–	130–180
Biotite	$(H,K)_2(Mg,Fe)_2Al_2Si_3O_{12}$	–	10–160
Clinochlore	$H_8(Mg,Fe)_5Al_2Si_3O_{18}$	–	100–120

Table 2 Calcium and magnesium fertilizers

Fertilizer	*Chemical formula*	*Solubility (g l^{-1})*	*Solubility product (pK$_{sp}$)*	*Calcium content (g kg^{-1})*	*Magnesium content (g kg^{-1})*
Monocalcium phosphate	$Ca(H_2PO_4)_2$	18	1.14	171	–
Dicalcium phosphate	$CaHPO_4$	0.14	6.6	294	–
FGD gypsum	$CaSO_4 \cdot 2H_2O$	–	–	232	–
Mined gypsum	$CaSO_4 \cdot 2H_2O$	–	–	387	–
Tricalcium phosphate	$Ca_3(PO_4)_2$	0.02	24.0	398	–
Hydroxyapatite	$Ca_{10}(PO_4)_6(OH)_2$	Insoluble	55.9	374	–
Carbonate apatite	$Ca_{10}(PO_4)_{6-x}(CO_3)_xF_2$	Insoluble	108.3	396	–
Fluorapatite	$Ca_{10}(PO_4)_6F_2$	Insoluble	110.2	217	–
Dolomite	$MgCO_3 \cdot CaCO_3$	0.038	17.09	258	100
Calcined dolomite	$MgO \cdot CaCO_3$	–	–	350	160
Hydrated dolomite	$MgO \cdot Ca(OH)_2$	–	–	–	170
Magnesite	$MgCO_3$	0.076	8.24	–	260
Brucite	$Mg(OH)_2$	0.091	11.41	–	360
Magnesia	MgO	0.150	–	–	560
Kieserite	$MgSO_4 \cdot H_2O$	710	–	–	160
Epsom salt	$MgSO_4 \cdot 7H_2O$	1270	–	–	90
Kainite	$MgSO_4 \cdot KCl \cdot 3H_2O$	–	–	–	70
Langbeinite	$2MgSO_4 \cdot K_2SO_4$	–	–	–	110
Fosterite	Mg_2SiO_4	0.067×10^{-3}	28.11	–	320–350

FGD, flue gas desulfurization.

to pure $CaCO_3$. The amount of liming material required to rectify soil acidity depends on the neutralizing value of the liming material and pH-buffering capacity of the soil. Recently the potential value of other Ca-containing compounds in overcoming the problems associated with soil acidification has been evaluated. Some of these materials include phosphate rocks, flue gas desulfurization (FGD) gypsum, fluidized bed boiler ash, fly ash, and lime-stabilized organic composts.

Magnesium deficiency in soils can be overcome by adding Mg fertilizers such as serpentine superphosphate, epsom salt, kieserite, dolomite, and calcined magnesite (magnesia) (Table 2). Epsom salt and kieserite are fast-release Mg sources, used both for soil and foliar applications. The other water-insoluble fertilizers are used as slow-release sources. Dolomite, which contains both Ca and Mg, is more effective in acid soils, because the Mg is brought into solution by the acid soil. Dolomite is the most widely used source of Mg, both as an ingredient of mixed fertilizers and as a separate amendment for liming. There will rarely be any need for additional Ca and Mg for any crop where continuous acidification of soils (e.g., legume-based pastures) requires a regular liming program where dolomite is the major liming material.

Reactions in Soils

Calcium and Mg are present in three major forms in soils: in solution, on exchangeable sites, and in minerals. This arbitrary division accounts for differences in their bioavailability in soils, which follows the pathway: solution > exchangeable > mineral. Unlike potassium (K^+) and ammonium (NH_4^+) ions, Ca^{2+} and Mg^{2+} ions do not get fixed in the interlayers of 2:1 silicate clay minerals. Only a small fraction of the total Ca and Mg is present in soil solution and, depending on the soil type, the majority of these elements are present in other forms. Plants absorb these two cations from soil solution, which is buffered by the readily exchangeable forms that, in turn, are slowly replenished by soil reserves containing slowly exchangeable and structural forms. There is equilibrium between the various forms of Ca and Mg that allows the release of these two from less-available forms to more-available forms (i.e., labile form).

Calcium and Mg^{2+} ions released from fertilizers undergo exchange reactions similar to K^+ ions. The adsorption sites of the soil mineral colloids are not very selective for Ca^{2+} and Mg^{2+} ions, though Ca^{2+} is slightly more preferred than Mg^{2+} because of the lower diameter of hydrated Ca^{2+} ion compared with hydrated Mg^{2+} ion. The adsorption of Ca^{2+} to organic colloids such as humic acids is, however, very specific. Thus in soils containing large amounts of Ca (calcareous soils), the humic acids are mainly present in the form of Ca humate. Ca^{2+} and Mg^{2+} ions adsorbed on to both mineral and organic colloids tend to equilibrate with these ions in solution. Most mineral soils contain sufficient concentration of Ca^{2+} in solution, and their exchangeable sites are well saturated with Ca^{2+} to meet crop demand adequately. However, in acid peat soils, the native Ca and Mg contents can be so low that plants suffer from their deficiency, requiring Ca- and Mg-containing fertilizers. In acid soils, most of the Ca and Mg would exist in soluble ionic forms. Most acid soils contain adequate Ca for most plants and only highly leached, low-CEC acid soils (sands, oxisols) may show Ca-deficiency symptoms. Plant root growth in highly acidic soils can be affected because of a high Al^{3+}-to-Ca^{2+} ratio in soil solution.

As with other cations, the requirements for Ca and Mg application to plants are based on the exchangeable soil Ca and Mg tests. Field calibrations of Ca soil tests are not available. However, field calibrations for a Mg soil test are available that enable the soil-testing service to predict the Mg status of the soils and to make necessary Mg fertilizer recommendations. A number of soil-test methods are used to predict Mg availability to plants, which include 1N ammonium acetate exchangeable Mg and percentage CEC saturated with Mg. The ability of these indices to predict the availability of Mg to plants varies depending on the relative concentration of Mg^{2+}, Ca^{2+}, and K^+ ions in soil solution. A simple guide that may be useful to overcome Ca and Mg deficiency in soils is to maintain an adequate saturation of the exchange sites. Critical Ca and Mg saturation levels are considered to be 30–50% and 5–10% of CEC, respectively. However, for perennial plants (e.g., forestry), soil tests based on dilute acid extractions have been found to predict Mg requirements better than the exchangeable-Mg soil test.

Leaching of nitrate (NO_3^-) sulfate (SO_4^{2-}) and chloride (Cl^-) ions induces the leaching of basic cations such as Ca^{2+} and Mg^{2+}. When an excessive amount of K fertilizer is added to soil, it is possible that some of the Ca^{2+} and Mg^{2+} ions retained on to the cation exchange sites will be replaced by K^+ ions, inducing the leaching of Ca and Mg. In addition, induced deficiency may occur in some soils as a result of nutrient imbalances. Soils that are heavily fertilized, particularly with materials lacking in Mg^{2+} or high in Ca^{2+}, K^+, and NH_4^+ can also induce Mg deficiency. By growing crops that require high levels of Mg throughout the growing season (e.g., tobacco, citrus, potato, cotton, and soybeans), Mg deficiency in soils may be exacerbated or induced.

Modeling the Dissolution Reactions of Calcium and Magnesium Compounds in Soils

Most of the modeling work on Ca and Mg compounds in agricultural soils is concentrated on their use as liming materials rather than essential plant nutrients. Liming materials are used to overcome Ca and Mg deficiency, and the problems associated with soil acidification. They produce alkaline hydroxyl ions (OH^-) that neutralize the acid H^+ ions, thereby decreasing the activity of Al^{3+} and Mn^{2+}, and increasing concentration of basic cations such as Ca^{2+} and Mg^{2+}. The amount of liming material required to correct soil acidity depends on a number of factors, which include the neutralizing value of the liming material, the pH-buffering capacity of the soil, the form and amount of fertilizer used, and the production capacity (e.g., stocking rate in grazed pastures) of the farm.

Based on the concept of equal-diameter reduction, a mathematical model has been developed to predict the rate of lime particle dissolution. The equal-diameter reduction hypothesis assumes that (1) the initial mass of limestone is present as spheres of uniform size, density and composition, and (2) the rate of loss of mass is directly proportional to the instantaneous surface area of the spheres. This suggests that as the surface area decreases with dissolution, the rate of dissolution per unit surface area remains constant, i.e., the rate of dissolution is proportional to the surface area, as represented in the following equation:

$$\mathrm{d}m/\mathrm{d}t = -kS_\mathrm{a} \qquad [1]$$

where m is the mass that has dissolved from the surface after time t, k is the dissolution rate constant, and S_a is the total area of the spheres.

The above equation can be solved for conditions required by model of lime particles of increasing complexity of shape and size. For a simple system of particles of spherical size with equal size, the equation can be simplified to:

$$M/M_\mathrm{o} = (1 - kt/rD_\mathrm{o})^3 \qquad [2]$$

where M_o is the initial mass, M is the mass at time t, r is the density, and D_o is the initial radius of the spherical particle.

Commonly known as the 'cube root equation,' this equation is widely used to predict the dissolution of other insoluble fertilizer materials such as elemental sulfur and Mg fertilizers. A conceptual model for the dissolution of lime in soils has been developed incorporating the following four chemical reaction systems in soils: (1) dissolution of lime in stagnant aqueous system; (2) cation exchange reactions with Ca^{2+} and exchangeable acidity; (3) leaching and accumulation of dissolved components; and (4) the carbonate equilibrium system.

The cube root equation (eqn 2) to predict the dissolution of a range of Mg fertilizers has indicated that, within each size class, the changing particle surface area controls the rate of dissolution of dolomite. The specific dissolution rates increase with decreases in particle size.

Plant Requirements and Deficiency Symptoms

Ca is an important constituent of the plant cell wall and is involved in maintaining the turgidity of plants. Ca is required for cell elongation and cell division; it activates enzymes, particularly those that are membrane-bound, and is important in membrane permeability and the maintenance of cell integrity. Therefore, Ca provides protection against drought, salinity, mechanical stress, and toxic elements. Low Ca levels in storage organs induce high membrane permeability and allow solute diffusion in these tissues, an important mechanism for accumulating large amount of sugars from phloem in fruits and other plant storage organs. Ca moves very slowly in plants and is not transported from older leaves to younger leaves. Therefore Ca-deficiency symptoms are often noticed first in the growing tips, which require a continuous supply of Ca.

To a large extent, the Ca level in plants is genetically controlled and is little affected by Ca fertilization in the root medium, provided Ca availability is adequate for normal plant growth. Calcium deficiency is characterized by a reduction in growth of meristematic tissues, where the affected tissue becomes soft due to the dissolution of the cell walls. Calcium deficiency is rarely seen in field crops but is often observed in fruit crops such as apples. Of all the mineral elements, Ca has the greatest impact on the postharvest quality of pipfruit. In apple, the disease is called 'bitter pit,' and in tomato it is known as 'blossom end rot.'

Mg is the central unit and the only metal ion of chlorophyll in plant leaves, and it cannot be substituted by other metal ions. An insufficient supply of Mg reduces chlorophyll formation, which is likely to affect the photosynthetic ability of the plants. Because Mg is a structural element of chlorophyll, it is assigned a dominant role in the life of the plant. But it is not only on this account that Mg is indispensable to plants; its capacity for forming complexes with water can be of even greater significance, since in this way Mg has a controlling action on the swelling of plasma. Magnesium is involved in the production of starch during photosynthesis and plays a major role in the functions of many enzymes in plants.

Mg-deficiency symptoms generally show up in the leaves, since the leaf is the primary plant organ in which assimilation of photosynthate occurs. As a rule, Mg deficiency leads to chlorosis of the leaves, the formation of pale yellow spots and streaks that result from local failure in the formation of green leaf pigment. The deficiency is initially characterized by an interveinal chlorosis, although, in acute stages, the leaf may be generally deficient in both green and yellow pigments (i.e., variegated), and there may be necrosis in the areas of the leaf first affected by the deficiency. A deficiency of Mg also induces the formation of anthocyanins in some plant species such as cotton. One of the most important Mg-deficiency diseases, commonly referred to as 'sand drown,' was identified in tobacco leaves with less than 2.5 g kg^{-1} Mg. Similarly, pastures with less than 2.0 g kg^{-1} Mg lead to 'grass tetany' in grazing animals.

Unlike Ca, Mg is readily mobile in plants; it moves from older to younger leaves under Mg deficiency, and the deficiency symptoms therefore show first on the older leaves of the plant. Plants differ markedly in their response to a Mg deficiency in soil. In general, buckwheat is sensitive, corn intermediate, and small grains, grasses, and clover are only slightly responsive to Mg fertilization.

Plant Uptake

Plants take up Ca and Mg as their respective cations (Ca^{2+} and Mg^{2+}). For plants to utilize Ca and Mg, these ions need first to move to the root surface, followed by uptake by the roots. As the concentration of Ca^{2+} in soil solution is generally high (0.25–12.5 $mmol\,l^{-1}$), Ca^{2+} moves in soil solution predominantly by mass flow along with water, whereas Mg^{2+} ions move in soil solution by both mass flow and diffusion, depending on the concentration in soil solution (0.125–8.3 $mmol\,l^{-1}$) and transpiration rate. Magnesium uptake increases with increasing transpiration rate, especially when the concentration of Mg in soil solution is high, which is attributed to increased mass flow. However, the rate of ion movement in soil solution by both mass flow and diffusion depends largely on soil moisture content.

Once Ca^{2+} and Mg^{2+} ions reach the root surface, these ions are absorbed by both passive and active processes. Passive absorption occurs at high concentration along electrochemical potential gradients through the apoplasmic pathway, where the ions move freely through the free spaces in the cell walls and into the epidermis. Active absorption occurs against electrochemical potential gradients through the symplasmic pathway, and this process requires respiration energy and involves carrier molecules. Ca^{2+}, Mg^{2+}, and K^+ ions compete with each other for absorption when passive absorption occurs at high concentration. However, when active absorption occurs at low concentration, competition between these ions is unlikely to occur, because different carrier molecules are involved in the absorption of these ions.

Interactions with Other Nutrients

Ca-deficiency symptoms are rarely seen in the field because of low levels of Ca required for plant growth functions. But most of the Ca in plants and soils acts as an excluder or detoxifier of other elements such as Al and heavy metals that might otherwise be toxic. Nutritional problems related to Ca and Mg are mainly caused by impaired translocation of, or antagonism in, the uptake of NH_4^+, K^+, and Al^{3+} ions rather than by a simple Ca- or Mg-deficiency. For example, excessive use of K through fertilizer and effluent application has been shown to result in Ca and Mg deficiency in pasture.

While the addition of Mg-free liming materials (e.g., calcite) increases the supply of Ca, it has the opposite effect on the availability of Mg. Lime-induced reductions in tissue Mg level and Mg uptake by plants result from an increase in Mg adsorption due to an increase in pH-dependent adsorption sites in soils containing variable-charge components. Liming not only creates new, amorphous Al polymers, but also changes the charge character of their surfaces. More negative-charge sites are formed on such variable surface-charge materials, and cation adsorption is favored. Mg is favored over other cations for continued adsorption by such materials because of the presence of a Mg sink. However, excessive addition of Ca-containing materials such as calcite and gypsum has been shown to increase the leaching potential of Mg, mainly because the Ca^{2+} added through lime exchanges with Mg^{2+} on the soil surfaces, leading to the leaching of Mg^{2+} in the soil solution.

Animal Requirements

Calcium, as one of the constituents of hydroxyapatite mineral, forms the matrix of bone and teeth and is involved in nerve function, contraction of muscles, and blood clotting in animals. Calcium-binding proteins such as calmodulin play a central role in cellular regulation. A decrease in blood Ca levels in recently calved dairy cattle causes a disorder known as ‘milk fever’ (i.e., hypocalcemia), which is characterized by restlessness, muscle tremors, and sometimes coma. Animals with milk fever are treated with calcium borogluconate, administrated subcutaneously or intravenously. The ratio of the different basic cations in herbage influences animals’ relative absorption of these cations. For example, increasing intake of K has been identified as the major cause of the decrease in the absorption of Ca in grazing animals that results in milk fever.

The supply of Ca in animals is monitored from the dietary cation–anion difference (DCAD) values in the pasture:

$$\mathrm{DCAD} = \left(\left[\mathrm{Na}^+\right] + \left[\mathrm{K}^+\right]\right) - \left(\left[\mathrm{Cl}^-\right] + \left[\mathrm{SO}_4^{2-}\right]\right) \qquad [3]$$

where concentrations are in milliequivalents per kilogram of dry matter. Dietary cation–anion balance is important in animals to maintain systemic acid–base balance and osmotic pressure in order both to protect the integrity of cells and membranes and to optimize biochemical and physiological processes. When herbage with excess cations over anions (positive DCAD) is fed to animals, the concentration of alkali ions, such as bicarbonate, increases in body fluids, resulting in alkalosis. Conversely, when feed with a surplus of anions (negative DCAD) is ingested, then the concentration of acidic hydrogen ions increases and metabolic acidosis occurs. There have been conflicting reports on the optimum levels of DCAD values required for dairy cattle. It has, however, been shown that diets with high DCAD values tend to increase the incidence of milk fever, and the supplementation of precalving rations with anionic salts (low DCAD values) reduces the incidence of milk fever.

Magnesium plays an important role in the enzymatic metabolism of carbohydrates, lipids, proteins, and nucleic acids. Generally, Mg is an activator for the numerous enzymes such as phosphatases and the enzyme-catalyzing reactions involving adenosine triphosphate (ATP), which split and transfer phosphate groups. It is also involved in nerve conduction and muscular contraction. Deficiency of Mg in blood plasma causes a disorder known as 'hypomagnesemia' (grass tetany or staggers), which usually occurs in dairy cows in the early part of lactation. Mg-deficiency in animals can be overcome by regular use of Mg salts as a drench or water-trough treatment, as a lick, or in pasture after foliar application. Increasing intake of K has been identified as the major cause for decreased absorption of Mg in animals that results in tetany and convulsion. Irrigation of pasture with dairy-shed effluents rich in K has been shown to increase the incidence of Mg-deficiency, which leads to grass staggers.

In pasture and fodder crops, the grass staggers index (GSI) (eqn [4]) is used to predict the chances for the occurrence of Mg deficiency:

$$\mathrm{GSI} = [\mathrm{K}^+]/([\mathrm{Ca}^{2+}] + [\mathrm{Mg}^{2+}]) \qquad [4]$$

where concentrations are in milliequivalents per kilogram of dry matter. A GSI value of greater than 2.2 has been suggested to enhance the risk of grass staggers, a condition generally linked with animal serum Mg levels less than 1.0–1.5 mg per 100 ml, compared with normal levels of 1.7–3.0 mg per 100 ml. GSI values increase with increasing levels of K addition. Application of Mg fertilizers such as epsom salt is likely to decrease GSI values, mainly due to an increase in the concentration of Mg.

See also: **pH**

Further Reading

Adams F (1984) *Soil Acidity and Liming*. Madison, WI: Soil Science Society of America Publishing.

Bangerth F (1979) Calcium-related physiological disorders of plants. *Annual Review of Phytopathology* 17: 97–122.

Barber SA (1984) *Soil Nutrient Bioavailability – A Mechanistic Approach*. New York: John Wiley.

Christiansen MN and Foy CD (1979) Fate and function of calcium in tissue. *Communications in Soil Science and Plant Analysis* 10: 427–442.

Foy CD (1992) Soil chemical factors limiting plant root growth. In: Hatfield JL and Stewart BA (eds) *Advances in Soil Science*. Berlin, Germany: Springer-Verlag.

Geogievskii VI, Annenkov BN, and Samokhin VT (1981) *Mineral Nutrition of Animals*. London, UK: Butterworth.

Huttl RF and Schaaf W (eds) (1997) *Magnesium Deficiency in Forest Ecosystem*. London, UK: Kluwer Academic.

Mayland HF and Wilkinson SR (1989) Soil factors affecting magnesium availability in plant–animal systems: a review. *Journal of Animal Science* 67: 3437–3444.

McLaughlin SB and Winner R (1999) Tansley review no. 104 – Calcium physiology and terrestrial ecosystem processes. *New Phytologist* 142: 373–417.

Mengel K and Kirkby EA (1982) *Principles of Plant Nutrition*. Worblaufen-Bern, Switzerland: International Potash Institute.

Peverill KI, Sparrow LA, and Reuter DI (eds) (1999) *Soil Analysis – An Interpretation Manual*. Collingwood, Australia: CSIRO Publishing.

Saggar S and Bolan NS (2003) Secondary nutrients. In: Benbi DK and Nieder R (eds) *Handbook of Processes and Modeling in the Soil–Plant System*, pp. 561–588. New York: The Howarth Press.

Swartzenruber D and Barber SA (1965) Dissolution of limestone particles in soil. *Soil Science* 100: 287–291.

Terblanche JH and Wooldridge LG (1979) The redistribution and immobilisation of calcium in apple trees with special reference to bitter pit. *Communications in Soil Science and Plant Analysis* 10: 195–215.

Tinker PB and Lauchil A (eds) (1984) *Advances in Plant Nutrition*. New York: Praeger Publishers.

Wilkinson SR, Welch RM, Mayland HF, and Grunes DI (1990) Magnesium in plants – uptake, distribution, function, and utilization by man and animals. *Metal Ions in Biological Systems* 26: 33–56.

CAPILLARITY

D Or, University of Connecticut, Storrs, CT, USA
M Tuller, University of Idaho, Moscow, ID, USA

Introduction

The coexistence of gaseous, liquid, and solid phases in soil pores gives rise to a variety of interfacial phenomena that, for example, lead to spreading of liquid droplets on solid surfaces, liquid rising in capillaries and soil pores, or the entrapment of liquid in crevices. These phenomena, partially attributed to capillarity, determine retention and movement of water and solutes through soils. Hence they are of great importance in a variety of environmental and agricultural problems.

Liquid Properties

The phenomenon of capillarity in porous media results from two opposing forces: liquid adhesion to solid surfaces, which tends to spread the liquid; and the cohesive surface tension force of liquids, which acts to reduce liquid–gas interfacial area. The resulting liquid–gas interface configuration under equilibrium reflects a balance between these forces. The phenomenon of capillarity is thus dependent on solid and liquid interfacial properties such as surface tension, contact angle, and solid surface roughness and geometry.

Surface Tension

At the interface between water and solids or other fluids (e.g., air), water molecules are exposed to different forces than are molecules within the bulk fluid. For example, water molecules in the bulk liquid are subjected to uniform cohesive forces whereby hydrogen bonds are formed with neighboring molecules on all sides. In contrast, molecules at the air–water interface experience net attraction into the liquid because of lower density of water molecules on the air side of the interface, with most hydrogen bonds formed at the liquid side. The result is a membrane-like water surface that has a tendency to contract and reduce the amount of its excess surface energy. The surface tension reflects the amount of interfacial energy per unit area, or the energy required to bring molecules from the bulk liquid to increase the surface (it is also useful to express surface tension as force per unit length of interface). Different liquids vary in their surface tension σ (Table 1).

Surface tension also depends on temperature, usually decreasing linearly as the temperature rises. Thermal expansion reduces the density of the liquid and therefore also reduces the cohesive forces at the surface as well as inside the liquid phase.

Soluble substances can increase or decrease surface tension. If the affinity of the solute molecules or ions to water molecules is greater than the affinity of the water molecules to one another, then the solute tends to be drawn into the solution and to cause an increase in the surface tension. This is the effect of electrolytic solutes. For example, a 1% NaCl concentration increases the surface tension of an aqueous solution by $0.17\,\mathrm{mN\,m^{-1}}$ at 20°C. If, on the other hand, the cohesive attraction between water molecules is greater than their attraction to the solute molecules, then the latter tend to be relegated toward the surface, reducing its tension. That is the effect of many organic solutes, particularly detergents.

Contact Angle

When a liquid drop is placed on a solid surface, the angle formed between the solid–liquid (SL) interface

Table 1 Liquid–vapor interfacial tensions for various liquids

Liquid	*Temperature (°C)*	*Surface tension ($mN\,m^{-1}$)*
Water	20	72.94
	25	72.13
Methylene iodide	20	67.00
Glycerin	24	62.6
Ethylene glycol	25	47.3
Dimethyl sulfoxide	20	43.54
Propylene carbonate	20	41.1
1-Methyl naphthalene	20	38.7
Dimethyl aniline	20	36.56
Benzene	20	28.88
Toluene	20	28.52
Chloroform	25	26.67
Propionic acid	20	26.69
Butyric acid	20	26.51
Carbon tetrachloride	25	26.43
Butyl acetate	20	25.09
Diethylene glycol	20	30.9
Nonane	20	22.85
Methanol	20	22.50
Ethanol	20	22.39
Octane	20	21.62
Heptane	20	20.14
Ether	25	20.14
Perfluoromethylcyclohexane	20	15.70
Perfluoroheptane	20	13.19
Hydrogen sulfide	20	12.3
Perfluoropentane	20	9.89

Reproduced from Adamson AW (1990) *Physical Chemistry of Surfaces*, 5th edn. New York: John Wiley.

and the liquid–gas (LG) interface (Figure 1) is referred to as the equilibrium (or static) contact angle (γ). Two equivalent approaches are commonly used to describe the equilibrium contact angle on smooth and chemically homogeneous planar surfaces: (1) a force balance approach, and (2) an interfacial, free-energy minimization. The force balance formulation considers interfacial tensions (σ_{ij}) as forces per unit length; hence the force balance at the contact line of a drop resting on a solid surface under equilibrium requires the vector sum of the forces acting to spread the drop (outward) to be equal to opposing cohesion and viscous forces. The free-energy minimization approach regards interfacial tension as energy per unit area, and calculates changes in surface free energy (ΔF) due to infinitesimal displacement (ΔA):

$$\Delta F = \Delta A(\sigma_{SL} - \sigma_{GS}) + \Delta A \cos\gamma\ \sigma_{LG} \qquad [1]$$

The result is identical whether considering the minimization of free energy, with $\Delta F/\Delta A = 0$, or taking a balance of forces tangential to the solid surface; both cases yield the Young equation:

$$\sigma_{LG} \cos\gamma + \sigma_{SL} - \sigma_{GS} = 0 \qquad [2]$$

with L, G, and S indicating liquid, gas, and solid, respectively, and σ_{ij} the respective interfacial surface tensions. The equilibrium contact angle is therefore:

$$\cos\gamma = \frac{\sigma_{GS} - \sigma_{SL}}{\sigma_{LG}} \qquad [3]$$

Liquids that are attracted to solid surfaces (adhesion) more strongly than to other liquid molecules (cohesion) exhibit a small contact angle, and the solid is said to be 'wettable' by the liquid (Figure 1a). Conversely, when the cohesive force of the liquid is larger than the adhesive force, the liquid 'repels' the solid and γ is large (Figure 1b).

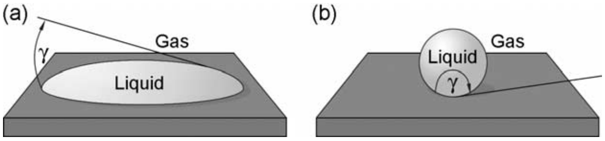


Figure 1 Liquid–solid–gas contact angles: (a) hydrophilic surface ($\gamma < 90°$) where liquid wets the surface; (b) hydrophobic surface ($\gamma > 90°$) where liquid 'repels' the surface.

Figure 2 illustrates differences in wettability of a silt soil. In Figure 2b a water droplet is resting on a soil surface that was treated to become water-repellent ($\gamma = 70°$). In contrast, Figure 2a depicts a wettable soil surface. In general, the contact angle of water on clean glass, and presumably on most soil minerals, is small, and for mathematical convenience is often taken as $\gamma = 0°$.

Curved Surfaces and Capillarity

When the forces that spread the liquid (adhesion and spreading on solids, or gas pressure within a bubble) are in balance with surface tension that tends to minimize interfacial area, the resulting liquid–gas interface is often curved. In porous media, the liquid–gas interface shape reflects the need to form a particular contact angle with solid(s) on the one hand, and the tendency to minimize interfacial area within the pore. A pressure difference forms across the curved interface, where the pressure at the concave side of the interface is greater by an amount that is dependent on the radius of curvature and the surface tension of the fluid. For a hemispherical liquid–gas interface having radius of curvature R, the pressure difference is given by the Young–Laplace equation:

$$\Delta P = \frac{2\sigma}{R} \qquad [4]$$

where $\Delta P = P_L - P_G$ when the interface curves into the gas (e.g., water droplet in air); or $\Delta P = P_G - P_L$ when the interface curves into the liquid (e.g., air bubble in water, water in a small glass tube). In many instances a bubble may not be spherical, or an element of liquid may be confined by irregular solid surfaces, resulting in two or more different radii of curvature such as water held in pendular rings between two spherical solid particles (Figure 3). The Young–Laplace equation for this case is given by:

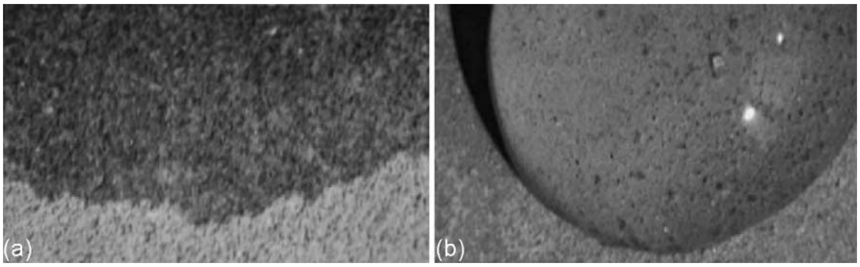


Figure 2 (a) Wettable silt surface ($\gamma \sim 0°$); (b) treated water-repellent silt soil surface ($\gamma = 70°$). (Reproduced from Bachmann J, Elliesb A, and Hartgea KH (2000) Development and application of a new sessile drop contact angle method to assess soil water repellency. *Journal of Hydrology* 231: 66–75.)

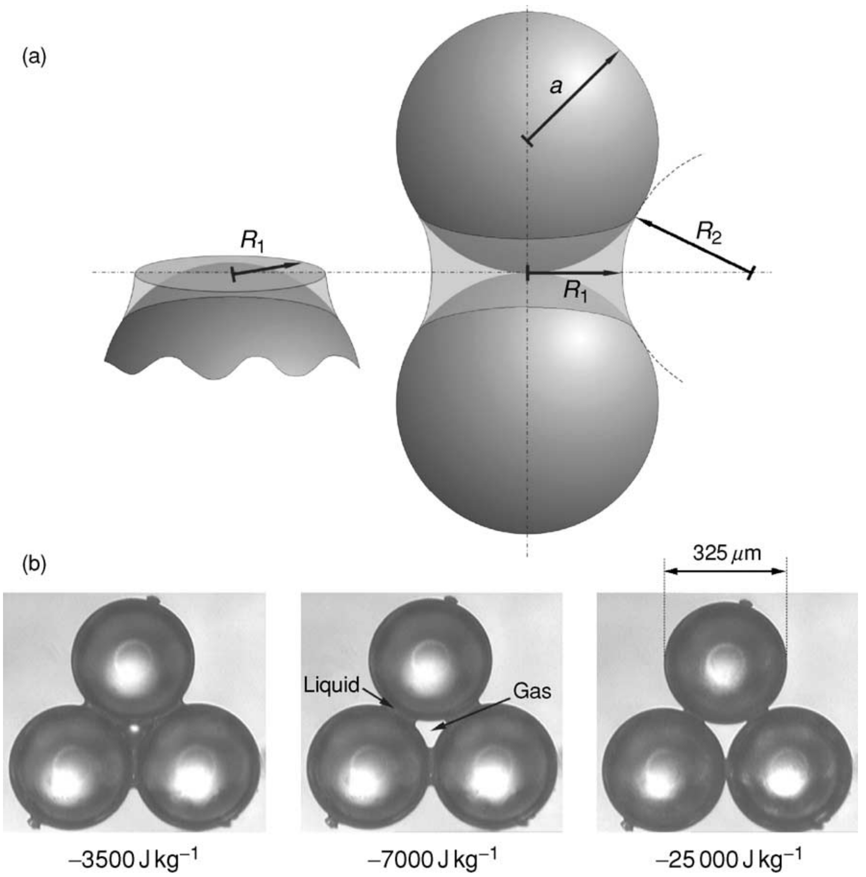


Figure 3 (a) Radii of curvature and shape of water held in pendular space between two spherical grains (note that for two equal spheres with radius *a*, the relationship between R_2 and R_1 is given as: $R_2 = R_1^2[2(a - R_1)]$. (b) Water menisci held between three spherical glass beads at different capillary pressures.

$$\Delta P = \sigma\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \qquad [5]$$

Note that this equation reduces to eqn [4] for spherical geometry with $R_1 = R_2$, and the sign of R is negative for convex interfaces ($R_2 < 0$) and positive for concave interfaces ($R_1 > 0$). For an interface forming in a linear crevice or within a fracture, $R_2 \to \infty$, hence eqn [5] reduces to: $\Delta P = \sigma/R_1$, where R_1 equals half the fracture aperture.

The Capillary Rise Model

When a cylindrical glass tube of small diameter (capillary) is dipped into free water, a meniscus forms in the tube owing to the contact angle between water and the tube walls, and minimum surface energy requirements. The smaller the tube radius, the larger the degree of curvature and the pressure difference across the air–water interface (Figure 4). The pressure at the water side (P_W) is lower than atmospheric

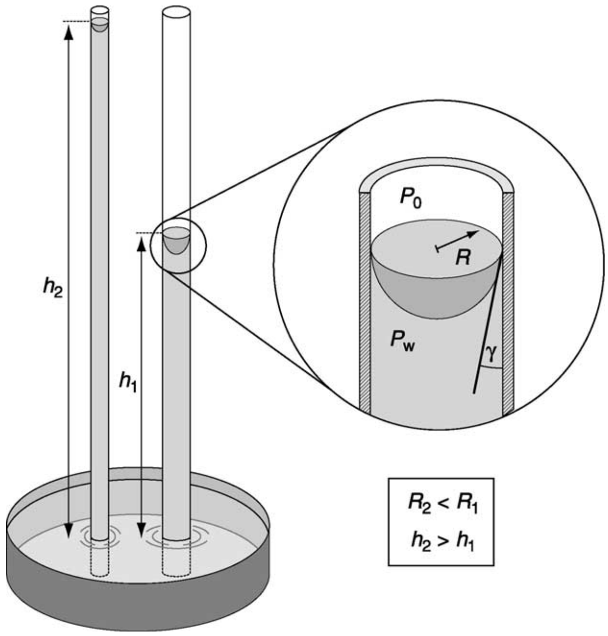


Figure 4 Capillary rise in cylindrical tubes with different radii.

pressure (P_0). This pressure difference causes water to rise into the capillary until the upward capillary force is balanced by the weight of the water column. In a cylindrical tube, the radius of meniscus curvature (R) is related to the tube radius r by $R = r/\cos\gamma$; consequently the equilibrium height of capillary rise in a cylindrical tube with contact angle γ is:

$$h = \frac{2\sigma \cos\gamma}{\rho_w g r} \qquad [6]$$

where g is the acceleration of gravity, and ρ_w is the liquid density. For water at 20°C in a glass capillary with $\gamma = 0°$, the capillary rise equation simplifies to: $h(\text{mm}) = 15/r(\text{mm})$.

Capillarity in Soils

The complex geometry of soil pore space creates numerous combinations of interfaces, capillaries, and wedges in which water is retained, and results in a variety of air–water and solid–water configurations. Water is drawn into and/or held by these interstices in proportion to the resulting capillary forces. In addition, water is adsorbed on to solid surfaces with considerable force at close distances. Due to practical limitations of present measurement methods, no distinction is made between the various mechanisms affecting water in porous matrices (i.e., capillarity and surface adsorption). Common conceptual models for water retention in porous media and matric potential rely on a simplified picture of soil pore space as a 'bundle of capillaries' (*See* **Water Retention and Characteristic Curve**). The primary conceptual steps made in such models are illustrated in **Figure 5**. The representation of soil pores as equivalent cylindrical capillaries greatly simplifies modeling and parameterization of soil pore space and relies heavily on the capillary rise equation (eqn[6]).

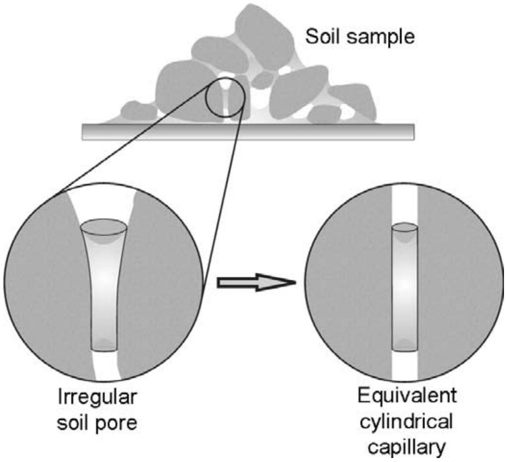


Figure 5 Idealization of the soil pore space as cylindrical capillaries.

Capillarity in Angular Pores

Cursory inspection of scanning electron micrographs of soils and other natural porous media (**Figure 6**) shows that pore spaces formed by aggregation of primary particles and mineral surfaces tend to be angular and slit-shaped, rarely resembling cylindrical tubes. Such observations and other shortcomings of the 'cylindrical capillary' model have led to development of new models for capillarity in angular and slit-shaped pores.

Capillarity in angular pores is quite different from the behavior in cylindrical pores with equivalent cross-sectional area. For example, when angular pores are drained, a fraction of the wetting phase (water) remains in the pore corners (**Figure 7a**). This aspect of 'dual occupancy' of wetting and nonwetting phases, not possible in cylindrical tubes, more realistically represents liquid configurations and the mechanisms for maintaining hydraulic continuity in porous media. Liquid-filled corners and crevices play an important role in displacement rates of oil and in other transport processes in partially saturated porous media. For all (regular and irregular) polygons with n corners, the total water filled area (A_{wt}) at a given matric potential is simply the sum of the

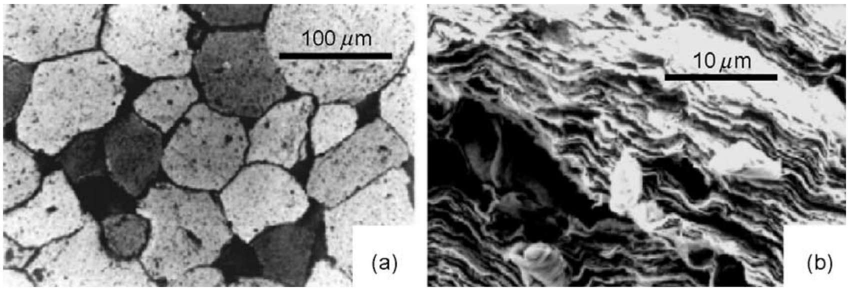


Figure 6 (a) Thin section of Devonian sandstone, revealing angular pore space. (Reproduced from Roberts JN and Schwartz LM (1985) Grain, consolidation and electrical conductivity in porous media. *Physical Review B* 31(9): 5990–5997.) (b) Scanning electron micrograph of calcium-saturated montmorillonite clay.

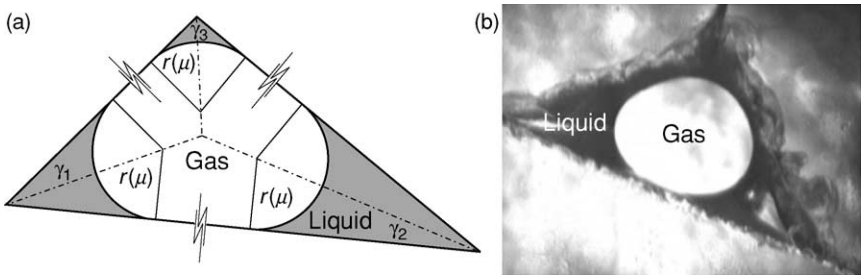


Figure 7 (a) Dual-occupancy of wetting and nonwetting phases in triangular pores; (b) liquid–vapor interfacial configuration in a triangular glass pore (~2 mm).

water-filled areas in each corner (**Figure 7a**). This sum is given by the simple equation:

$$A_{\mathrm{w_t}} = r(\mu)^2 \cdot F(\gamma) \qquad [7]$$

with

$$F(\gamma) = \sum_{n=1}^{\mathrm{i}} \left(\frac{1}{\tan\left(\frac{\gamma_\mathrm{i}}{2}\right)} - \frac{\pi \cdot (180 - \gamma_\mathrm{i})}{360} \right) \qquad [8]$$

where μ is the matric potential and $F(\gamma)$ is a shape factor dependent on pore angularity (corner angles γ_i) only.

In contrast to a piston-like filling or emptying of circular capillaries, angular pores undergo different filling stages and spontaneous displacement in the transition from dry to wet or vice versa. Under relatively dry conditions (low chemical potentials) liquid accumulates in corners due to capillary forces. An increase in chemical potential leads to an increase in the capillary radius of interface curvature until the capillary corner menisci contact to form an inscribed circle. At this critical potential, liquid spontaneously fills up the central pore (pore snap-off). The radius of interface curvature at this critical point is equal to the radius of an inscribed circle in the pore cross-section. If an angular pore is drained, liquid is displaced from the central region first, leaving some liquid behind in corners. Subsequent decrease in chemical potential results in incrementally decreasing amounts of liquid in the corners. The critical potentials at spontaneous liquid displacement differ for imbibition and drainage. (*See* **Water Retention and Characteristic Curve**.)

For completeness, one must also consider the role of liquid films due to adsorption to solid surfaces. (*See* **Water Potential; Water Retention and Characteristic Curve**.)

Dynamic Aspects of Capillarity

Dynamics of Capillary Rise

The equilibrium height of fluid rise in a capillary (eqn [6]) does not contain any information regarding the rate of rise and the associated time scale, which is often of significant importance in many industrial and natural processes. A simple force balance can be employed between a driving capillary force F_σ:

$$F_\sigma = 2\pi\ \mathrm{R}\sigma\ \cos\gamma \qquad [9]$$

and a retarding viscous force F_η (assuming Poiseuille flow):

$$F_\eta = 8\pi\eta x \frac{\mathrm{d}x}{\mathrm{d}t} \qquad [10]$$

to model the rate of capillary flow into a horizontal capillary. Inertial effects can be included, according to:

$$m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = F_\gamma - F_\mu \qquad [11]$$

where m is the mass of the liquid in the capillary, x is distance, and t is time. Substitution of the forces (eqns [9] and [10]) into eqn [11] and integration (neglecting higher-order terms) yields the so-called Lucas–Washburn–Rideal (LWR) equation:

$$x = \sqrt{\left(\frac{R\sigma}{\eta}\frac{\cos\gamma}{2}\right)t} \qquad [12]$$

which describes the rate of liquid penetration into a horizontal capillary with the dependency of x on $\sqrt{t}$. It is interesting to note that Washburn's neglect of inertial effects and Rideal's truncation of higher-order terms ($r^{-\mathrm{n}}$, $n > 2$) in his series solution yield the same solution (eqn [12]). Exact solutions have been provided that fully account for inertial effects and expand the LWR expression to consider flows into horizontal grooves and other capillary shapes.

Analytical solutions for dynamic capillary rise with gravity present a mathematical challenge. Several simplified analytical solutions for the rate of capillary rise in vertical capillaries have been proposed, such as the following implicit solution:

$$\frac{\rho g R^2}{8\eta} t = z(t) - z_\mathrm{e}\ \ln\left[1 - \frac{z(t)}{z_\mathrm{e}}\right] \qquad [13]$$

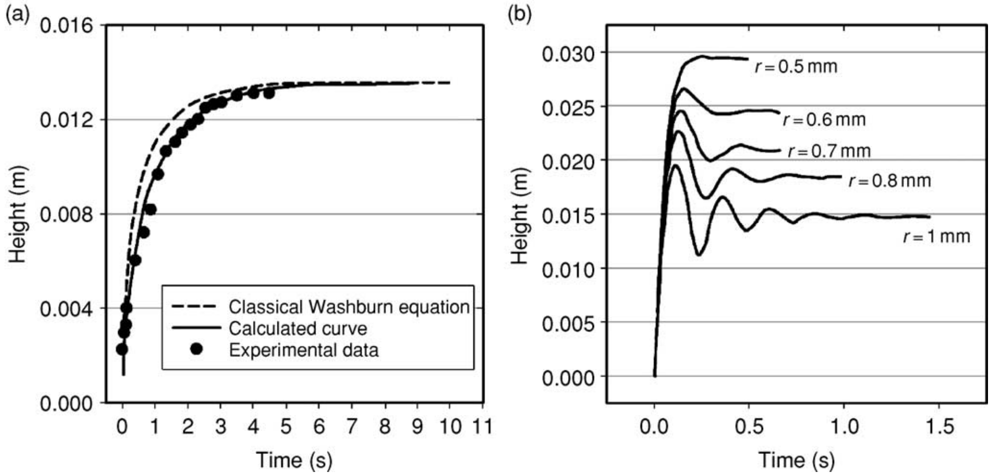


Figure 8 (a) Comparison of measurements and theoretical models for capillary rise dynamics of silicon oil (PDMS 10) in glass capillary with $r = 0.315$ mm (calculated curve from eqn [14]; classic Washburn equation from eqn [13]); (b) inertia-induced oscillations during capillary rise of water in different glass capillary sizes (numerical simulations). Note that inertial oscillations vanish for capillaries smaller than $r = 0.474$ mm according to eqn [15]. (Adapted from Hamraoui A and Nylander T (2002) Analytical approach for the Lucas–Washburn equation. *Journal of Colloid Interface Science* 250: 415–421.)

The solution diverges as $z(t)$ approaches the equilibrium capillary rise z_e (eqn [6]). Another approximate solution has been proposed, based on the introduction of a retardation coefficient (β):

$$z(t) = z_e\left(1 - \exp\left[-\frac{\sigma\cos\gamma}{\beta z_e}t\right]\right) \qquad [14]$$

The solution converges to the equilibrium capillary rise z_e (eqn [6]) for long periods of time. Figure 8a depicts comparison of eqns [13] and [14] with capillary-rise measurements of silicon oil (PDMS 10) in a glass capillary, with $r = 0.315$ mm ($\sigma = 20.1$ mN m^{-1}; $\eta = 10$ mPa; and $\rho = 0.935$ kg m^{-3}).

The nondimensional retardation coefficient for water in glass capillaries ranges from $\beta = 0.5$ for large radii ($r > r_c$), representing friction dissipation due to contact line motion and contact angle adjustment, to $\beta = 0.7$ for small radii ($r < r_c$) representing primarily viscous dissipation. The critical radius r_c is related to an interesting feature of capillary rise in the presence of gravity, namely inertia-induced oscillations in large capillaries, as depicted in Figure 8b. The inertial oscillations disappear in capillaries of radii smaller than r_c:

$$r_c = \frac{2\left(\sigma \cdot \cos(\gamma)\eta^2\rho^2 g^3\right)^{1/5}}{\rho g} \qquad [15]$$

Dynamic Contact Angle

The contact angle formed between a flowing liquid front (advancing or receding) and a solid surface is not constant but reflects the interplay between capillary and viscous forces. The relative importance of these forces is often expressed by the so-called capillary number, $\mathrm{Ca} = \eta\, v/\sigma$, with η the liquid dynamic viscosity, and v the contact line velocity. The dependency of the dynamic contact angle γ_D on the velocity of the contact line during complete wetting can be described by a nearly universal behavior according to the so-called Tanner law:

$$\gamma_D^3 = A\mathrm{Ca} \qquad [16]$$

where A is a constant (~94 for γ_D in radians). Eqn [16] fits the data of Hoffman for $\mathrm{Ca} < 0.1$ and $\gamma_D < 130°$ (Figure 9). The complete range of Hoffman's data fitted to the empirical expression:

$$\gamma_D(\mathrm{rad}) = \cos^{-1}\left\{1 - 2\tanh\left[5.16\left(\frac{\mathrm{Ca}}{1 + 1.31\mathrm{Ca}^{0.99}}\right)^{0.706}\right]\right\} \qquad [17]$$

is depicted by a continuous line in Figure 9.

For conditions of partial wetting ($\gamma_S > 0$), the relationships between contact angle and Ca are less universal. It has been postulated that at low Ca the apparent dynamic contact angle remains close to the static angle but rapidly deviates when Ca exceeds the value for γ_S (Figure 9). This postulate is formalized by the following expression:

$$\gamma_D^3 - \gamma_S^3 = A\mathrm{Ca} \qquad [18]$$

Additional examples of advancing and receding contact angle dependency on capillary number are shown in Figure 10. Note that for receding contact angle there is a critical Ca above which the contact angle vanishes.

The theoretical basis for the postulate in eqn [18] was first derived by Voinov, using hydrodynamic approximations near the moving contact line, resulting in:

$$\gamma_D^3 - \gamma_S^3 = 9\text{Ca}\ \ln(Y/Y_m) \qquad [19]$$

where Y/Y_m is a ratio of macroscopic length over which the contact angle is defined (~mm) to molecular length where continuum theories fail (~nm). Application of eqn [19] with $Y/Y_m = 10^5$ to the data of Hoffman is depicted in **Figure 9**. A key shortcoming of such hydrodynamic models for a dynamic contact angle is the lack of consideration of the effects and interactions with solid surface properties.

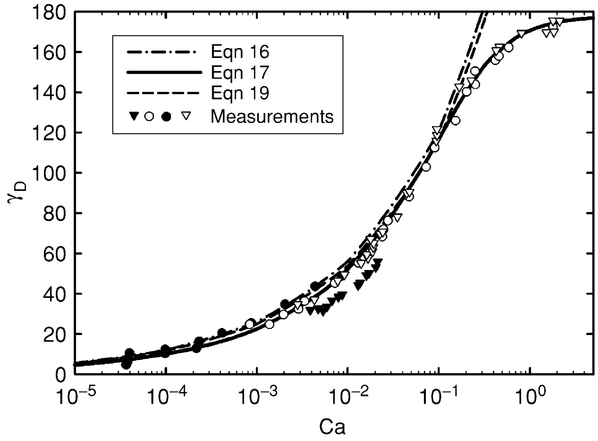


Figure 9 Experimental results of Hoffmann fitted with eqn [17] (Hoffmann RL (1975) A study of advancing interface. *Journal of Colloid Interface Science* 50: 228–241), and approximations given by eqns [16] and [19]. Note that, for water flow in soils, the capillary number Ca rarely exceeds the range of values between 10^{-6} and 10^{-4} (for $v = 1\ \text{mm s}^{-1}$, $\text{Ca} = 1 \times 10^{-5}$). (Adapted from Kistler SF (1993) Hydrodynamics of wetting. In: Berg JC (ed.) *Wettability,* pp. 311–429. New York: Marcel Dekker.)

Heterogeneous Surfaces and Microscale Hysteresis

Contact Angle on Chemically Heterogeneous and Rough Surfaces

Consider a chemically heterogeneous surface made up of patches of solids (or grains) with two different equilibrium contact angles γ_a and γ_b, and with the fraction of the area occupied by a solid given as f (**Figure 11**). The apparent equilibrium contact angle (γ_e) for the composite surface is given by the semi-empirical Cassie equation:

$$\cos\gamma_e = f\cos\gamma_a + (1-f)\cos\gamma_b \qquad [20]$$

An example of the Cassie law for contact angle of water on a sand surface with increasing amounts of hydrophobic grains is shown in **Figure 12**. The Cassie law (eqn [20]) is in remarkable agreement with experimental data for sand (**Figure 12**) and silt surfaces.

An interesting extension of the Cassie law for porous surfaces (soil, fabric, etc.) predicts that the apparent contact angle (γ_e) should be proportional to surface porosity (n):

$$\cos\gamma_e = (1-n)\cos\gamma_a - n \qquad [21]$$

The negative sign associated with porosity is due to the nonwetting properties of empty pores (i.e., air with $\cos\gamma_{air} = -1$).

These concepts of mixed wettability can be incorporated into the capillary rise model (eqn [6]) where capillary rise takes place in slits formed between two walls of different wettability. The same study applies

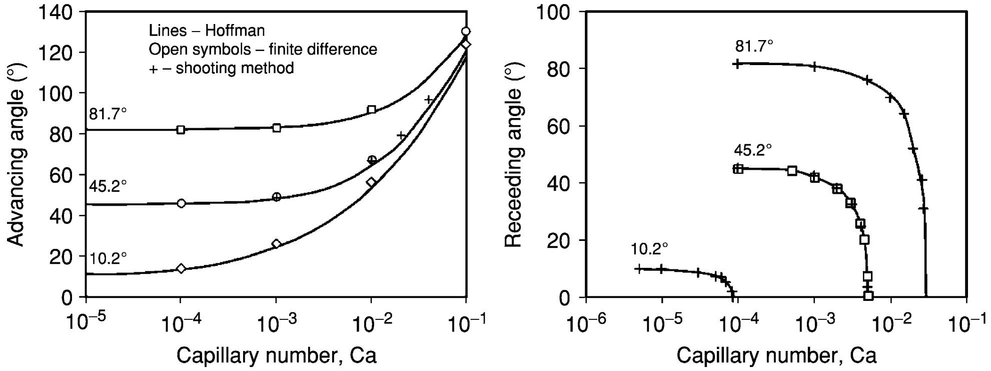


Figure 10 Finite difference computation versus eqn [18] and parameters from Kistler (Kistler SF (1993) Hydrodynamics of wetting. In: Berg JC (ed.) *Wettability,* pp. 311–429. New York: Marcel Dekker.) for advancing (left) and receding (right) contact angle as a function of Ca. (Adapted from Hirasaki GJ and Yang SY (2002) Dynamic contact line with disjoining pressure, large capillary numbers, large angles, and prewetted, precursor, or entrained films. In: Mittal KL (ed.) *Contact Angle, Wettability and Adhesion*, vol. 2, pp. 1–30.)

the Cassie law to liquid retention in porous media and demonstrates these effects on the hydraulic properties of unsaturated porous media with varying surface wettability.

In addition to surface chemical heterogeneity, the roughness of a surface is known to alter its wettability properties by increasing the wettable surface area per unit projected area, and by enabling a complex interplay between macroscopic contact angle and microscale geometry, leading to gas entrapment and a patchwork of microinterfaces underneath the wetting fluid. A spectacular demonstration of surface roughness-induced super hydrophobicity with a water drop resting on a fractal hydrophobic surface and forming a contact angle of about 170° is shown in Figure 13a. Such enhanced hydrophobicity is not only important for a variety of engineering and industrial treatments aimed at waterproofing of surfaces and fabrics, but it may also be important for explaining wettability properties of natural soil surfaces.

Assuming that surface roughness only affects the solid–liquid and solid–vapor interfacial areas, minimization of surface free energy results in the so-called Wenzel equation:

$$\cos\gamma_e = r\cos\gamma \qquad [22]$$

where γ is the static contact angle for a smooth surface of similar chemical composition (see scheme in Figure 11b).

The scope of surface influence is more complicated than predicted by simple expressions such as the Cassie and Wenzel equations. Other factors such as details of roughness geometry, interfacial pinning, and air trapping conspire to accentuate surface wetting properties as shown in Figure 13b. The scheme depicted in Figure 13b is based on experimental results showing the apparent contact angle on a rough surface plotted against the static contact angle on a smooth surface with similar chemical composition (to isolate the influence of surface roughness). Subsequent studies have shown a range of behaviors and asymmetry between the hydrophobic ($\cos\gamma < 0$) and hydrophilic ($\cos\gamma > 0$) sides of Figure 13b. It is interesting to note that certain roughness patterns

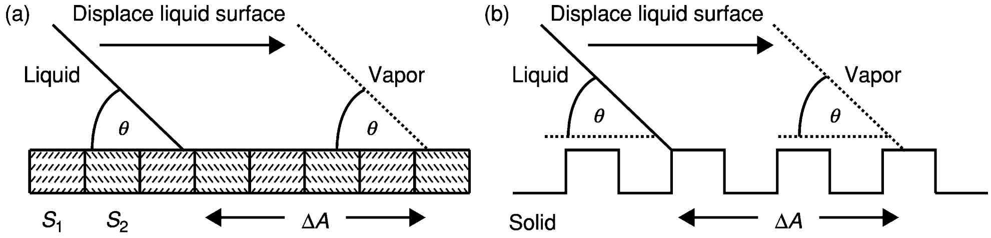


Figure 11 Definition sketch for contact angle formation on (a) chemically heterogeneous surface and (b) rough surface with $r = \Delta A/\Delta A_0$, where A_0 is the projected area over a smooth surface. (Reproduced from McHale G and Newton MI (2002) Frenkel's method and the dynamic wetting of heterogeneous planar surfaces. *Colloids and Surfaces A* 206: 193–201.)

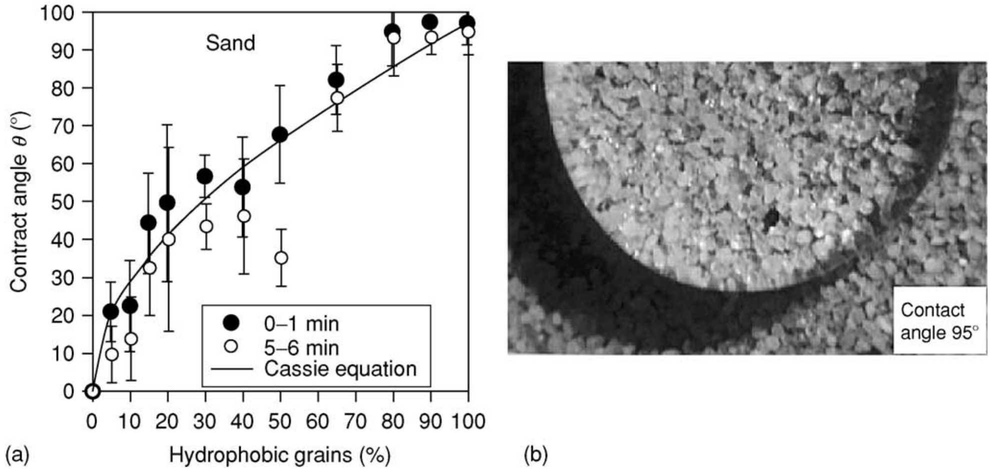


Figure 12 Application of the Cassie law to (a) experimental results of contact angle with sand surfaces containing different proportions of hydrophobic (treated) sand grains; and (b) an image of a water droplet on nonwetting sand forming a contact angle of 95°. (Adapted from Bachmann J, Elliesb A, and Hartgea KH (2000) Development and application of a new sessile drop contact angle method to assess soil water repellency. *Journal of Hydrology* 231: 66–75.)

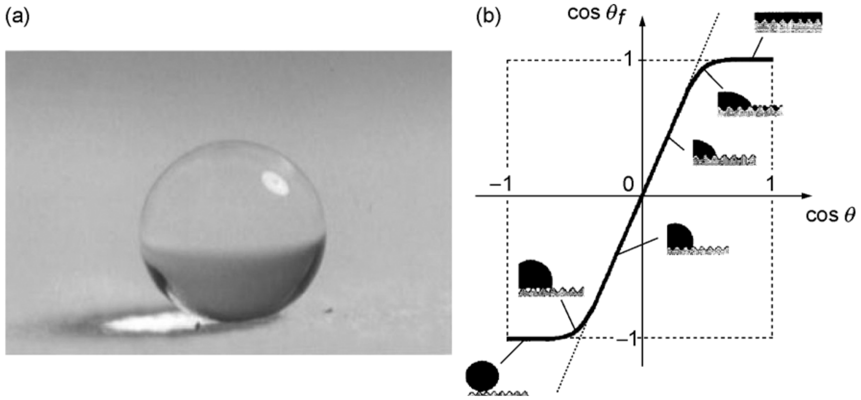


Figure 13 (a) Water drop ($r=1$ mm) resting on fractal rough surface with $r=4.4$ (eqn [22]); and (b) apparent contact angles as a function of surface microroughness for a range of surfaces with different wettability. (Reproduced with permission from Onda T, Shibuichi S, Satoh N, and Tsujii K (1996) Super-water-repellent fractal surfaces. *Langmuir* 12: 2125–2127.)

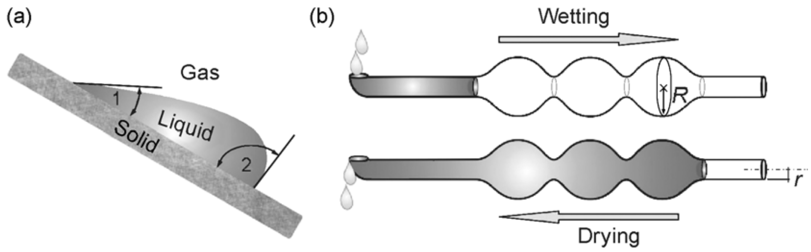


Figure 14 Two microscale mechanisms for hysteresis in capillary behavior: (a) differences between advancing and receding contact angle; and (b) the ‘ink bottle’ effect depicting two different amounts of liquid retained in identical pores under the same matric potential.

induce formation of air patches trapped underneath the liquid (similar to water drops resting on surfaces of some plant leaves).

Hysteresis

The amount of liquid retained in a porous medium is not uniquely defined by the value of the matric potential but is also dependent on the ‘history’ of wetting and drying. This phenomenon, known as hysteresis, is closely related to various aspects of pore geometry, capillarity, and surface wettability. The macroscopic manifestation of hysteresis in soil water retention (or soil water characteristic) is rooted in several microscale mechanisms, including: (1) differences in liquid–solid contact angles for advancing and receding water menisci (Figure 14a), which is accentuated during drainage and wetting at different rates; (2) the ‘ink bottle’ effect resulting from nonuniformity in shape and sizes of interconnected pores, as illustrated in Figure 14b, whereby drainage of the irregular pores is governed by the smaller pore radius *r*, and wetting is dependent on the larger radius *R*. Additional effects stem from pore angularity; (3) differences in air-entrapment mechanisms; and (4) swelling and shrinking of the soil under wetting and drying, respectively. From early observations to the present, the role of individual factors remains unclear, and hysteresis is a subject of ongoing research.

See also: **Water Potential**; **Water Retention and Characteristic Curve**

Further Reading

Adamson AW (1990) *Physical Chemistry of Surfaces*, 5th edn. New York: John Wiley.

Bachmann J, Elliesb A, and Hartgea KH (2000) Development and application of a new sessile drop contact angle method to assess soil water repellency. *Journal of Hydrology* 231: 66–75.

Bico J, Thiele U, and Quere D (2002) Wetting of textured surfaces. *Colloids and Surfaces A* 206: 41–46.

Blunt M and Scher H (1995) Pore-level modeling of wetting. *Physical Review E* 52(6): 6387–6403.

Dullien FAL, Lai FSY, and Macdonald IF (1986) Hydraulic continuity of residual wetting phase in porous media. *Journal of Colloid Interface Science* 109: 201–218.

Friedman SP (1999) Dynamic contact angle explanation of flow rate-dependent saturation-pressure relationships during transient liquid flow in unsaturated porous media. *Journal of Adhesion Science Technology* 13: 1495–1518.

Haines WB (1930) Studies in the physical properties of soil. V. The hysteresis effect in capillary properties, and the modes of moisture distribution associated therewith. *Journal of Agricultural Science* 20: 97–116.

Hamraoui A and Nylander T (2002) Analytical approach for the Lucas–Washburn equation. *Journal of Colloid and Interface Science* 250: 415–421.

Hirasaki GJ and Yang SY (2002) Dynamic contact line with disjoining pressure, large capillary numbers, large angles and pre-wetted, precursor, or entrained films. In: Mittal KL (ed.) *Contact Angle, Wettability and Adhesion*, vol. 2, pp. 1–30. Zeist, the Netherlands: VSP.

Hoffman RL (1975) A study of advancing interface. *Journal of Colloid Interface Science* 50: 228–241.

Kistler SF (1993) Hydrodynamics of wetting. In: Berg JC (ed.) *Wettability*, pp. 311–429. New York: Marcel Dekker.

Kool JB and Parker JC (1987) Development and evaluation of closed-form expressions for hysteretic soil hydraulic properties. *Water Resources Research* 23: 105–114.

Lucas R (1918) Ueber das Zeitgesetz des kapillaren Aufstiegs von Flussigkeiten. *Kolloid Zeitschrift* 23: 15–22.

Marmur A (1992) Wettability. In: Schrader ME and Loeb GI (eds) *Modern Approaches to Wettability: Theory and Applications*. New York: Plenum Press.

McHale G and Newton MI (2002) Frenkel's method and the dynamic wetting of heterogeneous planar surfaces. *Colloids and Surfaces A* 206: 193–201.

Morrow NR and Xie X (1998) Surface energy and imbibition into triangular pores. In: van Genuchten MT, Leij FJ, and Wu L (eds) *Proceedings of International Workshop on the Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media*. Riverside, CA: University of California Press.

Nitao JJ and Bear J (1996) Potentials and their role in transport in porous media. *Water Resources Research* 32: 225–250.

Onda T, Shibuichi S, Satoh N, and Tsujii K (1996) Super-water-repellent fractal surfaces. *Langmuir* 12: 2125–2127.

Quere D, Raphael E, and Ollitrault J-Y (1999) Rebounds in a capillary tube. *Langmuir* 15: 3679–3682.

Rideal EK (1921) On the flow of liquids under capillary pressure. *Philosophical Magazine* 44: 1152–1159.

Rye RR, Mann JA Jr., and Yost FG (1996) The flow of liquids in surface grooves. *Langmuir* 12: 555–565.

Sciffer S (2000) A phenomenological model of dynamic contact angle. *Chemical Engineering Science* 55: 5933–5936.

Shibuichi S, Onda T, Satoh N, and Tsujii K (1996) Super water-repellent surfaces resulting from fractal structure. *Journal of Physical Chemistry* 100: 19512–19517.

Tuller M, Or D, and Dudley LM (1999) Adsorption and capillary condensation in porous media: liquid retention and interfacial configurations in angular pores. *Water Resources Research* 35(7): 1949–1964.

Ustohal P, Stauffer F, and Dracos T (1998) Measurement and modeling of hydraulic characteristics of unsaturated porous media with mixed wettability. *Journal of Contaminant Hydrology* 33: 5–37.

Voinov OV (1976) Hydrodynamics of wetting. *Fluid Dynamics* 11: 714.

Washburn EW (1921) The dynamics of capillary flow. *Physical Review* 17: 273–283.

CARBON CYCLE IN SOILS

Contents

Dynamics and Management
Formation and Decomposition

Dynamics and Management

C W Rice, Kansas State University, Manhattan, KS, USA

Introduction

Carbon is the fundamental building block of all life on Earth. It is present in the atmosphere, plant and animal life, nonliving organic matter, fossil fuels, rocks, and is dissolved in oceans. Carbon is the sixth-most abundant element in the universe, after hydrogen, helium, oxygen, neon, and nitrogen. Concerns about increasing levels of carbon dioxide in the atmosphere have resulted in greater public and scientific interest in the global carbon cycle. Since the mid to late 1800s, fossil fuel use, expansion of cultivated agriculture, and forest clearing have led to an increase in atmospheric CO_2 from 260 ppm to current levels of approximately 370 ppm. Most of the recent increase in atmospheric CO_2 is attributed to burning of fossil fuels. Current levels of CO_2 are higher now than in the past 1000 years. The increase in CO_2 is of concern because CO_2 is one of the three primary greenhouse gases, the other two being methane (CH_4) and nitrous oxide (N_2O). These three gases help retain heat that normally radiates from the Earth's surface. The concern is that elevated levels of CO_2 could increase the heat retained in the atmosphere thus leading to global warming. Observations have recorded a 0.6°C increase in the global temperature since 1900.

The basic carbon cycle of life is: (1) the conversion of atmospheric carbon dioxide to carbohydrates by photosynthesis in plants; (2) the consumption and oxidation of these carbohydrates by animals and microorganisms to produce carbon dioxide and other products; and (3) the return of carbon dioxide to the atmosphere (**Figure 1**). On a global level, the total carbon cycle is more complex and involves carbon stored in fossil fuels, soils, oceans, and rocks.

We can organize all the carbon on Earth into five main pools, listed in order of the size of the pool:

1. Lithosphere (Earth's crust): This consists of fossil fuels and sedimentary rock deposits such as limestone, dolomite, and chalk. This is much the largest carbon pool on Earth. The amount of carbon in the lithosphere is 66–100 million Pg ($1\ \mathrm{Pg} = 10^{15}\ \mathrm{g}$). Of this amount, only 4000 Pg consists of fossil fuels;
2. Oceans: Ocean waters contain dissolved carbon dioxide, and calcium carbonate shells in marine organisms. The amount of carbon in oceans is 38 000–40 000 Pg;
3. Soil organic matter: The amount of carbon store in soil organic matter is 1500–1600 Pg;
4. Atmosphere: The atmosphere consists primarily of carbon dioxide, carbon monoxide, and methane. The amount of carbon in the atmosphere has increased from 578 Pg in 1700 to about 766 Pg in 1999, and continues to increase at the rate of about 6.1 Pg year^{-1};
5. Biosphere: The biosphere consists of all living and dead organisms not yet converted into soil organic matter. The amount of carbon that resides in the biosphere is 540–610 Pg.

In the terrestrial system, soils play a key role both in reservoirs and fluxes in the plant–soil–atmosphere continuum. Carbon in soils is present both in inorganic and organic forms. Carbon comprises up to 10% of the soil mass, with the exception of waterlogged soils such as Histosols that contain up to 30% carbon. The surface horizons of most soils contain less than 3% carbon. Most of the carbon that resides in soil is in the organic form, with the exception of arid soils and some soils formed from carbonate parent material. While carbon is a relatively minor component in terms of mass, it serves an important function in soil and the environment.

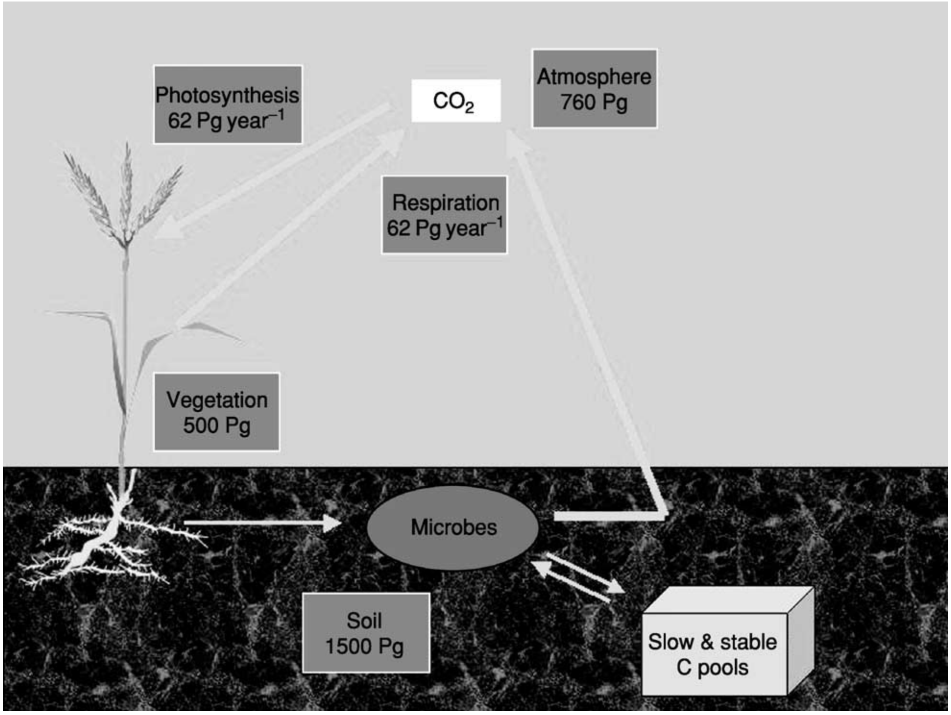


Figure 1 Simplified representation of the terrestrial carbon cycle between the atmosphere, plants, and soil.

Fluxes

The largest fluxes of the carbon cycle are those that occur between the atmosphere and the vegetation and oceans. Carbon enters the soil primarily through plants (Figure 1). During photosynthesis, plants convert CO_2 into organic carbon for growth. The estimated fixation by photosynthesis is approximately 62 Pg year^{-1}. This plant carbon enters the soil through root exudates, root turnover, and the addition of aboveground plant material. Upon their death, plant tissues decompose, primarily by soil microorganisms, and then much of the carbon in the plant material is eventually released back into the atmosphere as CO_2. Respiration is estimated to be 62 Pg year^{-1}, which balances photosynthesis. This compares with 5 Pg year^{-1} released by burning of fossil fuels; however the release of C by burning of fossil fuels is not offset by a sink, thus resulting in the dramatic increase in atmospheric CO_2. In some cases under anaerobic conditions, methane (CH_4) is produced by microorganisms. However, soil microorganisms also consume CH_4 under aerobic conditions. Approximately 7% of the atmospheric C is recycled between plant uptake and microbial respiration. Thus plant productivity and microbial respiration are the two key mediators between atmospheric CO_2 uptake and input.

Soil-Forming Factors

The level of organic carbon in soils is a function of climate, topography, biology (plant and microorganisms), parent material, and time. These are the same factors that affect soil development. The first few years after disturbance, soil carbon increases relatively dramatically but slows until equilibrium is attained under the conditions of the existing environment. This is easily seen after reclamation of mined lands or conversion of cultivated cropland to permanent grasslands. A change in the environmental conditions with climate or management can alter the equilibrium levels. The equilibrium is reached as plant inputs balance the outputs of microbial respiration. The soil organic C levels are a result of biological recalcitrance and physical and chemical stabilization of the carbon, to be discussed later.

Climate is probably the most important factor governing soil organic carbon levels. Climate through water and temperature affect plant productivity and microbial activity. Hot climates where water is restricted have the lowest levels of soil organic carbon, because plant production is limited. Deserts are characterized by this environment. In the USA, a west to east gradient from the Rocky Mountains east shows increasing levels of soil carbon due to greater precipitation and thus increasing plant productivity.

Vegetation type can also influence the soil organic carbon levels through compositions and production of the plant biomass. Plant materials vary in their decomposability due to chemical differences. Lignin and other polyphenolic substances decrease the decomposition of the plant material. Grasses tend to promote higher soil organic carbon. Grasses not only have high productivity but also allocate more photosynthate belowground. The high densities of grass roots tend to favor formation of soil organic carbon. While forests also sequester carbon, a greater proportion of that carbon is in woody biomass. There are differences among crop species in the ability to sustain soil organic carbon levels. In general high residue-producing annual crops such as corn, wheat, and grain sorghum replenish the supply of carbon to the soil. Low residue-producing crops such as cotton and soybeans tend to deplete soil organic carbon levels.

Parent material affects levels of soil carbon through the effect on texture, mineralogy, and pH. The amount of clay in the soils also influences the ability of the soil to form stable soil aggregates, which contribute to the protection and stabilization of soil organic carbon.

Topography affects soil organic carbon levels through effects on microclimate, runoff and erosion, and evaporation. Lower slope positions often have higher organic carbon levels because of deposition and greater plant production. Slope positions of excess water also restrict microbial decomposition, resulting in buildup of soil organic C.

Functions

Soil carbon is probably the most important component in soils as it affects the soil properties. Carbon as soil organic matter alters the physical, chemical, and biological properties of the soils (Figure 2a). Soil organic matter is a primary indicator of soil quality. Improvements in soil organic matter create a more favorable environment, leading to increases in plant growth.

Physically, soil organic matter improves aggregation of soil particles. Improved aggregation results in better soil structure, allowing for movement of air and water through the soil as well as better root growth. More stable soil structure results in less soil erosion, which retains nutrients on the land and protects water quality. Higher levels of soil organic carbon reduce bulk density, thus providing an improved rooting environment. In addition, soil organic matter holds soil water, which is an important

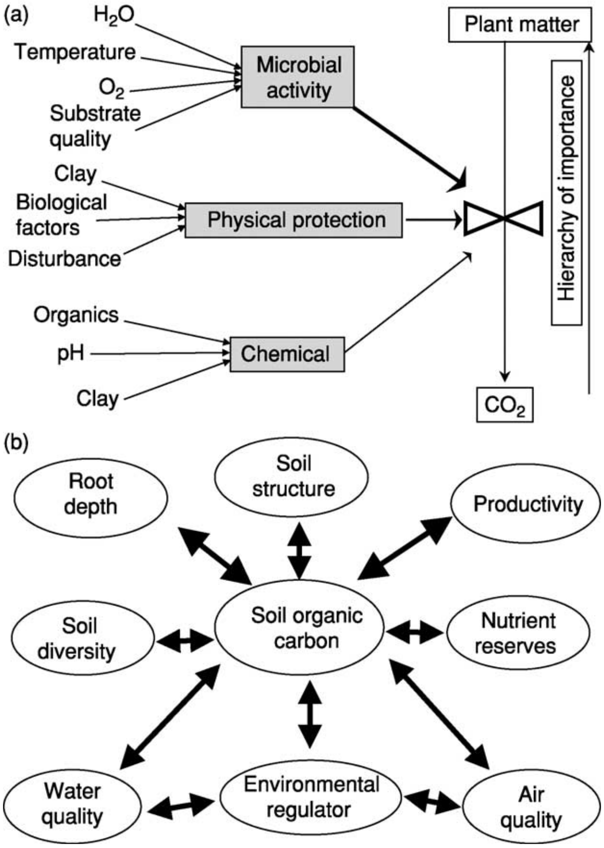


Figure 2 (a) Regulation of carbon stabilization in soil; (b) relationship between soil organic carbon and soil, water, and air quality. Reproduced with permission from Rice CW (2002) *Organic Matter and Nutrient Dynamics*. Encyclopedia of Soil Science. Dekker.

attribute for plant growth in arid and mesic environments. Higher soil organic matter decreases soil crusting and increases water infiltration rates which enhance plant productivity and thus the return of plant material to the soil.

Chemically, soil organic carbon increases the cation exchange capacity of the soil. Twenty to eighty percent of the cation exchange capacity of the soil is due to soil organic matter. These cation exchange sites are important for retention of nutrients. Associated with the organic carbon are organic-bound nitrogen, phosphorus, and sulfur which, upon decomposition, provide slow release of nutrients for plant production. In some cases, soil organic compounds enhance the chelation of metals, thus increasing the bioavailability of trace elements required for plant growth. Soil organic carbon often provides binding sites for many anthropogenic chemicals, thus minimizing leaching of hazardous chemicals through the soil profile or making them less available, which reduces toxicity.

Biologically, soil organic carbon is the source of carbon and energy for most soil microorganisms and fauna. Increased soil organic carbon enhances the biomass and diversity of the soil biota. Since the soil microbial community drives many of the microbial transformations in soil, plant nutrient availability is often enhanced with the increase in microbial biomass and activity of the soil. Some organic compounds in the soil also exhibit plant growth-promotion properties, further enhancing plant productivity.

Regulation of Soil Organic Carbon Dynamics

As much as 20% of plant carbon remains in the soil as organic matter, sometimes referred to as 'humus.' Some of this carbon can remain in soils for hundreds and even thousands of years. The quantity of organic carbon in soil is a function of the amount of plant material entering the soil, the decomposition rates of those residues, and the soil chemistry and mineralogy. Soil C may be stabilized due to biochemical recalcitrance, chemical stabilization, and physical protection (Figure 2b). Any environmental factor that affects microbial activity influences the decomposition of organic material. These factors include soil water, temperature, pH, and O_2.

Plant production and soil microbial activity are recognized as the biological processes governing soil carbon dynamics (Figure 2). The amount of plant material produced is a function of plant species, as well as climate (temperature and available water) and nutrients. The composition of the plant material entering the soil is also a key regulator of microbial decomposition. Typical plant components include soluble sugars, hemicellulose, cellulose, and lignin, which vary in proportions between plant species, within plant species, and within plant organs. Of these components, lignin is the most difficult to degrade. The quality of the carbon compounds (e.g., N content, C:N ratio, and lignin:N ratio) are important factors regulating decomposition processes (Figure 2). The C:N ratio often serves as a guide of decomposability; a ratio of more than 30 slows decomposition because of the lack of sufficient N for microbial growth; a ratio less than 20 provides sufficient N for microbial growth and allows microbial decomposition to proceed. Lignin content or lignin:N ratios are also used as a guide of organic matter degradability.

The organic carbon is transformed to greater biological recalcitrance as soil microorganisms process the C inputs from plant material. Soil carbon may be stabilized because of its biochemical recalcitrance, e.g., lignin derivatives or melanins produced by fungi and other soil organisms.

The rate of transformation is a function of temperature and soil water. As temperature increases microbial activity increases; generally, for every 10°C increase, microbial activity doubles ($Q_{10} = 2$). Soil-water content is also important; optimum microbial activity occurs at near 'field capacity,' which is equivalent to 60% water-filled spore space. As soils become waterlogged, decomposition slows and becomes less complete. Peat soils are a result of these waterlogged conditions. As soils dry below 60% water-filled pore space, decomposition is also slowed.

Soil structure plays a dominant role in controlling microbial access to substrates, and thus turnover (Figure 2). Relatively labile material may become physically protected from decomposition by incorporation into soil aggregates. Disruption of aggregates, either by natural forces (freeze–thaw, wet–dry) or human activity (tillage) stimulates decomposition of the organic C protected inside the aggregates. One of the primary causes of soil organic C loss is tillage. The role of soil structure protection of soil organic carbon from decomposition is demonstrated by increased C mineralization in disrupted aggregates relative to intact aggregates. Cultivation of soils strongly affects the structural stability and reduces the amount of soil organic C, and this loss reduces the proportion of soil macroaggregates.

Chemical stabilization of C refers to binding of organic C to clay surfaces or precipitation. Chemical stabilization by clay is regulated by clay type, pH, and other organics (Figure 3). Clay that have a high adsorption capacity such as 2:1 clays (e.g., montmorillonite), protect organic carbon to a greater extent than 1:1 clays (e.g., kaolinite), which have a lower adsorption capacity. Certain cations such as Ca, Al, and Fe are known to enhance stabilization of organic carbon through cation bridges with clay surfaces. Clay type is an inherent property of soil and therefore cannot be directly manipulated as easily as the biological and physical mechanisms of stabilization. Inorganic forms of carbon in soil are the result of equilibrium of CO_2 with water from carbonic acid (H_2CO_3), bicarbonate (HCO_3^{-1}), and carbonate (CO_2^{-3}) in the soil solution. Secondary minerals form in arid soils from the precipitation of Ca and Mg with carbonate. Dissolution of minerals also occurs depending on pH.

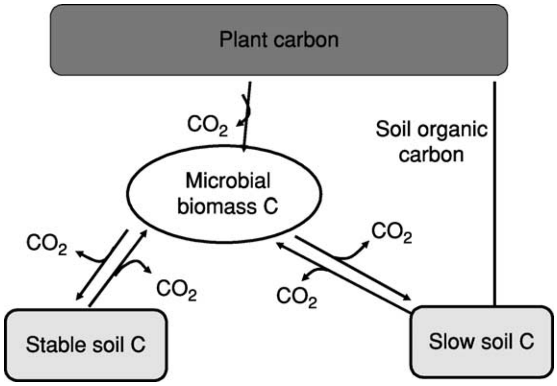


Figure 3 Conceptualization of soil carbon pools. (Adapted from Paul EA and Clark FE (1996) *Soil Biology and Biochemistry*. San Diego, CA: Academic Press, with permission.)

Composition of Soil Organic Carbon

The most widely accepted theory on the formation of soil humus carbon is that, as plant material is decomposed by microorganisms, the altered compounds and new compounds synthesized by soil microbes polymerize through chemical or enzymatic reactions. Soil organic carbon is undergoing constant transformation. Typically, most models separate soil organic carbon into pools of organic matter that differ in composition and decomposability. One model separates organic carbon into three pools (Figure 3). The 'active' pool, comprised of microbial biomass and labile organic compounds, makes up less than 5% of the soil organic C. The slow pool usually makes up 20–40% of the total organic C, and the recalcitrant pool makes up 60–70% of the soil C. The microbial biomass, while a smaller proportion of the organic C in the soil, is the processing agent. The recalcitrant pool is material that is difficult to degrade and contains what is referred in the literature as humic and fulvic acids – fractions obtained by chemical fractionation procedures. The active pool has turnover times on the order of months to years, the slow pool takes decades to turn over, while C in the recalcitrant pool takes from hundred to thousands of years to turn over completely. However, 2–5% of the recalcitrant pool is degraded annually. Since the recalcitrant pool is generally in equilibrium in natural systems, the rate for formation equals the rate of degradation.

Management

Much of the world's productive soils are now in cultivated agriculture. Historically, agriculture relied upon plowing the soil. In some cases, low crop yields and removal of crop residues reduced the amount of plant material returned back to the soil. This combination of agricultural practices resulted in reducing the replenishment of organic material (carbon) to the soil. As a result, soil C content has decreased by as much as 50% over a 50- to 100-year period in many agricultural soils. A typical soil carbon-loss curve is shown in Figure 4. Losses are rapid initially and then decline with time after the conversion to agriculture

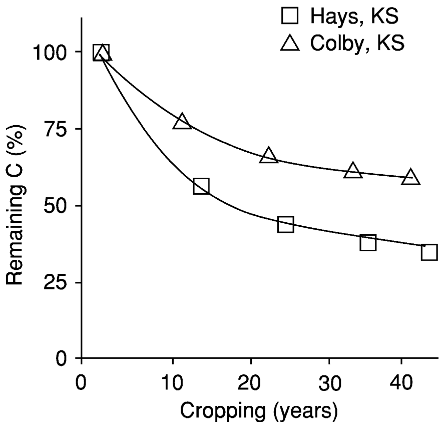


Figure 4 Loss of soil carbon at the onset of cultivation for two locations in Kansas, Hays and Colby.

Table 1 Management strategies for soil carbon sequestration

Land use	*Soil management*	*Crop management*
Cropland	Tillage management	Crop varieties
Rangeland	Residue management	Crop rotations
Forestry	Fertility management	Cover crops
	Water management	

Adapted from Lal R, Kimble JR, Follett RF, and Cole CV (1998) *The Potential of US Cropland to Sequester Carbon and Mitigate the Greenhouse Effect.* Ann Arbor, MI: Ann Arbor Press.

from native ecosystems. In recent decades, higher yields, return of crop residues, and development of conservation tillage practices have increased soil carbon. These and other advancements in crop and soil management practices have the potential to increase soil C. Table 1 lists several practices affecting the soil's ability to sequester C. These practices can be grouped into three guiding principles: (1) minimization of soil disturbance; (2) maximizing plant production through nutrients and water availability; and (3) maximizing return of plant biomass.

Cultivation or high tillage intensity is one of the primary reasons for lower organic carbon levels in soils. Tillage promotes disruption of soil aggregates, which exposes 'protected' carbon inside aggregates to microbial attack. In addition, tillage increases aeration, stimulating microbial activity, which results in conversion of organic carbon to CO_2. A reduction in tillage intensity, with no-tillage being the least disruptive in cropped soils, allows macroaggregate formation and reduces degradation of soil organic C. Minimum or conservation tillage also conserves carbon in soil. Losses of soil carbon by erosion are also reduced with conservation tillage.

Cropping intensity affects the equilibrium level of soil organic carbon. Crop rotations can impact soil carbon levels by varying the amount and composition of the plant material added to the soil. In agriculture, high residue-producing crops such as corn, wheat, and sorghum tend to sustain or increase soil C. Elimination of summer fallow, a practice used in the Great Plains, can increase soil C by providing more plant material on an annual basis. In those areas where sufficient soil water is available, double-cropping is a viable practice, where two crops are produced per year. In more humid regions, winter cover crops are also a viable option for returning plant material to the soil. In grassland systems, perennial crops also tend to increase soil C because of the lack of tillage and the addition of organic C through root turnover.

Table 2 Estimates of C sequestration potential of agricultural practices of US cropland

Agricultural practice	*(MTC ha^{-1} year^{-1})*
Conservation Reserve Program	0.3–0.7
Conservation tillage	0.24–0.40
Fertilizer management	0.05–0.15
Rotation with winter cover crops	0.1–0.3
Summer fallow elimination	0.1–0.3

Adapted from Lal R, Kimble JR, Follett RF, and Cole CV (1998) *The Potential of US Cropland to Sequester Carbon and Mitigate the Greenhouse Effect.* Ann Arbor, MI: Ann Arbor Press.

Those practices that directly affect the plant production are also important in increasing soil carbon. Improved water management alleviates short-term water stress during the growing season, which improves production of plant biomass and the return of plant material to the soil. In arid regions, irrigation can increase plant production and microbial activity. Water management also affects microbial activity, as discussed earlier. Fertility management eliminates nutrient limitations of plant growth. Proper management of nutrients provides sufficient nutrients for the plant while minimizing the excess that could lead to degradation of the environment.

For grazing lands, improved species, addition of legumes, and fertilizer management can increase soil C in tame pastures. For native lands, proper grazing rotation and management of burning are strategies to improve soil carbon, especially for previously poorly managed grazing lands.

Restoration of degraded soils and ecosystems offers high potential for storing carbon in soils. Marginal croplands could be reverted back to forests and grasslands. The Conservation Reserve Program of the US Department of Agriculture is designed to take marginal lands out of production (Table 2). Restoration of soils of mined land offers many opportunities for storing carbon in soil.

See also: **Carbon Emissions and Sequestration**; **Climate Models, Role of Soil**

Further Reading

Council for Agricultural Science and Technology (2004) *Climate Change and Greenhouse Mitigation: Challenges and Opportunities for Agriculture*. Task Force Report No. 141. Ames, IA: CAST.

Jastrow JD and Miller RM (1996) Soil aggregate stabilization and carbon sequestration: feed backs through organo-mineral associations. In: Lal R, Kimble JM, Follett RF, and Stewart BA (eds) *Soil Process and the Carbon Cycle*, pp. 207–223. New York: CRC Press.

Lal R, Kimble JR, Follett RF, and Cole CV (1998) *The Potential of U.S. Cropland to Sequester Carbon and Mitigate the Greenhouse Effect*. Ann Arbor, MI: Ann Arbor Press.

Paul EA and Clark FE (1996) *Soil Biology and Biochemistry*. San Diego, CA: Academic Press.

Paul EA, Paustian K, Elliot ET, and Cole CV (1997) *Soil Organic Matter in Temperate Agroecosystems: Long-term Experiments in North America*. Boca Raton, FL: CRC Press.

Rice CW and Owensby CE (2000) Effects of fire and grazing on soil carbon in rangelands. In: Follet R *et al.* (eds) *The Potential of U.S. Grazing Lands to Sequester Carbon and Mitigate the Greenhouse Effect*, pp. 323–342. Boca Raton, FL: Lewis Publishers.

Schlesinger WH (1997) *Biogeochemistry: An Analysis of Global Change*. San Diego, CA: Academic Press.

Smith JL, Lynch JM, Bezdicel DF, and Papendick RI (1992) Soil organic matter dynamics and crop residue management. In: Metting B (ed.) *Soil Microbial Ecology*, pp. 65–94. New York: Marcel Dekker.

Sundquist ET (1993) The global carbon dioxide budget. *Science* 259: 934–941.

Sylvia DM, Fuhrman JF, Hartel PG, and Zuberer DA (1998) *Principles and Application of Soil Microbiology*. Upper Saddle River, NJ: Prentice-Hall.

Formation and Decomposition

C A Cambardella, USDA Agricultural Research Service, Ames, USA

Introduction

Earth's ecosystems are immense thermodynamic machines that run on energy derived from the sun. Energy enters the biotic portion of Earth's ecosystems through the process of photosynthesis, where plants use the energy of sunlight to transform CO_2, the predominant form of oxidized carbon in the atmosphere, to reduced forms of organic carbon. During photosynthesis, sunlight energy is transformed to chemical energy, and stored in the chemical bonds of organic carbon compounds produced in the plant. Photosynthetically fixed carbon, either directly or indirectly, provides the energy for all other forms of life on Earth. Living organisms use the energy tied up in the carbon bonds for growth, reproduction, and biochemical maintenance. The organic carbon and its associated energy continue to be transferred among the living components of the ecosystem and are eventually released to the atmosphere by respiration or combustion (Figure 1). Carbon and energy are linked as they move through ecosystems because the same processes govern their entry into, transfer through, and loss from ecosystems. Although aboveground plant and animal communities are the most obvious aspects of terrestrial ecosystems, these surface communities rely on the existence of a dynamic, efficiently functioning belowground ecosystem. This article will examine the interrelationship between plants and soil in a discussion of carbon cycling in the soil/plant system.

Sources of Carbon

Nearly all carbon entering the soil ultimately comes from plants, with small amounts coming from photosynthetic soil bacteria. Green plant tissues are, on average, 75% water and 25% dry matter. The herbaceous parts of plants are about 40–45% carbon on a dry-weight basis, regardless of age or type. Woody tissue is slightly higher, at 50% carbon. Plant dry matter also contains, on average, 42% oxygen, 8% hydrogen, and 8% ash. About half of the carbon that is photosynthetically fixed by plants is lost as CO_2 through plant respiration. Most of the remaining carbon is transferred to the soil as dead above- and belowground plant litter, and as root exudates, and is ultimately converted to soil organic carbon through the process of decomposition. Leaching can also move soluble carbon compounds from plant leaves on the soil surface into the soil, especially in environments with high rainfall. The types and average amounts of compounds associated with plant litter in decreasing order of decomposability are sugars, starches, and simple proteins (5%), protein (8%), hemicellulose (18%), cellulose (45%), fats and waxes (2%), polyphenols (2%), and lignin (20%). Compound type and amount differ among various plant parts (leaves, stems, roots, etc.) and from one plant species to another.

Anthropogenically produced organic substances represent a relatively small proportion of total carbon inputs to soil, but localized impacts of these substances on ecosystem function can be quite dramatic. This group of substances includes synthetically produced biochemicals, such as pesticides, polycyclic aromatic hydrocarbons from the burning of fossil fuels, and a variety of xenobiotics, such as plastics.

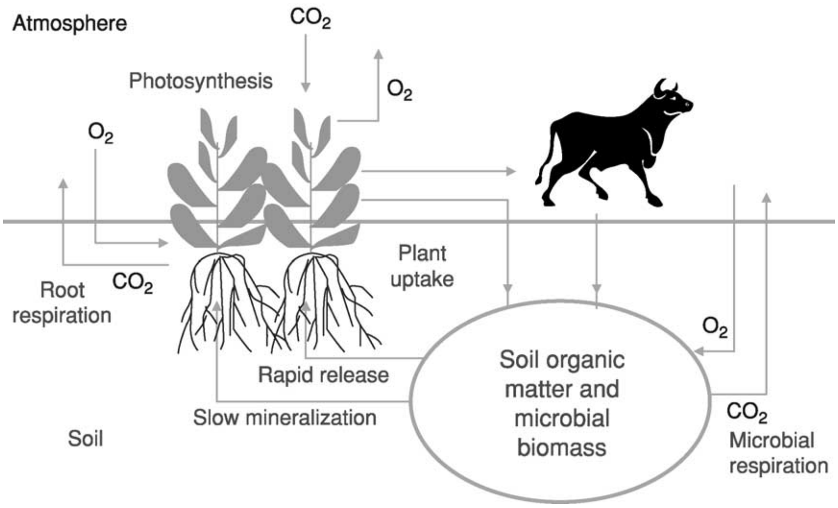


Figure 1 Soil carbon cycle. Source: USDA-NRCS, Soil Quality Institute, 2003.

Small amounts of atmospheric CO_2 can chemically react in the soil to produce carbonic acid and the carbonates and bicarbonates of calcium, potassium, and magnesium. The bicarbonates are readily soluble and can easily be lost through leaching. Carbonates produced in this manner represent a very small fraction of total carbon inputs to soil.

Aerobic Decomposition

Decomposition is the physical and chemical breakdown of organic material. In aerobic environments, decomposition is mediated by heterotrophic organisms, which derive their energy and carbon from organic matter produced by plants. Aerobic decomposition consumes oxygen, the preferred terminal electron acceptor, and returns carbon to the atmosphere as CO_2. Decomposition occurs both inside and outside living organisms and is the consequence of many interacting physical and chemical processes. The term mineralization refers specifically to the process that produces inorganic nutrient components from organic compounds. Most of the inorganic ions released by mineralization are readily available to higher plants and microorganisms.

The initial step in the decomposition process is physical fragmentation of the larger pieces of organic material into smaller ones by soil animals. There are three general categories of soil animals, classed by size. The largest are the soil macrofauna (>2 mm), such as insects, earthworms, and termites. The macrofauna impact decomposition directly through litter fragmentation and redistribution, and indirectly through burrowing. The smallest are the soil microfauna (<0.1 mm), such as flagellates, amebas, ciliates, and small nematodes. The microfauna are mobile organisms that inhabit the waterfilms on soil surfaces. They are very effective soil predators, feeding on bacteria, fungi, and other soil microfauna. Microfaunal activity impacts the temporal and spatial dynamics of nutrient mineralization and immobilization, which is an indirect, but important, control on decomposition. The soil animals that have the greatest direct effect on decomposition are the soil mesofauna (0.1–2 mm), such as collembola, mites, and other microarthropods. The soil mesofauna inhabit the air-filled pore spaces in soil, but, unlike the macrofauna, are unable to create new pore space. These organisms fragment and ingest plant litter, selectively feeding on material that has already been altered by microbial activity. They deposit large amounts of fecal material in the soil, which creates a more favorable environment for decomposition through enhanced soil water-holding capacity and surface area.

The next stage of decomposition is the biochemical alteration of the fragmented plant litter through the activity of soil bacteria and fungi. Since the compounds found in dead organic material are too large to pass through microbial membranes, the microbes secrete extracellular enzymes that convert the macromolecules into soluble products which can then be absorbed and utilized by the microbes. Numerous species of microorganisms are involved in the decomposition process because of the complex nature of the organic material.

Fungi are filamentous, aerobic soil organisms that constitute the major portion of the soil biomass. Soil fungi produce networks of filaments (hyphae) that enable them to grow into new substrates and transport materials through the soil. The hyphal network gives fungi a competitive advantage over bacteria in nutrient-poor environments, because they can transport nutrients over distances in the soil. Fungi have enzyme systems that are capable of breaking down virtually all classes of plant compounds. They are the principal decomposers of fresh plant litter, because they secrete enzymes that enable them to penetrate the cuticle of dead leaves or the suberized exterior of roots. Only a few specialized soil fungi are capable of decomposing lignin because the chemical linkages among the phenolic rings in lignin molecules are so strong and varied. Fungi account for 60–90% of the microbial biomass in forest soils, where pH and nutrient concentrations are low and plant litter has a high lignin content. In grassland soils, there are equal amounts of bacteria and fungi, and bacteria tend to dominate in frequently disturbed agricultural soils.

The bacteria are the smallest and most numerous of the organisms in soil. There is a wide range of bacterial types in soil. Rapidly growing species specialize on labile substrates secreted by roots in the rhizosphere. Their small size and large surface-to-volume ratio enable them to absorb soluble substrates rapidly. Bacteria are also important in lysing and breaking down live and dead bacterial and fungal cells. Slower-growing actinomycete species have a filamentous structure similar to fungal hyphae and can break down relatively recalcitrant substrates. Most soil bacteria are immobile and exist in complex communities on the surfaces of soil aggregates. The communities often secrete a carbon-rich extracellular mucilage that functions as a protective coat and buffers the populations against severe environmental change. The mucilage is carbohydrate in nature and is composed primarily of low-molecular-weight polysaccharides. An important consequence of immobility is that the community eventually uses up the available food supply. When competition for food becomes severe, microbial activity declines, respiration rates become negligible, and the colony becomes inactive. At any given point in time, 50–80% of the bacteria in soil are inactive. However, these bacteria can reactivate in the presence of newly introduced labile substrates, such as when a farmer plows in fresh crop residues or when deciduous trees drop their leaves in the autumn.

Fresh inputs of above- and belowground plant litter are the primary stimulus for initiation of biochemical decomposition. Soil microorganisms first begin to oxidize the most readily decomposable compounds, incorporating a portion of the organic carbon into new microbial biomass, releasing a portion as CO_2 during respiration, and using a portion for synthesis of new organic compounds. The microbial biomass carbon can account for as much as one-sixth of the total soil carbon at the peak of microbial activity. Normally, the microbial biomass is on average about 4–5% of total soil carbon. When the easily decomposable compounds are depleted, microbial activity declines. As the opportunistic microbial populations begin to die of starvation, their bodies become part of the organic substrate available for decomposition. One year after the plant litter input, 20–30% of the plant litter carbon will remain in soil as soil organic matter. Some of this carbon will be protected from further decomposition by occlusion inside soil aggregates, by adsorption to clay surfaces, and by transformation into soil humus.

Anaerobic Decomposition

The rate of decomposition under anaerobic conditions is much slower than when oxygen is plentiful. Anaerobic soils and sediments occur in water-saturated environments, such as wetlands, swamps, estuaries, and rice paddies. Anaerobic microsites can also occur under generally aerobic soil conditions as rapid decomposition depletes the oxygen supply in soil microhabitats, such as a colony of bacteria growing around a small piece of straw or root fragment. In anaerobic environments, oxygen supply frequently limits decomposition rate, so organisms must use other electron acceptors, such as NO_3^-, SO_4^{2-}, and CO_2, to derive energy from organic matter. These alternative electron acceptors provide less energy return per unit of organic matter oxidized than oxygen, so energy production is less efficient in anaerobic systems. The dominant microorganisms in anaerobic environments are soil bacteria. The most common and widespread of these bacteria are the denitrifiers. Denitrifying bacteria use NO_3^- as a terminal electron acceptor in the decomposition of organic carbon substrates and produce nitrous oxide and di-nitrogen gas as waste products. As the nitrate supply is depleted, decomposition shifts to fermentation, where organisms break down organic material to simple organic compounds and hydrogen. These fermentation products are then used by methanogens, to transfer electrons to CO_2 to produce methane. Since the supply of SO_4^{2-} is limited in terrestrial environments, sulfate reduction is a relatively less important process than methane production. Methane and nitrous oxide gas are extremely effective at absorbing infrared radiation, and are much more potent greenhouse gases than CO_2. However, methane is an effective energy source for organisms that have access to oxygen because of its highly reduced oxidative state. Methanotrophic bacteria occur in the surface layers of anaerobic soils and use methane as an energy source through the same enzyme system that converts ammonium to nitrate. Most of the methane is consumed before it escapes to the atmosphere. Even in wetland areas, methane accounts for a very small percentage of the carbon released to the atmosphere by decomposers.

Factors Controlling Decomposition

Decomposition is controlled by soil microclimate variables, such as moisture, temperature, and oxygen content; by intrinsic soil properties, such as soil texture and pH; and by the quality of the organic substrate available to the decomposers. In general, decomposition rates are highest when the temperature is between 25°C and 35°C and about 60% of the total soil pore volume is filled with water. Soil decomposers function best in soils with a pH that is close to neutral. Acidic soils generally have lower rates of decomposition compared with neutral or alkaline soils. Temperature directly affects decomposition through its effects on microbial activity. In general, microbial activity increases about 10-fold for every degree increase in temperature between about 5°C and 35°C. Typical soil temperature extremes seldom kill bacteria, and usually cause only temporary suppression of activity. Decomposition rates tend to decline if soil moisture levels are too low or too high. The optimum is approximately equal to the moisture content at field capacity, which is the point at which all of the larger soil pores have drained due to the force of gravity, and waters is held in the smaller capillary pores due to the matric forces in the soil. Soil microbes can function at soil moisture contents that are too low for plant growth but decomposition rates are reduced dramatically in wet, anaerobic soils. Considering the combined effects of moisture and temperature, soil carbon accumulation will be greatest in cool, wet soils, like those found in peat bogs and fens.

All else being equal, soils high in clay and silt are generally higher in organic matter than sandy soils. Finer-textured soils accumulate organic matter primarily because they lose less organic matter through decomposition than sandy soils. Clay particles bind with organic matter to form organomineral complexes that are very resistant to decomposition. Soils high in silt and clay tend to be structurally stable and can protect organic matter from decomposition through occlusion inside soil aggregates.

Substrate quality is defined as the susceptibility of the organic material to decomposition measured under standardized conditions. The quality of the organic substrate undergoing decomposition determines to a great extent the rate of biochemical oxidation. Differences in decomposition rate are a logical consequence of the types of carbon compound present in the organic material. Rapidly decomposing materials generally have higher concentrations of labile substrates and lower concentrations of recalcitrant compounds than do slowly decomposing materials. The ratio of carbon to nitrogen concentration, or C:N ratio, has frequently been used as an index of plant residue quality because residues with low C:N ratios generally decompose quickly. The C:N ratio in plant residue can vary widely with vegetation type. The range of C:N ratios of legumes and young green leaves is 10:1 to 30:1, whereas corn stalks, wheat straw, and other grass litter have a C:N ratio between 60:1 and 80:1. Wood can have C:N ratios as high as 600:1. In contrast, the C:N ratio of soil microorganisms varies between 5:1 and 10:1, with an average of 8:1. Bacteria generally contain more protein than soil fungi, and consequently have a lower C:N ratio.

Soil microorganisms require a balance of nutrients from which to build cells and extract energy. On average, soil microbes must incorporate into their cells about eight molecules of carbon for every molecule of nitrogen. Since soil microbes are not very efficient at converting substrate carbon to biomass C, only one-third of the carbon metabolized is incorporated into their cells. The remainder is lost as respired CO_2. This means that soil microbes need about 1 g of nitrogen for every 24 g of carbon that they metabolize. Most fresh plant residues do not have enough nitrogen to support the synthesis of new microbial tissue. If the C:N ratio of the plant residue is greater than 25:1, the soil microbes must scavenge the soil solution to obtain enough nitrogen. Assuming soil concentrations of inorganic nitrogen are sufficiently high, the incorporation of high C:N ratio residues will deplete the soil of soluble nitrogen. If soil concentrations of inorganic nitrogen are insufficient, then decomposition will stop, or at least be delayed.

Plant residues that are high in lignin and/or polyphenols are considered to be poor-quality substrates for the soil organisms that cycle carbon and nutrients. Lignin is degraded slowly because only some soil fungi produce the enzymes necessary to degrade lignin, and these enzymes are only produced when other more labile substrates are not available. The fungi invest more energy in the production of the enzymes to degrade lignin than they gain from the oxidation of the lignin molecule. Lignin is apparently degraded to provide access to labile compounds, such as cellulose and hemicellulose, protected within the interior of the lignin molecular structure. Phenolic compounds form highly resistant complexes with proteins in soil solution, which can slow the rates of both nitrogen mineralization and carbon oxidation.

The overall quality of plant residues can perhaps best be characterized by a combination of the indicators described in the previous paragraphs.

Poor-quality plant residues in general have a C:N ratio greater than 30:1, lignin contents greater than 20%, and polyphenol contents greater than 3%. This means that this material has a limited potential for microbial decomposition and subsequent mineralization of inorganic nutrients.

Soil Organic Matter Formation and Turnover

Soil organic matter is composed of living plant, animal, and microbial biomass, dead roots and other plant residues in various stages of decay, and soil humus. It is assumed that soil humus forms as a result of microbial activity but little is known about the exact mechanisms of humus formation. Humus contains numerous aromatic rings with attached phenolic and organic acid groups, but the remainder of the chemical structure is poorly understood. Some tentative models have been proposed for the complete molecular structure of humus but many scientists believe that a large portion of soil humus is amorphous, with no consistent molecular weight or repeating units of structure. Since humus comprises 60–80% of the total soil organic matter and has a relatively high nitrogen content, it constitutes a large reservoir of nitrogen in most ecosystems. Humus is tightly bound to the inorganic soil matrix and, as a result, decomposes extremely slowly because most of the structure is inaccessible to soil microbes and their enzymes.

The cycling of carbon in soils is the summation of the processes leading to rapid turnover of the majority of the plant litter in the surface soils, and processes leading to the slower production, accumulation, and turnover of humus deeper in the soil profile. Most decomposition occurs near the soil surface, where plant litter inputs are concentrated. Gravity carries most aboveground litter to the ground surface, where the initial decomposition and nutrient release occur. Since roots tend to grow into the surface soil to access these nutrients, most root litter is also produced primarily in the surface soil. There is a proportionally larger amount of aboveground plant litter carbon than root carbon that is lost as CO_2 during the transformation to stabilized soil organic matter. Soil mixing by animals and leaching of dissolved organic matter transfer surface carbon deeper into the soil profile. About half of the soil organic carbon is typically below 20 cm but deep-soil carbon is often older and more recalcitrant than surface soil carbon.

Soil disturbance increases soil organic carbon decomposition by exposing new soil surfaces to microbial attack and by increasing the oxygen content of the soil. The mechanism by which disturbance stimulates decomposition is basically the same at all scales. Disturbance disrupts soil aggregates so that the organic matter contained within them becomes more exposed to oxygen and microbial colonization. This disturbance effect explains why the introduction of European earthworms to the northeastern USA increased the decomposition rate of forest soils. It also explains why plowing causes rapid loss of organic matter from grassland and forest soils after conversion to agriculture, depleting the soil nutrient reserves and destroying soil structural integrity. However, it does not explain why disturbance results in a very rapid decline in soil organic matter in the first few years, followed by a much slower decline in subsequent years. In order to explain this observation, we must recognize that soil organic matter is a diverse mix of plant, animal, and microbial residues that vary in their susceptibility to microbial decomposition. This concept can be conceptualized by defining soil organic matter as a series of fractions that comprise a continuum based on decomposition rate. This can be extended further through the development of simulation models that mathematically describe the processes of organic matter formation and turnover. Models that best represent these processes include two to three organic matter pools that are kinetically defined with different turnover rates. Generally, these pools are conceptualized as one small pool with a rapid turnover rate (labile pool) and one to several pools of greater size and slower turnover rate (recalcitrant or stabilized pools).

Turnover rates of the labile pool of soil organic matter are measured in months to years. The size of the labile carbon pool can be increased very quickly through additions of fresh plant residues but carbon is also quickly lost from the labile pool when the soil is disturbed. The labile pool of organic matter is the first to be affected by soil disturbance and other changes in land management. Soil organic carbon is lost from this pool within the first few years after tillage of native grassland and forest soils. Intermediately labile pools of soil organic matter have turnover times typically measured in decades. These pools are an important source of mineralizable nitrogen and other plant nutrients, and provide a steady food source for soil microbial populations. Native grassland soils are rich in intermediately labile carbon, which accounts for the long-term fertility of these soils when they are brought under cultivation. Stabilized soil organic matter turns over every 100 to 1000 years and carbon losses from this pool are very gradual. This pool includes most of the soil humus and constitutes 60–90% of the soil organic matter in most soils. This pool is associated with the colloidal properties of humus and is involved in stabilization of soil structure.

Over the past few decades, considerable progress has been made towards understanding soil organic matter dynamics in temperate ecosystems. An important advance has been to integrate the structural and functional properties of soil organic matter and to relate them both to soil biological processes. Evidence suggests that soil organic matter turnover and consequently nutrient cycling and carbon storage potential are coupled to the formation and stabilization of soil aggregates. The most labile form of biologically active organic carbon is rapidly cycled in soil and represents a relatively small proportion of the total soil organic carbon (3–5%). The importance of the labile pool in maintaining soil fertility and controlling the supply of nutrients to plants has been extensively documented. However, the turnover time of the labile pool is so rapid that its direct effect on soil carbon storage potential may be limited. A significant proportion of biologically active organic carbon in soil is physically protected from decomposition through occlusion inside soil aggregates. Conceptually, this protected material comprises a pool of carbon that has an intermediate turnover time in soil of the order of 10–50 years and can represent up to 40% of the total soil carbon in native grassland systems. This intermediately labile pool of organic carbon is turning over at a timescale that is appropriate for assessing the impact of climate and land use change on ecosystem function, and is therefore the critical organic matter pool to monitor as a short-term relative indicator of longer term changes in soil carbon storage potential.

Further Reading

Brady NC and Weil RR (2002) *The Nature and Properties of Soils*, 13th edn. Upper Saddler River: Prentice Hall.

Chapin III FS, Matson PA, and Mooney HA (2002) *Principles of Terrestrial Ecosystem Ecology*. New York: Springer-Verlag.

Schlesinger WH (1997) *Biogeochemistry: An Analysis of Global Change*, 2nd edn, p. 588. San Diego: Academic Press.

Stevenson FJ (1986) *Cycles of Soil: Carbon, Nitrogen, Phosphorus, Sulfur, Micronutrients*. New York: John Wiley.

Tate III RL (1987) *Soil Organic Matter: Biological and Ecological Effects*. New York: John Wiley.

CARBON EMISSIONS AND SEQUESTRATION

K Paustian, Colorado State University, Fort Collins, CO, USA

Introduction

Soils hold more than half of the carbon contained in all the earth's terrestrial ecosystems, approximately 1500 billion tonnes in organic forms and 900 billion tonnes as inorganic carbonates (measured to a depth of 1 m). Thus soils are a large repository of the carbon (C) that cycles through the atmosphere, vegetation, and soil and interacts with groundwater and aquatic ecosystems. The annual turnover of soil C is approximately 50 Pg year^{-1}, where this amount of C on average enters the soil through primary production, balanced by an approximately equal amount of C emitted from soils through decomposition. On an annual basis, this represents one of the larger fluxes in the terrestrial C cycle. Many factors, including climate and human land use and management can alter the rates and relative balance of inputs and emissions from soils, which can affect whether soils are a net source or a net sink for carbon. Thus, there is both concern that the effects of climate and land-use change on soils will exacerbate the problem of increasing CO_2 in the atmosphere and a potential that, through better management, soils can play a part in mitigating increasing CO_2 levels.

Soils and Carbon-Cycling

Carbon is added to soils mainly through the death and subsequent deposition of biomass – leaves, stems, roots – from higher plants, which assimilate CO_2 from the atmosphere through photosynthesis. The organic compounds making up plant tissues, i.e., cellulose, hemicellulose, sugars, proteins, nucleic acids, lipids, waxes, and other compounds, are the food source for a vast array of heterotrophic organisms residing in the soil. These organisms include bacteria, fungi, and other microorganisms as primary decomposers of plant materials and a diverse collection of micro-, meso-, and macrofauna that consume the microorganisms and each other. Through the metabolism of the ingested organic materials, by everything from bacteria to earthworms, CO_2 is released in the process of respiration. (In oxygen-limited environments, carbon is also released in the form of methane (CH_4), through fermentation and anaerobic respiration.) Nearly all of the CO_2 released by respiration diffuses through the soil back to the atmosphere, although under certain conditions small amounts may enter groundwater as dissolved CO_2 or precipitate at lower soil depths to form secondary carbonates. Organic materials that are not fully metabolized (due to chemical or physical recalcitrance or inaccessibility to soils organisms), together with waste products and metabolites, can remain in soil and recombine into complex secondary compounds, generically referred to as humus. While much of the carbon entering the soil as plant (or other) biomass (collectively referred to as 'litter') is released back to the atmosphere as CO_2 within a year, some undecomposed material and especially secondary humic materials can reside in soils for hundreds to thousands of years. Sequestration refers to the buildup and storage of these C stocks in soils.

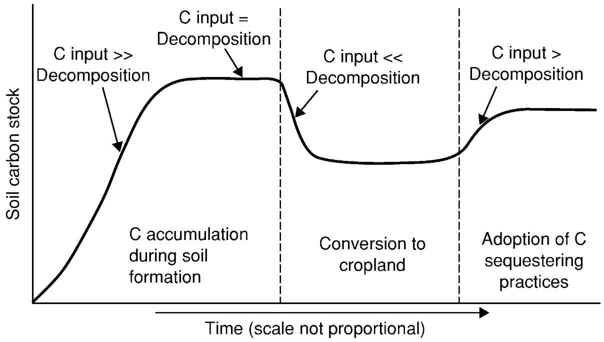


Figure 1 The buildup of soil C stocks initially during the course of soil formation and native vegetation succession, reaching an approximate equilibrium condition after several hundred to thousands of years. During this phase C inputs exceed C emissions from decomposition so that soil C stocks accumulate over time. Following human-induced disturbances such as deforestation and conversion to agriculture, soil C stocks can decline rapidly, as decomposition exceeds inputs, before leveling out toward a new, lower equilibrium state. Subsequent changes in land use (e.g., conversion back to perennial vegetation) or improved agricultural management practices can again change the balance in favor of C inputs exceeding emission, to increase soil C stocks.

Controls on Emissions and Sequestration

During the process of soil formation (pedogenesis), soils begin accumulating organic matter as vegetation becomes established and productivity and litter input increases. As ecosystems mature, soil organic C stocks tend to reach an approximate equilibrium or 'steady state,' when C inputs are balanced by C emissions. In simplified terms the rate of change in soil C stocks (C_s) can be expressed as:

$$\mathrm{d}C_s/\mathrm{d}t = I_c - kC_s, \qquad [1]$$

in which I_c is the rate of C addition from litter fall and k is a specific rate of organic matter mineralization, implying that losses (kC_s) are directly proportional to the amount of organic C present. It is also apparent from eqn [1] that, at a fixed level of C input (which is determined by the productivity and type of vegetation), soil C stocks will tend toward an equilibrium point where $C_s = I_c/k$, where there is no change in C stocks (i.e., $\mathrm{d}C_s/\mathrm{d}t = 0$). At equilibrium, soils are neither a source nor a sink of C to or from the atmosphere. However, it is easy to see that if conditions are altered and C input rates and/or the specific organic matter decay rate is altered, the soil can be moved toward a new equilibrium state, along a trajectory of either soil C losses or gains (Figure 1). Shifts in environmental conditions, such as changing climate, as well as alterations in land use and management can induce changes in soil carbon and hence rates of emission or sequestration of carbon.

The environmental factors that influence soil C stocks, emissions, and sequestration are relatively well understood, at least in terms of explaining broad geographic and climatic trends. In native ecosystems, C inputs to soils are largely determined by the amount and distribution of primary productivity and the type and life cycle of the vegetation. Thus, C inputs tend to increase from dry to wet regions and from cold to warm regions. Within a given climate zone, soil fertility can vary greatly, creating smaller-scale variability in C additions. Vegetation type affects the amount and timing of inputs. For example, perennial grasses tend to allocate a high proportion of total net productivity belowground to roots, which turn over and regenerate on relatively short time scales (months to years). The long-lived, woody biomass of forests means that relatively less of the annual productivity is returned to soils in the short term. Depending on forest disturbance regimes, large pulses of carbon can be added over longer time scales with tree fall and stand replacement.

The activity of decomposer organisms also reacts to a wide variety of factors including climate as well as soil physical and chemical properties and disturbance regimes. Decomposition rates change along the same climate gradients as primary productivity, increasing from dry to wet and cold to warm. However, the effects of climate on decomposition tend to 'lag' behind productivity. Soil C stocks initially increase as conditions become warmer and wetter, but further along the climate gradient decomposition responds relatively more to increasing temperature and

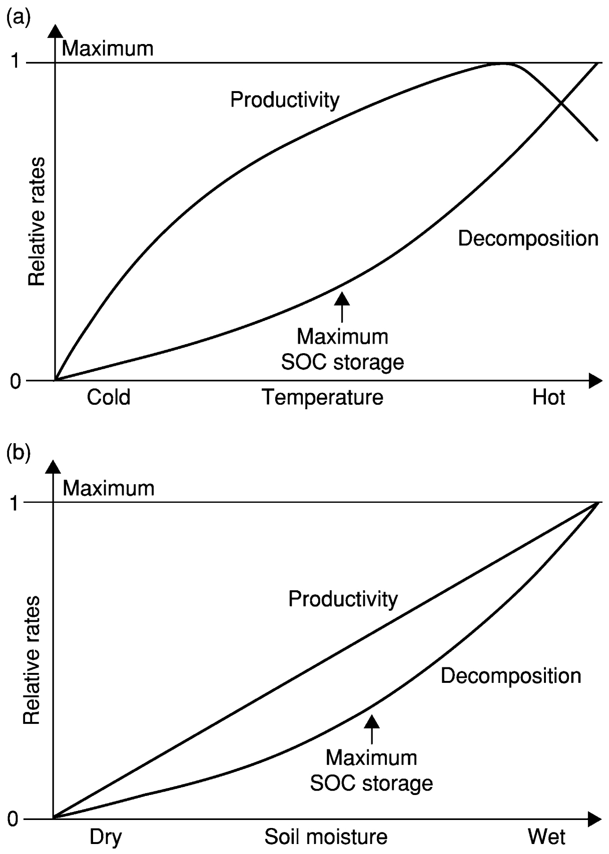


Figure 2 Climate controls on C additions and emissions through decomposition as determinants of soil C stocks. The relative responses of productivity (and C inputs) versus decomposition, across broad climate gradients relating to temperature (a) and moisture (b). Soil organic matter stocks tend to be greatest at points along the gradients where the greatest differences in the relative responses of productivity versus decomposition occur.

moisture, such that soil C stocks tend to level out or decline further along the gradients (Figure 2). Within this broad, climatically driven pattern of soil C stocks, productivity and decomposition rates (hence C storage and emissions) are influenced by soil chemistry, including nutrient availability and pH. Soil texture and mineralogy influence forces that bind organic matter to clay minerals, binding that tends to help stabilize organic matter in soils. Thus sandy soils tend to accumulate less organic matter than finer-textured soils containing more clay. Aeration provides one of the strongest controls on decomposition – in poorly drained (e.g., flooded) soils, decomposition rates are greatly reduced, which can lead to formation of so-called organic soils or peat that can extend to several meters in depth. These are some examples of the many environmental factors affecting soil C emission rates. The interaction of these factors, along with those controlling productivity and C inputs, can lead to a complex variability in soil C storage and C emissions at field to landscape scales.

Land Use and Land-Use Change

In managed ecosystems, human activities that manipulate vegetation- and soil-related processes such as land-use change and agricultural management can substantially alter C-cycling patterns of the ecosystem. Cropland represents a high degree of disturbance relative to native ecosystems, with wholesale vegetation removal (e.g., forest clearing) and replacement with annual crops and frequently intensive soil disturbance from tillage. The clearing of land for agriculture usually results in the rapid loss of organic matter from soils, often amounting to 20–40% of the total C in the top 30 cm of soil. Much of this loss occurs in the first years to decades following land conversion and plow out. On newly converted lands, management practices are often designed to exploit the nutrient reserves of soil, resulting in a ‘mining out’ of the organic matter. Historically, low crop productivity, harvest export, residue removal (including burning), intensive tillage, and bare fallowing have been the proximal causes of decreases in cropland soil C stocks. Wetland drainage and conversion to cropland, pasture, and forests are a special case in which decomposition and C emissions can be very high. Cultivated peat soils can lose as much as 20 Mg C ha^{-1} year^{-1} in tropical and subtropical areas and 5–15 Mg C ha^{-1} year^{-1} in temperate regions.

Over the past 200 years, agricultural areas have expanded greatly throughout North America and Europe, and especially in the past 100 years in tropical regions of Africa, Asia, and South America. Losses of C from these soils have amounted to as much as 50–100 Pg C since the introduction of agriculture. In many developing countries in the tropics, deforestation and land conversion to agriculture continues at a high rate, resulting in large net emissions of C. Global emissions from land-use conversion are estimated at 1–2 Pg C per year, mainly from biomass burning, but with soil C losses contributing also. In developed countries, in contrast, little new land is being converted to agriculture, and in parts of North America and Europe some cropland is being temporarily or permanently retired (i.e., government set-aside scheme), and urban and suburban development is the most significant land-use change involving agriculture. Thus, while the principles for managing land to reduce carbon emissions and sequester C in soils are similar across the globe, the most effective means and priority areas vary between developing and developed regions. In developing areas (predominantly in the tropics), this means reducing emissions by avoiding deforestation through systems providing alternatives to slash-and-burn and more productive cropping systems. In developed countries, emphasis is on alternative production systems that rebuild organic matter stocks that were lost in previous decades and in converting marginal croplands to conservation land uses based on perennial grassland or forest vegetation.

Management Practices for Soil C Sequestration

Conditions for C sequestration in soils are most favorable where C stocks have been previously depleted but where the productivity capacity remains high. Such a condition is typical for many cropland soils. In general, C sequestration will be favored under management systems that (1) minimize soil disturbance and erosion, (2) maximize the amount of plant residue return, and (3) maximize water- and nutrient-use efficiency of plant production. Although it may be impossible to optimize all these system attributes simultaneously, practices that effectively sequester C share one or more of these traits.

Decreasing tillage intensity, especially by using no-tillage practices, has been found to promote C sequestration in many soils. Intensive tillage tends to break down soil structure, disrupting aggregates (or 'soil crumbs') that help to protect organic matter from rapid decomposition. If soil structure has been substantially degraded by intensive tillage, the resulting reductions in productivity and C inputs can also contribute to lower soil C storage. In many long-term experiments, conversion to no-till practices has been shown to increase soil C storage at rates of 0.1–0.7 $\mathrm{Mg\,C\,ha^{-1}\,year^{-1}}$ or more, over periods of 20–30 years or more (Table 1). In semiarid regions, adoption of no-till can also enable higher crop C inputs by increasing water-use efficiency and allowing use of more intensified crop rotations, particularly elimination of bare (summer) fallows. Higher C inputs and more efficient water use helps to further increase soil C stocks under semiarid conditions.

Increasing the amount of residue returned to soil can be promoted through a variety of practices, including growing high residue-yielding crops, using hay crops in crop rotations, application of manure and biosolids, and by improved management of fertilizer, water, and pests. Most cropland soils show a clear response to increasing amounts of C return over time. The increase in soil C content is often directly proportional to the amount of C added to soil as crop residues, at least for a number of years or decades until a new equilibrium level is approached. Where production is water- or nutrient-limited, provision of these water and nutrient inputs can contribute to C sequestration. However, energy costs associated with manufacture and distribution of fertilizer, energy for irrigation pumping, as well as potential increased emissions of other greenhouse gases (N_2O and CH_4) from soil must be considered, for these costs may offset part or all the gains in C storage. The production-enhancing inputs (e.g., fertilizer, irrigation) are primarily used to meet objectives of food production and not as a means of mitigating greenhouse gas emissions. Practices promoting optimally efficient water and nutrient use, however, will probably have the greatest benefits in terms of decreased greenhouse gases.

Table 1 Representative soil C sequestration rates for some management practices in cropland, grazing lands, set-aside, and afforested cropland[a]

Management activities	*Sequestration ($\mathrm{Mg\,C\,ha^{-1}\,year^{-1}}$)*
Adoption of no-tillage practices on cropland	0.1–0.7
Conversion of cropland to perennial grassland (e.g., set-aside)	0.5–1.5
Afforestation of cropland	0.1–0.5
Grazing-land improvement	0.1–1.0

[a] Estimates based on data syntheses reported by: Conant RT, Paustian K, and Elliott ET (2001) Grassland management and conversion into grassland: effects on soil carbon. *Ecological Application* 11: 343–355; Paustian K, Andren O, Janzen H *et al.* (1997) Agricultural soil as a C sink to offset CO_2 emissions. *Soil Use and Management* 13: 230–244; Post WM and Kwon KC (2000) Soil carbon sequestration and land-use change: processes and potential. *Global Change Biology* 4: 67–79; West TO and Marland G (2002) A synthesis of carbon sequestration, carbon emissions, and net carbon flux in agriculture: comparing tillage practices in the United States. *Agriculture, Ecosystems and Environment* 91: 217–232.

Various management practices on grazing lands (pasture and rangeland) can increase soil C. On poorly managed grazing lands depleted of soil carbon, practices that increase production and C inputs can build up soil C. Such practices include improving grazing management, using improved species, sowing legumes, fertilizing, and irrigating (Table 1). As for annual crop systems, management of grazing lands for greenhouse gas mitigation needs to consider the net effects of practices on all greenhouse gases. For example, high nitrogen (N) fertilization rates in intensively managed pastures may cause large N_2O emissions that wipe out benefits from carbon sequestration, whereas phosphorus (P) fertilization and/or moderate N in highly P- or N-limited systems can yield large gains in productivity and C sequestration with little increase in N_2O emissions. Improvements in pasture productivity and forage quality through improved management can sequester C and also reduce methane emissions from grazing livestock by increasing forage quality.

Table 2 Global, regional and country estimates of soil C sequestration potential

Area	*Land-use sector*	*Sequestration (Tg C year^{-1})*	*Source*
Global	Improved management of cropland	260	IPCC (2000)[a]
	Improved management of grazing land	470	
	Improved management of forest land	700	
	Agroforestry	630	
USA	Cropland	83	Sperow *et al.* (2003)[b]
USA	Cropland	75–174	Lal *et al.* (1999)[c]
USA	Grazing land (excluding cropland set-aside to grassland)	7–75	Follett *et al.* (2001)[d]
UK	Cropland	6	Smith *et al.* (2000)[e]
EU	Increased manure application to arable land	13	Smith *et al.* (1997)[f]
	Sewage sludge application to arable land	8	
	Increased straw application to cereal land	14	
	Afforestation of 30% surplus arable land	50	
	Conversion of arable land to ley-arable farming	40	

[a] International Panel on Climate Change (2000) *Land Use, Land Use Change, and Forestry. Intergovernmental Panel on Climate Change Special Report*. Cambridge, UK: Cambridge University Press.

[b] Sperow M, Eve M, and Paustian K (2003) Potential soil C sequestration on US agricultural soils. *Climatic Change* 57: 319–339.

[c] Lal R, Follett RF, Kimble J, and Cole CV (1999) Managing US cropland to sequester carbon in soil. *Journal of Soil and Water Conservation* 54: 374–381.

[d] Follett RF, Kimble JM, and Lal R (eds) (2001) *The Potential of US Grazing Lands to Sequester Carbon and Mitigate the Greenhouse Effect*. Boca Raton, FL: Lewis Publishers.

[e] Smith P, Milne R, Powlson DS *et al.* (2000) Revised estimates of the carbon mitigation potential of UK agricultural land. *Soil Use and Management* 16: 293–295.

[f] Smith P, Powlson DS, Glendining MJ, and Smith JU (1997) Potential for carbon sequestration in European soils: preliminary estimates for five scenarios using results from long-term experiments. *Global Change Biology* 4: 67–79.

Restoring degraded soils and converting marginal cropland to perennial vegetation such as grassland and forest can yield relatively high rates of soil C accrual. Elimination of soil disturbance, reduced erosion and greater proportion of primary productivity going back to the soil contribute to soil C increases. Average rates of soil C gain under afforestation (0.1–0.5 Mg C ha^{-1} year^{-1}) have been reported and somewhat higher rates may occur with conversion to grassland vegetation (Table 1).

Regional and Country Estimates of C-Sequestration Potential

Estimates of the potential for soil C sequestration on cropland and grazing lands have been made, globally and for some countries and/or regions, particularly in Europe and North America (Table 2). A global survey by the Intergovernmental Panel on Climate Change has estimated that as much as 2.3 Pg year^{-1} could be sequestered through widespread adoption of improved management and land-use practices on cropland, grazing lands, and forests. Soil C sequestration has in principle been included as an allowable option within the framework of the Kyoto Protocol for countries to meet part of their emission-reduction targets. However, there are many outstanding questions regarding quantification and accounting procedures, C credit trading, and issues such as permanence and leakage that need to be addressed before soil C sequestration is included as a policy option.

To date, most estimates of potentials have been based on highly aggregated data and thus have considerable uncertainty, which has not been formally assessed. Moreover, these biophysical potentials do not consider the economic factors that limit the adoption of C sequestering practices. On the other hand, the development of new technology to specifically enhance C sequestration rates and thus increase biophysical potentials is in its infancy. Thus, the amounts of C sequestration that can be attained are still poorly known and will depend to a large degree on economic and policy incentives.

In addition to C sequestration, increasing soil organic matter levels generally enhances several biological, chemical, and physical attributes of soils. These improvements include enhanced water storage capacity, increased water infiltration, reduced runoff (and erosion), increased soil buffering capacity, and increased storage of essential plant nutrients. Thus, promoting practices for increasing soil organic matter as part of strategies for mitigating increasing greenhouse gas concentrations can be highly beneficial for improving soil quality and the sustainability of managed lands. If these additional benefits can be 'bundled' with C sequestration, incentives for implementing C sequestration policies could probably be increased.

List of Technical Nomenclature

1 billion tonnes = 1 petagram (Pg) = 10^{15} grams (g)

1 million tonnes = 1 teragram (Tg) = 10^{12} grams (g)

See also: **Carbon Cycle in Soils:** Dynamics and Management; **Climate Change Impacts**; **Greenhouse Gas Emissions**; **Sustainable Soil and Land Management**

Further Reading

Antle JM, Capalbo SM, Mooney S, Elliott ET, and Paustian K (2002) Economic analysis of agricultural soil carbon sequestration: an integrated assessment approach. *Journal of Agricultural and Resource Economics* 26: 344–367.

Conant RT, Paustian K, and Elliott ET (2001) Grassland management and conversion into grassland: effects on soil carbon. *Ecological Application* 11: 343–355.

Follett RF, Kimble JM, and Lal R (eds) (2001) *The Potential of U.S. Grazing Lands to Sequester Carbon and Mitigate the Greenhouse Effect.* Boca Raton, FL: Lewis Publishers.

Intergovernmental Panel on Climate Change (2000) *Land Use, Land Use Change, and Forestry. Intergovernmental Panel on Climate Change Special Report.* Cambridge, UK: Cambridge University Press.

Kimble JM, Lal R, and Follett RF (eds) (2002) *Agricultural Practices and Policies for Carbon Sequestration in Soil.* Boca Raton, FL: Lewis Publishers.

Lal R, Follett RF, Kimble J, and Cole CV (1999) Managing U.S. cropland to sequester carbon in soil. *Journal of Soil and Water Conservation* 54: 374–381.

Paustian K, Collins HP, and Paul EA (1997) Management controls on soil carbon. In: Paul EA, Paustian K, Elliott ET, and Cole CV (eds) *Soil Organic Matter in Temperate Agroecosystems: Long-term Experiments in North America*, pp. 15–49. Boca Raton, FL: CRC Press.

Paustian K, Andren O, Janzen H *et al.* (1997) Agricultural soil as a C sink to offset CO_2 emissions. *Soil Use and Management* 13: 230–244.

Paustian K, Cole CV, Sauerbeck D, and Sampson N (1998) CO_2 mitigation by agriculture: an overview. *Climatic Change* 40: 135–162.

Paustian K, Elliott ET, Six J, and Hunt HW (2000) Management options for reducing CO_2 emissions from agricultural soils. *Biogeochemistry* 48: 147–163.

Post WM and Kwon KC (2000) Soil carbon sequestration and land-use change: processes and potential. *Global Change Biology* 6: 317–327.

Smith P, Powlson DS, Glendining MJ, and Smith JU (1997) Potential for carbon sequestration in European soils: preliminary estimates for five scenarios using results from long-term experiments. *Global Change Biology* 4: 67–79.

Smith P, Milne R, Powlson DS *et al.* (2000) Revised estimates of the carbon mitigation potential of UK agricultural land. *Soil Use and Management* 16: 293–295.

Sperow M, Eve M, and Paustian K (2003) Potential soil C sequestration on U.S. agricultural soils. *Climatic Change* 57: 319–339.

West TO and Marland G (2002) A synthesis of carbon sequestration, carbon emissions, and net carbon flux in agriculture: comparing tillage practices in the United States. *Agriculture, Ecosystems and Environment* 91: 217–232.

CATION EXCHANGE

L M McDonald, West Virginia University, Morgantown, WV, USA
V P Evangelou[†], Iowa State University, Ames, IA, USA
M A Chappell, Iowa State University, Ames, IA, USA

Introduction

Soil chemistry is the study of how the elements and their compounds are distributed between and within the three principal phases that comprise the soil, the solid, liquid, and gaseous phases. By studying cation exchange reactions, we seek to understand and predict how positively charged ions are distributed between the solid and liquid phases. Because this distribution plays an important role in the flocculation and dispersion of soils and suspended sediments, the availability and transport of nutrient and contaminant cations, and the regulation of soil acidity, cation exchange is an essential and unifying concept in soil science.

J.T. Way is credited with the first systematic studies of cation exchange reactions in soils. Building on the observation by H.S. Thompson that $CaSO_4$ was leached out when $(NH_4)_2SO_4$ was applied to soil columns, Way established that equivalent amounts of Ca^{2+} were removed from soils when leached with NH_4^+, K^+, and Na^+. Since that time, a significant amount of work has been done to apply the concept of cation exchange to model the availability of nutrient ions in soils, particularly K^+, NH_4^+, and Ca^{2+} exchange. The relative concentration of sodium on soil surfaces directly affects the degree of colloid dispersion and therefore the formation of soil crusts and soil hydraulic conductivity. Therefore, principles of Na^{2+}–Ca^{2+} exchange have been used to reclaim and manage saline-sodic soils. Considerable progress has been made into the effects of sodium and solution composition, pH, ionic strength, and mineralogy on

[†] Deceased.

the dispersive properties of soil. Aluminum–calcium exchange reactions have been used to study the effects of acid rain and other anthropogenic inputs on soil acidification. Cation exchange reactions have been, and continue to be, an active area of soil chemistry research, as evidenced by the numerous research articles that have been published on the subject. Several excellent reviews are available, including those with details on the history, experimental methods, and kinetic aspects of cation exchange.

Cation exchange occurs in soils because of two general phenomena that are simply described and easily understood. First, except for the very acid and/or highly weathered, most soils have a net negative charge. Second, all natural macroscopic systems are electrically neutral. As a result, the net negative charge of a soil particle must be balanced by a charge-equivalent number of cations at or near the particle surface (interfacial region). When salts are added to the soil by natural mineral weathering or organic matter decomposition processes, in irrigation water, as fertilizer, acid rain, or other anthropogenic input, some fraction of the added ions will accumulate in the interfacial region and displace a charge-equivalent number of ions from the interfacial region into the soil solution. The simplicity of these concepts belies the complexity of the cation exchange process when applied to a system as varied and heterogeneous as soil.

Cation Exchange Capacity

The negative charge in soils comes from permanently charged clays or variably charged minerals and soil organic matter. Permanent charge is a consequence of isomorphic substitution (e.g., Al^{3+} for Si^{4+} or Mg^{2+} for Al^{3+}) within the clay lattice and results in negative charge that is independent of soil-solution properties. Variable negative charge results from the deprotonation of surface functional groups on soil organic matter and the surfaces of oxides and hydroxides. The magnitude of the variable charge depends on soil-solution pH, ionic strength, temperature, and the presence of other specifically adsorbing (potential determining) ions. Cation exchange capacity (CEC) can be defined as either the total quantity of this negative charge per unit mass of soil or as the sum total of the exchangeable cations neutralizing this charge per unit mass of soil, depending on whether anion exclusion is to be considered part of the definition. It is expressed in units of centimoles of cation charge per kilogram ($\text{cmol}_c\,\text{kg}^{-1}$) or milliequivalents of charge per 100 g soil ($\text{mEq}\,100\,\text{g}^{-1}$). Because of the effects of ionic strength and pH on the surface charge of variable-charge minerals, and complications from soluble and sparingly soluble salts and carbonates, CEC is explicitly defined by the procedure used in the determination. Therefore, a meaningful determination of CEC should consider the uses for which CEC is being determined and the unique properties of the soils being measured, including potential complications from salts, carbonates, pH, and ionic strength.

The basic procedure for determining CEC is to saturate the soil with an 'index' cation (e.g., NH_4^+, Ba^{2+}) and then either (1) determine the concentration of Na^+, K^+, Mg^{2+}, Ca^{2+}, and Al^{3+} in the supernatant, or (2) determine the concentration of the index cation in the supernatant after it has been displaced by some other cation (e.g., K^+, Mg^{2+}). Supernatant concentrations are converted to units of charge per unit mass of soil by considering the valence of the index cation and the solid-to-solution ratio used in the procedure.

This procedure can be performed with unbuffered salts (e.g., $0.2\,\text{mol}\,\text{l}^{-1}$ ammonium chloride), pH-adjusted salts (e.g., calcium chloride at pH 8.2 or ammonium acetate at pH 7), or using the compulsive exchange technique. Unbuffered salts are preferred when the CEC at the native soil pH is desired (effective cation exchange capacity, ECEC). Buffered or otherwise pH-adjusted salt solutions are used for calcareous and gypsiferous soils, or soils otherwise containing soluble or sparingly soluble salts and carbonates. It is also used when a single procedure is needed to characterize a wide variety of soils so that lime and fertilizer practices will not bias the results for soils with variable-charge mineralogy. The compulsive exchange technique is the most time-consuming of the three but has the advantage that it accounts for ionic strength effects, an important consideration for highly weathered soils. Detailed procedures for these methods can be found in any reference work on methods of soil analysis.

Because CEC is in an important property of soils for many disciplines, it has been determined for a wide variety of soils and soil minerals. The CEC of clay minerals is proportional to the amount of accessible layer charge: low CEC on talc but high CEC on vermiculite and montmorillonite (Table 1). The muscovites have comparatively low CEC, because the high layer charge is satisfied by nonexchangeable K (Table 1). Oxides and hydroxides have generally low CEC, especially at the pH of most agricultural soils. Because of the importance of clays as exchange sites, CEC increases as the fine soil fraction increases (Table 1). Histisols have the highest CEC of any soil order because of the large contribution from soil organic matter (Table 1). For the other soil orders, CEC decreases as the amount of weathering increases (Table 1) due to replacement of 2:1 permanent charge

Table 1 Cation exchange capacity of common soil minerals, textural classes, and soil orders

Clay mineral[a]		*Textural class*[b,c]		*Soil order*[b,c]	
Mineral	*CEC ($cmol_c\,kg^{-1}$)*	*Texture*	*CEC ($cmol_c\,kg^{-1}$)*	*Order*	*CEC ($cmol_c\,kg^{-1}$)*
Talc	<1	Loamy sand	3.4	Ultisol	3.5
Kaolinite	2–15	Sandy loam	5.6	Alfisol	9.0
Biotite	10–40	Fine sandy loam	7.6	Spodosol	9.3
Chlorite	10–40	Silt loam	13.4	Entisol	11.6
Halloysite	10–40	Loam	14.2	Inceptisol	14.6
Muscovite	10–40	Clay loam	20.3	Aridisol	15.2
Dioctahedral vermiculite	10–50	Silty clay loam	27.5	Mollisol	18.7
Montmorillonite	80–150	Silty clay	32.5	Vertisol	35.6
Trioctahedral vermiculite	100–200	Clay	36.6	Histisol	128.0
Allophane	5–350				

[a] Reproduced from: Sparks DL (1995) *Environmental Soil Chemistry*, p. 46. San Diego, CA: Academic Press.
[b] Holmgren GGS, Meyer MW, Chaney RL, and Daniels RB (1993) Cadmium, lead, zinc, copper and nickel in agricultural soils of the United States of America. *Journal of Environmental Quality* 22: 335–348.
[c] Geometric mean for US soils.
CEC, cation exchange capacity.

minerals with variably charged 1:1 clays, iron, and aluminum oxides.

Qualitative Description of Cation Exchange

The first issue to be decided when describing cation exchange reactions is what is to be considered an exchange reaction. Historically, reactions that result in the formation of new solid phases or alterations of surface properties are not considered cation exchange reactions, and 'exchange' is taken to mean reversible adsorption as an outer-sphere or diffuse layer complex; for example:

$$\mathrm{ExMg} + \mathrm{Ca}^{2+}(\mathrm{aq}) \leftrightarrow \mathrm{ExCa} + \mathrm{Mg}^{2+}(\mathrm{aq}) \qquad [1]$$

where Ex indicates a solid-phase exchanging surface with a charge −2. However, because all ions have some potential to form inner-sphere complexes with surfaces, this distinction is as much conceptual as physical.

Cation exchange reactions are classified by (1) the number of ions considered, e.g., binary, ternary, or quaternary, and (2) the valence of the ions, either homovalent (Na^+–K^+, Mg^{2+}–Ca^{2+}, etc.) or heterovalent (Na^+–Ca^{2+}, K^+–Al^{3+}, Ca^{2+}–Al^{3+}, etc.). For a generalized, binary, homovalent exchange reaction such as eqn 1, an equilibrium expression can be written:

$$K_{\mathrm{Ca-Mg}} = \frac{(\mathrm{ExCa})}{(\mathrm{ExMg})} \cdot \frac{(\mathrm{Mg}^{2+})}{(\mathrm{Ca}^{2+})} \qquad [2a]$$

$$K_{\mathrm{Ca-Mg}} = \frac{\gamma_{\mathrm{ExCa}} \cdot [\mathrm{ExCa}]}{\gamma_{\mathrm{ExMg}} \cdot [\mathrm{ExMg}]} \cdot \frac{\gamma_{\mathrm{Mg}}[\mathrm{Mg}^{2+}]}{\gamma_{\mathrm{Ca}} \cdot [\mathrm{Ca}^{2+}]} \qquad [2b]$$

where the subscript on K indicates that Ca^{2+} in solution is replacing Mg^{2+} on a surface, the parentheses indicate activities, the brackets indicate concentration (solution, moles per liter; exchanger, moles per kilogram or centimoles of cation charge per kilogram) and γ represents activity coefficients. A generalized, binary heterovalent reaction can be written:

$$(1/2)\mathrm{Ex_2Ca} + \mathrm{Na}^+ \leftrightarrow \mathrm{ExNa} + (1/2)\mathrm{Ca}^{2+} \qquad [3a]$$

$$K_{\mathrm{Na-Ca}} = \frac{\gamma_{\mathrm{ExNa}} \cdot [\mathrm{ExNa}]}{(\gamma_{\mathrm{Ex_2Ca}} \cdot [\mathrm{Ex_2Ca}])^{0.5}} \cdot \frac{(\gamma_{\mathrm{Ca}} \cdot [\mathrm{Ca}^{2+}])^{0.5}}{\gamma_{\mathrm{Na}} \cdot [\mathrm{Na}^+]} \qquad [3b]$$

where Ex is now defined as an exchanger with charge −1. Because of the difficulty of defining surface activity coefficients (γ_{ExCa}, γ_{ExMg} in eqn [2b] and γ_{ExNa}, $\gamma_{\mathrm{Ex_2Ca}}$ in eqn [3b]), the left-hand side of eqn [2b] and [3b] is called a selectivity coefficient rather than an equilibrium constant.

Selectivity is the tendency for a charged surface preferentially to adsorb one ion over another. If the exchanging surface in eqn 1 'preferred' Ca^{2+}, then the equilibrium would lie to the right and if the surface preferred Mg^{2+} then the equilibrium would lie to the left. Soil minerals (and soil organic matter) are not indifferent to the suite of ions in the soil solution and exhibit a preference for certain of these ions. This preference is a consequence of the properties of those ions, the bulk and surface properties of the soil particles, and the extent to which both of these are modified by the soil solution.

Properties of Ions

From Coulomb's Law, the interaction energy between a charged surface and an oppositely charged solute is proportional to the solute's charge and inversely

Table 2 Ionic and hydrated radii of some Group I and II cations, and ammonium

Cation	*Ionic radii (nm)*	*Hydrated radii (nm)*
Li^+	0.068	1.003
Na^+	0.098	0.790
K^+	0.133	0.532
NH_4^+	0.143	0.537
Rb^+	0.149	0.509
Cs^+	0.165	0.505
Mg^{2+}	0.089	1.080
Ca^{2+}	0.117	0.96
Sr^{2+}	0.134	0.96
Ba^{2+}	0.149	0.88

proportional to the solvated radius. Therefore, in the absence of any surface-specific effects, cation (group I and II) preference follows the lyotropic series:

$$Ba^{2+} > Sr^{2+} > Ca^{2+} > Mg^{2+} > Cs^+ > Rb^+ > K^+ > Na^+ > Li^+ \qquad [4]$$

Because of periodic trends in polarizability and hydrated radii (Table 2), negatively charged surfaces prefer divalent cations over monovalent cations, and (within a periodic group) the more weakly hydrated Ba^{2+} over Mg^{2+}, and Cs^+ over Li^+. The lyotropic series assumes that exchangeable cations retain their hydration spheres (i.e., adsorbed as outer-sphere complexes) and that free metal ion activities are being compared. That is, preference for Ca^{2+} over Na^+ (on a concentration basis) will decrease when sulfate is the dominant anion (compared with the chloride system) because Ca has more potential to form ion pairs with sulfate than does Na. Typically this difference will disappear when ion activities and ion pairing are accounted for. If it does not, it may suggest either competition from charged ion pairs (e.g., $CaCl^+$) or surface modification due to inner-sphere complexation. Competition effects from charged ion pairs can be significant, especially when one of the exchanging cations is Pb^{2+} or Cd^{2+} in chloride-containing solutions.

Solution-phase single-ion activity coefficients can be calculated using either the Debye–Huckel equation or the Davies equation. The advantage of the Davies equation is that it has no ion-specific parameters and so can be readily incorporated into equilibrium-speciation computer models. This is, however, also a disadvantage of the Davies equation in that it would calculate the same activity coefficient for two different cations of the same valence (e.g., Ca^{2+} and Cu^{2+}). Both the Debye–Huckel and Davies equations are widely used and accepted by the soil chemistry research community.

Properties of Surfaces

Surface-specific effects are always present in soils and so the lyotropic series is not a universal relationship. Surfaces that act as strong bases will prefer the stronger Lewis acid when cations have the same valence. Surfaces that exhibit weak acid behavior (higher pK_α) show stronger preference for heavy metals than hard (Group I and II) metals compared with surfaces with stronger acid behavior (low pK_α). For example, illite or kaolinite shows stronger preference for Cu^{2+} or Cd^{2+} than montmorillonite. The anion with the highest potential to form surface complexes controls the sorption potential of heavy metals. For example, Ni^{2+} in the presence of $Ca(NO_3)_2$ exhibits greater adsorption potential than Ni^{2+} in the presence of $CaSO_4$, because sulfate will produce surface sites with high specificity for Ca^{2+}. A surface that has the potential to form inner-sphere complexes with certain monovalent cations shows stronger preference for these cations than any other cation. For example, vermiculite exhibits more preference for K^+ or NH_4^+ than for Na^+ or Ca^{2+}. For surfaces that do not have the potential to form inner-sphere complexes, preference depends on the magnitude of the surface electrical potential. For example, a surface with high electrical potential shows higher preference for divalent cations in the presence of a monovalent cation than a surface with low electrical potential. Surfaces that have the potential to undergo conformational changes, e.g., humic acids, prefer higher-valence cations (e.g., Ca^{2+}) over lower-valence cations (e.g., K^+).

Properties of Solvents

Little work has been done on cation exchange reactions in solutions other than water. Whether in nonaqueous solvents (benzene, trichloroethylene (TCE)) or in mixtures of water and miscible solvents (low-weight alcohols, acetonitrile, etc.), low-dielectric solvents have the potential to affect the properties of ions and exchanging surfaces. Low-dielectric solutions are likely to promote ion-pairing, especially for polyvalent ions, due to coulombic attractions. The effects of nonaqueous solvents on the surface properties of minerals have been variable, with reports of decreased (rutile) and increased (goethite) surface-charge density. The most pronounced effect of low-dielectric solvents is on the interlayer spacing of 2:1 minerals. At some critical solvent concentration, dielectric saturation occurs and clay interlayers collapse. The solvent concentration at which this collapse occurs is proportional to the expected complexing ability of the solvent with the interlayer cation.

Quantitative Description of Cation Exchange

To determine cation exchange selectivity coefficients, the terms on the right-hand side of eqns [2b] and [3b] are determined over a range of equilibrium solution-phase cation activities. The solution-phase and solid-phase exchangeable cation concentrations are readily determined experimentally, but the activity coefficients must be calculated and so require one or more assumptions. Solution-phase activity coefficients can be calculated using the Davies or Debye–Huckel equation, but there are no analogous universal equations to calculate solid-phase exchangeable cation activity coefficients, and so some convention must be accepted. The simplest of these is the Kerr convention (Table 3), which uses exchanger concentration units of moles per kilogram and assumes that the two solid-phase exchangeable cation activity coefficients cancel from eqns [2b] (γ_{ExCa}, γ_{ExMg}) and [3b] (γ_{ExNa}, $\gamma_{\text{Ex}_2\text{Ca}}$). While this assumption is approximately true for the homovalent Ca–Mg system Kerr studied (eqn [2]), it would not be true generally, especially for heterovalent exchange (eqn [3]). In the Vanselow convention (Table 3), exchanger activities are assumed equal to their mole fractions (the ideal solid solution theory), so that for eqn 3:

$$(\text{ExNa}) = \text{X}_{\text{Na}} = \frac{[\text{ExNa}]}{[\text{ExNa}] + [\text{Ex}_2\text{Ca}]} \qquad [5a]$$

$$(\text{Ex}_2\text{Ca}) = \text{X}_{\text{Ca}} = \frac{[\text{Ex}_2\text{Ca}]}{[\text{ExNa}] + [\text{Ex}_2\text{Ca}]} \qquad [5b]$$

Note that because the solid-phase exchangeable cation concentrations are expressed in moles per kilogram the denominator in eqn [5] is not a constant even though the CEC is constant (sum of Na and Ca on a charge-equivalent basis). For the Gapon convention (Table 3), it is assumed that solid-phase exchangeable cation activities are equal to concentrations when concentrations are expressed on a charge-equivalent basis. To do this, eqn [3a] needs to be rewritten as:

$$\text{ExCa}_{0.5} + \text{Na}^+ \leftrightarrow \text{ExNa} + (1/2)\text{Ca}^{2+} \qquad [6a]$$

where Ex still indicates an exchanger with an average charge of −1, and

$$K_{\text{Na-Ca}} = \frac{[\text{ExNa}]}{[\text{ExCa}_{0.5}]} \cdot \frac{[\text{Ca}^{2+}]^{0.5} \cdot \gamma_{\text{Ca}}^{0.5}}{[\text{Na}^+] \cdot \gamma_{\text{Na}}} \qquad [6b]$$

where only the first term on the right-hand side of eqn [3b] has been changed. Note that the distinction between the conventions is relevant only for heterovalent exchange. For homovalent exchange, all the conventions in Table 3 are equivalent to the Kerr convention. Although the Kerr, Vanselow, and Gapon conventions are the most widely used, there are others, including Gaines–Thomas, Krishnamoorthy–Overstreet, and Rothmund–Kornfeld. Details and references for these models can be found in most soil chemistry texts.

Homovalent Binary Exchange

The homovalent cation exchange isotherm is obtained by solving eqn [2b] for exchangeable Ca. Given that

$$((1/2)\text{CEC}) = \text{ExMg} + \text{ExCa} \qquad [7a]$$

when CEC has the units centimoles of cation charge per kilogram and ExMg and ExCa have units centimoles per kilogram, we can write:

$$\text{CEC} = \text{ExMg}_{0.5} + \text{ExCa}_{0.5} \qquad [7b]$$

to indicate that all terms have units centimoles of cation charge per kilogram. Note that the exchanger is still defined as having a charge −2. It is assumed that other cations, e.g., exchangeable K^+ and H^+, are present in negligible quantities and do not interfere

Table 3 Selectivity symbols and expressions for the three most commonly used cation exchange conventions

Selectivity coefficient	*Symbol*	*Homovalent exchange*[a]	*Heterovalent exchange*[a]
Kerr[b]	K_K	$\frac{[\text{ExCa}][\text{Mg}^{2+}]}{[\text{ExMg}][\text{Ca}^{2+}]}$	$\frac{[\text{ExNa}][\text{Ca}^{2+}]^{0.5}}{[\text{Ex}_2\text{Ca}]^{0.5}[\text{Na}^+]}$
Vanselow[b,c]	K_V	$\frac{(\text{ExCa})(\text{Mg}^{2+})}{(\text{ExMg})(\text{Ca}^{2+})}$	$\frac{(\text{ExNa})(\text{Ca}^{2+})^{0.5}}{(\text{Ex}_2\text{Ca})^{0.5}(\text{Na}^+)}$
Gapon[d]	K_G	$\frac{[\text{ExCa}][\text{Mg}^{2+}]}{[\text{ExMg}][\text{Ca}^{2+}]}$	$\frac{[\text{ExNa}][\text{Ca}^{2+}]^{0.5}}{[\text{ExCa}_{0.5}][\text{Na}^+]}$

[a] Square brackets indicate concentration, parentheses indicate activities for both the solution phase and the exchanger.
[b] For the exchange reaction as written in eqns [1] or [3a] solution-phase units, moles per liter; exchanger units, moles per kilogram.
[c] Solution-phase activity coefficients defined using either the Debye–Huckel or Davies equation. Exchanger activities defined using eqns [5a] and [5b].
[d] For the exchange reaction as written in eqn [6a]. Solution-phase units, moles per liter; exchanger units, centimoles of cation charge per kilogram.

with the exchange reaction being studied, or H^+ is tightly bound to the charged surface, giving rise only to pH-dependent charge. Using eqn [7], and the definition for the calcium activity ratio (AR_{Ca}) or the calcium concentration ratio (CR_{Ca}, calcium adsorption ratio):

$$AR_{Ca} = \frac{(Ca^{2+})}{(Mg^{2+})} \quad [8a]$$

$$CR_{Ca} = \frac{[Ca^{2+}]}{[Mg^{2+}]} \quad [8b]$$

where the square brackets indicate solution-phase units moles per liter or millimoles per liter, eqn [2b] becomes

$$ExCa_{0.5} = \frac{K_{Ca-Mg} \cdot CEC \cdot AR_{Ca}}{1 + AR_{Ca} \cdot K_{Ca-Mg}} \quad [9]$$

Equation [7b] essentially invokes the Gapon convention so that $\gamma_{ExCa} = \gamma_{ExMg} = 1$ in eqn [2b]. A plot of $ExCa_{0.5}$ versus AR_{Ca} will produce a curvilinear line, asymptotically approaching CEC. The path of the line from $ExCa_{0.5} = 0$ to $ExCa_{0.5} = CEC$ depends on K_{Ca-Mg} and CEC (**Figure 1**).

An important component of homovalent exchange is the magnitude of the exchange selectivity coefficient. With some exceptions, homovalent cation exchange reactions in soils or soil minerals exhibit a selectivity coefficient of approximately 1. This indicates that the surface does not show any particular adsorption preference for either of the two homovalent cations.

Cation preference can be demonstrated using fractional isotherms. Fractional isotherms are plots of equivalent fraction on the exchange phase versus equivalent fraction in the solution phase. To determine the nonpreference isotherm, it follows from eqn [2b] that if there is no preference for Ca over Mg, then $K_{Ca-Mg} = 1$, so that eqn [9] can be rewritten:

$$(ExCa_{0.5}/CEC) = \frac{CR_{Ca}}{1 + CR_{Ca}} \quad [10a]$$

This is the diagonal line in **Figure 2** when CR_{Ca} is converted to a mole or equivalent fraction. An experimentally determined isotherm above the diagonal line reveals that the surface prefers Ca^{2+} ($K_{Ca-Mg} > 1$), while any line below the nonpreference line reveals that the surface prefers Mg^{2+} ($K_{Ca-Mg} < 1$). Stated another way, for a surface that prefers Ca^{2+}, a relatively small increase in the concentration of Ca^{2+} in solution results in a proportionally large concentration of Ca^{2+} on the surface. Introducing solution single-ion activities, γ_i, into eqn [10a]:

$$ExCa_{0.5}/CEC = \frac{([Ca^{2+}]/[Mg^{2+}]) \cdot \left(\gamma_{Ca}/\gamma_{Mg}\right)}{1 + ([Ca^{2+}]/[Mg^{2+}]) \cdot \left(\gamma_{Ca}/\gamma_{Mg}\right)} \quad [10b]$$

it can be shown that the homovalent nonpreference isotherm is independent of ionic strength because γ_{Ca} is nearly equal to γ_{Mg}.

Heterovalent Binary Exchange

The equation that is most commonly used to describe heterovalent cation exchange is the Gapon equation. For Na^+–Ca^{2+} exchange, eqn [8b] can be rearranged to solve for ExNa:

$$ExNa = \frac{K_G \cdot CEC \cdot SAR}{1 + K_G \cdot SAR} \quad [11a]$$

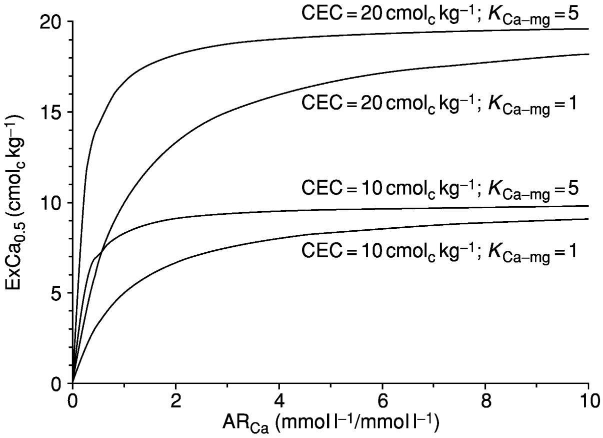


Figure 1 Effect of selectivity coefficient (K_{Ca-Mg}) and cation exchange capacity (CEC) on homovalent cation exchange isotherm shape. Plotted using eqn [9].

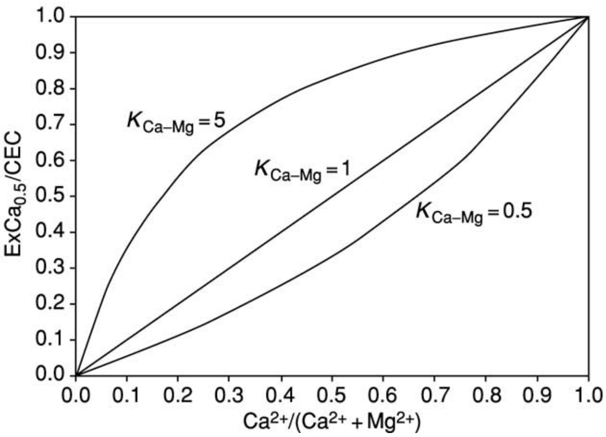


Figure 2 Calcium–magnesium fractional isotherms for three values of the homovalent cation exchange selectivity coefficient (K_{Ca-Mg}). Plotted using eqn [10a].

where K_G is the Gapon selectivity coefficient, SAR is the sodium adsorption ratio:

$$\mathrm{SAR} = \frac{[\mathrm{Na}^+]}{[\mathrm{Ca}^{2+}]^{0.5}} \qquad [11b]$$

The square brackets denote concentration in moles per liter or millimoles per liter in the solution phase, and

$$\mathrm{CEC} = \mathrm{ExCa}_{0.5} + \mathrm{ExNa} \qquad [11c]$$

with units centimoles of cation charge per kilogram. A plot of ExNa versus SAR will produce a curvilinear line, asymptotically approaching CEC. Because of the exponent in the denominator of eqn [11b], the K_G of a heterovalent exchange depends on solution concentration units, whereas the K_G of a homovalent exchange is independent of solution units. A heterovalent K_G obtained with solution units in millimoles per liter when multiplied by $(1000)^{0.5}$ gives the K_G in units of (moles per liter)$^{-0.5}$.

Rearranging eqn [6b] gives:

$$(\mathrm{ExNa}/\mathrm{ExCa}_{0.5}) = K_G \cdot \left([\mathrm{Na}^+]/[\mathrm{Ca}^{2+}]^{0.5}\right) \qquad [11d]$$

Theoretically, a plot of $\mathrm{ExNa}/\mathrm{ExCa}_{0.5}$ or exchangeable sodium ratio (ESR) versus SAR will produce a straight line with a slope equal to K_G. The mean magnitude of K_G for arid US soils is approximately 0.015 $(\mathrm{mmol\,l}^{-1})^{-0.5}$. However, experimentally determined K_G are found to depend on pH, salt concentration, clay mineralogy, and sodium load. As sodium load increases, K_G increases, and, as pH increases in variable-charge soils, K_G decreases. The K_G of Na^+–Ca^{2+} exchange may vary from 0.50 to 2$(\mathrm{mol\,l}^{-1})^{-0.5}$, whereas the K_G of K^+–Ca^{2+} exchange may vary from 2 to 30 $(\mathrm{mol\,l}^{-1})^{-0.5}$, depending on mineralogy and pH. Illite has significantly greater affinity for K^+ than does smectite or organic matter.

Cation preference in heterovalent exchange is also demonstrated through fractional isotherms. Heterovalent fractional cation preference isotherms differ from homovalent cation preference isotherms in that the nonpreference line in heterovalent exchange is not the diagonal line and depends on ionic strength ('square-root effect'). The nonpreference line for a heterovalent exchange reaction is shown as one of the solid lines in Figure 3. Heterovalent exchange data occurring above the nonpreference line reveal that the surface prefers the monovalent cation, while exchange data occurring below the nonpreference line reveal that the surface prefers the divalent cation (Figure 3).

It is suggested that the Vanselow equation is consistent with thermodynamics of chemical reactions because it uses units of moles for solution phase and the exchanger. Solving the Vanselow equation for ExNa gives:

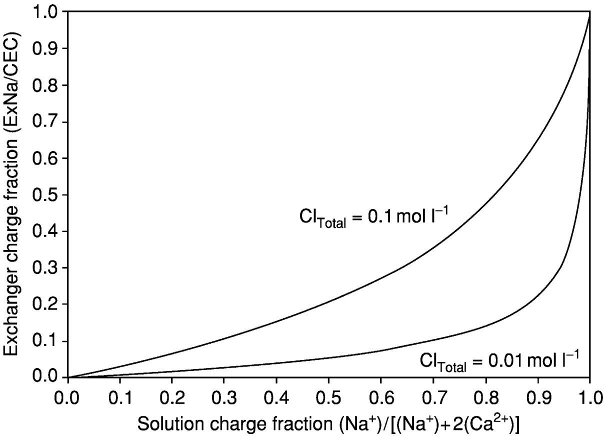


Figure 3 Nonpreference isotherms for heterovalent sodium–calcium exchange at two total chloride concentrations ($\mathrm{Cl_{Total}}$). Plotted using eqn [14].

$$\mathrm{ExNa} = \frac{\mathrm{CEC} \cdot K_V \cdot \mathrm{AR_{Na}}}{\left[4 + (K_V \mathrm{AR_{Na}})^2\right]^{0.5}} \qquad [12a]$$

where $(\mathrm{Na}^+)/(\mathrm{Ca}^{2+})^{0.5}$ is the sodium activity ratio, $\mathrm{AR_{Na}}$. Considering that for the nonpreference isotherm, as $I \to 0$, $K_v = 1$ and $\Delta G^0 = 0$, then:

$$(\mathrm{ExNa}/\mathrm{CEC}) = \frac{\mathrm{AR_{Na}}}{\left[4 + (\mathrm{AR_{Na}}^2)\right]^{0.5}} \qquad [12b]$$

Using a number of $\mathrm{AR_{Na}}$ values, as $I \to 0$, to cover the entire exchange isotherm, equivalent fractional loads for Na^+ can be estimated using eqn [12b]. A plot of ExNa/CEC versus $(\mathrm{Na}^+)/[(\mathrm{Na}^+) + (2\mathrm{Ca}^{2+})]$ will produce a curvilinear line representing the nonpreference isotherm. Upon introducing activity coefficients in the $\mathrm{Na}^+/[\mathrm{Na}^+ + 2\mathrm{Ca}^{2+}]$ term such that:

$$(\mathrm{ExNa}/\mathrm{CEC}) = \frac{(\mathrm{Na}^+)/\gamma_{\mathrm{Na}}}{(\mathrm{Na}^+)/\gamma_{\mathrm{Na}} + 2 \cdot (\mathrm{Ca}^{2+})/\gamma_{\mathrm{Ca}}} \qquad [13]$$

(again assuming $K_V = 1$), it can be shown that as I increases, γ_{Na} and γ_{Ca} decrease disproportionately, with γ_{Ca} decreasing significantly more than γ_{Na}. Therefore, as I increases, eqn [13] decreases, indicating that the nonpreference isotherm is ionic-strength dependent. Alternatively, an equation for the heterovalent nonpreference isotherm has been derived:

$$\mathrm{ExNa}/\mathrm{CEC} = \left[1 + \left(\frac{2\gamma_{\mathrm{Ca}}}{(\mathrm{TN})\gamma_{\mathrm{Na}}^2}\left(\frac{1}{(E'_{\mathrm{Na}})^2} - \frac{1}{E'_{\mathrm{Na}}}\right)\right)\right]^{-0.5} \qquad [14]$$

where E'_{Na} is the charge fraction of Na^+ in solution and TN is the total cation normality. Eqn [14] is plotted in Figure 3 for two total cation normalities ($Cl_{Total} = Na^+ + 2Ca^{2+}$). Increasing the salt concentration in solution decreases the preference the surface shows for the divalent cation, Ca^{2+}. At sufficiently high salt concentrations, the surface may actually show slight preference for the monovalent ion. The influence of γ_i, induced by increasing I, on apparent ion preference, however, is very small in comparison with the square-root effect on apparent ion preference also induced by increasing I.

In general, the selectivity coefficient (K_v) for a binary exchange reaction depends primarily on the ionic strength of the solution, on the proportion of cations in the soil-absorbing complex, and the proportion of the cations in the soil solution phase. Exchange reactions of Na^+ with trace metal cations (Cd^{2+}, Co^{2+}, Cu^{2+}, Ni^{2+}, and Zn^{2+}) on Camp Berteau montmorillonite show that K_v is constant and independent of exchanger composition, suggesting ideal solid-solution behavior, up to an equivalent fraction of trace metal cations of 0.70. Sodium–calcium exchange studies on similar clay minerals show that there is a more pronounced selectivity of clay for Ca^{2+} at the calcium-rich end of the isotherm. That experimentally determined selectivity coefficients are not constant across the entire exchange isotherm indicates that the exchanger activity coefficients (γ_{ExNa}, γ_{Ex_2Ca} in eqn [3b]) change as a function of surface composition. Mass balance constraints mean that a change in any one exchanger activity coefficient must be compensated by an equal change in the other exchanger activity coefficient (the Gibbs–Duhem equation). These ideas lead to the derivation of the thermodynamic exchange equilibrium constant, K_{eq}, from the Vanselow convention as:

$$\ln K_{eq} = \int_0^1 \ln K_V \, dE_{Na} \qquad [15]$$

where E_{Na} is the equivalent fraction of Na on the exchanger (ExNa/CEC). Homovalent and heterovalent thermodynamic exchange constants (K_{eq}) for a number of exchangers and exchange reactions are given in Table 4.

The Future of Cation Exchange

The preceding discussion, and nearly all of the research into cation exchange reactions, has been conducted on binary exchange reactions. However, soils are multi-ion systems. In order for binary exchange reactions accurately to describe multi-ion exchange, the binary selectivity coefficients have to be independent of the exchanger composition, at least over the range of relevant solution and exchanger concentrations. In some studies, binary reactions sufficiently describe higher-order cation exchange; in others, they do not. Considerably more work needs to be

Table 4 Thermodynamic equilibrium exchange constants (K_{eq}) for binary homovalent and heterovalent exchange (all experiments conducted at 298 K except where noted)

	Homovalent		*Heterovalent*	
Exchanger	*Process*	K_{eq}	*Process*	K_{eq}
Calcareous soil	Ca–Mg	0.89–0.75	Ca–Na	0.38–0.09
Camp Berteau montmorillonite	Ca–Mg	0.95	Ca–Na	0.72
			Ca–NH_4	0.035
Wyoming bentonite	Ca–Cu	0.96		
	Na–Li	1.08		
	K–Na	1.67		
World vermiculite	Na–Li	11.42	Ca–Na	0.98
			Mg–Na	1.73
Chambers montmorillonite	Na–Li	1.15	Ca–K	0.045
	K–Na	3.41		
Kaolinitic soil clay (303 K)	Mg–Ca	0.65	K–Ca	16.16
Soil	K–Na	4.48–6.24	Ca–Na	0.42–0.043
	Mg–Ca	0.61	Mg–Na	0.75–0.053
			K–Ca	5.89–323.3
			K–Ca	12.09
			K–Ca	19.92–323.3
			K–Ca	0.46
			K–Ca	0.64–6.65
			K–Ca	6.42–6.76
			K–Mg	5.14

Adapted from Sparks DL (1995) *Environmental Soil Chemistry.* San Diego, CA: Academic Press.

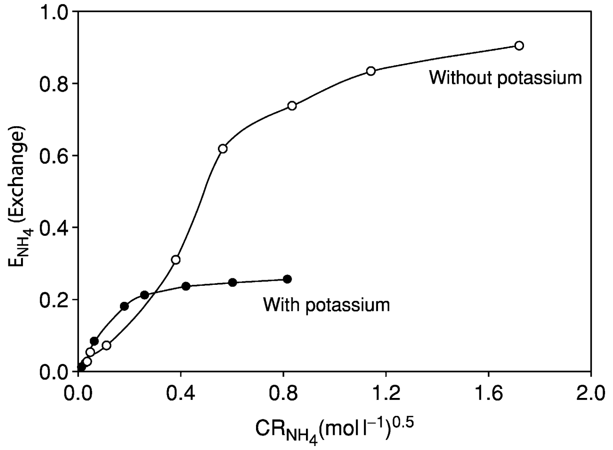


Figure 4 Effect of potassium on $NH_4^+ - Ca^{2+}$ exchange isotherm for vermiculite. Reproduced with permission from Evangelou VP and Lumbanraja J (2002) Ammonium–potassium–calcium exchange in the absence and presence of potassium on vermiculite and hydroxy-Al vermiculite. *Soil Science Society of America Journal* 66: 445–455.

done to identify the conditions where ternary exchange can be predicted from binary exchange data. One situation where binary data are not likely to be accurate is when surfaces undergo conformational changes. In circumneutral agricultural soils, Ca^{2+} and K^+ are present in large quantities, and NH_4^+ is added as fertilizer. Figure 4 shows an NH_4–Ca exchange isotherm for vermiculite, with and without added K^+. At low NH_4 concentrations (the agriculturally important region), adding K^+ increases the selectivity coefficient toward preference for NH_4^+. That is, in addition to the well-known effect of K^+ on ammonium fixation (nonexchangeable NH_4^+, K^+ appears to influence the availability of exchangeable NH_4^+.

As discussed in the Introduction, cation exchange is assumed to be a reversible, adsorption reaction as outer-sphere or diffuse-layer complexes. However, all cations possess some ability to form inner-sphere surface complexes and so a complete understanding of cation distribution in soils will require consideration of these reactions. Cation exchange equilibria can be incorporated into existing computer-based surface complexation models. The advantage to this approach is that ternary and quaternary exchange reactions can be included. Although the data and calibration requirements are substantial, this approach holds considerable promise for providing a comprehensive description of cation distribution in soils.

List of Technical Nomenclature

γ	Activity coefficient
ϵ	Dielectric constant
I	Ionic strength
K (Kelvin)	Temperature
z	Ion valence

See also: **Calcium and Magnesium in Soils**; **Chemical Equilibria**; **Crusts:** Structural; **Flocculation and Dispersion**; **Salt-Affected Soils, Reclamation**; **Sodic Soils**

Further Reading

Evangelou VP and Lumbanraja J (2002) Ammonium–potassium–calcium exchange on vermiculite and hydroxy-Al vermiculite. *Soil Science Society of America Journal* 66: 445–455.

Evangelou VP, Wang J, and Phillips RE (1994) New developments and perspectives in characterization of soil potassium by quantity–intensity (Q/I) relationships. In: Sparks DL (ed.) *Advances in Agronomy.* vol. 52, pp. 173–227. Orlando, FL: Academic Press.

Holmgren GGS, Meyer MW, Chaney RL, and Daniels RB (1993) Cadmium, lead, zinc, copper and nickel in agricultural soils of the United States of America. *Journal of Environmental Quality* 22: 335–348.

Rhee H and Dzombak (1998) Surface complexation/Gouy–Chapman modeling of binary and ternary cation exchange. *Langmuir* 14: 935–943.

Sparks DL (1989) *Kinetics of Soil Chemical Processes*. San Diego, CA: Academic Press.

Sparks DL (1995) *Environmental Soil Chemistry*. San Diego, CA: Academic Press.

Sposito G (1981a) Cation exchange in soils: a historical and theoretical perspective. In: Dowdy RH (ed.) *Chemistry in the Soil Environment*, pp. 13–30. Madison, WI: Soil Science Society of America.

Sposito G (1981b) *The Thermodynamics of Soil Solutions*. New York: Oxford University Press.

Suarez DL (1999) Thermodynamics of the soil solution. In: Sparks DL (ed.) *Soil Physical Chemistry*, pp. 97–134. Boca Raton, FL: CRC Press.

Sumner ME and Miller WP (1996) Cation exchange capacity and exchange coefficients. In: Sparks DL, Page AL, Helmke PA *et al.* (eds) *Methods of Soil Analysis, part 2. Chemical Methods*, pp. 1201–1229. Soil Science Society of America Book Series No. 5. Madison, WI: Soil Science Society of America and American Society of Agronomy.

Thomas GW (1977) Historical developments in soil chemistry: ion exchange. *Soil Science Society of America Journal* 41: 230–238.

CHEMICAL EQUILIBRIA

A P Schwab, Purdue University, West Lafayette, IN, USA

Introduction

Although soils are very complex living systems, they are composed of identifiable minerals, amorphous compounds, and discrete ionic species that can be quantified. Solubilities, solid-phase transformations, dissolution of gases, and changes in oxidation states of these chemical constituents can be predicted by chemical thermodynamics. However, implicit in the application of these principles to soils is that the soil system (or, at the very least, the constituents in question) are in chemical equilibrium. If the assumption of equilibrium is valid, then we have very powerful tools at our disposal to describe the chemical behavior of the solid, liquid, and gas phases in soil.

Chemical Equilibrium

Before discussing chemical equilibrium in soils, we must clearly understand the concept. Perhaps the most famous discussions of equilibrium were presented in the 1880s by Henri LeChatlier, for whom LeChatlier's Principle was named. Two basic definitions of chemical equilibrium will be discussed: kinetic and free energy. The two definitions simply express the same phenomenon in different ways.

Kinetic Definition

Consider the generalized chemical reaction with reactants A and B, and forming products C and D with stoichiometric coefficients a, b, c, and d:

$$a\mathrm{A} + b\mathrm{B} \Leftrightarrow c\mathrm{C} + d\mathrm{D}$$

The rate of the forward reaction is given by:

$$v_\mathrm{f} = k_\mathrm{f}[\mathrm{A}]^a[B]^b$$

in which $[x]$ represents the solution concentration of component x. The rate expression for the reverse reaction is given by:

$$v_\mathrm{r} = k_\mathrm{r}[\mathrm{C}]^c[D]^d$$

When only the reactants are present, the forward reaction proceeds quickly. As products begin to accumulate, the reverse reaction begins to occur simultaneously. Eventually, the point will be reached in which the forward and reverse reactions are proceeding at the same rate and the concentrations of the reactants and products will not change with time. This is the point of a dynamic, chemical equilibrium. When the forward and reverse rates are equal:

$$k_\mathrm{f}[\mathrm{A}]^a[\mathrm{B}]^b = k_\mathrm{r}[\mathrm{C}]^c[D]^d$$

Rearranging

$$K_\mathrm{eq} = \frac{k_\mathrm{f}}{k_\mathrm{r}} = \frac{[\mathrm{C}]^c[D]^d}{[\mathrm{A}]^a[B]^b}$$

in which K_eq is the equilibrium constant.

Free-Energy Definition

The chemical potential of any species in solution (μ_i) is given as:

$$\mu_\mathrm{i} = \mu_\mathrm{i}^\mathrm{o} + RT\ln(i)$$

in which $\mu_\mathrm{i}^\mathrm{o}$ is the standard chemical potential of species i, R is the ideal gas constant, T is the temperature in Kelvin, and (i) is the activity of species i. The expression for the free-energy change of any chemical reaction is:

$$\Delta G = \sum_\mathrm{i} v_\mathrm{i}\mu_\mathrm{i}$$

in which ΔG is the change in Gibbs free energy during the course of the reaction, and v_i is the stoichiometric coefficient for the components of the equilibrium expression (v_i is positive for products, negative for reactants). Substituting:

$$\Delta G = \sum_\mathrm{i} v_\mathrm{i}\mu_\mathrm{i}^\mathrm{o} + RT\sum_\mathrm{i} v_\mathrm{i}\ln(i)$$

or:

$$\Delta G = \Delta G^\mathrm{o} + RT\ln\prod_\mathrm{i}(i)^\mathrm{vi}$$

where:

$$\Delta G^\mathrm{o} = \sum_\mathrm{i} v_\mathrm{i}\mu_\mathrm{i}^\mathrm{o}$$

and ΔG^o is the standard Gibbs free energy change of the reaction. The reaction quotient Q is represented as:

$$Q = \prod_\mathrm{i}(i)^\mathrm{vi}$$

such that:

$$\Delta G = \Delta G^\mathrm{o} + RT\ln Q$$

By definition, the free-energy change of reaction equals zero ($\Delta G = 0$) at equilibrium, and the value of Q at equilibrium becomes the equilibrium constant K_{eq}:

$$K_{\mathrm{eq}}\ Q = \prod_{\mathrm{i}} (i)^{\mathrm{vi}}_{\mathrm{eq}}$$

and:

$$\Delta G^{\mathrm{o}} = -RT\ln K_{\mathrm{eq}}$$

Thus, we have the free-energy definition of equilibrium ($\Delta G = 0$) and the thermodynamic derivation of the equilibrium constant.

Applicability of Chemical Equilibrium Principles to Soils

Invoking the assumption of chemical equilibrium in soils, particularly for soils in natural conditions, is difficult because soils are highly complex and forever changing. Soils continually have inputs of external energy, water, gases, and dissolved constituents. Losses of these same components are equally possible through radiation of heat, leaching, and biological activities. These changes perturb the soil system and upset any equilibrium that may be present. With this in mind, a reasonable question is, "Can the principles of chemical equilibrium be applied to soils?"

Despite the dynamic nature of soils and continuous inputs and removal from the system, chemical equilibrium can be applied to soils if care is taken, the limitations are clearly understood, and expectations are reasonable.

Limitations to Equilibrium in Soils

Most soils are open systems with frequent inputs and removal of mass and energy. Even the simple act of rain falling on a soil results in dilution of all dissolved soil constituents and disruption of all previous equilibria. Changes in temperature, plant assimilation of nutrients, evaporation of water, and diffusion of gases into or out of soil have a similar disruptive effect. These perturbations in equilibrium may be temporary or may have long-term implications, depending upon the kinetics of these reactions. For example, the half-life of complexation/dissociation reactions vary from 10^{-9} to 10^{3} s (Table 1). Thus, outside disturbances to equilibria will have only a small impact on complexation reactions. However, oxidation–reduction and solid-phase dissolution–precipitation reactions can require a very long time to achieve equilibrium. A classic example of this is the dissolution of gibbsite (γ-Al(OH)$_3$). Monitoring the activity of Al^{3+} and pH in a suspension of soil generally reveals that the concentration of Al^{3+} in the soil solution would be more than 10-fold less than equilibrium after 200 h, based on a comparison of the known equilibrium constant with measured parameters. Solubilities would be within a factor of 2 of equilibrium after 400 h and equilibrium should be attained after approximately 1000 h. Slight shifts in pH, temperature, or moisture content would result in nonequilibrium conditions for γ-Al(OH)$_3$(gibbsite) that would again require hundreds of hours to correct.

Table 1 Half-lives of selected complexation reactions in soil solutions

Reaction	*Half-life (s)*
$\mathrm{MnSO_4^o} \Leftrightarrow \mathrm{Mn^{2+}} + \mathrm{SO_4^{2-}}$	10^{-9}
$\mathrm{Fe^{3+}} + \mathrm{H_2O} \Leftrightarrow \mathrm{Fe(OH)^{2+}} + \mathrm{H^+}$	10^{-7}
$\mathrm{Mn^{2+}} + \mathrm{SO_4^{2-}} \Leftrightarrow \mathrm{MnSO_4^o}$	10^{-5}
$\mathrm{NiC_2O_4^o} \Leftrightarrow \mathrm{Ni^{2+}} + \mathrm{C_2O_4^{2-}}$	10^{-1}
$\mathrm{CO_2} + \mathrm{OH^-} \Leftrightarrow HCO_3^-$	10
$\mathrm{Al^{3+}} + \mathrm{F^-} \Leftrightarrow \mathrm{AlF^{2+}}$	10^{3}

Data from Sposito G (1994) *Chemical Equilibria and Kinetics in Soils*. New York: Oxford University Press.

Some reactions are thermodynamically favorable but are associated with a large energy barrier (e.g., energy of activation). Unless that energy barrier can be overcome, equilibrium will not be realized. The conversion of N$_2$(g) to nitrate should be a spontaneous reaction, but the presence of approximately 78% N$_2$(g) in the atmosphere is clear evidence that this reaction does not proceed as predicted. Redox couples with oxygen-containing anions (e.g., $\mathrm{AsO_4^{3-}/AsO_3^{3-}}$, $\mathrm{S^{2-}/SO_4^{2-}}$) attain equilibrium slowly in the absence of a catalyst.

Measurement of critical system variables can be subject to serious limitations that restrict rigorous application of thermodynamics to natural soils. Detection limits of soluble constituents may preclude quantification, redox-sensitive electrodes have shortcomings, and distinguishing oxidation states of soluble elements can be troublesome. Redox is a crucial system parameter, but chemical limitations of redox electrodes are well documented, and not all oxidation–reduction couples establish reversible equilibria at the surface of the sensing electrode.

The final limitation is ensuring that all products can form under soil conditions. Thermodynamic data and/or equilibrium constants are available for hundreds of minerals found in soils, but many of these minerals can only be formed under conditions of high temperatures and pressures. This is particularly true of the primary minerals in soils. Thermodynamics might predict that these minerals form in soils, but this prediction is meaningless if the solids simply cannot form.

Soil Systems in Which Equilibrium Concepts May Apply

Chemical equilibrium principles can be applied with reasonable confidence to most (but not all) reactions occurring exclusively in the solution phase. As illustrated in Table 1, most reactions attain equilibrium in a matter of a few seconds or less; therefore, typical perturbations to equilibrium relax quickly. Notable exceptions are oxyanion redox couples and cleaving the $N_2(g)$ molecule to form nitrate.

Attempts to apply equilibrium to the solid phase in the soil require more consideration. The kinetics of dissolution of typical soil minerals are slow. Calcium phosphate and sulfate minerals can achieve equilibrium in hours; aluminum phosphates, carbonates, and most oxides in a matter of days or weeks; and complex aluminosilicates in years. If the soil solution has been in contact with soil minerals for the time necessary to establish equilibrium while experiencing minimal disturbances, application of equilibrium principles is possible. Complete equilibrium among all species in the soil solution and all soil minerals is not essential; however, these principles cannot be applied if even one component ion of the mineral being studied is out of equilibrium.

Consider the example of variscite, $AlPO_4{\cdot}2H_2O$. Phosphate ions may reach constant concentrations in a matter of hours in the soil solution, but Al^{3+} may require days or weeks due to the slow reaction kinetics of aluminum oxides. Therefore, the system will be considered to be in equilibrium only when both Al^{3+} and PO_4^{3-} have reached time-independent concentrations.

Reasonable Expectations from Application of Equilibrium

Universal adherence to chemical equilibrium in soils is unrealistic. However, employing thermodynamics and associated solubility and complexation constants in soil systems can be fruitful if one has reasonable expectations.

Predict endpoints of reactions Thermodynamics can be useful in determining possible endpoints versus impossible outcomes of chemical reactions in soils. For example, we can predict that metallic iron will readily convert to ferric oxides under surface soil conditions:

$$2Fe(s) + 1.5O_2(g) + 3H_2O \Leftrightarrow 2Fe(OH)_3(amor)$$

Determine ionic activities in soil solution The activities of individual species in the soil solution are often the driving forces behind many chemical and biological reactions in soils and can be calculated using chemical equilibrium constants.

Calculating final conditions for systems that attain equilibrium rapidly When changes to the soil system occur infrequently relative to the timescale required for equilibrium, then equilibrium concepts can be applied.

Experimental Approach to Equilibrium in Soils

Thermodynamics may be used to describe chemical reactions in soils regardless of whether the application is interpretation of experimental results or the prediction of changes in soil chemistry induced by environmental perturbations. In either case, the user must be acutely aware of the limitations of chemical equilibrium in soils and must have knowledge of the systems to which these principles may be applied with confidence. The steps taken are generally the same: critical system variables must be identified; key reactions of all the contributing and interacting components are established; reliable equilibrium constants are located and assigned to each reaction; and, finally, the system of equilibrium equations is set up as an mathematical array and solved. All steps in this process require careful consideration.

System Variables

The most important variables in a soil system are similar to those in any other geochemical environment. The list of parameters to be considered is driven in part by the interest and focus of the user and is strongly influenced by other variables that are intimately linked to the systems of interest. For example, if one is considering only the solubility of ferric oxides in soils with low pH, then the obvious system parameters are pH and soluble Fe. However, Fe^{3+} can form important solution complexes with Cl^-, SO_4^{2-}, and F^-, therefore, one must account for these ancillary anions before the Fe system can be fully defined.

Reactions and Associated Equilibrium Constants

For any given system, all relevant reactions and the corresponding constants must be identified. This is not an easy task and requires databases of reliable information. Much of this work was done in the development of geochemical models, as will be discussed below. For a very simple system such as amorphous Fe oxide in equilibrium in soil, at least the following solution species need to be considered: Fe^{3+}, FeF^{2+}, FeF_2^+, $FeCl^{2+}$, $FeOH^{2+}$, $Fe(OH)_2^+$, $Fe(OH)_3^o$, $Fe(OH)_4^-$, $FeHPO_4^+$, and $FeSO_4^o$. A typical equation and constant would be:

$$Fe^{3+} + H_2O \Leftrightarrow FeOH^{2+} + H^+$$

$$\log K_{eq} = -2.19$$

Equations and constants are assembled for all species, including the dissolution reaction for the controlling solid phase:

$$FeOH_3(amor) + 3H^+ \Leftrightarrow Fe^{3+} + 3H_2O$$

$$\log K_{eq} = 3.54$$

After all species have been associated with an independent equation or an actual measurement of a system parameter (e.g., total soluble Fe or pH), they are assembled in a mathematical array of simultaneous equations; the array is solved, and the system can be defined exactly. For small systems, this can be done by hand, but larger and more comprehensive systems require a full geochemical model to provide a solution.

Equilibrium Modeling of Soil Systems

Solving the array of simultaneous equations that results from defining a soil system in equilibrium provides a serious challenge. Hand solutions are highly time-consuming, and writing small computer programs for each new system is not productive. Fortunately, this problem was recognized in the mid 1970s, and several research groups set about the task of writing generalized computer models for geochemical systems.

Geochemical models In 1974, the WATEQ was published by the US Geological Survey. WATEQ was a fairly comprehensive model with a carefully reviewed database of thermodynamic constants. The program was large and somewhat slow to compile and execute. The MINEQL model, published in 1974, has an elegant mathematical structure and allows for very fast execution times and flexibility. However, the database had not been rigorously reviewed. Since then, geochemical modeling has evolved in many directions with many new models: PHREEQE; GEOCHEM; SOILCHEM; MINTEQ; and many others. Many review articles have been written comparing various models. The calculation of the equilibrium between the gas, solid, and aqueous phases differs only by speed of execution, the database of thermodynamic constants, and associated functions.

Model inputs Each model requires that the user define the systems to be modeled. The computer models are fully capable of accepting experimental data for the purpose of calculating ionic activities and comparing them with mineral solubilities. Similarly, the models may be used completely in a predictive mode in which assumptions are made concerning equilibrium with solid phases, pH, temperature, gas partial pressures, etc. The models then will report resulting activities, total elemental solubilities, residual masses of solid phases, and the distribution of mass of a given element across all phases.

Table 2 Modeling results using MINTEQA2 and inputs of pH 5.5, equilibrium with $Fe(OH)_3$(amorphous), and total solubilities of 0.001 mol l^{-1} for chloride, fluoride, and sulfate

Species	*log (Activity)*	*Species*	*Fe mass (%)*
Cl^-	−3.05	$Fe(OH)_2^+$	96.3
F^-	−3.05	$FeOH^{2+}$	1.3
SO_4^{2-}	−3.18	$Fe(OH)_3^o$	1.1
Fe^{3+}	−12.96		
Total soluble Fe	−7.59		

Continuing with the amorphous Fe oxide example, let us assume that the system in question is in equilibrium with amorphous $Fe(OH)_3$(amor) at pH 5.5, with total soluble chloride, fluoride, and sulfate of 0.001 mol l^{-1}. Using this information as input into MINTEQA2 resulted in the data given in Table 2.

Application of Chemical Equilibrium to Soils: Examples

This section will explore the application of equilibrium to soils in which soil solution measurements were used as inputs into a geochemical model. The output of the model was then used to determine ion activity products and compare them with corresponding equilibrium constants. The first examples are successes, followed by apparent nonequilibrium examples and controversies.

Iron Oxides in Reduced Systems

An ideal system to study in soils is the Fe(II/III) oxide group because of well-characterized thermodynamics and rapid equilibrium. In a recent study, amorphous Fe(III)oxide was precipitated under reducing conditions and exposed to either small or large increments of O_2(g) to encourage oxidation. When O_2(g) was added slowly, equilibrium was established with Fe_3O_4(magnetite). When O_2(g) was added rapidly, equilibrium Fe^{3+} activities were consistently elevated relative to Fe_3O_4(magnetite), suggesting the formation of a different solid phase, Fe_3O_4(amorphous) (Figure 1). The proximity of the measured data to the predicted solubility products is striking. For Fe_3O_4(amorphous), the close correspondence is observed over a range from pE + pH 5 to pE + pH 12. Equilibrium with magnetite is suggested in the low-redox environments, and simultaneous equilibrium with $Fe(OH)_3$(amorphous) and magnetite is suggested at higher redox potentials. The formation of an amorphous mixed Fe oxide has been observed previously, but usually with higher solubility.

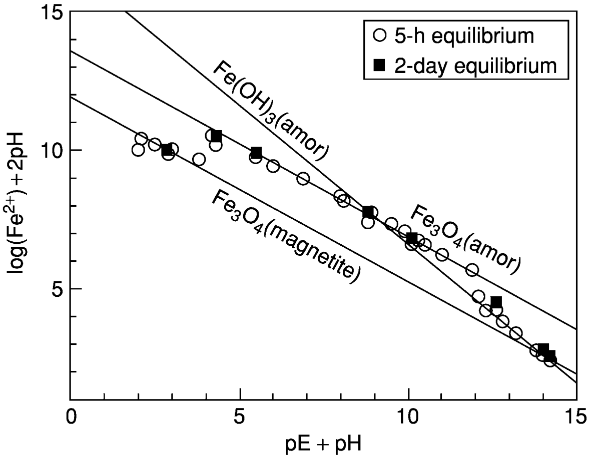


Figure 1 Measured solution activities of Fe (data points) as compared to theoretical solubilities (lines) of selected Fe solid phases. Reproduced from Brennan EW and Lindsay WL (1998) Reduction and oxidation effect on the solubility and transformation of iron oxides. *Soil Science Society of America Journal* 62: 930–937.

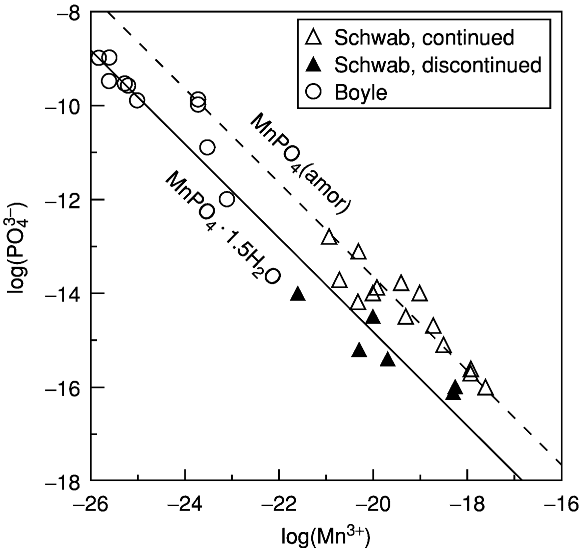


Figure 2 Measured solution activities of Mn^{3+} and PO_4^{3-} (data points) as compared to theoretical solubilities (lines) of $MnPO_4 \cdot 5H_2O$.

Manganese(III) Phosphate

Manganese is a critical element in soils in terms of plant and animal nutrition and can be a pollutant if Mn concentrations become elevated. Manganese phosphates are known to form in natural environments but are too soluble to persist, with the possible exception of $Mn(III)PO_4 \cdot 1.5H_2O$. Although the equilibrium constant for this mineral ($\log K_{eq} = -34.4$) suggests very low solubilities, Mn(III) is present in vanishingly small concentrations in soil (10^{-15} mol l^{-1} as Mn^{3+}). Nevertheless, two studies seem to suggest that the mineral can form in fertilized soils.

Activities of PO_4^{3-} and Mn^{3+} were determined in calcareous soils, and the correspondence between these activities and the theoretical solubility of $Mn(III)PO_4 \cdot 1.5H_2O$ was readily apparent (Figure 2). A similar study was conducted in nonalkaline soils, and two trends were observed in solubility, depending upon whether or not the soil had been recently fertilized. In soils fertilized annually, activities were slightly elevated relative to $Mn(III)PO_4 \cdot 1.5H_2O$, suggesting the presence of a less-crystalline, more-soluble form of the mineral. However, in those soils in which P applications had ceased several years prior to soil sampling, the correspondence between measured and theoretical activities for $Mn(III)PO_4 \cdot 1.5$ H_2O were closer.

Calcite in Soils: Equilibrium versus Nonequilibrium

Precipitation in soils of calcium carbonate, usually considered to be calcite, is easily reversible and quickly establishes equilibrium. Although many studies examining this mineral have been conducted and support the notion of calcite equilibrium, others have found varying degrees in nonequilibrium.

When at equilibrium, the solubility of calcite may be expressed by the following equation:

$$CaCO_3(\text{calcite}) \Leftrightarrow Ca^{2+} + CO_3^{2-}$$
$$\log K_{eq} = -8.48$$

Determining the single ion activities of Ca^{2+} and CO_3^{2-} in soil solution makes it possible to compare measured activity products to the thermodynamic constant. If care is taken in determining the ionic activities, then correspondence between measured and predicted activities can be observed. In one such study, the measured ion activity products ranged from $10^{-8.22}$ to $10^{-8.58}$ with a mean of $10^{-8.41}$.

Although some calcite-containing soils maintain apparent equilibrium with calcite, others show supersaturation. Despite following the diligent protocols to ensure accuracy of measurement, the measured ion activity products are as high as $10^{-7.0}$, indicating supersaturation by a factor of about 30. Nonequilibrium appears to have been induced by high levels of soluble organic matter that stimulates microbial activity (creating a dynamic system) and acts as a long-term source for soluble calcium.

The Outlook for Studying Chemical Equilibrium in Soils

Advances in geochemical modeling and computing capacity have established new ways of viewing and studying soils. Using numerical methods to predict the fate and transport of chemicals in soils is a compelling notion, and models are being used worldwide. The applications are nearly limitless and include plant nutrition, fate of fertilizers, leaching of soluble constituents, examining the mobility of contaminant metals, and volatilization of organics from soils. Armed with the powerful geochemical models and the knowledge of the chemical and physical limitations of applying chemical equilibria in soils, today's soil scientists can be fully prepared to tackle the most challenging and interesting problems of chemical behavior in soil systems.

List of Technical Nomenclature

μ_i	Chemical potential of species *i*
μ_i^o	Standard chemical potential of species *i*
Equilibrium	A state of balance in a chemical reaction in which the concentrations of the reactants and products are no longer changing and in which minimum free energy has been attained
k	Rate constant
Redox	Oxidation/reduction potential
Solubility constant	Thermodynamic equilibrium constant for a reaction describing the dissolution of a solid phase
Steady state	A balance in an experimentally observed chemical reaction in which the concentrations of the reactants and products appear to be unchanging with time. Steady state differs from equilibrium in that the minimum free energy for a given reaction has not been achieved
Stoichiometry	A description of the relative quantities (in moles) of products and reactants in a chemical reaction as indicated by the coefficients in the balanced chemical equation

Further Reading

Boyle FW and Lindsay WL (1985) Preparation, X-ray-diffraction pattern, and solubility product of manganese (III) phosphate hydrate. *Soil Science Society of America Journal* 49: 758–760.

Brennan EW and Lindsay WL (1998) Reduction and oxidation effect on the solubility and transformation of iron oxides. *Soil Science Society of America Journal* 62: 930–937.

Crawford MB (1996) PHREEQEV: The incorporation of a version of model V for organic complexation in aqueous solutions into the speciation code PHREEQE. *Computers and Geosciences* 22: 109–116.

Gaskova OL, Azaroual M, Bodenan F, and Gaucher E (1999) Evidenced redox disequilibrium from iron and arsenic behavior in tailing leachate (Cheni Site, France). In: *Proceedings of the Ninth Annual V.M. Goldschmidt Conference*. Cambridge, MA.

Kittrick JA (1977) Mineral equilibria and the soil system. In: Dixon JB and Weed SB (eds) *Minerals in Soil Environments*, pp. 1–25. Madison, WI: Soil Science Society of America.

Lindsay WL (1979) *Chemical Equilibrium in Soils*. New York: Wiley Interscience.

Schwab AP (1989) Manganese–phosphate solubility relationships in an acid soil. *Soil Science Society of America Journal* 53: 1654–1660.

Sposito G (1994) *Chemical Equilibria and Kinetics in Soils*. New York: Oxford University Press.

Sposito G and Coves J (1988) *Soilchem, a Computer Program for the Calculation of Chemical Speciation in Soils*. Berkeley, CA: University of California at Berkeley.

Stumm W and Morgan JJ (1996) *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters*. New York: Wiley Interscience.

Truesdell AH and Jones BF (1974) WATEQ, a computer program for calculating chemical equilibria in natural waters. *US Geological Survey Journal of Research* 2: 233–248.

Westall JC, Zachary JL, and Morel FMM (1976) *MINEQL, a Computer Program for the Calculation of Chemical Equilibrium Composition of Aqueous Systems*. Massachusetts Institute of Technology, Technical Note 18. Cambridge, MA: MIT Press.

Whitfield M (1974) Thermodynamic limitations on the use of the platinum electrode in Eh measurements. *Limnology and Oceanography* 19: 857–865.

Wolt JD (1989) SOILSOLN: a computer program for teaching equilibria modeling of soil solution composition. *Journal of Agronomic Education* 18: 40–42.

Chemical Speciation Models *See* Surface Complexation Modeling

Chernozems *See* Grassland Soils

CHILDS, ERNEST CARR

E G Youngs, Cranfield University, Silsoe, Bedfordshire, UK

From early times it has been recognized that the physical state of soil is important in the practice of agriculture. Soil physics has thus become a subject studied as part of agriculture, not as a subject in mainstream physics. Progress in soil physics owes much to the few physicists in the first part of the twentieth century who left careers in pure physics to work in agricultural research, and these few laid the foundations of the scientific discipline of soil physics. Ernest Carr Childs was one of the physicists who left a career in pure physics to help pioneer a basic physical approach to soil behavior (Figure 1). Born 6 November 1907 in Forest Gate, London, he attended East Ham Grammar School and then studied for his BSc degree in physics at King's College in the University of London. After graduating with first-class honours in 1927, he continued at King's as a research student under E.V. Appleton and later as a demonstrator, with an interruption of a year with the Cambridge Instrument Company. There he was in charge of small-instrument development and this led to the collaboration with A. Campbell in writing a book, *The Measurement of Inductance, Capacitance and Frequency*, which was published in 1935. He was awarded a PhD by the University of London in 1931 for a thesis on the radiofrequency properties of ionized air. He then became a research student at the Cavendish Laboratory in Cambridge in the 'golden years' under Lord Rutherford. His research on the diffraction of slow-speed electrons in certain metal vapors earned him a second PhD, this time from the University of Cambridge, in 1934.

Figure 1 Ernest Carr Childs 1907–73.

With this physics background, in the forefront of research in one of the most prestigious physics departments in the world, Childs crossed the road to the School of Agriculture in the neighboring museum's site in Downing Street, Cambridge, to join H.H. Nicholson in 1934 to study water movement in heavy clay soils, particularly as it related to the newly awakened interest in land drainage in the UK. Why he left such an eminent band of pure physicists to work largely by himself on a rather obscure (at least in the world of physics) subject was never really discussed by him, although it was said that he found the pressure in a competitive environment for quick publication of results at the expense of scientific thoroughness to be unacceptable. The result of the move was that someone with an uncompromising and critical mind for thoroughness in physics was thrown into the practical world of agriculture. Throughout his life, Childs' research philosophy was to develop a physical understanding of soil-water phenomena and to apply this to practical problems in agriculture, hydrology, and engineering. Two topics dominated Childs' work: one was water movement in unsaturated soils, and the other was groundwater movement to land drains.

In considering water in unsaturated soils, Edgar Buckingham, at the beginning of the century, had postulated that water was held in the interstices between solid particles by means of surface tension forces, and had defined the capillary potential dependent on the water content. A gradient of this capillary potential would produce movement of the water. L.A. Richards, in 1931, argued that Darcy's law would be obeyed in unsaturated soils because the air-filled pores could be regarded as the same as the solid soil particles as far as water conduction was concerned. However, the hydraulic conductivity would decrease with the water content, since the pathways for water movement would be more tortuous, fewer, and narrower, with the larger pores draining first. The work of W.B. Haines at this time on the physics of hysteresis in the relationship between water content and soil-water pressure in unsaturated soils was a complicating factor. Childs saw that these basic physics concepts were of little use to agriculture unless they could be applied to practical situations.

The study of water-table heights in drained agricultural land had been pursued by groundwater hydrologists by applying Darcy's law through the work of Jules Dupuit, Philipp Forchheimer, and J. Boussinesq. These studies resulted in the now-familiar drainage equations, the first obtained by the Danish engineer L.A. Colding in 1864, and in the 1930s developed for practical drainage engineering by S.B. Houghoudt in the Netherlands. However, these developments were not rigorous enough for Childs, and he pursued a more basic analytical approach, as did Don Kirkham in the USA.

Childs brought his physics discipline to the ongoing work at Cambridge on the mole drainage of heavy clay soils in which unlined channels, formed by a bullet drawn through the soil, act as drains. In his first soil physics papers, he recognized that gravity played a minor role in the movement of soil water when these soils were unsaturated, and gradients of soil-water pressure were all-important. In order to be able to predict the soil-water movement, he hypothesized that water moves in soils according to the diffusion equation that is amenable to mathematical solution for known boundary conditions. He was thus able to demonstrate how the moisture profile reacts to various surface conditions of evaporation and rainfall, predicting soil-water behavior and comparing analytical results with those observed in measurements on the Gault clay soil at the Cambridge University Farm, as well as with other published results. A decade later, Childs introduced the concept of soil-water diffusivity that depends on the soil-water content. Nevertheless, in the series of papers on water movement in heavy clay soils that used a constant diffusion coefficient to analyze soil-water profiles, he concluded "that the theory of diffusion of water through soil accounts qualitatively for the water movement from wet to dry soil, fits the best experiments numerically, and is at present alone in doing this." This still summarizes the situation so long as it is recognized that the diffusion coefficient is dependent on soil-water content and that difficulties associated with hysteresis occur in using diffusion theory, difficulties that were apparent to Childs even in this series of papers.

This analytical work on the movement of water through heavy clay soils was accompanied by an interest in practical problems of mole drainage, the common method of draining heavy clay soils. He developed an automatic recorder to measure the rapid increase and decrease of flow rates from mole drains and employed it in studies of drain deterioration with the annual weather cycle. This deterioration depends on the stability of clay soils to cycles of wetting and drying. Childs used moisture-retention curves, which he termed 'moisture characteristics,' after the characteristics of the thermionic valve with which he was familiar through his earlier work in pure physics, to indicate the pore-size distribution of soil samples. He devised a test to determine the stability of clay soils to wetting cycles, and hence their suitability for mole drainage, by subjecting samples of soil crumbs to different times of wetting and then measuring their moisture characteristics. The stability of a soil was shown by its ability to maintain large pores that drained at small tensions. By plotting the slope of the moisture characteristic against the pressure deficiency, he was thus able to show the stability of the Gault clay and the instability of London clay when left undisturbed in water. Besides this work on soil behavior during mole drainage, Childs discussed the mechanics of the design of the mole plough itself. For this early work, he was awarded the ScD degree of the University of Cambridge in 1945.

Childs' studies on moisture characteristics of soil crumbs in the context of the stability of mole drains led to more fundamental studies connected with the interpretation of such curves in terms of their pore-size distribution that determines the conduction of water through soils. Measurement of the hydraulic conductivity in unsaturated soils was a time-consuming and difficult task. Childs and N. Collis-George led the way in modeling the hydraulic conductivity function from the pore-size distribution obtained from the moisture characteristic. They compared their modeled results with accurately measured values of the hydraulic conductivity of unsaturated sands obtained using the method of flow to a water table down long columns that they devised.

While the Richards equation gives the basis for describing the moisture-profile development in soil profiles under unsaturated conditions, computing such profiles from this equation was practically impossible with the tools available in the middle of the twentieth century. Research workers, if they were lucky, used electromechanical calculators, but most relied on slide rules, logarithmic tables, and pencil and paper to do their calculations, so little progress seemed possible. Childs' recollection of his earlier work, in which he hypothesized that water movement could be described by the diffusion equation, led him to the diffusion form of the Richards equation with a soil-water diffusivity dependent on the water content. This was the beginning of the theoretical calculation of moisture profile developments from the Richards equation.

It is interesting to note that Don Kirkham and C.L. Feng concluded in 1949 that "the differential equation of diffusion theory is not valid" while producing moisture profiles during horizontal infiltration that can be analyzed to give a moisture content-dependent diffusivity. They were clearly of the opinion that the diffusion coefficient needed to be a constant.

Advances were being made in the 1950s in the numerical solution of the nonlinear diffusion equation in other scientific disciplines. Arnold Klute, using a soil water-dependent diffusivity, used these advances to compute the moisture profiles with the steep wetting fronts that occur during horizontal infiltration. Later, John Philip developed a very efficient method of solving numerically the highly nonlinear diffusion equation, and he extended this to the situation of vertical infiltration, when the effect of gravity has to be considered. Computers have allowed great improvements in numerical techniques of solving the Richards equation in two and three dimensions, as well as the simpler one-dimensional case, making the use of the diffusion approach redundant. Nevertheless, the diffusion approach that Childs introduced produced the progress in soil-water studies that occurred in the middle part of the last century. A particular short-coming of the diffusion analysis of soil-water flow is that it is limited to profiles either being entirely in the wetting state or in the draining state, because at a reversal of trend the soil-water diffusivity is discontinuous. In describing the soil-water distribution during the redistribution of infiltration water, Childs and his colleagues at Cambridge were quick to point out that the diffusion approach was unsuitable in situations where a reversal of wetting or draining takes place.

Childs' introduction to the physics of soil water through the drainage of heavy soils that was ongoing in Cambridge in the 1930s led to an interest in land-drainage theory in general. He was critical of the general acceptance of work based on applying the approximate Dupuit–Forchheimer analysis that properly relates to shallow flow to ditches that penetrate down to a horizontal impermeable floor. He was particularly critical of drainage equations incorporating the concept of radial flow to drains to account for a depth of soil below drain level. He thus turned his attention to a study of land drainage using a fundamental physics approach.

Childs noted that the water flow in soils was analogous to the flow of electricity in conductors, and this led him to pioneer the use of electric analogues in the study of the two-dimensional problem of steady-state drainage of uniform rainfall to pipe drains. To do this he produced a conducting paper that simulated the soil, by carefully spraying filter paper with a graphite suspension in water to produce a paper with uniform conductivity. The marketing of commercial Teledeltos graphited paper was later a great help in this work. In this way, Childs was able to trace the streamlines and equipotentials in drained lands under given rainfall conditions and obtain the position and shape of the water table. 'Rain' was introduced through electrodes along the 'water table' whose position was found by trial and error and cut to shape to give the requirement that the potential there was equal to the height above the 'drain' formed by an electrode cemented to the paper. This work was of great significance in correcting earlier viewpoints regarding land drainage.

It seemed to Childs that there were limits to the insights to be gained by rigid mathematical analysis of problems relating to the flow of fluids in porous materials, especially those relating to the drainage of agricultural lands, and further progress was possible only by model or a full-scale study of porous materials themselves. Childs therefore had constructed a laboratory containing a large sand tank with the aim of furthering the understanding of the physics of land drainage. This was built on Cambridge University Farm. It formed the nucleus of the Agricultural Research Council Unit of Soil Physics, affectionately known as the 'Tank,' which was formed in 1951 and disbanded in 1977. Childs, an Assistant Director of Research at the University of Cambridge and later the Reader in Soil Physics, was the Unit's Director. The tank, which measured 10 m × 10 m × 1.5 m, was filled with uniform Leighton Buzzard sand. There was the facility to provide uniform 'rain' of various intensities on to the sand surface. Portholes in the side of the tank provided outlets for drains that could be installed in the sand. Pressure cells were developed and installed in the sand so that pressure heads could be continuously recorded during experiments.

Experiments in the sand tank were used to develop and test methods of measuring hydraulic conductivity below the water table as well as for its main purpose of advancing our understanding of water-table behavior in drained lands. From work done in the Unit's sand tank, advances were made concerning: the influence of depth of soil below drain level on water-table heights; the non-steady-state problem of the moving water table above drains with changing rainfall intensity, including a reappraisal of the concept of specific yield; the hydraulic behavior around gappy drains; and the water-table behavior in three-dimensional drainage situations that was inspired by mole drains being laid above lateral drain channels, recalling Childs' early work when he joined the School of Agriculture. Present-day modeling of water regimes in drained lands depends much on the basic understanding that emerged from these studies.

Having built the large sand-tank laboratory, Childs left others at Cambridge to use it. Childs' attention was diverted to the use of conformal mapping techniques to solve the problem in potential theory of the difficult and unusual boundary conditions presented by land drainage. This technique of solving two-dimensional groundwater problems had long been used by Russian scientists, but in general their work had gone unnoticed in the West. Childs learned of the solutions given by the Dutch mathematician J.J. van Deemter in 1950, and the Danish mathematician Frank Engelund in 1951, to the steady-state drainage problem of uniform rainfall and upward artesian water flow to uniformly spaced cylindrical drains installed in an infinitely deep soil. These gave the shape and height of the water table for drains of given diameter. Van Deemter also considered the problem of ditch drainage when a surface of seepage occurs on the ditch wall above ditch-water level. The analysis of the drainage problem using conformal mapping techniques gave a much more rigorous approach to the determination of water-table heights in drained lands than theory developed from the Dupuit–Forchheimer analysis.

Childs made an attempt to extend the analysis to the situation where there was an impermeable barrier at some depth, but perhaps the more significant use of the theory given by van Deemter and by Engelund was its use in the consideration of the nature of the drain channel. It was accepted that drains did not conform to the uniform sinks that theory assumed, because they were mostly of a gappy nature, being constructed initially from tiles butted together and then from slotted plastic pipes with water entry taking up a limited area of the drain surface. It was argued, and confirmed experimentally in the Unit's sand tank, that the converging flow to the gaps in the drain produced almost cylindrical equipotentials within a short radial distance, thus allowing the oncept of 'equivalent drain radius' to be applied. This concept could simply be used in van Deemter's and Engelund's analyses to show the effect on the water-table height due to the resistance effect of the gappy nature of the drain channel. In particular, it was noted that the gappy nature of the drain had a relatively small effect on the water table midway between drain lines, but raised the water table above the drains significantly so that the water table became flatter the greater the drain resistance. All this time, during which Childs was attracted to more theoretical studies, work progressed in the Unit's sand tank.

From early times in his career in soil physics research, Childs was very mindful of the role hysteresis played in the way water redistributed itself in soil profiles. It is impractical to undertake the amount of measurement required to obtain a complete trace of the hysteresis in soil-water relationships, so that bringing hysteresis into a rigorous analysis of soil-water movement seemed impossible. It was with some excitement that Childs' attention was drawn to the independent-domain theory proposed by L. Néel in the context of magnetic hysteresis. This excitement was short-lived, however, when it was realized that pores that could be viewed as the 'domains' behaved far from independently. Further work, particularly by Ed Miller and Alex Poulovassilis, led to the dependent-domain theory of hysteresis that would apply to soil-water relationships. At the time of his death, Childs was working on a unified-domain theory of hysteresis that remained uncompleted. Generally, however, it appears that the amount of experimental measurements required to obtain the necessary data for the dependent-domain theory is practically the same as that required for the direct tracing of the hysteresis paths, so that little is achieved. Despite this, what domain theories may give are insights into the pore structure of soils.

Childs died at the age of 65 on 24 May 1973. His legacy was a fundamental physics approach to soil physics that inspired those who worked with him, as well as those who followed in his footsteps. His work emphasizes the importance of a basic understanding of soil physics phenomena in tackling practical soil problems.

See also: **Capillarity**; **Darcy's Law**; **Diffusion**; **Drainage, Surface and Subsurface**; **Hydrodynamics in Soils**; **Hysteresis**; **Infiltration**; **Macropores and Macropore Flow, Kinematic Wave Approach**; **Porosity and Pore-Size Distribution**

Further Reading

Childs EC (1940) The use of soil moisture characteristics in soil studies. *Soil Science* 53: 239–252.

Childs EC (1942) Stability of clay soils. *Soil Science* 53: 79–92.

Childs EC (1950) The water table, equipotentials, and streamlines in drained land. *Soil Science* 56: 317–330.

Childs EC (1969) *An Introduction to the Physical Basis of Soil Water Phenomena*. London, UK: John Wiley.

Childs EC and Collis-George N (1950) The permeability of porous materials. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* 201: 392–405.

Childs EC and Youngs EG (1974) Soil physics: twenty-five years on. *Journal of Soil Science* 25: 408–419.

Gardner WH (1986) Early soil physics into the mid-20th century. *Advances in Soil Science* 4: 1–101.

Nicholson HH (1942) *The Principles of Field Drainage*. London, UK: Cambridge University Press.

Youngs EG (1977) *The Agricultural Research Council Unit of Soil Physics, 1951–1977*. University of Cambridge Department of Applied Biology Memoir, 49, pp. 4–10. Cambridge, UK: Cambridge University Press.

Youngs EG (1983) The contribution of physics to land drainage. *Journal of Soil Science* 34: 1–21.

Youngs EG (1984) Developments in the physics of land drainage. *Journal of Agricultural Engineering Research* 29: 167–175.

Youngs EG, Towner GD, and Poulovassilis A (1974) In memoriam: Ernest Carr Childs. *Soil Science* 117: 241–242.

CIVILIZATION, ROLE OF SOILS

D Hillel, Columbia University, New York, NY, USA

Human Management of the Soil

Manipulation and modification of the environment was a characteristic of many societies from their very inception. Long before the advent of earth-moving machines and toxic chemicals, even before the advent of agriculture, humans began to affect the land and its biota in ways that tended to destabilize natural ecosystems. In many of the ancient countries, where human exploitation of the land began early in history, we find disturbing examples of once-thriving regions reduced to desolation by human-induced degradation. Some of the early civilizations succeeded all too well at first, only to set the stage for their own eventual demise. The poor condition of the Fertile Crescent today is due not simply to changing climate or to the devastation caused by repeated wars, though both of these may well have had important effects. It is due in large part to the prolonged exploitation of this fragile environment by generations of forest cutters and burners, grazers, cultivators, and irrigators, all diligent and well-intentioned but destructive nonetheless.

An example of soil abuse on a large scale can be seen in the rainfed parts of the Mediterranean region, which has borne the brunt of human activity more intensively and for a longer period than any other region on earth. Visit the hills of Israel, Lebanon, Greece, Cyprus, Crete, Sicily, Tunisia, and southeastern Spain. There, rainfed farming and grazing were practiced for many centuries on sloping terrain, without consistent or fully effective soil conservation. The land has been denuded of its natural vegetative cover, and the original mantle of fertile soil has been raked off by the rains and carried down the valleys toward the sea. That may have been the reason why the Phoenicians, Greeks, Carthaginians, and Romans, each in turn, were compelled to venture away from their own country and to establish far-flung colonies in pursuit of new productive land. The end came for each of these empires when it had become so dependent on distant and unstable sources of supply that it could no longer maintain central control or ward off growing competition from other land-hungry nations.

Consider, for another example, the southern part of Mesopotamia. Aerial and satellite photographs of this area, now part of Iraq, reveal wide stretches of barren, salt-encrusted terrain. Long ago, these were fruitful fields and orchards, tended by enterprising irrigators whose very success inadvertently doomed their own land. The once-prosperous cities of Mesopotamia are now 'tells,' mute time capsules in which the material remnants of a civilization that lived and died there are entombed. Similarly ill-fated was the ancient civilization of the Indus Valley in present-day Pakistan.

There were, on the other hand, some societies that did better than others. The more successful ones were those who were able to develop modes of soil and land management that enabled them to thrive in the long run. Impressive evidence exists regarding the terrace-building farmers of eastern Asia and the Near East, as well as the wetlands-based societies of Meso-America and South America. Remarkably productive wetland management systems have survived in China and other parts of Southeast Asia. In contrast with the irrigation-based civilization of Mesopotamia, the similarly based civilization of Egypt sustained itself for more than five millennia – though it too (owing to its intensified management, impelled by its 20-fold increase of population in the last two centuries) is now beset with problems of waterlogging and salinity.

Historical Attitudes Toward the Soil

Early societies generally revered the earth and tended to deify it. The earth was held sacred as the embodiment of a great spirit, the creative power of the universe, manifest in all phenomena of nature. The earth spirit was believed to give shape to the features of the landscape and to regulate the seasons, the cycles of fertility, and the lives of animals and humans. Rocks, trees, mountains, springs, and caves were recognized as receptacles for this spirit.

The cult of the earth is perhaps the oldest and most universal element in all religions. The Australian aborigines and the African Bushmen, among the last to have maintained the preagricultural hunter-gatherer mode of life, have regarded the earth as the Great Provider, the source of all sustenance. So did – and still do – the native Americans. The ancient Egyptians represented the earth as the god Geb, who mated with the sky goddess Nut. The sexual roles were reversed in the culture of the ancient Canaanites, who worshiped the male Baal as the god of the sky, who provided the rain that fructified the earth goddess Ashera. To the ancient Greeks, the earth was Gaea, the maternal goddess who, impregnated by her son and consort Uranus (god of the sky), became mother of the Titans and progenitor of all the many gods of the Greek pantheon.

In the Hebrew Bible, there are two very different accounts of creation and the role granted or assigned to humanity in the scheme of life on earth. In the first chapter of Genesis, we read that God (called 'Elohim') decided to "make man in our own image, and let them rule over the fish of the sea, and over all the earth, and over every creeping thing that creepeth upon the earth." And God blessed man and woman and said unto them: "Be fruitful, and multiply, and fill the earth, and conquer it… Here, I have given you every herb yielding seed and every tree with fruit… to you shall it be for food."

But the divine injunction to humanity is defined quite differently in the second chapter of Genesis. In this version, God (called 'Yahweh') "formed man out of the soil of the Earth and blew into his nostrils the breath of life, and man became a living soul." Then God planted a garden in Eden in the east and placed the man therein… to serve and preserve it." (Those words are this author's translation of the Hebrew words 'l'ovdah ul'shomrah' (Genesis 2:15), usually rendered "to dress it and keep it" (King James Version; or "to till it and keep it," Revised Standard Version.) Here, humanity is not given license to rule over the environment for self-gratification, but – quite the contrary – is charged with the responsibility to nurture and protect it.

Thus, latent in one of the main founts of Western Civilization we have two opposite perceptions of humanity's destiny. One is anthropocentric: humans are set above nature, to be its omnipotent masters. They are endowed with the power and the right to dominate all other creatures, toward whom they have no obligations. The other view is more modest. The human earthling is made of soil and is given a "living soul." There is no mention of being made "in the image of God." Humanity's appointment is not an ordination but an assignment, which is to serve as the custodians of God's garden.

Over the generations, unfortunately, it has generally been the arrogant and narcissistic view, implied in the first Biblical account, that has prevailed. It has repeatedly been cited and used as a religious justification or rationale for the unbridled and relentless exploitation of the environment. The imperative now is to accept and realize the long-ignored second view of our proper role in relation to nature.

Readers of the Bible in translation miss much of the imagery and evocative verbal associations in the original language. The indissoluble link between humanity and soil is manifest in the very name 'Adam,' derived from 'adamah,' a Hebrew noun of feminine gender meaning earth or soil. Adam's name encapsulates the notion that his existence is derived from the soil, to which he is tethered throughout life and to which he is fated to return at the end of his days. Likewise, the name assigned to Adam's mate, 'Hava,' rendered Eve in transliteration, literally means 'living.' In the words of the Bible, "Adam called his wife Hava because she was the mother of all living." Together, therefore, Adam and Eve signify "Soil and Life."

The ancient association of humanity with soil is echoed in the Latin name for man, 'homo,' derived from 'humus,' the stuff of the soil. This powerful metaphor suggests an early realization of a profound truth that has since been too often disregarded. Since the words 'humility' and 'humble' also derive from humus, it is rather ironic that we should have assigned our species so arrogant a name as *Homo sapiens sapiens* ('Wise Wise Man'). Perhaps a more appropriate and certainly more modest name would be *Homo sapiens curans*, with the last word denoting caring or caretaking, as in 'curator.'

Human Origins

Our species' birthplace was evidently in the continent of Africa, and its original habitat was probably the subtropical savannas that constitute the transitional areas of sparsely wooded grasslands lying between the zone of the humid and dense tropical forests and the zone of the semiarid steppes. We can infer the warm climate of our place of origin from the fact that we are naturally so scantily clad, or furless; and we can infer the open landscape from the way we are conditioned to walk, run, and gaze over long distances.

For at least 90% of its career, the human animal existed merely as one member of a community of numerous species who shared the same environment. Humans were adapted to subsist within the bounds defined by the natural ecosystem. By and large, our ancestors led a nomadic life, roaming in small bands, foraging wherever they could find food. They were gatherers, scavengers, and opportunistic hunters. Unlike their primate cousins who remained primarily vegetarian, humans diversified their diet to include the flesh of whichever animals they could find or catch, as well as a variety of plant products such as nuts, berries and other fruits, seeds, some succulent leaves, bulbs, tubers, and fleshy roots.

The story of how humans gradually ventured far from their original birthplace to range over a variety of climates and landscapes is a remarkable saga of audacity, ingenuity, perseverance, and adaptability. The mode of human adaptation was not entirely genetic or physical: there was not enough time for that. Rather, their adaptation was in large part behavioral.

Instead of relying on physical prowess, they had to use inventiveness to survive the elements and to compete successfully against stronger animals. The increase in brain size and manual dexterity, as well as the invention of various tools and stratagems, gradually enabled humans to overcome the constraints of their ancestry.

By 1 million years ago, hominids had become taller (approximately 1.5 m in height) and had acquired a larger brain. Some time later, the so-called *Homo erectus* learned to set and use fire, probably at first only for cooking and softening food. That achievement, along with the fashioning of stone tools, was a momentous innovation, celebrated in the Greek myth of Prometheus. Eventually, it had a great effect on the environment. Evidence has been found in southern and eastern Africa of repetitive occurrences of brush fires, whether purposeful or accidental, apparently set by humans nearly a million years ago. This early manifestation of pyrotechnology signifies the beginning of human manipulation of the earth's ecosystems. The use of fire became even more important when humans moved out of the tropics into colder climes, where bonfires and hearths were needed to warm their shelters in winter, as well as to cook their food.

The Paleolithic Transformation

At some point, humans began to use fires deliberately and systematically to flush out game and to modify the vegetation. The resultant suppression of woody plants and the fertilizing effect of ash encouraged the growth of herbaceous plants and improved their nutritional quality. This benefited foraging species and raised the carrying capacity for game animals. It also facilitated travel and hunting by humans. In time, the practice of clearing woodlands and shrublands by repeated firings also set the stage for the advent of agriculture.

As vegetation is affected by fire-setting hunters, so are soils. Following repeated fires and deforestation, soil erosion and landslides often result in the greatly increased transport of silt by streams, and in the deposit of that silt in river valleys and estuaries. The dating of fluvial sediments in river valleys in England, for example, suggests that they were the products of erosion caused by anthropogenic clearings in the originally closed deciduous forest during the Late Paleolithic period.

By approximately 40 000 years ago, modern humans, evidently indistinguishable from us today in physical features and in intelligence, had gained dominance. Clad in sewn garments made of animal skins, able to make and use a variety of implements and weapons, humans were able to range and settle in locations and climes far from their ancestral home. All the while they continued to evolve by natural selection, increasingly aided by cultural and technological development. They also contrived increasingly sophisticated methods of obtaining and storing foods, including the selective gathering, processing, and preservation of biological products, and eventually the domestication of plants and animals.

The described series of changes has been termed the Paleolithic (Early Stone Age) Transformation. It was marked by the development of adaptive mechanisms and modes of social organization suited to exploiting potentialities within the environment. Each modification of the environment entailed additional human responses, which in turn further modified the environment, so that a process of escalating, dual metamorphosis was instigated. Human intelligence and culture were both cause and effect in that fateful interplay. The peculiarly dynamic and progressive evolution of human ecology is the true history of our species.

The Agricultural Transformation

The gradual intensification of land use continued throughout the Paleolithic period, so that by its later stages nearly all the regions of human habitation had experienced some anthropomorphic modification of the floral and faunal communities. At some stage, humans began to delineate sections of the environment that they could control and manage to suit their special needs, and in which they could find secure shelters for habitation.

The process of intensification of land use can be seen as an adaptation to increasing population pressure. Several millennia of occupation by hunter-gatherers, even at a very low density and slow rate of population growth, filled up the terrain and decimated the natural forageable resources to the point where subsistence was difficult. The choice was then between migration and some form of intensification aimed at inducing the same area to yield a greater supply. The selective eradication of undesirable animal species and the encouragement of desirable ones led eventually to domestication and herding. Similarly, selective manipulation of plant communities involved suppressing some species and promoting the growth of others. The entire series of activities quite logically led to plant domestication and propagation, and to purposeful soil management aimed at creating favorable conditions for crop production – that is to say, these activities culminated in the development of agriculture and the agricultural way of life.

The Agricultural Transformation is very probably the most momentous turn in the progress of humankind, and many believe it to be the real beginning of civilization. Often called the Neolithic Revolution, this transformation apparently first took place in the Near East, approximately 10 000 years ago, and was based on the successful domestication of suitable species of plants and animals. The ability to raise crops and livestock, while resulting in a larger and more secure supply of food, definitely required attachment to controllable sections of land and hence brought about the growth of permanent settlements of larger, coordinated communities. The economic and physical security so gained accelerated the process of population growth and necessitated further expansion and intensification of production. A self-reinforcing and self-perpetuating pattern thus developed, so the transition from the nomadic hunter-gatherer mode to the settled farming mode of life became in effect irreversible.

The Agricultural Transformation radically changed almost every aspect of human life. Food production and storage stimulated specialization of activities and greatly enhanced the division of labor that had already started in hunting-gathering societies. The larger permanent communities based on agriculture required new forms of organization, both social and economic. Domestication affected family structure and the roles and status of men, women, and children. With permanent facilities such as dwellings, storage bins, heavy tools, and agricultural fields came the concept of property. The inevitably uneven allocation of such property resulted in self-perpetuating class differences. Religious myths and rituals, as well as moral and behavioral standards, developed in accordance with the new economic and social constellation and the new relationship between human society and the environment.

The evolution of agriculture left a strong imprint on the land in many regions. The vegetation, animal populations, slopes, valleys, and soil cover of land units were radically altered. The processes of tillage and fallowing, of terracing, of irrigation and drainage have had considerable consequences for such processes as the erosion of slopes and the aggradation of valleys, as well as the formation of deltas in seas and lakes where silt from the land surface naturally comes to rest. Soil lost from deforested and subsequently cultivated slopes is unlikely to be regenerated unless the land is allowed to revert to its forest cover for many scores, perhaps even many hundreds, of years.

Soil Husbandry and Ceramics

An important factor in the evolution of agriculture in the Near East, as elsewhere, was the development of the tools of soil husbandry. Seeds scattered on the ground are often eaten by birds and rodents, or subject to desiccation, so their germination rate is likely to be very low and uneven. Given a limited seed stock, farmers would naturally do whatever they could to promote germination and seedling establishment. The best way to accomplish this is to insert the seeds to some shallow depth, under a protective layer of loosened soil, and to eradicate the weeds that might compete with the crop seedlings for water, nutrients, and light.

The simplest tool developed for the purpose was a paddle-shaped digging stick, by which a farmer could make holes for seeds. The use of this simple device was extremely slow and laborious, however, so at some point the digging stick was modified to form the more convenient spade, which could not only open the ground for seed insertion but also loosen and pulverize the soil and eradicate weeds more efficiently. In time, the spade developed a triangular blade, at first made of wood, but later made of stone, and eventually of metal. Such a spade, initially designed to be used by one person, was later modified so that it could be pulled by a rope so as to open a continuous slit, or furrow, into which the seeds could be sown. A second furrow could then be made alongside the first, to facilitate seed coverage. In some cases, the rows were widely enough separated to permit a person to walk between the rows, weeding the cultivated plot.

The human-pulled traction spade or ‘ard’ gradually metamorphosed into an animal-drawn plow. The first picture of such a plow, dating to 3000 BCE, was found in Mesopotamia, and numerous later pictures have been found both there and in Egypt, as well as in China. It was not long before these early plows were fitted with a seed funnel, so that the acts of plowing and sowing could be carried out simultaneously. The same ancient implement is still in use today in parts of the Near and Middle East.

While the development of the plow represented a huge advance in terms of convenience and efficiency of operation, it had an important side effect. As with many other innovations, the benefits were immediate, but the full range of consequences took several generations to play out, long after the new practice became entrenched. The major environmental impact was that plowing made the soil surface – now loosened, pulverized, and bared of weeds – much more vulnerable to accelerated erosion. In the history of civilization, contrary to the idealistic vision of the prophet Isaiah, the plowshare may well have been far more destructive than the sword.

As farming induced sedentary living in villages, there also developed an important new industry that depended directly on the soil – pottery, which began in the Near East at approximately 6000 BCE. The shaping and baking of clay to form hardened vessels for grain, for liquid storage and conveyance, and for cooking, represented the first transmutation of material by humans. Such an innovation could not have been possible, owing to the fragility of the ceramic objects, during the nomadic hunting-gathering phase.

From Rainfed to Irrigated Farming

The Mediterranean-type climate of the Near East is at best semihumid, but more typically semiarid, with a rather high incidence of drought. Hence the practice of rainfed farming could not provide anything like total food security. The early farmers, who depended only on seasonal rainfall to water their crops, were always at the mercy of a capricious and highly unpredictable weather regime. The Hebrew Bible, for instance, is replete with references to the ever-present threat of drought and consequent famine. In time of need, therefore, it was only natural for farmers located near river courses to attempt to augment the water supply to crops by diverting water from the river. It was also reasonable to try to raise crops on riverine flood plains that were naturally inundated, and thereby irrigated, periodically.

Thus, many centuries after its advent, farming was extended from the relatively humid centers of its origin toward the extensive river valleys of the Tigris-Euphrates, the Nile, and the Indus. As the climate of these river valleys is quite arid, a new type of agriculture based primarily or even entirely on irrigation came into being. With a practically assured perennial water supply, an abundance of sunshine, a year-round growing season, deep and fertile soils, and relative security from the hazards of drought and erosion that beset rainfed farming, irrigated farming became a highly productive enterprise. However, behind its success lurked an insidious problem that could not have been foreseen initially: the problem of soil waterlogging and salination.

Silt and Salt in Ancient Mesopotamia

Ancient Mesopotamia owed its prominence to its agricultural productivity. The soils of this alluvial valley are deep and fertile, the topography is level, the climate is warm, and water is provided by the twin rivers Euphrates and Tigris. However, the diversion of river water onto the valley lands led to a series of interrelated problems.

The first problem was sedimentation. Early in history, the upland watersheds were deforested and overgrazed. The resulting erosion was conveyed by the rivers as suspended silt, which settled along the bottoms and sides of the rivers, thus raising their beds and banks above the adjacent plain. During periods of floods, the rivers overflowed their banks, inundated large tracts of land, and tended to change course abruptly. The silt also settled in channels and clogged up the irrigation works.

The second and more severe problem was salt. Seepage from the rivers, the irrigation channels, and the flood-irrigated fields caused the water table to rise throughout southern Mesopotamia. Because all irrigation waters contain some salts, and because crop roots normally exclude salts while extracting soil moisture, the salts tend to accumulate in the soil and groundwater. As the undrained water table rose, it took the salts back into the soil.

The farmers of ancient Mesopotamia attempted to cope with the process of salination by periodically fallowing their land, and by replacing the salt-sensitive wheat with relatively salt-tolerant barley. However, the process proceeded inexorably; so the ancient hydraulic civilizations of Sumer, Akkad, Babylonia, and Assyria each, in turn, rose and then declined, as the center of population and culture shifted over the centuries from the lower to the central to the upper parts of the Tigris-Euphrates valley.

The Sustainability of Egyptian Agriculture

In contrast to Mesopotamia, the civilization of Egypt thrived for several millennia in the same location. What explains the persistence of irrigated farming in Egypt in the face of its demise in southern Mesopotamia? The answer lies in the different soil and water regimes of the two lands. Neither clogging by silt nor poisoning by salt was as severe along the Nile as in the Tigris-Euphrates plain.

The silt of Egypt is brought by the Blue Nile from the volcanic highlands of Ethiopia, and it is mixed with the organic matter brought by the White Nile from its swampy sources. It was not so excessive as to choke the irrigation canals, yet was fertile enough to add nutrients to the fields and nourish their crops. Whereas in Mesopotamia the inundation usually comes in the spring, and summer evaporation tends to make the soil saline, the Nile rises in the late summer and crests in autumn. So in Egypt the inundation comes at a more favorable time: after the summer heat has killed the weeds and aerated the soil, just in time for the prewinter planting of grain.

The ancient Greek name for Egypt was ‘Khemia,’ from the word ‘Khami’ signifying black soil, which was what the Egyptians themselves called their land. So fabulously fertile was that dark deposit of the Nile (which contrasted vividly with the yellowish color of the nearby desert sand) that the Greeks considered it the mother lode of all material substances and apparently named the science of materials after it. That name has been transmuted into our term ‘chemistry.’

The narrow floodplain of the Nile (except in the Delta) precluded the widespread rise of the water table. Over most of its length, the Nile lies below the level of the adjacent land. When the river crested and flooded the land, the seepage naturally raised the water table. As the river receded and its water level dropped, it pulled the water table down after it. The all-important annual pulsation of the river and the associated fluctuation of the water table under a free-draining floodplain created an automatically repeating, self-flushing cycle by which the salts were leached from the irrigated land and carried away by the Nile itself.

The basis of Egypt's civilization was the nearly optimal combination of water, soil, nutrients, and organic matter, provided by the regular annual regime of the river, which was more dependable and timely than the relatively capricious floods of Mesopotamia. It enabled Egyptian farmers to produce a surplus that fed the artisans, scribes, priests, merchants, noblemen, and, above all, the Pharaohs, who used their power to order the building of monuments. Those monuments still stand today, less in testimony to the kings who ordered them than to the diligence and organization of a society of labor rooted in the soil.

See also: **Desertification**; **Erosion:** Water-Induced; **Irrigation:** Environmental Effects; **Salination Processes**

Further Reading

Carter VG and Hale T (1974) *Topsoil and Civilization.* Norman, OK: University of Oklahoma Press.

Hillel Daniel (1992) *Out of the Earth: Civilization and the Life of the Soil.* Berkeley, CA: University of California Press.

Hyams Edward (1976) *Soil and Civilization.* New York: Harper & Row.

Jacobsen Thorkild (1982) *Salinity and Irrigation in Antiquity.* Malibu, CA: Udena Publications.

Mitchell John (1975) *The Earth Spirit: Its Ways, Shrines, and Mysteries.* New York: Thames and Hudson.

Redman CL (1978) *The Rise of Civilization: From Early Farmers to Urban Society in the Ancient Near East.* San Francisco, CA: W. H. Freeman.

Rindos David (1984) *The Origins of Agriculture: An Evolutionary Perspective.* San Diego, CA: Academic Press.

Thirgood JV (1981) *Man and the Mediterranean Forest: A History of Resource Depletion.* New York: Academic Press.

Classification of Land Use *See* **Land-Use Classification**

CLASSIFICATION OF SOILS

R W Arnold, Formerly with USDA–Natural Resources Conservation Service, Washington, DC, USA

A classification is an organized body of knowledge about something of interest. It is intended to show relationships among and between entities, and to help recall important properties of these entities. Three principles deal with the setup of a classification: 'purpose' states the reasons for wanting to organize soil knowledge; 'domain' specifies the universe of objects relevant to the purpose; and 'identity' defines and names the individual members of the domain. Four additional principles deal with the organization of a system: 'differentiation' specifies a protocol-guided hierarchical structure with categories and classes within categories; 'prioritization' is evident by sequencing categories and sequencing classes within categories; 'diagnostics,' whereby selected soil properties and features (diagnostics) are quantified, provide objectivity; and 'membership' is based on quantified class limits and described central tendencies. A final principle of certainty recognizes change as inevitable and the driving force for continual testing of a system.

Soils occur in most terrestrial environments and commonly have observable properties such as color and structure, arranged as layers; that is, soils have morphology. There are patterns of soil morphology throughout the world that are systematic enough to suggest causal relationships with other features of location.

In the late 1800s, 'pedology' originated from 'genetic soil science,' as termed by V.V. Dokuchaev, a Russian. He envisioned soils to be natural bodies of transformed materials at or near the interface of the lithosphere, biosphere, and atmosphere (Figure 1). Soils were recognized as cause-and-effect results of processes that were, and are, influenced by natural environmental factors and conditions: the factors of climate, biota, parent materials, and relief (landscape features) interact over time to produce soils. This was the most fundamental change in the concept of

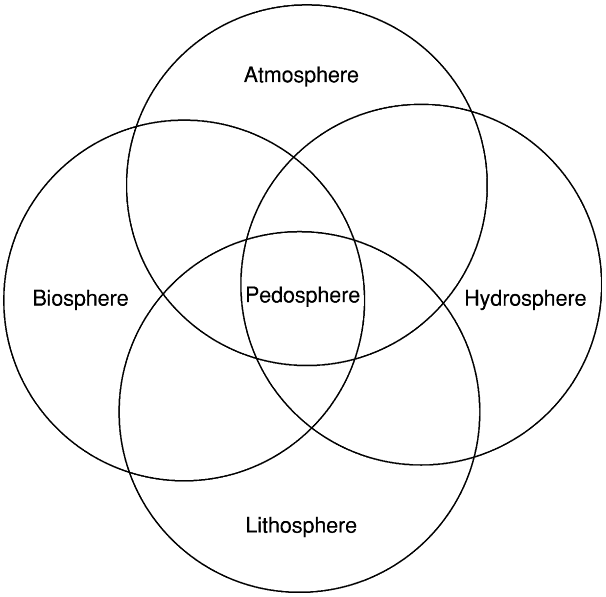


Figure 1 The pedosphere results from the interactions of associated spheres of influence. Reproduced with permission from the international journal cover of *Pedosphere*.

soil in history and significantly altered the manner in which soils were classified.

Although most soil interpretation schemes are also classifications, the more popular soil classification schemes of the twentieth century have been morphogenetic ones based on hypotheses and theories of soil development and pedological transformation of so-called parent materials into the entities called soils. Soil properties were described and genetic hypotheses proposed to explain the presence and spatial occurrence of soils. Names were given to different morphologies, e.g., Podzols and Chernozems, and they served as short-hand identifiers of the natural bodies in the environment.

Soil classification has seldom been static, consequently each system is an abstract of knowledge about soils at a given moment. Over time, technology has provided opportunities to make more precise measurements, enabling a major shift from qualitative definitions to quantitative ones. Field studies and soil survey have revealed the complexity and spatial intricacies of global pediment formation in unconsolidated materials. Such evidence greatly influenced soil genesis research and models of soils as landscapes. The significance of the time factor has become more apparent as details of paleopedology have been unraveled. The realities of polygenetic cycles of soil genesis have challenged prevailing theories of when and how soils form and develop. Not only have some soils developed in preconditioned geologic materials, but others have formed in previously developed soils, thus complicating the issue of inherited properties versus pedogenically formed ones.

Taxa limits in most classifications are direct consequences of theories, models, and experience; however, quantitative definitions of limits and even of central concepts of taxa are coldly factual, and departures of theory from facts become readily apparent. Some cherished concepts have received harsh treatment at the hands of precise definition.

As a source from which a thing proceeds, a principle serves as a foundation for the framework of knowledge about a population being classified. The following principles are fundamental statements upon which a hierarchical system of classification is developed and, as such, provide answers to commonly asked questions about classification. Eight principles are listed: the first three, purpose, domain, and identity, are concerned with the rationale of the system. The next four, differentiation, prioritization, diagnostics, and membership, are concerned with the organization of the information. The last principle, certainty, is concerned with the future viability of a system.

Principle of Purpose

Why do we want to organize soil knowledge? Each classification is dependent on the purpose, the rationale, of arranging current information in a systematic structure. With the new emphasis on genetic soil science, it was desirable to group soils according to their pathway or mode of formation. The groups could be related to each other based on the understanding of how soils obtained their properties, and by assigning names to these groups it helped recall many of their properties. Consequently the principle of purpose for most soil classification schemes is to show order in nature and to remember major features of soils. The purpose of the American Soil Taxonomy is to serve a soil survey program that is conducted to provide information about soils relative to their use and management. Considerable overlap with a purely genetic system is likely, but the selection of many properties is associated with a different purpose.

Principle of Domain

What do we want to include, or exclude, in our classification? In other words, what is the domain of interest? The realm of soils has many connotations, consequently it is necessary to specify what is to be included, and what is to be excluded, in a classification. At one time organic deposits were not considered to be soils and so were excluded. For some people soils must have a vegetative cover, for others they only need to be capable of supporting plant life.

The principle of domain requires that the universe of interest must be specified and, as such, will have to be defined. Instead of the collection of natural soil bodies, a domain might consist of terrestrial entities defined by subjective functional characteristics and associated environmental parameters. Thus, use-oriented groupings, in contrast to genetic ones, are not as concerned with order in nature; rather they generally show rankings of suitabilities of soils for specific uses. The choice of the domain depends on the purpose of the classification.

Principle of Identity

Who are the members of the domain of interest? How are they identified? A domain indicates the universe, or population, that is being considered; however, it does not define the members that will be the source of data for the classification. Where detailed soil survey has been important, the individuals are members of the taxa of the lowest category in a hierarchical system. Identity is a way of providing a name for a nondivisible component, or individual, that would not otherwise be recognizable. A major difference between some systems is the selection of the basic unit; in some systems, members of high-level taxa are the basic units, like the Russian Soil Type, whereas in others they are members of the lowest category, like the American Soil Series.

Examples of individuals of interest include: polypedon, pedon, profile, arbitrary body, soil landscape unit, and segment of a continuum. In use-oriented classifications, the members may be functional units or capability units rather than soil bodies themselves. Arbitrary soil individuals are meaningful reference units to compare classification schemes if their boundaries are independent of soil properties and class limits. Although precise definitions may be difficult and rather cumbersome, it is nearly impossible to proceed without resolving the principle of identity.

Identifying taxa at all categorical levels assists in visualizing relationships and patterns within the structure of the classification framework. A systematic nomenclature that indicates location within the system and also provides mnemonic links to important properties promotes understanding and use of the classification. The use of connotative elements has been successfully demonstrated in global systems, including Soil Taxonomy and the World Reference Base, and in several national schemes.

Principle of Differentiation

What kind of structure is needed? How should the domain be divided into groups, and how many groups are needed? How should the categories be defined? The sheer numbers of soils in the world, or even a small area of it, suggest that a hierarchical structure would be appropriate. There would be few classes in the higher categories and many more in the lower categories. It is possible to conceive of grouping individual soils together, and then combining those groups into more comprehensive groups, until the whole domain has been combined into only a few classes. It is more common to start with the domain and divide it into groups, then subdivide those groups into smaller, less-inclusive classes until the classes at the lowest level contain individuals with many features in common. It is sometimes forgotten that once a population is organized into a hierarchy it is only possible to employ the system from the top down, that is, by differentiating more and more specific classes. At the present time, the 'rules of engagement' for soil classification have only been devised, or clearly stated, for the processes of separation.

A hierarchical system of classification is a separation of a domain into successively more specific classes. It divides the domain into smaller groups of members that are also mutually exclusive classes. Consequently the differentiae (criteria) that separate classes focus on the limits or boundaries of the classes, and subordinate the central concepts of the classes. This is done for operational reasons; nevertheless the essence of a class is its central concept, and its description serves as the focus and image of the class.

The first separation of a domain into classes creates the highest category. The definition of this particular set of classes must be quite abstract to capture the intent of the classification. It may be genetic, geographic, interpretive, or functional, but it must apply to all members of the population (a rule sometimes forgotten). The definition of the category suggests possible indicators that are consistent with, and satisfy, the definition. Such marks or evidence are soil properties assumed to be closely related to the category definition. It is usual to expect classes of the highest category to reveal large areas when displayed on a global map indicating where soil-forming conditions are, or have been, similar enough to produce common tendencies of soil features. The number of classes of the highest category is subjective, that is, it depends on the designers of the system; for example, Soil Taxonomy has 12 and the World Reference Base has approximately 30.

The separation of each class into subclasses at the next-lower level is guided by the definition of this next-lower category. It must be less abstract than the definition of the category above. For example, difference of kind among soils is more abstract than difference in degree. An important feature of a hierarchy is that successively lower-level classes accumulate the limits and descriptions that apply to the classes above which it is a subdivision. This means there is an accumulation of defining characteristics with each successive subdivision. In addition it is appropriate also to make statements about those properties and features that are correlated with the defining characteristics; thus each class has a set of defining characteristics and a set of associated, or accessory, characteristics which permit many useful statements to be made about the classes in the lower categories.

Two important aspects of a hierarchy are: (1) that all members of the domain being considered are included in each categorical level; and (2) that differentiating criteria accumulate in each lower set of classes.

Defining categories is a difficult challenge. The set of category definitions may refer to morphogenesis of soil profiles, geographic associations of kinds of soils, or even potentials for use, but they must be derived from the purpose intended for the classification. The set of category definitions establishes a priority among the concepts and theories of the paradigm (Table 1).

Table 1 Definitions of categories of Soil Taxonomy

Category	*Definition*	*Abstraction*
Order	Soils whose properties result from major conditions of soil formation and evolution	Genetic pathways (soil evolution)
Suborder	Soils within an order whose additional properties or conditions are major controls of the current set of soil-forming processes	Major controls of current processes
Great group	Soils within a suborder whose additional properties constitute subordinate or additional controls, or reflect such controls, of the current set of soil-forming processes	Subordinate controls of current processes
Subgroup	Soils within a great group whose additional properties result from a blending or overlapping of sets of processes in space and time that cause one kind of soil to develop from, or toward, another kind of soil (intergrades), or whose conditions have not previously been recognized (extragrades)	Merging of processes or has specific conditions
Family	Soils within a subgroup whose additional properties characterize parent material and ambient conditions	Constraints on further change; affect use
Series	Soils within a family whose additional properties reflect relatively narrow ranges of soil-forming factors and localized processes	Smaller range of factors and processes

For example, the definition of the highest category of Soil Taxonomy, the Order, can be stated as 'soils whose properties result from, or reflect, major conditions of soils formation and evolution.' The processes cannot be measured, but soil morphology and some additional properties are assumed to correlate with the formation and presence of particular features. It is easy to produce confusion when the definitions are violated.

Due to the many areas of uncertainty in the measurement of properties and the relevance of different combinations of sets of properties, in addition to the lack of precision of concepts and relationships, it is obvious that classification is not a truth that can be discovered. There is no obvious way to determine the correct details of structure of a hierarchical system. The number of categories required to represent adequately a specific domain of millions of individuals is not known, and whether exclusive classes and those that overlap are equally relevant and proper is still unclear. Evaluating the adequacy of a system relies heavily on understanding the definitions of categories and their classes. This is where most 'pet theories and sacred cows' are led to slaughter.

Principle of Prioritization

How are priorities set? How are categories, and classes within a category, sequenced? It is usually assumed, or clearly stated, that measurable soil properties and features are the source of data, and not the concepts or theories themselves. A category definition is commonly given in abstract terms rather than by naming soil properties; that is, the concepts of soils' development, or use, guide the structure of a classification to show how the groupings are related to each other. In morphogenetic-based classifications, the concepts of genesis are the bases for the definitions of categories; however, the indicators or evidence of those concepts are specific soil properties and features that are thought be to appropriately correlated with the conceptually based definitions. In Soil Taxonomy the suborder category (division of the classes of the order category) can be stated as 'soils whose additional properties result from or reflect major controls of the current soil-forming processes.' As a concept, 'constraints on present day processes' is thought to be less abstract than 'pathways of development,' which is used to define the order category.

Soil-forming processes are important to an understanding of how and when soil features develop, but most processes cannot be adequately measured to confirm or deny such relationships. Which processes are more important? That is subjective and at the discretion of the designers of systems. For example, some conditions such as coldness and dryness are believed to restrict soil-forming processes. The sequence used in a key indicates the priority given to them. Because many soil schemes are morphogenetic, cause-and-effect relationships and concepts guide the selection of soil properties. Consequently, properties are surrogates for the concepts of genesis or soil behavior as the case may be.

The sequence of classes in the highest category establishes the precedent for the remainder of the framework and serves as the entry into the system. Most hierarchical arrangements rely on concepts of exclusion to reduce the volume of repetitive definitions. Such a technique also retains flexibility in what is being included in a given class. The list of classes at the order level of Soil Taxonomy demonstrates the possibilities:

1. Gelisols
2. Histosols
3. Spodosols
4. Andisols
5. Oxisols
6. Vertisols
7. Aridisols
8. Ultisols
9. Mollisols
10. Alfisols
11. Inceptisols
12. Entisols

Gelisols have permafrost near the surface during most years and, without further specification, it is implied that this group includes both organic and mineral soils. All soils that do not meet the permafrost definition are excluded from this group. Histosols are the remaining organic soils (those without permafrost) that may or may not be saturated with water most of the time. All other mineral soils are excluded from this group. This procedure is also evident in the structure of the classes of the first level in the World Reference Base used to correlate and compare major soil classification systems.

The sequence of classes within each category is also a priority scheme that promotes consistency and facilitates using the system as an identification key. The list of subdivisions of the class of Alfisols (a particular kind of texture-differentiated soil) at the suborder level (Table 2) indicates a preference for the major controls of the current processes.

The Alfisols are divided into Aqualfs, Cryalfs, Ustalfs, Xeralfs, and Udalfs. The presence of an aquic moisture regime is considered to be a stronger control of processes than temperature and is commonly used as a high priority in groups of soils that have well-expressed properties recognized at a higher categorical level. Subdivisions of the Ustalf suborder of Alfisols into great groups (Table 2) are a key to the preferences given to subordinate controls of current soil-forming processes: first is the presence of a duripan; then horizons dominated by plinthite; next, those with a natric horizon; followed by thin and thick kandic horizons; followed by a thick argillic or a petrocalcic horizon; then those having red colors associated with active iron oxides; and finally Ustalfs, which have a common argillic horizon.

Because of prioritization of the categories, and of prioritization of classes that are subdivisions of a class in a higher category, it is critical to evaluate alternatives and to maintain consistency in the development of a classification scheme. A key is a simplified exclusion technique to show relationships and minimize long, complicated definitions of each taxon in a system. The first class in a sequence that accepts a soil being considered is the correct placement. To place a soil in an Alfisol requires that all of the criteria of the first nine orders have been considered and rejected.

Table 2 Suborders of Alfisols and great groups of Ustalfs in Soil Taxonomy

	Description
Suborders	Major control of processes (affects biological activity)
Aqualfs	Aquic moisture regime
Cryalfs	Cryic temperature regime
Ustalfs	Ustic moisture regime
Xeralfs	Xeric moisture regime
Udalfs	Udic moisture regime
Great groups	Subordinate control of processes (affects water movement)
Durustalfs	Has a duripan (Si-cemented)
Plinthustalfs	Plinthite-dominated horizons
Natrustalfs	A natric horizon (columnar, sodium-affected)
Kandiustalfs	A thick kandic horizon (medium-textured, low CEC and ECEC)
Kanhaplustalfs	A thin kandic horizon
Paleustalfs	Petrocalcic (cemented) or thick argillic horizon
Rhodustalfs	Dark, dusky red argillic horizons, reflect active iron oxides
Haplustalfs	Has an argillic horizon

CEC, cation exchage capacity; ECEC, effective cation exchange capacity.

The principle of prioritization allows others to visualize the rationale applied throughout a system. It also provides a check on what kind of properties are appropriate at different categorical levels and promotes consistency in placing soils. Prioritization does not, however, eliminate differences of opinion about the correctness of the selection of properties nor the placement of particular soils.

Principle of Diagnostics

Must all soil properties be quantified? Why are methods of measurement needed? In most classifications, it is assumed or clearly stated that measurable soil properties and features are the data source and not the concepts or theories themselves. The definitions of soil properties and sets of properties to be used as criteria must be based on the methods of measurement. The main reason is objectivity. There are many choices, such as color by Munsell color charts or by spectrometers, and readings may vary according to conditions outside or inside a laboratory. General availability of methods and equipment often restrict widespread acceptance of the parameters selected. If the properties and methods of measurement are specified then other scientists can repeat the procedures and observe the same, or very similar, values. This removes part of the bias of the classification architects.

Any property selected for measurement is obviously subject to the understanding of relationships between the concepts used to explain the development of the property and the data obtained for that soil property. These connections give substance to the framework of the classification and provide meaning to the pattern of order that is displayed.

Often it is thought best to present some information as ratios or percentages, and others as weights or concentrations. Sets of properties may define horizon sets and be given specific names, such as a 'mollic epipedon,' or a 'kandic horizon.' Depths of occurrence of features may be diagnostic for some features. Quantification facilitates consistency, and diagnostics represent a short-hand way to aid recognition and placement of individuals.

The principle of diagnostics refers to quantified soil properties and features and not to the rationale of generalities or abstractions used to define categories. For example, the nomenclature of placement within Soil Taxonomy is built with syllables mostly from Latin and Greek. It is common in most systems to recognize diagnostic properties by mnemonic names to assist in recalling these important features (see Table 2 and Principle of Prioritization, above, for examples). The presence or absence of features may also be diagnostic insofar as the features are defined and are relevant to the differentiae of the classes.

The orders all end with a '-sol' syllable, indicating 'soil.' All suborder classes have two syllables, whereas great group classes have an additional one or two syllables prefixed to the suborder name. The linguistics are also mnemonic to aid in recalling important features of the soils: Alfisols are soils having clay-enriched subsoils (argillic horizons) and modest reserves of basic ions; Ustalfs (two syllables) are Alfisols having a wet–dry (ustic) moisture regime; Durustalfs (three syllables) are Ustalfs with a water-restrictive duripan in the subsoil.

Principle of Membership

What determines membership? What if properties overlap taxa boundaries? Regardless of how the individuals are defined, their membership into taxa is characterized by two concepts of classes. One is the central tendency that is like a mean, mode, or idealized abstract entity. The second is the boundary of a class with other classes. The limits of a class include properties with adjacent classes in the same category and with classes in other categories that share a common property. Although mutually exclusive groups of members are desirable and theoretically possible, the uncertainties of measuring properties and the relevance of precise limits indicate that actual membership acceptance may be more probabilistic than deterministic.

Keying out the placement of an individual relies on the limits of a class more than on the properties of the central concept. It is a process of elimination and, when the properties fall within the first defined class of the priority sequence, that is the proper placement. For example, to classify a soil as a Xeralf, first its placement as an Alfisol must be verified. Then within the Alfisols the criteria of Aqualfs, Cryalfs, and Ustalfs must be considered and rejected to arrive finally at the Xeralfs.

In systems that do not rely on mutually exclusive classes, such as fuzzy c-means classification, membership in a class can be expressed on a continuous scale from 0 to 1, thus partial class membership is common. Membership values are interpolated with kriging techniques, and thresholds are commonly set for class boundary detection. Currently guidelines and applications for using these techniques in morphogenetic soil classifications are not common.

Because of ambiguity in some taxa definitions, improved techniques for decision-making are desirable. The principle of membership is fraught with doubt because there are usually only 'yes' or 'no' answers, with no options for 'almost,' or 'not quite' and other 'near misses.' There appear to be opportunities for innovation.

Principle of Certainty

What happens in the future? Should a classification be viable and flexible? How can we minimize prejudicing the future? It has been found prudent to accept yet another principle to assure that a classification does not stagnate before its time. The principle of certainty suggests that change will occur, new facts and improved knowledge will be available, and continual evaluation for relevancy is a valid construct. Thus provision should be made to test each modification; check for inconsistencies and ambiguities; and decide whether the changes are merely busy work or whether they enhance the agreement with the purpose of the classification. Using the principle of certainty enables a classification to remain flexible and open to the changing world of science.

Summary

Soil classification is a means of organizing knowledge about soils. It is common to use hierarchical schemes because there are many soils and they have numerous physical, chemical, and biological properties. Protocols for designing such systems have been widely used and are summarized in eight principles. Three deal with the rationale of classification: purpose, domain, and identity of entities and their nomenclature. Another four principles guide the organization of the available information: differentiation, prioritization, diagnostics, and membership of individuals into the defined taxa. An additional principle of certainty is concerned with the future viability of a classification scheme.

List of Technical Nomenclature

Diagnostics	Quantified soil features and properties used in definitions
Differentiation	Separation according to defined criteria
Hierarchy	Structured arrangement of information into multiple categories
Key	Priority listing of subclasses belonging to the same class
Pedon	A small, arbitrary volume of soil (1 m^2 in area) that represents a type of soil; its properties are compared with defined properties for classifying a soil
Polypedon	A small segment of landscape dominated by a single type of soil; a landscape unit composed of similar pedons
Taxa	Classes of organized information, which can be formed at different categorical levels

See also: **Classification Systems:** Australian; FAO; Russian, Evolution and Examples; USA

Further Reading

Cline MG (1949) Principles of soil classification. *Soil Science* 67: 81–91.

Cline MG (1961) The changing model of soil. *Soil Science Society of America Proceedings* 25(6): 442–446.

Cline MG (1963) Logic of the new system of soil classification. *Soil Science* 96: 17–22.

Cooperative Research Group, CST (2001) *Chinese Soil Taxonomy*. Beijing, China: Science Press.

Deckers JA, Nachtergaele FO, and Sapargaren OC (eds) (1998) *World Reference Base for Soil Resources. Introduction*, 1st edn. ISSS Working Group RB. Leuven, Belgium: Acco.

Food and Agriculture Organization of the United Nations, International Soil Reference and Information Centre, and International Society of Soil Science (1998) *World Reference Base for Soil Resources*. World Soil Resources report 84. Rome: FAO.

Rozanov BG (ed.) (1990) *Soil Classification*. Reports of the International Conference, Alma Ata. USSR State Committee for Environmental Protection. Moscow, Russia: Publication Centre for International Projects.

Shishov LL, Tonkonogov VD, Levedeva II, and Gerasimova MI (2001) *Russian Soil Classification System*. Moscow, Russia: VV Dokuchaev Soil Science Institute.

Smith GD (1963) Objectives and basic assumptions of the new soil classification system. *Soil Science* 96: 6–16.

Smith GD (1986) *The Guy Smith Interviews; Rationale for Concepts in Soil Taxonomy*. Soil Management Support Services, Technical Monograph No. 11. Washington, DC: USDA–SCS.

Soil Survey Staff (1975) *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 1st edn. USDA–NRCS, Agriculture Handbook No. 436. Washington, DC: US Government Printing Office.

Soil Survey Staff (1999) *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 2nd edn. Agricultural Handbook No. 436. Washington, DC: US Government Printing Office.

Stremski M (1975) *Ideas Underlying Soil Systematics*. (Translation of 1971 Polish edn.) Warsaw, Poland: Foreign Scientific Publications Department of the National Center for Scientific, Technical and Economic Information.

Sumner ME (ed.) (2000) *Handbook of Soil Science*, sect. E, Pedology. Washington, DC: CRC Press.

Wilding LP, Smeck NE, and Hall GF (eds) (1983) *Pedogenesis and Soil Taxonomy. I. Concepts and Interactions*. Amsterdam, the Netherlands: Elsevier Science.

CLASSIFICATION SYSTEMS

Contents
Australian
FAO
Russian, Background and Principles
Russian, Evolution and Examples
USA

Australian

R W Fitzpatrick, CSIRO Land and Water
Glen Osmond, SA, Australia

Introduction

Classifying soils (for a particular purpose) involves the ordering of soils into groups with similar properties and for potential end uses. The historical evolution of soil-classification systems currently used in Australia that have national, regional, and special-purpose applications can be traced as follows:

- General-purpose soil classifications that have been used in Australia since 1931 to communicate soil information and soil distributions at national scales;
- State and regional 'user-friendly' soil classifications designed both to assist with communication of soil information and to account for the occurrence of soils that impact on existing and future industry development and prosperity;
- Special-purpose and more-technical classification systems for single-purpose applications that involve using detailed soil-assessment criteria with recommendations for soil-management practices have been developed for a range of specific Australian industries.

The wide occurrence in Australia of relict erosional and depositional landscapes has given rise to soil materials not encountered in younger glaciated landscapes of the northern hemisphere. Some of these materials are: ferricretes, silcretes, and companion strongly weathered zones; the alluvial and eolian deposited materials on the plains of the Murray–Darling basin in eastern Australia; and the deeply weathered soils of the erosional terrain of low gradient in northern Queensland. A large proportion of soils in Australia have strong texture contrasts in the profile and their B horizons are susceptible to dispersion and erosion because of sodicity. More than 60% of the 20 million hectares of cropping soils in Australia are sodic, and dryland farming is mainly practiced on these soils. More than 80% of sodic soils in Australia have dense clay subsoils with high sodicity (exchangeable sodium percentage (ESP) more than 6%) and alkalinity ($pH > 8.5$). There is also high potential for dryland salinity to develop through shallow saline groundwater tables. Thus, compared with soils in other major agricultural regions in the world, many Australian soils present problems for plant growth due to one or more of the following:

- Poor capacity to store water;
- Salinity, sodicity, and alkalinity;
- Low nutrient availability;
- Susceptibility to wind and water erosion;
- Poor physical status of surface and subsoil horizons (e.g., hard, compact, or slowly permeable).

Contemporary objectives of soil and land resource survey programs in Australia now deal with practical issues of land and soil evaluation for a broad range of land-management issues required by landholders, private enterprises, researchers, and government agencies, which is aided by an improved system adopted in 1996. Less emphasis is given to soil genesis. The 1996 *Australian Soil Classification* officially replaced the 1968 *Handbook of Australian Soils* and *The Factual Key for the Recognition of Australian Soils* (1979). Various special-purpose classification systems have also been developed for utilitarian ends. Development of numerical methods for soil classification has also been investigated by researchers.

History of General-Purpose National Soil Classifications

The first classification of Australian soils was devised by Prescott in 1931 and was subsequently revised in 1947. This classification system incorporated concepts of the Russian system of soil classification, an approach that was followed broadly by Stephens. These systems laid emphasis on soil genesis. This was followed by the *Handbook of Australian Soils*

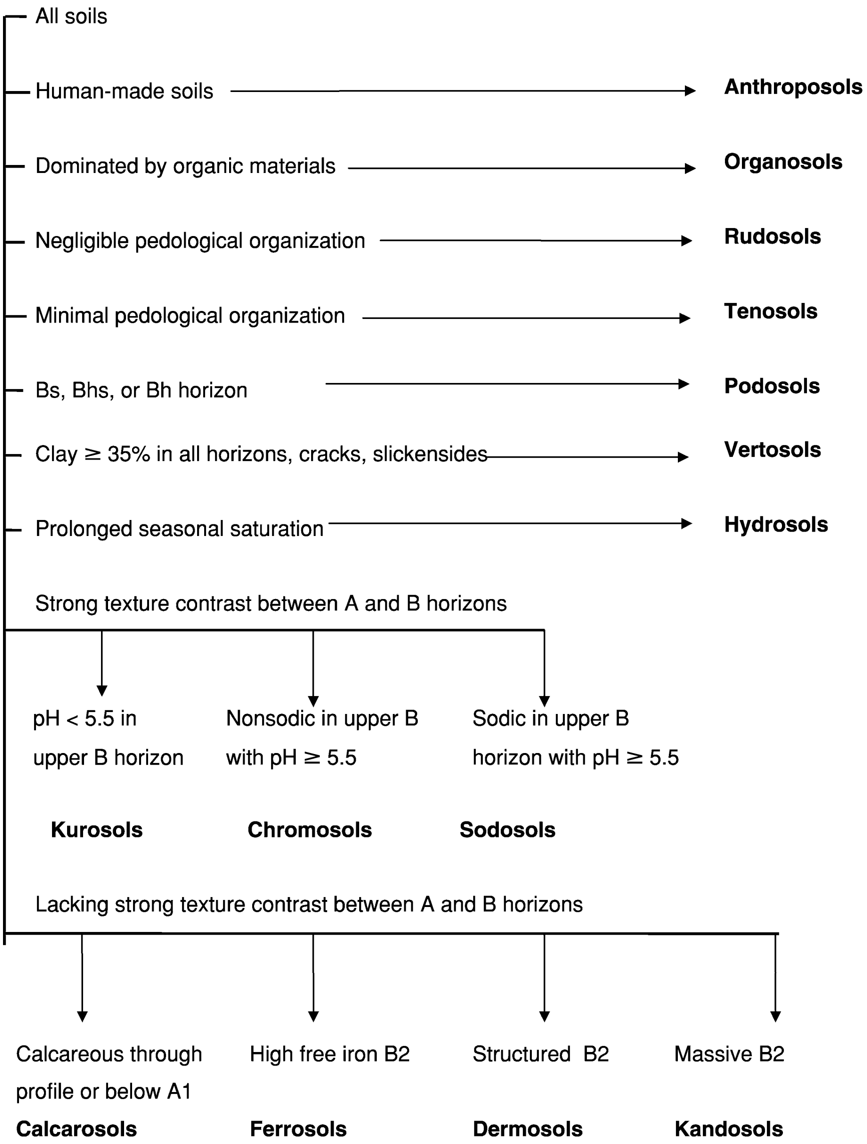

All soils
Human-made soils
Anthroposols
Dominated by organic materials
Organosols
Negligible pedological organization
Rudosols
Minimal pedological organization
Tenosols
Bs, Bhs, or Bh horizon
Podosols
Clay ≥ 35% in all horizons, cracks, slickensides
Vertosols
Prolonged seasonal saturation
Hydrosols
Strong texture contrast between A and B horizons
pH < 5.5 in upper B horizon
Nonsodic in upper B with pH ≥ 5.5
Sodic in upper B horizon with pH ≥ 5.5
Kurosols
Chromosols
Sodosols
Lacking strong texture contrast between A and B horizons
Calcareous through profile or below A1
High free iron B2
Structured B2
Massive B2
Calcarosols
Ferrosols
Dermosols
Kandosols

Figure 1 Structure of the Australian Soil Classification. Modified from Isbell RF (1996) *The Australian Soil Classification*. Copyright CSIRO 1996. Reprinted 1998. Published by CSIRO Publishing, Melbourne, Australia.

in 1968, which had many features in common with the 40 Great Groups of the American Great Group system. These genetic systems were criticized because they did not contain morphological information and did not emphasize profile criteria such as marked texture contrast features. Such features were used by Northcote in his *Factual Key for the Recognition of Australian Soils* to develop a more objective system. This classification was the basis for mapping soils of the Australian continent through the *Atlas of Australian Soils*. Forty-three Great Soil Groups were described and supported by representative profile data in *The Handbook of Australian Soils*. However, because separation between soil classes was not always defined clearly, many soils were inconsistently classified. For this reason Isbell devised a new national soil-classification system, which combined concepts from previous Australian systems and overseas classification schemes.

The New National Classification

The new national Australian soil-classification system is broad in its application and its hierarchical structure allows for unambiguous allocation of soils to particular classes. The system has 14 soil orders at the highest level that reflect important features of the soil continuum in Australia (Figure 1). The new system deals with agricultural and rangelands (arid zone) soils, tidal soils, and human-modified soils (Anthroposols, which are poorly understood). For example, the occurrence of the following two extreme soil orders is reflected by aridity and strong weathering in a continent with an absence of modern glaciation:

- The Sodosol order is defined by a high exchangeable-sodium percentage in the subsoil and by soils that have also been affected by salt during formation, sometimes under past semiarid or arid conditions;
- The Kandosol and Ferrosol orders usually have a history of strong weathering and leaching.

In addition, at the lowest level (family level), properties such as soil depth, thickness, texture, and gravel content of the A horizon and the maximum texture of the B horizon can be used to predict incisively soil- and land-management responses. While the format of this system is suited for pedologists, it is considered inappropriate for teaching and broader public adoption. Consequently, a user-friendly, CD-based interactive key has been developed as a tool to communicate its use to people who lack expertise in soil classification.

Soil Classifications Used on Australian State and Regional Scales

In Australia, there is a strong trend at state and regional levels toward mapping and describing land based on specific soil and landscape attributes. The Australian soil-classification system has proved useful in developing a 'labeling system' for these soil–landscape mapping units, but it is considered too complex for land resource assessors, who lack pedological skills. Consequently, there remains a need for a form of soil classification which distinguishes key soils at the state or regional level. This is considered essential for public communication, and for some geographic information system (GIS)-based modeling applications.

The Basis for State and Regional Soil-Classification Systems

Soil profile classes were defined at various levels of generalization (e.g., series, family, great soil group or phase), depending on the information available, purpose of survey, and scale of mapping. A class of soil profiles is a 'group,' not necessarily contiguous, which may be grouped on morphological similarities and possibly some laboratory-determined properties. For example, Butler in 1980 supported the concept to "develop and use the soil classification that arises from the landscape itself," and this approach takes into account national classification systems in some instances, but also uses soil features and terms that are recognized and used locally. For example, this has been done in various ways in each state by combining a locality descriptor with the taxonomic name, e.g., Dorrigo Red Ferrosols. These systems are easy to use and understand, and have practical implications for managers of agricultural land, while also encapsulating the higher resolution of the national system.

State and Regional Examples

A soil-classification system was developed in 2001 that defined several 'soil groups and soil classes,' which reflected the morphological characteristics having the greatest impact on land use and soil management in the agriculture regions of South Australia – an area covering about 16 million hectares (Table 1). Soil groups and soil classes were defined using the presence of calcium carbonate (i.e., calcareous soils or depth to a calcrete horizon) at the highest level of classification, because these characteristics are a major component of South Australian soils (59%; Table 1) and are particularly important for agricultural management.

In contrast, many agricultural soils in Western Australia have been distinguished by the relative abundance of ironstone gravels, which again affect land use and agricultural management. These soil groups were developed to satisfy a need for a simple, standardized and easy-to-understand way of classifying the most common soils in Western Australia – an area covering approximately 250 million hectares. Soil groups can be allocated by nontechnical people

Table 1 Soil groups for South Australia's agricultural districts, covering 15.7 million hectares

Soil groups for South Australia	*Area (%)*
Calcareous soils	23.9
Shallow soils on calcrete	20.2
Gradational soils with highly calcareous lower subsoils	4.5
Hard, red-brown texture-contrast soils with highly calcareous lower subsoils	10.7
Cracking clay soils	1.4
Deep loamy texture-contrast soils with brown or dark subsoils	2.8
Sand over clay soils	9.0
Deep sands	14.4
Highly leached sands	0.8
Ironstone soils	1.5
Shallow to moderately deep acidic soils on rock	2.4
Shallow soils on rock	4.0
Deep uniform to gradational soils	1.5
Wet soils	2.8
Volcanic ash soils	0.1

Adapted from: PIRSA Land Information (2001) *Soils of South Australia's Agricultural Lands;* and *Atlas of Key Soil and Landscape Attributes – Agricultural Districts of South Australia* (CD-ROMs). Primary Industries and Resources South Australia.

who may have difficulty using the new national Australian Soil Classification. Similar systems have been devised in other states and regions of Australia to assist with the public communication of soils information and are suitable for mapping at broader state and regional scales. Each classification system takes into account the most important features of that state or region. For example, in the western region of New South Wales, the devised soil-classification system uses soil physical properties (inferred from soil texture) and microtopography indices (gilgai) as an indicator of infiltration rate, water-holding capacity, and presence of root-restricting layers.

Special-Purpose Soil-Classification Systems

Special-purpose, technical classification systems have been devised and designed to cover a wide spectrum of practical issues, and are required for finer scales of resolution. These include: matching soils for viticulture and forestry (hardwoods and softwoods), engineering applications (defining best options for installing optical fiber cables), rehabilitation of disturbed mine sites, saline soils, coastal acid sulfate soils (with direct links to policy and jurisdiction), soil tillage (abrasive soils), top-dressing soil for turf, urban planning for infrastructures and mineral exploration (Table 2). These special-purpose classification systems all involve using soil-assessment criteria and also provide recommendations for improving soil-management practices. These classification systems mainly rely on soil attributes but invariably also include relevant landscape features such as geology, terrain, vegetation, hydrology, or soil chemical features. These together provide a more complete understanding of how soils and their properties vary and behave within landscapes, and how this variability needs to be managed satisfactorily. Two contrasting case studies are presented to illustrate how special-purpose soil-classification systems are being developed and utilized by Australian industries.

Viticultural Soils

All the Australian general-purpose or national soil-classification systems were found to be inadequate and could not be adapted for identifying soil profiles and soil properties within vineyards by managers lacking pedological skills. Accordingly, the Australian viticulture industry called for the development of a user-friendly soil key that could be adopted by viticulturists to select and match grapevine rootstocks with appropriate Australian soils. Subsequently, an Australian Viticultural Soil Key was developed in 2002. The key essentially uses nontechnical terms to categorize soils based on attributes important for vine growth and also correlates these attributes with the new national soil classification, great soil groups, and several international soil-classification systems (Soil Taxonomy, World Soil Reference Base, and South African). The soil features used in the key are easily recognized and focus on the following soil diagnostic features: depth to mottling caused by waterlogging; consistency; color; structure; calcareousness in different restrictive layers; cracks; texture trends down profiles (e.g., texture contrast at A/B horizon boundary or duplex character) (Table 2). The key layout is bifurcating, being based on the presence or absence of particular soil-profile features. The key is used for correlating rootstock performance with soil properties and as a vehicle for delivering soil-specific land development and soil-management options to viticulturists.

Engineering Applications: Laying of Telecommunication Optical Fiber Cables

A special-purpose soil-classification system was developed to minimize soil damage to Australia's network of telecommunication optic fiber cables. Buried optical fiber cables can develop transmission faults by soil movements caused from soil shrink–swell properties or by corrosion from saline soil solutions. Such faults are very costly to repair and if avoided can save millions of dollars. Close liaison between soil scientists and engineers ensured that research investigations led to the development of a practical soil-classification system, comprising a 1–10 rating of soil shrink–swell risk. The rating is derived logically by using a series of questions and answers set out in a manual entitled *Soil Assessment Manual: A Practical Guide for Recognition of Soils and Climatic Features with Potential to Cause Faults in Optical Fiber Cables*. The manual, used in planning optical fiber routes, describes practical, surrogate methods to assist engineers estimate shrink–swell indices in soils using either published soil maps in office assessments (*Atlas of Australian Soils*) or by undertaking simple visual observations and chemical measurements of soil properties in the field (Table 2). Guided by this manual, telecommunication engineers have learnt how to integrate pedological, climatic, and soil chemical information. Firstly, shrink–swell and corrosive soils are avoided along optical fiber cable routes. Secondly, soils are matched to appropriate types of cable, thereby circumventing the need to lay expensive, heavy-duty cables along the entire route. Thirdly, problems affecting cables previously installed in troublesome soil types have been rectified.

Table 2 Examples of special-purpose soil-classification systems used in Australia

Industrial issue	*Soil (and other) attributes used in classification*
Viticultural soils: Identification of restrictive soil layers that limit effective root depth. (9 categories; 36 subcategories)	Depth to waterlogging (mottling), hard (nonrippable) or soft rock (rippable), rockiness and stoniness, soil consistence, color, texture and structure, calcareousness in different restrictive layers, cracks; three types of texture change with depth: contrast (duplex character), uniform (little change) or gradational (gradual change)[a]
Engineering: soil damage to telecommunication optic fiber cables. (10 soil shrink–swell risk classes/soil-assessment key to select cable type)	Soil shrink–swell and corrosion risk classes: rock type (geology), cracks, gilgai, soil color, structure (slickensides), texture, soil depth, dispersibility (sodicity), soil salinity. Soil assessment key: shrink–swell and corrosion risk, soil maps, vegetation, climate hydrology[b]
Minesite rehabilitation: soils on waste-rock and spoil dumps. (14 new subgroup classes of Spolic, Anthromorphic Anthroposols)	Rockiness and stoniness, rock type, soil color and mottling, structure, texture, depth, dispersibility (sodicity), soil salinity, pH, acid sulfate soils, impermeable crusts, watertables
Salinity hazard: soil salinity: Linked hydrology and soil chemical hazards. (29 categories or classes of primary, secondary, and transient salinity)	Halitic (sodium chloride dominant); gypsic (gypsum or calcium sulfate dominant); sulfidic (pyrite dominant); sulfuric (sulfuric acid dominant); and sodic (high exchangeable sodium on clay surfaces); hydrology (presence or absence of groundwater); water status (natural or primary as opposed to induced or secondary status)
Soil degradation within catchments: waterlogging, salinity, and sodicity. Linked to options for land use and remediation. (8 soil classes)	Rockiness and stoniness, soil consistence (ease of excavation), color and mottling, structure, texture, depth, dispersibility (sodicity), soil salinity (EC), pH, sulfidic material, topography, watertables, vegetation type[c]
Coastal acid sulfate soils: Identify actual (AASS) and potential acidification hazard (PASS). (2 soil classes)	AASS: soil pH <4, shells, yellow, jarositic horizons, water of pH <5.5, iron stains, scalds PASS: waterlogged, unripe muds, black to blue-gray color, pH >7, positive peroxide test, shells
Forestry soils for future development: site productivity for hardwood and softwood plantations in Tasmania. (4 site productivity classes)	Soil color, texture, depth of each soil layer to a minimum depth of 80 cm or to an impeding layer if shallower, native vegetation type and species, and rock type (geology), elevation, rainfall, soil drainage, tree-rooting conditions, and nutrient availability
Tillage: abrasive wear of cultivation equipment. Abrasive soils. (3 soil classes: highly and moderately abrasive; nonabrasive)	Highly: hard-setting, high bulk density, ploughpan, many rough-surfaced, magnetic ironstone gravels, high silt and sand content. Moderately: few ironstone gravels; moderate organic matter, calcareous gravel, silt, and sand contents. Nonabrasive: friable, no gravels; high clay, fine carbonate and organic matter contents
Topdressing soil for turf. (3 classes: suitable, restricted, and unsuitable)	Soil structure, soil coherence, soil mottling, macrostructure, ped strength, soil texture, gravel and sand content, acidity, salt content, soil color, cutans, other toxic features (sulfides, metals, etc.)
Urban planning. Capabilities and limitations. (5 primary classes and several subclasses)	Soil properties: depth, permeability, shrink–swell potential, Gilgai, bearing strength, drainage properties, erodibility, salinity, and pH Terrain properties: mass movement, watertables, subsidence, and flooding
Mineral exploration: soil sampling medium. (6 soil classes of soil material)	Saline seepages, acid sulfate soils, iron- and aluminum-rich precipitates, sulfidic material, mottles in sulfuric horizons; salinity, pH, geochemical analyses

Adapted from: Fitzpatrick RW, Powell B, McKenzie NJ *et al.* (2002) Demands on soil classification in Australia. In: Eswaran H, Rice T, Ahrens R, and Stewart BA (eds) *Soil Classification: A Global Desk Reference*, pp. 77–100. Boca Raton, FL: CRC Press.

[a] Adapted from Maschmedt DJ, Fitzpatrick RW, and Cass A (2002) *Key for Identifying Categories of Vineyard Soils in Australia*. CSIRO Land and Water, Technical Report No. 30/02.

[b] Adapted from Fitzpatrick RW, Slade PM, and Hazelton P (2001) Soil-related engineering problems: identification and remedial measures. In: Gostin VA (ed.) *Gondwana to Greenhouse: Australian Environmental Geoscience*, pp. 27–36. Geological Society of Australia Special Publication 21. Australia: GSA.

[c] Adapted from Fitzpatrick RW, Cox JW, Munday B, and Bourne J (2003) Development of soil-landscape and vegetation indicators for managing waterlogged and saline catchments. *Australian Journal of Experimental Agriculture* 43: 245–252.

Future Directions

Transfer of soil information in Australia to land managers, regional and urban planners, and researchers has become less ‘pedocentric’ and is now better tailored to suit client needs (end users). This has involved improvements in the communication of general-purpose or national classification systems and the development of several special-purpose schemes for particular industries. These soil-classification systems continue to be developed and become more dominant within Australia. The new national Australian soil-classification system also continues to evolve and have a role in developing landscape models with attached spatial soil attributes. Finally, the need to develop an integrating system or approach for describing soils in terms of landscape dynamics is being actively researched.

Further Reading

Butler BE (1980) *Soil Classification for Soil Survey.* Oxford, UK: Oxford University Press.

Charman PEV and Murphy BW (eds) (2002) *Soils: Their Properties and Management*. Soil Conservation Commission of New South Wales. Australia: Oxford University Press.

Fitzpatrick RW, Slade PM, and Hazelton P (2001) Soil-related engineering problems: identification and remedial measures. In: Gostin VA (ed.) *Gondwana to Greenhouse: Australian Environmental Geoscience*, pp. 27–36. Geological Society of Australia Special Publication 21. Australia: GSA.

Fitzpatrick RW, Powell B, McKenzie NJ, Maschmedt DJ, Schoknecht N, and Jacquier DW (2002) Demands on soil classification in Australia. In: Eswaran H, Rice T, Ahrens R, and Stewart BA (eds) *Soil Classification: A Global Desk Reference*, pp. 77–100. Boca Raton, FL: CRC Press.

Fitzpatrick RW, Cox JW, Munday B, and Bourne J (2003) Development of soil-landscape and vegetation indicators for managing waterlogged and saline catchments. *Australian Journal of Experimental Agriculture* 43: 245–252.

Isbell RF (1992) A brief history of national soil classification in Australia since the 1920s. *Australian Journal of Soil Science* 30: 825–842.

Isbell RF (1996) *The Australian Soil Classification.* Melbourne, Australia: CSIRO Publishing.

Jacquier DW, McKenzie NJ, Brown KL, Isbell RF, and Paine TA (2001) *The Australian Soil Classification. An Interactive Key.* Melbourne, Australia: CSIRO Publishing.

Maschmedt DJ, Fitzpatrick RW, and Cass A (2002) *Key for Identifying Categories of Vineyard Soils in Australia.* CSIRO Land and Water, Technical Report No. 30/02. Melbourne, Australia: CSIRO Publishing.

McBratney AB (1994) Allocation of new individuals to continuous soil classes. *Australian Journal of Soil Research* 32: 623–633.

Moore AW, Isbell RF, and Northcote KH (1983) Classification of Australian soils. In: *Soil – An Australian Viewpoint*, pp. 253–266. CSIRO Division of Soils. CSIRO: Melbourne, Australia: CSIRO Publishing/London, UK: Academic Press.

Northcote KH (1979) *A Factual Key for the Recognition of Australian Soils*, 4th edn. Adelaide, Australia: Rellim.

PIRSA Land Information (2001) *Soils of South Australia’s Agricultural Lands;* and *Atlas of Key Soil and Landscape Attributes – Agricultural Districts of South Australia* (CD-ROMs). Primary Industries and Resources South Australia

Schoknecht NR (ed.) (2001) *Soil Groups of Western Australia*. Resource Management Technical Report 193. Australia: Agriculture Western Australia.

Stace HCT, Hubble GD, Brewer R, Northcote KH, Sleeman JR, Mulcahy MJ, and Hallsworth EG (1968) *A Handbook of Australian Soils*. Glenside, South Australia: Rellim.

FAO

F O Nachtergaele, FAO, Rome, Italy

Introduction

The development of the Food and Agriculture Organization (FAO) soil classification took place in three distinct stages. The first one occurred between 1960 and 1981, when FAO was the lead agency for the preparation of the FAO–UNESCO *Soil Map of the World*, for which the legend was published in 1974. A second stage involved the refinement of the original legend into a more comprehensive system, which took place between 1981 and 1990 and resulted in the *Revised Legend of the Soil Map of the World.* During the last stage, the FAO revised legend was taken as a basis for the development of a universal soil nomenclature by the International Society of Soil Science (ISSS), which had been working since 1981 on such a worldwide reference system. This work, the *World Reference Base for Soil Resources*, was published in 1998 and was subsequently accepted as the preferred international soil nomenclature in a motion by the Sixteenth Congress of the ISSS, in Montpellier, France, in 1998.

The Development of the FAO *Legend* 1960–1981

The Seventh Congress of the ISSS recommended in 1960, in Madison, Wisconsin, USA, that a world soil map be prepared. The project started in 1961, when,

Table 1 Descriptive overview of diagnostic horizons[a]

Horizon	*Description*
Histic surface	Significant amounts of organic matter
Mollic surface	Well-structured, dark, thick nutrient and organic matter-rich topsoil
Umbric surface	Well-structured, dark, thick, base-desaturated topsoil, with moderate to high amounts of organic matter
Ochric surface	Surface horizon which is light-colored, or thin, or has a low organic matter content, or is massive and hard when dry
Argillic	Subsurface horizon showing a clear accumulation of clay
Natric	Subsurface horizon with more clay than the overlying horizon and having a high content of exchangeable sodium. Usually dense with a prismatic or columnar structure
Cambic	Young subsurface horizon showing evidence of alteration such as modified colors, removal of carbonates, or soil structure
Spodic	Dark-colored subsurface horizon with accumulation of amorphous substances composed of organic matter and aluminum with or without iron
Oxic	Strongly weathered; the clay fraction is dominated by low-activity clays
Calcic	Distinct calcium carbonate enrichment
Gypsic	Distinct calcium sulfate enrichment
Sulfuric	Extremely acid subsurface horizon in which sulfuric acid has formed through oxidation of sulfides
Albic	Bleached eluviation horizon with the color of uncoated soil material, usually overlying an accumulation horizon

[a]The actual definitions are fully quantitative and for those the following should be consulted: FAO/UNESCO (1974) *FAO-UNESCO Soil Map of the World*, vol. 1, *Legend*. Paris, France: UNESCO.

in a meeting at FAO, Rome, an advisory panel laid the basis for the preparation of an international legend, the organization of field correlation, and the selection of the scale of the map and its topographic base. The first draft of definitions of soil units was issued in 1968. The legend of the soil map was finalized in 1974. The first draft of the soil map of the world was presented to the Ninth Congress of the ISSS, in Adelaide, Australia, in 1968. The map itself was completed over a span of 20 years. The first sheets, those covering South America, were issued in 1971; the final map sheets for Europe appeared in 1981.

At the global level, the 1:5 million scale *FAO–UNESCO Soil Map of the World* is still, 20 years after his finalization, the only worldwide, consistent, harmonized soil inventory that is readily available in digital format and comes with a set of estimated soil properties for each mapping unit. The FAO legend for the soil map of the world had, as its main characteristic, that it was based to the maximum extent possible on factual information derived from actual soil surveys.

The legend was based on the diagnostic horizon approach adapted under the soil taxonomy classification system developed by the US Department of Agriculture (USDA) during the 1950s and 1960s. Similar measurable and observable diagnostic horizons to those in *Soil Taxonomy* were defined. The diagnostic horizons all have a set of quantitatively defined characteristics. An overview and brief description of the 13 diagnostic horizons are given in Table 1.

A number of soil characteristics that are used to separate soil units cannot be considered as horizons. These are considered as diagnostic features of horizons or of soils which, when used for soil classification purposes, are quantitatively defined. There were approximately 20 of these defined in the FAO legend. The most important ones are summarized in Table 2.

As the legend was the outcome of a vast international collaboration of soil scientists, it was by necessity a compromise. Certain historical soil names were retained to accommodate some national sensitivities. Examples of these at the highest level are Rendzinas, Solonetzes, Solonchaks, and Chernozems. Some of the names had a dubious scientific connotation (such as the Podzoluvisols in which no podzolization in the sense of accumulation of organometal complexes takes place), while others were defined nearly identically to those developed in *Soil Taxonomy*, for example the Vertisols.

In contrast with *Soil Taxonomy*, climatic characteristics were not retained in the FAO legend, with the exception of the Xerosols and Yermosols, which largely coincided with soils developed under an aridic moisture regime and classified as Aridisols in *Soil Taxonomy*.

The FAO legend of 1974 recognized 26 Great Soil Groups, subdivided in 106 Soil units, which were the lowest category recognized on the world soil map. In addition, 12 soil phases were recognized, three general texture classes (fine, medium, and coarse), and three general slope classes (with slopes less than 8%, 8–30%, and more than 30%). Most soil mapping units were in fact soil associations, the composition of which were indicated at the back of each paper map sheet. The dominant soil unit gave its name (and appropriate color) to the mapping unit, followed by a number unique for the associated soils and inclusions.

Table 2 Descriptive overview of the main diagnostic properties in the FAO legend[a]

Property	*Description*
Abrupt textural change	Considerable increase in clay content over a very short distance
Aridic moisture regime	Moisture is not stored in the soil for more than half of the year and less then 3 months consecutively within the year
Dominance of amorphous materials	Combination of low bulk densities and a high cation exchange capacity of the clay
Hydromorphic properties	Visible evidence of prolonged waterlogging either by groundwater or by a perched water table
Plinthite	Iron-rich, humus-poor mixture of kaolinitic clay with quartz which changes irreversibly to a hardpan on exposure to air after repeated wetting and drying
Permafrost	Horizon layer in which the temperature is perennially less than 0°C
Soft powdery lime	Significant amounts of translocated lime soft enough to be readily cut with a finger nail
Sulfidic materials	Waterlogged soil materials which contain sulfur in the form of sulfides. They occur in brackish water and, when the soil is drained, they oxidize to form sulfuric acid
Tonguing	Penetration with greater depth than width of an albic horizon into an argillic one
Vertic properties	Cracks that are more than 1 cm wide and go 50 cm deep in clayey Cambisols and Luvisols

[a] The actual definitions are fully quantitative and for those the following should be consulted: FAO–UNESCO (1974) *FAO/UNESCO Soil Map of the World*, vol. 1, *Legend*. Paris, France: UNESCO.

Texture (1, 2, 3) and slope symbols (a, b, c) were included in the mapping-unit symbol.

The 26 Great Soil Groups can be further aggregated based on the major soil-forming factors that determine their characteristics.

Set 1

Set 1 holds all soils that are characterized by the accumulation of organic matter. These soils, which receive their parent material from the top (from the vegetation), rather than from the bottom (parent rock), have a histic horizon and are called Histosols (from Greek *histos*, tissue).

Set 2

Set 2 contains all mineral soils whose formation and characteristics are determined by the nature of their (mineral) parent material. These include:

1. The Andosols (from Japanese, *An*, dark, and *do*, soil), derived from volcanic ash dominated by amorphous materials;
2. The sandy Arenosols (from Latin, *arena*, sand) of desert areas, beach ridges, inland dunes and areas where soils are derived from sandstone, all characterized by coarse-textured materials;
3. The swelling and shrinking clayey Vertisols (from Latin, *verto*, to turn) of backswamps and river basins in (sub)tropical areas with an expressed dry season and dominated by expanding 2:1 lattice clays.

Set 3

Set 3 accommodates mineral soils whose formation was markedly influenced by their topography and physiography. Particularly soils in low terrain positions suffer form recurrent floods and prolonged wetness. Others on steep slopes in particular are continuously subjected to erosion processes, and remain shallow. The first two are in low-lying positions; the next four, in elevated and eroding areas:

1. Young alluvial Fluvisols (from Latin *fluvius*, river), which show stratification of recent seasonal sedimentation;
2. Nonstratified Gleysols (from Russian *gley*, mucky soil mass) occur in waterlogged areas and show hydromorphic properties linked to the process of oxidation and reduction of iron;
3. The very shallow Lithosols (from Greek *lithos*, stone) occur in the most eroding positons of the landscape and are less than 10 cm thick;
4. The slightly deeper Rendzinas (from Polish *rzedic*, noise) develop on calcareous rocks and have a mollic surface horizon;
5. The Rankers (from Austrian, *Rank*, steep slope) are also shallow, develop on acid rocks, and are characterized by an umbric horizon;
6. The deeper Regosols (from Greek *regos*, blanket) occur in unconsolidated materials, which have only superficial profile development because of prevailing low temperatures, prolonged dryness, or erosion.

Set 4

Set 4 holds soils that are only moderately developed because of their limited pedogenetic age or because of rejuvenation of the soil material. These soils are called Cambisols (from Latin, *cambiare*, change) and may occur in all climates and all landscapes.

Set 5

Set 5 accommodates the typical red and yellow soils of wet tropical and subtropical regions where the high soil temperatures and the relatively abundant rainfall result in a more profound and faster weathering of the parent material and a rapid decay of organic matter. This reults in deep and pedogenetically mature soils grouped in three major Great Soil Groups:

1. The deeply weathered Ferralsols (from Latin, *ferrum* and *aluminium*) have an oxic horizon in which 1:1 lattice clays (kaolinite) dominate;
2. The deep Nitosols (from Latin, *nitidus*, shiny) occur often in more-basic rocks and are more fertile than other soils in the tropics. They are characterized by shiny, nut-like structural elements and high clay content;
3. The strongly leached Acrisols (from Latin, *acris*, very acid) are low in bases and have an expressed clay accumulation (argillic horizon) with depth.

Set 6

Set 6 groups soils of arid and semiarid regions. Redistribution of calcium carbonate, gypsum, and more soluble salts is common under dry climatic conditions:

1. The Solonchaks (from Russian, *sol*, salt) are characterized by their high content of soluble salts;
2. The Solonetz (from Russian, *sol*, salt) contain high levels of sodium on the clay complex and have a natric horizon;
3. The Xerosols (from Greek, *xeros*, dry) are characterized by an aridic (but not cold) climate and moderate contents of organic matter in the topsoil;
4. The Yermosols (from Spanish, *yermo*, desert) are also characterized by an aridic (but not cold) climate and very low contents of organic matter in the topsoil.

Set 7

Set 7 groups soils that are mainly found in steppe regions (pampa in South America, prairie in Northern America), characterized by a vegetation of ephemeral grasses and dry forests, where accumulation of organic matter and less-soluble salts dominates over leaching processes. This set holds four different Great Soil Groups, all characterized by a mollic horizon:

1. The Chernozems (from Russian, *chern*, black, and *zemlja*, earth), characterized by their great depth and dark surface horizon, rich in organic matter, with a pronounced accumulation of calcium carbonate with depth;
2. The Kastanozems (from Latin, *castaneo*, chestnut, and Russian, *zemlja*, earth) that generally occur in the driest part of the steppe zone, with lighter-colored topsoils and stronger accumulations of calcium carbonate or gypsum;
3. The Phaeozems (from Greek, *phaios*, dusky and Russian, *zemlja*, earth) that occur in the wettest part of the steppe and that show no signs of enrichment of calcium carbonate, but remain rich in bases;
4. The Greyzems (from Anglo-Saxon, *grey*, and Russian, *zemlja*, earth) that occur in the colder parts of the steppe and that show gray coatings of silica powder in the topsoil.

Set 8

Set 8 contains the brownish and grayish soils of humid temperate areas. The soils in this set show evidence of redistribution of clay and organic matter. The cool climate and the short pedogenetic history of these soils explain why some are still rich in bases in spite of an expressed leaching process (Luvisols). In more sandy material, eluviation and illuviation of metal–humus compounds result in the grayish colors of the eluvial horizons and the blackish colors where the compounds accumulate (Podzols).

1. Acid Podzols (from Russian, *pod*, under and *zola*, ash), with a bleached eluvial horizon over an accumulation horizon of organic matter with aluminum and iron (spodic horizon);
2. Planosols (from Latin, *planus*, flat, level), with a bleached topsoil overlying abruptly a dense, slowly permeable subsoil;
3. The base-rich Luvisols (from Latin, *luo*, to wash), with a distinct clay accumulation and argillic horizon with a high base saturation;
4. The base-poor Podzoluvisols (from Podzols and Luvisols), with a bleached eluviation horizon tonguing into a clay-enriched subsurface layer.

Although initially developed as a legend for a specific map, not as a soil classification system, the FAO *Legend* found a quick acceptance as an international soil correlation system and was used, for example, as a basis for national soil classifications and regional soil inventories as in the *Soil Map of the European Communities*.

The FAO *Revised Legend* 1981–1990

With the applications of the FAO *Legend* as a soil classification, numerous comments and suggestions were received to improve the coherence of the system. In fact some combinations of diagnostic horizons could not be classified, while one soil unit identified in the key (gelic Planosols) did not occur in the *Soil Map of the World*. The revision effort undertaken in the 1980s finally resulted in the publication of the revised legend of the *FAO/UNESCO Soil Map of the World*. This revised legend was applied to the world soil resources map at 1:25 000 000 scale, accompanied by a report, and presented at the Fourteenth ISSS Congress in Kyoto in 1990 (Figure 1).

The revised legend retained many features of the original FAO legend, but definitions of diagnostic horizons and properties were much simplified. Where drastic simplifications were made, the diagnostic horizons were given another name; for example the argillic horizon was renamed 'argic,' the oxic horizon was renamed 'ferralic.'

The number of Great Soil Groups increased from 26 to 28. The Rankers and Rendzinas were grouped with the Leptosols (formerly called Lithosols), the 'aridic' Yermosols and Xerosols were not retained, while new Great Soil Groups of Calcisols and Gypsisols, characterized by respectively calcic and gypsic horizons, were created.

The Luvisols and Acrisols, both characterized by clay accumulation but with different base status, were further divided according to the activity of the clay fraction, resulting in four symmetric groups (Luvisols, high base saturation, high-activity clays; Acrisols, low base saturation, low-activity clays; Lixisols, high base saturation, low-activity clays; and Alisols, low base saturation, high-activity clays).

The revised legend created the Anthrosols at the highest level, grouping in this way soils strongly influenced by human activities. The number of soil units increased from 106 to 152. Texture and slope classes remained unchanged but were not represented on the map, for obvious reasons of scale.

A start was made with the development of a set of units at the third level. The latter were elaborated in a comprehensive set of third-level 'qualifiers' and presented at the Fifteenth ISSS Congress in Acapulco, in 1994.

As had happened with the original FAO *Legend*, the *Revised Legend* was used as a basis for the national map legends, e.g., in Botswana. The *Revised Legend* was also successfully used as a basis for university teaching of soil science.

The *World Reference Base for Soil Resources* 1981–1998

In a parallel development, FAO in cooperation with the ISSS had been involved in the development of an internationally acceptable soil classification system. In 1980, an international meeting of soil scientists, held in Sofia, launched a joint project, the International Reference Base for Soil Classification. This initiative was endorsed by the ISSS at its Twelfth Congress in New Delhi, in 1982. A Working Group RB (reference base) was set up in the framework of Commission V of the ISSS. In 1992, at a meeting of this working group, the recommendation was made that, rather than developing a fully new soil classification system, the working group should consider the FAO *Revised Legend* as a base and give it more scientific depth and coherence. This principle was accepted, and a first draft of the *World Reference Base* (WRB) appeared in 1994 that still showed large similarities with the FAO *Revised Legend*.

The WRB aims to satisfy at the same time two very different kind of users of soil information: first, the occasional interested user, who should be able to differentiate the main reference soil groups; and second, the professional soil scientist who needs a universal nomenclature for soils in a simple system that enables communication about the soils covered by a national soil classification.

The first version of the WRB was presented in 1998 at the 16th ISSS congress in Montpellier and was endorsed, in a historical motion, as the official soil correlation system of the International Union of Soil Sciences (IUSS). Major differences from the FAO *Revised Legend* include:

- The development of a two-tier system in which the first level describes and defines, in simple terms, 30 soil reference groups and classifies them with a key. The second level is composed of a list of uniquely defined qualifiers (121 of them) that can be attached to the soil reference groups. For international soil correlation purposes, a specific, preferred priority ranking is given for each soil reference group. This approach results in a very compact system that can be summarized in fewer 20 pages;
- The definition of a number of prefixes that permit the precise description where a soil phenomena occurs (epi-, endo-, bathy-) or how strongly it is expressed (hypo-, hyper-, petri-). This results in the ability to give a very precise characterization of the soil;
- The explicit emphasis put on morphological characterization of soils rather than on analytical procedures. Granted this aim is not always fully realized, but at least it is an attempt to permit soil classification to take place in the field rather than in the laboratory;

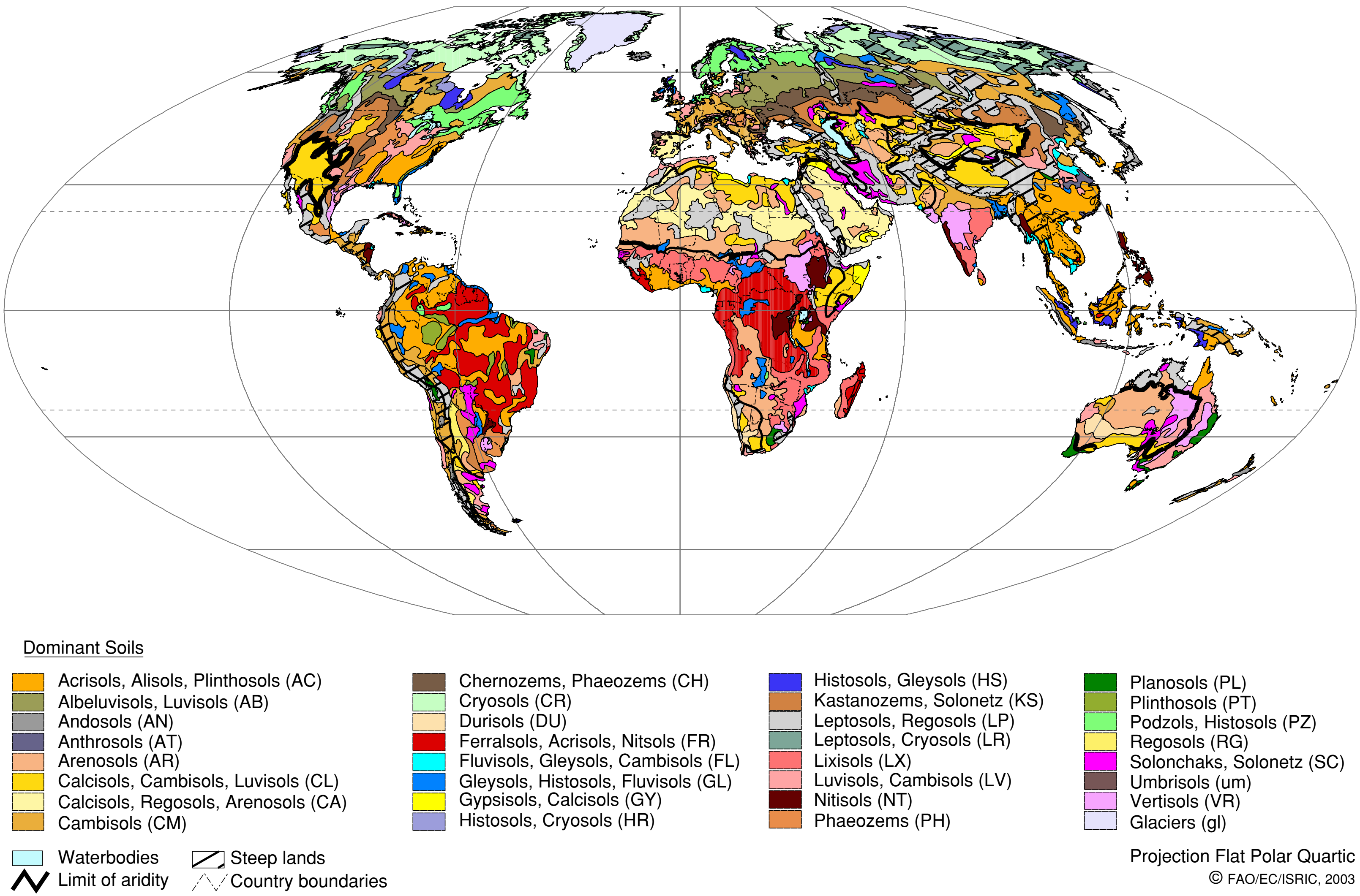


Figure 1 World soil resources map. Original scale 1:25 000 000. Flat polar quartic projection.

- The parallel publication of a WRB topsoil characterization system with approximately 70 different topsoil combinations recognized, permits classification in a much more elaborate way than is possible with six epipedons, the most important portion of the soil for most of its uses.

The WRB system as such has met with considerable success since its official endorsement by the IUSS in 1998. In the three years since its appearance, 3000 WRB books have been distributed, a website established, the system has been translated into 10 languages, and it has been adopted in several regions (European Union, West Africa, global soil and terrain database) and in individual countries (Italy, Lithuania, Georgia, South Africa), for national correlations or classification purposes, which has led to a revival of interest in soil classification and soil nomenclature.

Summary

The developments described above and the association of the FAO with the WRB work indicate that the latter has effectively replaced the FAO *Legend* with an FAO-backed and internationally endorsed world soil reference base for soil resources.

Further Reading

Brammer H, Antoine J, Kassam AH, and van Velthuizen HT (1988) *Land Resources Appraisal of Bangladesh for Agricultural Development*. Report 3. Land Resources Database, vol. II. Soil, Landform and Hydrological Database. Rome, Italy: UNDP/FAO.

Commission of the European Communities (1985) *Soil Map of the European Communities 1:1 000 000*. Luxembourg: Directorate General for Agriculture, Coordination of Agricultural Research.

Deckers J, Driessen P, Nachtergaele F, Spaargaren O, and Berding F (2003) Anticipated developments of the World Reference Base for Soil Resources. In: Eswaran H, Rice T, Ahrens R, and Stewart BA (eds) *Soil Classification: A Global Desk Reference*, pp. 245–256. CRC Press LLC.

Driessen P and Dudal R (1991) *The Major Soils of the World*. Lecture notes on their geography, formation, properties and use. Wageningen, the Netherlands: Wageningen University/Leuven.

Driessen P, Deckers J, Spaargaren O, and Nachtergaele F (2001) *Lecture Notes on the Major Soils of the World*. World Soil Resources Report 94. Rome, Italy: FAO.

Dudal R (1968) *Definitions of Soil Units for the Soil Map of the World*. World Soil Resources Report 30. Rome, Italy: FAO.

Dudal R (1980) Towards an international reference base for soil classification. *Bulletin of the International Society of Soil Science* 57: 19–20.

FAO (1993) *World Soil Resources*. An explanatory note on the FAO world soil resources map at 1:25 000 000 scale, 1991. Rev 1993. World Soil Resources Reports 66. Rome, Italy: FAO.

FAO (1995) *The Digitized Soil Map of the World Including Derived Soil Properties* (version 3.5). FAO Land and Water Digital Media Series 1. Rome, Italy: FAO.

FAO/UNESCO (1971–1981) *The FAO–UNESCO Soil Map of the World, Legend*. Paris, France: UNESCO.

FAO/UNESCO (1974) *FAO–UNESCO Soil Map of the World*, vol. 1, *Legend*. Paris, France: UNESCO.

FAO/UNESCO/ISRIC (1988) *FAO–UNESCO Soil Map of the World. Revised Legend*. World Soil Resources Reports 60. Rome, Italy: FAO.

Nachtergaele FO, Remmelzwaal A, Hof J *et al.* (1994) Guidelines for distinguishing soil subunits. In: Etchevers BJD (ed.) *Transactions of the 15th World Congress for Soil Science*, vol. 6a, pp. 818–833. Commission V: Symposia. Mexico: Instituto Nacional de Estadistica, Geografia e Informatica.

Nachtergaele FO, Spaargaren O, Deckers JA, and Ahrens R (2000) New developments in soil classification. World reference base for soil resources. *Geoderma* 96: 345–357.

Nachtergaele FO, Berding FR, and Deckers J (2001) *Pondering Hierarchical Soil Classification Systems*. Proceedings, Soil Classification 2001. Godollo, Hungary: University of Godollo.

Purnell MF, Nachtergaele FO, Spaargaren OC, and Hebel A (1994) Practical topsoil classification – FAO proposal. In: Etchevers BJD (ed.) *Transactions of the 15th World Congress for Soil Science*, vol. 6b. Commission V: Symposia. Mexico: Instituto Nacional de Estadistica, Geografia e Informatica.

Remmelzwaal A and Verbeek K (1990) *Revised General Soil Legend of Botswana*. AG/BOT/85/011. Field Document 32. Gabarone, Botswana: Ministry of Agriculture.

Soil Survey Staff (1960) *Soil Classification, a Comprehensive System,* 7th approximation. SCS, USDA. Washington, DC: US Government Printing Office.

Soil Survey staff (1999) *Soil Taxonomy. A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 2nd edn. Agricultural Handbook 436. Washington, DC: NCS, USDA.

Sombroek WG, Braun HMH, and van der Pouw BJA (1982) *Exploratory Soil Map and Agro-climatic Zone Map of Kenya*. Nairobi: Kenya Soil Survey.

Spaargaren O (ed.) (1994) *World Reference Base for Soil Resources. Draft*. Rome, Italy: ISSS/ISRIC/FAO.

Working Group RB (1998) *World Reference Base for Soil Resources. Introduction*. Leuven, Belgium: Acco.

Working Group RB (1998) *World Reference for Soil Resources. Atlas*. Leuven, Belgium: Acco.

Working Group RB (1998) *World Reference Base for Soil Resources*. (ISSS/ISRIC/FAO). World Soil Resources Reports 84. Rome, Italy: FAO.

Russian, Background and Principles

M Gerasimova, Moscow State University, Moscow, Russia

Introduction

The first classification systems in Russia were elaborated at the end of the nineteenth century – in the period when pedology was born as a new branch of natural sciences. They were developed by the great founders of this science in accordance with their perception of soil as an important part, or 'mirror,' of the landscape, and this environmental approach exerted a strong influence on further activities in the area of soil classification. Thus, the majority of Russian classification systems have a 'factor-genetic' background, which presumes that the origin of soils, soil-forming processes, and agents of soil formation serve as criteria to group and subdivide soils at almost all taxonomic levels. This means that soil diagnostics is based more on the analysis of environmental conditions and concepts of soil-forming processes at present or in the past than on ranking soil properties, unlike many western systems.

Despite similar priorities in principles, diverse systems were proposed by a number of individual soil scientists and official bodies in Russia. In terms of conceptual background, the systems may be conventionally subdivided into the groups genetic, evolutionary-genetic, factor-genetic, and substantive-genetic. The most recent system is a member of the latter group, and its substantive principles were partially inspired by the international experience in the area of soil classification.

The hierarchy of soil classification systems in Russia comprised three groups of categories. The low-level categories, or soil systematic, were concerned with texture, intensity of manifestation of processes or properties, erosion phases, etc.

The central level was always regarded as basic: it was presented by soil types, or genetic soil types in present-day versions, which were practically the same in all systems, although their definitions had some differences depending on the classification principles.

The high-level categories, or above-type classes, were specified in accordance with the approach used, thus indicating the concepts of the author(s). Most of the early systems dealt with the upper categories and embraced either the world soils, or soils of Russia/USSR. Unlike American soil classification systems, there were no elementary soil units similar to soil series.

Principles, Categories, and Classes in Major Classification Systems Developed in Russia

Early Systems

The first system was that of V.V. Dokuchaev. It was published as a simple descriptive zonal scheme in 1886 after his famous expedition to Nizhniy Novgorod. The system was in a tabulated form with short comments. All soils, 14 groups, were subdivided into 'normal' soils, 'transitional' (boggy-meadow soils, rendzinas, and solonetzes), and 'abnormal' ones (swamped, alluvial, and aeolian). The system was slightly changed by N.M. Sibirtsev in 1893, who introduced the idea of zonality for grouping the same few soils known in those times. High-category classes were named 'zonal soils, intrazonal soils, incomplete soils, and surface geologic formations,' lower categories were provided for soil texture and parent rock properties. Based on his system, Sibirtsev gave the first description of soils of Russia in 1898. However, in the field of soil classification he always emphasized the priority of Dokuchaev, who issued another version in 1896 with allowance for the remarks of his pupil.

The most popular version (*Soil Classification of Professor Dokuchaev for the Northern Hemisphere*) was published in 1900 in *Pochvovedenie* journal. In this version, normal soils were regarded as synonymous with zonal ones. Seven zones with one or two dominant soils were characterized in terms of pedogenetic processes, parent rocks, climate, vegetation, fauna, and relief. The classes of transitional (between 'normal' and 'abnormal' soils), and abnormal soils were the same as in the first version. 'Surface geologic formations' were mentioned by Dokuchaev as climate-dependent neighbors of 'abnormal' soils gradually merging with them.

Parent rocks and climate as criteria of equal importance to specify classes at the upper taxonomic level were introduced by K.D. Glinka in his first system of 1915 (exo- and endo-dynamomorphic soils, respectively); later (in 1921) it was rearranged in accordance with the idea of soil-forming processes. Five main types of soil formation (lateritic, podzolic, steppe, solonetzic, boggy) corresponded to seven soil groups with soil types as their members.

The idea of pedogenetic processes was supported by S.S. Neustruev, who proposed (in 1924) a more detailed scheme derived from the concept of elementary soil processes. Soil formation was represented there by two classes: automorphic and hydromorphic, with a further subdivision of the former in accordance with the intensity of processes. The criteria for the third level were a broad set of chemical characteristics (R_2O_3 behavior, humus properties, acidity/alkalinity) of soil constituents. The resulting 17 classes of major processes roughly corresponded to soil types; combinations of processes were provided as well. Information on landscapes was scarce and complementary.

There were other classification systems in the first decades of the last century using different approaches to combine the same few soils at the upper level: zonality (G.N. Vysotskii, 1906), vegetation (A.N. Sabanin, 1909), soil-forming agents (S.A. Zakharov, 1927).

NonOfficial Systems in the Second Half of the Twentieth Century

Many years later, I.P. Gerasimov, who was proud to be Neustruev's pupil, revived the idea of elementary pedogenetic processes – EPP (1973, 1975). Five broad groups of processes, such as 'pedomorphism of organic material, removal or accumulation of substances, pedoturbation', etc., were identified for soils of the USSR; major processes were further subdivided into elementary ones, whose number reached 25. Combinations of EPP produced soil types. These combinations, additionally differentiated in terms of process intensity in a given soil, were named 'process codes'; they produced 'profile codes' – sequences of genetic horizons, or profile formulas. For example, humus-illuvial podzols had the following combination of processes: hydration of primary and secondary minerals, formation of raw humus in an acid medium, podzolization and iron pan formation. This corresponded to the profile formula: At–**A2**–**Bih**–Cf, where major diagnostic horizons are shown in boldface.

It is worth emphasizing that this was the first system of soil classification to use profile formulas, which means that it combined genetic and substantive principles, namely, soil-forming processes and soil properties. It was simple but not exhaustive.

Soil chemical properties as criteria to group soils at high taxonomic levels were used in the system proposed by M.A. Glazovskaya (1966). Being a geochemist, she subdivded the world soils in accordance with acidity/alkalinity and redox conditions into 'geochemical associations' at the upper level, then into 36 'classes' by major pedogenetic processes. For example, the association of acid fulvic subaerial soils comprised classes of acid humus-enriched, acid metamorphic, acid metamorphic ferruginated, acid cryogenic, lateritic, and acid eluvial-illuvial soils (Podzols).

The criteria for the third category of 'families' presumed chemical properties of soil formation-derived products: humus forms, segregations, and genetic horizons. Families were subdivided into 'soil types' (fourth category), which were traditional in their essence, but were defined in terms of heat regimes within families. Thus, in the example with podzols, two families were specified: humus-illuvial podzols and podzolized iron-clay illuvial soils with different horizons and chemical compounds; the former family embraced types of dwarf podzols of the north, several 'normal' podzols of the temperate climate, tropical giant podzols, and podzols in the mountains.

Despite the emphasis placed on geochemical criteria by Glazovskaya, the system dealt more with geographic units than with soil bodies. Its elements were implemented in the world maps published in 1981 and 1998 in Russia.

V.A. Kovda presented an original evolutionary-genetic approach to soil grouping at the above-type level. The latest and most popular version of his system was used for the legend of the World Soil Map, 1975. Twelve classes of the upper category are ' soil-geochemical formations' differentiated by major trends of weathering, types of humus, clay minerals, acid–base properties, neoformations, and supplemented by characteristics of climate and vegetation. The second category, 'stadial groups of soils', corresponded to hypothetical evolution stages of pedogenesis from the submerged sites to excessively drained ones. The following stadial groups were proposed: hydroaccumulative → hydromorphic → mesohydromorphic → paleohydromorphic → proterohydromorphic → automorphic (including mountain soils) → neoautomorphic. The number of stadial groups, as well as that of soil-geochemical formations, varied in different versions of this system; moreover, climatic zones or belts were introduced in the early ones. Stadial groups of soils comprised soil types and subtypes in final version, whereas the definition of soil type had no importance.

Authors of all these systems were concerned with the upper taxonomic categories; they arranged soil types in different ways in accordance with their ideas and paid less attention to the type level. However, it should be emphasized that the notion, or 'central image' (integrity of properties, occurrence) of the soil type was the same in all systems, whereas its definition was different and depended on the conceptual background of the system. Lower categories, 'soil systematic,' were beyond the scope of these systems.

Official Systems

Unlike the systems of the previous group, official systems are 'complete' hierarchical structures with more emphasis put on soil systematics. The first official system for the purposes of soil mapping, inventory, and research in the USSR was published as a draft in 1967, then in final version in 1977; its English translation appeared in 1986. Both versions are regarded as based on soil ecology and genesis. The new system is now under preparation; its first version appeared in 1997, and the next one is planned to replace the old official system of 1977. It is assessed as substantive-genetic in its ideology because of soil horizons and soil properties serving as criteria to categorize soils.

Classification and diagnostics of soils of the USSR – 1977 Conceptually, this classification system was based on the 'neo-Dokuchaev triad': factors → proces- processes → properties, and is regarded as adhering to national traditions in genetic soil science and soil terminology. However, among the agents of soil formation, climate, or 'bioclimate', as this combination was termed in the book, had the priority in grouping soils.

The system started with the type category; any higher levels were absent. 'Genetic soil types' were differentiated with respect to bioclimatic conditions or, more precisely, to the place of soils in the system of geographic zones. At the same time, soil type was regarded as a direct result of the main pedogenetic process inherent to the zone; superposition of processes in adjacent zones was responsible for the formation of subzonal subtypes – intergrades. Thus, geographically, subtypes partially corresponded to subzones and/or to climatic facies, which were specified by soil temperature regimes, hence, there were subzonal and facial subtypes. The degree of soil hydromorphism was also taken into account at both taxonomic levels, and along with automorphic (zonal) types, semihydromorphic and hydromorphic types were recognized. This principle was strictly followed.

Genetic soil types were characterized in a descriptive manner, ABC horizons were mentioned in a general way, and some quantitative parameters with flexible and overlapping boundaries were used for both type and subtype definitions.

The third level (genus) was most severely criticized for its inconsistency. The criteria to specify genera were the following: parent rock and groundwater properties, paleopedological events, depth of effervescence, solonetzic features, faunal activity, and base saturation. Quantitative criteria were designed for the fourth level (species); they were elaborated in detail for many properties and many soils. Basically, the category was believed to characterize the degree of development of the major and superimposed pedogenetic processes. For example, chernozems were differentiated at the species level in accordance with the depth of humus horizon and the humus content in the topsoil as indicators of humus-accumulative process development. There are two more categories in the system: varieties for topsoil texture and phases for erosion phenomena. Thus, humus-illuvial podzol is referred to as: (1) the type of podzolic soils under taiga vegetation, having a percolative water regime..., acid reaction, profile differentiation...; (2) the subtype of typical (lacking humus accumulation and gley features) podzolic soils; (3) the group of genera on light-textured materials, among which there is a special genus of humus-illuvial soils. At the species level, this podzol is characterized by the depth of podzolic horizon, and content of illuviated humus.

The system was highly appreciated by users in the former Soviet Union for its logical structure, especially at lower levels, and suitability for survey because of easy soil interpretation by analyzing the soil-forming agent patterns. It has been efficiently used in large-scale soil surveys performed for arable soils by regional institutes for land management ('Giprozem') for more than 20 years. The classification system is tightly bound to manifestations of pedogenesis in geographic zones and catenas within them, and allows forecasting soil development in changing environments. One more advantage for users is its adherence to the traditional nomenclature of Russian soils.

The main drawbacks of the classification system of 1977 were the following:

- Its factor-oriented diagnostics could not provide adequate and reliable identification of soils; moreover, it contained some virtual soils
- Many soils of Russia were not included in the system, e.g., most Siberian soils, and those of tundra areas
- The system was not open to newly discovered soils, since all the 'ecological niches' in the matrix with bioclimatic entries were filled
- Human transformed soils were weakly presented.

These and some other reasons promoted the development of the next version of the national classification system.

Russian soil classification system – 1997 The new classification system is based on the principles suggested by V.M. Fridland; it was developed in the Dokuchaev Soil Institute. The system presumes the priority of soil profile morphology as a result of pedogenetic processes. That is why soil properties and soil genesis are regarded as differentiating criteria. Environmental characteristics are no longer involved in the definitions of taxa.

Genetic soil types as central units of the system are preserved, and lower categories remain, although slightly changed in comparison with the former system. However, two above-type categories are introduced, and the system now comprises eight hierarchical levels. At the uppermost level, trunks, soils are grouped in accordance with the ratio of pedogenesis versus lithogenesis (i.e., parent rock formation); hence, there are postlithogenic, synlithogenic, and organogenic soils. Criteria for the next category are soil-forming processes in a broad meaning; recognized by the character of the profile composition. Thus, orders of texturally differentiated soils, humus-accumulative, alluvial, and volcanic soils may serve as examples. The classes of the third category are represented by genetic soil types. Although the essence of soil type central images remains close to the traditional one, the criteria for identifying soil types are different in this system. Genetic soil type is defined as a soil body with a certain sequence of diagnostic horizons; consequently, each soil type has its own formula of the profile. These formulas are complemented by indices corresponding to diagnostic features serving as criteria for the fourth category, subtypes.

Therefore, diagnostic horizons and diagnostic features as manifestations of soil properties acquire great importance, and many diagnostic horizons are specified and defined in terms of morphological, and to a lesser extent, chemical, parameters and/or regimes. The full horizon definition in the English version of the system is the following: "Diagnostic horizons are genetic horizons specified by the integrity of their properties, which derive of soil-forming processes. Among these properties the major ones are regarded as diagnostic; that is, they are inherent to the 'central image' of a horizon and serve for differentiating among horizons."

Soil genesis is regarded as a tool to control the separation of horizons and features along with the choice of differentiating criteria for them. Such an attitude to the contribution of genetic approaches is very close to that in western soil classification systems, although the horizons themselves are similar but not identical.

Diagnostic features are mostly modifications of genetic horizons produced by superposition of several mechanisms responsible for horizon development, specific characteristics of present-day soil regimes, or are inherited from the earlier pedogenesis, or correspond to peculiar human-produced features in horizons that were formerly natural. Consequently, many subtypes are intergrades among types.

There are four lower categories after the subtype one. They retained their names: genus, species, variety, phase. Few changes were introduced at the 'quantitative' species level (concerning gradations, boundaries), still less at the 'textural' variety. Criteria related to parent rock properties, as it was in the former system, were completely removed from the generic level; they were partially accounted for in the classes of types and orders, and referred to at the lowest level – phase. Hence, new criteria were introduced to discriminate among genera: base saturation and salinity parameters (Table 1).

Special attention is paid to human-modified soils, which are considered to be the result of soil evolution under the impact of human activities, basically farming. Several stages of such agrogenic evolution are distinguished; they are identified by the occurrence of human-modified horizons and features, and are reflected by the position of soils in the system. Thus, strongly modified soils form separate orders (second level); they may be correlated with Anthrosols of western systems. Moderately modified soils (with one new horizon and the remainder of the natural profile) are qualified as types within natural soil orders, whereas weakly modified ones (no changes in the sets of horizons) are included as subtypes into natural type classes (Figure 1). Goals and

Table 1 Taxonomic units and criteria for classes in the new Russian Soil Classification System

Category	*Criteria*
Trunk	Pedogenesis-to-lithogenesis (peat formation) ratio
Order	Similar major elements of the profile and similarity of pedogenic processes
Type	System of diagnostic horizons
Subtype	Modifications of diagnostic horizons (recorded by diagnostic features), intergrades between types
Genus	Features of cation exchange capacity or salinity
Species	Degree of diagnostic soil properties development (quantitative parameters)
Variety	Texture, amount of stones
Phase	Parent material, depth of the solum

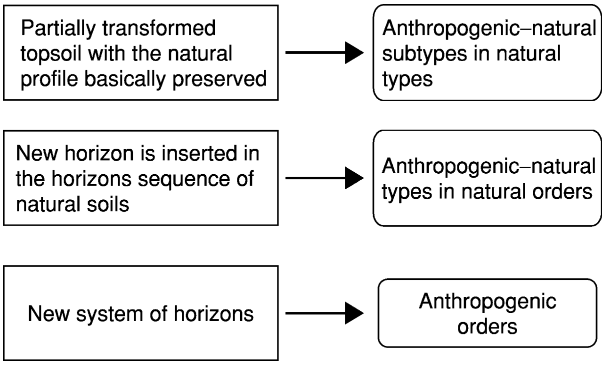

Partially transformed topsoil with the natural profile basically preserved
Anthropogenic–natural subtypes in natural types
New horizon is inserted in the horizons sequence of natural soils
Anthropogenic–natural types in natural orders
New system of horizons
Anthropogenic orders

Figure 1 Major elements of profile composition and taxonomic level of human-modified soils.

Table 2 Criteria of genetic soil-type definition in diverse systems as presenting their main principles

	Above-type levels		
Classification system, author	*Number*	*Criteria*	*Genetic soil type*
Early systems	1	Zonality, processes	No strict definitions
Gerasimov			Set of processes
Glazovskaya	3	Geochemical	Regimes within geochemical classes
Kovda	2	Evolutionary	Factors + morphology + regimes
Official – 1977			Factors + processes + profile + regimes
New – 1997	2	Profile properties	Set of diagnostic horizons

character of human impacts on soil and the level of soil fertility are not taken into account.

The new classification retains most of the soil names that have been traditionally used in Russia, with two exceptions. First, all soil names indicating environmental conditions (e.g., forest soils, meadow soils, hydromorphic soils, etc.) are excluded. Second, new names were introduced for a few 'new' soils that were absent in the former system, as well as for human-modified soils. The prefix 'agro' was proposed for the agrogenic soil types: agro-chernozem, agro-solonetz, etc.

Conclusion

Soil classification in Russia started with environmental-genetic ideas, and the systems developed applied the genetic concepts to identify high- and medium-level taxa in different ways. Soils were grouped in accordance with the combinations of soil-forming factors (climate, time), pedogenetic processes, or soil properties controlled by theories on soil genesis. Most of these approaches were implemented at higher taxonomic levels. Since the early 1980s there has been a shift to substantive-genetic principles in classifying soils. After a series of publications in journals, the first version of the *Russian Soil Classification System* was published in Russian (1997, 2000) and in English (2001). The second version is awaited in July 2004. Preserving the basic principles, it presents a broader group of soils (mainly natural); hence, the list of genetic horizons and diagnostic features is considerably enlarged.

The definition of the basic level, soil types, depends on the principles applied (**Table 2**). In the *Russian Soil Classification System* of 1997 genetic soil types are defined by combinations of diagnostic horizons, which have much in common with the diagnostic horizons of western systems. In order to summarize the information on this system, an example of a full soil name is presented.

Example: Soil type – soddy-podzolic soil (Albeluvisol in the WRB system; soil near Moscow). Upper categories: postlithogenic trunk → order of texturally differentiated soils. Lower categories: typical subtype → unsaturated genus → nondeeply podzolic species → sandy loamy variety → on mantle loam phase.

See also: **Classification of Soils**; **Classification Systems:** Russian, Evolution and Examples

Further Reading

Classification and Diagnostics of the Soils of the USSR (1977) Moscow: Kolos. English translation (1986) Washington, DC: USDA and National Science Foundation.

Classification and Diagnostics of the Soils of the USSR (1986) Washington, DC: USDA and National Science Foundation.

Dokuchaev VV (1900) Soil classification of Professor Dokuchaev for the northern hemisphere. *Pochvovedenie*, appendix – special issue.

Russian Soil Classification System (1997) Tonkonogov VD, Lebedeva II, Shishov LL, and Gerasimova MI (eds) Moscow: Dokuchaev Soil Institute (English version edited by RW Arnold (2001)).

Russian, Evolution and Examples

D Konyushkov, V.V. Dokuchaev Soil Science Institute, Moscow, Russia

Introduction

"Every soil classification is a philosophical system of pedology formulated in logical categories and symbols. It reflects the general credo and the current advancement of science. A series of classifications reflects the evolution of science, consecutive stages of its development" (Afanasiev JN (1927) The classification problem in Russian soil science. In: *Uspekhi Pochvovedeniya* (*Advances in Soil Science*). Moscow Izd, Akad, Nauk SSSR, pp. 49–108 [in Russian]). Classification has to organize our knowledge about things so that "they are thought of in such groups, and those groups in such an order, as will best conduce to the remembrance and to the ascertainment of their laws ..." (Mill JS (1891) *A System of Logic*, 8th edn. New York: Harper). For practical reasons, it is important to subdivide the universe of soils into nonintersecting groups with objective diagnostics.

The revolutionary concept of soil as an independent natural body, the function of soil-forming agents (climate, relief, biota, parent material, and time) was formulated by V.V. Dokuchaev in the 1870s. It provided excellent grounds for developing diverse systems of soil classification. Soil type – a group of soils forming in similar conditions and thus having similar genesis and properties – is the central taxon of Russian soil classifications. The subdivision of soil types into lower taxonomic categories is generally based on substantive soil properties (the humus content, texture, and acidity) and has much in common in all classifications. The difference in approaches is seen at high levels of the taxonomic hierarchy.

In factor–genetic (factor–ecological) classifications, the highest taxonomic categories of soils are distinguished on the basis of the analysis of the factors of soil formation; usually, with emphasis on bioclimatic conditions. This approach was advanced by N.M. Sibirtsev (1895) and realized in full measure in the classification by Ye.N. Ivanova and N.N. Rozov (1967) and in the official Classification and Diagnostics of Soils of the USSR (1977). (All dates are given for the original publication: the dates for translated versions may differ.)

In substantive–genetic (profile–genetic) classifications, the morphology of soil profiles and essential substantive soil properties reflecting the character of pedogenesis and the influence of geodynamic processes (erosion and deposition of sediments) on the soils are used as classification criteria at all taxonomic levels. This approach was clearly stated in the early classifications by V.V. Dokuchaev (1879, 1886) and P.S. Kossovich (1910).

The idea of a polycomponent basic substantive–genetic classification of soils was advanced by I.A. Sokolov (1978, 1991) and V.M. Fridland (1979, 1982). They considered it as a system of three complementary components: (1) the substantive profile–genetic component that describes soil genesis as reflected in the morphology of soil profiles; (2) the lithologic–mineralogical–textural component that classifies the features of parent material inherited by the soil bodies; and (3) the regime component that considers data on soil water and temperature regimes. The new Classification of Russian Soils (1997) combines the substantive profile–genetic component with the lithologic–textural component.

It can be expected that the further development of soil classification in Russia will follow this line. Along with soil classification, the ecological classification of landscapes is being developed as an integral system that takes into account not only climatic and lithological characteristics but also the geomorphic position of soils, the character of vegetation, the presence of geochemical barriers, and other indices important from the viewpoint of predicting soil behavior and elaborating the strategy of sustainable soil management.

Pre-Dokuchaev Period

Agroproductive Soil Groups, Folk Soil Nomenclature, and First Scientific Soil Classifications

The first data on Russian soils were registered in Pistsovye knigi (descriptive books) that appeared in the fifteenth century and were aimed at the evaluation of lands for taxation purposes. They contained qualitative estimates of soil fertility (rich (kind), intermediate, poor, and utterly poor soils) and scarce information on soil properties (sandy, clayey, swampy, water-logged soils). In some areas, yield-based semiquantitative estimates of soil fertility were applied to croplands and hayfields.

Extensive geographical explorations of Russia in the eighteenth century enriched scientific literature in folk soil names (Chernozems, Podzols, Solonchaks, Meadow soils, Birch soils, etc.), first concepts of soil genesis ("Chernozem derives from rotting remains of plants and animals," M.V. Lomonosov), and the notion of nature zones with specific climatic, soil, and vegetation conditions. The notion and the nomenclature of soil types – the main taxa in Russian soil classifications – are derived from local folk names of soils. However, it should be stressed that the same local soil name (e.g., Chernozem or black earth) could mean different soils in different places.

The first cartographic inventories of Russian soils were undertaken in the middle of the nineteenth century. Soil maps were compiled on the basis of questionnaires sent to local administrations. As noted by Dokuchaev, these questionnaires were greatly influenced by the first scientific classification of soils suggested by the German agronomist A.D. Thaer (1810). In the map compiled by V.I. Chaslavskii (1875), the list of soils included 32 names that can be roughly grouped into several categories: (1) textural (sandy, loamy, silty, clayey, pebbly soils); (2) local soil names (Podzols, Gray Northern soils, Chernozems, Solonchaks, Bogs, Tundra); (3) soil fertility (rich Chernozems, poor Chernozems); (4) petrographic features (Marl soils); and combinations of (1) and (2) (loamy Chernozems) and (2) and (4) (calcareous Chernozems).

The agrogeological approach to soil classification that was advanced by German scientists F.A. Fallou (1862) and F.P. Richthofen (1882, 1886) also influenced the development of soil classification in Russia. These scientists distinguished between the classes of (1) nontransferred (residual, eluvial, original) and (2) transferred (alluvial, eolian, colluvial) soils with further subdivision with respect to their petrographic composition and/or the genetic type of sediment. In 1886, P.A. Kostychev suggested an original classification system on the basis of petrographic (textural) and chemical (the content of mineral oxides and humic substances) properties of soils. Kostychev classified soils into five orders: (1) silty clayey; (2) loess; (3) sandy loamy; (4) sandy; and (5) gravelly (pebbly) soils. They were subdivided into 25 classes (classes of quartz or silicate; clayey; marl, calcareous, and dolomitic; and humus-rich soils (swampy, mucky, and chernozemic) were separated in the first four orders; quartz or silicate and calcareous soils were separated in orders 4 and 5).

Dokuchaev's Period

Genetic Approach to Soil Classification

A critical review of these initial classification systems was made by V.V. Dokuchaev (1886). He noted that they: (1) consider soil from different viewpoints (chemical, physical, geological, agricultural) instead of giving the definition of soil *per se* as an independent natural historical body; (2) do not pay attention to the fact that many soil properties are tightly linked to particular climatic, relief, geological, and vegetative conditions; and (3) give preference to some artificially selected soil properties instead of classifying soils on the basis of the total integrity of their essential properties. Dokuchaev suggested that natural scientific classification of soils as independent natural bodies should be based on their genesis studied in relation to soil-forming factors. He advanced a concept of the soil profile and an A-B-C system of horizon designation. His ideas predetermined the further development of soil classifications in Russia and abroad.

Initially (1879), Dokuchaev's scheme of soil classification looked as follows:

Division A. Normal soils lying in the place of their origin and unaffected by other dynamic processes:
- Class 1. Terrestrial vegetative soils
  - a. Northern gray soils
  - b. Chernozemic soils
  - c. Chestnut soils
  - d. Reddish solonchakous soils
- Class 2. Terrestrial swampy soils

Division B. Abnormal soils, i.e., soils that are strongly transformed by geodynamic processes:
- Class 3. Outwashed (eroded) soils
- Class 4. Inwashed (aggraded, sedimentary) soils

Dokuchaev stressed that the laws of soil formation should be established on the basis of studying normal soils, the main object of pedology. He foresaw that the number of soil types should increase in the future in parallel with the general progress of science, which, however, should not disturb the general structure of his classification. He considered color-based names of soil types as some labels reflecting the thickness of soils, their essential properties, and their relation to vegetation and climate. Finally, he believed that lower soil taxa should be distinguished with respect to the character of parent materials. It was the first substantive (based on soil properties) genetic classification of soils with special emphasis on the effect of proper pedogenic and geomorphic (geodynamic) processes on the character of soils. In 1886, Dokuchaev modified this classification and separated three groups of normal, transitional, and abnormal (sedimentary alluvial) soils by the manner of soil occurrence (soil bedding).

In the 1890s, Dokuchaev advanced the idea of natural horizontal and vertical zonality in the world of soils. It was a brilliant geographical concept. The magic of natural zonal regularities in the pedosphere was very attractive. Dokuchaev's disciple N.M. Sibirtsev (1895, 1900) introduced the zonal perception of soils into the classification scheme. At the highest level, he distinguished between the classes of: (1) zonal soils, or full-profile mature fine-earth soils forming continuous belts (zones) on the Earth's surface and developing under the impact of typical zonal soil-forming agents; (2) intrazonal soils, that appear within soil zones due to the impact of some specific local factors; and (3) azonal soils, or immature soils with not fully developed profile.

Zonal soils included the types of: (1) Lateritic (red earth); (2) Eolian – dust (loess); (3) Desert – steppe; (4) Chernozemic; (5) Gray forest; (6) Soddy – podzolic; and (7) Tundra soils. Intrazonal soils comprised the types of: (1) Solonetzic; (2) Bog (peat); and (3) Mucky-calcareous (Rendzic) soils. Azonal soils were separated into subclasses of nonriparian soils with the types of skeletal and coarse (dune) soils and flood-plain soils with the type of alluvial soils. In essence, this classification puts the emphasis on soil-forming factors (mainly climate) rather than on soils proper. It can be referred to as the first factor–genetic classification, the main branch of soil classification systems in Russia.

The last classification scheme by Dokuchaev (1900) developed the scheme by Sibirtsev and gave a concise description of typical zonal soil-forming conditions (predominant rocks, climate, vegetation, fauna, and relief) and the character of zonal pedogenetic processes. In this scheme, the concepts of normal, transitional, and abnormal soils were considered as synonyms for the concepts of zonal, intrazonal, and azonal soils, respectively.

Post-Dokuchaev Period

Diversification of Classification Decisions

Some ambiguity in the factor–genetic approach to soil classification was quite obvious for the followers of Dokuchaev. There were several attempts to make the classification 'closer to the soil,' i.e., to base classification decisions on proper soil properties rather than on soil-forming factors.

The concept of the types of pedogenesis (major directions of soil formation that manifest themselves in essential inner soil properties and are governed by some similarity in the character of soil-forming factors) as the highest level of soil classification was advocated by K.D. Glinka, the leader of soil geographical studies in Russia. In his latest (1921–1927) works, Glinka stressed that the concepts of zonal, intrazonal, and azonal soils should be considered geographical rather than classification concepts. He noted that some isolated areas of Chernozem can occur in the forest zone and, in this case, should be placed into the category of intrazonal soils. Thus, the same type of soil (Chernozem) can be found in two different groups of soils (zonal and intrazonal) at a higher level of 'zonal' classification, which is inconsistent with the general principle of classification. Glinka distinguished between the five types of pedogenesis (Lateritic, Podzolic, Steppe, Bog, and Solonetzic) and subdivided them into 25 soil types. In fact, these types of pedogenesis represented ectodynamomorphic soils (pedogenesis governed by external (climatic) conditions) that had been earlier (in 1908) separated by Glinka from endodynamomorphic soils (pedogenesis governed by the specificity of parent rocks). In the system of 1908, Glinka divided ectodynamomorphic soils into six classes on the basis of the degree of climatic moistening.

G.N. Vysotskiy (1906) tried to improve Sibirtsev's classification by introducing into it the factors of relief (orography) and parent materials. He suggested that zonal soils should be defined as the soils of particular climatic zones developing on flat surfaces composed of loamy sediments. Intrazonal soils were subdivided into subclasses of: (1) soils that become zonal in neighboring climatic zones; (2) absolutely intrazonal soils (affected by groundwater or by the surface water stagnation in depressions); and (3) skeletal (sandy and calcareous) soils. Azonal immature soils were subdivided into subclasses of denuded (eroded) and aggraded (accumulative, sedimentary) soils. Vysotskiy distinguished between the soils of flat well-drained warm sites (e.g., soils of southern slopes that become zonal in a more arid climate), and poorly drained cold sites (e.g., soils of northern slopes that become zonal in a more humid climate). This was an orographic–climatic factor–genetic classification.

A.G. Sabanin (1909) suggested the classification system, with special emphasis on the role of vegetation as the main factor that dictates the difference between parent material and soil. He separated all soils into six divisions by the character of vegetation: (1) soils of evergreen deciduous forests; (2) soils of coniferous–deciduous forests; (3) soils of dark forests; (4) soils of meadow forests; (5) soils of wormwood-grass communities (semideserts); and (6) soils of swamps. At the second level (soil classes), characteristic chemical and physical properties of soils were taken into account. This was a vegetation-based factor–genetic classification.

S.A. Zakharov (1927) grouped soil types into suborders and orders by the character of predominant soil-forming factors. He distinguished between the orders of (1) climatogenic soils (zonal soils of Sibirtsev on plains); (2) orogenic (or oroclimatogenic) soils (zonal soils in the mountains); (3) hydrogenic soils (excessive moistening, soils of depressions); (4) halogenic and hydrohalogenic soils (salt-affected soils); (5) fluvigenic soils (alluvial soils); and (6) lithogenic soils (soil properties are governed by the character of soil-forming rocks). The zonal sequence of climatogenic soils included the types of Red-Earth soils, Sierozems, Chestnut soils, Chernozems, Forest-Steppe soils, Podzolic soils, and Tundra soils. The niches for the soils from orders (3–6) were found in every soil zone belonging to orders (1) and (2). Thus, every soil zone was characterized by a series of soils from different orders. These series were called analogous soil series. Zaharov's system was substantive–genetic (morphogenetic) at lower levels and factor–genetic at higher levels of soil taxonomy.

The idea of analogous soil series in different climatic conditions was also developed by J.N. Afanasiev (1927) and D.G. Vilenskii (1925, 1945). Afanasiev's classification was based on the idea that analogous combinations of parent rocks, topography, and vegetation can occur in different climatic zones, which specifies the development of analogous series of soils in these zones. Climatic zones were differentiated in coordinates of temperature belts (cold, temperate cold, temperate warm, subtropical, and tropical) and sectors of climatic continentality (maritime, moderately continental, continental, and extremely continental). The resulting cells of the classification table were then subdivided into columns with different vegetation types (forest, forest-meadow (herbaceous), and grassy) and rows characterized by different kinds of parent material (silicate, calcareous, saline). In general, it was a climate-oriented factor–genetic classification.

Vilenskii developed several classification schemes using the idea of analogous series of soils. In 1925, he distinguished between thermogenic (soils of hot deserts), phytogenic, hydrogenic, halogenic, and compound soil divisions by the predominant factor of soil formation, within which analogous series of soils were distinguished on the basis of soil moistening and vegetation conditions. Later (1945), soil divisions were distinguished with respect to the type of weathering: lithogenic (weak weathering), H-siallitic, Ca-siallitic, Na-siallitic, Fe-siallitic, and Ferrallitic (for zonal soils); and hydrogenic, halogenic, fluvigenic, and orogenic divisions (for intrazonal and azonal soils). Analogous soil series within these divisions reflected the stages of soil development (from embryonic mineral soils to organomineral, organoaccumulative, and, finally, organoeluvial soils). Soil types were placed into this system of coordinates that can be referred to as an evolutionary factor–genetic soil classification.

S. S. Neustruev (1924) developed an original classification of soil-forming processes characteristic of the particular types of pedogenesis (as defined by Glinka) and soil types. The divisions of automorphic and hydromorphic (affected by groundwater) processes were separated at the highest level. Automorphic processes were subdivided into three classes by the intensity of transformation of the mineral mass of soils (strong, moderate, and weak). At the third level, the fate of residual products was considered. Hydromorphic processes were subdivided into subgroups with (1) the accumulation of precipitates from the groundwater and (2) the stagnation of water and the development of anaerobic reducing conditions. In essence, it was a process-oriented substantive–genetic classification of soils.

The importance of substantive characteristics in soil classification was clearly stated by P. S. Kossovich (1906, 1910). He considered pedogenesis as a combination of the processes of transformation of mineral (the source of bases) and organic (the source of acids) substances that is always accompanied by either the removal or accumulation of these substances. Special attention was paid to the input of substances into the soil with ground or surface waters. The classes of genetically independent (autonomous) and genetically dependent (geochemically conjugated, subordinate, heteronomous) soils were separated at the highest level. At the second level, soils were grouped with respect to: (1) the rate and character of mineral weathering; (2) the character of migration and redistribution of substances within the soil profile; (3) the rate of decomposition of organic substances; and (4) the character of accumulation of humic substances within the soil profile. Kossovich characterized seven types of autonomous pedogenesis (Desert, Semidesert, Chernozemic (or Steppe), Podzolic, Tundra, Peat-moss, and Lateritic) and four types of genetically dependent pedogenesis linked with corresponding autonomous soils. Kossovich showed that the genetically dependent soils accumulate the substances leached from the autonomous soils. His system can be referred to as the first process-oriented geochemical–genetic soil classification.

K.K. Gedroits (1924) tried to substantiate the separation of the main types of pedogenesis on the basis of data on the properties of the soil adsorption complex (SAC). At the highest level, he distinguished between the groups of (1) base-saturated soils (not containing H^+ ions in the SAC) and (2) unsaturated soils (with H^+ ions in the SAC). The major types of pedogenesis (Chernozemic, Solonhakous, Solonetzic, Solodic, Podzolic, and Lateritic) were characterized with respect to the composition of adsorbed cations, the state of the SAC, and the character or redistribution of pedogenetic products in the soil profile. Gedroits believed that the further development of physicochemical studies would make it possible to expand his system, that can be referred to as a physicochemical substantive–genetic classification.

B.B. Polynov (1933) tried to combine the substantive–genetic principle of soil classification with the idea of consecutive stages of soil evolution. At the highest level, he distinguished between the soils of the eluvial (automorphic, zonal) and the lacustrine-bog-solonchakous soil series. The first series was subdivided into the groups of alkaline carbonate-containing soils and acid soils. Then, these groups were subdivided into eight types of pedogenesis that were believed to be evolutionary-linked (primitive alkaline, prechernozemic, and chernozemic types were separated in the alkaline pedogenesis).

Official Soil Classification and Unofficial (Authors') Systems

In the 1930s, the needs of soil mapping required the development of more detailed soil classification systems encompassing the real geographical diversity of soils in the former Soviet Union. The factor–genetic approach was laid in the basis of classification decisions. The prototype of the new system was suggested by I.P. Gerasimov, A.A. Zavalishin, and Ye.N. Ivanova in 1939. The basic unit of soil classification – soil type – was defined as "a group of soils developing in similar conditions and characterized by common origin and common processes of the transformation and migration of substances." The system of 1939 listed 10 zonal soil types. In the 1950s, this list was expanded to over 100 soil types. In 1956, Ye.N. Ivanova and N.N. Rozov suggested the classification of world soils on the basis of the ecologic–genetic approach.

In 1958, the definition of soil taxa at the below-type level was officially adopted. The work on compiling the unified systematic list of soils of the Soviet Union was entrusted to the Dokuchaev Soil Science Institute and headed by Professor Ye.N. Ivanova. In 1966, after several revisions, this list was officially adopted by the Ministry of Agriculture of the USSR for use in large-scale soil surveys and it included 110 soil types. The principles of their grouping at higher taxonomic levels are shown in Table 1. The position of soil types in the coordinate system of bioclimatic parameters and the degree of soil hydromorphism were very rigid; some deviations (e.g., the appearance of soluble salts in some soils of the humid climate) were 'shifted down' to the soil genus level. It was a very consistent and fully developed system. Virtually all soil niches could be characterized.

Further work on the development of soil diagnostics resulted in publication of the only official Classification and Diagnostics of Soils of the USSR (1977). Soil types were characterized by particular sequences of genetic horizons. The diagnostics of genetic horizons and a system of indices for their designation were suggested. The classification of 1977 considered only the soils of the agricultural area; soils of vast northern and Siberian regions with permafrost were not included.

In parallel with the development of the official classification on the basis of the factor–genetic approach, several prominent Russian pedologists suggested their own original classification schemes of world soils. The most influential classification decisions were suggested by M.A. Glazovskaya (1966, 1972) and by V.A. Kovda, Ye.V. Lobova, and B.G. Rozanov (1967).

Glazovskaya renewed the ideas of Kossovich and Polynov and suggested the substantive geochemical–genetic system of soil classification at the above-type level (Table 2). In the 1990s, Glazovskaya developed a series of applied classifications of soils by their tolerance toward pollutants (heavy metals, acid rain). The soil classification of 1972 proved to be very useful for this purpose, as it was based on the parameters affecting the migration of substances in the soil profile (*Eh*—pH conditions, the character of organic matter and humic substances, mineralogy of the clay fraction, and the presence of geochemical barriers in the soil profile).

Kovda, Lobova, and Rozanov combined the evolutionary and geochemical approaches to soil grouping. The authors tried to arrange a system of soils that would "reflect the history of the balance of substances in the course of pedogenesis taking place under the particular energy potential that depends on the radiation balance and the humidity factor." By the energy potential, soils were grouped into 14 energy orders: humid tropical, arid–humid tropical, arid tropical, humid subtropical, and so on. Provisional evolutionary stages of soil formation were separated on the basis of the hypothesis that soil evolution on great plains proceeds from hydroaccumulative (subaqual) soils (e.g., mangrove soils) to hydromorphic (hydrobioaccumulative), mesohydromorphic, paleohydromorphic (with relic hydromorphic features), protohydromorphic (with weak features of paleohydromorphism), primitive automorphic (bioaccumulative), automorphic, paleoautomorphic, and mountainous (erosional) soils. The authors also considered soil–geochemical formations as the highest taxon of their system. Initially, there were eight formations (acid allitic, acid allitic–kaolinitic, acid kaolinitic, acid siallitic, neutral and slightly alkaline siallitic, neutral and slightly alkaline montmorillonitic, alkaline and saline, and volcanic soils). Later, in the legend to the Soil Map of the World (1:10 M scale), soil formations were considered as particular manifestations of the energy potential of weathering and pedogenesis.

The systems by Glazovskaya and Kovda (who was the main initiator of the work on the evolutionary–geochemical soil classification) were designed to classify soils at high taxonomic levels; they could not be applied to large-scale soil mapping.

The Problem of Basic Substantive–Genetic Soil Classification and the new Russian Classification

The official soil classification of 1977 did not consider soils lying beyond the main agricultural area of

Table 1 Principles of soil types grouping in the factor–genetic soil classification by Rozov and Ivanova (fragment, abridged)

		Biophysicochemical soil orders				
Ecologic–genetic (bioclimatic) soil classes	*Genetic soil orders*[a]	*Unsaturated fulvatic, leached from soluble salts and carbonates*	*Slightly unsaturated fulvate-humatic, leached from soluble salts*	*Saturated Ca-humatic, leached from soluble salts in the upper horizons*	*Ca-humate-fulvatic, with soluble salts in the middle part of the profile*	*Na-humate-fulvatic; soluble salts can occur at any depth*
Groups of soil classes: Arctic and tundra, Boreal, Subboreal, Subtropical						
Permafrost-affected taiga soils, $\Sigma T^\circ > 10^\circ C$ 600–800, humid climate, siallitic weathering	A	Taiga frozen (pergelic), soddy taiga pergelic	Palevye Taiga pergelic, sod-calcareous pergelic, meadow-forest pergelic			
	AH	Alluvial soddy pergelic				
	SH	Taiga swampy pergelic	Palevye swampy pergelic, muck-calcareous pergelic, taiga solod pergelic	Meadow-chernozemic pergelic		
	H	Bog (high-moor peat) pergelic	Meadow pergelic, meadow-bog pergelic			Meadow solonetzes and solonchaks pergelic
		Low-moor peat pergelic	Low-moor peat pergelic			

[a] Genetic soil orders (series) are distinguished with respect to the character of soil water supply and the degree of soil hydromorphism: A, automorphic soils (atmospheric water supply; deep groundwater); AH, alluvial hydromorphic (short-term inundation by flood water with simultaneous deposition of alluvium); SH, semihydromorphic soils (periodical waterlogging with surface and/or groundwater; groundwater 3–6 m); H, hydromorphic soils (permanent waterlogging; ground water <3 m).

Table 2 High-level taxonomic categories of soil classification

Rank	*Taxonomic category of soils*	*Criteria for separation*
I	Geochemical associations of soils	Soil reaction (acid, acid/alkaline, neutral/alkaline) and redox conditions (oxidative, oxidative/reducing, reducing) as related to the soil water regime in the upper/lower horizons
II	Soil generations (classes) (within geochemical associations of soils)	Manifestation of major soil processes: (1) organic matter accumulation; (2) biochemical weathering and mineral neoformation; (3) translocation of the products of pedogenesis in the soil profile (soil differentiation); (4) gleyzation; and (5) hydrogenic accumulation of substances on oxidative and evaporative barriers
III	Soil families (within soil generations)	Qualitative composition of pedogenetic products: (1) humus and organic matter (fulvate, humate, Ca-humate, muck, peat, etc.); (2) secondary minerals (carbonates, gypsum, soluble salts, allophanes, sesquioxides, etc.); (3) character of eluvial and illuvial horizons (gley-eluvial, Ca-humus argillic, etc.); (4) character of the horizons of ancient or recent hydrogenic accumulation (hydrogenic calcareous, hydrogenic saline, etc.)
IV	Soil types (within soil generations)	The degree of development of characteristic family features as conditioned by the soil temperature regime and the intensity of biological turnover (analogous soil types in different climatic zones, e.g., shallow humus-illuvial podzols of humid boreal climate and deep (tropical) humus-illuvial podzols of humid tropical and subtropical climate)

the former Soviet Union. The principles of this classification were not formulated. Soil taxa above the type level were absent. The dominance of climatic rather than proper soil characteristics was evident, especially at the level of facial (provincial) subtypes of soils. Anthropogenic modifications of soils were not properly considered. To improve this situation, the Classification Commission was organized by the All-Union Soil Science Society.

Preliminary schemes of a polycomponent basic substantive–genetic classification of soils were advanced by I.A. Sokolov (1978, 1991) and V.M. Fridland (1979, 1982). The idea of this classification is grounded in the fact that soil bodies combine stable features acquired in the course of pedogenesis (pedogenic) and inherited from the parent material (lithogenic), as well as dynamic soil characteristics reflecting the current state of soils (first of all, temperature and water regimes). Though all these characteristics are interrelated, the regularities governing them have a different nature, and there is no strict correspondence between stable lithogenic and pedogenic soil properties and dynamic soil regimes. It can be expedient to classify them in a system of three complementary classification components. The substantive profile–genetic component describes soil genesis as reflected in the morphology of soil profiles. The lithologic–mineralogical–textural component classifies the features of parent material inherited by the soil bodies. The regime component considers data on soil water and temperature regimes. The new Classification of Russian Soils (1997, 2001) compiled by L.L. Shishov, V.D. Tonkonogov, and I.I. Lebedeva (in collaboration with M.I. Gerasimova) is a realization of the substantive profile–genetic component combined with the lithologic–textural component. This classification differs significantly from the classification of 1977.

First, soils are separated on the basis of the soil profile morphology, i.e., the system of soil horizons that reflects in its properties the genesis of soils. Factors of soil formation are not taken into account. Soil names, which were indicative of bioclimatic conditions (e.g., desert soils, taiga soils, meadow soils, and so on), are replaced by the new names consistent with the logic of the profile–genetic system.

Second, anthropogenically modified soils are separated from natural soils at a high taxonomic level. Provisional principles and criteria for classifying human-created (technogenic) surface nonsoil formations are also suggested.

Third, the new system of diagnostic genetic horizons and genetic features of natural and human-transformed soils is suggested; a special section of the classification considers the correlation of major genetic horizons with diagnostic horizons of the World Reference Base (1998) system.

Fourth, the taxonomic structure of the classification is modified. At high levels, it came closer to Dokuchaev's classification of 1886. The highest taxon of the new classification is soil trunk reflecting the division of soils by the relationship between proper pedogenic and sedimentation (or organic matter accumulation) processes. The orders of Postlithogenic (soil develops from the already existing parent material, and pedogenesis is not disturbed by the deposition of new portions of sediments), Synlithogenic (pedogenesis is synchronous to sediment deposition, as in alluvial and volcanic soils), and Organogenic (organic) soils are distinguished. Soil divisions – the second taxonomic level – comprise the soils characterized by the similar trend of pedogenesis and having the same major diagnostic horizon (as a rule, subsurface horizons are taken into account). Considerable changes are made at the level of soil types that are distinguished by the similarity in the system of the main diagnostic horizons, i.e., by the similar horizonation of the soil profile. Thus, the previously single type of Chernozem is differentiated into two types (with and without textural differentiation of the profile). Some new soil types (Gleyzems, Cryozems, Dark Vertic soils) have been added. The types of mountainous soils having the same genetic horizons as the corresponding soils of plains are united with the latter. Overall, 181 types of natural and agronatural soils are distinguished.

Soil types are characterized by the similarity of the system of the main diagnostic horizons, except for the character of the parent material. Soil subtypes are distinguished as the soils having qualitative modifications of the main diagnostic horizons; as a rule, they represent intergrades between soil types. Quantitative criteria are used as soil differentiate at lower taxonomic levels (similar to the classification of 1997). Soil genera are separated by the peculiarities of their exchange complex and the chemistry of salinization. Soil species are distinguished on the basis of the degree of development of soil features taken into account at the type, subtype, or genera levels. Soil varieties take into account soil texture and stoniness. Soil phases are distinguished by the character of soil-forming and underlying rocks and thickness of the fine-earth part of the soil profile.

Further development of this system implies the creation of classification schemes for the mineralogical and textural peculiarities of soils and soil-forming rocks and for soil temperature and water regimes. However, the authors of the new classification argue that it would be difficult to develop appropriate systems for soil regimes because of the lack of adequate data for Russian soils. The development of a separate lithologic–mineralogical component reflecting soil features inherited from the parent rock is hampered by the lack of adequate criteria allowing us to distinguish between proper pedogenic alteration and the initial state of the parent material. At the same time, the need for ecological (environmental) characterization of soils is evident. Special ecological–landscape classifications should be developed, taking into account not only climatic and lithological indices but also the geomorphic position of soils, the character of natural and anthropogenic vegetation, the presence of geochemical barriers in soils, and other indices important from the viewpoint of predicting soil behavior and elaborating the strategy of sustainable soil management.

See also: **Classification of Soils**; **Classification Systems:** Australian; FAO; USA

Further Reading

Afanasiev JN (1927) *The Classification Problem in Russian Soil Science*. Russian Pedological Investigations, commission V. Leningrad: Publishing Office of the Academy of Sciences of the USSR.

Classification and Diagnostics of Soils of the USSR (1986). New Delhi: Oxonian Press.

Gedroits KK (1966) *Genetic Soil Classification Based on the Absorptive Soil Complex and Absorbed Soil Cations*. Jerusalem: Israel Program for Scientific Translation.

Glazovskaya MA (1983) *Soils of the World*, vol. 1. *Soil Families and Soil Types*. New Delhi: Amerind.

Goryachkin SV, Tonkonogov VD, Gerasimova MI *et al.* (2002) Changing concepts of soil and soil classification in Russia. In: Eswaran H, Rice T, Ahrens R, and Stewart BA (eds) *Soil Classification – A Global Desk Reference*, pp. 187–200. Boca Raton, FL: CRC Press.

Ivanova EN and Rozov NN (eds) (1970) *Classification and Determination of Soil Types*. Nos. 1–5. Jerusalem: Israel Program for Scientific Translation.

Neustruev SS (1967) *A Tentative Classification of Soil-Forming Processes as Related to Soil Genesis*. Jerusalem: Israel Program for Scientific Translation.

Shishov LL, Tonkonogov VD, Lebedeva II, and Gerasimova MI (2001) *Russian Soil Classification System*. Moscow: V. V. Dokuchaev Soil Science Institution.

Strzemski M (1975) *Ideas Underlying Soil Systematics*. Warsaw, Poland: Foreign Scientific Publications Department of the National Center for Scientific, Technical and Economic Information.

USA

D J Brown, Montana State University, Bozeman, MT, USA

Introduction

> "People who like sausage and respect the law shouldn't watch either being made" (commonly attributed to Otto von Bismarck, German chancellor, 1871–1890).

The construction of the US Department of Agriculture's Soil Taxonomy system has more in common with the legislative process and sausage-making than with the development of a carefully reasoned philosophical doctrine. To be sure, there are philosophical justifications and rationalizations for Soil Taxonomy. But the key to understanding Soil Taxonomy, as with most systems of laws or rules, lies in understanding the history and evolution of the contemporary system.

There are many thorough reviews of Soil Taxonomy, with summaries of the soil orders found in virtually every pedology text. US Soil Survey staff have published documents on the details of Soil Taxonomy and how it is supposed to work. This chapter is intended to complement and explain these sources, particularly the Keys to Soil Taxonomy. To the uninitiated, the Keys can appear overwhelmingly complex and difficult to interpret. This article provides a simple, clearly stated introduction for the uninitiated – though, to be fair, many experienced soil scientists struggle to understand Soil Taxonomy. The emphasis is on explanation, not description, with insight valued over completeness. Toward this end, I take a pragmatic stance – with both an appreciation for historical contingency and an irreverent focus on 'where the rubber meets the road,' how Soil Taxonomy is actually employed in practice.

The fundamental sampling unit for Soil Taxonomy provides an illustration of the pragmatic approach. Officially, the pedon is the smallest soil unit, with a 1–3.5 m^2 surface area and approximately 2 m depth. The polypedon, a contiguous set of pedons, is the fundamental unit for Soil Taxonomy. While there are philosophical arguments both for and against the pedon and polypedon concepts, these have little bearing on the way soils are actually classified. With few exceptions, in practice the exposed side of a soil pit (the profile) serves as the basic unit for description, sampling, and classification.

Soil Classification Criteria

The criteria in Soil Taxonomy are designed to classify soils "on the basis of their characteristics and not on the basis of the supposed or partly proved causes which have produced the characteristics," (a mandate from the early survey leader Curtis Fletcher Marbut). Yet while the classification system is superficially based on contemporary soil characteristics, historic soil-forming processes and regional atmospheric climate play a fundamental role in shaping the structure of Soil Taxonomy. The current role of climate and soil-forming processes can be traced to past classification systems. An appreciation of the history of soil classification in the USA greatly clarifies the seemingly obscure contemporary system.

The most important taxonomic requirements for soil materials, and diagnostic horizons and diagnostic characteristics are outlined below. For complete requirements, the reader is referred to the Keys to Soil Taxonomy.

Soil Composition

The amount and type of chemically active surface in a soil are controlled by: (1) secondary clay minerals; (2) organic material; and (3) noncrystalline Fe and Al materials (imogolite, ferrihydrite, Al–humus complexes, and the semicrystalline allophane included by convention). We can subdivide the secondary clay minerals into: (a) high-activity clays (2:1 layer silicates, smectites, and vermiculites); and (b) low-activity clays (kaolinite, Fe-oxyhydroxides, and Al-oxides). Likewise, organic materials can be classified according to the degree of decomposition, from least to most: (a) fibric; (b) hemic; (c) sapric. These differences in soil composition, which greatly affect soil management, are captured through various criteria in Soil Taxonomy and summarized in Figure 1.

All soil materials can be classed into two mutually exclusive categories: (1) mineral soil materials; and (2) organic soil materials. Most accumulations of organic soil materials can be found in locations where the soil is saturated for at least 30 days year^{-1}. For these materials to be classed as organic, they must have more than 12–18% organic C, depending on clay content (0–60% respectively). Organic requirements depend on clay content because the contribution of organic matter to overall soil activity becomes more important as clay content declines. Organic soil materials can be further classified according to fiber content as fibric, hemic, or sapric (important for classifying organic soils).

The term 'andic soil properties' refers to materials with significant amounts of noncrystalline Fe and Al materials, formed by the weathering of volcanic glass and usually associated with volcanic ash deposits. Key requirements for the andic designation include low bulk density, high phosphate retention, and significant amounts of noncrystalline Al and Fe (as measured by oxalate extraction). For coarser soils (30% coarse silt and sand), there is a tradeoff between volcanic glass

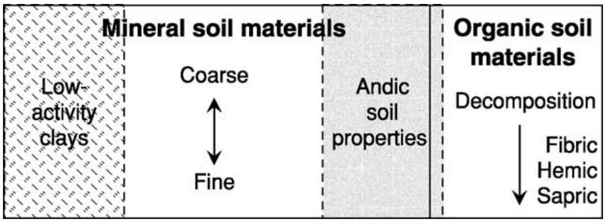


Figure 1 Schematic diagram of important soil constituents, with Soil Taxonomy designations where appropriate.

content and noncrystalline Al and Fe content required. To be termed andic, a soil must also have less than 25% organic C, with the result that both organic and mineral soils can qualify. Because the largely noncrystalline Fe and Al materials are very 'active,' they can be more important to the overall soil management than even large amounts of organic matter. While the laboratory measures required to identify andic properties are demanding, in practice they are rarely undertaken unless a volcanic ash parent material has been identified in the field.

The types and amounts of clay minerals in a soil are vital for soil management. Low-activity clays (1:1 kaolinite, Fe-oxyhydroxides, and Al-oxides) predominate in highly weathered soils, often found in warmer and wetter climates. Within Soil Taxonomy these minerals are not identified with a distinctive class. However, low-activity clays (cation exchange capacity or CEC per kg clay ≤ 16 cmol) are key criteria for two diagnostic horizons, discussed in the next section. Smectitic clays can cause soils to shrink and swell, but they are identified in Soil Taxonomy largely through field observation of shrink–swell features in a profile – not mineralogy. Subsoil clay accumulation is a defining feature for two soil orders, and there are also texture modifiers in many diagnostics, as with the clay content for the determination of organic soil materials.

Genesis and Soil Taxonomy

> "Soil surveys have created a new branch of soil science, that of soil anatomy" (C F Marbut (1921) The contribution of soil survey to soil science. *Society for the Promotion of Agricultural Science Proceedings* 41: 116–142).

Most of the diagnostics for Soil Taxonomy have a basis in soil formation theory. Understanding and using Soil Taxonomy effectively requires at least a rudimentary understanding of soil-forming processes. Soil properties important for land management play a secondary role in this system. For example, surface texture is one of the most important soil properties from a management perspective, yet influences classification primarily at the lowest series level. The emphasis on soil formation in Soil Taxonomy can be traced to the introduction of the biological metaphor in early twentieth-century soil survey and classification.

Curtis Fletcher Marbut, the inspirational leader of the US Soil Survey from 1913 until his death in 1935, modeled the science of soil survey on biology. Due largely to the tremendous success and influence of Darwin's ideas, many natural sciences drew upon a biologic metaphor from the late nineteenth through the early twentieth century, including geomorphology, ecology, entomology, and sociology. The study of profile formation was tied to the biological subfield of morphology – the study of embryonic development. And from this morphology, a hierarchical, 'genetic' soil taxonomy was developed analogous to biological taxonomy. Soils were grouped according to their mode of formation, not necessarily their contemporary properties. "No deviation from strict scientific [genetic] considerations for the sake of the so called practical use of the soil can safely be permitted." For example, texture was not considered important for classification because it was considered "not primarily a product of soil development." Like an old wine in new skins, the current Soil Taxonomy has evolved in many ways from early soil classification attempts, yet still retains the core of Marbut's system.

The genetic basis of Soil Taxonomy can be clearly observed in the requirements for diagnostic subsurface illuvial horizons (Table 1). (To present these horizons more clearly the requirements have been summarized: see Soil Taxonomy for precise requirements.) All but a few of these horizons require some evidence of illuviation in addition to compositional requirements. The argillic and natric require clay skins or an increasing fine-to-total clay ratio. The spodic requires an overlying albic or chemical evidence of illuviation. The calcic or gypsic requires observation of secondary (formed in soil) $CaCO_3$ or gypsum, or an increase in $CaCO_3$ relative to the parent material below. Among the nonilluvial diagnostic subsurface horizons (Table 2), the oxic horizon requires low CEC clays, and a paucity of weatherable minerals in the fine sand fraction (a genetic requirement). The cambic is a diagnostic horizon with indications of weak soil development. It is not necessary to demonstrate the genesis of these horizons, but the quantitative requirements for most diagnostic horizons are designed to capture genesis as best evidenced by contemporary soil properties.

There is no diagnostic subsurface horizon in Soil Taxonomy to describe soil layers that are periodically saturated and chemically reduced. The term 'aquic conditions' refers to soils which are periodically saturated and reduced. In most cases, redoximorphic features ('mottles' or a 'gleyed' horizon) are used to indicate reducing conditions. An accumulation of organic soil materials can also indicate saturated conditions. Direct measurements of soil hydrology and/or a chemically reduced state can confirm aquic conditions, but due to the cost and time involved are rarely employed. Aquic conditions are deliberately vague, with more detailed requirements specified for each soil order. Confusing matters further, the term 'aquic' is also employed for the aquic soil moisture regime.

Table 1 Illuvial–eluvial diagnostic subsurface horizons

Diagnostic horizon	*Field*	*Genesis*	*Key requirements*
Albic	E	Eluviation/leached	Light-colored horizon below A ● High value, low chroma
Argillic	Bt	Clay illuviation	Clay-enriched subsoil ● Clay 3–8% (absolute) > A/E horizons ● Evidence of illuviation: – clay skins (macro/micro), or – increased fine : total clay ratio
Natric	Btn	Clay dispersion and illuviation	High sodium argillic ● Argillic requirements ● Columnar or prismatic structure ● Sodic soil (high proportion of Na)
Spodic/ortstein	Bs Bh Bhs Bhsm	Chelation and illuviation	Illuvial Al-humus materials, often with Fe ● pH $\leq$ 5.9, organic C $\geq$0.6% ● Dark and/or red color with overlying albic ● If cemented, then ortstein
(Petro) calcic	Bk, Bkm	Illuviation	Secondary $CaCO_3$ accumulation ● $\geq$15% $CaCO_3$ (or 5% if sandy) ● Evidence of illuviation: – 5% identifiable secondary $CaCO_3$, or – 5% (absolute) > an underlying horizon ● If cemented/indurated → petrocalcic
(Petro) gypsic	By, Bym	Illuviation	Secondary gypsum accumulation ● 5% gypsum ($CaSO_4 \cdot 2H_2O$) ● Thickness (cm) × % gypsum $\geq$150 ● Evidence of illuviation: – 1% identifiable secondary gypsum ● If cemented/indurated → petrogypsic
Duripan	Bqm	Illuviation	Cemented/indurated silica accumulation ● Does not slake in weak acid ● Does slake in alkali solution

All diagnostic horizons have minimum thickness requirements, 2.5–15 cm.
All pans or cemented horizons require lateral continuity, vertical cracks $\geq$10 cm apart.

The surface diagnostic horizons, called epipedons, also have a genetic basis (Table 3). The histic epipedon, for example, not only requires a surface accumulation of organic materials, but also has a genetic requirement that this organic material accumulated under saturated conditions. The mollic epipedon is designed to capture all surface soils formed under grasslands. Trees generally deposit organic detritus on the soil surface, resulting in surface 'duff' layers and thin accumulations of organic matter in the soil relative to grasslands where annual root turnover adds organic material to the rooting depth. The mollic epipedon criteria (e.g., 0.6% organic C) are designed to capture the lowest prairie organic matter accumulation, such as in Montana, so prairie soils in Iowa usually exceed the requirements many times over.

One important point should be added to this discussion of diagnostic horizons: formal identification of these layers requires extensive, highly specified laboratory characterization. For a calcic horizon, $CaCO_3$ must be greater than 15% as determined by the evolution of CO_2 gas with acid treatment. For argillic horizons, clay films can be observed in the field or through a microscopic analysis of thin sections. For the mollic epipedon, the organic C content must be measured in the laboratory, no matter how thick and dark the surface soil. In fact, there are only a handful of diagnostic horizons that can be determined based on field observations alone. In this way, the contemporary Soil Taxonomy differs greatly from the earlier US Soil Survey classification systems where "the criteria used to classify are those that can be observed or determined rapidly by simple tests in the field."

In practice, however, laboratory characterization is rarely performed. Given the time and expense of laboratory analyses, soil surveyors and others rely on experience, field observations, and a familiarity with similar profiles to make reasonable assumptions as to soil composition.

Climate Zones in Soil Taxonomy

> "without soil climate as a criterion at some level in the taxonomic system, for example, Vertisols from Texas could be in the same class as Vertisols from North

Table 2 Other common diagnostic subsurface horizons or characteristics

Diagnostic horizon	*Field*	*Genesis*	*Key requirements*
Aquic conditions	Bg	Redox	Periodically saturated-reduced soil ● Redoximorphic features (gley, 'mottles') ● Saturated soil (directly measured) ● Chemical reduction (directly measured)
Kandic	Bt	? Clay, mineral weathering	Clay-enriched, weathered subsoil ● Clay increase relative to A/E ● Low-activity clays (kaolinite, oxides) – CEC $\leq$ 16 cmol kg^{-1} clay
Oxic	Bo	Mineral weathering	Highly weathered soil ● Low-activity clays (kaolinite, oxides) – CEC $\leq$ 16 cmol kg^{-1} clay ● Weatherable minerals $\leq$10% in fine sand ● No clay increase (not Kandic)
Salic	Bz	Multiple pathways	Saline layer ● EC $\geq$ 30 dS m^{-1} ● EC × thickness (cm) $\geq$900
Fragipan	Bx	? Unclear	High-bulk-density brittle pan ● Dry material slakes in water ● Very coarse or weak structure, or massive ● At least firm rupture resistance ● Brittle, with no roots
Placic	Bsm	Redox	Thin Fe/Mn/organic pan ● Cemented with Fe, Mn, or organic matter ● Thin: 1 mm minimum, usually less than 25 mm
Cambic		Multiple pathways	Default subsurface horizon ● Some soil formation, but not enough to meet any other diagnostic requirements

All diagnostic horizons have minimum thickness requirements, 2.5–15 cm.
All pans or cemented horizons require lateral continuity, vertical cracks $\geq$10 cm apart.
CEC, cation exchange capacity; EC, electrical conductivity.

Table 3 Most common diagnostic epipedons (surface horizons)

Epipedon	*Genesis*	*Key requirements*
Mollic	Prairie grasses	Thick, dark, organic-rich, fertile topsoil ● Dark: moist Munsell value and chrome $\leq$3 ● Organic C $\geq$0.6% ● Base saturation $\geq$50% ● Thickness $\geq$18 or 25 cm (see conditions), or meets requirements when mixed to 18 cm depth
Umbric	?	Low base saturation mollic ● Meets mollic requirements, except: ● Base saturation $\leq$50%
Histic	Wetland	Organic surface horizon ● Saturated $\geq$30 days year^{-1} ● Organic soil material $\geq$20 cm thick ● Not thick enough to be an organic soil (40–60 cm)
Folistic	Litter layer	Nonwetland organic surface horizon ● Saturated <30 days year^{-1} ● Organic soil material $\geq$15 cm thick, or $\geq$20 cm if sphagnum or bulk density <0.1
Melanic	Volcanic	Thick, dark volcanic surface horizon ● Andic soil properties $\geq$30 cm thick ● Organic C $\geq$4% throughout, $\geq$6% average ● Dark: moist Munsell value and chroma $\leq$2
Ochric	Forest/shrub	Default surface epipedon ● Does not meet the requirements of other epipedons

> Dakota" (Soil Survey Staff (1999) Soil Taxonomy. Washington, DC: US Department of Agriculture, Natural Resources Conservation Service, p. 94).

The classification of a dogwood (*cornus*) would not change at any level if we dug up a species in Texas and replanted it in North Dakota. If we dug up a Vertisol in Texas, and 'planted' it in North Dakota, however, the classification of the soil would change by virtue of location, and likely at one of the highest levels of Soil Taxonomy. This geographic–climatic dimension of Soil Taxonomy reflects the origins of soil classification in nineteenth-century Russia.

V.V. Dokuchaev and his student Sibertsev proposed that soils be studied from a 'geologic–geographic' perspective, laying the theoretical groundwork for contemporary pedology, the science behind soil survey. In the nineteenth century, soils were studied from a fertility perspective or assumed to be an extension of the geologic rock below. Dokuchaev argued for the study of soils from a natural history perspective, examining their formation and distribution over the surface of the Earth. From this vantage point, he further outlined five environmental factors that controlled the formation and distribution of soils: "climate, country age, vegetation, topography and parent rock."

Based on these ideas, Sibirtsev published a 'zonal' classification system in 1900 whereby typical profiles were described for the major 'physiographic zones' of Russia. 'Intrazonal types' deviated from the zonal norm due to the influence of one or two soil-forming factors, commonly topography (e.g., bog soils) and/or parent rock (e.g., solonetzes, salt-affected soils). 'Azonal' soils were dominated by parent materials (e.g., fresh alluvium). This classification system was brought into Europe and the USA through the writings of K.D. Glinka, who strongly emphasized the role of climate in soil formation and classification. Determined to free pedology from the discipline of geology, Glinka argued that, unlike geologic strata, soil formation and geography were controlled primarily by climate, which "provides justification for singling out the soils as a particular group of natural bodies, with which a special branch of science should be concerned." The 'zonal' or 'climatic' concept had a strong influence on the early development of pedology in the USA and this influence can still be found in contemporary Soil Taxonomy.

Climate is an important and diagnostic characteristic for both mineral and organic soils. One soil order (Aridisols) and most of the suborders are governed by soil moisture regimes, while soil temperature regimes become important at lower levels of Soil Taxonomy. Another soil order, the Gelisols, is founded on a soil feature (permafrost) that is highly correlated with atmospheric climate.

The requirements for determining soil moisture regimes (Table 4) are excruciatingly complex, with extensive soil monitoring required. In practice these measurements are rarely made because, according to soil taxonomy, "the intent in defining the soil moisture control section is to facilitate the estimation of soil moisture regimes from climatic data." In other words, soil moisture regimes are typically determined from regional atmospheric climate data, without need for soil measurements. For the vast bulk of the Earth's land surface, the soil moisture regime can be mapped in large swathes or 'zones' reminiscent of the Russian system. Local topography and hydrology can also, in some situations, influence the soil moisture regime. In particular, the aquic and peraquic moisture regimes (saturated soils) are usually controlled by local topography and groundwater hydrology and identified by the presence of redoximorphic features or an accumulation of organic soil materials.

Soil temperature regimes (Table 4) are also estimated for the most part from atmospheric climate data, though if necessary, temperatures can be taken at a depth of 50 cm at monthly intervals. Adjustments are available to convert mean annual temperature (MAT) to mean annual soil temperature (MAST). The temperature and moisture regimes for a soil to be classified can also be obtained from other soils in the region. Nonalpine soils in Montana, for example, are generally assumed to have ustic moisture and frigid temperature regimes.

Structure and Nomenclature

Most Soil Taxonomy users will never classify a soil themselves. However, understanding the processes by which soil is classified can greatly enhance the interpretation of soils already classified. Toward this end, the procedure for classifying and naming a soil is described below, followed by a worked example using observations and data from a profile in south Dakota, USA.

How it Works

Modeled on biological taxonomy, Soil Taxonomy has a hierarchical organization (Figure 2). The order, suborder, great group, and subgroup levels are primarily governed by climate and soil genesis, with the family and series levels capturing important physical and chemical properties for crop growth, engineering, and land management. In some cases, soil genesis and management are related: for example, with

Table 4 Brief descriptions of temperature and moisture regimes

Moisture regimes	*Brief description*	*Temperature regimes*	*Brief description*
Aridic/torric	Dry most of the year		
Xeric	Pronounced dry season – summer	Cryic	MAT < 8°C Summer temperature < 6–15°C No permafrost
Ustic	Pronounced dry season – winter	(Iso)frigid	MAT < 8°C Isofrigid; can also be cryic
Udic	Humid climate, no pronounced dry season	(Iso)mesic	8°C ≤ MAT < 15°C
Perudic	Wet climate, year-round precipitation > evapotranspiration	(Iso)thermic	15°C ≤ MAT < 22°C
Aquic	Saturated and reduced part of the year	(Iso)hyperthermic	22°C ≤ MAT
Peraquic	Saturated and reduced year-round		

MAT, mean annual temperature.
'Iso' prefix indicates that the mean summer and mean winter temperatures differ by less than 6°C.

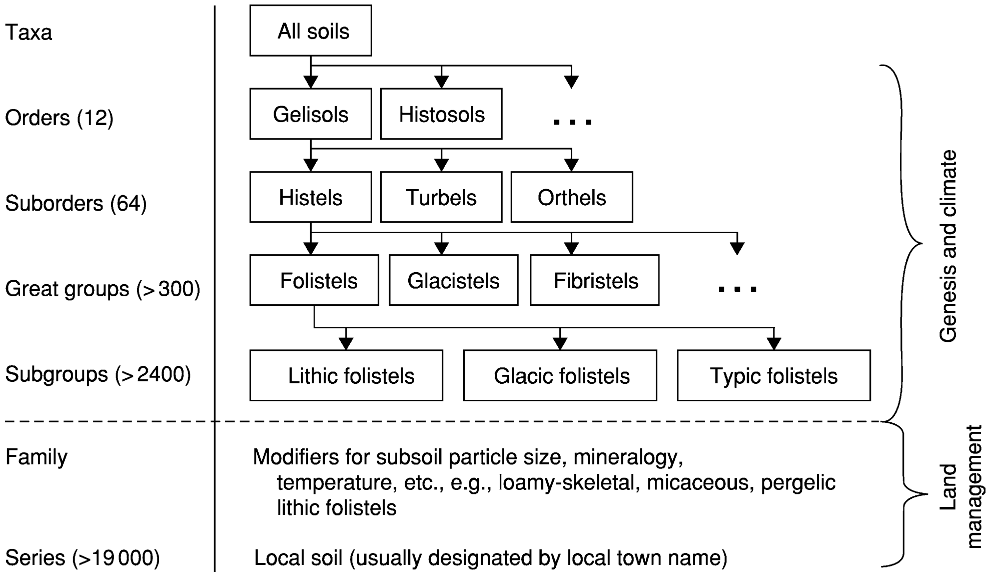


Figure 2 Hierarchical organization of Soil Taxonomy, with climatic–genetic versus management levels identified separately.

Vertisols, where shrink–swell processes are key for both profile genesis and land management. Unfortunately, this not always the case.

The soil orders are summarized in Table 5. Most of the orders can be related closely to genetic factors of soil formation, though most are related to more than one factor. Soils are keyed out in order, starting with Gelisols and ending with Entisols. Implicit in the requirements for each order is the requirement that the soil not meet the requirements for preceding orders. For example, a soil with permafrost at 75 cm, a mollic epipedon, and base saturation greater than 75% for all horizons would be classified as a Gelisol as this order takes precedence over the Mollisol. This sequential approach to 'keying out' a soil profile is used for all levels of soil taxonomy except the lowest, the soil series.

In practice, where Soil Survey activity is largely complete, soils are often classified from the bottom up. At the local level, a soil profile can be described and compared to known soil series. The series that best 'fits' the examined profile can then be applied, and the established taxonomic designation for that series thus derived.

The Soil Taxonomy nomenclature follows the hierarchical structure of the system, and allows users to glean basic soil information from the name alone. Soil names are constructed from right to left, with segments drawn primarily from the diagnostic criteria discussed previously. This nomenclature system can best be illustrated with an example:

Order: Alfisol Suborder: Udalf Great group: Kandiudalf
Subgroup: aquic kandiudalf → Aquic | kandi | ud | alf

Table 5 Summary of soil orders and key factors of soil formation

Order	*Important features and properties*	*Key factors of soil formation*
Gelisols	Permafrost within 1–2 m of surface	Climate: frozen soils
Histosols	Organic soils (usually in wetlands)	Topography: usually in wetlands
Spodosols	Illuvial Fe-Al-humus subsurface horizon, usually sandy	Vegetation and parent material: coniferous trees, sandy deposits
Andisols	Amorphous 'minerals'; volcanic glass	Parent material: volcanic ash
Oxisols	Highly weathered soils; dominated by low-CEC clays (kaolinite and oxides)	Climate and time: mineral weathering
Vertisols	Shrink–swell activity, as evidenced by cracks and 'slickenslides'	? Source of shrink–swell clays
Aridisols	Arid climate	Climate
Ultisols	Clay-enriched subsoil; low base saturation	Climate and time: base leaching, acidification
Mollisols	Thick, dark, organic-rich surface layer; high base saturation	Vegetation: prairie soils
Alfisols	Clay-enriched subsoil, high base saturation	Vegetation, climate, and time: forest soils, moderately leached
Inceptisols	Minimal horizon development	Time and climate: adolescent development
Entisols	Unaltered parent material; usually geomorphic deposits	Time: young

CEC, cation exchange capacity.

Climate, genesis, and key soil properties can be discerned from the four nomenclature segments above. The 'alf' term indicates that this profile has a clay increase in the subsoil, and does not meet the criteria for all of the preceding nine orders. The 'ud' term refers to the udic moisture regime. 'Kandi' is an abbreviation for the kandic diagnostic horizon, a clay-enriched, low-activity (low CEC per kg clay) subsoil. 'Aquic' in the subgroup tag refers to aquic conditions lower in the profile. (By comparison, an 'aqualf' would have aquic conditions higher in the profile.) Given a familiarity with soil moisture regimes, order names, and diagnostic horizons, most Soil Taxonomy names can be similarly interpreted. Default terms are 'orth,' 'hapl,' and 'typic' for the suborder, great group, and subgroup respectively. For example, a 'typic haplorthod' is a spodosol with no additional features to describe in the second to fourth levels of Soil Taxonomy.

Example Profile

The site description, field description, and essential lab characterization data for a Ferney profile in South Dakota are provided in Table 6. Using this information, the key diagnostic horizons are outlined with bold lines in Table 7. As can be seen from this example, diagnostic horizons do not necessarily correspond directly to field horizons. The mollic epipedon encompasses both the Ap and Bt horizons, while the argillic and natric diagnostic horizons are comprised of the Bt and Btkz horizons identified in the field. Furthermore, field designations do not directly correspond with the more precisely defined diagnostics. There are three B horizons with a 'k' designation for illuvial carbonate accumulation, for example, only one of which qualifies as a calcic horizon. There are also three 'z' designations in the subsoil for salt accumulation, though the electrical conductivities of these horizons do not meet the salic horizon requirement.

Both the Ap and Bt horizons meet the color value and chroma requirements as well as the organic carbon requirements for a mollic epipedon. The combined depth is 25 cm, which satisfies the thickness requirement for this epipedon, though the Ap alone would not. Similarly, the Bt and Btkz field horizons both meet the illuvial (observed clay skins) and clay increase requirements (1.2 × eluvial horizon) for an argillic horizon, with a combined thickness of 51 cm. Horizons below could also meet the illuvial requirement (despite the lack of clay skins) through the examination of thin sections under a microscope, but these data are not available. The calcic horizon has both the overall quantity of $CaCO_3$ required, and a 5% increase over an underlying horizon (evidence of illuviation). The bottom two horizons (C and Bkz2) might also meet the requirements for a gypic horizon, but confirmation of 1% secondary gypsum is lacking. Finally the Bt and Btkz horizons both meet the requirements for a natric horizon: (1) argillic requirements; (2) columnar or prismatic structure; and (3) sodium absorption ratio (SAR) ≥ 13.

Using the diagnostic horizons obtained, together with climatic data, we can readily key out the soil profile (Table 8), indicating the degree to which the designated moisture regime borders on another regime (e.g., ustic bordering on aridic). Care must be taken to distinguish between 'and' and 'or' requirements, but with practice the Keys can be systematically applied to available data, yielding a Leptic Natrustoll in this case. Space constraints do not

Table 6 Ferney profile (based on Soil Survey Pedon no. 86P0004)

Site description

Physiography: glaciated uplands
Slope: 1% concave south-east facing
Parent material: glacial till from mixed material
Temperature regime: frigid

Drainage: moderately well drained
Moisture regime: Ustic
Moisture control section:
dry 3/10 days with soil temperature >5°C

Clay (%)	*Silt (%)*	*Sand (%)*	*Organic C (%)*	*$CaCO_3$ (%)*	*Gypsum (%)*	*Base sat. (%)*	*pH (H_2O)*	*CEC*	*CEC/clay*	*SAR*
								($mmol_c$ kg^{-1})	*($mmol_c$ kg^{-1})*	
Ap – 0–13 cm; black (10YR 2/1) interior moist clay loam; moderate fine and medium subangular blocky structure; few fine accumulations of carbonate; strongly effervescent; neutral (pH = 7.0); abrupt smooth boundary										
23.7	47.7	28.6	3.24	–	–	100	6.6	26.3	111	5
Bt – 13–25 cm; very dark brown (10YR 2/2) interior moist clay loam; strong medium columnar structure; continuous faint coats on tops of columns and continuous faint clay films on faces of peds; mildly alkaline (pH = 7.6); clear wavy boundary										
40.3	33.6	26.1	1.31	tr	–	100	8.1	33.2	82	13
Btkz – 25–64 cm; dark grayish-brown (2.5Y 4/2) interior moist clay loam; moderate medium and coarse prismatic structure; continuous faint clay films on faces of peds; few fine carbonate concretions and many fine and medium salt masses; strongly alkaline (pH = 8.8); clear wavy boundary										
40.7	33.7	25.6	0.64	13	2	100	8.2	23.7	58	16
Bkz1 – 64–99 cm; dark grayish-brown (2.5Y 4/2) interior moist clay loam; few fine distinct mottles; weak medium and coarse prismatic structure; common fine carbonate concretions and common fine salt masses; strongly effervescent; strongly alkaline (pH = 8.8); clear wavy boundary										
38.2	35.0	26.8	0.27	17	1	100	8.4	21.2	56	21
Bkz2 – 99–127 cm; olive brown (2.5Y 4/4) interior moist clay loam; common fine prominent and many medium and coarse distinct mottles; weak coarse prismatic structure; common fine carbonate concretions, many coarse salt masses; strongly effervescent; moderately alkaline (pH = 8.2); gradual wavy boundary										
33.3	34.8	31.9	0.22	3	10	100	8.2	19.2	58	19
C – 127–152 cm; olive brown (2.5Y 4/4) interior moist clay loam; common fine prominent and common fine and medium distinct mottles; massive; common fine carbonate concretions, few salt masses; strongly effervescent; strongly alkaline (pH = 8.8)										
33.9	35.9	30.2	0.16	8	7	100	8.2	19.2	55	17
Note: Maximum electrical conductivity for profile is 14.2 dS m^{-1}.										

CEC, cation exchange capacity; SAR, sodium absorption ratio.

Table 7 Diagnostic horizons for Ferney profile

Horizon	*Mollic*	*Argillic*	*Calcic*	*Natric*

Ap
13 cm
	Value=2 ≤ 3
	Chroma=1 ≤ 3
	OC=3.2 ≥ 0.6%
		Eluvial horizon
		23.7% × 1.2=
		28.4% cutoff
			▨	▨

Bt
25 cm
	Value=2 ≤ 3
	Chroma=2 ≤ 3
	OC=1.3 ≥ 0.6%
		Clay films
		40.3 > 28.4% clay
			$CaCO_3$=trace< 15	Argillic +
				SAR=13 ≥ 13
				columnar structure

Btkz 64 cm	Value=4 > 3 XXXXX	Clay films 40.7 > 28.4% clay	$CaCO_3$=13 < 15	Argillic + SAR=16 ≥ 13 prismatic structure
Bkz1 99 cm		No clay films	$CaCO_3$=17 < 15 17% > 3%+5% $CaCO_3$ (3% underlying)	Not argillic
Bkz2 127 cm		No clay films	$CaCO_3$ = 13 < 15%	Not argillic
C 152 cm		No clay films	$CaCO_3$ = 12 < 15%	Not argillic

OC, Organic carbon; SAR, sodium absorption ratio.

Table 8 Keying out Ferney profile

Order	*Yes/no*	*Reason*
A. Gelisol	No	No permafrost or gelic materials
B. Histostol	No	No organic soil materials
C. Spodosol	No	No spodic horizon or ortstein
D. Andisol	No	No volcanic ash or andic materials
E. Oxisol	No	No oxic or kandic horizon
F. Vertisol	No	No surface cracks or slickenslides
G. Aridisol	No	Ustic, not aridic moisture regime, no salic
H. Ultisol	No	Base sat. >35% for all horizons
I. Mollisol	Yes	1a. Mollic epipedon 2. Base sat. >50% for all horizons
Suborder		
IA. Alboll	No	No albic horizon
IB. Aquoll	No	No aquic conditions at 40–50 cm depth
IC. Rendoll	No	2. Argillic and calic horizons present
ID. Xeroll	No	Have Ustic, not Xeric moisture regime
IE. Ustoll	Yes	Ustic moisture regime
Great group		
IFA. Durustoll	No	No duripan
IFB. Natrustoll	Yes	Natric horizon present
Subgroup (Natrustolls)		
IFBA. Leptic Torretic	No	3a. Dry only 3 days out of 10 in moisture control section with temperature >5°C (not torric)
IFBB. Torrertic	No	2a. Dry only 3 days out of 10 in moisture control section with temperature >5°C (not torric)
IFBC. Leptic Vertic	No	2a. No slickenslides or cracks 2b. Linear extensibility not 6.0 cm or greater
IFBD. Glossic Vertic	No	No vertic properties (see above)
IFBE. Vertic	No	No vertic properties (see above)
IFBF. Aridic Leptic	No	2a. Dry only 3 days out of 10 in moisture control section with temperature >5°C (not aridic)
IFBG. Leptic	Yes	Visible salt crystals at 25 cm depth <40 cm
Classification of profile to subgroup level: Leptic Natrustoll		

allow classification to the family level, though assuming mineralogical data are available, this can also be accomplished by systematically following the Keys.

At the subgroup level, there are often statistically precise climatic requirements. These requirements indicate the degree to which the moisture regime borders on another regime (e.g., Ustic bordering on Aridic). To simplify the classification for this example, the frequency of dry days during the growing season was provided in the site description, though this is rarely available in practice. As with climate regimes generally, these subgroup climatic modifiers are established regionally and can be obtained from the local soil survey staff.

Closing Thoughts

In this chapter, I explained: (1) how US soil classification took its present form, and (2) how the contemporary system is actually used. Tracing the roots of Soil Taxonomy explains many peculiar features of the contemporary system. Past climatic and genetic classification ideas can be found beneath the utilitarian veneer of the current system. Understanding these historical ideas is vital to understanding contemporary US Soil Taxonomy, and how to use it.

List of Technical Nomenclature

Base sat. Percentage base saturation, exchangeable bases × $(\mathrm{CEC})^{-1} \times 100$

$CaCO_3$ Percentage of $CaCO_3$ equivalent (measured by CO_2 evoluation) measured relative to fine earth fraction (<2 mm)

CEC Cation exchange capacity ($\mathrm{mmol_c\ kg^{-1}}$)

Clay (%) Percentage of fine earth (<2 mm) with particle size less than 2 μm

EC Electrical conductivity ($\mathrm{dS\ m^{-1}}$)

Gypsum Percentage of $CaSO_4 \cdot 2H_2O$ equivalent measured relative to fine earth fraction (<2 mm)

Org. C Percentage of organic C measured relative to fine earth fraction (<2 mm)

pH Measure of solution acidity, $\log[H^+]$

SAR Sodium absorption ratio $= Na^+ \cdot \{0.5 \cdot (Ca^{2+} + Mg^{2+})^{0.5}\}$, based on saturated paste extract

Sand (%) Percentage of fine earth (<2 mm) with particle size between 50 μm and 2 mm

Silt (%) Percentage of fine earth (<2 mm) with particle size less than 50 μm

See also: **Classification of Soils**; **Classification Systems:** Australian; Russian, Evolution and Examples

Further Reading

Ahrens RJ and Arnold RW (1999) Soil taxonomy. In: Summer ME (ed.) *Handbook of Soil Science*, pp. E117–136. Boca Raton, FL: CRC Press.

Allen G (1978) *Life Science in the Twentieth Century.* Cambridge: Cambridge University Press.

Baldwin M, Kellogg CE, and Thorp J (1938) Soil classification. In: *Soils and Men*, pp. 979–1001. Washington, DC: United States Government Printing Office.

Bockheim JG and Gennadiyev AN (2000) The role of soil-forming processes in the definition of taxa in Soil Taxonomy and the World Soil Reference Base. *Geoderma* 95: 53–72.

Buol SW, Hole FD, McCraken RJ, and Southard RJ (1997) *Soil Genesis and Classification.* Ames, IA: Iowa State University Press.

Cline MG (1949) Basic principles of soil classification. *Soil Science* 67: 81–91.

Dokuchaev VV (1967) (originally published 1883)*Russian Chernozem. Selected Works of V.V. Dokuchaev 1.* Jerusalem Israel: Israel Program for Scientific Translations.

Glinka KD (1963) (originally published 1931)*Treatise on Soil Science (Pochvovedenie).* Jerusalem, Israel: Israel Program for Scientific Translations.

Haskett JD (1995) The philosophical basis of soil classification and its evolution. *Soil Science Society of America Journal* 59: 179–184.

Marbut CF (1921) The contribution of soil survey to soil science. *Society for the Promotion of Agricultural Science Proceedings* 41: 116–142.

Marbut CF (1927) *A Scheme for Soil Classification*, pp. 1–31. Washington, DC: First International Congress of Soil Science. American Organizing Committee.

Sibirtsev NM (1966) *Selected Works, 1.* Jerusalem, Israel: Israel Program for Scientific Translations.

Soil Survey Staff (1998) *Keys to Soil Taxonomy.* Washington, DC: US Department of Agriculture, Natural Resources Conservation Service.

Soil Survey Staff (1999) *Soil Taxonomy, A Basic System of Soil Classification for Making and Interpreting Soil Surveys.* Washington, DC: US Department of Agriculture, Natural Resources Conservation Service.

CLAY MINERALS

D G Schulze, Purdue University, West Lafayette, IN, USA

Introduction

The clay-size fraction of soils consists of mineral particles that are less than 2 μm in equivalent diameter. This is the realm of exceedingly small, crystalline particles dominated by planar arrays of SiO_4 structural units and many structural hydroxyls and water. These 'clay minerals' crystallize in the aqueous environment at the Earth's surface from the constituent ions released by dissolving (weathering) 'primary minerals' such as olivines, pyroxenes, feldspars, micas, quartz, and others that were formed under extreme heat and pressure deep within the Earth. Clay minerals are responsible for many of the soil's most important and characteristic physical and chemical properties. Fundamental soil properties such as cation exchange and shrink–swell properties, as well as practical considerations such as how well a particular soil will attenuate a specific pollutant, or how much fertilizer phosphorus will be fixed and unavailable to crops, are all influenced by molecular-scale differences in soil clay minerals.

Clay minerals are distinguished on the basis of their different crystal structures, and there is a close relationship between the crystal structure and the corresponding bulk physical and chemical properties of a particular type of clay. We begin by considering some of the properties of the major chemical elements that make up the clay minerals.

Major Element Composition of Clay Minerals

Most of the mass and volume of the Earth's crust is made up by only a few chemical elements. O and Si alone account for almost 75% of the mass, with most of the remainder, in order of decreasing abundance, consisting of Al, Fe, Ca, Na, K, Mg, Ti, H, P, and Mn. On a volume basis, oxygen alone accounts for more than 90% of the total volume. O, as O^{2-}, is the only abundant anion, while the other abundant elements are all cations. Most of these cations have only one stable oxidation state at the Earth's surface (Al^{3+}, Ca^{2+}, Na^{+}, K^{+}, Mg^{2+}, Ti^{4+}, H^{+}, P^{5+}); Fe (Fe^{2+}, Fe^{3+}) and Mn (Mn^{2+}, Mn^{3+}, Mn^{4+}) are the exceptions. The O^{2-} anions are much larger than most of the positively charged cations. The Earth's crust, therefore, can be characterized as large O atoms in an approximately close-packed arrangement held together by attraction to smaller cations located in the interstitial space.

Most of the elements in the crust and in soils occur in minerals, and the elements listed above are major constituents of the most abundant minerals, including clay minerals. Building on the concept of packing O atoms in space, we will consider atoms as rigid spheres, realizing that this is an oversimplified but convenient model for developing the key structural concepts.

Basic Structural Concepts

Tetrahedra and Octahedra

Two distinct structural features occur within the crystal structures of soil clay minerals as a consequence of packing the large O^{2-} ions together in space. The first consists of four O^{2-} ions packed closely together, and can be described as three O^{2-} ions arranged in a triangle with the fourth O^{2-} occupying the dimple formed by the other three (Figure 1). The centers of the four O^{2-} ions form the apices of a regular tetrahedron, and the small space in the center is called a 'tetrahedral site.' Cations located in tetrahedral sites are in fourfold or tetrahedral coordination, because they are surrounded by and bonded to four O^{2-} ions.

The second structural feature consists of six closely packed O^{2-} ions. Three of them are arranged in a triangle in one plane, and the other three, also in a triangle but rotated 60° relative to the first three, are in a second plane so that the two triangular groups intermesh (Figure 1). The centers of the six O^{2-} ions form the apices of a regular octahedron, and the small space in the center is called an 'octahedral site.' Cations located in the octahedral site are said to be in sixfold or octahedral coordination because they are surrounded by and bonded to six O^{2-} ions.

Tetrahedral and octahedral sites differ in another important way. The space that can be occupied by a cation in a tetrahedral site is smaller than the space that can be occupied in an octahedral site. Since cations vary in size, smaller cations tend to occur in tetrahedral sites, somewhat larger cations tend to occur in octahedral sites, and the largest cations must fit into spaces that are even larger than octahedral sites. Cations with sizes intermediate between the optimum for two sites can occur in either site. The Al^{3+} ion, for example, can occur in either octahedral or tetrahedral sites. Table 1 summarizes the structural sites in which cations tend to occur in clay minerals.

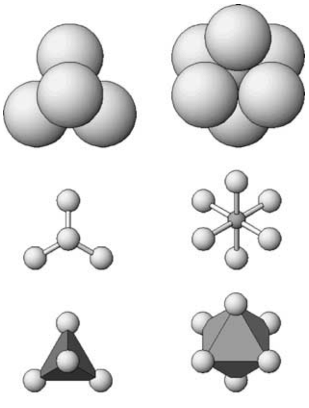


Figure 1 Spheres closely packed to form a tetrahedron and an octahedron. Note the appearance for the three different ways of drawing the model: as a sphere-packing model (top row), ball-and-stick model (middle row), and a polyhedral model (bottom row). (Adapted from Schulze DG (2002) An introduction to soil mineralogy. In: Dixon JB and Schulze DG (eds) *Soil Mineralogy with Environmental Applications*, pp. 1–35. Madison, WI: Soil Science Society of America, with permission.)

Representing Crystal Structures

Octahedra and tetrahedra are commonly represented using different types of models, each of which portrays the same concept, but highlights different structural features. Sphere-packing models give an impression of the space occupied by the atoms, ball-and-stick models highlight the bonds, while polyhedral models emphasize the tetrahedral and octahedral units (Figure 1). There is some ambiguity associated with each representation, and it is important to understand the correspondence between, and limitations of, each type of model. Polyhedral models are used here, along with some sphere-packing models.

Tetrahedral and Octahedral Sheets

The most common and abundant clay minerals belong to a group of minerals called phyllosilicates (from Greek 'phyllon,' meaning 'leaf') or sheet silicates. A common structural theme in all phyllosilicates is the presence of SiO_4 tetrahedra arranged into sheets. Octahedra arranged into sheets are also present in the structures of phyllosilicates and in some hydroxide minerals as well. Different combinations of the two sheets give rise to the different clay mineral structures.

Tetrahedral Sheet

Table 1 Type of structural sites in which common cations tend to occur in phyllosilicate mineral structures

Type of site	*Cation*
Tetrahedral only	Si^{4+}
Tetrahedral or octahedral	Al^{3+}, Fe^{3+}
Octahedral only	Mg^{2+}, Ti^{4+}, Fe^{2+}, Mn^{2+}
Interlayer sites	Na^{+}, Ca^{2+}, K^{+}

The tetrahedral sheet consists of SiO_4 tetrahedra arranged such that three of the four O^{2-} ions of each tetrahedron are shared with three nearest-neighbor tetrahedra (Figure 2). These shared O^{2-} ions are all in the same plane and are referred to as basal oxygens. Note that adjacent tetrahedra share only one O^{2-} between them (the tetrahedra share apices or corners). The fourth O^{2-} ion of each tetrahedron is not shared with another SiO_4 tetrahedron and is free to bond to other polyhedral elements. These unshared O^{2-} ions are referred to as apical oxygens. Since each basal oxygen contributes a charge of -1 to each Si^{4+} ion, the addition of H^+ ions to the apical oxygens to form hydroxyls should result in an electrically neutral tetrahedral sheet. Such individual tetrahedral sheets do not form stable mineral structures by themselves and only occur in combination with octahedral sheets, as described below.

Figure 2 shows all of the apical oxygens pointing in the same direction, namely, out of the plane of the paper toward the reader. This is the most common arrangement, but structures also occur in which the apical oxygens point alternately in opposite directions.

Octahedral Sheet

Analogous to the tetrahedral sheet, we can consider the octahedral sheet illustrated in Figure 3 as an assemblage of octahedra in which adjacent octahedra share two oxygens with one another. In other words, adjacent octahedra share edges. For the arrangement of octahedra shown in Figure 3, the octahedral sites are occupied by trivalent cations, typically Al^{3+}, and for charge balance, a proton (H^+) must be associated with each O^{2-}. (The H^+ takes up very little space, and the OH^- ion can be considered a sphere of roughly the same size as an O^{2-} ion.) Each OH^- contributes one-half a negative charge to each cation because each OH^- is shared between two octahedra. Each Al^{3+} cation is therefore effectively surrounded by $6 \times 0.5 = 3$ negative charges and the sheet is electrically neutral. Note the pattern of empty and filled octahedral sites. Two of every three possible octahedral sites are filled when trivalent cations are present in the octahedral sites. This arrangement is called dioctahedral and is the most common in soil clay minerals. If the octahedral sites are filled with divalent cations such as Mg^{2+} then every possible octahedral site must be occupied to produce an electrically

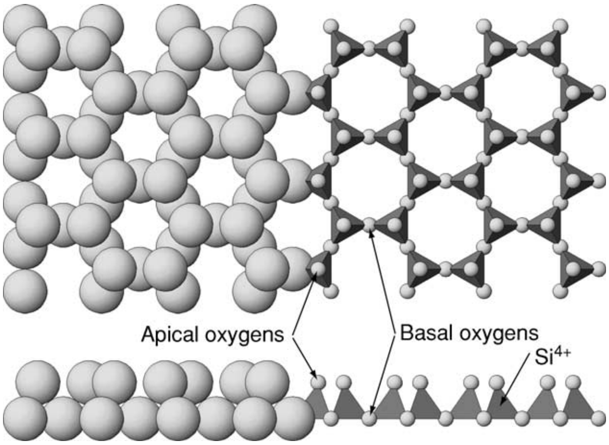


Figure 2 The tetrahedral sheet as a sphere-packing model (left half) and a polyhedral model (right half). (Adapted from Schulze DG (2002) An introduction to soil mineralogy. In: Dixon JB and Schulze DG (eds) *Soil Mineralogy with Environmental Applications*, pp. 1–35. Madison, WI: Soil Science Society of America, with permission.)

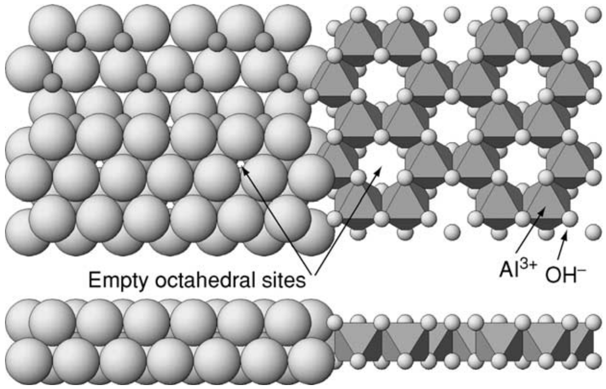


Figure 3 The octahedral sheet as a sphere-packing model (left half) and a polyhedral model (right half). The top three rows of spheres have been omitted from the sphere-packing model so that the underlying cations, represented here as Al^{3+}, can be seen more easily. The dioctahedral arrangement is shown here, in which two-thirds of the possible octahedral sites are filled with cations and one-third are empty. (Adapted from Schulze DG (2002) An introduction to soil mineralogy. In: Dixon JB and Schulze DG (eds) *Soil Mineralogy with Environmental Applications*, pp. 1–35. Madison, WI: Soil Science Society of America, with permission.)

neutral structure. This arrangement is called trioctahedral.

Octahedral sheets, stacked one on top of the other and held together by hydrogen bonds, make up the structure of gibbsite, $Al(OH)_3$, an aluminum hydroxide mineral that occurs in intensively leached soils. The structure of gibbsite is the simplest in a series of structures containing octahedral sheets.

Phyllosilicate Minerals Common In Soil Clays

Phyllosilicates are divided into two groups, 1:1- and 2:1-type minerals, based on the number of tetrahedral and octahedral sheets in the layer structure.

1:1-Type Minerals

1:1 Layer structure The 1:1 layer structure consists of a unit made up of one octahedral and one tetrahedral sheet, with the apical O^{2-} ions of the tetrahedral sheets being shared with (and part of) the octahedral sheet (Figure 4). There are three planes of anions (Figure 4b). One plane consists of the basal O^{2-} ions of the tetrahedral sheet, the second consists of O^{2-} ions common to both the tetrahedral and octahedral sheets (marked ‘a’ in Figure 4a) plus OH^- belonging to the octahedral sheet (‘b’ in Figure 4a), and the third consists only of OH^- belonging to the octahedral sheet.

Kaolinite The structure of kaolinite consists of 1:1 layers stacked one above the other. Kaolinite contains Al^{3+} in the octahedral sites and Si^{4+} in the tetrahedral sites (Figure 5). The 1:1 layer is electrically neutral and adjacent layers are held together by hydrogen bonding between the basal oxygens of the tetrahedral sheet and the hydroxyls of the exterior plane of the adjacent octahedral sheet.

Kaolinite is a common mineral in soils and is the most common member of this subgroup. It tends to be particularly abundant in more weathered soils such as Ultisols and Oxisols. Kaolinites have very little isomorphous substitution in either the tetrahedral or octahedral sheets and most kaolinites are close to the ideal formula $Al_2Si_2O_5(OH)_4$. The 1:1 layer has little or no permanent charge because of the low amount of substitution. Consequently, cation exchange capacities and surface areas are typically low. Soils high in kaolinite are generally less fertile than soils in which 2:1 clay minerals dominate.

Kaolinite can form in soils from Al and Si released by the weathering of primary and other secondary minerals. For example, feldspars often weather to kaolinite in soils formed from igneous rocks. Kaolinite can also be inherited from clayey, sedimentary soil parent materials.

Halloysite Halloysite has a 1:1 layer structure similar to kaolinite except that the 1:1 layers are separated by a layer of H_2O molecules when fully hydrated (Figure 5). This water is probably present as hydration shells around a small number of interlayer cations (cations that reside between two adjacent 1:1 or 2:1 layers), although the presence of interlayer

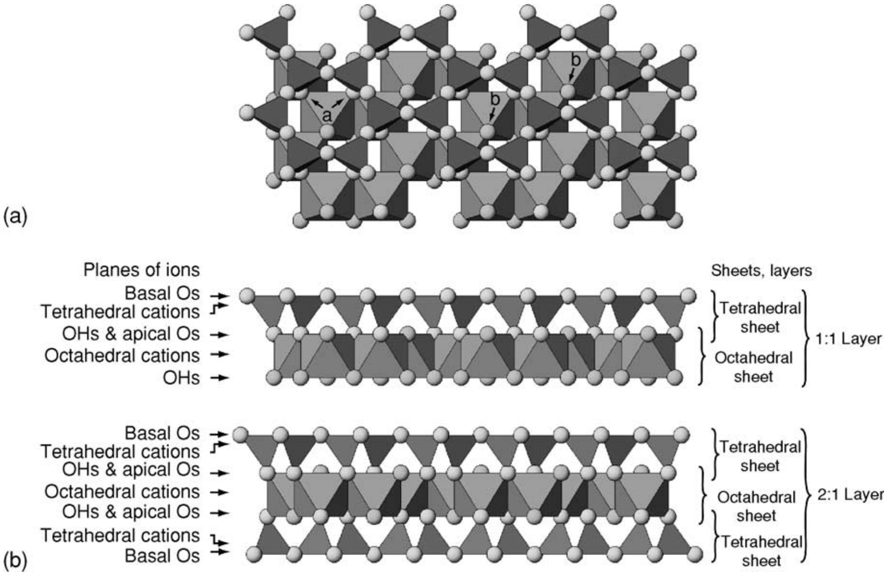


Figure 4 (a) Oblique view of the 1:1 layer structure illustrating the relationship between the tetrahedral and octahedral sheets. Note that adjacent apical oxygens of the tetrahedral sheet (arrows 'a') also define corners of octahedra in the octahedral sheet. Arrows marked 'b' point to OHs that lie directly in the center of the hexagonal rings of tetrahedra, although they appear off-center in this oblique view. (b) Edge view of the 1:1 and 2:1 layer structures, illustrating phyllosilicate nomenclature. (Adapted from Schulze DG (2002) An introduction to soil mineralogy. In: Dixon JB and Schulze DG (eds) *Soil Mineralogy with Environmental Applications*, pp. 1–35. Madison, WI: Soil Science Society of America, with permission.)

cations and the existence of layer charge to attract them has been difficult to confirm. Most clay silicates occur as thin plates, but halloysite often occurs as tubular or spherical particles.

Halloysite is usually found in soils formed from volcanic deposits, particularly volcanic ash and glass. It is a common clay mineral in the Andisol soil order. Halloysite forms early in the weathering process but it is generally less stable than kaolinite and gives way to kaolinite with time.

2:1-Type Minerals

In contrast to the 1:1 minerals, which are represented in soils by only two major minerals, the 2:1 minerals are structurally more diverse and are represented by several mineral species.

2:1 Layer structure The 2:1 layer structure consists of two tetrahedral sheets, with one bound to each side of an octahedral sheet (Figure 4b). There are four planes of anions. The outer two consist of the basal oxygens of the two tetrahedral sheets, while the two inner planes consist of oxygens common to the octahedral sheet and the two tetrahedral sheets, plus the hydroxyls belonging to the octahedral sheet.

Pyrophyllite The simple structure of pyrophyllite is a good starting point for discussing 2:1 structures. Pyrophyllite consists of 2:1 layers stacked one above the other. The tetrahedral sheets contain only Si^{4+} and the octahedral sheet contains only Al^{3+}, resulting in the ideal formula $Al_2(Si_4)O_{10}(OH)_2$. The charge is balanced completely within the 2:1 layer, making the layer electrically neutral, and adjacent 2:1 layers are held together only by weak van der Waals forces. Pyrophyllite occurs only rarely in soils, usually only when it is inherited from low-grade metamorphic rocks.

Micas Mica minerals have the 2:1 layer structure described for pyrophyllite but with two important differences. First, instead of having only Si^{4+} in the tetrahedral sites, one-quarter of the tetrahedral sites are occupied by Al^{3+}. Because of this substitution, there is an excess of one negative charge per formula unit in the 2:1 layer. Second, this excess negative charge is balanced by monovalent cations, commonly K^+, that occupy interlayer sites between two 2:1 layers (Figure 5). This gives an ideal formula of $KAl_2(AlSi_3)O_{10}(OH)_2$ for a mica mineral with Al in the octahedral sites.

The octahedral sheet can contain either Al^{3+} (the dioctahedral case; Figure 3) or Mg^{2+} (the trioctahedral case), and there are several different mica species, because Fe^{2+} and Fe^{3+} can substitute for Mg^{2+} and Al^{3+} in the octahedral sheet and Na^+ and Ca^{2+} can substitute for K^+ in the interlayer.

Mica in soils is usually inherited from the parent rock and is likely to occur in soils derived from various

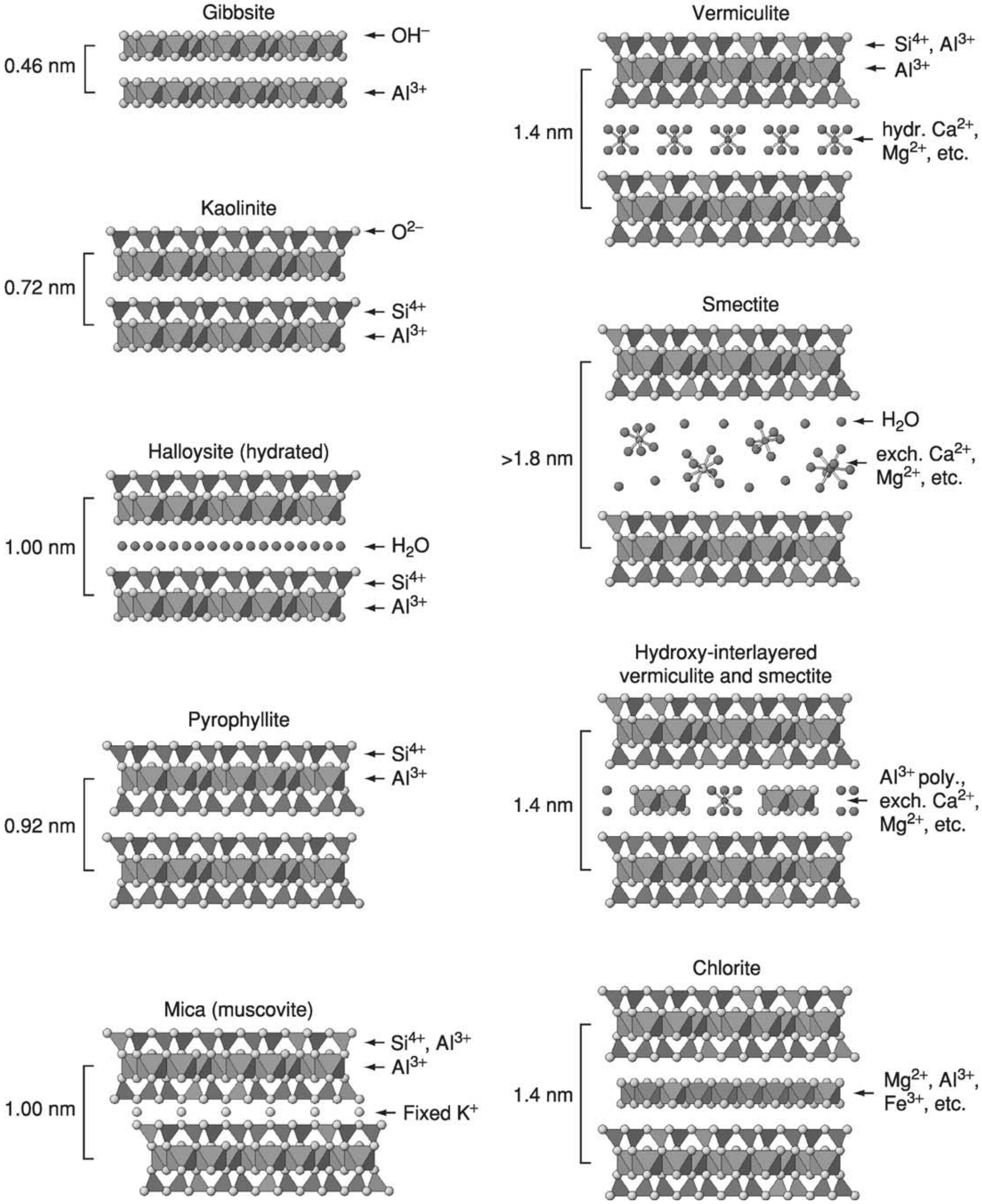


Figure 5 Structural scheme of soil minerals based on tetrahedral and octahedral sheets. (Adapted from Schulze DG (2002) An introduction to soil mineralogy. In: Dixon JB and Schulze DG (eds) *Soil Mineralogy with Environmental Applications*, pp. 1–35. Madison, WI: Soil Science Society of America, with permission.)

igneous and metamorphic rocks, as well as from sediments derived from them. Muscovite, biotite, and phlogopite are the three most common mica group minerals in rocks, and consequently in soils. All three contain K in the interlayer, but they differ in the composition of the octahedral sheet and whether they are di- or trioctahedral. Mica in the clay fraction of soils and sediments differs somewhat from the macroscopic muscovite mica it most closely resembles. This clay-size mica is often referred to as illite. Glauconite is another mica mineral that is similar to illite, but it contains more Fe and less Al in its octahedral sheet than illite.

Micas weather to other minerals, particularly to vermiculites and smectites, and the K^+ released during weathering is an important source of K for plants. As a rule, the dioctahedral micas such as muscovite are more resistant to weathering than trioctahedral micas. Thus, muscovite tends to be the most common mica mineral in soils.

Vermiculites Vermiculite has a 2:1 layer structure as described for mica, but, instead of having a layer charge of ~1 per formula unit and K^+ in interlayer positions, vermiculite has a layer charge of 0.9–0.6 per formula unit and contains hydrated exchangeable cations, primarily Ca and Mg, in the interlayer (Figure 5). A typical formula for an idealized vermiculite weathered from muscovite is: $M^+_{0.75}$ $Al_2(Si_{3.25}Al_{0.75})$ $O_{10}(OH)_2$, where M^+ represents exchangeable cations. The high charge per formula unit gives vermiculite a high cation exchange capacity and causes vermiculte to have a high affinity for weakly hydrated cations such as K^+, NH_4^+, and Cs^+. Fixation of K^+ by vermiculite can be significant in soils that are high in vermiculite and that have not received large amounts of chemical fertilizers.

Vermiculites in soils are believed to form almost exclusively from the weathering of micas and chlorites. The weathering of micas to vermiculite (or smectite) is believed to occur by replacement of K^+ in the interlayer sites with hydrated exchangeable cations. The integrity of the 2:1 layer is preserved, but there is a reduction in the layer charge. Vermiculite does not swell as extensively as smectite and this is shown in Figure 5 by the presence of only two planes of water molecules surrounding the hydrated cations in the interlayer space.

Smectites The smectite group consists of minerals with the 2:1 structure already discussed for mica and vermiculite, but with a still lower charge per formula weight, namely 0.6–0.2. As in vermiculite, the interlayer contains exchangeable cations (Figure 5). An idealized formula for a common soil smectite, the mineral beidellite, is: $M^+_{0.33}Al_2$ $(Si_{3.67}Al_{0.33})$ $O_{10}(OH)_2$, where M^+ represents exchangeable cations, typically Ca^{2+} and Mg^{2+}.

The most common smectite minerals range in composition between three end-members: montmorillonite, beidellite, and nontronite. All are dioctahedral, but they differ in the composition of the tetrahedral and octahedral sheets. Smectites do not fix K^+ as readily as do vermiculites because smectites have a lower layer charge, but smectites swell more extensively than vermiculite. This is illustrated in Figure 5 by the larger spacing between the 2:1 layers.

Smectites are important minerals in temperate-region soils. Many plant nutrients are held in an available form on the cation exchange sites of soil smectites. Soils rich in smectite tend to be very effective at attenuating many organic and inorganic pollutants because of the high surface area and adsorptive properties of the smectites. Smectites shrink upon drying and swell upon wetting. This shrink–swell behavior is most pronounced in the Vertisol order and in vertic subgroups of other soil orders. The shrink–swell properties lead to cracking and shifting problems when houses, roads, and other structures are built on smectitic soils.

Chlorites Like mica, chlorite minerals have a 2:1 layer structure with an excess of negative charge. In contrast to mica, however, the excess charge is balanced by a positively charged interlayer hydroxide sheet (Figure 5), rather than K^+. The interlayer hydroxide sheet is an octahedral sheet as shown in Figure 3 and can be either di- or trioctahedral. Instead of being electrically neutral as in gibbsite, the hydroxide sheet has a positive charge caused by substitution of higher-valence cations for lower-valence ones, for example, $Mg_2Al(OH)_6^+$. Either octahedral sheet – the one that is part of the 2:1 layer or the interlayer hydroxide sheet – can be di- or trioctahedral, and can contain Mg^{2+}, Fe^{2+}, Mn^{2+}, Ni^{2+}, Al^{3+}, Fe^{3+}, and Cr^{3+}, giving a large number of different mineral species.

Chlorite minerals in soils are often primary minerals inherited from either metamorphic or igneous rocks. They may also be inherited from sedimentary rocks such as shales, or from hydrothermally altered sediments. Chlorites are rather infrequent minerals in soils and when present they generally occur in small amounts. Chlorite weathers to form vermiculite and smectite, and the ease with which chlorites break down makes them sensitive indicators of weathering.

Hydroxy-interlayered vermiculite and smectite Hydroxy-interlayered vermiculite and smectite can be considered a solid solution with vermiculite or smectite as one end-member and chlorite as the other. Hydroxy-interlayered minerals form as Al^{3+} released during weathering hydrolyzes and polymerizes to form large polycations with a postulated formula of $Al_6(OH)_{15}^{3+}$ (or similar) in the interlayers of vermiculite and smectite. These polycations balance some of the charge of the 2:1 layer. The combination of a 2:1 layer with hydroxy Al in the interlayer gives a structure similar to that of chlorite (Figure 5). Thus, these minerals are also called secondary chlorites. The degree of filling of the interlayer with hydroxy Al can vary from none to almost complete, with properties of the clay varying accordingly. The interlayer hydroxy Al is not exchangeable, therefore it lowers the cation exchange capacity of smectite or vermiculite almost linearly as a function of the amount of Al adsorbed in the interlayer.

Interlayer hydroxy Al prevents smectite from shrinking and swelling as it normally would. In vermiculite, it reduces K$^+$ fixation by lowering the exchange capacity and by preventing the interlayer from collapsing around the K$^+$. The positively charged hydroxy interlayers also provide potential sites for anion adsorption. Hydroxy-interlayered vermiculite and smectite are most common in Alfisols and Ultisols. Within a given profile, they tend to be most abundant near the soil surface.

Interstratification in phyllosilicates Because of the structural similarities of all of the phyllosilicate minerals just discussed, phyllosilicates in soils do not always occur as discrete particles of mica, vermiculite, smectite, chlorite, or kaolinite. For example, instead of being made up of a stack of identical 2:1 vermiculite layers, one physically discrete particle may consist of a mixture of both mica and vermiculite layers instead. Such minerals are referred to as 'mixed-layer' or 'interstratified minerals.'

Different types of interstratified minerals have been identified. Two-component systems include: mica-vermiculite, mica-smectite, mica-chlorite, kaolinite-smectite, and others. Three-component mixed layer systems can also occur. The sequence of layers can be either regular or random. A regularly interstratified mineral consisting of two types of layers denoted by A and B could have a sequence such as ABABAB . . . , or ABBABBABB . . . , or any other repeating sequence. In a randomly interstratified mineral, the sequence of layers is random – for example, ABBABAABBAAA . . . . Random interstratification of layer-silicates is more common in soils than regular interstratification, though regular interstratification, especially in weathering micas, is not rare.

Partial removal of interlayer K from micas or of interlayer hydroxide from chlorite is one way that interstratified minerals can form in soils. Other possibilities include (1) fixation of adsorbed K$^+$ by some vermiculite layers to give mica-like layers, and (2) the formation of hydroxide interlayers to produce chlorite-like layers.

Palygorskite and sepiolite Palygorskite and sepiolite are considered phyllosilicates, but are distinct structurally from the typical 1:1 and 2:1 layer structures. Both minerals have continuous tetrahedral sheets, but adjacent bands of tetrahedra within one tetrahedral sheet point in opposite directions rather than in one direction as in the 1:1 and 2:1 structures. The result is a structure that can be described as ribbons of 2:1 layers joined at their edges, as illustrated in Figure 6. Water molecules occur in the spaces between the ribbons. The 2:1 ribbons are wider in sepiolite than in palygorskite.

Palygorskite and sepiolite are often found in soils of arid and semiarid environments. Both minerals have a fibrous morphology in contrast to the platy morphology of most 1:1 and 2:1 minerals.

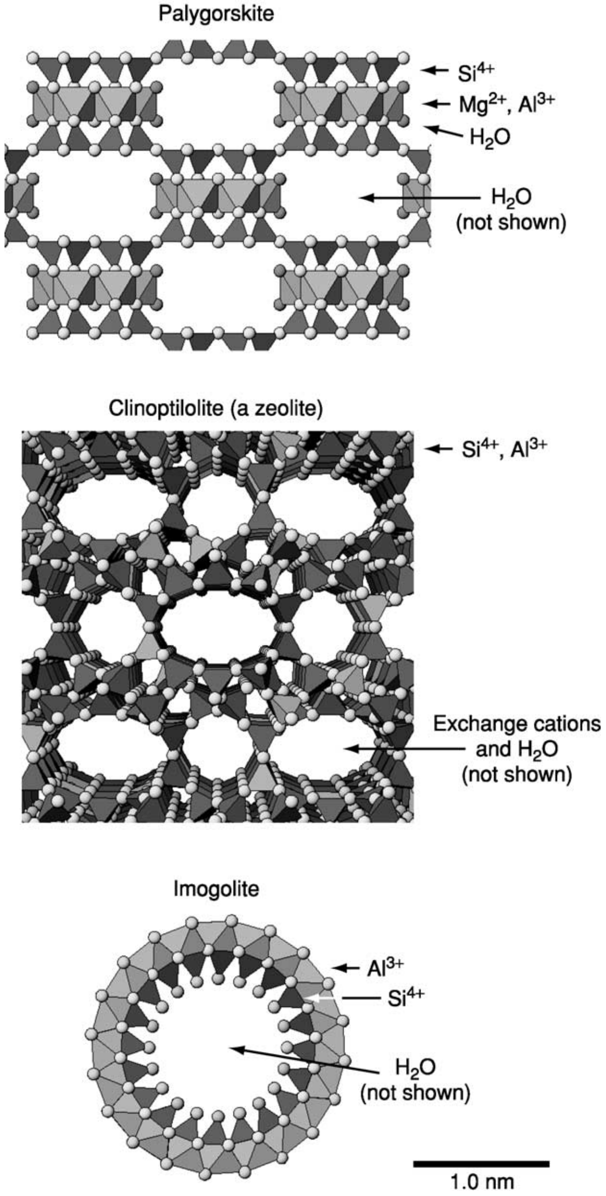


Figure 6 Structural models of representatives of three other aluminosilicate mineral groups that occur frequently in soils. All three are drawn to the same scale. (Adapted from Schulze DG (2002) An introduction to soil mineralogy. In: Dixon JB and Schulze DG (eds) *Soil Mineralogy with Environmental Applications*, pp. 1–35. Madison, WI: Soil Science Society of America, with permission.)

Other Minerals that Occur in Soil Clays

The phyllosilicate clay minerals described above are the most abundant and common in most soils, and they are the minerals that are usually considered to make up the 'clay minerals' group. Several additional minerals or mineral groups require mention as well. Some occur only in particular soils, others occur in many soils, but usually only at low concentrations.

Zeolites

Zeolites are a large group of aluminosilicate minerals that consist structurally of SiO_4 tetrahedra arranged in ways that result in large amounts of pore space within the crystals (Figure 6). Aluminum substitutes for Si in the tetrahedral sites and, as a result, the $(Si,Al)O_4$ framework has a net negative charge. The charge is balanced by cations that reside in the channels and pores along with water molecules. Because the cations are exchangeable, zeolites have cation exchange properties similar to the phyllosilicates, but, because the tetrahedral framework of the zeolites is rigid and the size of the pores is fixed, small cations can move into and out of the pores freely, while larger cations are excluded. Thus, zeolites are often referred to as 'molecular sieves' because of their very selective cation exchange properties. Zeolites are relatively rare in soils because they weather easily in humid regions, but they occur in some soils in arid regions.

Allophane and Imogolite

The aluminosilicate minerals discussed above have three-dimensional crystal structures, with atoms packed together in a more or less regular manner over relatively long distances (10s of nanometers). They exhibit long-range order. Two other aluminosilicates, allophane and imogolite, exhibit short-range (or local) order. Structures with short-range order exhibit order over several nanometers, but on a larger scale the structure is disordered.

Allophane is a material consisting chemically of variable amounts of O^{2-}, OH^-, Al^{3+}, and Si^{4+}, and characterized by short-range order and a predominance of Si-O-Al bonds. It consists of small (3.5–5.0 nm) spheres, the structure of which has not been determined. The spheres clump together to form irregular aggregates. Imogolite consists of tubes several micrometers long with an outer diameter of 2.3–2.7 nm and an inner diameter of approx. 1.0 nm. The tubes consist of a single dioctahedral sheet with the inner surface OH replaced by SiO_3OH groups (Figure 6). Several individual tubes are arranged in bundles 10–30 nm across to give thread-like particles several micrometers long.

Allophane and imogolite usually occur as weathering products of volcanic ash and are important minerals in the Andisol soil order. Imogolite has also been identified in the Bs horizons of Spodosols. Allophane and imogolite can specifically adsorb many inorganic and organic compounds. Andisols, for example, usually fix large amounts of phosphate, making it unavailable to plants, and the large amounts of organic matter common in Andisols may be due, in part, to adsorption of organic molecules by allophane and imogolite. Soils containing large amounts of allophane and imogolite usually have unique physical properties such as a low bulk density, high water-holding capacity, high liquid and plasticity limits, and a thixotropic consistence.

Aluminum Hydroxide Minerals

Gibbsite, which has already been mentioned as simply a stack of octahedral sheets containing Al^{3+} in the octahedral sites (Figure 5), occurs in situations where weathering and leaching have been intense or long. Gibbsite often occurs in Ultisols and Oxisols and can be important in the fixation of phosphate fertilizers.

Iron Oxide Minerals

Fe is almost as abundant as Al in the Earth's crust, and one or more strongly colored iron oxide minerals are ubiquitous accessory minerals in almost all soil clays. Goethite (FeOOH) is the most common soil iron oxide mineral and accounts for the yellowish-brown colors of many soils. Hematite (Fe_2O_3) is common also and accounts for the red colors of many soils. Iron oxide surfaces are highly reactive and can sorb and fix phosphate, many transition metals, and organic compounds. Redox reactions are an important aspect of soil iron oxide mineralogy.

Manganese Oxide Minerals

Manganese is about 50 times less abundant in the Earth's crust than Fe, so minerals containing large amounts of Mn are proportionally less abundant. Nevertheless, some Mn oxide minerals occur in almost all soils, with birnessite ($(Na,Ca)(Mn^{3+},Mn^{4+})_7O_{14}\cdot 2.8H_2O$) being one of the more common ones. The Mn^{4+} and Mn^{3+} in manganese oxide minerals is reduced to soluble Mn^{2+} under relatively mild reducing conditions, making Mn oxide minerals important in many soil redox reactions.

Titanium Oxide Minerals

Ti is slightly more abundant than Mn in the Earth's crust, and two Ti minerals, anatase and rutile (both TiO_2) occur widely in soil clays. The Ti-oxide minerals, however, are relatively inert chemically and have a negligible impact on most soil properties.

Carbonates, Sulfates, and Soluble Salts

In semiarid to arid climates, calcite ($CaCO_3$), gypsum ($CaSO_4\cdot 2H_2O$), and an array of evaporite minerals (minerals with solubilites greater than or equal to gypsum) often occur coassociated with the aluminosilicate and oxide minerals discussed above. The impact of these minerals varies greatly, depending on the amount and type of mineral present.

List of Technical Nomenclature

Minerals	gibbsite, kaolinite, halloysite, pyrophyllite, mica, illite, glauconite, muscovite, smectite, montmorillonite, beidellite, nontronite, vermiculite, chlorite, hydroxy-interlayered smectite and vermiculite, interstratified clay minerals, allophane, imogolite, palygorskite, sepiolite, zeolites, iron oxide minerals, goethite, hematite, aluminum hydroxide minerals, manganese oxide minerals, birnessite, titanium oxide minerals, anatase, rutile, calcite, gypsum, evaporite minerals, biotite, phlogopite

See also: **Kinetic Models**; **Thermodynamics of Soil Water**

Further Reading

Brindley GW and Brown G (eds) (1980) *Crystal Structures of Clay Minerals and Their X-ray Identification.* London, UK: Mineralogical Society.

Dixon JB and Schulze DG (2002) *Soil Mineralogy with Environmental Applications*. Madison, WI: Soil Science Society of America.

Dixon JB and Weed SB (1989) *Minerals in Soil Environments*. Madison, WI: Soil Science Society of America.

Klein C and Hurlbut CS Jr (1993) *Manual of Mineralogy* (after James D. Dana), 21st edn. New York: John Wiley.

Moore DM and Reynolds RC Jr (1997) *X-ray Diffraction and Identification and Analysis of Clay Minerals*, 2nd edn. New York: Oxford University Press.

CLIMATE CHANGE IMPACTS

P Bullock, Cranfield University–Silsoe, Silsoe, UK

Introduction

Soils form through the interaction of a number of influences, including climate, relief and/or landscape, parent material, organisms (including fauna, flora, and humans) and time. The nature of this interaction varies in different parts of the world, resulting in several thousand types of soil worldwide. It takes thousands of years for a soil to form, and most soils are still evolving as a result of changes in some of these soil-forming factors, particularly climate and vegetation, over the past few millennia. Changes in any of the soil-forming factors, such as climate, will impact directly and indirectly on current soils, with important implications for their development and use.

Unraveling the likely extent and impact of climate change on soils is a complex process and one in which progress has been slow. It is made all the more complicated by the fact that not only can soils be strongly affected directly and indirectly by climate change, but soils themselves can act as both source and sink for greenhouse gases and thus have the potential for either positive or negative feedback to climate change. The lack of specificity of the global circulation models (GCMs) at present, combined with the complexity of the interaction of the various soil-forming processes and the fact that there is still a limited knowledge of many of them, particularly biological ones, makes it difficult to quantify the changes that will ensue. On the basis of current knowledge, it is only possible to describe the likely impacts of climate change on soils in a qualitative or semiquantitative way and highlight the key changes, their direction where there is adequate climate change information, and the implications of them.

Climate Change Predictions

Estimates of global climate change are continually being revised as models are being improved and new data collected. The current main sources of information available on the likely extent of climate change are the Third Assessment Reports of the Intergovernmental Panel on Climate Change (IPCC).

These reports conclude that the globally averaged surface temperature will increase by between 1.4 and 5.8°C over the period 1990 to 2100 and that nearly all land areas will warm by more than this global average, particularly in northern high latitudes in the cold season (Table 1). Increased summer continental drying and associated risk of drought is likely over most midlatitude interiors, the main areas for which the models are consistent with one another.

Table 1 Projected changes in extreme weather during the twenty-first century

Changes in weather phenomenon	*Confidence in projected changes*
Higher maximum temperatures and more hot days over nearly all land areas	Very likely
Higher minimum temperatures, fewer cold days and frost days over nearly all land areas	Very likely
Reduced diurnal temperature over most land areas	Very likely
Increase in heat index[a] over land areas	Very likely, over most areas
More intense precipitation events	Very likely, over many areas
Increased summer continental drying and associated risk of drought	Likely over most midlatitude continental interiors (lack of consistent projections in other areas)
Increase in tropical cyclone wind intensities	Likely, over some areas
Increase in tropical mean and peak precipitation intensities	Likely, over some areas

[a] Heat index: a combination of temperature and humidity that measures effects on human comfort.

Adapted from Houghton JT, Ding Y, Griggs DJ *et al.* (eds) (2001) *Climate Change 2001: The Scientific Basis*, p. 72. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press, with permission.

Global average precipitation is projected to increase during the twenty-first century, with larger year-to-year variations in precipitation and more intense precipitation events very likely. Depending on the regional scale, both increases and decreases in precipitation are likely to be seen. Models predict an increase in precipitation in both summer and winter in high-latitude regions. In winter, increases in precipitation are predicted over northern midlatitudes and tropical Africa, and in summer in south and east Asia. There is likely to be a decrease in rainfall in Australia, Central America, and South Africa. Increased intensity of midlatitude storms, tropical peak wind intensities, and mean and peak precipitation intensities are also likely. Global mean sea level is projected to rise by 0.09–0.88 m between 1990 and 2100, and this will impact on soils of coastal regions. These changes, particularly in combination, are likely to have a major impact on soil formation, soil processes, land use, and rates of land degradation, especially soil erosion, desertification, and salinization. The Third Assessment Reports, although providing much-needed synthesis of model predictions of climate change, give little attention to the actual impacts of these changes on soils.

Timescale for Change

The diverse range of physical, chemical, and biological processes that affect soil formation and modify soil properties will respond to climate change according to varying timescales (Table 2). Parameters such as bulk density, porosity, infiltration rate, permeability, nitrate content, and composition of soil air can change on a daily basis, depending on the weather. At the other end of the timescale, weathering of minerals as part of soil formation and changes in soil texture are more likely to be on millennial timescales.

The effect of climate change will be to modify the rates of these processes and lead to changes in soil properties, with a range of implications for soil formation, soil genesis, and the way in which soils can be used.

Climate Change Impacts on Soil Water and Soil Temperature

Soil Water

The main effects of climate change on soils will be through changes to soil-moisture regimes. Soil moisture is a key driver to most soil processes and is instrumental in the use that can be made of soils. As climate changes, soil-moisture levels will be influenced by direct climatic effects (precipitation, temperature effects on evaporation), climate-induced changes in vegetation, different plant growth rates and different cycles, different rates of soil-water extraction, and the effect of enhanced CO_2 levels on plant transpiration. Changes in soil-water fluxes may also feed back to the climate itself and even contribute to drought conditions by decreasing available moisture, altering circulation patterns, and increasing air temperatures.

Soil water can be influenced in a number of ways by climate change. Changes in precipitation will rapidly affect soil water, since the timescale for response to rainfall in the soil is usually within a few hours. Increasing temperatures will also lead to greater evapotranspiration and hence loss of water from the soil. Much will also depend on land use and vegetation cover, as these will influence water use.

Several soil-forming processes, including organic matter turnover, structure formation, weathering, podzolization, clay translocation, and gleying, are strongly affected by soil-moisture contents. The type

Table 2 Timescale for changes in soils with change in climate

Timescale categories	*Soil parameter*	*Properties and characteristics*	*Horizons and phases*	*Regimes*
$<10^{-1}$ years	Temperature; moisture content; bulk density; total porosity; infiltration rate; permeability; composition of soil air; nitrate content	Compaction; drainage; workability		Aeration; heat regime
10^{-1}–10^{0} years	Total water capacity; field capacity; hydraulic conductivity; pH; nutrient status; composition of soil solution	Microbiota		Microbial activity; human-controlled plant-nutrient regime; erosion
10^{0}–10^{1} years	Wilting percentage; soil acidity; cation exchange capacity; exchangeable cations	Type of soil structure; annual roots biota; mesofauna; litter, fluvic, gleyic, stagnic properties; slickensides	Sulfuric horizon; gelundic, inundic, salic, yermic phases (fine earth properties only)	Moisture; natural fertility; salinity–alkalinity; desertification; permafrost
10^{1}–10^{2} years	Specific surface; clay mineral association; organic matter content	Tree roots; soil biota; salic, calcareous, sodic, vertic properties	Histic (<20 cm), ochric, gypsic, albic, and immature natric and spodic horizons (Podsols); gilgai, placic, sodic, takyric phase	
10^{2}–10^{3} years	Primary mineral composition; chemical composition of mineral part	Tree roots; color (yellowish/reddish); iron concretions; soil depth; cracking; soft powdered lime; indurated subsoil	Histic, mollic, umbric, calcic, albic, natric, cambic, spodic, and nitic horizons; plinthite, placic, yermic phases (stone surfaces)	
$>10^{3}$ years	Texture; particle-size distribution; particle density	Parent material; depth; abrupt textural change	Argic, oxic, petrocalcic, petrogypsic horizons; duripan, fragipan, skeletic, petroferric, lithic, rudic phases	

Adapted from Varallyay GY (1990) Influence of climatic change on soil-moisture regime, texture, structure, and erosion. In: Scharpenseel HW, Schomaker M, and Ayoub A (eds) *Soils on a Warmer Earth*, Development in Soil Science 20, p. 46. Amsterdam, the Netherlands: Elsevier, with permission.

of soil structure that develops under a particular climatic regime is particularly important, because it affects the processes of runoff, infiltration, percolation, and drainage, processes that are vital in the distribution of water across the landscape.

Those areas predicted to have warmer temperatures and less rainfall will have less soil moisture, with potentially large implications for the crops that can be grown and the natural and seminatural ecosystems that can continue to exist. Given these circumstances, arid areas which are already marginal for agriculture may become totally unsuitable for agriculture, new areas will become classed as arid lands, with the associated difficulties of maintaining an agricultural base and survival of its ecosystems, and a larger proportion of the world's land will become unsuitable for agricultural production. However, in view of the many different interacting influences on soil-moisture levels, it is difficult to predict the impact of climate change on soil water at regional or local level.

Soil Temperature

There is a close relationship between air temperature and soil temperature, and an increase in air temperature leads to an increase in soil temperature. The temperature regime of the soil is governed by gains and losses of radiation at the surface, the process of evaporation, heat conduction through the soil profile, and convective transfer via the movement of gas and water. As with soil moisture, soil temperature is a prime factor in most soil processes. Warmer soil temperatures everywhere will accelerate soil processes, leading to more rapid decomposition of organic matter, increased microbiological activity, quicker release of nutrients, increased rates of nitrification, and generally increased chemical weathering of minerals. However, this effect could be minimized or reversed if soils become drier.

Gelisols (*see* **Cold-Region Soils**), one of the two orders of soils largely defined on a climatic basis, are particularly vulnerable to increases in temperature. Gelisols occupy large areas of the northern latitudes where thickening of the seasonally thawed layer above permafrost is predicted. Large areas of permafrost are likely to begin to thaw, with consequent changes in soil drainage, increased mass movement, and thermal erosion likely. There is thus likely to be a significant change in the nature and the distribution of Gelisols. Aridisols, the other group of soils with a strong link to climate, cover more than 12% of the globe. Already such soils are very difficult to manage, are at the margin of cultivation, and are subject to soil erosion, desertification, and salinization. Increasing temperatures, accompanied by increased evaporation, would exacerbate these problems.

Changes in Soil-Forming Processes and Properties

Soil Organic Matter

Soil organic matter is arguably the most important soil component, influencing soil structure, water-holding capacity, soil stability, nutrient storage and turnover, and oxygen-holding capacity, properties that are fundamental in maintaining and improving soil quality. (*see* **Organic Matter:** Principles and Processes.) A decline in organic matter content increases the susceptibility to soil erosion. Organic matter is particularly important as the prime habitat for immense numbers and variety of soil fauna and microflora, which play a critical role in the health and productivity of soils. It is highly susceptible to changes in land use and management and to changes in soil temperature and moisture. In the last decades of the twentieth century, changes in land use and management, particularly conversion of forest and grassland to agriculture, have led to a significant decline in organic matter levels in some parts of the world.

Soil organic matter is one of the major pools of carbon in the biosphere and, unlike most other soil properties, is important both as a driver of climate change (*see* **Carbon Emissions and Sequestration**) and as a response variable to climate change, capable of acting both as a source and sink of carbon. How climate change will impact on soil organic matter is a matter of considerable debate. On the one hand it is recognized that global warming and increasing CO_2 levels in the atmosphere can favor increased plant growth, which in turn would increase organic inputs to the soil. On the other hand, a rise in air temperature and that of the soil are likely to increase decomposition and loss of soil organic matter. Soil organic matter represents a major pool of carbon in the biosphere, estimated at about 1500 Gt of carbon, double that in the atmosphere at present. There is thus significant interest in the fate of such carbon, particularly the extent to which soils and land use may sequester carbon from the atmosphere or lose organic carbon to the atmosphere. The balance of opinion currently is that, in the absence of mitigating action, losses through organic matter decomposition are likely to exceed levels gained from increased plant growth, thus adding to atmospheric CO_2 levels and the greenhouse gas effect, and to lower levels of soil organic matter.

A group of world soils that are particularly vulnerable to climate change are peats (Histosols). (*see*

Organic Soils.) These are soils that are dominantly composed of organic matter throughout their whole depth. Already they have been under threat worldwide because of drainage for use in crop production. Further drying-out of the soils in a warmer, drier climate with concomitant oxidation could lead to losses of this important, highly productive soil type, in addition to releasing a large amount of carbon to the atmosphere as CO_2.

Soil Structure

The structure of the soil, that is, the way in which the soil particles combine together (*see* **Structure**), is an important property of soils, affecting the movement of gases, water, nutrients, soil fauna, and the emergence of crops. The nature and quality of the structure of particular soils are strongly influenced by the amount and quality of organic matter present and also by the inorganic constituents of the soil matrix, cultivation methods, and by natural physical processes such as shrink–swell and freeze–thaw. A decline in soil organic matter levels as could occur under climate change would lead to a decrease in soil aggregate stability, an increase in susceptibility to compaction, lower infiltration rates, increased runoff, and hence an increased susceptibility to erosion.

The structure of the soil is also important for water quality, seepage, and building foundations. Soils with high clay contents, particularly those with smectitic mineralogy, have the potential to shrink when dry, resulting in formation of large cracks and fissures. When the soils rewet again, the cracks close. Drier climatic conditions would be expected to increase the frequency and size of crack formation. The importance of soil structure in determining pathways for the movement of water, pollutants, and contaminants through the soil is now well recognized and future management of soil water, movement of nutrients within the soil and the landscape, and the movement of contaminants once in the soil are important aspects to be considered in a climate change environment. In some areas, as a result of increased cracking and change in structure, there could be an increase in flash-flooding as water moves more rapidly through or over the soil to rivers.

Areas that are likely to experience increased drought may also find that buildings, roads, etc., built to particular specifications relating to current conditions, have foundations which become unstable as soils dry out more. For example, the increase in subsidence to buildings in southeast England alone in the decade to 1997, the warmest recorded, cost the insurance industry several billion euros.

Clay soils with a high shrink–swell potential are termed Vertisols and occupy some 260 million hectares globally. They are important agricultural soils in Africa and Asia but are renowned as some of the problem soils of the world, owing to difficulty in managing them for cultivation. They become very hard and difficult to cultivate when dry and too plastic for trafficking or cultivation when wet. Changes in climate, particularly increased drying of the soil, will lead to increased difficulties in managing these high-clay-content soils and may also lead to other world soils developing the properties of this soil type.

Soil Fauna and Soil Flora

Soil fauna and flora are essential components of all soils. Particularly vital is their role in the retention, breakdown, and incorporation of plant remains, nutrient cycling, and their influence on soil structure and porosity. There are thousands of species in a square meter of most soils but, despite these numbers and their importance, little is known about the roles of the species. Global warming may not have a direct effect on the ecological composition, because soil fauna and flora have a relatively broad temperature optimum. However, changes in ecosystems and migration of vegetation zones are likely in some areas as a result of increased temperature and changes in rainfall. Soil flora and fauna may be seriously affected by such changes, because their migration rates are likely to be too slow.

A further significant impact of climate change on soil fauna and flora is through enhanced CO_2 levels in the atmosphere, leading to enhanced plant growth and in turn increased allocation of carbon belowground. The microbial population and its activity under this regime would increase, potentially leading to higher rates of nitrogen fixation, nitrogen immobilization and denitrification, increased mycorrhizal associations, increased soil aggregation, and increased weathering of minerals. However, as noted above, much will depend on what balance between increased plant growth on the one hand and increased decomposition of soil organic matter on the other under a changing climate will emerge.

Acidification and Nutrient Status

While temperature increases are forecast for most parts of the world, there is less certainty about precipitation changes. Significant increases in rainfall will lead to increases in leaching, loss of nutrients, and increasing acidification, depending on the buffering pools existing in soils. Decreases in rainfall coupled with warmer, more evaporative regimes as are forecast for Australia, Southern Africa, Central America, and southern Europe will increase evaporation, making increased salinization a major risk. The direction of change toward increased leaching or increased evaporation will depend on the extent to which rainfall and temperature change. In either case the situation could lead to important changes in world soils.

Changes in Land-Degradation Processes

Land degradation is already one of the major problems affecting the world. Currently some 6–7 million hectares are lost annually through soil erosion, desertification affects about one-sixth of the world's population and one-quarter of the world's land, and salinization affects some 20 million hectares of irrigated land. Land degradation through damage to the soil is a serious problem and its causes are often complex and interwoven. Severe damage has already been done to the world's soils, and the impact of climate change needs to be considered in parallel with the effect of the existing pressures on the land. It is difficult to separate the effects of these various impacts and their cumulative impact on soils is often greater than a simple summation.

Soil Erosion

Soil erosion is the movement and transport of soil by various agents, particularly water, wind, and mass movement; hence climate is a key factor. It has been recognized as a major problem since the 1930s and, although there has been some 70 years of research into the causes and processes, it is still increasing and of growing concern. Global rates of soil erosion have been exceeding those of new soil formation by 10- and 20-fold on most continents of the world in the last few decades. The increase in soil erosion to date is strongly linked with the clearance of natural vegetation, to enable land to be used for arable agriculture, and the use of farming practices unsuited to the land on which they are practiced. This, combined with climatic variation and extreme weather events, has created ideal conditions for soil erosion. The main climatic factors influencing soil erosion are rainfall (amount, frequency, duration, and intensity), and wind (direction, strength, and frequency of high-intensity winds), coupled with drying-out of the soil. Land use, soil type, and topography are the other key factors.

Soil erosion by water is more widespread and its impact greater than that by wind. Climate change is likely to affect soil erosion by water through its effect on rainfall intensity, soil erodibility, vegetative cover, and patterns of land use. General circulation models predict for many areas seasonally more intense drying, coupled with increased amounts and intensity of precipitation at other times, conditions that could lead to large increases in rates of erosion by water.

Soil erosion also occurs by wind transport of soil particles by suspension, surface creep, or saltation over distances ranging from a few centimeters to hundreds of kilometers. Wind erosion not only transports soil particles around arid and semiarid landscapes but inputs into ecosystems and may even alter global climatic patterns. Wind erosion is particularly a problem on sandy and organic soils, which are subject to intermittent low-moisture contents and periodic winds. Currently wind erosion is mainly a feature of arid and semiarid conditions. Those areas where climate change is predicted to lead to more droughty soils under increasing temperatures will become increasingly vulnerable to wind erosion. Although general circulation models have in the past been unable to predict changes in wind speed and frequency with any certainty, the latest models are predicting increased summer continental drying and risk of drought in midlatitude areas and an increase in tropical cyclone peak intensities in some areas, both sets of conditions favoring an increase in soil erosion by wind. However, it is important to note that erosion is site-specific, and different permutations of conditions can increase or decrease it.

In the last few decades of the twentieth century, significant advances have been made in modeling erosion risk under different climate change scenarios. Research in the USA, using the EPIC model for two different sets of climatic conditions at 100 sites in the US Corn Belt, has shown that mean water erosion varies approximately linearly with mean precipitation, with approximately a 40% change for a 20% change in mean precipitation (Figure 1). By contrast, with a 20% increase in wind speed, erosion increased fourfold. This suggests that wind erosion is potentially more sensitive to climate change than is water erosion and secondly that, for wind-erosion predictions, it is important to understand and predict wind-speed threshold. These results and others using different models suggest that increased wind speed, rainfall amount and intensity and increased frequency of high-wind events are likely to lead to significant increases in soil erosion, thus exacerbating the already serious situation.

Desertification

Desertification is the process of ecological degradation by which economically productive land becomes less productive, in some cases leading to the development of a desert-like landscape. There is a huge literature on the nature and causes of desertification,

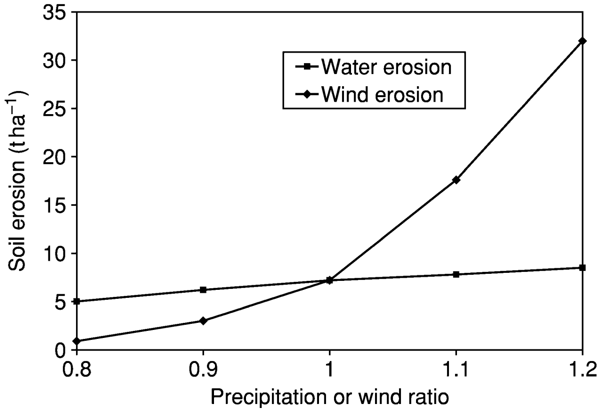


Figure 1 Sensitivity of soil erosion in the US Corn Belt to climate change as estimated using EPIC. For water erosion, temperature, CO_2, and wind speed were held at current values, while precipitation volume (expressed as ratio to current volumes) was varied. For wind erosion, temperature, CO_2, and precipitation volume were held at current values, while wind speed (expressed as ratio to current speeds) was varied. Each point represents the 100-year mean of 100 randomly selected sites. (Reproduced with permission from Ingram J, Lee J, and Valentin C (1996) *Journal of Soil and Water Conservation* 51(5): 378.)

some of which indicates that human impacts arising from overstocking, overcultivation, and deforestation are primarily responsible for the process; some lays the blame on the impact of extended droughts over the last few decades. In reality, desertification is likely to be due to a combination of drought and mismanagement of land, particularly where there is a lack of harmony between land use and management on the one hand and prevailing climate on the other.

Desertification occurs mainly in hyperarid, arid, semiarid, and subhumid climatic zones, ranging from precipitation:potential evapotranspiration indices (P:PET) of less than 0.05 to 0.70. The area of land occupied by these four zones in which true or induced deserts can occur is 47% of the planet. With the increased temperatures and evaporation predicted, this percentage could increase.

Salinization

The accumulation of salts in soils can negatively affect soil properties and processes. It can lead to degradation of soil structure, decline in porosity, and increase in bulk density, and impede nutrient dynamics and nutrient-holding capacity. It leads to a decline in productivity and can cause land to become unsuitable for agricultural production. Soil salinity is already a major global problem in Australia, Africa, Latin America, and the Near and Middle East. It is typically found in areas where evaporation exceeds precipitation. In the USA alone, some 50 million hectares of cropland and pasture is affected by salinity, and the area of land so affected is growing by some 10% annually. Salinization is likely to be prevalent in two situations:

- *Coastal zones*. In these situations salinity is likely to depend primarily on the sea level and its tidal, seasonal, and long-term fluctuations. Given that the global mean sea level is projected to rise by between 9 and 88 cm between 1990 and 2100, there is likely to be a territorial extension of coastal salinity under the direct and indirect effects of saline seawater. Impacts are likely to include flooding of coastal plains by saline seawater, including low-lying coastal fringes, marshlands and swamps, and deltas and estuaries of some of the world's big rivers, increase in storm tides affecting areas several meters in elevation around the coast, with penetration of saline and brackish water inland, and rapid erosion of coastlines.
- *Continental salt transport and salt accumulation processes*. There are three principal mechanisms of this form of salinization: salt accumulation, seepage, and wind deposition. Salinization by salt accumulation happens when leaching is reduced and salt accumulates at or near the surface. It can also occur when salt is leached into a perched water table and the water moves through the landscape to areas of lower elevation, where the water evaporates, leaving the salts. Given suitably exposed deposits of salts, wind erosion can transfer the salts elsewhere in the landscape.

All of these causes of salinity are likely to be enhanced with climate change. Zones subject to rising temperatures during the summer will experience more evapotranspiration and aridity and thus higher concentrations of salt in the soil solution. There may also be enhanced capillary rise from shallow water tables, with the potential to carry salt into the overlying soil horizons. Where climate change leads to increasing wetness as well as temperature, there will be a reduction in the concentration of salts in the salt solution. However, where there is periodic drying also involved, capillary upward transport of salts can take place from shallow groundwater to overlying soil horizons.

Given the predictions for climate change, the balance will be for the area of salt-affected soils to increase, although there may be an improvement to a few of the existing saline areas of the world, albeit a very slow one.

Climate Change Impacts on Soil Functions

The most recognized function of soils is their use for agriculture and the implications this has for feeding the global population. Soil properties, as well as

Table 3 Soil functions and climatic change

Function	*Use of soils*	*Impacts of climate change*
Economic	Food crops, energy crops, timber Sand, gravel, minerals Foundations for buildings, roads, etc.	Changes in land-use capability Changes of productivity Erosion and salinization Nitrate leaching and increased use of fertilizers (wetter conditions) Need for increased irrigation (drier conditions) Changes to foundations of buildings Higher building insurance premiums
Ecologic	Habitats for soil fauna and microflora Food for ground feeders, e.g., birds Nutrient supply and storage Cycling of water and air	Loss of some habitats Stress on many habitats Changes in soil biodiversity Likely loss of soil organic matter Changes in soil fertility Acidification (wetter conditions) Loss of peat habitats (drier conditions)
Hydrologic	Water storage, flow control and runoff, absorption, amelioration	Water resources problems (drier conditions) Transfer of salts Transfer of nutrients Increased bypass flow (drier conditions) Erosion sediment transfer
Pollutant control	Source and sink for pollutants Waste disposal medium	Increased bypass flow (drier conditions) Erosion and movement of pollutants Changes between sources and sinks
Gaseous exchanges	Source and sink of greenhouse gases	Increased decomposition of organic matter (CO_2 release) Sequestration of carbon Increased/decreased N_2O release (wetter/drier conditions) Increased/decreased CH_4 release (wetter/drier conditions)
Heritage protection	Protection of buried archeological sites	Effects of erosion and sedimentation Loss of peat Cracking

climate, are the main factors governing which crops can be grown, where they can be grown, and their productivity. Changes in the physical, chemical, and biological properties of soil have major implications for agriculture. Changes in soil and air temperature and in rainfall will affect the ability of crops to reach maturity and their potential harvest. Reductions in amounts of soil water may be compensated by irrigation, but scarcity of water may preclude the use of water for irrigation. The increases in land degradation, whether in the form of soil erosion, desertification, salinization, or loss of peat soils, will further impact on the capability of soils to supply the needs of agriculture. If the current predictions of climate change are borne out there is likely to be a shift in agricultural zones, requiring a movement of population, new farming methods, development of new skills, or some combination of these. The rate of change may be sufficiently slow for readjustment to take place in line with the pace of climate change. The main problem areas are likely to be those experiencing increasing aridity where the options for successful change are limited.

There are, however, several other soil functions (Table 3) that have a major influence on the quality of life. Soils are fundamental to land-based ecosystems and, as with agriculture, in association with climate, govern the nature and distribution of the world's natural and seminatural ecosystems, providing water, nutrients, and a growing medium. Although most ecosystems are naturally variable with time and adapt reasonably well to small climatic fluctuations, changes to soils as a result of climate change are likely to be reflected in a change in ecosystems. The form the change will take depends much on the degree of warming and the changes to precipitation, and prediction of the changes will require significantly improved climate change models at regional level.

As discussed earlier, soil is the habitat for an immense number and variety of organisms, important in the decomposition of soil organic matter and the recycling of nutrients on which plant productivity and ultimately life depend.

Soil properties are important in controlling the fate and behavior of water once it reaches the surface of the soil. The soil system influences runoff, flow, storage, and regional distribution of water. There are likely to be major changes to the distribution of water under a changed climate, and the soil will be a central issue in managing the water regime under these new conditions. The soil also plays an important part as a medium supporting buildings and other infrastructure, and protecting our archeological heritage. Current infrastructure will almost certainly need to change to meet the new climatic conditions. Finally, climate change concerns associated with the levels of greenhouse gases in the atmosphere have focused on the large amounts of carbon stored in the soils, and the ability of the soil to act as a source and sink for carbon. The fact that the soil can act as both a source and sink for several greenhouse gases, e.g., CO_2, N_2O, and CH_4, has already become an important issue in the quest to reduce the impacts of climate change.

See also: **Carbon Emissions and Sequestration**; **Cold-Region Soils**; **Desertification**; **Organic Matter:** Principles and Processes; **Organic Soils**; **Structure**

Further Reading

Houghton JT, Ding Y, Griggs DJ *et al.* (eds) (2001) *Climate Change 2001: The Scientific Basis*. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press.

IPCC (1996) *Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scientific–Technical Analyses*. Cambridge, UK: Cambridge University Press.

McCarthy JJ, Canziani OF, Leary NA, Dokken DJ, and White KS (eds) (2001) *Climate Change 2001: Impacts, Adaptation and Vulnerability.* Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press.

Metz B, Davidson O, Swart R, and Pan J (eds) (2001) *Climate Change 2001: Mitigation*. Contribution of Working Group III to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press.

Parry ML (2000) *Assessment of Potential Effects and Adaptations for Climate Change in Europe: The Acacia Project*. Jackson Environmental Institute. Norwich, UK: University of East Anglia Press.

Rosenzweig C and Hillel D (1998) *Climate Change and the Global Harvest. Potential Impacts of the Greenhouse Effect on Agriculture*. Oxford, UK: Oxford University Press.

Rounsevell MDA and Loveland PJ (1994) *Soil Responses to Climate Change*. NATO ASI Series. Berlin, Germany: Springer-Verlag.

Rounsevell MDA, Bullock P, and Harris JA (1996) Climate change, soils and sustainability. In: Taylor AG, Gordon JE, and Usher MB (eds) *Soils, Sustainability and the Natural Heritage*. Edinburgh, UK: HMSO.

Rounsevell MDA, Evans SP, and Bullock P (1999) Climate change and agricultural soils: impacts and adaptations. *Climatic Change* 43: 683–709.

Scharpenseel HW, Schomaker M, and Ayoub A (eds) (1990) Soils on a warmer earth. *Developments in Soil Science* 20.

CLIMATE MODELS, ROLE OF SOIL

P Smith, University of Aberdeen, Aberdeen, UK

Introduction

Models of climate change usually focus on fluxes of carbon between land (vegetation and soil) and the atmosphere, because carbon dioxide is the most prevalent greenhouse gas. Consequently, the soil components of these models focus on soil organic carbon (SOC) dynamics. Though methane and nitrous oxide are also important biogenic greenhouse gases, and some models include these gases, it is models of SOC dynamics that are described here. Ecosystem models are increasingly used to predict the impacts of climate change on soils and other components of ecosystems, and to investigate the feedbacks between climate change and soil or other ecosystem processes. Most of the modeling of climate change impacts has occurred independently of the modeling of climate change itself, but recently climate models have included fully coupled biospheric carbon models to allow the feedback between biospheric carbon and atmospheric processes to be explored.

Approaches to Modeling SOC Dynamics for the Study of Climate Change

There are a number of approaches to modeling the turnover of SOC, including: (1) process-based, multicompartment models – this approach is by far the most common; (2) models that consider each fresh addition of plant debris as a separate cohort which decays in a continuous way; and (3) models that account for C and N transfers through various trophic levels in a soil food web, often termed 'food-web models' or 'organism-oriented models.'

Process-Based, Multicompartment Models of SOC Dynamics

These models focus on the processes mediating the movement and transformations of matter or energy and usually assume first-order rate kinetics. Early models treated SOC as one homogeneous compartment having a single, first-order rate constant. Later two-compartment models were proposed and, as computers became more accessible, multicompartment models were developed. Of the 33 SOC models currently represented within the Global Change and Terrestrial Ecosystems (GCTE)–Soil Organic Matter Network (SOMNET) database, 30 are multicompartment, process-based models. Each compartment or SOC pool within a model is characterized by its position in the model's structure and its decay rate. Decay rates are usually expressed by first-order kinetics with respect to the concentration (C) of the pool:

$$\mathrm{d}C/\mathrm{d}t = -kC \qquad [1]$$

where t is time. The rate constant (k) of first-order kinetics is related to the time required to reduce by half the concentration of the pool when there is no input. The pool's half-life ($h = (\ln 2)/k$), or its turnover time ($\tau = 1/k$) are sometimes used instead of k to characterize a pool's dynamics: the lower the decay rate constant, the higher the half-life, the turnover time, and the stability of the organic pool.

The flows of C within most models represent a sequence of C moving from plant and animal debris to the microbial biomass, then to soil organic pools of increasing stability. Figure 1 shows the flow of C in the Rothamsted C model.

Some models also use feedback loops to account for links between decomposition (catabolic processes) and the synthesis of new polymers from the breakdown products (anabolic processes) mediated by different groups of microbes. The output flow from any organic pool is usually split, some carbon being directed to a microbial biomass pool, some to other organic pools, and, under aerobic conditions, some to CO_2. This split simulates the simultaneous anabolic and catabolic activities and growth of a microbial population feeding on one substrate. Two parameters are required to quantify the split flow. They are often defined by microbial (utilization) efficiency and stabilization (humification) factors, which control the flow of decayed C to the biomass and humus pools, respectively. The sum of the efficiency and humification factors must be less than 1 to account for the release of CO_2. Unlike food-web models, multicompartment, process-based models generally ignore the separate roles of different organisms involved in organic matter turnover in soil. Instead, they concentrate on the overall processes and quantities of C being transformed.

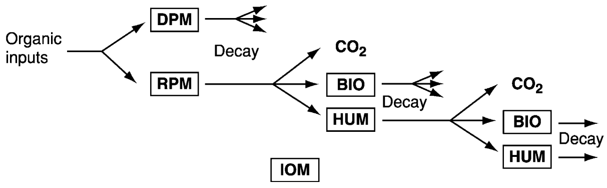


Figure 1 Flows of carbon in the Rothamsted carbon model. BIO, microbial biomass; DPM, decomposable plant material; IOM, inert organic matter; HUM, humified OM; RPM, resistant plant material.

Cohort Models Describing Decomposition as a Continuum

Another approach to modeling SOC turnover is to treat each fresh addition of plant debris into the soil as a cohort. Such models consider one SOC pool that decays with a feedback loop into itself. Some models are represented by a single rate equation with the organic C pool divided into an infinite number of components, each characterized by its 'quality' with respect to degradability, as well as impact on the physiology of the decomposers. The rate equations for such models represent the dynamics of each organic C component of quality q, which is quality-dependent. Exact solutions to the rate equations are obtained analytically.

Food-Web Models

Another type of model simulates C and N transfers through a food web of soil organisms; such models explicitly account for different trophic levels or functional groups of biota in the soil. Some models have been developed which combine an explicit description of the soil biota with a process-based approach. Food-web models require a detailed knowledge of the biology of the system to be simulated and are usually parameterized for application at specific sites.

Factors Affecting the Turnover of Soil Organic Carbon in Models

Rate 'constants' (k), used in all models, are constant for a given set of biotic and abiotic conditions. To adjust a rate constant for nonoptimum environmental circumstances, the simplest way is to modify the maximum value of k using a reduction factor μ, ranging from 0 to 1. Environmental factors considered by SOC models include temperature, water, pH, nitrogen, oxygen, clay content, cation exchange capacity, type of crop/plant cover, and tillage.

Many studies show the effect of temperature on microbially mediated transformations in soil, either expressed as a reduction factor or the Arrhenius equation. Water and oxygen have a major impact on microbial physiology. While some models simulate O_2 concentrations in soil explicitly, many define the extent of anaerobiosis based on soil pore space filled with water. Soil clay content and total SOC are usually correlated: in general, a soil with a high clay content will contain more organic matter than a soil with a lower clay content if management and climate are similar. This is because clay surfaces have a major role in the stabilization of organic matter through absorption on surfaces. Various schemes simulate the effect of clay on rate equations to obtain SOC accumulation. N is an essential element for microbial growth which will be maximal when enough N is assimilated to maintain the microbial C:N ratio. Table 1 presents an overview of the 33 models represented in the GCTE-SOMNET, including the factors affecting SOC turnover.

Model Performance

There are many reasons for evaluating the performance of an SOC model. Model evaluation shows how well a model can be expected to perform in a given situation. This is important if, for example, a scientist, farmer, or policy-maker wishes to use a model to predict the impact of climate change or a change in land management on SOC content. A model can only be used with confidence in this predictive way if it has been tested previously against real data. In a scientific sense, it is also instructive to evaluate a model against data, because this can help to improve understanding of the system – especially when the model fails, as this may highlight a process, or a factor controlling a rate, that is not properly simulated in the model. Models can be evaluated at a number of different levels. They can be evaluated at the individual process level or at the level of a subset of processes (e.g., net mineralization), or the models' overall outputs (e.g., changes in total SOC over time) can be tested against measured laboratory and field data. Models can also be evaluated for their applicability in different situations, e.g., for scaling-up simulated net C storage from a site specific to a regional level.

Figure 2 shows an example of the simulation of two commonly used SOC models using the same data set. In both cases, the models can be compared with measured data graphically and statistically to indicate the overall error and any evidence of systematic bias.

Models Used to Study the Impacts of Climate Change

Many ecosystem models have been used to examine the potential impact of climate change on ecosystem components. Many of these models include some description of the soil, though the detail with which soil processes are described tends to decrease with the scale of application. Site-specific applications of ecosystem models allow detailed mechanistic descriptions of soil processes to be described, whereas models applied at continental, biome, or global scales (e.g., the Vegetation/Ecosystem Modelling and Analysis Project (VEMAP) exercise) tend to have much more simplistic descriptions of the soil. Models applied at the global scale often have a very simple description of the soil. Many examples exist of the use of models to examine the impacts of climate change, but, in most of these studies, climate change is used to drive the models, with no explicit feedback between the biosphere and the climate system. In the next section, some recent work in which climate models have been linked to biospheric C cycle models to allow the feedback between climate and biosphere to be examined.

Climate Models with Coupled Description of Biospheric Carbon Feedbacks

Until recently, climate (general circulation) models (GCMs) had no soil C component at all, relying instead upon a purely physical description of the Earth's climate system. Recently, however, climate models with a fully coupled biospheric C module (including soil C) have been developed. The group of P. Cox at the UK Hadley Centre and the group of P. Freidlingstein of the Laboratoire des Sciences de l'Environnement, Paris, have independently developed climate models that include biospheric C feedback.

Though the models are very different, both suggest (to different degrees) that the feedback between climate and biospheric carbon leads to accelerated loss of C from the biosphere, which in turn leads to accelerated climate change. One of the key processes leading to this acceleration is the response of soil respiration to increased temperature. The findings from studies using coupled biospheric C and climate models are subject to many uncertainties, but do highlight the need for including soils in assessments of future climate. Future developments require that the process level understanding developed at the site level, using mechanistic process-based models be incorporated in these coupled GCMs so that the full impact of the feedback of soil C change on climate can be quantified.

Table 1 Overview of soil organic matter models represented within the Global Change and Terrestrial Ecosystems–Soil Organic Matter Network (GCTE-SOMNET) database in August 2002

	Inputs					
Model	*Time step*	*Meteorology*[a]	*Soil and plant*[b]	*Management*[c]	*Factors affecting decay rate constants*[d]	*Soil outputs*[e]
ANIMO	Day, week, month	P, AT, Ir, EvW	Des, Lay, Imp, Cl, OM, N[f], pH	Rot, Ti, Fert, Man, Res, Irr, AtN	T, W, pH, N, O	C, N, W, ST, gas
APSIM	Day	P, AT, Ir	Lay, W, C, N, BD, Wi, PG, PS	Rot, Ti, Fert, Irr	T, W, pH, N	C, N, W, ST, gas
Candy	Day	P, AT, Ir	D, Imp, W, N, C, Wi, PD, Nup	Rot, Ti, Fert, Man, Res, Irr, AtN	T, W, N, Cl	C, N, W, ST, gas
CENTURY	Month	P, AT	W, Cl, OM, pH, C, N	Rot, Ti, Fert, Man, Res, Irr, AtN	T, W, N, Cl, pH, Ti	C, BioC, 13C, 14C, N, W, ST, gas
Chenfang Lin Model	Day	ST	OM, BD, W	Man, Res	T, W, F	C, BioC, gas
DAISY	Hour, day	P, AT, Ir, EvG	Lay, Cl, C, N, PG, PS	Rot, Ti, Fert, Man, Res, Irr, AtN	T, W, N, Cl	C, BioC, N, W, ST, gas
DNDC	Hour, day, month	P, AT	Lay, Cl, OM, pH, BD	Rot, Ti, Fert, Man, Res, Irr, AtN	T, W, N, Cl, Ti	C, BioC, N, W, ST, gas
DSSAT	Hour, day, month, year	P, AT, Ir	Des, Lay, Imp, W, Cl, PS, OM, pH, C, N	Rot, Ti, Fert, Man, Res, Irr	T, W, N, Cl, Ti	C, BioC, N, W, ST
D3R	Day	P, AT	Y, PS	Rot, Ti, Res	T, W, N, Cv, Ti	Decomposition of surface and buried residue
Ecosys	Minute, hour	P, AT, Ir, WS, RH	Lay, W, Cl, CEC, PS, OM, pH, N, BD, PG, PS	Rot, Ti, Fert, Man, Res, Irr, AtN	T, W, N, O, Cl, Cv	C, BioC, N, W, ST, pH, Ph, EC, gas, ExCat
EPIC	Day	P, AT	Lay, Imp, W, Cl, OM, pH, C, BD, Wi	Rot, Ti, Fert, Man, Res, Irr, AtN	T, W, N, pH, Cl, Ce, Cv	C, BioC, N, W, ST
FERT	Day	P, AT, WS	Des, Lay, W, Cl, OM, pH, C, N, BD, W, Ph, K, Nup, Y, PS	Rot, Ti, Fert, Man, Res, Irr	T, W, N, pH, Cv	C, N, Ph, K
ForClim-D	Year	P, AT	W, AG	None	T, W	C
GENDEC	Day, month	ST, W	W, InertC, LQ	Can be used; not essential	T, W, N	C, BioC, N, gas, LQ
HPM/EFM	Day	P, AT, Ir, WS	W, Cl, PS	Rot, Fert, Irr, AtN	T, W, N	C, BioC, N, W, gas
ICBM	Day, year	Combination of weather and climate	Many desirable; none essential	C inputs to soil	T, W, Cl	C
KLIMAT-SOIL-YIELD	Day, year	P, AT, ST, Ir, EvG, EvS, VPD, SH	Des, Lay, Imp, W, Cl, PS, OM, pH, C, N	Fert, Man, Res, Irr	T, W, N, Cl	C, BioC, N, W, ST
CNSP Pasture Model	Day	P, AT, Ir	Lay, Imp, W, Cl, CEC, OM, pH, C, N, PS, AS	Fert	T, W, N, pH	C, N, W, ST
Humus Balance	Year	Climate based on P and AT	Des, Lay, PS, OM, pH, C, N	Rot, Fert, Man	N, H, Cl, Cv	C, N
MOTOR	User-specified	P, AT, EvG	Des, OM	Rot, Ti, Fert, Man	T, W, N, Cl, Ti	C, BioC, 13C, 14C, gas
NAM SOM	Year	P, AT	Des, PS, OM, Ero	Man, Res	T, W, Cl, Cv	C, BioC
NCSOIL	Day	ST (P, AT)	W, OM, C, N	Fert, Man, Res	T, W, N, pH, Cl, Ti	C, BioC, 14C, N, 15N, gas

(*Continued*)

Table 1 (*Continued*)

Model	*Inputs*				*Factors affecting decay rate constants*[d]	*Soil outputs*[e]
	Time step	*Meteorology*[a]	*Soil and plant*[b]	*Management*[c]		
NICCE	Hour, day	P, AT, Ir, WS	Imp, OM, C, N, W, TC, PG	Fert, Man, Res, Irr, AtN	T, W, Cl, N	C, BioC, 13C, 14C, N, 15N, W, ST, gas
O'Brien Model	Year	None	Lay, C, 14C	None	None	C, 14C
O'Leary Model	Day	P, AT	Lay, W, Cl, pH, N	Ti, Fert, Res	T, W, N, Cl, Ti	C, BioC, N, W, ST, gas, ResC, ResN
Q-Soil	Year	Optional	C, N	Rot, Fert, Man, Res, AtN	T, W, N	C, BioC, 13C, N
RothC	Month	P, AT, EvW	Cl, C, InertC (can be estimated)	Man, Res, Irr	T, W, Cl, Cv	C, BioC, gas, 14C
SOCRATES	Week	P, AT	CEC, Y	Rot, Fert, Res	T, W, N, Cv, Ce	C, BioC, gas
SOMM	Day	P, ST	OM, N, AshL, NL	Man	T, W, N	C, N, gas
Sundial	Week	P, AT, EvG	Imp, Cl, W, Y	Rot, Fert, Man, Res, Irr, AtN	T, W, N, Cl	C, BioC, N, 15N, W, gas
Verberne	Day	P, AT, Ir, WS, EvS	Des, W, Cl, PS, OM, C, N	Man, AtN	T, W, N, Cl	C, BioC, N, W
VOYONS	Day, week, month	P, ST	Cl, OM, C, N	Fert, Man, Res, Irr, AtN	T, W, Cl	C, BioC, 13C, 14C, N, gas
Wave	Day	P, AT, Ir, EvG	Lay, OM, C, N, W, PG	Rot, Ti, Fert, Man, Res, Irr, AtN	T, W, N	C, N, W, ST, gas

[a] P, precipitation; AT, air temperature; ST, soil temperature; Ir, irradiation; EvW, evaporation over water; EvG, evaporation over grass; EvS, evaporation over bare soil; WS, wind speed; RH, relative humidity; VPD, vapor pressure deficit; SH, sun hours.

[b] Des, soil description; Lay, soil layers; Imp, depth of impermeable layer; Cl, clay content; OM, organic matter content; N, soil nitrogen content/dynamics; C, soil carbon content/dynamics; InertC, soil inert carbon content; pH, pH; W, soil-water characteristics; Wi, wilting point; PD, soil particle-size distribution; CEC, cation exchange capacity; Ero, annual erosion losses; BD, soil bulk density; TC, thermal conductivity; PG, plant growth characteristics; PS, plant species composition; AS, animal species present; AG, animal growth characteristics; Y, yield; Nup, plant nitrogen uptake; LQ, litter quality; AshL, ash content of litter; NL, N content of litter.

[c] Rot, rotation; Ti, tillage practice; Fert, inorganic fertilizer applications; Man, organic manure applications; Res, residue management; Irr, irrigation; AtN, atmospheric nitrogen inputs.

[d] T, temperature; W, water; pH, pH; N, nitrogen; O, oxygen; Cl, clay; Ce, cation exchange capacity; Cv, cover crop; Ti, tillage; F, Fauna.

[e] BioC, biomass carbon; 13C, ^{13}C dynamics; 14C, ^{14}C dynamics; 15N, ^{15}N dynamics; gas, gaseous lossess (e.g., CO_2, N_2O, N_2); ResC, surface residue carbon; ResN, surface residue nitrogen; Ph, phosphorus dynamics; K, potassium dynamics; EC, electrical conductivity; ExCat, exchangeable cations.

[f] N in the soil inputs and outputs section is used to denote all aspects of the N cycle. (*See* **Civilization, Role of Soils**.)

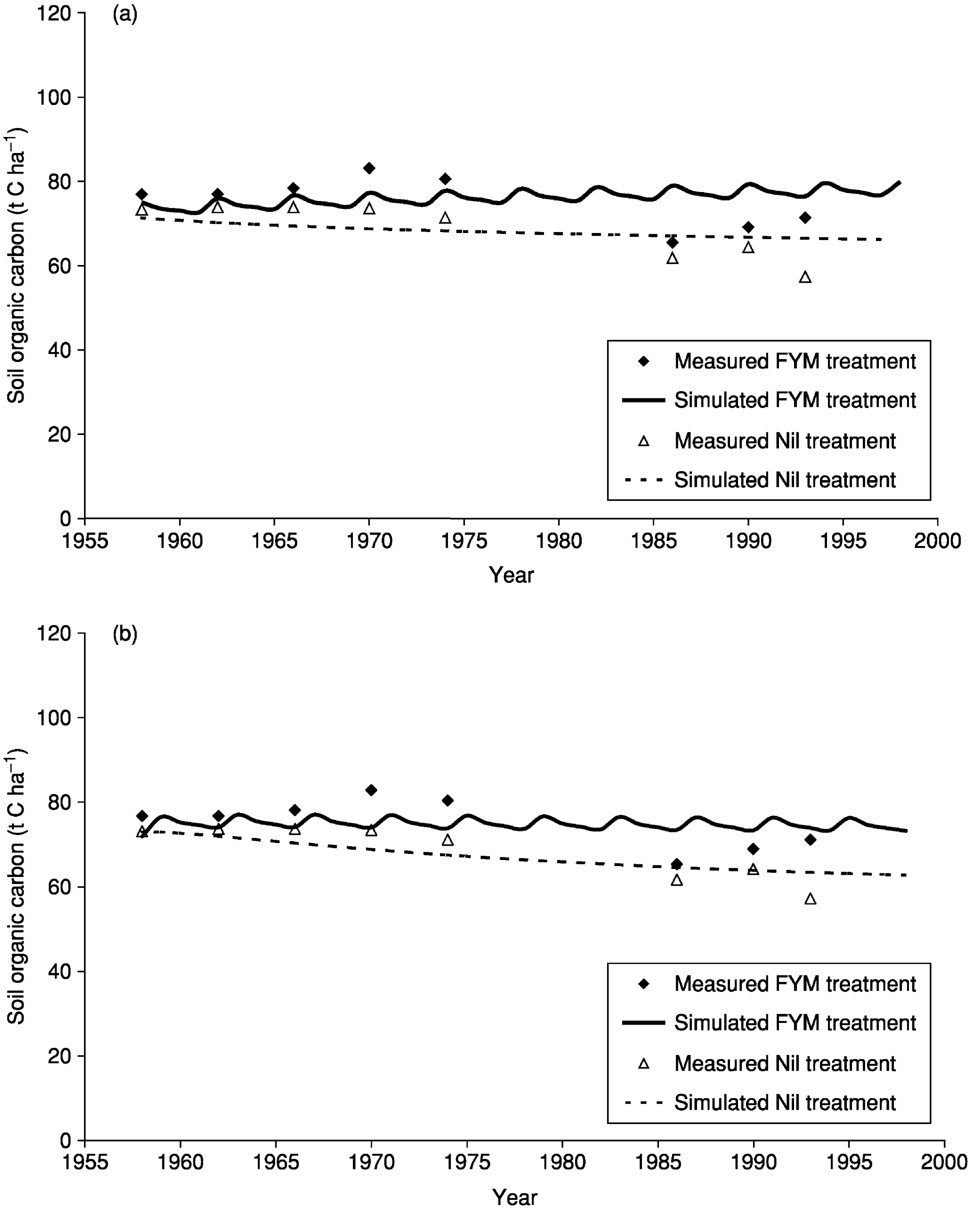


Figure 2 Simulated (lines) and measured (symbols) values of soil organic carbon for two treatments (farmyard manure (FYM) and nil inputs), at the Martonvasar long-term cropping experiment in Hungary, as simulated by two models: Century (a) and RothC (b). Root-mean-square error (RMSE) values for each simulation are Century: FYM = 7.8, Century: nil = 5.3, RothC: FYM = 8.0 and RothC: nil = 7.3.

List of Technical Nomenclature

GCM	general circulation model
GCTE	Global Change and Terrestrial Ecosystems
k	rate constant
RMSE	root-mean-square error
SOC	soil organic carbon
SOMNET	Soil Organic Matter Network

See also: **Carbon Cycle in Soils:** Dynamics and Management; **Carbon Emissions and Sequestration**; **Civilization, Role of Soils**; **Climate Change Impacts**; **Fertility**; **Organic Matter:** Principles and Processes; **Soil–Plant–Atmosphere Continuum**; **Sustainable Soil and Land Management**

Further Reading

Bosatta E and Ågren GI (1995) Theoretical analyses of the interactions between inorganic nitrogen and soil organic matter. *European Journal of Soil Science* 76: 109–114.

Cox PM, Betts RA, Jones CD, Spall SA, and Totterdell IJ (2000) Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* 408: 184–187.

Falloon P and Smith P (2002) Simulating SOC dynamics in long-term experiments with RothC and Century: model evaluation for a regional scale application. *Soil Use and Management* 18: 101–111.

Friedlingstein P, Bopp L, Ciais P *et al.* (2001) Positive feedback between future climate change and the carbon cycle. *Geophysical Research Letters* 28: 1543–1546.

Hunt HW, Trlica MJ, Redente EF *et al.* (1991) Simulation model for the effects of climate change on temperate grassland ecosystems. *Ecological Modelling* 53: 205–246.

McGill WB (1996) Review and classification of ten soil organic matter (SOM) models. In: Powlson DS, Smith P, and Smith JU (eds) *Evaluation of Soil Organic Matter Models Using Existing, Long-Term Datasets*, pp. 111–133. NATO-ASI 138. Berlin, Germany: Springer-Verlag.

Melillo JM, Kicklighter DW, McGuire AD, Peterjon WT, and Newkirk KM (1995) Global change and its effect on soil organic carbon stocks. In: Zepp RG and Sonntag CH (eds) *Role of Nonliving Organic Matter in the Earth's Carbon Cycle*, pp. 175–189. New York: John Wiley.

Molina JAE and Smith P (1998) Modeling carbon and nitrogen processes in soils. *Advances in Agronomy* 62: 253–298.

Powlson DS, Smith P, and Smith JU (eds) (1996) *Evaluation of Soil Organic Matter Models Using Existing, Long-Term Datasets*. NATO ASI 138. Berlin, Germany: Springer-Verlag.

Smith P, Powlson DS, Smith JU, and Elliott ET (eds) (1997) Evaluation and comparison of soil organic matter models. *Geoderma* 81: 1–225.

Smith P, Andrén O, Brussaard L *et al.* (1998) Soil biota and global change at the ecosystem level: describing soil biota in mathematical models. *Global Change Biology* 4: 773–784.

Smith P, Falloon P, Powlson DS, and Smith JU (2001) *Soil Organic Matter Network (SOMNET): 2001 Model and Experimental Metadata*, 2nd edn. GCTE Report No. 7. Wallingford, UK: GCTE Focus 3 Office.

VEMAP (1995) Vegetation/Ecosystem Modelling and Analysis Project: comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial responses to climate change and CO_2 doubling. *Global Biogeochemical Cycles* 9: 407–437.

COLD-REGION SOILS

C-L Ping, University of Alaska Fairbanks, Palmer, AK, USA

Introduction

Soils of cold regions are those soils formed in areas with a mean soil temperature of generally less than 8°C and having soil formation controlled by cryogenesis (or cryogenic processes). Cryogenic processes involve the formation of ice and subsequent frost heave caused by the increased volume. These soils are distributed throughout high latitudes or areas with high altitudes. The soil temperature strongly influences geochemical and biological processes within the pedosphere. Soil organic matter accumulation is one of the most important features in soils of a cold region. The rate of organic matter accumulation increases with latitudes and altitude, and maximizes at approximately 50–65° N. Above latitude 65°, the rate of accumulation decreases more as a result of lower inputs, but at the same time the rate of decomposition decreases with the increasing latitude. The cold temperatures in high latitudes and altitudes cause freeze–thaw cycles that result in unique cryogenic fabrics and cryoturbated soil horizons. Many of these soils have permafrost that dominates the soil processes in the tundra zone and the northern boreal zone. The polar zone is less affected because it is too dry.

Definition of Cold Soils

Cold soils generally refer to the soils formed in high latitudes and at high altitudes, and they are identified based on vegetation belts by ecologists and on soil temperature by soil scientists. The northernmost region is the polar desert in the High Arctic, where only sparse cushion plants and lichens grow. In these regions, the climate is severe, with mean annual precipitation (MAP) less than 200 mm and air temperatures in the warmest summer month only about 4°C, and mean annual soil temperatures (MAST) at 50 cm being less than −8°C. The region below the High Arctic is the tundra zone where the ericaceous plants and mosses dominate the upland tundra; moss and sedges dominate the coastal plains. In the tundra zone, the annual precipitation ranges from 200 to 400 mm and the warmest summer air temperatures are less than 10°C, with MAST ranging from −2 to −8°C. The region below the tundra zone is the boreal zone, where the summer air temperature may reach 20°C, the MAST is less than 8°C, and the annual precipitation is more than 200 mm. The vegetation in the boreal zone is dominantly boreal forest, or taiga. Generally, the polar and the arctic zones correspond with the zone of continuous permafrost and the boreal zone corresponds to the zone of discontinuous permafrost. Permafrost is defined as a thickness of soil or other superficial deposit, or even of bedrock, that remains frozen for a consecutive 2 years or more.

Characteristics and Genesis of Cold-Region Soils

What separates the cold-region soils from those of the temperature and tropical regions is their cryogenic nature and the unique properties resulting from freeze–thaw cycles and the formation of ground ice. Due to ice-lens formation during the freeze cycle, a granular structure usually forms in the A horizons of soils associated with earth hummocks, and lenticular and reticulate structures form in subsoils (Figure 1). Commonly an ice-rich layer occurs just above the permafrost table. Thirdly, the presence of permafrost has a profound impact on soil morphology due to its impact on many of the normal soil-forming factors and the fact that it leads to cryoturbation. Cryoturbation is a very common phenomenon not only in permafrost-affected areas, but also in the alpine zone, where the soils are subjected to strong freeze–thaw cycles but lack permafrost. Cryoturbation results in warped soil horizons, mixing of horizons and, most importantly, frost-churning of surface organic matter downward to the upper permafrost and thus sequesters carbon (Figure 2).

Physical Properties

Boreal soils Most of the soils in the northern boreal zone have permafrost, but all soils in this zone are subjected to seasonal freeze–thaw cycle. Those without permafrost are defined as soils with cryic soil temperature regimes in *Soil Taxonomy*. These soils have MAST of less than 8°C with a cool summer. Most of these soils have vegetation dominated by conifer forest or boreal forest. When soil freezes, soil water moves to the freezing front, thus forming a thin layer of ice called an ice lens. The progressive freezing creates a stratified fabric that consists of alternate layers of soil and ice lens. After repeated freeze–thaw cycles, the finer soil particles become orientated along the plate surface. Thus a distinctive, thin platy structure forms in the early stage of this process. With time, the expansion of ice deforms the flat plates into discontinuous, small curved plates referred to as 'lenticular' fabrics by geocryologists (Figure 3). It is also common in the seasonal-frozen as well as some of the permafrost soils; a crusty, or crumb-like structure forms on the surface of the soils. This is caused by needle-ice formation during early winter when water is extracted from belowground and freezes on the soil surface to form needle ice below the very top, thin layer of the soils.

Arctic soils In areas strongly affected by permafrost, the soils have more strongly expressed cryogenic structures. Generally the lenticular structures form in mineral soil horizons in the upper and lower sections of the active layer, directly above the permafrost. The active layer is the portion of the soil profile above the permafrost that is subjected to seasonal freeze and thaw. Soils in the middle section of the active layer have weaker and thicker lenticular or coarse blocky structures. The reason for this is that during freezing the water not only moves to the top due to a descending freezing front, but also moves to the bottom, where the permafrost serves as another freezing front. On top of the permafrost table, there is a zone that has horizontal and vertical cracks, like an ice net (Figure 1). This is caused by the freeze cracks during the freeze-up in early winter. An ice-rich layer is formed by repeated freezing and thawing, and during each thawing more water enters the cracks and later becomes frozen. Thus, in the upper part of this zone, an angular, blocky structure forms and, in the lower part of this zone, the ice-rich layer often contains more than 70% ice. The mineral soil blocks seem to be 'floating' in the ice matrix. When the permafrost table drops and the water drains, a rectangular-shaped lattice forms reticulate structures (Figure 4).

Alpine and high arctic soils Some of the cold-region soils are in the alpine region, where the MAST may be the same as those in the boreal and the arctic regions, but they experience such strong diurnal temperature fluctuations that freeze–thaw cycles are common even in the midst of the summer. In the alpine and the High Arctic regions, rock fragments not only are subject to physical weathering due to a strong freeze–thaw cycle, but also grind against each other during the process, thus producing finer particles, especially silt. This process is referred to as 'underground glaciation' by Russian scientists. Thus many rock fragments in the soil profile are silt-capped.

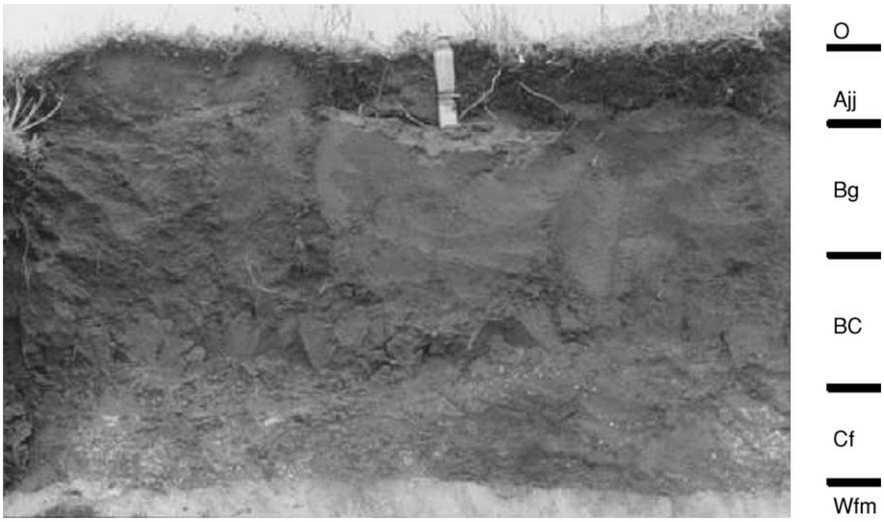


Figure 1 Tundra soil formed in Yedoma deposit near the Arctic Ocean coast, NE Russia. Note the thin organic horizon (O), cryoturbated A horizon (Ajj), gleyed mineral horizon (Bg), ice-net formation (BC), reticulate structure, ice-rich layer (Cf), and ice wedge (Wfm).

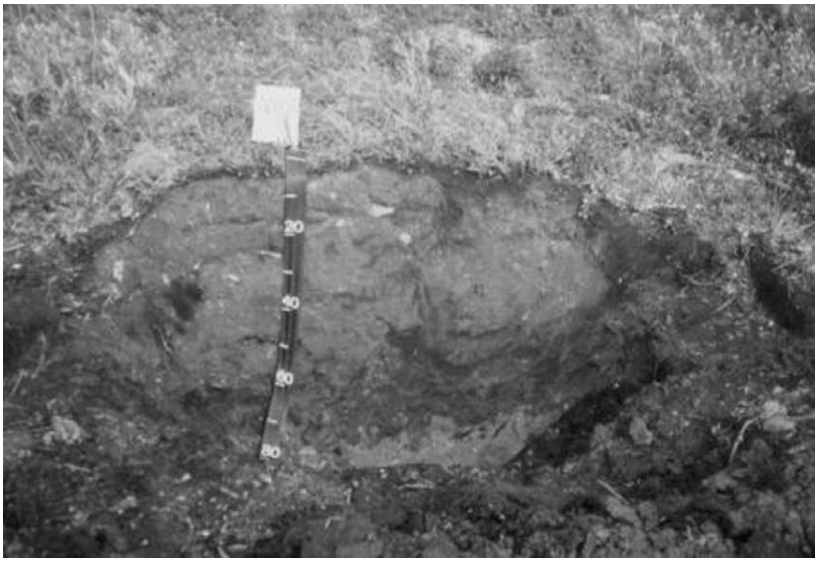


Figure 2 Cryoturbated soil formed in earth hummocks, Northwest Territory, Canada. Note the bare soils at the center of the hummock, thick organic horizon in the trough between the hummocks, and the frost-churned organic matter at 50–70 cm.

Morphological Properties

In soils affected by seasonal frost with or without permafrost, the top horizons usually contain more ice then the underlying horizons, owing to the desiccation process described earlier. During thawing, if the soil drainage is impeded, the topsoil remains saturated for longer then the underlying zone, thus often a reduced or a gleyed layer forms near the top of the profile. However, the permafrost plays an even bigger role in soil morphology. During the growing season, the permafrost table acts as a barrier not only to water movement but also to roots, thus creating a reduced zone, because the water table is positioned on top of the permafrost. Thus the lower active layer is often gleyed (Figures 1 and 3). In areas affected by permafrost, the ground cracks during the winter. The intervening cracks create a unique land pattern called 'polygons.' Each year the melting water enters into the cracks and then freezes in the winter. The ice vein eventually thickens and becomes an ice wedge. The formation of ice wedges increases the internal volume of the soils, and the soils buckle up at the edge of the

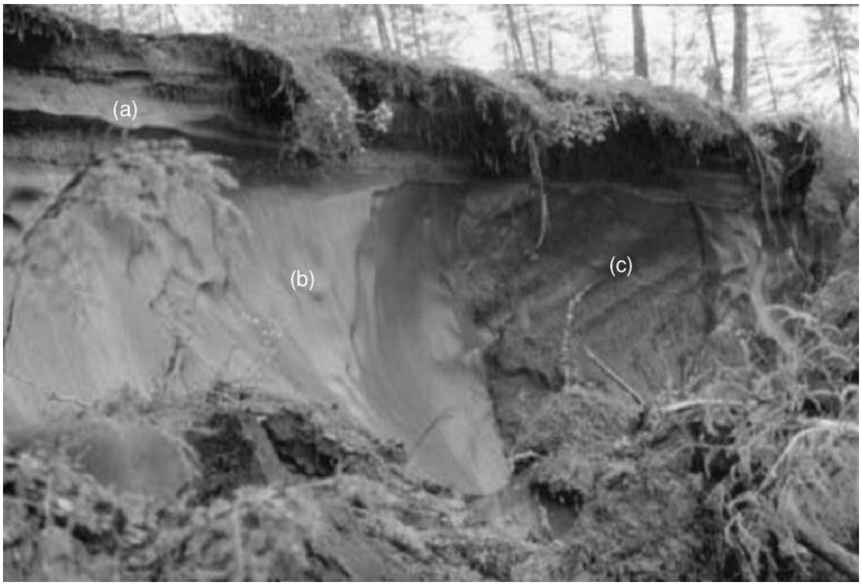


Figure 3 Riverbank erosion caused by thawing of ice wedges under boreal forest, Duvany Yar, Lower Kolyma region, NE Russia. The surface horizontal layers belong to the active layer (a), ice wedge at left (b), and distorted frozen soil at right (c).

polygon. Eventually, the center of the polygon raises and forms an earth hummock (Figure 2). The center of the hummock becomes better drained than the surrounding low areas, where it is poorly drained, and thus contains more water, which drives the process even more. More organic matter accumulates and portions of the organic matter get squeezed between two hummocks and eventually are 'frost-churned' into the lower part and mixed with the underlying mineral horizons (Figure 2). In the process, the horizontal soil horizons become warped and distorted by the process called cryoturbation.

Another phenomenon unique to the soils with cryogenic processes is the frost boil. Frost boils appear on the ground surface as areas of the mineral soil material 'boiled' from underneath the surface, forming circles. The vegetation is pushed to the edge of the frost boil. Although there are many theories as to how frost boils form, field morphology suggests that it is due to diaparism; that soils have different viscosity and when the surrounding soils become frozen the unfrozen soils in the center get squeezed out. In a well-developed frost-boil soil profile, it is clear that organic matter has been cryoturbated into the lower soil horizons (Figures 5 and 6). Even though the formation processes involved in earth hummocks and frost boils may not be the same, they all result in similar soil morphology: frost-churned organic matter in the lower horizons; and warped and broken soil horizons.

Soils formed in high mountain, alpine, and high plateau regions have characteristic morphological features. Although these soils also have cryogenic fabrics such as lenticular and reticulate structures, due to the general arid environment, these structures are more abundant in the lower active layers. In the exposed ridges and slopes, sorted circles are very common. In sorted circles, soils form in the center of the circles, whereas the rock fragments form the circles.

Biological and Chemical Properties

Temperature and moisture conditions control soil-profile development, microbial respiration in soil, and hence the character of soil organic matter. Temperature and moisture affect the accumulation and distribution of C stocks in the soil profile through controls on vegetation, cryic soil processes (such as cryoturbation) and the presence of or depth to permafrost. One of the most important features in cold soil formation is the accumulation of organic matter. Based on recent studies, the estimated total carbon stored in the peatlands in the Arctic and the boreal zones of Alaska ranges from 32–80 kg m^{-2} to more than 130 kg m^{-2}, respectively. This agrees with Ovenden's finding that peat accumulation is a function of climate and landscape position and it maximizes in the boreal region. Most of the organic soils are in the boreal regions in the Northern Hemisphere: northern Russia, the Canadian Shield, and interior Alaska. Thus the widest distribution of organic soils is in the cryic zone and the boreal zone. Recent study has indicated that, in the arctic tundra soils, nearly 50% of the total organic carbon is stored in the upper

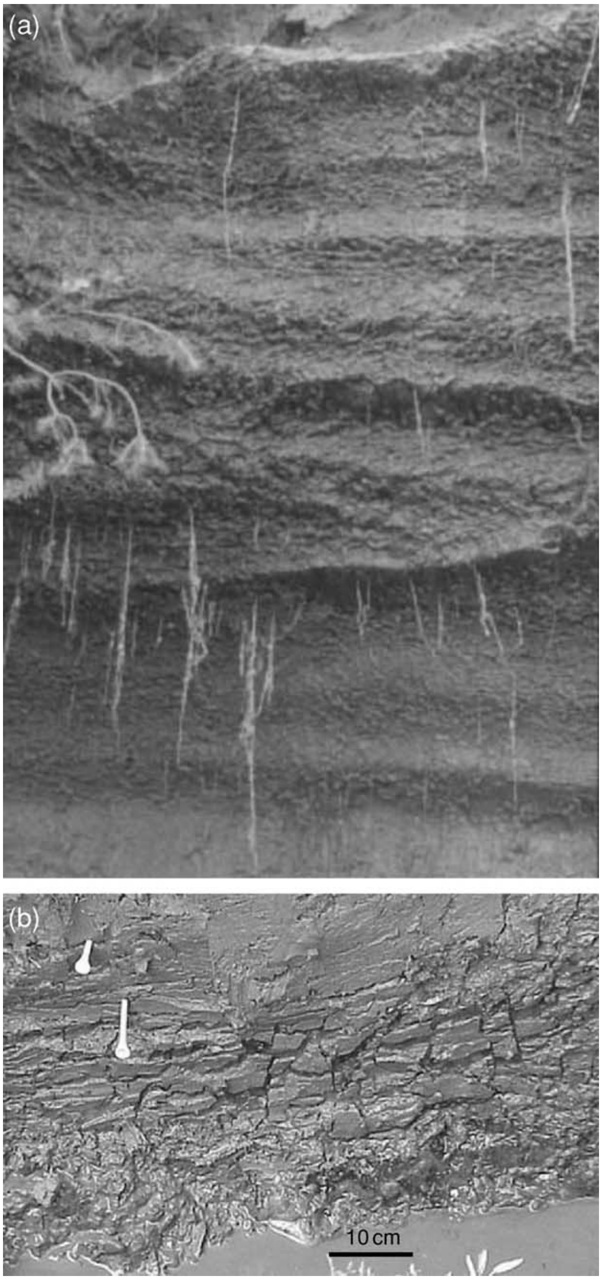


Figure 4 An enlargement of the active layer identified in Figure 3: (a) note the lenticular structure at the top and bottom, and reticulate structure in the middle; (b) the cryogenic structure can be better observed after the frozen soil is partially thawed.

permafrost as a result of cryoturbation. Thus cryoturbation contributes to the sequestration of atmospheric carbon into the upper permafrost. But if there is global climate change and warming of these regions, this portion of carbon could be a major source of CO_2 and methane into the atmosphere and a major perpetrator of more climate change.

Due to the presence of permafrost and seasonal freeze–thaw cycles, leaching is limited in most of the cold soils. In the high alpine regions, the leaching is further retarded by the arid conditions. In the high mountains and plateau, such as the northern Qinghai-Tibet Plateau in west China, the aridity (the ratio of potential evaporation to precipitation) exceeds 10 and is as high as 40 in some areas. Thus, in these soils, soluble salts often accumulate on the soil surface and the pH value ranges from 7.5 to 8.5. As the elevation drops, and with increased precipitation, the soluble salts and carbonates moved to the subsoils, whereas, in the arctic tundra regions, the lithology of the substrates has more influence on soil chemistry. Soils with a higher base status generally form in carbonate-rich loess or glacial drift and support nonacidic tundra vegetation. But with time, the vegetation succession gradually changes the soil chemistry by protonation or acidification of the surface soils, thus the landscape becomes acidic tundra. However, the leaching in these soils is generally weak and the soils are slightly acidic to slightly alkaline. But on the coastal plain and lowlands, base-rich fen forms and the soils are slightly to moderately alkaline.

The character of soil organic matter (SOM) in cold soils is influenced by its cryogenic environment, its position in the soil profile, the surface, subsurface soil active layer, and the presence of permafrost. Generally the SOM in cold soils is less decomposed than that in the temperate and tropical regions. In a recent study, the SOM in arctic soils was separated into extractable (EF) and nonextractable fractions (NEF) based on alkali solubility. The NEF, containing cell-wall and related constituents, such as hemicellulose and soluble fibers, dominate the SOM of cold soils. In a pattern consistent with a process of cold-temperature preservation, these NEF are found in largest proportions in the SOM of the upper permafrost. This fraction has greater potential to influence carbon cycling than EF in these ecosystems with climate warming.

Alpine and Plateau Regions

The high-plateau cold desert exists below the snowline on the high slopes on the mountains rising above the broad plateau in the Qinghai-Tibet Plateau, the Andes, Central Asia, and the High Arctic. The annual evaporation is more than 3000 mm, thus the climate is extremely arid and the aridity index is 23–24. Although the MAST is less than −3°C, the active layers are generally more than 2 m thick due to strong diurnal temperature fluctuations, dry conditions, and coarse-textured soils. The parent materials consist of

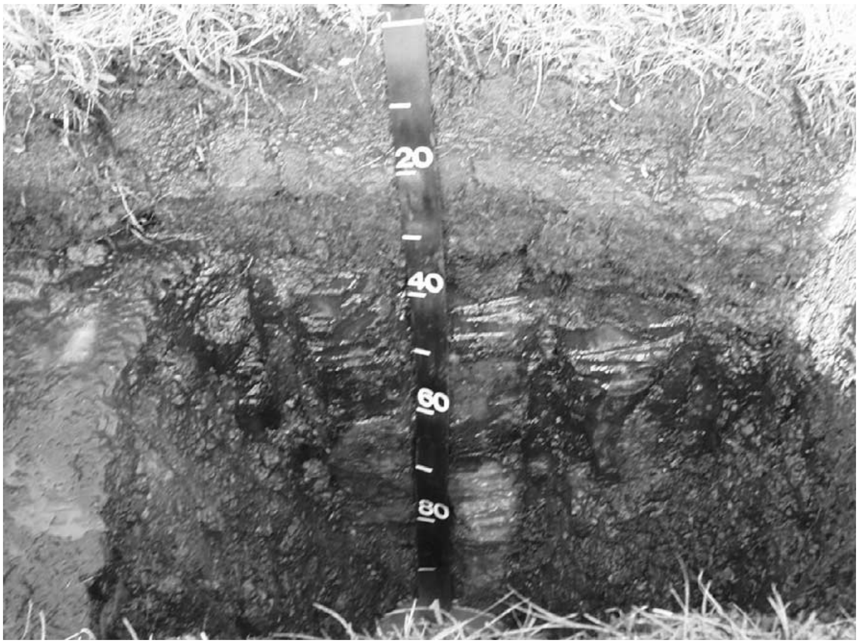


Figure 5 An organic soil formed in the arctic coastal marsh, Prudhoe Bay, Alaska. Note the cryoturbated underlying mineral horizon.

glacial deposits or colluvium. At elevations below 5000 m, ice-cemented permafrost may be present between 1 and 3 m below the surface. This area also experiences strong diurnal and seasonal freeze–thaw cycles that have resulted in granular structures in surface horizons and reticular or platy structures in subsurface horizons. Thus, cryoturbation occurs in these soils.

The alpine meadow zone lies below the high mountain zone and occupies the upland slopes on both sides of the mountains as well as the lowlands in the alpine zone. The climate is cold, semiarid to subhumid; with MAST of −0.2 to −6.0°C, MAP of 300–400 mm, and 85–90% falling between June and September, and annual evaporation of more than 2000 mm; thus the aridity index is more than 10. The active layer is generally more than 1.5 m. In addition to seasonal freeze–thaw, the area also experiences strong diurnal freeze–thaw cycles. The well-drained soils have a dense, black surface horizon, with many grass roots overlying a strong brown subsoil, often with an accumulation of carbonates, Fe, manganese oxides, and humus (to a lesser extent) coatings on soil particles, and well-developed reticular and fine, subangular blocky structures. The strong freeze–thaw cycle results in frost-heaved grass mounds, frost cracks, and gelifluction lobes. In lowlands and depressions of the alpine meadow zone, where drainage is limited, gleization features are prominent in the subsurface horizons. Organic soils form in depressions with restricted drainage.

Vegetation and Fire

The boreal forest is subjected to frequent lightning-ignited fires during the summer. Fires are an important factor in controlling soil properties through altered vegetation succession and permafrost dynamics. Vegetation mosaics and soil morphology in this region strongly reflect the combined effects of fire, permafrost, and slope. It is common to find evidence of past fire events such as accumulations of charcoal particles in soil profiles, especially at the bottom of organic horizons. In areas of soils recently affected by fire of moderate or severe intensity, vegetation is commonly altered to willow shrub and aspen or mixed white spruce–aspen forest types. Combustion and consolidation of the insulating organic mat can result in warming of underlying soils, lowering of the permafrost table resulting in a thicker active layer, and transition to a well-drained and permafrost-free state before eventual return to preburn conditions. Soil reaction in the surface organic horizons ranges from moderately alkaline (pH 7.9–8.4) immediately after fire to slightly acid levels (pH 6.1–6.5) several years later.

Land Use of Cold-Region Soils

The cold soils present a special challenge to land-use managers owing to the effects of frost heave. It has caused highway-maintenance problems, structure-foundation instability, inconvenience to people, and

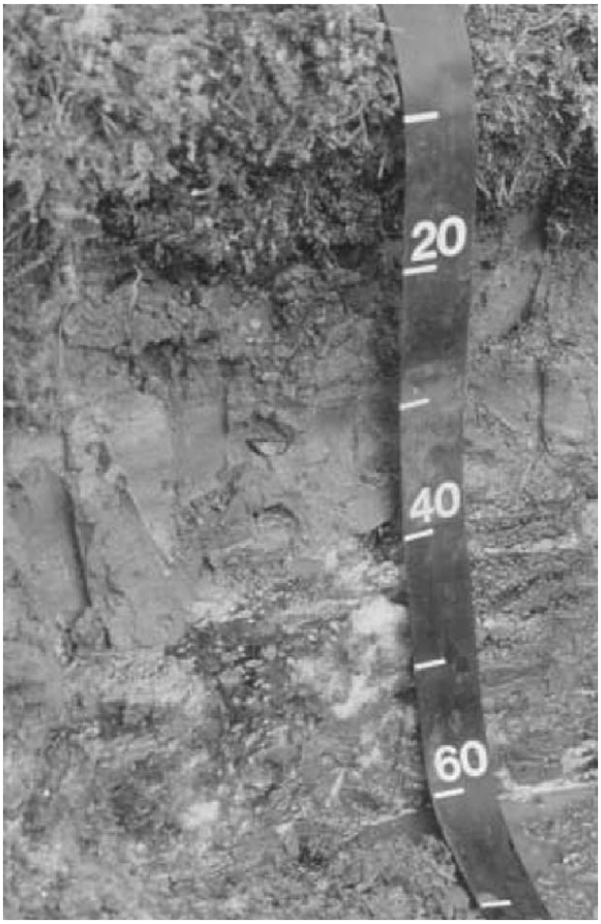


Figure 6 A cryogenic soil formed under moist acidic tundra, Toolik Lake Field Research Station, in the Arctic Foothills, Alaska. Note the frost-churned organic matter in upper permafrost at 45–60 cm.

inhibition of agricultural development. Because of the cold temperature, vegetation takes longer to recover after disturbance, and certain species of wildlife are unique to the cold regions; the ecosystems associated with cold soils are considered fragile. In past decades, natural resources extraction, mainly oil, gas, and minerals, has raised concerns. The optimum solution lies within our understanding of these cold soils and their environment.

Engineering

In the boreal regions, where permafrost is discontinuous and thin, the general practice is either to thaw the permafrost or to insulate the permafrost to minimize the effects of frost damage. But in the arctic region, the best practice is to insulate the permafrost to avoid thawing. Thick beds of gravel are applied to serve as an insulator, and lighter-color pavement is used to decrease solar radiation absorption. For permanent structures and houses, gravel pads or insulated foundations are commonly used in soils containing only ice lenses. But in areas where the soils contain a high content of ice, piling is used for structures and houses. In some cases, the piling or the foundation is refrigerated to avoid heat conduction to the permafrost.

Agriculture Development

Soil affected by permafrost accounts for more then 26% of the global land surface. Most of these areas are, in their natural state, tundra, boreal forest, cold deserts, and bogs. However, some areas have been cultivated for a long time. Cold soils present a great challenge to farm development not because of the low temperatures, but because of the ice content. In the subarctic or boreal regions, farms newly cleared from native forest underlying permafrost often face a subsidence problem. The land surface may become hummocked because of the melting of ice wedges and even thermokarst sinkholes, where large volumes of ice are present. Such problems are more severe on Pleistocene surfaces because of the higher ice content. The Holocene terraces are generally more favorable for land-clearing and farming.

The natives of the circumpolar regions have used the land for reindeer grazing for centuries. These people include the Samis in northern Scandinavian, the Eskimos in northern Alaska and Canada, and the Chukchi people in northern Russia. In the alpine and plateau regions, the Tibetans raise yaks, sheep, and cattle on the Tibetan Plateau; the Indians of the Andes have long raised llamas and other domesticated animals. In recent years there have been concerns of overgrazing in these fragile ecosystems. Once the organic layers or the surface vegetation covers are disturbed or destroyed, the quality of the land deteriorates owing to topsoil erosion.

There is intensive farming as well as grazing. Farming in the circumpolar region is made possible because of the long daylight hours in the summer that enable crops to mature in a shorter time than in the temperate region. In the boreal regions, where the soil conditions are favorable, small grains, especially barley, can mature within 110 days and oats within 130 days. Vegetables, such as carrots, lettuce, and cabbages can be harvested in less than 60 days. Potatoes are grown more than 200 km north of the Arctic Circle, where the permafrost table persists at less than 1 m from the surface. Thus in the boreal regions the potential for agriculture is greater than some of the northern temperate regions, where the summer always remain cool. However, such intensive farming can only exist in high-latitude environments owing to the prolonged daylight, but not in alpine and plateau environments due to the diurnal temperature changes; the temperature can reach more than 20°C at midday but drops below freezing at night.

It should be noted that the development of these areas leads to an increased active layer and deeper permafrost, and this results in changes in leaching and oxidation and reduction as well as cryogenic processes. Consequently, there will be more methane and CO_2 emission owing to increased wetlands from thawing of the permafrost and thickening of the active layers. Not withstanding all of these problems, these areas can be highly productive because of the long day length and radiation inputs. This also raises the question as to why these soils are classified as high-latitude soils when they have much higher summer temperatures than ones found at high altitudes. One area can be used for crop production and the other cannot.

The greatest concern of farming in soils affected by permafrost is subsidence. Since these soils contain variable amounts of ice, once the soils are stripped of their natural vegetation covers the thermal region changes and the soils will warm. In soils with a high ice content and ice wedges, thermokarst will occur, and the cleared field becomes hummocky and presents difficult management problems. Another concern is the nature of the substratum. When the land is cleared, water from the melted ice has to be drained. Waterlogging or ponding will occur if there is restricted layer or the farm is on a topographic low. Thus good understanding of the soil and the geographic environment is the key to successful farming in cold soils.

Further Reading

Agriculture Canada Expert Committee on Soil Survey (1998) *The Canadian System of Soil Classification*, 3rd edn. Ottawa, ON: NRC Research Press.

Bockheim JG, Walker DA, Everett LR, Nelson FE, and Shiklomanov NI (1998) Soils and cryoturbation in moist nonacidic and acidic tundra in the Kuparuk River Basin, arctic Alaska, USA. *Arctic and Alpine Research* 30: 166–174.

Brown J, Ferrians OJ Jr, Heginbottom JA, and Melnikov ES (1997) *Circum-Arctic Map of Permafrost and Ground-Ice Conditions*. US Geological Survey, Map CP-45, 1:10 000 000. Washington, DC: US Government Printing Office.

Clymo RS (1983) Peat. In: Gore AJP (ed.) *Ecosystems of the World*, vol. 4A, *Mires: Swamp, Bog, Fen, and Moor*, ch. 4, pp. 159–224. Amsterdam, The Netherlands: Elsevier Scientific.

Cryosol Working Group (2004) *Cryosols*. New York: Springer Verlag.

Dai XY, Ping CL, and Michaelson GJ (2002) Characterization of organic matter in Arctic tundra soils by different analytical approaches. *Organic Geochemistry* 33: 407–419.

French HM (1996) *The Periglacial Environment*, 2nd edn. White Plains, NY: Longman.

Gubin SV (1993) Late Pleistocene soil formation on coastal lowlands of northern Yakutia (in Russian). *Pochvovedeniye* 10: 62–70.

Hoefle CM and Ping CL (1996) Properties and soil development of late-Pleistocene paleosols from Seward Peninsula, northwest Alasaka. *Geoderma* 71: 219–243.

Lal R, Kimble JM, and Stewart BA (eds) (2000) *Global Climate Change and Cold Regions Ecosystems*. Boca Raton, FL: Lewis.

Michaelson GJ, Ping CL, and Kimble (1996) Carbon storage and distribution in tundra soils of Arctic Alaska, USA. *Arctic and Alpine Research* 28: 414–424.

Muller SW (1947) *Permafrost or Permanently Frozen Ground and Related Engineering Problems*. Ann Arbor, MI: JW Edwards.

Ovenden L (1990) Peat accumulation in northern wetlands. *Quaternary Research* 33: 377–386.

Péwé TL (1954) *Effect of Permafrost on Cultivated Fields, Fairbanks Area, Alaska*. US Department of Interior, Geological Survey Bulletin 989-F, pp. 315–351. Washington, DC: US Government Printing Office.

Péwé TL (1975) *Quaternary Geology of Alaska*. US Department of Interior, Geological Survey Professional Paper 835. Washington, DC: US Government Printing Office.

Ping CL, Michaelson GJ, and Kimble JM (1997) Carbon storage along a latitudinal transact in Alaska. *Nutrient Cycling in Agroecosystems* 49: 235–242.

Ping CL, Bockheim JG, Kimble JM, Michaelson GJ, and Walker DA (1998) Characteristics of cryogenic soils along a latitudinal transect in arctic Alaska. *Journal of Geophysical Research* 103(D22): 917–928.

Qiu GQ and Cheng GD (1993) Development condition of Cryosols in alpine and plateau regions of western China. In: Gilinchinsky DA (ed.) *Post-Seminar Proceedings. Joint Russian–American Seminar on Cryopedology and Global Change*, pp. 59–75. Pushchino, Russia: Russian Academy of Sciences.

Qiu GQ and Cheng GD (1995) Permafrost in China: past and present. *Permafrost and Periglacial Processes* 6: 3–14.

Reiger S (1983) The genesis and classification of cold soils. New York: Academic Press.

Richardson JL and Vepraskas MJ (2001) *Wetland Soils: Genesis, Hydrology and Classification*. Boca Raton, FL: Lewis.

Samson-Liebig SE, Kimble JM, and Ping CL (1995) Improvements in definition of cryic and pergelic soil temperature regimes in soil taxonomy using daylength/solar radiation. *Soil Survey Horizons* 36(1): 20–25.

Viereck LA (1970) Forest succession and soil development adjacent to the Chena River in Interior Alaska. *Arctic Alpine Research* 2: 1–26.

Viereck LA, Dyrness CT, and Foote MJ (1993) An overview of the vegetation and soils of the floodplain ecosystems of the Tanana River, interior Alaska. *Canadian Journal of Forest Research* 23: 889–898.

Walker DA, Auerbach NA, and Backheim JG (1998) Energy and trace-gas fluxes across a soil pH boundary in the Arctic. *Nature* 394: 469–472.

COLLOID-FACILITATED SORPTION AND TRANSPORT

R Kretzschmar, Institute of Terrestrial Ecology, Zurich, Switzerland

Introduction

Colloid transport in soils has attracted the attention of soil scientists since the 1930s, as they have tried to understand the formation of clay skins, argillic horizons, and clay pans in soils. Research on colloidal stability and aggregation of soil clays and reference clay minerals was later stimulated by problems due to soil dispersion in irrigation agriculture, where the release and transport of colloids can lead to drastic reductions in soil permeability. Since the early 1990s, increasing concern has been raised that mobile soil colloids may also serve as carriers for strongly sorbing contaminants such as heavy metals, hydrophobic organic compounds, and radionuclides, thereby facilitating the transport of contaminants which would otherwise be highly immobile. This colloid-facilitated transport mechanism is depicted schematically in Figure 1. A large number of laboratory and field studies on colloid transport in subsurface porous media have been conducted since then, providing a better understanding of the release, transport, and deposition behavior of colloidal particles. Today, many of the observed phenomena are understood qualitatively, but quantitative theories are still lacking and are the subject of ongoing research in environmental soil chemistry and physics.

Four key conditions must be fulfilled for colloid-facilitated contaminant transport to become an environmentally important factor: (1) mobile colloidal particles must be present in sufficiently large concentrations; (2) the contaminant must sorb strongly to the colloidal particles and desorb only slowly; (3) the particles and associated pollutants must be transported over significant distances through less contaminated zones of a soil or subsurface sediment; and (4) the concentrations of contaminants transported via colloids as carriers must exceed the tolerable limits. This chapter discusses the environmental factors controlling the release, transport, and deposition of colloids in soils and their potential influence on the transport of strongly sorbing contaminants.

Nature and Stability of Soil Colloids

Colloids are often defined as solid particles with an equivalent spherical diameter between 1 and 1000 nm dispersed in a liquid phase. The size limits of colloidal particles are to a certain degree arbitrary and should not be considered sharp boundaries between dissolved molecules, colloids, and larger suspended particles. However, colloids have two unique properties: (1) they have a very large specific surface area (more than $10\,m^2\,g^{-1}$) and therefore can be efficient sorbents for inorganic and organic contaminants; and (2) they exhibit extremely low gravitational settling velocities

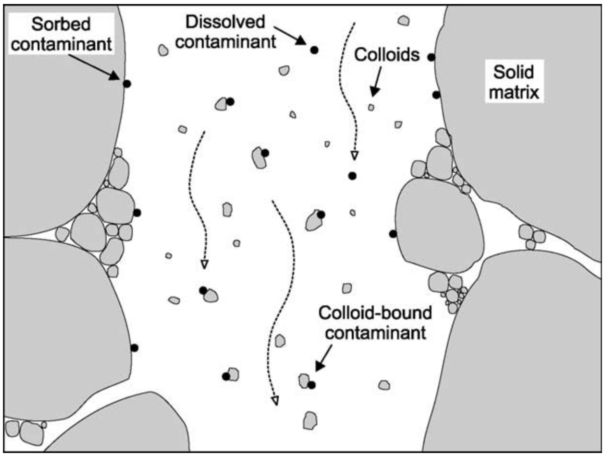


Figure 1 Schematic of colloid-facilitated contaminant transport in porous media. A contaminant (black sphere) can be present as a dissolved species, sorbed to the solid matrix, or bound to colloidal particles that move with the flowing water. For strongly sorbing contaminants, such mobile colloids may serve as carriers and provide a rapid transport pathway. Reproduced from Kretzschmar R, Borkovec M, Grolimund D, and Elimelech M (1999) Mobile subsurface colloids and their role in contaminant transport. *Advances in Agronomy* 66: 121–194 with permission from Elsevier.

in water and therefore remain stable in dispersion over long time periods, unless they coagulate to form larger aggregates or deposit on to liquid–solid or liquid–gas interfaces.

In soils and sediments, the most important source of mobile colloids is the release of small, submicron-sized mineral or organic particles existing in virtually every soil. In fact, the particle size fraction smaller than 1 μm can contribute significantly to the total soil mass and it usually contributes most of the specific surface area, even in sandy soils. As a result, the composition of mobile colloidal particles found in soils and underlying aquifers often reflects the composition of the fine particle size fractions present in the source horizons. Typical components include phyllosilicate minerals (e.g., kaolinite, illite, smectite), oxyhydroxides of Fe and Al (e.g., goethite, hematite, ferrihydrite), colloidal silica, and natural organic matter (NOM, e.g., humic substances). Mostly these components occur in mixtures or as mineral–organic complexes. The size ranges of potentially mobile particles in soils are depicted in Figure 2, along with the conventional particle size classes for comparison.

In some cases, mobile colloids can also be formed by precipitation from oversaturated solutions or from the release of colloids from soil amendments such as sewage sludge. One example where precipitation might play a role is the formation of colloidal calcite particles in the solution, for example, as a result of changes in CO_2 partial pressure. However, little is known about the colloidal stability of calcite particles and their mobility in soils. Viruses and bacteria also exhibit in some respects colloidal properties and are therefore sometimes considered biocolloids (Figure 2).

The stability of colloids in aqueous dispersion strongly depends on their surface charge properties and the ionic composition of the solution. In the vast majority of soils, mobile colloidal particles possess predominantly negative surface charge. Positively charged particles deposit very effectively on to negatively charged matrix grain surfaces due to electrostatic attraction. They are therefore highly immobile, unless the entire soil is dominated by positively charged mineral components. This is only the case in strongly acidic subsoil horizons of highly weathered Oxisols or Ultisols with low organic matter contents. All other soils are dominated by the negatively charged surfaces of clay minerals, feldspars, quartz, and soil organic matter. The surface charge of soil components can have different origins. Firstly, most phyllosilicates in soils have isomorphic substitution of cations either in tetrahedral (e.g., Al^{3+} for Si^{4+}) or octahedral (e.g., Fe^{2+} for Al^{3+}) sheets, resulting in permanent, pH-independent negative surface charge. Secondly, the edge surfaces of clay minerals and oxyhydroxide and oxide minerals have reactive surface hydroxyl groups which can protonate or deprotonate, resulting in pH-dependent negative or positive surface charge. At pH values below the point of zero charge (PZC), the surfaces are positively charged, while at pH values above the PZC the surfaces are negatively charged. Typical PZC values for different soil minerals are given in Table 1.

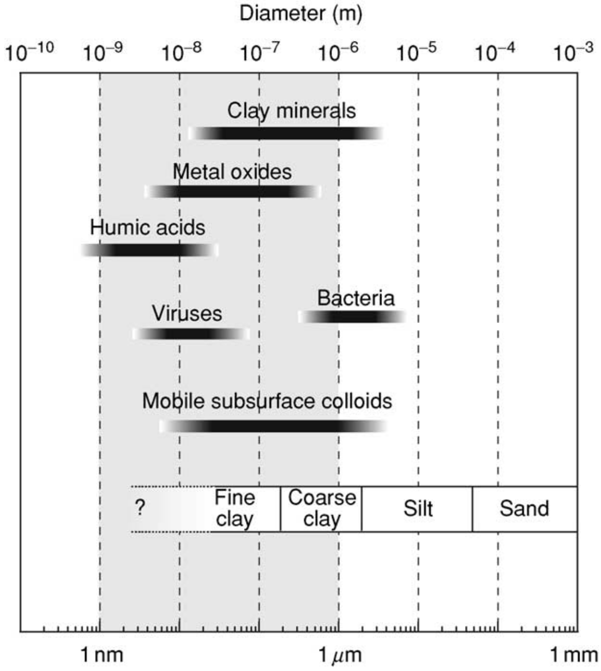


Figure 2 Size range of colloidal particles in soils. The classic size fractions commonly used in soil science are shown for comparison. Reproduced from Kretzschmar R, Borkovec M, Grolimund D, and Elimelech M (1999) Mobile subsurface colloids and their role in contaminant transport. *Advances in Agronomy* 66: 121–194 with permission from Elsevier.

Table 1 Typical points of zero charge (PZC) of various soil minerals in the absence of specifically adsorbed ions. Values reported in the literature vary by up to ±0.5 pH units, depending on mineral source, purity, method, and experimental conditons used

Solid	*PZC*
Calcite	9.5
Amorphous aluminum oxide	9.0
Amorphous iron oxide	9.0
Hematite	9.0
Goethite	8.5
Magnetite	6.5
Kaolinite	5.5
Montmorillonite	2.5
Feldspars	2.2
Quartz	2.0
Amorphous SiO_2	1.8

It is important to note, however, that adsorption of NOM or specifically adsorbing anions such as phosphate can drastically alter the surface charge of oxide and clay surfaces, resulting in a decrease in PZC. Under acidic conditions, one can observe surface charge reversal of oxide particles from positive to negative upon adsorption of NOM. Other ions which significantly reduce the PZC of oxide minerals to varying degrees include adsorbed silica, phosphate, arsenate, and sulfate.

The stability of a colloidal dispersion is determined by the balance between long-ranged electrostatic repulsive forces and short-ranged London–van der Waals attractive forces between colloidal particles. According to the classic Derjaguin–Landau–Verwey–Overbeek (DLVO) theory of colloidal stability, the sum of these forces results in an interaction force profile as a function of separation distance which strongly depends on the ionic composition of the solution and the surface charge density of the particles. At low ionic strength, the colloidal particles must overcome a repulsive energy barrier due to electrostatic repulsion. Therefore, not every particle collision results in coagulation (attachment efficiency $\alpha < 1$) and the process is termed slow or reaction-limited coagulation. With increasing ionic strength, the repulsive energy barrier diminishes and finally disappears due to charge screening and compression of the diffuse double layer. Under such conditions every particle collision results in aggregation (attachment efficiency $\alpha = 1$) and the process is termed rapid or transport-limited coagulation. The attachment efficiency α is the fraction of particle collisions that results in coagulation. In colloid chemistry, the critical coagulation concentration (CCC value) is defined as the minimum electrolyte concentration required for rapid coagulation under given pH and environmental conditions. The relationships between the cation concentration in monovalent, mixed bivalent/monovalent, and bivalent cation solutions and the attachment efficiency α are shown schematically in Figure 3.

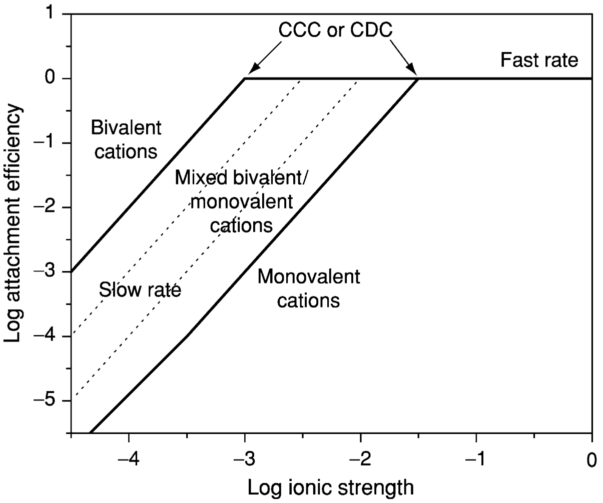


Figure 3 Influence of ionic strength of monovalent, mixed bivalent/monovalent, and bivalent cation solutions on the attachment efficiency (α) of negatively charged colloidal particles during aggregation or deposition on negatively charged surfaces. The transition from slow (log $\alpha < 0$) to fast (log $\alpha = 0$) deposition is called the critical coagulation concentration (CCC) for aggregation and critical deposition concentration (CDC) for deposition, respectively.

The CCC values of soil clays and clay minerals have often been determined by coagulation series tests in a simple batch system, in which the colloidal stability of clay suspensions is evaluated optically after a given time period allowed for coagulation and settling. Such CCC values are strictly operationally defined and one must be aware that they decrease with increasing initial clay concentration and increasing coagulation and settling time. One method to determine CCC values more quantitatively is by direct measurement of the relative coagulation rate using dynamic light scattering (DLS). So far, this method has only been used by a few researchers studying the colloidal stability of soil clays and reference clay minerals, but DLS has great potential.

For several different clay minerals, including illite, kaolinite, and montmorillonite, CCC values of typically less than 40 mol m^{-3} are observed in mixed Na/Ca systems with a sodium adsorption ratio (SAR) up to 60 and for pH values ranging from 6 to 9. The SAR is defined as

$$\mathrm{SAR} = \frac{[\mathrm{Na}^+]}{([\mathrm{Ca}^{2+}] + [\mathrm{Mg}^{2+}])^{0.5}} \qquad [1]$$

where square brackets denote solution concentrations in moles per cubic meter. Substantially higher CCC values than in mixed Na/Ca systems are measured in pure Na systems, where the strong coagulating effect of the divalent Ca^{2+} is absent. Since even sodic and saline-sodic soils contain considerable amounts of Ca^{2+} and Mg^{2+}, CCC values for pure Na systems are of limited practical relevance. Typical CCC values for colloidal dispersions of reference clay minerals in solutions of monovalent and divalent cations are given in Table 2. DLVO theory would predict that the CCC values are roughly 64 times higher for monovalent cations than for divalent cations. Similarly, for trivalent cations the CCC value would be expected to be roughly 700 times lower than for monovalent cations, illustrating the strong coagulating power of Al^{3+}. Experimental data are at least in qualitative agreement with this aspect of the DLVO

Table 2 Typical critical coagulation concentrations (CCC) of monovalent ($CCC_{monovalent}$) and divalent ($CCC_{divalent}$) cations for reference clay minerals

	$CCC_{monovalent}$ (mmol l^{-1})	$CCC_{divalent}$ (mmol l^{-1})	$CCC_{divalent}/CCC_{monovalent}$
Soil mineral			
Illite	48 ± 11	0.14 ± 0.02	0.003
Kaolinite	10 ± 4	0.3 ± 0.2	0.030
Montmorillonite	8 ± 6	0.12 ± 0.02	0.015
DLVO theory prediction			0.0156

DLVO, Derjaguin–Landau–Verwey–Overbeek.
Adapted from Sposito G (1989) *The Chemistry of Soils.* Oxford, UK: Oxford University Press.

theory, and deviations may partly be explained by specific interactions of cations with the mineral surface. In colloid science, the general dependency of the CCC value on cation valence is also known as the Schulze–Hardy rule, which has been recognized since the end of the nineteenth century.

Soil clays very often exhibit much higher CCC values than corresponding reference clay minerals. This is particularly true for soil clays isolated from surface soil horizons. This higher colloidal stability of soil clays compared to their reference clay counterparts is mainly attributed to the influence of adsorbed NOM such as humic substances. As mentioned earlier, adsorbed humic substances add negative surface charge and can cause charge reversal of clay edges from positive to negative, for example in kaolinite under acidic pH conditions. Adsorbed humic substances therefore lead to increased electrostatic stabilization of soil clay dispersions. In addition, adsorbed humic substances may also cause steric stabilization of clay colloids, which is due to a thermodynamic repulsion that occurs between interpenetrating organic polymer chains attached to the colloid surface. However, steric stabilization by adsorbed NOM may only be relevant if the thickness of the adsorbed organic layer exceeds the thickness of the diffuse double layer. Such conditions are most likely to occur at high ionic strengths or in the presence of multivalent cations.

Due to pH-dependent charge, CCC values of soil colloids are also strongly pH-dependent. For example, oxyhydroxide and oxide mineral colloids coagulate even in deionized water at pH values close to their respective PZC, because electrostatic repulsive forces are absent or too weak to cause electrostatic stabilization. Well above or below the PZC, such dispersions are stabilized by electrostatic repulsive forces. Kaolinite colloids exhibit a different behavior. At pH values less than the PZC of the edge surfaces (near pH 5.5), pure kaolinite particles coagulate rapidly even in deionized water, due to attractive interactions between positively charged edge surfaces and negatively charges face surfaces (edge-to-face coagulation). Above the PZC of the edge surfaces, the entire particles are negatively charged and electrostatic stabilization occurs. Addition of NOM leads to edge charge reversal from positive to negative at pH values below the PZC and strongly increases colloidal stability. Therefore, kaolinite colloids from topsoils often exhibit much higher colloidal stability than kaolinites from subsoils or reference clay sources.

Colloid Transport in Porous Media

The transport of colloidal particles in soils appears to be limited to larger pores, especially to preferential flow paths such as old root channels, earthworm burrows, and other interaggregate pores. Evidence for colloid transport in field soils stems primarily from micromorphological investigations showing oriented clay skins on pore walls and faces of soil peds. The formation of such features is attributed to the translocation and deposition of fine clay material within soil profiles. It has been shown that the composition of clay skin material in subsoil horizons has greatest similarities with the fine clay fraction in the corresponding surface soils, suggesting colloid mobilization near the surface (A and E horizons) and transport through macropores into the subsurface (e.g., Bt horizon).

Laboratory experiments have shown that colloidal particles in natural porous media can be transported faster than a conservative solute tracer such as tritiated water. This effect is termed 'size exclusion effect,' in analogy to size exclusion chromatography. Unlike a conservative solute tracer, colloidal particles are restricted to larger pores, and therefore the colloid peak elutes earlier than the tracer peak. Such effects have been observed both in undisturbed soil and saprolite columns, as well as in repacked soil columns.

The mobility of colloidal particles in soils strongly depends on solution and surface chemistry and on physical factors including water flow velocity, water saturation, and pore size distribution. Colloid transport in soils has been studied mostly in laboratory column experiments with repacked or undisturbed soil columns. Most experiments were done under water-saturated conditions, which makes it easier to control and vary the flow velocity. Under such conditions, colloid transport can be described by a convective-dispersive transport equation, including terms that account for colloid deposition and release:

$$(\partial C/\partial t) = D_\mathrm{p}\frac{\partial^2 C}{\partial x^2} - \nu_\mathrm{p}\frac{\partial C}{\partial x} - \frac{\rho_\mathrm{b}}{\epsilon}\frac{\partial S}{\partial t} \qquad [2]$$

and

$$((\rho_\mathrm{b}/\epsilon)(\partial S/\partial t)) = k_\mathrm{d}C - \frac{\rho_\mathrm{b}}{\epsilon}k_\mathrm{r}S \qquad [3]$$

Along with the appropriate boundary conditions, eqns [2] and [3] describe the colloidal particle concentration in suspension $C(x,t)$ and the amount of deposited particles per unit mass of the porous matrix $S(x,t)$ as a function of travel distance x and time t. Here, D_p is the hydrodynamic dispersion coefficient for colloidal particles, ν_p is the mean interstitial velocity of colloidal particles, ρ_b the solid matrix bulk density, ϵ the porosity, and k_d and k_r the colloid deposition and release rate coefficients, respectively. The deposition of colloids is often observed to follow a first-order kinetic rate law, while the colloid release rate is often extremely low under steady-state flow and constant chemical conditions. Therefore, many researchers have used a simplified equation that neglects particle release to describe colloid transport in laboratory-scale columns:

$$(\partial C/\partial t) = D_\mathrm{p}\frac{\partial^2 C}{\partial x^2} - \nu_\mathrm{p}\frac{\partial C}{\partial x} - k_\mathrm{d}C \qquad [4]$$

The factors influencing the colloid release (k_r) and deposition rates (k_d) are discussed in the following sections.

Factors Influencing Colloid Release

The release of colloidal particles in soils is still poorly understood, and predictive models or theories are currently lacking. However, a qualitative understanding of colloid release has been developed by controlled laboratory column studies, field observations, and theoretical considerations of particle–surface interactions. The release of significant concentrations of colloids in soils is mainly a result of physical or chemical perturbations. Physical perturbations can result, for example, from raindrop impact on bare soil, rapidly infiltrating water, drying and rewetting of soil aggregates, or soil tillage. Chemical perturbations can occur naturally due to atmospheric deposition and leaching of salts, or can be a result of management practices such as fertilizer application or soil irrigation. Possible mechanisms of colloid release due to chemical changes include the dissolution of cementing agents and, much more importantly, the dispersion of soil particles due to electrostatic repulsive forces at low ionic strength and/or a high percentage of exchangeable Na^+ on the cation exchange complex.

Figure 4 shows a conceptual view of the colloid-release process. The kinetics of colloid release in porous media are influenced by particle–surface interactions and the hydrodynamics of the flow field. The release of colloidal particles attached to a surface involves two subsequent steps: (1) detachment of colloidal particles and transport across the interaction boundary layer, and (2) diffusional transport of detached colloidal particles through a stationary water film surrounding the matrix grains, the so-called diffusion boundary layer. The detachment step is strongly influenced by the surface charge of colloids and matrix grains and by solution chemistry. At sufficiently high ionic strength, the particles must escape from a primary energy minimum in order to detach from the surface. Under these conditions, detachment of particles is the rate-limiting step (detachment-limited release) and the release rates are often extremely low. With decreasing ionic strength, the depth of the primary energy minimum decreases. At sufficiently low ionic strength, the interaction forces between colloids and matrix grains can become repulsive at all separation distances, resulting in rapid detachment. The diffusion through the stationary water film is now rate-limiting (transport-limited release). Since the thickness of the diffusion boundary layer decreases with increasing flow velocity, colloid release rates are expected to increase with flow velocity. Additional shear and drag forces can act on colloidal particles and further increase the colloid release rates at very high flow rates, for example, during rapid infiltration of water through macropores.

Chemical changes that can induce colloid release include decreases in soil-solution ionic strength, increases in the percentage of Na^+ on cation exchange sites, increases in soil pH, and adsorption of ions or molecules which alter the mineral surface charge. Decreases in ionic strength can result from rainfall or irrigation with low-ionic-strength water. However, experimental results and field observations show that a decrease in ionic strength alone does not result in significant colloid release if the soil's cation exchange capacity is saturated with multivalent cations such as Ca^{2+} or Al^{3+}. If the soil is partly saturated with monovalent cations, especially Na^+, decreases in

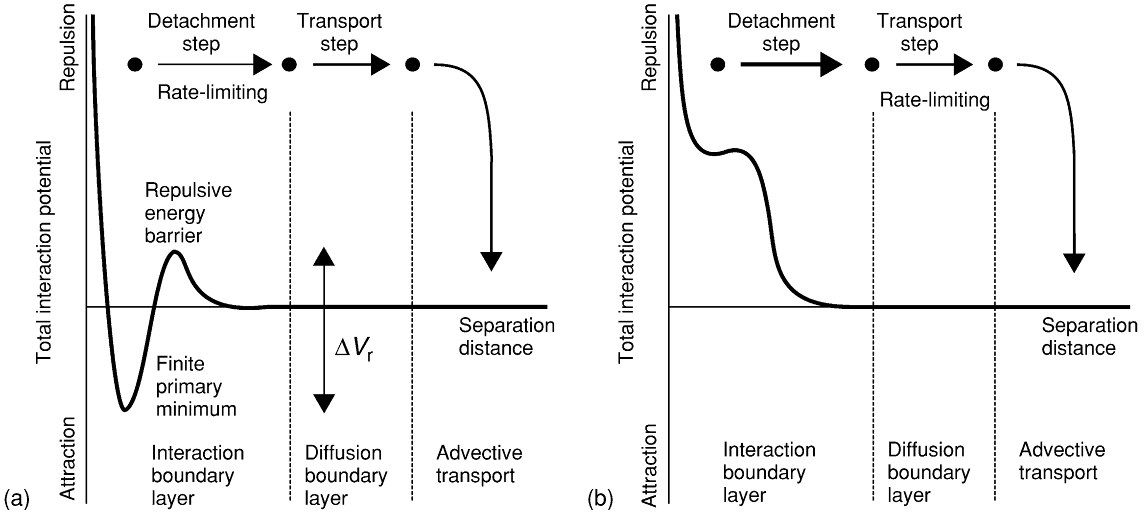


Figure 4 Conceptual view of the colloid-release process consisting of two steps. In the first step, colloidal particles must escape from the primary minimum in which they are attached. The release energy corresponds to ΔV_r. In the second step, the colloidal particles must be transported across the diffusion boundary layer. (a) At medium ionic strength, $\Delta V_r > 0$ and the release kinetics is detachment limited; (b) at very low ionic strength, the barrier height diminishes and the release kinetics becomes transport-limited. Reproduced from Kretzschmar R, Borkovec M, Grolimund D, and Elimelech M (1999) Mobile subsurface colloids and their role in contaminant transport. *Advances in Agronomy* 66: 121–194 with permission from Elsevier.

ionic strength can result in significant colloid release, which usually leads to drastic reductions in soil-water permeability. This problem is rather common in soils where water with a relatively high SAR and low salinity is used for irrigation. In irrigation agriculture, an exchangeable sodium percentage (ESP) of more than 15% of the cation exchange capacity is considered as a critical value. However, also surface soils dominated by kaolinite clay (e.g., in Ultisols) can be highly dispersive and generate mobile soil colloids, which are probably stabilized in suspension by adsorbed NOM.

In strongly acidic mineral soils with high Al^{3+} saturation of the cation exchange capacity, colloid release rates are usually low due to the strong aggregating power of trivalent cations. Exceptions may occur in strongly weathered Ultisols or Oxisols, where positively charged colloidal particles may be mobilized at low ionic strength. However, this has only rarely been observed in practice. Another important factor leading to colloid release is the adsorption of ions or molecules, which make the colloid surfaces more negatively charged. For example, phosphate additions to soils can lead to a decrease in the PZC of oxides and cause charge reversal from positive to negative charge at given pH. As a result, colloidal particles can be more easily released and transported through the soil. Similarly, organic molecules which adsorb to soil particles can add negative surface charge and thereby increase colloid release rates. Examples include humic substances, malonate, citrate, oxalate, and salicylate. Also anionic surfactants and strong complexing ligands such as EDTA can lead to increased colloid release in soils, which has to be taken into consideration when developing in-situ soil remediation technologies. It has been demonstrated that carboxylic acids and other organic constituents released by plant roots can cause kaolinite particle dispersion, suggesting that roots and microorganisms can induce colloid release under natural conditions. On the other hand, many soil organisms are also known to release polysaccharides which tend to stabilize soil aggregates and thereby prevent soil dispersion and release of colloids.

Factors Influencing Colloid Deposition

The most important mechanism of particle removal from solution in porous media is colloid deposition, that is, the collision and attachment of colloidal particles on surfaces of larger grains within the soil matrix. If the chemical conditions are favorable for colloid deposition, granular porous media are extremely efficient particle filters and they are therefore widely used in water treatment. Much research on colloid transport and deposition has been conducted to understand the performance and failures of granular filter beds. Colloid transport and deposition in soils follow basically the the same principles. Particle deposition is commonly assumed to take place in two steps: (1) transport of colloidal particles to matrix

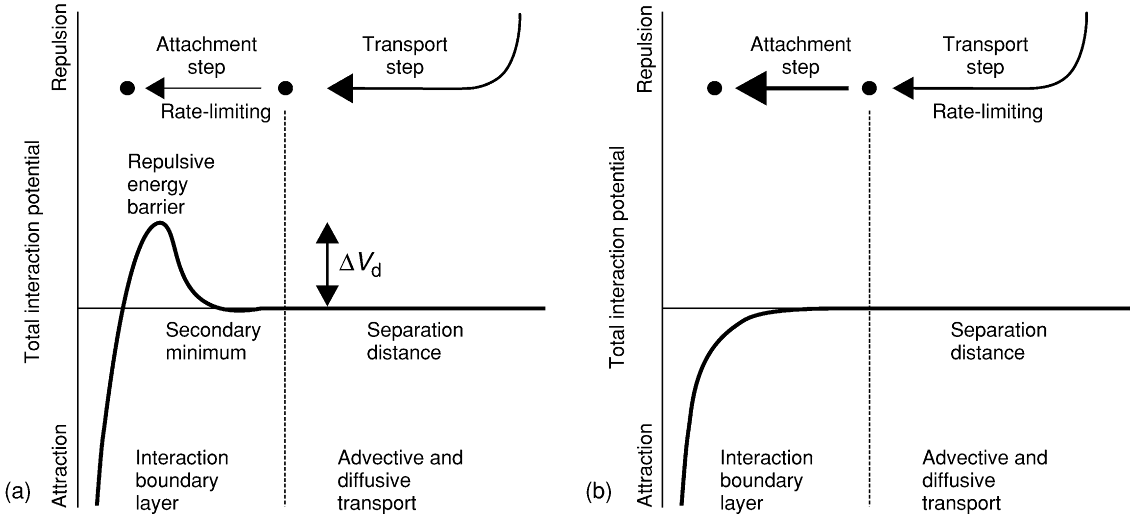


Figure 5 Conceptual view of the colloid-deposition process, consisting of two steps. In the first step, colloidal particles must be transported to the matrix surface by diffusion, interception, or gravitational sedimentation. In the second step, the colloids must be transported across the repulsive energy to be attached in the primary energy minimum. (a) At low ionic strength, a pronounced repulsive energy barrier with height ΔV_d develops between similarly charged colloids and surfaces. As a result, colloid deposition is slow or attachment-limited; (b) at high ionic strength, the repulsive energy barrier disappears and the kinetics of colloid deposition becomes fast or transport-limited. Reproduced from Kretzschmar R, Borkovec M, Grolimund D, and Elimelech M (1999) Mobile subsurface colloids and their role in contaminant transport. *Advances in Agronomy* 66: 121–194 with permission from Elsevier.

surfaces within the porous medium by Brownian diffusion, interception, or gravitational sedimentation, resulting in colloid–matrix collisions, and (2) attachment of colloidal particles to the matrix surfaces. A conceptual view of the particle-deposition process is shown in Figure 5.

The kinetics of the transport step depends on physical factors such as size and density of colloidal particles, accessible surface area for colloid deposition, pore structure, and flow velocity. The particle diffusion velocity decreases with increasing particle size, while interception and gravitational settling velocities increase. Therefore, particles with diameters in the order of 0.1–1 μm are expected to be most mobile in porous media, but this also depends on the specific mass of the colloidal particles.

The kinetics of the attachment step depends strongly on the charge density of colloidal particle and matrix surfaces and on the ionic composition of the solution. In principle, colloid attachment is analogous to heterocoagulation, where the radius and surface charge of two interacting particles are different. If the colloids and matrix surfaces are similarly charged, a repulsive energy barrier develops due to electrostatic repulsive forces and slows down the colloid attachment rate ($\alpha < 1$). With increasing ionic strength, the repulsive energy barrier is diminished and the deposition rate increases until every particle collision results in attachment ($\alpha = 1$). At this point, the deposition kinetics is transport-limited and is largely independent of chemical factors (Figure 3). If the colloidal particles and matrix surfaces are oppositely charged, the deposition rate is always rapid ($\alpha = 1$) due to attractive electrostatic forces.

Experiments with pure hematite colloids have shown that positively charged colloids are highly immobile in soils dominated by negatively charged mineral components, such as 2:1 type clay minerals, quartz, and feldspars. Only if the hematite colloids were coated with humic substances, resulting in charge reversal, could significant colloid transport be observed. This and other research results suggests that NOM coatings play a crucial role in the stabilization and transport of colloidal particles in soils.

In column experiments on colloid transport and deposition through granular porous media, colloid deposition often follows a first-order kinetic rate law, leading to an exponential decrease in suspended colloid concentration with travel distance:

$$\frac{C(x)}{C_0} = \exp(-\lambda x) \qquad [5]$$

where C_0 is the colloid concentration in the column inlet and λ is the so-called filter coefficient. The filter coefficient can be expressed in terms of the macroscopic deposition rate, k_d, by:

$$\lambda = \frac{k_d}{\nu_p} \qquad [6]$$

Colloid deposition rate coefficients k_d can be determined from the colloid breakthrough curves resulting from step-input or pulse-input column experiments.

If large amounts of colloidal particles are introduced into a porous medium the colloid deposition rate can change due to so-called blocking or filter-ripening effects. Again, these terms stem from research in deep-bed filtration, but the phenomena have also been demonstrated in soils and sediments. When colloid–colloid attachment is unfavorable (i.e., stable dispersion) because of interparticle repulsion, the colloid deposition rate tends to decrease with increasing amounts of deposited colloids on matrix surfaces (blocking effect). When colloid–colloid attachment is favorable (i.e., unstable dispersion), the colloid deposition rate tends to increase with increasing amounts of deposited colloids and multiple layers of colloid can form on matrix surfaces, ultimately leading to pore clogging. Pronounced blocking effects have been observed for transport of soil fine clays through undisturbed porous saprolite columns at low ionic strength, where colloid deposition results in a rapid decline in the deposition rate. Filter ripening may occur at high ionic strength, but under such conditions colloid release is usually low.

Another factor which strongly affects colloid deposition rates in porous media is the surface-charge heterogeneity which inherently occurs in soils, sediments, and weathered bedrocks. Surface-charge heterogeneity leads to preferential sites for colloid deposition, that is, sites that are particularly favorable for attachment. For example, a weathered granitic saprolite contains mainly negatively charged mineral components such as quartz, feldspars, and biotite. However, it also contains secondary minerals such as kaolinite, goethite, hematite, and ferrihydrite. Such components can create patches of positively charged surfaces which act as favorable deposition sites for negatively charged colloidal particles. The consequence of surface-charge heterogeneity is that colloid mobility cannot be predicted from streaming-potential or zeta-potential measurements of the matrix and colloidal particles, respectively, because such measurements usually provide only an average zeta potential for the entire sample. Charge heterogeneity may also explain the rapid blocking effects observed in weathered saprolites when small amounts of negatively charged soil colloids are introduced at low ionic strength.

So far, the discussion has been limited to the case of saturated flow. Since soils are often unsaturated, it is also important to consider the influence of a gas phase on colloid transport and deposition. Generally, the presence of air-filled pores leads to a lower connectivity of the water-filled pores and to the presence of water–gas interfaces in addition to water–solid interfaces. Hydrophobic colloidal particles are effectively removed by attachment to water–gas interfaces, while hydrophilic colloidal particles are less affected. Therefore, unsaturated conditions lead to a lower mobility of colloids in porous media, in particular of hydrophobic colloidal particles. For example, it has been shown experimentally that hydrophilic and hydrophobic bacteria have similar mobility in a water-saturated sand, but the hydrophobic bacteria are much less mobile than the hydrophilic bacteria under unsaturated flow conditions.

Colloid-Facilitated Transport of Contaminants

Strongly sorbing contaminants such as heavy metals, oxyanions, radionuclides, certain pesticides, and other hydrophobic organic compounds can be detected much earlier in the subsoil or groundwater than anticipated from traditional solute transport models. This discrepancy can have several different causes, including preferential flow through macropores and colloid-facilitated transport. In many cases, both transport mechanisms probably occur simultaneously, since the mobilization of colloids is most pronounced when dry soils are rewetted during a rainstorm event and dispersed particles are transported into deeper layers of the soil through rapidly infiltrating water along macropores. In contrast, during dry periods or periods with slow unsaturated water flow, the mobility of colloids in soil is small and preferential flow plays a minor role. Thus, colloid-facilitated transport and preferential flow typically occur together and the contributions of both effects are not easily separated.

In agricultural soils, colloidal transport of pesticides and phosphate are the greatest concern. For example, facilitated transport of the herbicides Napropamide (2-(α-naphthoxy-*N*,*N*-diethylpropionamide) and Glyphosate (*N*-(phosphonomethyl)glycine) by dissolved or colloidal NOM and other soil colloids has been demonstrated in column experiments. Drying and wetting of the soil surface and rainstorm events with rapidly infiltrating water enhance colloid mobilization and can lead to colloid-facilitated leaching of pesticides. The result is a rapid appearance of the pesticide in the leachate in association with dispersed colloids or NOM. Colloid-facilitated transport of phosphate is also a major concern on sandy soils that receive regular applications of animal manure, sometimes at excessive levels from the perspective of plant nutrition. Field lysimeter studies have shown that large percentages of total phosphorus in soil solution can be present in association with colloidal material, probably as colloidal organic matter–metal–orthophosphate complex or occluded in other colloidal particles. Excessive application of animal manure could increase the release of colloids and therefore the risk of groundwater pollution with phosphate and pesticides.

In metal-contaminated soils, colloid-facilitated transport of heavy metals may occur, depending on the vegetation cover, soil management, climate, and soil properties. In soil solutions, some metals such as Cu are often predominantly present as dissolved or colloidal organic complex. Heavy metals such as Pb, Cu, and Zn have also been shown to be associated with clay colloids and biocolloids (bacteria), respectively. Although colloid-facilitated transport of heavy metals has been demonstrated in laboratory experiments, quantitative field studies are lacking and the importance of this process under field conditions is still uncertain.

One important aspect which must be addressed in future research is the role of slow-desorption kinetics of heavy metals or radionuclides from surfaces of colloidal particles. Model calculations show that if sorption–desorption reactions are at local equilibrium (reversible sorption), then colloid-facilitated transport is not important. Contaminants associated with colloidal particles would desorb and be retained by the large excess of binding sites associated with the soil matrix. However, if sorption is irreversible or desorption is at least extremely slow compared with the residence time of the colloidal particles, then colloid-facilitated transport would theoretically be feasible. In soils, most colloids are probably transported during rainstorm events with rapid infiltration of water, and local equilibrium is therefore the exception. However, in the context of safety assessment of geologic disposal of nuclear waste, the flow velocities are extremely slow and local equilibrium may be a valid assumption. In any case, the behavior of mobile colloidal particles and associated contaminants in soils must be further studied in order to assess the importance of colloid-facilitated transport under field conditions.

Further Reading

Buffle J and van Leeuwen HP (1993) *Environmental Particles,* vol. 2. Boca Raton, FL: Lewis Publishers.

Elimelech M, Gregory J, Xia X, and Williams RA (1995) *Particle Deposition and Aggregation. Measurement, Modelling, and Simulation.* London, UK: Butterworth-Heinemann.

Goldberg S, Lebron I, and Suarez DL (2000) Soil colloidal behavior. In: Sumner ME (ed.) *Handbook of Soil Science*, Ch. 6, pp. B195–B240. Boca Raton, FL: CRC Press.

Hiemenz PC and Rajagopalan R (1997) *Principles of Colloid and Surface Chemistry.* New York: Marcel Dekker.

Kretzschmar R, Borkovec M, Grolimund D, and Elimelech M (1999) Mobile subsurface colloids and their role in contaminant transport. *Advances in Agronomy* 66: 121–194.

McCarthy JF and Degueldre C (1993) Sampling and characterization of colloids and particles in groundwater for studying their role in contaminant transport. In: Buffle J and van Leeuwen HP (eds) *Environmental Particles*, vol. 2, pp. 247–315. Boca Raton, FL: Lewis Publishers.

McCarthy JF and Zachara JM (1989) Subsurface transport of contaminants. *Environmental Science and Technology* 23: 496–502.

McDowell-Boyer LM (1992) Particle transport through porous media. *Water Resources Research* 22: 1901–1921.

Ryan JN and Elimelech M (1996) Colloid mobilization and transport in groundwater. *Colloids and Surfaces A, Physicochemical and Engineering Aspects* 107: 1–56.

Sposito G (1989) *The Chemistry of Soils*. Oxford, UK: Oxford University Press.

COMPACTION

J J H van den Akker, Alterra Wageningen UR, Wageningen, The Netherlands
B Soane, Formerly with Scottish Agricultural College, Edinburgh, UK

Introduction

Soil compaction is a form of physical degradation in which soil biological activity and soil productivity for agricultural and forest cropping are reduced, resulting in environmental consequences away from the immediate area directly affected. Compaction is a process of densification and distortion in which total and air-filled porosity and permeability are reduced, strength is increased, soil structure partly destroyed, and many changes induced in the soil fabric and in various characteristics. The term 'compaction' is used to identify a process and should be distinguished from the term 'compactness,' which indicates for a given time and position the state of packing of the solid soil constituent.

The compaction process can be initiated by wheels, tracks or rollers, traffic of cultivation machinery, and passage of draft or grazing animals. Soil densification can also be caused by heavy overburdens of ice and snow and by illuviation of clay from the A-horizon into the B-horizon. These processes may take place slowly over long periods of time and result in the B-horizon having much higher bulk density and strength than the A-horizon. Densification also occurs during structural collapse near the soil surface due to the impact of rain, resulting in formation of surface crusts.

In arable land with annual ploughing, both topsoil and subsoil compaction should be considered. We define the subsoil as the soil below the loosened layer (about 20–35 cm thick). This definition of the subsoil includes the panlayer as the upper part of the subsoil. This panlayer is, in many cases, less permeable for roots, water, and oxygen than the soil below it and is the bottleneck for subsoil functions. In contrast to the topsoil, the subsoil is not loosened annually, compaction is cumulative and, in the long run, a more or less homogeneous compacted layer is created. The resilience of the subsoil for compaction is low, and subsoil compaction is at least partly persistent.

Problems of compaction are widely distributed throughout the world, but tend to be most prevalent where heavy machinery is used in agriculture or forestry, in both temperate and tropical areas. Soils that are naturally fragile in structure, such as soils of the humid tropical forest and light-textured soils in areas of low but erosive rainfall, are particularly prone to problems arising from compaction and subsequent high risks of erosion due to reduction of permeability. Compaction is now included in surveys of soil degradation, and preliminary estimates have suggested that the area of degradation attributable to soil compaction may equal or exceed 33 Mha in Europe and 18 Mha in Africa. A project on mapping of Soil and Terrain Vulnerability in Central and Eastern Europe (SOVEUR) by the United Nations Food and Agriculture Organization (FAO) and International Soil and Reference Information Centre (ISRIC) showed that compaction is the most widespread kind of soil physical soil degradation in these countries. About 25 Mha proved to be lightly and about 36 Mha moderately compacted.

Factors and Processes Affecting Distribution and Intensity of Compaction

Compaction under Running Gear

Compaction under wheels and caterpillar tracks is a dynamic process in which a soil volume under the wheel undergoes normal and shearing stresses resulting in normal and shear strains. Neighboring soil volume elements at a certain depth under a wheel are considered in Figure 1. During a wheel pass volume element 1 at a certain depth endures not only compaction but also deformation, as depicted by the volume elements 2–26. The impact of the wheel load on the physical properties of the soil depends on the strength of the soil (*see* **Stress–Strain and Soil Strength**). If the exerted soil stresses exceed the precompression stress and shear strength of the soil, then compaction will be accompanied by large deformations and macropores and structure will be remolded, resulting in degradation of soil physical qualities. Degradation of soil physical qualities will be less severe if only the precompression stress is exceeded, because then the soil will mainly compact with limited deformations, and the remnants of the biomacropores and the intra-aggregate space will still retain a certain continuity.

Loading Characteristics of Individual Wheels or Caterpillar Tracks

For practical purposes a useful parameter is the average ground contact pressure *P* (kPa), which is defined as the wheel load divided by the ground contact area.

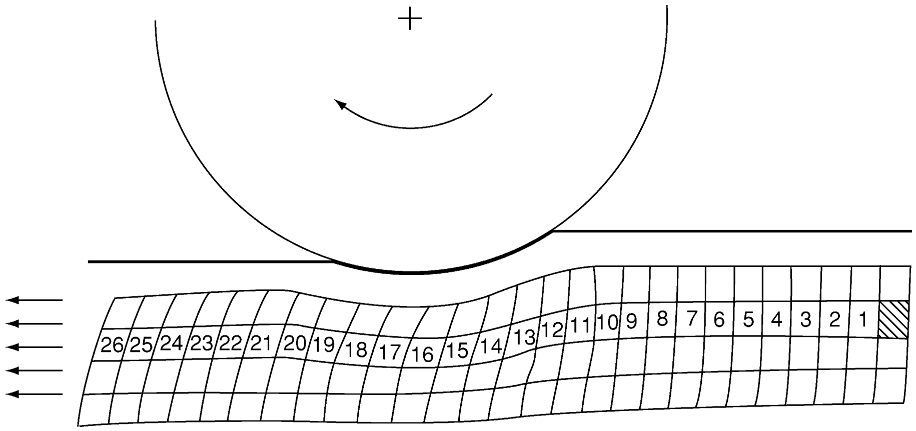


Figure 1 Displacement and deformation in a vertical section of a soil volume element under a wheel. Reproduced with permission from Koolen AJ and Kuipers H (1983) Agricultural soil mechanics. *Advanced Series in Agricultural Sciences* 13, Springer-Verlag, Heidelberg.

The average ground contact pressure can also be calculated as the sum of tire inflation pressure P_i and a pressure P_c for carcass stiffness:

$$P = c\,P_\mathrm{i} + P_\mathrm{c} \qquad [1]$$

The range for c is 0–1.25 and for P_c is 0–50 kPa. The factor c also depends on the carcass stiffness, and therefore eqn [1] is mostly reduced to $P = c\,P_\mathrm{i}$ or $P = P_\mathrm{i} + P_\mathrm{c}$. At inflation pressures larger than 200–300 kPa, the influence of carcass stiffness diminishes and the factor c can become smaller than 1. The higher the inflation pressure, the more the strength and firmness of the soil determine the ground pressure. The peak stresses in the ground contact area determine the peak stresses in the soil that result in compaction and distortion of the soil structure if the soil strength is exceeded. Peak stresses under lugs can be two to five times higher than the average ground contact pressure. However, the resulting soil stresses decrease rapidly with depth because the ground contact area is small. If soil stresses at a depth of 0.2–0.3 m are considered, then the peak stresses under the lugs can be neglected and a parabolic stress distribution in the ground contact area with a peak stress of one-and-a-half to two times the average ground contact pressure can be assumed. Peak stresses under caterpillar tracks are two to four times the average ground contact pressure and depend strongly on the firmness of the soil and design of the track system.

Examples of load characteristics for various vehicles and animals are shown in Table 1.

Compaction by agricultural machinery extends well into the subsoil (Figure 2). An important cause of subsoil compaction is moldboard plowing in which the furrow-side tractor wheels apply appreciable loads directly to the upper surface of the subsoil, causing a plow pan. Attempts have been made to specify maximum recommended average ground contact pressures in order to minimize compaction, especially in the subsoil. Suggested maximum values range from 80 kPa for wet soils in spring to 200 kPa for dry soils in summer, but progress in gaining official acceptance for such standards has been slow.

Table 1 Maximum loads and ground contact pressures applied by various sources

Source	*Total load (kN)*	*Load per wheel/track/ hoof (kN)*	*Average ground contact pressure (kPa)*
Large-wheeled tractor (120 kW)	100	50	250–350
Small-wheeled tractor (40 kW)	40	20	200–300
Sugarbeet harvester (loaded)	300–600	50–120	200–400
Slurry tanker (loaded)	100–300	25–60	200–500
Track-laying tractor	140	70	40
Horse	8	2–8	75–300
Cow	4–5	1–4	120–480

Field Traffic Intensity and Distribution

The overall incidence of soil compaction within a given field depends upon the distribution of traffic for each field operation and the cumulative value throughout the life of the crop. The weight of the crop to be transported from the field is an important factor in the compaction risk. The traffic intensity and compaction risk for a potato crop and a sugar beet crop are, respectively, more than twice and almost twice that of a winter wheat crop. As the number and width of wheels fitted to vehicles increase, so does the overall proportion of the field area covered

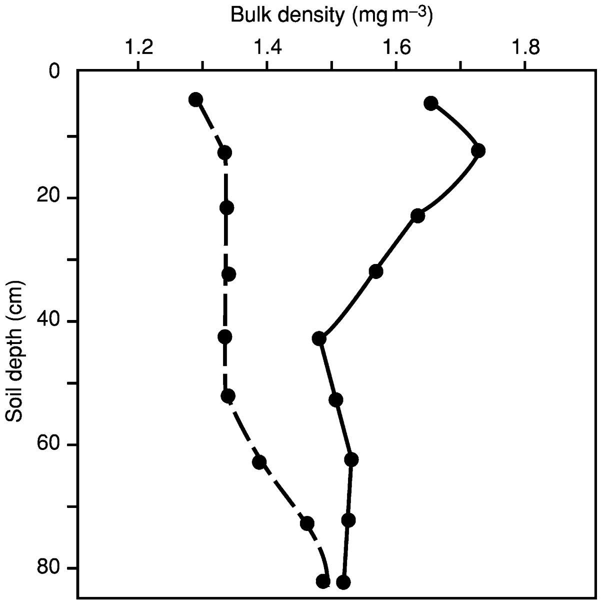


Figure 2 Increase of bulk density to considerable depth when a bulldozer was used for forest clearing in Surinam (---no vehicle traffic; — bulldozer traffic). Reproduced with permission from Van der Weert R (1974) *Tropical Agriculture (Trinidad)* 51(2): 325–331.

by wheels. However, this effect does not counteract the benefits from a reduction in compaction due to reduced ground contact pressure.

Soil Compactibility

Soils can vary from being sufficiently strong to resist all likely applied loads (low compactibility) to being so weak that they are compacted by even low loads (high compactibility). Well-structured soils combine good physical soil properties with high strength. Sandy soils with a single-grain structure and compacted massive soils can be very strong. Rootability and soil physical properties are then often poor. Roots have a binding action and increase the elasticity and resistance of a soil to compaction.

Soil moisture content and soil water suction have a dominant influence on soil compactibility. In soils where capillarity is the main cohesive force, strength increases with drying until it reaches a maximum, and then again decreases upon further dehydration because the cross-sectional area of the menisci decreases more than capillary suction between grains increases. This cohesive force based on soil water suction is called apparent cohesion. Drying also increases the true cohesion. Soil water suction increases the cohesion and so the strength and resistance to compaction of soils vary considerably. In dry structured (aggregated) soils, soil water concentrates in the aggregates, and cohesion and soil strength in the aggregate increases considerably. Dry structured soils shrink and turn into an assembly of individual aggregates that fit together rather neatly. This assembly of aggregates has a very high interaggregate angle of internal friction and moderate interaggregate cohesion. This soil is strong and has a low compactibility. However, if such a soil is overloaded and compacted, the aggregates will be crushed and the interaggregate space will be filled up, resulting in a dense compact soil.

Dry soils resist loads readily. However, extremely dry sandy soils can be deformed and compacted rather easily. As the moisture content increases, soil compactibility increases until the moisture content is approximately at the field capacity point, when a condition known as the optimum moisture content for compaction is reached. At still higher moisture contents, the soil becomes increasingly incompactible as the moisture tends to fill ever more of the total porosity and further loss of air-filled porosity becomes impossible. However, although the compaction may be minimal, the plastic flow of a wet soil results in a complete destruction of soil structure and macropores with accompanying diminishment of soil physical qualities. The reaction of dry and strong soils may be largely elastic, whereas at high moisture content and low strength the reaction may be plastic flow.

Increases in organic matter content tend to reduce soil compactibility and to increase elasticity. For example, peat soils are quite resistant to compaction. The value of the optimum moisture content for compaction increases as the organic matter content increases.

Effects on Soil Physical and Mechanical Properties

Bulk Density, Porosity, and Packing State

Bulk density (the mass of soil solids per unit volume) is the most direct and easy-to-measure indicator of changes in compactness, but changes in packing state can be better quantified by total porosity or void ratio. The relation between porosity (especially macroporosity) and soil physical properties such as saturated and unsaturated hydraulic conductivity and gas diffusivity is much closer than the relation of these soil physical properties with bulk density. Compaction causes major reductions in macroporosity ($>50\,\mu$m), reduction of total porosity, and often an increase in microporosity, resulting in a major impact on soil physical properties. Relative terms for packing state (e.g., relative density, degree of compactness) enable the same threshold values to be found for overcompaction for soils with different texture.

Hydraulic Properties

Saturated hydraulic conductivity is very sensitive to the compaction process by wheels that includes shearing and kneading of the soil. Figure 3 shows that wet soils are easier to compact and that the effect on the saturated hydraulic conductivity is more than proportional. The destruction of soil structure under wheels explains the greater degradation of soil qualities than observed in uniaxial tests.

In the process of compaction, the macroporosity decreases, whereas the microporosity often increases. This results in larger water contents for a wide range of matric potentials in compacted versus uncompacted soil. This results in a higher unsaturated hydraulic conductivity of a compacted soil versus uncompacted soil. However, near saturation meso- and macropores are also filled and contributes to the hydraulic conductivity, and thus the hydraulic conductivity of an uncompacted soil becomes higher than that of a compacted soil.

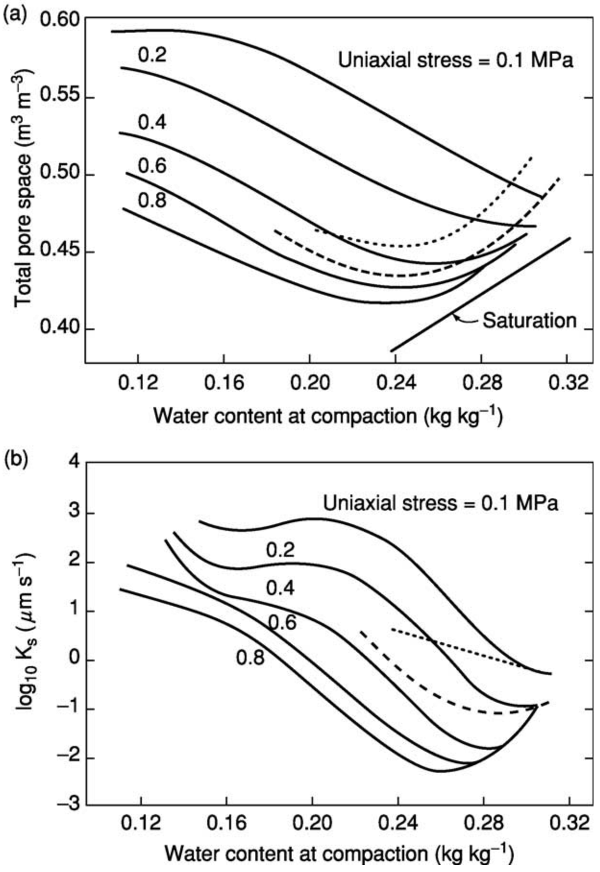


Figure 3 Changes in (a) total porosity and (b) saturated hydraulic conductivity of a sandy clay loam subjected to field traffic with low ground pressure (..... LGP, tire inflation pressure 80 kPa) or high ground pressure (- - - HGP, tire inflation pressure 240 kPa) or subjected to uniaxial stresses of 0.1, 0.2, 0.4, 0.6, and 0.8 MPa at various water contents. Solid lines refer to compression of aggregated mixtures. (After Dawidowski and Lerink, 1990.) Reproduced with permission from Horton R, Ankeny MD, and Allmaras RR (1994) *Soil Compaction in Crop Production.* Amsterdam: Elsevier.

Aeration Characteristics

An essential feature of compaction is reduction in the air-filled pore space. This property is an indication of the aeration status of the soil for plant growth and microbial activity. At air-filled porosity values less than 10%, oxygen deficiency is likely, especially in warm weather, while values less than 5% probably indicate incipient anaerobiosis. Analogous to saturated hydraulic conductivity, aeration status depends strongly on structure and continuous macropores. In a poorly structured compact soil, the threshold values of 10% and 5% should be doubled. In a well-structured soil, these threshold values may be halved. Other indices of aeration status, such as oxygen diffusion rate, oxygen content of soil air, and air permeability, provide more precise indicators of the aeration level.

Strength Characteristics

In most cases, the strength of a soil is increased by compaction. However, a well-structured soil is stronger than a poorly structured soil and a wet soil is weaker than a dry soil. In periods with wet-weather conditions, compacted poorly structured soils tend to become wetter than uncompacted well-structured soils due to limited infiltration capacity, lower saturated hydraulic conductivity, and high microporosity of compacted soils. In these circumstances, the strength and trafficability of compacted soils become lower than that of well-structured soils. After a wet period, a compacted soil will stay wet longer with limited workability. After a dry period, the strength of a compacted soil will increase considerably and result in high trafficability. Soil tillage then requires more powerful equipment and a higher energy input to loosen the soil and reduce clods to acceptable sizes.

Compaction increases the penetration resistance for roots, resulting in limited rooting depth and reduced crop growth. At penetration resistances (measured with a cone with a diameter of 12.7 mm and top angle of 30 degrees) of 1.5 MPa and 3.0 MPa, root growth rates are reduced to 50% and 0% respectively. However, in well-structured soils, roots can make use of continuous macropores to penetrate deeply into the soil. In drying soils, strength and penetration resistance increase.

Compaction in Crop Production

Compacted soils usually have unsatisfactory physical conditions for plant growth, but the extent of loss of productivity depends on soil type, plant species, and weather conditions.

Effects on Germination and Establishment

Compact soils result in cloddy seedbeds after primary tillage, poor soil/seed contact, and reduced germination. Soils may develop a compact surface layer due to crusting after heavy rainfall, which has sufficient strength to restrict or even to inhibit seedling emergence, especially of dicotyledonous species.

Effects on Root Growth and Distribution

Macropores (>50 μm diameter), through which roots can generally proliferate readily, are much reduced in compacted soil. As a result, root growth is restricted or even inhibited. In Figure 4 the limitations to root growth are conceptualized as relations between soil porosity and soil water potential at which soil aeration and mechanical resistance meet specified root requirements. In a soil with a certain pore volume, the soil water suction determines whether root growth is limited by too high mechanical resistance, by aeration problems, or is not limited. In Figure 4 the two thin lines show the situation in the case of a poorly structured soil. A poorly structured soil with few continuous macropores needs more air-filled pores for aeration than a well-structured soil. Also much lower mechanical resistances are allowed in a poorly structured soil than in a well-structured soil because roots can follow the macropores in the well-structured soil.

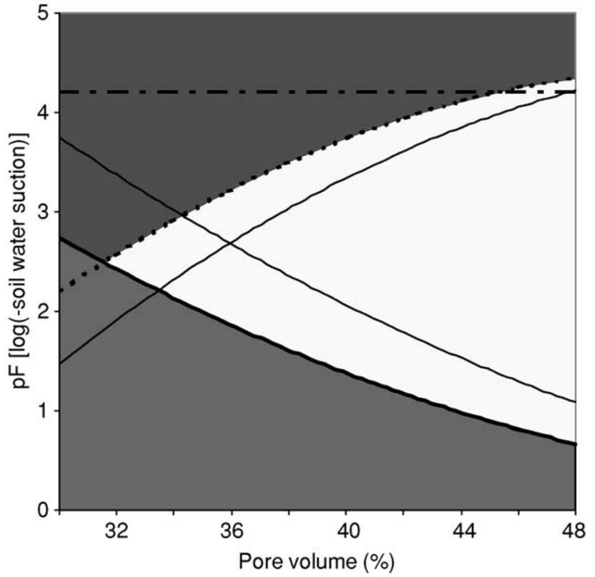


Figure 4 A conceptual relationship between soil porosity and soil water potential in which soil aeration and mechanical resistance (PR, penetration resistance) meet specified root requirements. Root growth is insufficient in the shaded areas and impossible beyond a soil water potential of 1600 kPa. Added to the figure are the same relationships if the structure is deteriorated resulting in a strong reduction of macropores. ▬ PR too high; ▭ too wet, aeration too low; □ rootable; - - - - PR limiting; ——— poorly structured soil; ━━━ aeration limiting; — - - too dry. (After Boone, 1988.) Reproduced with permission from Lindstrom MJ and Voorhees MB (1994) *Soil Compaction in Crop Production.* Amsterdam: Elsevier.

Effects on Plant Growth and Yield

The level of compaction (often expressed in dry bulk density) influences the growth, yield, and quality of crops, depending on crop species, soil type, and weather conditions (Figure 5). Optimum level of compaction tends to be higher for sandy soils, in dry seasons, and for monocotyledonous species. At a compaction level less than optimum, crops suffer from reduced soil/root contact, reducing germination and nutrient transfer, while at a compaction level higher than optimum, root growth and aeration are restricted and denitrification can lead to N losses.

Soil biota and biological processes are influenced by soil compactness. This is partly because of the influence of pore size distribution on spatial habitats for bacteria and fungi and partly because compacted soils may suffer from anaerobiosis, which in turn affects microbial metabolism markedly. The burrowing abilities of soil fauna, particularly earthworms, are reduced in compacted soils.

Interactive Crop Responses to Compaction and Fertilizer Application

Where farmers perceive the growth of crops to be adversely affected on compacted soils, a common action is to apply additional N fertilizer. However, the growth responses obtained may be less significant than those that would accrue at lower levels of compactness (Figure 6), and the additional nitrogen applications may be inefficient economically and detrimental to the environment.

Crop Responses to Subsoil Compaction

Amelioration of subsoil compaction is much more expensive and less effective than loosening compacted topsoils. Avoidance of compaction in the subsoil is therefore a much greater need than in the topsoil. Seventeen years after a single full-field compaction action, crop and nitrogen yield losses can still be significant on the compacted sites (Figure 7). After this compaction experiment only moderate axle loads were allowed on the fields. In practice, most subsoil in agricultural fields are partly compacted every year, and the subsoil quality is often worse than in these long-term experiments. Notice that the compaction effect was more pronounced in all years for harvested nitrogen than for grain dry matter. The strong decline of the effect in the first few years was mainly due to

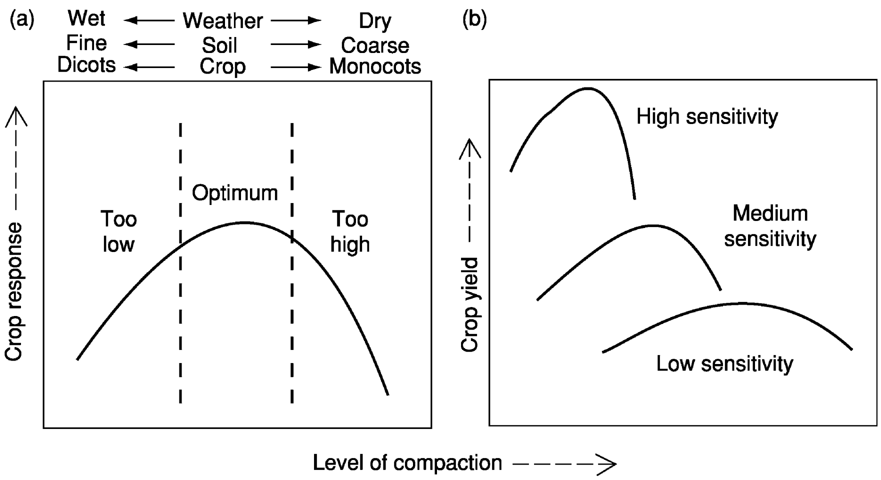


Figure 5 Conceptual relationship of crop response to level of soil compaction in relation to weather, soil texture and (a) crop type and (b) crop sensitivity. Reproduced with permission from Lindstrom MJ and Voorhees WB (1994) *Soil Compaction in Crop Production*. Amsterdam: Elsevier.

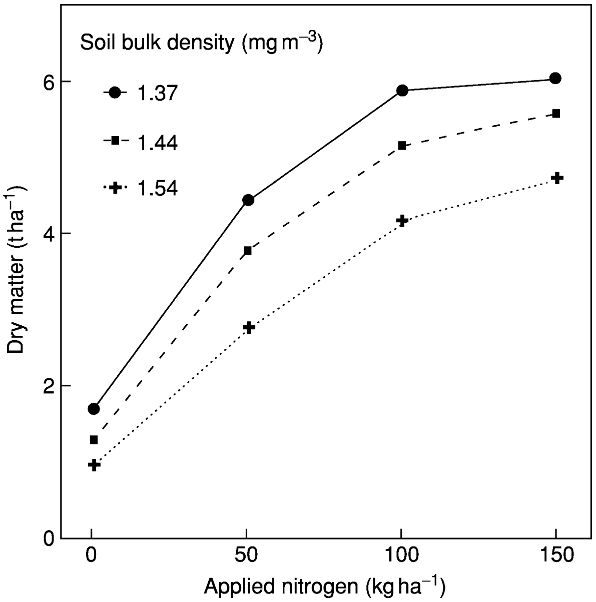


Figure 6 Dry-matter yields of ryegrass (first cut, mean of 3 years) in response to applied nitrogen fertilizer at three levels of bulk density (3–12 cm depth). Adapted from data in Douglas JT and Crawford CE (1993) *Grass Forage Science*. Oxford: Blackwell Science.

the recovery of the topsoil. The effect was influenced by rainfall during the subsequent growing seasons. In dry seasons a moderately compacted subsoil may result in yield advantages due to reduced loss of soil water percolating below and out of the reach of the root zone, whereas in wet seasons, the reduced permeability of the subsoil can lead to anaerobic conditions within the topsoil, resulting in direct damage to the crop and loss of nitrogen by denitrification.

Modeling of Crop Responses to Soil Compaction

Crop growth is less than potential when the uptake of water, oxygen, or nutrients is less than the demand of the crop. Potential crop growth is determined considering the prevailing weather conditions. Reduced crop growth may be caused by reduction of the length of the growing period, low temperature, limited supply from the soil of water, oxygen, and nutrients to the root system, and a limited activity of the root system. Soil water plays a central role in these limiting factors, and effects of soil compaction on crop growth and biological functioning should be modeled in relation to water. In Figure 8 a scheme is presented of the interrelationships among soil tillage, field traffic, soil structure and soil physical, chemical, and biological properties. In Figure 9 a part of this scheme is considered in more detail. To simulate crop responses to soil compaction all aspects presented in Figures 8 and 9 should be included in the model. However, up to now no model exists that includes all aspects, such as limitations of root growth, the role of macropores, reduced availability of nutrients (e.g., loss of nitrogen by denitrification), and effects of reduced biological activity. The existing models in general underestimate the impact of compaction on crop growth.

Effects on Environmental Components

Soil compaction influences a number of environmental parameters even at considerable distance from the original location at which the compaction occurred (Figure 10). Compaction may change the fluxes of greenhouse gases from the soil to the atmosphere

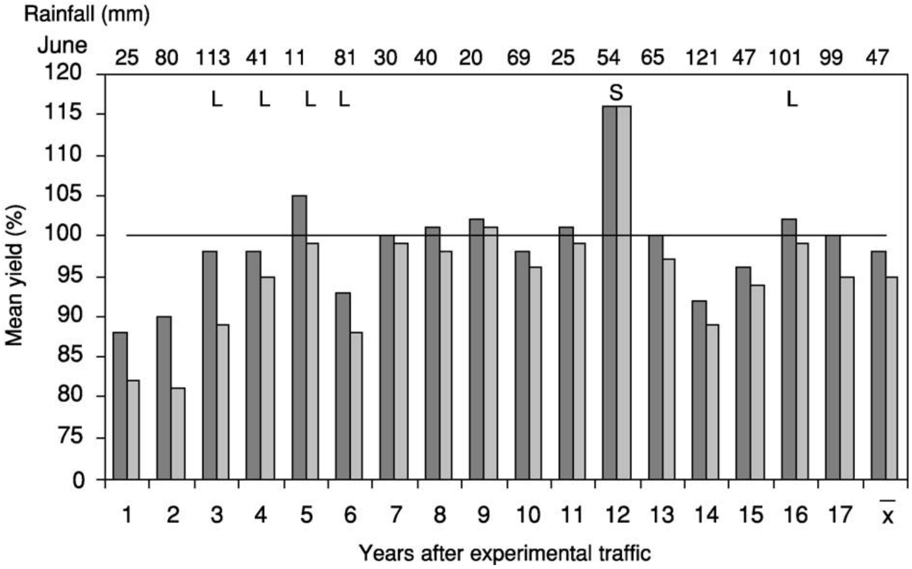


Figure 7 Mean grain and nitrogen yields of annual crops in control treatment (= 100%), and relative to the control in loading treatment of four passes with a 50 kN axle load in 1981 on clay soil, for 17 successive years after the loading. L, lodging; S, sprouting; ■ relative grain yield (%); □ relative nitrogen yield (%). Reproduced with permission from Alakukku L (2000) *Advances in GeoEcology* 32. Reiskirchen: Catena Verlag.

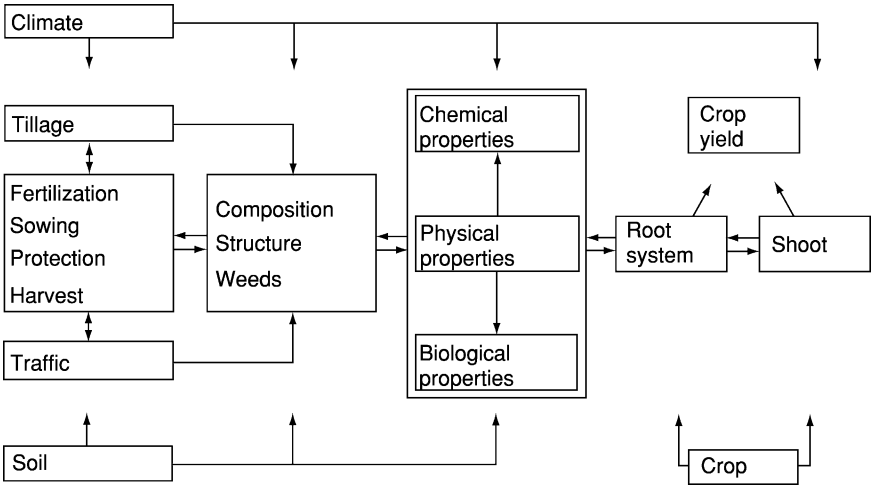


Figure 8 Interrelationships among climate, soil management, soil properties, and crop growth. Reproduced with permission from Boone FR (1988) *Soil and Tillage Research*. Amsterdam: Elsevier.

through mechanisms associated with effects on soil permeability, aeration, and crop development. Compaction increases CO_2 emissions because cultivation of compacted soils requires appreciably more energy than cultivation of uncompacted soils. Approximately 90% of the global N_2O emissions to the atmosphere comes from soils. Compacted soils tend to be wetter than noncompacted soils and denitrification is enhanced. The flux of N_2O increases rapidly as air-filled porosity declines (Figure 11). Compaction from vehicle traffic prior to the establishment of cereal crops can cause marked increases in the N_2O flux during the early growth period in spring (Table 2).

Due to reduced permeability, compacted soils usually show greater runoff and hence greater erosion than noncompacted soils. Surface rills and even gullies are sometimes directly associated with wheel tracks, particularly over seedbeds following periods of highly erosive rainfall. Surface waters may thus carry additional burdens of clay and silt, fertilizer, and pesticides when runoff occurs from areas of compacted soil.

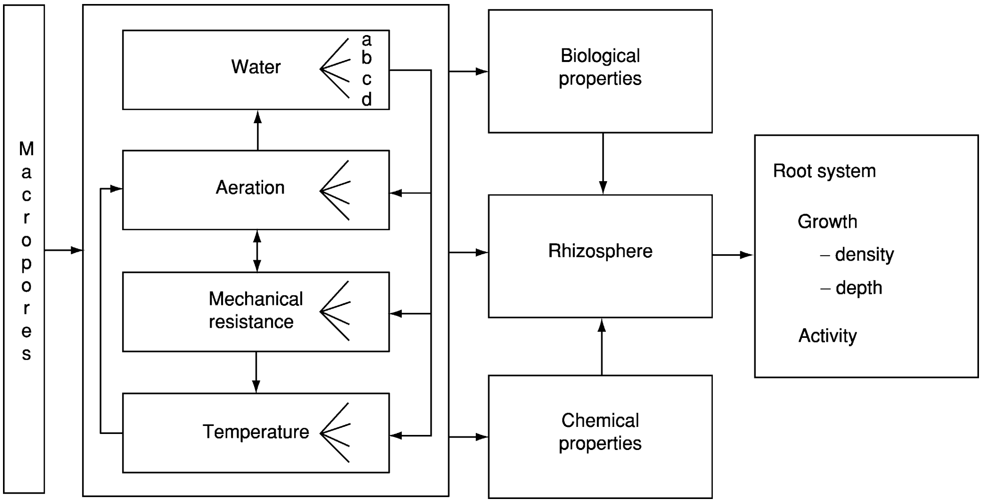


Figure 9 Relationships among soil macropores, soil physical, chemical, and biological properties, rhizosphere and root system. a, surface boundary; b, storage; c, transport; d, sink or source aspects. Reproduced with permission from Boone FR (1988) *Soil and Tillage Research*. Amsterdam: Elsevier.

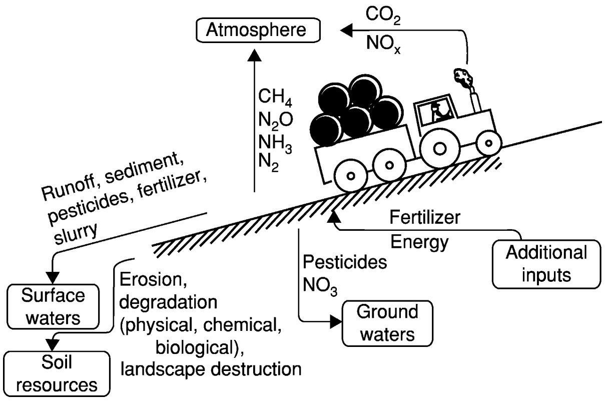


Figure 10 A conceptual diagram showing the major pathways whereby compacted soil conditions may influence components of the environment. Reproduced with permission from Soane BD and Van Ouwerkerk (1995) *Soil and Tillage Research*. Amsterdam: Elsevier.

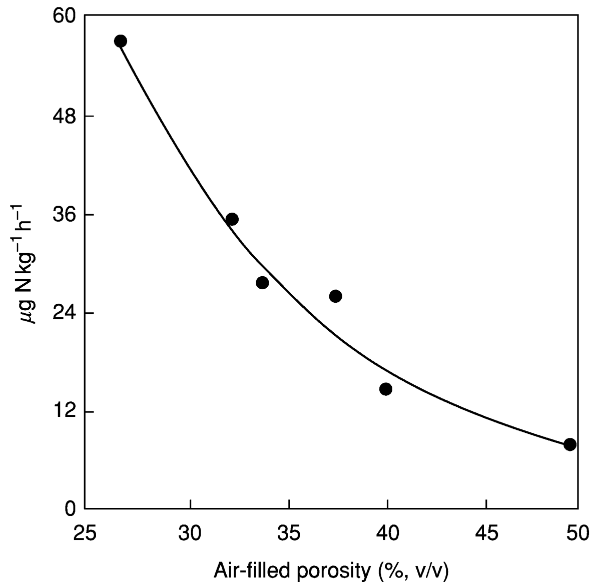


Figure 11 Increased emission of nitrogen by denitrification as air-filled porosity decreases. Reproduced with permission from Sextone AJ, Parkin TB, and Tiedje JM (1988) *Soil Biology and Biochemistry*. Amsterdam: Elsevier.

Table 2 Influence of vehicle traffic (zero, light, heavy) prior to the establishment of wheat and spring barley on N_2O flux during the early spring growth period.

		Cumulative N_2O flux ($g\,N_2O\text{-}N\,ha^{-1}$)		
Crop	*Period (dates and days)*	*Zero*	*Light*	*Heavy*
Spring barley	16 May–14 July = 60 days	320	310	401
Winter wheat	8 March–8 May = 62 days	245	210	578

Adapted from data in Ball BC, Parker JP, and Scott A (1999) *Soil and Tillage Research*. Amsterdam: Elsevier.

Techniques for the Reduction of Compaction

The compactive capability of vehicles can be minimized by reductions in overall mass, increases in ground contact area, and reduction of ground contact pressure. Tires of greater width and diameter will increase ground contact area, as will reduction in inflation pressure. Dual wheels, multiple axle systems, especially tandem axles for trailers and tankers with low-pressure tires, provide extra contact area.

Vehicle traffic can be reduced by combining different operations into a single-pass operation. Toolcarriers (gantries), up to 12 m width, have been found capable of providing traffic-free zones in which, because of the lack of compaction, tillage requirement is reduced and crop yield and quality may be improved.

Trafficking and working soil when it is too wet should be avoided. Tillage should be reduced as much as possible to optimize biological and physical processes that improve soil structure, in particular macroporosity.

Improving drainage will result in drier and stronger soils. Increases in soil organic matter, either as a surface mulch or incorporated, will increase soil elasticity and reduce compaction.

Amelioration of Compacted Soils

Natural weathering due to freezing and thawing is usually restricted to the top few centimeters of the surface soil and rarely penetrates to the subsoil. Swelling and shrinking arising from changes in water content can cause the loosening of compacted soils. Roots of certain species can penetrate compacted layers and soil biota can slowly increase macroporosity. In this way, gradual improvements can be made in the soil physical quality of compacted layers. However, highly compacted parts, in particular layers that cannot be penetrated by roots, will not or only slowly recover.

Compacted topsoils can be loosened by tillage. However, loosened compacted soils are cloddy and require additional secondary cultivation to achieve suitable seedbed tilth. Loosening subsoils requires high-draft special equipment and high traction, destroys still existing continuous macropores, and weakens soil structure and strength. Subsoiling should be done when the subsoil is dry with a minimum of loosening and with the aim to form cracks (*see* **Subsoiling**). Many loosened subsoils recompact within a few years with worsened soil physical properties and rootability. The result is that subsoils that have been loosened once must be loosened regularly every 4–5 years. Therefore the subsoil should be inspected beforehand to determine rootability, aeration status, and drainability, to avoid unnecessary subsoiling.

See also: **Conservation Tillage**; **Cultivation and Tillage**; **Stress–Strain and Soil Strength**; **Structure**; **Subsoiling**

Further Reading

Håkansson I (ed.) (1994) Subsoil compaction by high axle load traffic. *Special Issue of Soil Tillage Research* 29: 105–306.

Horn R, van den Akker JJH, and Arvidsson J (eds) (2000) Subsoil compaction: distribution, processes and consequences. *Advances in GeoEcology* 32.

Koolen AJ and Kuipers H (1983) Agricultural soil mechanics. *Advanced Series in Agricultural Sciences* 13.

Larson WE, Blake G, Allmaras RR, Voorhees WB, and Gupta S (eds) (1989) Mechanics and related processes in structured agricultural soils. *NATO ASI Series E: Applied Sciences* 172.

Monnier G and Goss MJ (eds) (1987) *Soil Compaction and Regeneration*. Rotterdam, Netherlands: A. A. Balkema.

Oldeman LR, Hakkeling RTA, and Sombroek WG (1991) *World Map of the Status of Human-induced Soil Degradation, an Explanatory Note*. Wageningen, Netherlands: ISRIC.

Pagliai M and Jones RJA (eds) (2002) Sustainable land management–environmental protection: a soil physical approach. *Advances in GeoEcology* 35.

Soane BD and Van Ouwerkerk C (eds) (1994) *Soil Compaction in Crop Production*. Amsterdam: Elsevier.

van den Akker JJH, Arvidsson J, and Horn R (eds) (2003) Experiences with the impact and prevention of subsoil compaction in the European Community. *Special Issue of Soil Tillage Research* 73: 1–186.

Van Ouwerkerk C (ed.) (1988) Tillage and traffic in crop production. *Special Issue of Soil Tillage Research* 11: 197–372.

Van Ouwerkerk C and Soane BD (eds) (1995) Soil compaction and the environment. *Special Issue of Soil Tillage Research* 35: 1–113.

COMPOST

T L Richard, Iowa State University, Ames, IA, USA

Introduction

Compost has been used to improve agricultural soils for hundreds of years, but only in the last few decades have we begun to understand the science behind this practice. This increase in scientific research parallels a dramatic growth in compost use, motivated both by growing demand from organic and conventional farmers, landscapers, and home gardeners, and also by a burgeoning supply of compost from municipal, industrial, and agricultural waste treatment. The resulting diversity of compost feedstocks, processing technologies, and marketing and utilization strategies results in a wide range of compost qualities and characteristics. Each compost has its own attributes, so that generalizations about compost behavior must be fine-tuned with knowledge of the particular compost and application.

When a compost is applied to soil, it initiates a cascade of changes in soil physical, chemical, and biological properties. While the direction of many of these changes is predictable based on soil and compost characteristics, their magnitude varies as a result of the dynamic interactions between these two diverse and biologically active porous media. Understanding these complex interactions is important for both scientific advances and practical management of compost in agronomic and horticultural systems.

Compost Production

Just as soil genesis impacts the structure and function of soil, the composting process plays a key role in the subsequent structural and functional attributes of compost. Compost can be produced at scales ranging from small backyard piles to large industrial factories, with management ranging from benign neglect to intensive engineering. Most composting attempts to encourage aerobic decomposition, and includes an initial thermophilic period that lasts a few days to a few weeks. However, there are exceptions to even this basic characterization of composting as an aerobic, thermophilic process. Anaerobic composting, a form of high-solids anaerobic digestion, is used to produce methane gas as well as stabilize various organic wastes. And ambient temperature processes include both sheet composting, where thin layers of organic material are applied to the soil surface, and vermicomposting, where worms are involved in the stabilization process.

The fundamental requirements for composting are relatively simple: fresh organic substrate, a diverse microbial population, adequate aeration and moisture, and a strategy for managing excess heat. The substrate provides readily biodegradable energy and nutrient sources for the microorganisms, which in the presence of oxygen and moisture will catabolize the substrate to sustain their populations and support active biomass growth. Enzymatic hydrolysis decomposes simple and complex carbohydrates, crude protein, and lipid molecules, some of which are used for microbial growth, while others are mineralized. If oxygen is available the mineralized carbon will be released almost entirely as carbon dioxide, with the hydrogen and oxygen released as water.

Aerobic decomposition releases large amounts of energy, about $14\,\mathrm{kJ\,g^{-1}}$ O_2 consumed, and this energy heats the pile to well above ambient temperature. Higher temperatures increase the rate of microbial decomposition, with the maximum rate typically observed around 60°C. At this temperature degradation rates can be five or more times as fast as at 20°C. This thermophilic phase of composting destroys most weed seeds and pathogens, with pathogen reduction requirements for regulated feedstocks or products typically ranging from 3 to 15 days at 55°C or above. These time–temperature combinations have been shown to reduce many types of human, animal, and plant pathogens to below infectious levels and, when combined with adequate stabilization of the feedstocks, prevent pathogen regrowth in the finished product. Although higher temperatures destroy pathogens more quickly, temperatures above 65–70°C cause a steep decline in reaction kinetics and a dramatic reduction in microbial diversity. The combination of advantages and risks associated with operating in the optimum thermophilic range make temperature management a critical factor in controlling the composting process.

Temperature management and oxygen supply are coupled because most heat is removed by the convective flow of oxygen-supplying air, and because the airflow rates needed for adequate heat removal normally provide sufficient oxygen to maintain aerobic conditions. Most composting systems rely on passive aeration by wind, diffusion, and natural convection. Other systems use mechanical aeration, often with temperature-feedback controllers operating a system of blowers and fans. In either case adequate porosity and permeability of the pile are necessary for uniform distribution of air, as excess moisture or compaction leads to anaerobic pockets where odorous gases can form. This requirement places practical limits on pile dimensions and moisture content, with systems greater than 2–3 m in height or less than 40% solids potentially at risk.

Degradation follows first-order kinetics after a short initial lag period, which can be minimized by recycling mature compost as a microbial inoculant. Oxygen demand and heat production are greatest during the first few days and weeks, when degradation rates are at their maximum. If oxygen becomes limiting during this period anaerobic fermentations form organic acids, which accumulate and result in a measurable but usually ephemeral reduction in pH. These acids normally degrade aerobically as oxygen becomes available, but if they persist their phytotoxicity can affect the compost product. During the thermophilic phase, compost can lose considerable quantities of nitrogen, with overall losses ranging from less than 10% to greater than 60% of the initial mass of N. These losses include volatilization of NH_3, N_2, and N_2O, and in outdoor sites can also include runoff and leaching of NH_4^+ and NO_3. Total N losses are inversely correlated with C:N ratio, and are minimized for C:N ratios greater than 30:1. The ammonia volatilization component of total N loss has been positively correlated with increasing airflow rates, temperatures, and pH.

After a thermophilic period ranging from a few weeks to months, temperatures decline to the mesophilic range as readily degradable compounds become increasingly scarce. The compost then begins a curing or maturation phase, which is characterized by continuing decomposition of hemicellulose, cellulose, and other slowly degradable fractions, reductions in particle size, slowing respiration rates, and increasing humic acid content. Lower temperatures can facilitate nitrification, which increases the plant-available N but can also denitrify and increase N losses. Over extended curing periods of 6 months to 2 years, nitrate can accumulate to a significant fraction of the total compost N. Figure 1 shows these N transformations over time during the composting process.

Compost will continue to stabilize indefinitely, but is often marketed and applied to soil when cumulative carbon mineralization is in the range of 60–70%. Depending on the original bulk density and ash content, as well as factors such as soil or inert matter incorporated during composting, the total volume reduction may be in this same range or somewhat higher, while dry-matter reduction may be somewhat less than 60%. P, K, and other mineral elements are normally largely conserved, although K and other soluble elements can leach from systems in outdoor piles. For all the reasons mentioned above, macronutrient contents vary widely in different compost products, as shown in Table 1. Moisture content can also vary from product to product, although the final moisture content from industrial composting systems is often around 40% (wet basis), which allows easy application without excessive dust.

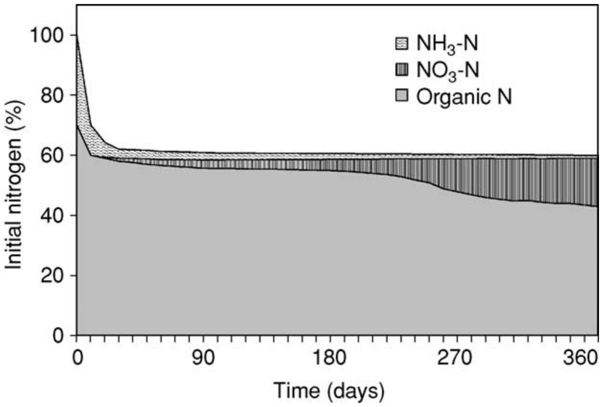


Figure 1 Progressive changes in N partitioning during the composting process. In this example total N loss during composting was 40%.

Compost Application and Incorporation

For any particular application, appropriate compost application strategies are a function of both a particular compost's characteristics and specific soil-management goals. Compost application rates vary widely, from as little as 1 Mg ha^{-1} for a nutrient-rich compost in agronomic applications, to more than 60 Mg ha^{-1} for remediation of disturbed sites or severely depleted soils. Compost can also be used to manufacture artificial soils, for landscaping, for ecologic restoration of mines, brownfield sites, and wetlands, or in high-value commercial media such as nursery or greenhouse potting soils. The compost fraction used in these blends is often in the 10–50% range, with higher ratios possible in artificial wetlands and some other specialized situations. Although most of the benefits of compost increase with application rate, some caution must be exercised to insure that both annual rates and cumulative applications are not excessive with respect to nutrients, trace metals, problematic biomolecules, and potential phytotoxic effects. Heavy metal and nutrient loading rates are sometimes regulated to prevent negative environmental impacts, while phytotoxicity concerns can be addressed by insuring adequate compost maturity, sometimes combined with intentional leaching to remove soluble salts.

Table 1 Ranges of nutrient composition in compost products (multisite means in parentheses)

Compost type	*N (g kg^{-1} (dry weight))*	*P (g kg^{-1} (dry weight))*	*K (g kg^{-1} (dry weight))*
Livestock manure	11–37 (20–31)	3–20 (9–20)	2–27 (15–26)
Yard trimmings	5–42 (6–21)	0.5–12 (2–3)	0.4–27 (3–10)
Biosolids	7–38 (18–35)	5–41 (7–21)	0.7–10 (4–5)
Municipal solid waste	4–17 (7–13)	0.8–6 (2–3)	2–10 (3–5)

Compost is normally applied to soil using equipment and techniques designed to handle solid manure, although specialized equipment is available for targeted applications. With drier, more friable composts, spinning and rotating brush spreaders can be used on cropland and turf, while blower hoses can be used for landscaping and erosion control. As with manure, the application rate and degree of incorporation affect the potential for short-term environmental impacts through volatilization, leaching, and runoff, as well as long-term dynamics as the compost interacts with the soil. Surface application of compost is less problematic than with manures, as nutrients are in more stable forms, and a surface layer of compost increases infiltration rates, reduces erosion, and in thicker layers can suppress weed growth. However, incorporation is also common in both agricultural and horticultural applications to distribute the compost product throughout the surface horizons.

Compost–Soil Interactions

The application of compost to soil initiates a series of physical, chemical, and biological transformations that affect both soil properties and processes. Some of these effects are strong initially but decrease over time, others are more persistent, while some are latent and may only manifest under certain soil, crop, or climatic conditions. Table 2 lists some of the most important impacts of compost on soil and indicates the typical direction of change.

Table 2 Results of compost–soil interactions

	Typical change
Physical	
Infiltration	Increases
Erosion	Decreases
Aggregate stability	Increases
Water-holding capacity	Increases
Total porosity	Increases
Permeability	Increases
Bulk density	Decreases
Chemical	
pH	Buffers near neutral
Cation exchange capacity	Increases
Electrical conductivity	Increases
Nutrient concentration	Increases
Nutrient availability	Varies
Trace elements and metals	Varies
Biological	
Microbial activity	Increases
Microbial biomass	Increases
Microbial diversity	Increases
Enzymatic activity	Increases
Phytotoxicity	Varies
Phytostimulation	Increases
Plant disease suppression	Varies

Physical Effects

Compost affects several critical soil physical functions, including water transport and storage, gas exchange, and heat transfer. At the soil surface, compost provides a barrier to raindrop impact, reducing surface sealing and allowing rapid infiltration of rainfall or irrigation water. These factors can delay the onset of runoff dramatically, with 5-cm-thick blanket applications entirely eliminating runoff and thus erosion in all but the most severe storms. When extreme events do occur, compost blankets can reduce erosion by one to three orders of magnitude. These benefits occur immediately after application and thus protect the soil even prior to vegetation establishment.

Deeper in the soil profile, the structural effects of compost particles typically increase both soil porosity and permeability, enhancing gas transfer as well as saturated water flow. Total porosity can increase by 50%, while permeability can increase by up to two orders of magnitude at high compost application rates. However, these potential porosity and permeability gains can be compromised by excessive compaction under wet conditions, when the mechanical strength of these particles is low. Depending on its organic matter content, compost particle density can range from 1600 to 2000 kg m^3, and if porosity has not been compromised this relatively low particle density will reduce the dry bulk density of the soil.

Over time, organic compounds introduced with the compost and produced by subsequent biological activity work to promote soil aggregation and enhance aggregate stability. These soil aggregates, as well as the compost itself, increase water-holding capacity, which can improve irrigation efficiency and mitigate drought stress. These effects occur not just in the surface layers where compost is typically incorporated, but also in subsurface layers where soluble organic carbon can penetrate by leaching from above. As with other organic amendments, the compost encourages earthworm activity, generating biological macropores, which facilitate deep percolation and drainage of excess moisture.

Many of these physical improvements associated with compost application peak within a few weeks or months after compost is applied and then slowly attenuate over time. While a single high-rate compost application (>20 Mg ha^{-1}) can have significant, lasting effects on soil physical properties, these are more commonly achieved by repeated compost applications over a period of years.

Chemical Effects

When compost is incorporated in soil, there are immediate, calculable changes in concentrations of nutrients, trace metals, and other chemical compounds that result from the application rate and composition of the two materials. However, the organic matter in compost also stimulates changes in biological activity of the soil that have longer-term impacts on soil chemical properties. These compost–soil dynamics are complex, and few studies have characterized the precise mechanisms or kinetics of the resulting chemical transformations and fate. Nonetheless, there are several areas where current understanding can provide important insights and practical guidance.

Carbon is the dominant element in composts, primarily occurring as the structural backbone of organic biomolecules. These biomolecules have open ion exchange sites that can react with other compounds in the soil system, increasing the overall cation exchange capacity (CEC) and chelating minerals and heavy metals in the soil. Some composts also include significant amounts of mineral carbonates, either from feedstocks or accidentally blended in from the surface of compost pads.

Both carbonates and organic carbon in compost can affect the soil pH. Carbonates raise pH through their affinity for hydrogen ions, while organic carbon tends to buffer the system somewhere near neutral. Addition of an immature or anaerobic compost can temporarily lower pH by introducing organic acids, although these degrade rapidly if the soil system is aerobic. In general, mature composts will modify soil pH toward neutral to slightly alkaline levels (pH 6.8–7.8), with larger increases typically associated with composts made with biosolids or other wastewater sludges that have had lime added during dewatering or other treatment processes.

One of the potential negative impacts of compost on soil is the addition of soluble salts. This is a particular concern with manure and biosolids-derived composts but is also possible with composts that include particular food-processing feedstocks and even leaves from municipalities where salt is used for roadway deicing. Electrical conductivity (EC) provides a measure of salt concentration, with levels of less than 4 dS m^{-1} in composts or 1.2 dS m^{-1} in soil–compost mixtures tolerated by all but the most sensitive plants. In situations where high application rates of a high-EC compost may cause problems, leaching by either irrigation or natural rainfall can reduce levels significantly. However, some of the soluble ions that register as EC are also beneficial minerals and nutrients that will be removed by excess leaching.

The CEC of soils facilitates retention of positively charged minerals and nutrients (Ca^+, K^{2+}, NH_4^+, etc.) for subsequent use by microorganisms and plants. Organic matter, along with clay minerals, is known to have a positive impact on CEC, with humic and fulvic acids playing a particularly important role. While the compost application rates common in agronomic situations are generally not high enough to result in measurable increases in soil CEC, significant increases are observed at high application rates. These increases complement the increased water-holding capacity of compost-amended soils, reducing the potential for minerals and nutrients to leach out of the plant root zone.

Nitrogen remains the critical nutrient in most agricultural and horticultural systems, so its availability from compost is of particular importance. Nitrogen dynamics in compost-amended soils can result in a wide range of outcomes, from net mineralization and a significant fertility boost to net immobilization resulting in crop nitrogen stress. The primary factor influencing this result is the compost C:N ratio, as excess carbon can stimulate microbial growth that immobilizes mineral N from both compost and soil sources. Other compost-specific factors include C and N availability, which can modify the impact of a particular C:N ratio, and particle size, as smaller particles will mineralize more rapidly. Soil conditions also have a strong influence on nitrogen mineralization, especially soil organic matter status, temperature, moisture, oxygen, salinity and pH, with optimum conditions similar to those in soil without compost. The higher biological activity associated with compost application can, however, cause negative perturbations. For example, unstable composts with high respiration rates can deplete oxygen levels in soil microenvironments, increasing denitrification rates and volatile nitrogen loss. The risks of potential negative impacts are minimized by use of stable compost products with a C:N ratio of less than 15:1.

When soil conditions are supportive of microbial growth, compost can prime the decomposition of existing soil organic matter and stimulate additional N mineralization, providing fertility benefits beyond those associated with the added compost itself. Similarly, synergistic effects can occur when both compost and synthetic fertilizers are applied to soil, with higher net N mineralization rates observed than with either amendment alone. A promising area of current research is in the synchronization of organic N mineralization with crop nutrient demand, which may demonstrate additional advantages for the use of compost as a component of soil-fertility management.

The interactions between these soil and compost factors result in variable mineralization rates and outcomes for nitrogen as well as other macronutrients. Figure 2 shows the range of mineralization rates for N, P, and K, reported in several experiments using different types of compost. Although there will be widespread exceptions to any generalizations from these specific compost–soil experimental results, some of the patterns represented in Figure 2 can be explained by mechanistic considerations. Many manure–crop residue composts are limited by available crop residue in the initial feedstocks, and thus tend to have high concentrations of nutrients (Table 1) that facilitate the relatively high mineralization rates for N and P (no data were found for K mineralization of manure composts). Municipal solid waste (MSW) has a relatively low N mineralization rate, reflecting the relatively high C:N ratio in typical MSW feedstocks. Biosolids, while they can have a high P concentration, are often processed in ways that bind the P to iron and calcium during wastewater treatment, resulting in reduced availability of this element. For any particular compost feedstock, transformations during the composting process (illustrated for N in Figure 1) will also affect subsequent mineralization rates. P availability may be enhanced by organic acids formed during the early stages of the composting process, and then later decline as the compost matures. If compost piles are exposed to precipitation, soluble K may leach out during the process, resulting in reduced availability for the K that remains. The wide range of observed mineralization rates, often varying by a factor of 3 or more, illustrates the need for additional tools and techniques to facilitate accurate and reliable compost nutrient management.

The difficulty of predicting initial compost mineralization rates is compounded as nutrients accumulate from repeated applications over a period of years. Although mineralization rates typically decline over time, the buildup of this pool of moderately available nutrients should also be considered in nutrient-management plans. Over time, mineralization from this 'nutrient bank' will increasingly substitute for current-year application requirements, as shown in the example presented in Figure 3. In this example, 20% of compost N mineralizes in the initial year, and 10% of the remainder mineralizes in each subsequent year. Initial applications must be heavy to supply the 150 kg N ha^{-1} required by a hypothetical crop, but over time these applications can drop off to a steady-state replacement rate as accumulated compost nitrogen from prior applications increases to greater than 1300 kg N ha^{-1}. Heavy initial compost requirements, followed by a gradual decline to steady state, are a common practice in agricultural and horticultural applications.

Environmentally sound compost nutrient management requires consideration of the balance among nutrients also. For many manure and biosolids composts, application at an N-based rate results in an overloading of P relative to crop demand. This imbalance is exacerbated relative to the original manure or sewage sludge feedstocks, because significant amounts of N are lost during composting, while P

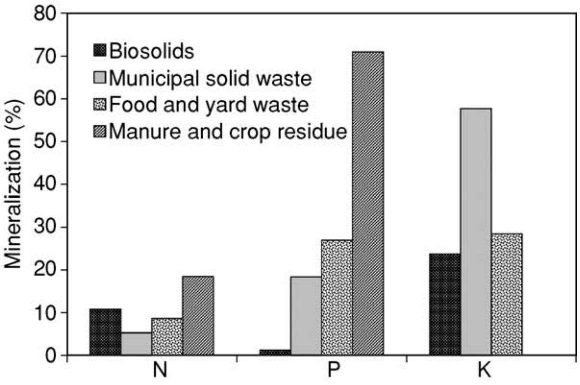


Figure 2 N, P, and K mineralization rates reported for various compost types.

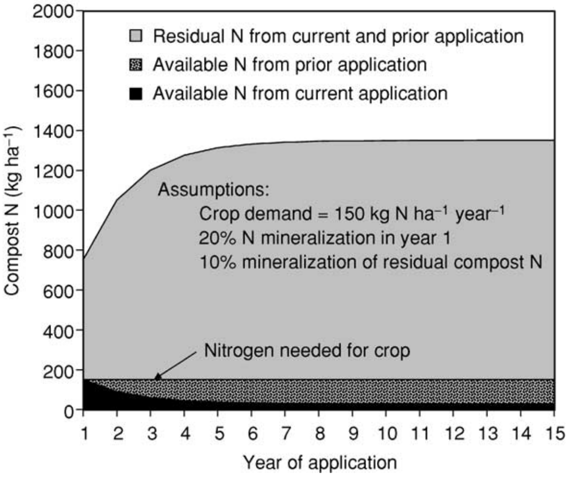


Figure 3 Changes in annual N supply and cumulative N reserve when compost is used to supply crop N requirements.

is largely conserved. This imbalance is particularly problematic on sites where the soil phosphorus levels are already elevated, either from prior manure use or synthetic fertilizer applications. Integration of nitrogen-fixing legumes or other alternative N sources in a crop rotation can be an excellent way to address this issue, while use of synthetic N fertilizers to supplement compost N can also work well. With either approach, compost application rates would then need to be reduced to match P demand, through either lower annual rates or alternate-year applications.

Heavy metals are present in all composts, but only sometimes at levels of toxicological concern. Governmental regulatory agencies have set concentration and cumulative loading limits for metals, particularly for composts whose feedstocks include biosolids and/or MSW, where concentrations of these elements are often higher than in source-separated organic feedstocks. Different metals pose varying types of ecosystem risks, from phytotoxic effects on sensitive plants to bioaccumulation in human and animal food chains. As with macronutrients, the bioavailability of these trace elements is strongly influenced by interactions with the compost and soil matrix. Bioavailability tends to decrease with time, due to chelation of metals and ligand formation. These binding and chelation capabilities of compost have been used to remediate soils with high heavy metal concentrations, and are particularly effective at lower pH. Compost can also be used to remediate toxic organic chemicals, both by enhancing biodegradation and by reducing their mobility and bioavailability.

Biological Effects

Several biological processes have already been mentioned, because they affect the physical and chemical characteristics of compost–soil systems. Organic carbon added through compost provides a major energy resource for microbial activity and growth, while the added nutrients and minerals cycle through both microorganisms and plants. These processes and others operate in multiple feedback loops to regulate nutrient availability, the microbial ecosystem, and plant growth and decay.

One of the dominant features of compost-amended soils is a significant increase in soil microbial activity, the extent of which depends on properties of both the compost and the soil. The microbial biodegradation that fuels this activity is indicated by the carbon mineralization rate, often measured in soil by CO_2 respiration studies. Several similar methods have been developed to measure and classify the degree of compost stability, and results are often available for commercial compost products. Although increased stability reduces the potential for odors and various phytotoxic compounds, it also reduces the carbon available for soil microbial processes. Thus while fairly stable composts may be required for residential markets, greenhouse applications, seedlings, and sensitive plants, agronomic applications may be better served by less-stable composts that can better stimulate microbial activity in the soil. The optimum degree of stability for any particular situation will depend on the application rate, soil system, and plants intended to be grown.

In addition to increasing overall microbial activity, compost also increases the activity of specific enzymes, the amount of active microbial biomass, and the overall diversity of the microbial ecology in soil–compost systems. These effects result from increases in both the amount and diversity of organic constituents introduced with the compost, including both partially decomposed organic matter and living microbial biomass.

Although most of the biological effects of compost are entirely positive, some composts can contain constituents that are detrimental to plants. Potentially phytotoxic compounds include short-chain organic acids, alcohols, and other organic compounds, ammonia, and the soluble salts and heavy metals previously described. Ammonia and the organic compounds are formed by protein degradation and anaerobic fermentations, respectively. Because these conditions are most likely during the early stages of composting, phytotoxicity is most commonly associated with immature composts, but can also occur with older composts that have been stored anaerobically, or that have high EC or heavy metal concentrations. These compounds have the strongest impact during the early stages in a plant's life cycle, with different plant species varying in their response. In general, smaller-seeded plants are more sensitive to the organic acids and phytotoxic biomolecules, while grasses are more sensitive to soluble salts. Laboratory bioassays have been developed to test for reduced seedling germination and root elongation, using species sensitive to specific types of effects. These bioassays can provide assurance to compost producers and users that a compost is safe to use on high-value, sensitive crops.

In weed control applications, the phytotoxicity associated with immature composts can potentially be put to positive use. Because many weeds are small-seeded, they are generally more sensitive to these phytotoxic compounds than large-seeded crops. Blanket surface applications of compost have also been shown to reduce weed emergence, primarily through the physical mulching effect, but biochemical interactions may contribute as well. However, once weeds germinate, the effect of compost on crop and weed competition becomes less well defined. While compost applications appear to reduce competition between crops and certain weed species, other weed species are luxury feeders on nutrients, and compost may provide these weeds a competitive advantage against certain crops. The use of compost in weed management remains an important and intriguing area for additional research.

The interaction between compost and plants is further complicated by the potential of composts to have phytostimulatory effects, beyond those caused by improvements in soil physical properties, moisture, or nutrient effects. Several biomolecules in compost have been shown to promote plant growth, including indole-3-acetic acid, humic and fulvic acids, and tend to increase in concentration as the compost matures. Humic substances have been linked with the following positive results: increased membrane permeability and nutrient uptake, improved photosynthesis and protein synthesis, enhanced enzyme activity, and other biochemical changes similar to those induced by plant growth regulators. While additional research is needed to confirm these results and explore possible mechanisms, such synergistic effects may help explain increases in plant growth and yield from compost treatments, at levels beyond those predicted by measured physical and chemical effects.

Another important interaction between compost, soils, and plants is in the area of plant disease suppression. Certain composts have been known for decades to suppress particular plant pathogens, including many of the soil-borne pathogens that plague nursery and greenhouse growers. The efficacy of different composts varies, with some having no effect while others perform even better than chemical treatments, depending on the compost, pathogen, and specific mechanisms of suppression involved. In some cases the mechanisms of disease suppression are narrow, such as when composts include specific microorganisms that are antagonistic to or parasitize particular plant pathogens, or somewhat broader when pathogens are suppressed by fungal antibiotic inhibitors or become less competitive when compost changes the nutrient status of the soil. Compost feedstock, processing strategy, and maturity can all contribute to these suppressive mechanisms, with an adequate but not excessive level of maturity particularly important in promoting biocontrol. In some cases biocontrol organisms are inoculated into compost, often just after the thermophilic period, when ecologic niches are relatively open. Certain composts can also induce systemic acquired resistance, stimulating the plant's immune system to ward off pathogens more effectively than it could before. This general mechanism of plant disease suppression is highly effective, suppressing even foliar diseases that have no direct contact with the compost or the soil. Relatively few composts produced today consistently induce systemic acquired resistance, making this a promising area for future research and development.

Compost can have a range of positive effects on soils, microbial communities, and plants. Although the complexity of these interactions and the diversity of compost types and behaviors pose continuing challenges for both scientific understanding and practical management, considerable progress has been made in recent years. With greater emphasis on ecologic management of both organic and conventional crop production, research on the interactions of compost with soil and crops is expected to expand. As more is understood about these physical, chemical, and biological mechanisms and relationships, compost is likely to play an increasing role in improving soil tilth, soil fertility, and managing plant growth and disease.

List of Technical Nomenclature

EC	Electrical conductivity (dS m^{-1})
MSW	Municipal solid waste

See also: **Organic Farming**; **Organic Matter:** Genesis and Formation; **Organic Residues, Decomposition**

Further Reading

Avnimelech Y, Kochva M, Yotal Y, and Shkedy D (1992) The use of compost as a soil amendment. *Acta Horticulturae* 302: 217–236.

Dick WA and McCoy EL (1993) Enhancing soil fertility by addition of compost. In: Hoitink HAJ and Keener HM (eds) *Science and Engineering of Composting*, pp. 622–644. Worthington, OH: Renaissance Publications.

Epstein E (1997) *The Science of Composting*. Lancaster, PA: Technomic Publishing.

Fraser DG, Doran JW, Sahs WW, and Lesoing GW (1988) Soil microbial populations and activities under conventional and organic management. *Journal of Environmental Quality* 17: 585–590.

Hoitink HAJ, Krause MS, and Han DY (2001) Spectrum and mechanisms of plant disease control with composts. In: Stoffella PJ and Kahn BA (eds) *Compost Utilization in Horticultural Cropping Systems*, pp. 263–273. Boca Raton, FL: Lewis Publishers.

Inbar Y, Chen Y, and Hadar Y (1990) Humic substances formed during the composting of organic matter. *Soil Science Society of America Journal* 54: 1316–1323.

Kashmanian RM, Kluchinski D, Richard TL, and Walker JM (2000) Quantities, characteristics, barriers, and incentives for use of organic municipal by-products. In: Power JF and Dick WA (eds) *Land Application of Agricultural, Industrial, and Municipal By-Products*, pp. 127–167. SSSA Book Series No. 6. Madison, WI: Soil Science Society of America.

Kuter GA, Blackwood KR, Diaz LF *et al.* (1995) *Biosolids Composting*. Alexandria, VA: Water Environment Federation.

Liebman M and Mohler CL (2001) Weeds and the soil environment. In: Liebman M, Mohler CL, and Staver CP (eds) *Ecological Management of Agricultural Weeds*, pp. 210–268. Cambridge, UK: Cambridge University Press.

Miller FC (1991) Biodegradation of solid wastes by composting. In: Martin AM (ed.) *Biological Degradation of Wastes*, pp. 1–31. London, UK: Elsevier Applied Science.

Ozores-Hampton M, Obreza TA, and Stofella PJ (2001) Weed control in vegetable crops with composted organic mulches. In: Stoffella PJ and Kahn BA (eds) *Compost Utilization in Horticultural Cropping Systems*, pp. 275–286. Boca Raton, FL: Lewis Publishers.

Rynk R and Richard TL (2001) Commercial compost production systems. In: Stoffella PJ and Kahn BA (eds) *Compost Utilization in Horticultural Cropping Systems*, pp. 51–93. Boca Raton, FL: Lewis Publishers.

Sikora LJ and Szmidt RAK (2001) Nitrogen sources, mineralization rates, and nitrogen nutrition benefits to plants from composts. In: Stoffella PJ and Kahn BA (eds) *Compost Utilization in Horticultural Cropping Systems*, pp. 287–305. Boca Raton, FL: Lewis Publishers.

Zucconi F, Monaco A, Forte M, and De Bertoldi M (1985) Phytotoxins during the stabilization of organic matter. In: Gasser JKR (ed.) *Composting of Agricultural and Other Wastes*, pp. 73–86. London, UK: Elsevier Applied Publishers.

CONDITIONERS

R E Sojka and J A Entry, USDA Agricultural Research Service, Kimberly, ID, USA
W J Orts, USDA Agricultural Research Service, Albany, CA, USA

Published by Elsevier Ltd.

Introduction

The use of naturally occurring materials as soil-stabilizing conditioners has been part of agriculture and general land management for millennia. Some of the most familiar conditioners in use since ancient times include animal and green manures, peat, crop residues, organic composts, and lime. These early uses of conditioners resulted from knowledge gained from trial and error long before there was scientific understanding of how efficacy was derived. Other conditioners in use for centuries or decades include composted manures, various organic debris, including sawdust or other milling residues, food, textile, and paper-processing wastes and other organic industrial wastes, as well as mineral materials such as rock phosphates, gypsum, coal dust, rock flour, and sand.

Soil conditioner use and technology, since ancient times, has, in great part, been a marriage of convenience between the agricultural necessity for chemical and physical maintenance or enhancement of the land, and for the disposal or management of waste materials from the full spectrum of human activities. However, since about the early nineteenth century, as modern physics and chemistry emerged and were applied systematically to agriculture, soil conditioner identification, development, and use became more creative and deliberate.

With the development of soil science as a specific discipline, the terminology and concept of soil amendments and conditioners was gradually assigned primarily a physical-conditioning connotation. Chemical conditioning, *vis-à-vis* supplying plant nutrients to soil, has been largely ascribed to materials termed fertilizers. Clearly, however, there is substantial overlap. Many fertilizers affect soil physical properties, both directly and indirectly, and many soil conditioners affect soil fertility both directly and indirectly. The overlap of physical and chemical effects occurs because of the intimate association of all soil physicochemical process and their coupling, as well, to soil-supported biotic processes, cycles, and functions. The designation of fertilizer versus conditioner is often based on the dominant effect intended. Categories are often assigned by law, based on the chemical analysis and/or the proof of claims for the materials.

Early Use of Mineral and Organic Materials

This article provides a brief history of early and traditional conditioner technologies and then focuses on recent developments in inexpensive and highly effective synthetic conditioner materials and use strategies. Organic conditioners have generally been applied to increase infiltration and soil water retention, promote aggregation, provide substrate for soil biological activity, improve aeration, reduce soil strength, and resist compaction, crusting, and surface sealing. The effects of organic conditioners often occur bimodally. That is, some effects, such as improved infiltration and water retention, are evident immediately upon soil incorporation, whereas other effects, such as improved aggregation, depend on chemical and biological processes over time.

Mineral conditioners are often used to affect soil chemical processes as well as soil physical processes. Lime, for example, raises soil pH. Gypsum or lime is often used to increase base saturation, or reduce the exchangeable sodium percentage (ESP) of retained cations. Because the divalent calcium ion has a compact hydrated radius, it also promotes flocculation of clays and increases aggregate stability. These effects help to reduce particle dispersion and detachment – which reduce erosion and surface sealing. Similarly, the calcium ion promotes flocculation and aggregation. These effects can be particularly important in arid soils with low soil organic matter (SOM) contents. The physical properties of such soils are often impaired when the exchange complex is dominated by the sodium ion, which has a much larger hydrated radius, and thus impedes flocculation and aggregation and favors dispersive phenomena. The physical benefits of calcium addition in low SOM saline soils provide for improved leaching of salts and removal of sodium, especially under irrigated conditions.

Mineral conditioners are especially important for the management of arid or tropical soils where high temperatures promote rapid bio-oxidation of incorporated organic material. A variety of other strategies are used with mineral conditioners to exploit soil physicochemical processes, directly or indirectly improving soil physical and/or chemical status. While the uses of lime and gypsum have ancient origins, another interesting approach in recent decades has been the use of various oxides of iron to promote aggregation in low-organic-matter soils. In the 1970s researchers added iron oxides to increase aggregate stability of soils and found peak aggregation at a 2% addition rate, with aggregation favored by acidic conditions. Others in the 1980s found promising results for addition of ferrihydrite compounds to calcareous soils, with the formation of weak quasicrystalline structures. Recent work shows potential for adding ferric hydrides to low-organic-matter soils for structure improvement and wind erosion resistance. Ferric hydrides are common water-treatment and industrial process waste products.

Soil conditioner research to the present has explored the use of many naturally occurring organic and mineral materials, agricultural and industrial waste products, or by-products of other processes. Materials that have been used as conditioners have included crushed rock, ground coal, gypsum (mined or from ground plasterboard), wood chips, bark, sawdust, food-processing wastes, cheese whey, various manures, composts of manures and/or other organic materials, and, as discussed more fully below, a wide range of synthetic polymer materials, including copolymers of synthetic and naturally occurring substances. All these materials have shown varying capacities to modify soil conditions or soil processes.

Use of Waste Materials as Conditioners

The extent of soil conditioner use has often been limited by economics. The cost of conditioner use has commonly been due more to transportation and application expenses of bulky materials than to the price of the materials *per se*. In many instances conditioner material is available gratis from waste streams of various processes where disposal is an expense. Use of waste materials as soil conditioners eliminates the disposal expense and in some cases creates profitable products. Because of material bulk, transportation, application, and related costs, the widespread use of traditional soil conditioners in mainstream production agriculture has been limited to only a few very highly efficacious materials such as lime, gypsum, and manure. Exceptions have occurred in high-value nursery, cash crops, turf and landscape applications, or in proximity to sources, where transport costs have been minimal.

Advent of Synthetic Conditioners

Since the early 1950s soil scientists have explored using synthetic polymeric conditioners to alter drastically soil physical and, in some cases, chemical properties. During World War II water-soluble polymers were used to stabilize soils in order to hasten the construction of roads and runways. The use of polymeric soil-conditioning chemicals was introduced to agricultural research and the farming community following World War II. In 1949 an industrial process for polymerizing acrylamide molecules was patented. This ultimately enabled a vast new array of water-soluble polymer compounds with thousands of industrial and environmental uses. Sixteen scientific reports of water-soluble polymer soil conditioning appeared by 1952, and 99 reports by 1955.

By enhancing the formation of soil aggregates and prolonging their longevity, water-soluble polymeric conditioners improve soil physical properties, including root penetration, erosion resistance, infiltration, aeration, and drainage. These direct physical improvements usually promote rooting and plant interception of nutrients and water, indirectly improving plant nutrition. The synthetic materials perform immediate conditioning and structural stabilization that would ordinarily require weeks, months, or years to achieve via a program of organic matter incorporation. Furthermore, synthetic materials can effectively condition soil to the depth of tillage with one to two orders of magnitude less material application than required with traditional conditioners, and can be zone- or spot-applied for even more efficient, targeted application and efficacy. Despite these performance advantages, however, in the early years of synthetic soil conditioner use, cost usually restricted use to high-value crops or specialty applications.

The most common strategy for water-soluble polymeric soil conditioner use from the early 1950s until the early 1990s was the application of sufficient conditioner material to affect significant physical modification of the soil to the depth of tillage. Depending on the nature of the polymer conditioner material, these application amounts often reached hundreds of kilograms per hectare. Generally, this mode of treatment entails multiple application operations, either as bulk solid materials, or as sprayed liquids or slurries. Each application usually requires tillage to incorporate the material to a desired depth. Because the mass of soil in a typical hectare-plow-layer is great (typically 2 000 000 kg per hectare 15-cm slice), many tons per hectare of traditional physical amendments and hundreds of kilograms per hectare of water-soluble polymeric soil amendments are usually necessary to overcome the physical or chemical buffering effect of the large mass of soil being treated.

Some of the most commonly used water-soluble synthetic soil-conditioning polymers since the 1950s have been: hydrolyzed polyacrylonitrile (HPAN), isobutylene maleic acid (IBM), polyacrylamide (PAM), polyvinyl alcohol (PVA), sodium polyacrylate (SPA), and vinylacetate maleic acid (VAMA). Commercial formulations of these compounds sometimes combined polymers and extenders or solubility-enhancing agents. Perhaps the most successful water-soluble soil-conditioning polymer marketed commercially before the 1990s was the Monsanto product Krilium which combined VAMA with a clay extender for improved application uniformity. Krilium and similar products were marketed in the 1950s at costs of \$4–5 kg^{-1}. Then-current application techniques required the application of tens to hundreds of kilograms per hectare, depending on the depth and extent of the soil zone to be treated. This precluded use on all but high-value crops or in specialty situations. After initial enthusiasm for these conditioners, most products have been withdrawn from general marketing to mainstream agriculture because of economic realities.

Hydrogels and Super Water-Absorbent Polymers

There was also interest over the years in super water-absorbent polymers for use in soils. These polymers are not water-soluble, but instead are strongly hydrophilic gel-forming materials that easily absorb hundreds or even 1000–2000 times their weight in water (Figure 1). Hydrolyzed starch-polyacrylonitrile graft polymers (H-SPANs), patented by the US Department of Agriculture in 1975 under the market name Super Slurper, and cross-linked polyacrylamides (gel-forming PAMs) have been the most common polymers for this application. They are used to improve the water retention of soils with low-water-retention properties, or that experience prolonged and untimely drought, especially immediately after planting. Spot placement of gel polymers in proximity to seeds, seedlings, or transplants prolongs the opportunity for emergence and seedling establishment without having to irrigate the entire soil profile. Again, because of cost and application amounts required, it is usually not economically feasible or logistically practical to attempt to modify an entire profile or even tillage zone, even when conditioner cost is as little as \$2 per kilogram.

Polymer chemistry, prior to about the 1980s, generally limited available soil-conditioning polymers to molecules with chain links of a few thousand monomer units. In addition, the purity of the preparations was not always good, sometimes carrying safety or environmental risks from reaction by-products or incompletely reacted base chemicals. Since the 1980s and early 1990s polymer purity and molecular size have increased, greatly improving the efficacy, safety, and affordability of environmental polymers. These changes, coupled with new application strategies that only target critical portions of the soil for treatment,

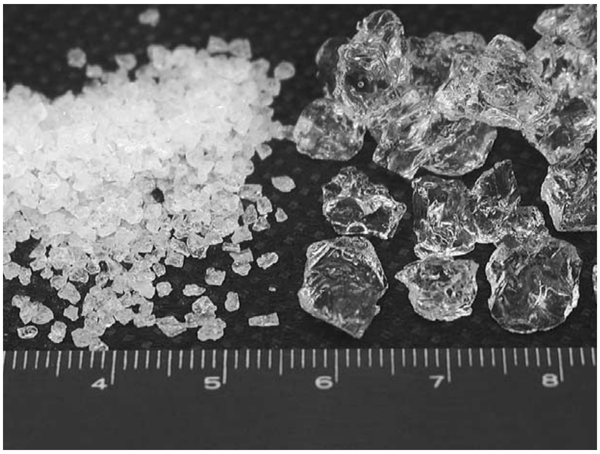


Figure 1 Superabsorbent cross-linked polyacrylamide in the dry state (left) and hydrated state (right); scale is in centimeters.

and that do not depend on expensive field operations for chemical application, have produced a sustained renewal of interest in environmental polymers for a growing number of uses. Perhaps the best example of this advancement has been the use of PAM for erosion control in irrigated agriculture.

Recent Advances Using Polyacrylamide

Isolated reports in the 1970s and 1980s provided a hint that very small amounts of PAM in irrigation water, flowing over soil in irrigation furrows, virtually eliminated detachment and transport of soil particles. These reports, however, were either anecdotal with respect to erosion or did not adequately identify the polymer used. Thus, the potential importance of the PAM-treatment erosion effects went unnoticed for several years. The foundation was laid for the practical use of PAM to halt erosion in furrow irrigation in a series of studies through the 1990s. The success of this new research came from the realization that the best way to treat soil structure to prevent erosion was to use the eroding water to deliver the soil conditioner. Irrigation is perfectly suited to this mode of application. In this mode of application only 1–2 kg ha^{-1} of PAM was needed to halt an average of 94% of erosion from irrigation furrows (Figure 2). The treated soil was restricted to about 25% of the field surface area and was only treated to a depth of a few millimeters. Inflows only needed to be dosed as water crossed the field, shutting off applicators when runoff began.

This strategy relies on the use of a highly specific class of food-grade PAM to ensure both efficacy and human and environmental safety. These PAMs are anionic, with a charge density of typically 18%; they are what are today regarded as moderately large molecules, having over 150 000 chained monomer segments per molecule for molecular weights of 12–15 million g mol^{-1}. The molecules are manufactured to a high purity, and are actually identical to PAMs used for food processing and drinking water treatment, with residual unreacted actylamide monomer (AMD) contents of <0.05%. The low AMD content and anionic nature of the molecule ensures safety for humans handling the PAM and for aquatic species in the event PAM is lost via runoff to surface

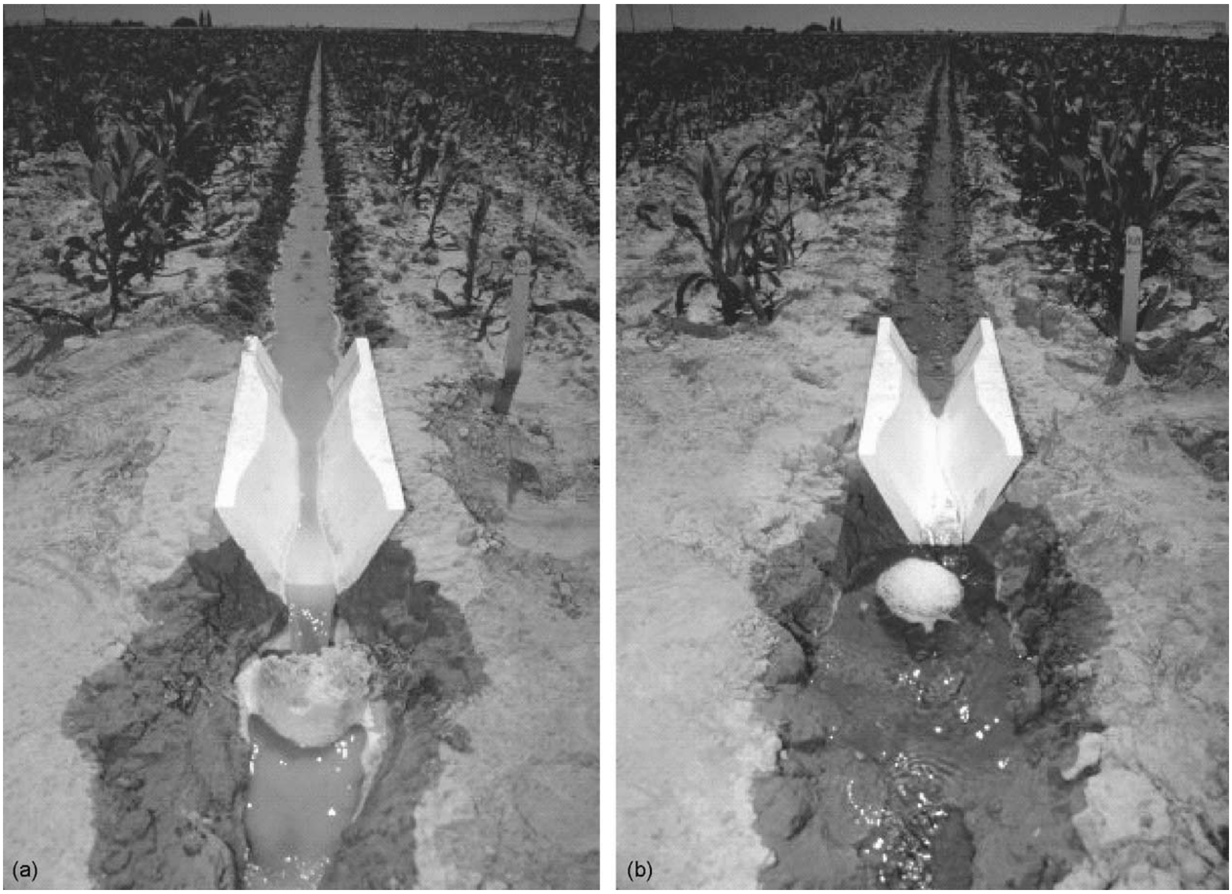


Figure 2 Runoff from irrigation furrows where (a) water is untreated and (b) water is treated with polyacrylamide; note the lack of turbidity, and thus absence of erosion from the polyacrylamide-treated furrow.

waters. However, the anionic charge imparts the need for bridging cations in the solvating water to link the anionic polymer to the anionic surfaces of soil minerals. Waters and soils containing dissolved calcium enable better efficacy than low-electrolyte (pure) water, and efficacy is best when there is little or no sodium present.

PAM is so effective at stabilizing surface structure, even at these small application amounts, that, in most fine- to medium-textured soils, infiltration is increased compared to untreated water, which induces surface sealing. While initial uses of PAM were focused mainly on erosion control, farmers are equally interested in using PAM for infiltration improvement. This is especially true as the technological barriers to use of PAM in sprinkler irrigation are overcome. With proper application strategies, PAM can be used both to increase infiltration amounts or rates as well as to improve infiltration uniformity. Since, with PAM in the water, soil structure is improved and surface sealing is reduced, water droplets enter the ground where they land, rather than causing seals and inducing runoff and ponding.

PAM use with irrigation for erosion control benefits water quality in a number of ways. By preventing erosion it also reduces the desorption opportunity for sorbed nutrients and pesticides, and limits dissolution of soil organic matter in runoff that elevates dissolved organic carbon (DOC) and biological oxygen demand (BOD). PAM-treated irrigation water has also proven highly effective at reducing movement off-site of soilborne microorganisms and weed seed, greatly reducing the likelihood of downstream inoculation and, ultimately, reducing the need for pesticides.

The prospects for the future development of PAM technology remain good. Because PAM increases the viscosity of water flowing through soil pores, the effects on infiltration are a balance of seal prevention allowing greater infiltration and viscosity slowing the passage of water. Experiments are currently underway to use the viscosity effects with other management strategies for canal sealing, improved infiltration uniformity along long irrigation furrows, and better water retention in soils where infiltration is not a problem.

Natural gas is the cheap abundant raw material from which PAM is currently made. However, current supplies and economics may not reflect the future. Work to develop new copolymers of PAM using chitin and starch as building blocks has proven promising, although results are yet to match those achievable currently with PAMs. Use of these materials as building blocks for effective flocculents and soil stabilizers carries the added benefit of using another agricultural waste stream to produce value-added products. In this case products may eventually be provided that can add to our inventory of environment-protecting and production-improving agricultural tools.

The field of water-soluble polymers for environmental protection and agricultural management is growing rapidly. These polymers are inexpensive, effective, and safe. They can be easily used in many settings and provide nearly instantaneous results in most instances. They can be used effectively in combination with more traditional land management and water quality protection techniques, such as reduced tillage or riparian buffer strips, either enhancing the effectiveness of these more familiar approaches or providing additional 'insurance' for situations when alone they are less reliable. The work of the last decade has emphasized that agricultural and environmental polymers cannot be regarded as 'silver bullets,' but when used in a well-planned approach to agricultural land and water management or environmental protection, offer a significant new capacity for better resource utilization and environmental protection.

Extensive additional information on the current use of PAM as an environmental polymer for erosion and pollution prevention and for irrigation water management improvement can be found at the website www://kimberly.ars.usda.gov/pampage.shtml.

See also: **Aggregation:** Physical Aspects; **Crusts:** Structural; **Structure**

Further Reading

Aase JK, Bjorneberg DL, and Sojka RE (1998) Sprinkler irrigation runoff and erosion control with polyacrylamide – laboratory tests. *Soil Science Society of America Journal* 62: 1681–1687.

Agassi M, Letey J, Farmer WJ, and Clark P (1995) Soil erosion contribution to pesticide transport by furrow irrigation. *Journal of Environmental Quality* 24: 892–895.

Bjorneberg DL and Aase JK (2000) Multiple polyacrylamide applications for controlling sprinkler irrigation runoff and erosion. *Applied Engineering in Agriculture* 16: 501–504.

Entry JA and Sojka RE (2000) The efficacy of polyacrylamide and related compounds to remove microorganisms and nutrients from animal wastewater. *Journal of Environmental Quality* 29: 1905–1914.

Lentz RD and Sojka RE (1994) Field results using polyacrylamide to manage furrow erosion and infiltration. *Soil Science* 158: 274–282.

Lentz RD, Sojka RE, and Robbins CW (1998) Reducing phosphorus losses from irrigated fields. *Journal of Environmental Quality* 27: 305–312.

Malik M and Letey J (1992) Pore-sized-dependent apparent viscosity for organic solutes in saturated porous media. *Soil Science Society of America Journal* 56: 1032–1035.

Orts WJ, Sojka RE, Glenn GM, and Gross RA (2001) Biopolymer additives for the reduction of soil erosion losses during irrigation. In: Gross RA and Scholz C (eds) *Biopolymers from Polysaccharides and Agroproteins.* ACS Symposium 786, pp. 102–116. Washington, DC: American Chemical Society.

Rhoton FE, Romkens MJM, Bigham JM, Zobeck TM, and Upchurch DR (2003) Ferrihydrite influence on infiltration, runoff and soil loss. *Soil Science Society of America Journal* 67: 1220–1226.

Sojka RE and Entry JA (2000) Influence of polyacrylamide application to soil on movement of microorganisms in runoff water. *Environmental Pollution* 108: 405–412.

Sojka RE, Lentz RD, Trout TJ *et al.* (1998) Polyacrylamide effects on infiltration in irrigated agriculture. *Journal of Soil Water Conservation* 53: 325–331.

Sojka RE, Lentz RD, and Westermann DT (1998) Water and erosion management with multiple applications of polyacrylamide in furrow irrigation. *Soil Science Society of America Journal* 62: 1672–1680.

Stewart BA (ed.) (1975) *Soil Conditioners*. Special Publication 7. Madison, WI: Soil Science Society of America.

Trout TJ, Sojka RE, and Lentz RD (1995) Polyacrylamide effect on furrow erosion and infiltration. *Transactions of the ASAE* 38: 761–765.

Wallace A (ed.) (1995) *Soil Conditioner and Amendment Technologies,* vol. 1. El Segundo, CA: Wallace Laboratories.

Wallace A (ed.) (1997) *Soil Conditioner and Amendment Technologies,* vol. 2. El Segundo, CA: Wallace Laboratories.

Wallace A and Terry RE (eds) (1998) *Handbook of Soil Conditioners: Substances That Enhance the Physical Properties of Soil.* New York: Marcel Dekker.

Conservation *See* **Erosion:** Water-Induced; Wind-Induced; **Sustainable Soil and Land Management**; **Terraces and Terracing**

CONSERVATION TILLAGE

M R Carter, Agriculture and Agri-Food Canada, Charlottetown, PE, Canada

Introduction

Conservation tillage (CT) is an umbrella or generic term used to describe tillage systems that have the potential to conserve soil and water by reducing their loss relative to some form of conventional tillage. Precise definitions of conservation tillage are only possible within the context of known crop species, soil types and conditions, and climates. A well-accepted operational definition of CT is a tillage or tillage and planting combination that retains a 30% or greater cover of crop residue on the soil surface. Generally, there are four main types of CT: mulch tillage, ridge tillage, zone tillage, and no-tillage. A main variant of the latter is direct drilling (sometimes termed zero-tillage), while other variants of CT are reduced tillage and minimum tillage. Conservation tillage can provide several benefits for agricultural systems such as soil conservation, economic advantages associated with reductions in crop establishment time and energy use, reduction in soil sheet erosion and nonpoint pollution, and enhanced storage or retention of soil organic matter and improvement of soil quality at the soil surface.

Evolution of Conservation Tillage Systems

Tillage involves the mechanical manipulation of the soil. In an agricultural, horticultural, or forestry context, it involves manipulation of the soil profile to modify soil conditions and to manipulate plant residues, and to control or remove unwanted plant growth. In agricultural systems, tillage functions as a subsystem that influences crop production mainly through crop establishment, modification of soil structure, incorporation of fertilizer and soil amendments (e.g., lime and manure), and weed control. Tillage is also used to alleviate both climatic and soil constraints.

The evolution of conservation tillage is a complex phenomenon with many varied themes. First, excessive soil tillage is associated with soil degradation processes such as compaction, a decrease in soil stability and structure, and increased soil erosion. Thus, one component of CT is a trend towards reducing or minimizing tillage events to address concerns with tillage-induced soil degradation. Second, most arable farming systems developed in climates where potential evapotranspiration is relatively high and precipitation is moderate to low. Under such conditions, a need arose to conserve soil water and reduce the propensity for soil erosion. Third, the advent of herbicides allowed the weed control aspect of tillage to be drastically curtailed and, in some cases, the need for tillage itself could be eliminated (i.e., adoption of no-tillage). Fourth, the cost of tillage in regard to tillage implement depreciation, tractor fuel use, and labor input is a major aspect of crop production costs. Thus, there is an economic incentive to reduce tillage events.

Within any one specific region, the speed and degree of CT adoption are related to a multitude of factors such as profitability and costs, socioeconomic concerns, the risk involved with adopting a new practice, and farmer skills and awareness of the need for soil conservation, and control of soil erosion. In the majority of reported surveys, concerns about conservation of the soil resource or pollution of the environment are not the foremost reasons for adoption of CT. The main reasons are economic feasibility and cultural compatibility. In many cases the potential economic advantages serve as the main incentive. Table 1 outlines the main reasons for adoption of CT in North America.

Defining Conservation Tillage

An acceptable all-embracing definition of CT is difficult to draft due to the varied conditions that underlie its development and the constraints presented by different soil types, cropping rotations, and climate. However, a common characteristic of CT is the potential to reduce soil erosion and water loss relative to some form of conventional tillage. A general operational indicator used to define CT, mainly in subhumid to semiarid climates, is the maintenance of a 30% soil surface cover with crop residue after seeding or planting. A well-accepted definition for CT, which reflects the above, is as follows: “Any tillage sequence, the object of which is to minimize or reduce loss of soil and water; operationally, a tillage or tillage and planting combination that leaves 30% or greater cover of crop residue on the surface.”

Table 1 Reasons, in order of importance, for the adoption of conservation tillage in North America

Reason for adoption of conservation tillage	*Major variables associated with adoption of conservation tillage*
Profitability	Low financial risk; fuel savings; no new equipment needed or savings on equipment cost/depreciation; potential for increase in crop yield and quality
Social and cultural benefits	Level of farmer education; form of land tenure, land owned rather than rented; greater farm size; full-time farmer
Control of soil erosion or degradation	Identification of soil degradation or soil erosion risk on farm
Conservation of soil water	Need to conserve soil water in semiarid warm areas
Timeliness	Need to improve time efficiency for crop establishment due to climate restrictions or labor shortage

Variants of Conservation Tillage

In many cropping situations and climates, innovations in tillage practices lead to soil and water conservation, although the operational definition for CT (i.e., crop residue indicator at 30% cover) may not be met or readily applicable. This has led to the development of various forms of tillage systems closely related to CT (Table 2). For example, in some agricultural systems the adoption of ‘reduced tillage’ allows a reduction in depth, degree, and frequency of tillage. Another variant of CT is ‘minimum tillage,’ which mainly refers to the approach or aim of achieving the minimum soil manipulation necessary for crop production, or that required to meet tillage requirements under specific soil and climatic conditions. A third variant is ‘shallow tillage’ or ‘non-inversion tillage,’ where the tillage is mainly restricted to a shallow (<15 cm) depth and the soil not turned

Table 2 Forms of tillage systems that meet some of the requirements of conservation tillage

Form of tillage system	*Characteristics of the tillage system*
Reduced tillage	Reduction in total number of primary[a] and secondary[b] tillage operations usually used in conventional[c] tillage, to prepare a soil for crop establishment
Minimum tillage	The minimum use of primary and secondary tillage necessary to meet crop production requirements under existing soil and climatic conditions, usually resulting in fewer tillage operations than for conventional tillage
Shallow or non-inversion tillage	Primary tillage confined to shallow soil depth (<15 cm). Absence of soil inversion[d] tillage

[a] Primary tillage is the initial major soil manipulation usually employed to loosen soil and bury or incorporate crop residue.

[b] Secondary tillage is any sequence of tillage operations that follow primary tillage generally used to prepare the soil for seeding/planting operations.

[c] Conventional tillage, a relative term, are those tillage operations normally used for crop production in a specific geographic area.

[d] Inversion tillage involves primary tillage that ‘inverts’ or ‘turns over’ soil causing much soil mixing. The main tillage implement used for inversion tillage is the moldboard plow.

over or inverted. This latter tillage system is sometimes termed 'plowless tillage.' In most cases, tillage systems that utilize moldboard plowing (tillage depth normally >15 cm; soil inverted) can approach the operational definition of CT if supported by use of cover crops and mulches to provide soil surface cover. Other variants use relatively deep tillage, but restrict the area or width of soil disturbed to a small proportion of the field.

Overall, the above variants of CT may not provide a 30% surface cover of crop residue, although most of the residue is still retained near the soil surface.

Common Types of Conservation Tillage Practices

The most common CT systems are 'mulch tillage,' 'ridge tillage,' 'zone tillage,' and 'no-tillage' (Table 3). The latter corresponds to 'direct drilling' and 'zero tillage.' Mulch tillage generally involves disturbance of the whole soil surface, while ridge tillage and zone tillage (sometimes termed 'strip tillage') often only disturb one-third or less of the soil surface. No-tillage restricts disturbance of the soil to that involved with crop seeding or planting. Depending on the type of seeder or planter, the proportion of soil surface disturbed can vary from 30 to 100%.

Constraints to the Adoption of Conservation Tillage Practices

The form of CT adopted in any one area is influenced by soil tillage requirement, climate, and type of farming system and crop rotation. Surveys (conducted in the 1990s) indicate that CT adoption in North America varies according to the above factors. Less than 25% of the coastal plains, delta areas, and humid northeastern regions of the USA utilize CT, while >40% of the southern corn (i.e., maize, *Zea mays* L.) belt, eastern uplands, Piedmont, and central Great Plains have adopted CT. The utilization of extreme forms of CT, such as no-tillage, in the above regions varies from 1 to 23%. In Canada, the adoption of CT ranges from 35% in the prairies to 13% in eastern Canada, with an overall adoption level of 24%.

Economic and sociocultural factors are the main reasons that influence the adoption of CT (Table 1). Generally, agronomic reasons are less important but still provide important incentives for the adoption of CT. Timeliness in crop establishment (especially in areas that use double cropping), the absence of soil compaction risk, and the range or diversity of crops conducive to the use of reduced-tillage practices are the main incentives. The great diversity of CT systems, as expressed by the various types of tillage (Tables 2 and 3), ensures that CT can be applicable to all soil types and farming systems; however, not every variant of CT is suitable for every situation. Several major constraints are associated with climate and soil type, high levels of crop residue, and mixed cropping systems.

Climate and Soil

No-tillage is generally difficult to use in cool, wet soils, or in soils with a relatively high tillage requirement. The latter refers to a soil's need for tillage based on its aggregate stability, potential for shrink-swell, and ability for self-structure. Tillage requirement is also influenced by climate and extrinsic factors such as site drainage and soil moisture class (Table 4). Generally, soils with an imbalance in particle size distribution (i.e., high clay or sand content) and/or those with poor permeability tend to have a high tillage requirement and are more difficult to manage when tillage is reduced.

High Levels of Crop Residue

Excessive amounts of crop residue may restrict the use of CT practices, especially no-tillage. Burning of

Table 3 The dominant types of conservation tillage used in agricultural systems[a]

Form of tillage system	*Some other terminology used*	*Main characteristics of the tillage system*
Mulch tillage	Stubble mulching, trash farming, sod farming, live mulch system	Some form of full-width[b] shallow primary tillage used prior to crop planting. High percentage of crop residues retained at the soil surface
Ridge tillage		Primary tillage confined to formation of raised ridges or beds in rows often on a contour. Planting occurs on the ridges
Zone tillage		Primary partial-width[b] tillage for row crops confined to bands, separated by bands of undisturbed soil, to form a seedbed for each row
No-tillage	Direct seeding, zero tillage	Soil undisturbed with no primary or secondary tillage. Crop seeded/planted directly into the soil

[a] Forms of conservation tillage mainly dominant in subhumid to semiarid climatic regions.

[b] Full-width tillage, tillage conducted over the entire soil surface; partial-width tillage, tillage conducted over a proportion of the soil surface.

Table 4 Soil wetness and physical constraints for adoption of conservation tillage

Constraint	*Unsuitable characteristic*	*Principal constraint*
Soil wetness problems	Imperfectly drained soils	Low soil strength
	Surface water-logging	Reduced trafficability
	Slow subsoil permeability	Excess soil compaction
Imbalance in soil particle size	Sandy soils with low organic matter	Excess soil compaction and poor structure
	Soils with >32% clay	Excess soil compaction
	Soils with high amounts of silt and fine sand	Excess soil compaction and poor structure

crop residue may not be an option because of environmental concerns, while removal of residue may not be feasible. The main concerns with high residue levels are mechanical interference with seeding operations, delayed drying and warming of the soil surface, potential for allelopathic or residue-derived toxins, increased potential for crop disease and pests, and reduction in the efficiency of fertilizers and pesticides. Generally, some of these residue constraints can be overcome by adoption of technologies that improve seed drilling, harvesting techniques, chaff and straw distribution, and use of banding or placement techniques for fertilizer and pesticides. A switch from no-tillage to forms of mulch and noninversion tillage can be successful under conditions of high crop residue.

Mixed Cropping Systems

Crop rotation is a basic strategy for sustainable cropping systems, but use of varied crops can pose a difficulty for CT as the tillage needs of each crop may differ. Root crops, such as potato (*Solanum tuberosum* L.), have a greater tillage need than cereals. Under these situations, CT needs to be integrated into the rotation by using variations of tillage (i.e., rotational tillage) suited for specific crops.

Beneficial Influence of Conservation Tillage

Many CT systems can improve soil quality over time and reduce the soil erosion risk, improve soil properties, and reduce tillage costs. Soil health may also be improved, as indicated by increased microbial activity and competition and, in some cases, the potential for amelioration of plant pathogen activity and survival. Such microbial antagonism in the root zone can lead to the formation of disease-suppressive soils. Beneficial improvements in soil quality and health are usually associated with organic matter increase and concomitant improvements in the soil physical condition, especially soil aggregation and tilth. Use of CT can also influence storage of organic matter in the soil, which has implications for agricultural greenhouse gas emissions. Overall, these potential effects have important implications for environmental health.

Improving Soil Quality and Health

Adoption of CT can cause an increase in organic matter at the soil surface, relative to a conventional tillage, as a result of maintaining a greater proportion of crop residue near the soil surface. This stratification with depth of soil organic matter has important implications for soil quality and the environment because the soil surface receives much of the fertilizer and organic amendments and rainfall impact, and serves as an interface for gaseous exchange. Enhanced levels of organic matter result in concomitant increases in soil biological activity. This phenomenon, in combination with a reduction in soil disturbance, often results in an increase in the size and stability of soil aggregates, and an increase in the continuity (but not always the size) of the soil pore space. This can have important implications for the way in which soil holds and partitions water, and holds and releases nutrients. For example, most surface soils under CT can better resist the harmful impact of raindrops and have a greater capacity to accept infiltrating water, than their conventional-tilled counterparts. One important factor, as a result of the above, is the development of a continuous pore network down the soil profile in loam to clay soils under CT, which can have implications for rapid movement or leaching of nutrients and pesticides.

Enhancing Soil Organic Matter Storage

The organic matter storage capacity of a soil is dependent upon factors that affect organic matter inputs and outputs such as climate, soil type, landscape position, and management. The latter influences the store of organic matter in the soil by type of cropping system and soil tillage. Combination of cropping practice and reduced tillage can lead to soil organic matter gains and thereby provide a net sink for atmospheric CO_2, although such gains may be of finite magnitude and duration.

Generally, CT itself does not cause an increase in stored soil organic matter when compared to conventional tillage, under conditions of similar carbon inputs (i.e., most studies indicate that crop yield and crop residue are similar between CT and conventional tillage). Various scenarios are evident for tillage-induced organic matter storage in soil. CT can enhance the organic matter content of surface soils (0–15 cm) in semiarid climates, in comparison to conventional tillage; however, such gains are usually due to labile forms of organic matter (e.g., particulate organic matter). In some climates, extensive tillage can place crop residues at a depth in the soil where decomposition proceeds at a slower rate than that observed for surface soils. Intensive tillage may also enhance organic matter association with soil clay and silt particles and thus encourage aggregation and consequently organic matter storage. Thus, differential placement and incorporation of crop residue and the degree of tillage intensity between tillage systems could result in differences in amounts of stored soil organic matter over time.

The beneficial influence of CT on reducing CO_2 emissions can involve more than increases in soil organic matter content. Full carbon cycle analysis takes into consideration carbon used or emitted from the manufacture and use of agricultural inputs such as machinery, fuel, and other inputs. Most full carbon cycle studies to date suggest that CT provides overall reduction in net CO_2 emissions, compared to intensive forms of tillage.

Reducing Soil Erosion and Environmental Risk

Nonpoint or diffuse pollution is the water pollution associated with land-use activities. In North America, CT is considered to be a beneficial practice to control nonpoint pollution. In general, due to the presence of surface crop residue, CT can slow the rate of water runoff, increase the rate of water infiltration, and reduce soil movement or erosion. Runoff volume (but not always runoff concentration), sediment losses, and losses of sediment-associated materials (e.g., adsorbed phosphorus and pesticides) can be decreased under CT. The leaching potential under CT, however, can be increased, which has implications for groundwater contamination by nitrogen and the limited number of pesticides that are highly mobile. In addition, runoff volume may not decrease in the presence of traffic/tillage-induced compaction, poor internal drainage, or the presence of fragipans and subsurface clay horizons. Reduced runoff can decrease sedimentation losses but not always the amount (i.e., increased concentration) of soluble phosphorus and pesticides in the runoff, resulting in no overall beneficial effect.

Table 5 summarizes the results of many CT studies in regard to range of runoff and leaching parameters. Conservation tillage can markedly reduce sedimentation but the effects on runoff volume are less clear, while the reduction in the nitrogen, phosphorus, and pesticide content of the runoff is variable. Leaching volume is greater under CT. Generally, most of the studies that measured pesticide loss in runoff or leachate indicated that only minor amounts were lost. Overall, CT is a useful management practice for control or reduction of sedimentation, but less effective for decrease of phosphorus movement. Generally, P resides in both inorganic and organic forms in the soil, and the latter is related to increasing levels of organic matter. Organically bound phosphorus is more soluble and therefore more mobile than phosphorus adsorbed on to soil minerals. Generally, CT needs to be integrated with both nutrient and pest management techniques to reduce nitrogen, phosphorus, and pesticide applications to the soil.

Table 5 Influence of CT on sedimentation, runoff and leaching compared to conventional tillage[a]

Measurement	*Decrease (%)*	*No effect (%)*	*Increase (%)*
Runoff volume	38	37	25
Sediment in runoff	100	0	0
Nutrients[b] in runoff	70	11	19
Leaching volume	20	0	80
Nitrogen leached	36	27	36

[a] Comparison of 40 studies (1991–1996) in North America comparing CT and some form of conventional tillage on a range of soil types.
[b] Nitrogen and phosphorus.

Tillage Costs

Use of fuel is generally lower under CT because of the reduction in tillage operations. Since fuel can account for 10% or more of farm energy use, CT clearly presents an economic and environmental benefit. However, general estimates of full production costs are problematic since CT often has higher fertilizer and pesticide requirements than the conventional tillage comparison. Overall, general principles in regard to tillage costs are not possible. Due to the wide variation in type of CT systems, in regard to tillage implements and crop production systems, conclusions on full tillage costs can only be site-specific.

See also: **Aggregation:** Physical Aspects; **Biocontrol of Soil-Borne Plant Diseases**; **Carbon Emissions and Sequestration**; **Crop-Residue Management**; **Degradation**; **Plant–Water Relations**; **Pollution:** Groundwater; **Sustainable Soil and Land Management**; **Zone Tillage**

Further Reading

Allmaras RR, Langdale GW, Unger PW, Dowdy RH, and Van Doren DM (1991) Adoption of conservation tillage and associated planting systems. In: Lal R and Pierce FJ (eds) *Soil Management for Sustainability*, pp. 53–83. Ankeny, Iowa: Soil and Water Conservation Society.

Cannell RQ and Hawes JD (1994) Trends in tillage practices in relation to sustainable crop production with special reference to temperate climates. *Soil Tillage Research* 30: 245–282.

Carter MR (ed.) (1994) *Conservation Tillage in Temperate Agroecosystems*. Boca Raton, Florida: CRC Press.

Carter MR (1994) A review of conservation tillage strategies for humid temperate regions. *Soil Tillage Research* 31: 289–301.

Carter MR (1998) Conservation tillage practices and diffuse pollution. In: Petchy AM, D'Arcy BJ, and Frost CA (eds) *Diffuse Pollution and Agriculture 11*, pp. 51–60. Aberdeen: Scottish Agricultural College.

Lal R (1989) Conservation tillage for sustainable agriculture: tropics versus temperate environments. *Advances in Agronomy* 42: 85–197.

Langdale GW, Alberts EE, Bruce RR, Edwards WM, and McGregor KC (1994) Concepts of residue management: infiltration, runoff, and erosion. In: Hatfield JL and Stewart BA (eds) *Crop Residue Management*, pp. 109–123. Boca Raton, Florida: CRC Press.

Logan TJ, Davidson JM, Baker JL, and Overcash MR (eds) (1987) *Effects of Conservation Tillage on Groundwater Quality*. Chelsea, Michigan: Lewis Publishers.

Mannering JV and Fenster CR (1983) What is conservation tillage? *Journal of Soil and Water Conservation* 38: 140–143.

Paustian K, Elliott ET, and Carter MR (eds) (1998) Tillage and crop management impacts on soil carbon storage. *Soil Tillage Research* 47: 181–349.

Pierce FJ (1985) A systems approach to conservation tillage: introduction. In: D'Itri FM (ed.) *A Systems Approach to Conservation Tillage*, pp. 3–14. Chelsea, Michigan: Lewis Publishers.

Rasmussen KJ (1999) Impact of ploughless soil tillage on yield and soil quality: a Scandinavian review. *Soil Tillage Research* 53: 3–14.

Uri ND (1998) Trends in the use of conservation tillage in US agriculture. *Soil Use and Management* 14: 111–116.

COVER CROPS

L Edwards, Agriculture and Agri-Food Canada, Charlottetown, PE, Canada
J Burney, Dalhousie University, Halifax, NS, Canada

Introduction

In a global climate of unease over the loss of ecological capital due to accelerated soil erosion and other forms of soil degradation under traditional cropping systems, attention is increasingly turning towards sustainable approaches that will protect the soil surface year-round and enrich it in the long term. Although the principal crop in most annual systems affords adequate ground protection at full canopy, interim cover cropping is a necessity, particularly where the soil is subject to erosion by water or wind. Water erosion takes its toll on sloping land under the influence of rain or snowmelt (which makes humid areas particularly prone), while wind erosion mostly affects areas that are flat, exposed, and dry. The gravity of this process is as great as it is ancient, and, likewise, are the steps taken to correct it. The need for its correction is a pressing urgency as a globally expanding green consciousness demands measures to promote: (1) a soil environment where the surface (most fertile) layer stays intact and builds rather than loses fertility; and (2) an aquatic environment that is free of imported sediment (and chemical pollutants), mostly from unprotected farmland. The judicious use of cover crops (sometimes known as green covers) is central in attaining these achievements.

Cover crops are the most practical means of providing ground cover, and thus the principal agronomic approach to erosion control and fertility building on crop land. Mulching also has a historical place, but can be limited where the material is unsuitable for soil enrichment, or where it is easily dislodged by wind or frequent runoff. The merits of cover crops are, however, well established and are mainly attributed to their relative ease of agronomic control. There is, for example, a wide choice of species and planting techniques for ensuring adequate plant establishment in most temperate or tropical settings, thus making cover cropping feasible under most soil, climatic, and cultural conditions.

The goals of cover cropping have historically varied between single-purpose and multipurpose, and may thus range from pure stands (Figure 1) principally for ground protection, to mixed stands of food (Figure 2) or feed species. Whereas these latter are, in large part, established towards economic ends (e.g., grain or hay in north-temperate climates and pulses in tropical climates), eventually all systems manifest in soil quality improvement that may take the form of: (1) physical conditioning; (2) nitrogen building; or (3) organic matter enhancements to the benefit of any arable terrain, be it sloping or flat.

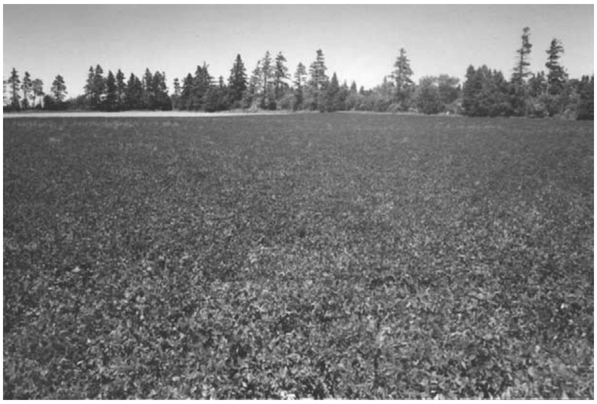


Figure 1 Pure stand of red clover (*Trifolium pratense*) as a rotation crop with potatoes (*Solanum tuberosum*).

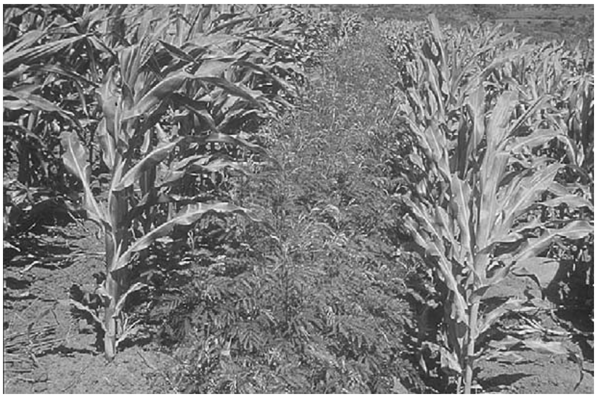


Figure 2 Leucaena cover (*Leucaena* spp.) with alternating rows of maize (*Zea mays*). Courtesy of P.K.R. Nair.

The conservation and nutrient-enrichment roles of cover crops in orchards and other perennial cultures are recognized, but are not dealt with specifically in this article.

This article examines the agronomy of cover crops, thus addressing species selection and husbandry, and describing their function in soil quality improvement. It also considers the role of cover crops in soil erosion control because of the body of evidence that puts this relationship into focus.

Agronomy

Cover crops are found among both monocotyledonous and dicotyledonous species. The former botanical group, which comprises mostly cereals and grasses, are doubly useful in soil erosion control because, in addition to the direct role of their vegetation in shielding the soil surface against aggregate breakdown by raindrop impact, their fibrous root systems bind the soil, thus increasing its resistance to entrainment by overland flow. On the other hand, the soil-protection attributes of a dicotyledonous cover are due, overwhelmingly, to the broad-leaf nature of this group of plants – specifically their characteristically overlapping system of leaves geared to intercepting rainfall and minimizing splash.

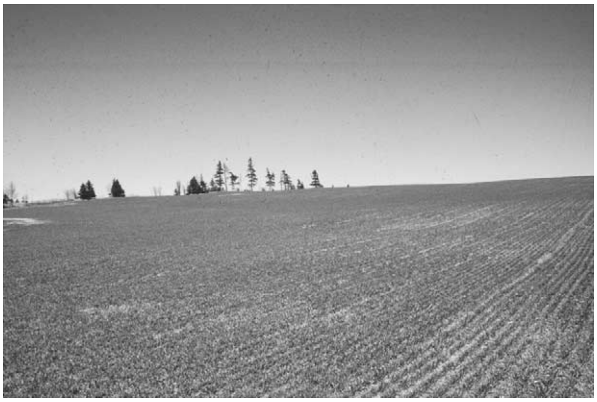


Figure 3 Winter rye cover (*Secale cereale* L.) following main- crop potatoes (*Solanum tuberosum*).

Success in cover-crop establishment and function depends greatly on species choice; and, accordingly, there is wide variation in both botanical groups. Thus, species suitability is hardly ever limiting for the range of cover cropping needs that exist.

Species Selection

Within the limits of adaptability of any cover crop to prevailing climate and soil, foremost among the desired characteristics is establishment vigor, persistence to the end of the required period, and ease of eradication to make way for the succeeding crop – usually the main crop. For instance, in the humid Atlantic region of Canada where cover cropping is practiced against cool-season erosion, cold temperature is the major limitation. Thus, winter rye (Figure 3) is the appropriate cover-crop choice under these circumstances because of its superior autumn-establishment vigor which allows it to fit into the narrow window between autumn potato (*Solanum tuberosum* L.) harvest and the onset of cold temperatures. Its cool-period persistence and excellent regrowth in early spring give this species a place in farming tradition as a cover crop in this region as in parts of the USA where it is planted after soybeans (*Gycine soja* L.) or after cotton (*Gossypium hirsutum* L.) to increase infiltration and reduce erosion on compacted soil. Seeding rates can even be manipulated to compensate for late-autumn seeding, which is its fate where potato harvesting is delayed to increase profitability. Generally, the later the potato harvest, the larger and more suitable are the tubers for both the fresh and processed (mainly french fries) markets.

In sharp climatic contrast to temperate situations, in many tropical areas where dry impoverished soils are all that exist for cover cropping after main crops of grain, cover-crop choice is restricted to hardy, low-growing species of legumes (like *Leucaena* spp.) because of their ease of establishment and spread, deep roots, and prospects of soil enrichment. Forage legumes are generally much hardier and better-adapted to these conditions than are pulses. Drought-tolerant grasses may be cultivated where longer-term cover is appropriate or where erosion is a severe threat. For this purpose, species such as *Chloris* spp. and *Cynodon* spp. have proven valuable.

Not always are cover crops confined to the period of most critical need, but might be deliberately taken to maturity and harvested for their economic products, usually within rotation systems. This is normally a business decision in which species versatility and adaptability to market opportunity play an important part. For example, many eastern Canadian farmers seed winter cereals as cover crops and cultivate them to full maturity if grain market prospects are good, even choosing to plant winter wheat (*Triticum aestivum*) where winter rye is preferable as cover. Under these conditions, winter rye is only an occasional grain crop because of the danger to livestock of fungus-infected grain.

Where the farmer wishes to hedge against cover-performance uncertainty and opts for mixed instead of pure stands, species compatibility is of prime importance in exercising choice, whether for herbage mixtures or herbage-grain companion cropping (as seen in Figure 2).

Seeding and Husbandry

Where the object of cover cropping is solely or primarily vegetative production, seeding method and spacing are far less exacting than where the object is an economic product like grain or pulses. Grass species, for example, are usually broadcast-seeded (surface-spread) for cover while legumes and cereals might be row-seeded (by drilling). This latter is a relatively demanding operation, but is an insurance against loss to pests or poor weather. However, where the system calls for companion cropping with grasses, this component might be interrow-seeded either simultaneously or in relay. Within this system, grass–legume mixtures are common, as are herbage–cereal mixtures, and usually require two operations to accommodate differences in seed size and planting-depth. Broadcast seeding may be accompanied by light harrowing to cover the seed.

Seedbed preparation is traditional for cover-crop seeding, but not always necessary depending on soil moisture conditions, terrain, seed availability, and the availability of planting equipment. Cultural circumstance might also play a part. Some US dairy farmers have been known to seed winter rye successfully in silage corn stubble and attain good winter cover with only light disking. Tea producers in India have successfully used a system of contour-staggered trenching (to conserve water) and leguminous cover crops, and have had yield increases exceeding 30% from the companion tea crop.

A variant of interseeding for cover was realized where winter cereal was successfully broadcast into standing crops of potatoes just days before harvesting in the autumn. Good winter cover was achieved without seedbed preparation. In this system, which was meant to achieve a measure of cover-crop growth (Figure 4) by the time the host crop was harvested, the winter cereal had germinated on undisturbed, relatively compact ground surface (but in a favorable microclimate), and was buried during potato harvesting with negligible damage to the emerging seedling. In this circumstance, soil moisture was adequate for growth – typical of this period in humid temperate climates.

Pregermination techniques to save establishment time are widely used, but are only lucrative for high-value crops (especially vegetables), enabling them to take early advantage of good field conditions in spring and then make a profitable entry to the market upon maturity. However, this technique is less practical for a low-value commodity like ground cover – whether it be in a temperate or tropical setting. Furthermore, short-term cover crops, upon germination, normally face a less-than-ideal period for establishment and growth since they usually follow the main crop and develop under progressively limiting conditions (i.e., cold or dry).

The most practical procedure to pretreat and seed cover crops for quick establishment, particularly

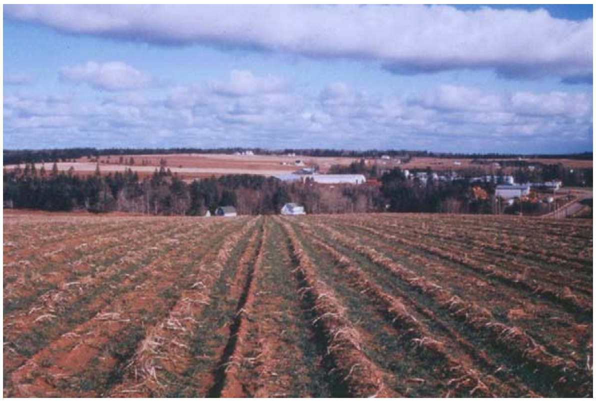


Figure 4 Early stand of winter rye (*Secale cereale* L.) broadcast-seeded into a standing crop of potatoes (*Solanum tuberosum*) just before harvest in the autumn.

under less-than-ideal soil-surface conditions, is the hydroseeding procedure used by highway departments which broadcast a liquid slurry of seeds, lime, fertilizer, and a fine, fibrous material that serves as a mulch, once applied. Straw or hay may also be used as a carrier if finely chopped, but is less capable of uniform spread. In the Atlantic region of Canada, this approach was tested to autumn-seed winter rye as a ground cover, but was constrained by uneven spread due mainly to equipment limitations. For example, standard on-farm equipment like manure spreaders proved inadequate, thus bringing into question: (1) the practicality (for a given farming enterprise) of even attempting to establish cover crops under poor ground-surface conditions; or, more directly, (2) the wisdom of investing in specialized equipment to facilitate cover cropping.

Cover-crop maintenance management naturally depends upon crop type and may, basically, be carried out as for any other crop husbandry to optimize growth. Where crop mixtures or companion cropping is called for, management becomes tricky and must be geared to the more sensitive of the crop components. However, in practical terms, since cover crops are usually relegated to secondary-crop status, postseeding inputs are often minimal. For example, fertilizers might not be used at all where cover cropping is short-term or if legumes are in the system. In any case, cover crops that follow heavily fertilized main crops (e.g., grain in Brazil and central or western Canada, or potatoes in the Atlantic region of Canada) could benefit adequately from nutrient carry-over.

Where cover crops are short-term and sandwiched by main crops into periods of unfavorable growth, husbandry attempts have, nevertheless, been made to boost performance through targeted fertilizer use, particularly with potash, with the object of enhancing winter- or drought-hardiness. However, success has been limited since the plants are unlikely to respond significantly during difficult periods of growth or when following a crop like potatoes that are normally overfertilized with K to improve eating or processing quality. Thus, it was found that increasing K dosages had no effect on autumn-seeded winter rye cover after main-crop potatoes in the Atlantic region of Canada, either in terms of leaf weight or the content of K in the leaves (Tables 1 and 2) at any time between cover establishment and termination in spring.

Termination of the cover-crop phase, in preparation for the main crop, is usually a tillage procedure that incorporates the plant material into the soil. This may range from near-surface harrowing of a short-duration cover of easily manageable plant material to plowing-under the vegetation, which may find itself below the root zone – as is the case with cereal stubble or swards. In either case, the procedure invariably aims to facilitate seedbed preparation and minimize interference with seeding machinery. It also aims to maximize contact between the incorporated material and the soil.

Termination of the cover-crop phase, in preparation for the main crop, may also be achieved through herbicide use, which is more common in large-scale operations or where soil-surface conditions are difficult. Undoubtedly, operational ease is the object of this approach, but it does little or nothing for soil improvement.

In situations of mixed farming, which characterizes many small to medium enterprises in the tropics, it is a common practice to harvest the cover crop periodically as cattle feed, although it may mean weakening the stand to the point where it contributes little to soil conservation or enrichment – even to the point of decimation.

Role in Soil Quality Improvement

Cover crops contribute to soil quality improvement principally through their decomposition by soil microbes. The products of decomposition, while

Table 1 Late autumn response (dry weight) of winter rye to applied K

	Trial					
	Year 1			*Year 2*		
K level[a]	*Plant count (tillers m^{-1})*	*Shoot (g m^{-1})*	*Root (g m^{-1})*	*Plant count (tillers m^{-1})*	*Shoot (g m^{-1})*	*Root (g m^{-1})*
K_0	127	4.46	0.93	145	4.46	1.27
K_1	140	4.82	1.01	162	4.36	1.57
K_2	141	4.92	0.85	148	4.11	1.45
LSD ($P = 0.05$)	33	1.23	0.23	37	1.44	0.43
($P = 0.01$)	46	1.70	0.31	51	1.99	0.59

[a] Year 1: K_0, control, K_1, 50 kg ha^{-1}, K_2, 75 kg ha^{-1}; year 2: K_0, control, K_1, 100 kg ha^{-1}, K_2, 150 kg ha^{-1}.

LSD, least significant difference.

Reproduced with permission from Edwards LM (1986) *Canadian Journal of Soil Science* 66: 31–35.

Table 2 Growth recovery (dry weight) of winter rye in spring with autumn-banded K

	Trial			
	Year 1		Year 2	
K level[a]	Shoot ($g\,m^{-1}$)	Root ($g\,m^{-1}$)	Shoot ($g\,m^{-1}$)	Root ($g\,m^{-1}$)
K_0	60.9	19.9	10.4	6.0
K_1	65.0	17.1	10.3	6.1
K_2	57.6	21.3	11.7	7.2
LSD ($P=0.05$)	17.2	5.9	2.2	1.4
($P=0.01$)	23.8	8.2	3.0	1.9

[a] Year 1: K_0, control, K_1, 50 kg ha^{-1}, K_2, 75 kg ha^{-1}; Year 2: K_0, control, K_1, 100 kg ha^{-1}, K_2, 150 kg ha^{-1}.

LSD, least significant difference.

Reproduced with permission from Edwards LM (1986) *Canadian Journal of Soil Science* 66: 31–35.

generally adding to the soil organic matter (SOM) reservoir, benefit the soil in two specific ways, i.e., through soil physical conditioning and through fertility building. The degree of enrichment depends on the quantity and quality of cover-crop biomass. Cellulose-rich plants or plant parts degrade far more rapidly than if they were ligneous – as is the nature of mature grasses. Hence, leafy portions of the shoot system degrade far more rapidly than stemmy portions. In any case, with decomposition, the resulting increases in SOM benefit the soil by improving its structure, thus enhancing gaseous exchange, internal drainage, nutrient exchange and water retention. Increased SOM also means greater benefits to soil biodiversity, which is at the center of all SOM dynamics and increased aggregate stability which, in turn, means a reduced tendency to erode.

From the standpoint of soil nutrient enrichment, cover crops contribute to the nutrient pool through the mineralization of decaying biomass under the influence of soil biota, which release the full range of plant nutrients in absorbable form. Where the decaying material is leguminous, nitrogen enrichment is predominant because of the nitrogenous nature of this family of plants and, particularly, the capacity of their roots to fix atmospheric nitrogen in symbiotic association with bacteria. Thus, by root rupture or necrosis, the soil eventually benefits from this nitrogen-fixing activity. *Pueraria phaseoloides* L. (Figure 5), a leguminous cover crop popular in the tropics, accumulated 150 and 250 kg N ha^{-1} within 4–18 months in parts of West Africa. In the end, the resulting soil nutrient enrichment following cover-crop fallows generally means increased crop production and savings on fertilizer costs, particularly under the poor-fertility conditions long-experienced and reported for Latin America and Africa. Tropical leguminous cover crops are also known for their foliar abundance. Some species contribute over 95% to total litter fall and are particularly favored in orchards and tree plantations if they are shade-tolerant and noncompetitive.

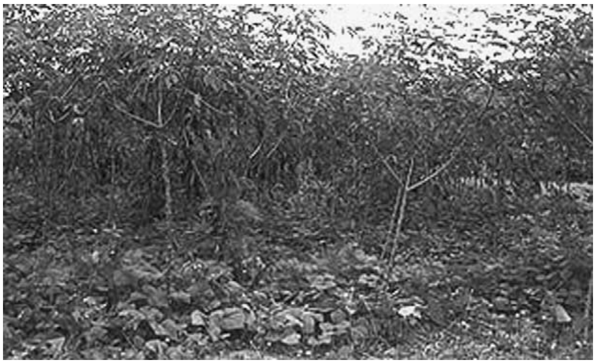


Figure 5 Stand of *Pueraria phaseoloides* under a crop of cassava (*Manihot* spp.). Courtesy of IITA, Nigeria.

Managing cover crops for soil physical improvement is best achieved with gramineous species. For example, grass cover crops, by virtue of the sheer volume and mass-distribution of their fibrous root systems within the root zone, bind the soil in a manner that gives it excellent structural attributes, the most outstanding of which are internal drainage and mechanical resistance. The vast root mass of which grasses are capable also contributes substantially to soil structure improvement via the SOM reservoir upon decay.

The ultimate cover-crop manipulation towards building soil fertility is the age-old practice of green manuring, the previously mentioned tillage activity that incorporates the plant material with the soil and stands out as the most useful form of crop destruction so far. Since most cover species are herbaceous, this material is usually not difficult to handle and biomass decomposition can be rapid. Because the speed of decomposition depends on the degree of contact between the plant material and the soil, and further depends upon the size of the incorporated material and the thoroughness of its incorporation, green-manuring tillage may require attention to be fully effective.

Role in Soil Erosion Control

Nature of Soil Erosion

There will never be any doubt of the beneficial role of cover crops in the control of soil erosion, which can be devastating to farmland (its customary point of origin) and neighboring environments. Its impact, however, is far more easily seen or felt than understood by either the land-user or the user of the

Table 3 Effect of winter rye (*Secale cereale* L.) cover on soil erosion using a laboratory rainfall simulator

	Mean value			
	Runoff		*Splash*	
Cell	*Volume (ml)*	*Sediment (g)*	*Volume (ml)*	*Sediment (g)*
Crop				
Bare	4191*a*	60.2*a*	434*a*	1.50*a*
Early-seeded	3917*a*	10.1*b*	390*a*	0.54*b*
Late-seeded	4943*a*	17.0*b*	437*a*	0.86*b*

a–b Means followed by the same letter in any column for any treatment are not significantly different at $P = 0.05$.

Reproduced from Edwards LM and Burney JR (1987) *Canadian Agricultural Engineering* 29: 109–115.

Table 4 Comparison of treatments: relative soil loss under outdoor simulated rainfall

Treatment	*No. of runs*	*Calculated relative soil loss*
Fallow	24	1.00*a*
Potato 0% cover	6	1.00*a*
Potato 5% cover	6	1.00*a*
Potato 70% cover	6	0.50*b*
Potato 90% cover	6	0.04*c*
Clover	12	0.06*c*

a–c Means in any column for any treatment followed by the same letter are not significantly ($P \leq 0.05$) different.

Reproduced from Parsons TS, Burney JR, and Edwards LM (1994) *Canadian Agricultural Engineering* 36: 127–133.

affected off-farm site or resource, be it a culvert, road, stream, river, lake, or estuary.

Fundamentally, the magnitude of soil erosion is contingent on a causative effect, a modulating effect, and a resistance effect. In the context of conserving topsoil on sloping farmland, the resistance effect can be attributed in large measure to live cover.

Undoubtedly, cover must be adequate, and becomes particularly crucial during periods of vulnerability to erosion on the basis that the occurrence, in combination, of high-intensity rainfall and weak soil resistance means that soil erosion on any given land surface would be otherwise high. Since contributing physical factors like weather and the degree of slope are uncontrollable, cover-crop usage becomes invaluable – particularly in tropical cultures where mulching material has to meet a more pressing need as fuel or building material than as ground cover.

Increased foliar mass of the cover crop reduces splash erosion predictably with increasing ground coverage, and this effect is incorporated as a crop-management factor in calculating soil loss. Cover crops work by reducing splash detachment and by intercepting the splashed soil, thus limiting its trajectory movement.

In work done in the Atlantic region of Canada to compare the effectiveness of winter rye (*Secale cereale* L.) cover in splash control using a laboratory rainfall simulator and splash equipment, a 47% reduction (compared to bare soil) was recorded for a highly erodible fine sandy loam (Table 3).

Runoff and Erosion Control

Over longer distances, the progressive build-up of runoff generates an increasing shear stress on the soil surface. As the effects of runoff increase relative to the constant effects of splash downslope, the root structure of the cover crop becomes increasingly important for keeping the soil in place.

A comparison of cover-management effectiveness under varying degrees of canopy, using a field rainfall simulator, showed that soil loss was not affected by a canopy of potatoes of up to 5%. It took a 70% canopy before the crop cover could keep the soil reasonably in place, and a 90% canopy before it was as effective as a mature stand of red clover (*Trifolium aestivum* L.) in the same comparison (Table 4).

Runoff comparisons, using a laboratory rainfall simulator on a fine sandy loam subjected to freezing and thawing cycles, showed that a cover of winter rye decreased runoff sediment to 23% of that from the bare soil, provoking significant deposition, particularly of the coarser-size grades. In taking a close look at the cover parameters and in assessing relative contribution to variations in sediment concentration, this study also showed that leaf area was by far the most important. Leaf area index, together with long-term climatic data, is a key component of many models that are used to determine safe planting dates, especially for winter cover crops.

Perspective

Present approaches to managing cover crops within a farm production system for soil enrichment or to minimize soil-surface destabilization are, for the most part, inadequate as farm management schedules mostly allow only: (1) short-term cover sandwiched between main crops; or, at best, (2) a growing season, where longer-term cover is required to bring the soil to full health. However, anything more than short-term cover could mean economic loss even if the cover is commercializable (e.g., as hay, cereal, or pulse).

The question still arises as to why cover cropping is ignored in the face of its agronomic and ecological merits, and why the pressure of commerce (minimum expenditure and maximum profit) overtakes most other considerations. For instance, monocropping grain crops or potatoes is common commercial practice in parts of North America and South America, with dire soil-health consequences in the long term. On the other extreme where, instead of cover cropping, bare fallowing is done to rest the soil from commercial production – a common practice with cereals in the Canadian Prairies – the result is a noxious salinization of the surface due to the upward movement of surplus soil nutrients – ultimately wasted. The use of cover crops to minimize nutrient wastage, whether through surface salinization or leaching, especially after intensive cropping, is widely recommended husbandry.

The practice of bare fallow in temperate or tropical regions is a missed opportunity to maintain (even enhance) the soil biological activity built up by the preceding crop. Thus, judicious selection of fallow cover crops can do much to enrich the soil microbial, earthworm and microarthropod populations – with benefits particularly needed under marginal conditions.

The danger to soil environmental health of ignoring cover cropping has struck a chord with many ecologists who press for its adoption at all levels of farming, and are ready to link its neglect, more widely, to global environmental decay. This linkage may, therefore, place cover cropping squarely in the realm of biological agriculture – remedial or sustainable – particularly based on its potential to sequester substantial amounts of soil organic carbon – estimated at nearly 30 millions of tonnes of carbon per year in the USA. Thus, it could only be to the advantage of cover cropping that it is identified with a 'green culture' that seeks a balance between agricultural enterprise and environmental capital.

Cover cropping is seen as green culture, given its role in SOM building, soil structure improvement, soil fertility increase, and disease suppression with minimal or no chemical inputs. Thus, it stands to gain expanding popularity and increasing political importance, considering that sustainable agriculture is successfully moving, as a recognized farming system, from temperate regions (Europe and North America) to tropical regions (Asia, South America, and Africa).

In view of the exigencies of cover cropping and the associated sacrifices, the question further arises of its relative place as a fertility builder at this time when the livestock industry produces, on average, a near excess of excrement whose benefit to soil health is unquestionable – having a greater capacity than cover crops to build SOM and enrich the soil rapidly. Furthermore, its effects are longer-term under intense cropping. In comparison, benefits derived from cover cropping (e.g., in terms of structure improvement) are only fully realized after several years of continual occupancy, and can disappear after a single year of commercial cropping under intense tillage. Therein lies a debate that could become particularly crucial where cropland is scarce.

But beyond any debate on the economy of cover cropping for soil-fertility building and erosion control on farmland is the importance of this practice in the wider context of terrestrial sustainability and ecological balance. Thus cover-crop species find important use in stabilizing fragile sand dunes that serve as habitat for rare coastal birds, in helping to restore surface-mine wastes to pre-mining productivity, and concealing dumps of solid industrial wastes that may pose a danger to animal life where easily accessible. Even the aesthetic value of cover crops is high considering their role in rendering unsightly or near-barren landscapes visually pleasing, even without mentioning their contribution to wildlife habitat and still broader biodiversity in temperate and tropical regions alike.

Having entered into the realm of green culture, cover cropping stands to gain status as a bona fide environmental ethic. It is, therefore, anticipated that the environmentalist movement will increasingly put pressure on farm operators and other land-users to recognize its ecological value and adopt it, at least in the name of environmental stewardship.

See also: **Carbon Emissions and Sequestration**; **Crop-Residue Management**; **Mulches**; **Nitrogen in Soils:** Cycle

Further Reading

Beasley RP, Gregory JM, and McCarty TR (1984) *Erosion and Sediment Pollution Control*. Ames, Iowa: Iowa State University Press.

Edwards LM (1986) Late-fall application of potash to winter rye to improve ground cover effectiveness. *Canadian Journal of Soil Science* 66: 31–35.

Edwards LM (1989) Dry matter growth performance of red clover and Italian ryegrass as cover crops spring-seeded into fall-seeded winter rye in relation to soil physical characteristics. *Journal of Soil and Water Conservation* 44: 243–247.

Edwards LM (1998) Comparison of two spring seeding methods to establish forage crops cover in relay with winter cereals. *Soil Tillage Research* 45: 227–235.

Edwards LM and Hergert G (1990) Establishing winter rye as a cover crop after potatoes. *Canadian Agricultural Engineering* 32: 183–187.

Foster GR (1982) Modelling the erosion process. In: Haan CT, Johnson HP, and Brakensiek DL (eds) *Hydrological Modelling of Small Watersheds*, pp. 297–370. ASAE Monograph Number 5. St Joseph, MI: ASAE.

Govers G (1991) Spatial and temporal variations in splash detachment: A field study. *Catena* Supplement 20: 15–24.

Kay BD (1990) Rates of change in soil structure under different cropping systems. *Advances in Soil Science* 12: 1–52.

Morgan RPC (1985) Establishment of plant cover parameters for modelling splash detachment. In: El-Swaify SA, Moldenhauer WC, and Lo A (eds) *Soil Erosion and Water Conservation Engineering*, pp. 377–383. Ankeny, IA: Soil Conservation Society of America.

Schwab GO, Fangmeier DD, Elliot WJ, and Frevert RK (1993) *Soil and Water Conservation Engineering*, 4th edn. New York: John Wiley.

Siegrist S, Schaub D, Phiffner L, and Mäder P (1998) Does organic agriculture reduce soil erodibility? The results of a long-term field study on loess in Switzerland. *Agriculture, Ecosystems and Environment* 69: 253–264.

Ward AD and Elliot WJ (1995) *Environmental Hydrology*. Boca Raton: CRC Press.

CROP ROTATIONS

C A Francis, University of Nebraska, Lincoln, NE, USA

Introduction

Since early in the history of agriculture, farmers have recognized the value of crop rotations. Sequences of different crops, in contrast to continuous cultivation of the same crop in the same field, were observed to have higher yields and were used thousands of years ago in China. In prehistoric Europe, when human population was small, land-clearing for several years of planting crops and pastures was often followed by allowing regrowth of natural vegetation and a repetition of this cycle of different species, but little is known about precise sequences. It is likely that there were rotations of annual crops with pastured areas, although this is speculation. The system is similar to the more familiar slash-and-burn or swidden agriculture still practiced in some parts of the tropics.

There is written evidence of the recognition of importance of crop rotations as well as putting specific crops on unique soil types during the Roman times. Although Cato the Censor (second century BC) dedicated most of his writing on farming to culture of grapes and olives, he also discussed the importance of applying manure and of fallowing land before planting annual or perennial crops. His writings describe the importance of vetch, beans, and lupines in helping to build the soil or fertilize the land, and that cereal crops following these legumes benefited and produced higher yields. Varro and Virgil (first century BC) recognized the importance of alfalfa, a deep-rooted perennial legume, in providing improved fertility for crops that followed, and they recommended rotations, fallow periods, and specifically a bean–spelt wheat rotation. Pliny (first century AD) wrote extensively on farming, including crop rotations in the context of Italy and the Mediterranean region. Likewise, there are written records of early effects of crop rotations and advice against the continuous planting of cereals in the literature from the Muslim era in the Iberian Peninsula.

In the Middle Ages before the Renaissance, there was apparently a reversion to crop–fallow sequences. There are limited reports of perennial pasture-grass plantings in rotation with cultivated cereal crops. Use of animal manure was prevalent. Perhaps the effects of application of manure were more apparent in a given year, if there were side-by-side areas with and without manure, compared with the more subtle differences that were the result of rotation, effects not easily observed. In the seventeenth and eighteenth centuries, more organized legume cereal rotations began to appear in reports from England. They often consisted of 2–4 years of vegetable or fodder crops, 1 year in legumes, and 1 or 2 years in cereals. Mixed farming with crops and livestock was prevalent, and the pastures and animal manure undoubtedly contributed to cereal yields. These rotation systems were taken to North America by early settlers from northern Europe.

Cropping diversity in many early systems included multiple species as well as crop rotations. In Europe the small grains were often mixed cultures of cereal species that were used for grain and for fodder. The well-known maize/bean/squash systems of Central America and Mexico as well as the intercrop and relay crop systems of maize/bean–potato in the Andean Zone are examples of diverse mixtures that evolved over centuries as people selected these New World crop species.

With an emerging understanding of soil chemistry and the responses of plants to applied manures and other soil amendments, crop rotations in the first half of the twentieth century focused again on cereal–legume sequences. In long-term experiments in

Missouri, USA, increased crop yields were found in rotations even when the cereals were supplied with adequate nutrients from animal manure application. According to various state experiment station reports, yields of maize were increased from 20 to 60% in rotations with legumes; in experiments in Ohio, there were 25% higher maize yields in rotation even when nitrogen, potassium, and phosphorus were applied to continuous maize. These popular rotation systems were phased out of most farming operations after World War II when large supplies of synthetic nitrogen and other multiple-nutrient fertilizers became popular. The specialization in field crops, especially cereals, and disappearance of livestock from most farms further pushed the systems into continuous cropping or simplified 2-year cereal–legume rotations. Only in the past two decades, with the interest and growth of organic farming and more integrated systems, are longer and more complex rotations coming back into popularity in the USA and Europe.

A summary of general types of crop rotations is presented in Figure 1, showing the array of different levels of genetic diversity that have been exploited in farmers' systems across the centuries. Rotating different varieties of the same cereal, rotating cereals, and rotating cereals with legumes are all strategies to provide some of the soil-fertility and plant-protection benefits that come from diverse cropping sequences. More diverse are the sequences of summer–winter crops and rotations of pastures with cropping fields, and most diverse of all are the annual–perennial sequences and combinations such as alley cropping.

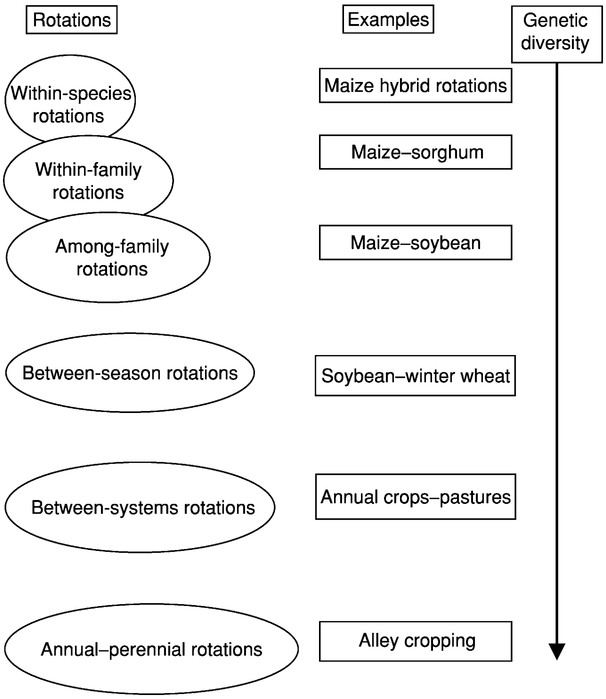


Figure 1 Different types of crop rotation, crop–pasture rotation, and long-term perennial alternatives, showing the increasing genetic diversity of specific systems (from top to bottom).

It is important to note that crop rotations are influenced by a wide range of factors, including biological interactions and contributions to yield on the farm, relative prices of differently adapted crops, and potentials for production and marketing from crop/animal systems. Farmers' potential to use rotations may depend on the agreements in place with landowners, since, for example, about half of the land currently farmed in the US Midwest is not owned by the people farming that land. Further, the political decisions in the EU and USA on crop commodity support payments and promotion for global markets for some selected crops narrow the farmers' options. Political decisions in the Baltic countries to position themselves for entry into the EU caused major changes in cropping systems during the 1990s, as farmers began to compete in a regional and global market and abandoned the guaranteed markets that were common under the former Soviet system. It is within this complex economic, biological, and political framework that crop rotations must be discussed.

Using current improved hybrids and varieties of major field crops, yields are generally improved by an average of 10% or more owing to crop rotation alone. This is a consistent observation, even when all apparent nutrient needs of crops are satisfied and there are no obvious limiting pest problems. If there are serious limiting factors that cannot be remedied by other means, for example a severe insect infestation or plant pathogen infection, crop rotations may increase yields much more. Thus the mechanisms of crop rotation contributions to yields and economic return are many and complex, and still poorly understood; those factors that have been researched relate to soil fertility, crop protection against pests, and the economic and social dimensions of farming.

Soil Fertility and Crop Yields

Cereal–legume rotations are among the most frequently used systems, dating back to before Roman times, and they still provide the foundation for successful design of crop sequences today. Crop rotations among crops from different families are among the most frequently used today in the temperate zone. In areas of adequate rainfall, above 600 mm year^{-1}, a maize–soybean rotation is prevalent in the Corn Belt of the US Midwest. In areas with lower rainfall, irrigation is used to assure a crop, and rotations are less common in these systems. Crop rotation effects are

Table 1 Examples and mechanisms for crop rotation contributions to soil fertility and nutrient cycling

Examples or mechanisms	*Contribution to soil fertility*
Cereal–legume rotation	Improves soil structure, texture, tilth, quality Increases water percolation, aeration, nitrification Promotes more efficient nutrient and water use
Different root structures	Shallow–deep species exploit more nutrients
Legume–grass sequence	Increases soil organic matter, reduces pests
Biennial and perennial legumes	Increases soil organic matter, manages weeds
Longer rotations and diversity	Improves soil quality, increases soil organic matter
Pasture–crops sequence	Increases organic matter, improves nutrient efficiency
Manure application	Stimulates soil mycorrhizae
Unlike crop sequences	Stimulates root exudates, increases mycorrhizal spores
No-tillage	Keeps earthworm burrows intact, improves soil mixing

often attributed to improved soil fertility that accrues from legumes in the sequence and the increased soil organic matter that improves soil quality. A number of specific effects of crop rotations are summarized in Table 1.

Maize in a 2-year rotation with soybean in the US Midwest consistently yields 5–20% more than continuous maize; yield increases in the same experiment in the same field often vary from 5 to 15% across years, depending on different growing conditions and what factors are limiting yields in the weather and soil situation in that array of years. Maize yields after a perennial crop such as alfalfa not only have higher yields than continuous maize, but the need for additional nitrogen is minimal in the 1st year. There are inconsistent results for maize yields in longer crop rotations with other annual crops. Similar yield increases are observed in other cereals grown in rotation with legumes, for example the grain sorghum–soybean rotation. Soybean in these rotations also shows approximately a 10% increase in yields, in spite of the fact that soybean fixes much of its needed nitrogen and fertilizer is most often not applied to this crop.

In areas of less rainfall, wheat is grown in rotation with a fallow year to accumulate enough soil moisture to make the crop profitable. An average of 25% of the fallow-year moisture is stored for a succeeding crop, with the rest lost to evaporation, yet this is enough in traditional systems to make the wheat–fallow rotation consistently profitable. Recently such systems as ecofallow, with minimal tillage and a sequence of two crops in 3 years (e.g., wheat–proso millet, wheat–sunflower, or wheat–maize) with short fallow periods, have intensified cropping and made better use of available rainfall.

Rotation of summer and winter crops, annual and perennial crops, and crops with pastures for grazing all reduce the need for applied fertilizers. This is due to rotations in crop root systems, nutrient-uptake patterns, timing of nutrient needs, diversity in crop residues, and different nutrient needs.

Cultural Management

Alternative soil-management methods, choices of crop species, and harvesting techniques can contribute to soil fertility, pest management, and amount of water available to crops in rotations. Reduced tillage maintains more residue on the soil surface, reduces erosion, and conserves soil moisture compared with plowing and other intensive land preparation and cultivation methods. The additional stored moisture opens up a farmer's options for crop rotation and increasing cropping intensity through the growing season. Catch crops of short-cycle legumes may thus be planted following a principal cereal grain crop, or a green manure crop planted with a cereal, in combinations not possible without this additional moisture.

Reduced tillage increases earthworm populations and soil quality, and slows oxidation and breakdown of organic matter to increase long-term nutrient supply to crops. Rotations of tillage practices that reach different soil depths can reduce the potential for soil compaction that leads to formation of a plow or disk pan. Rotating unalike crops may also require use of different equipment at different times of the year, and this diversity should promote a more diverse and reduced pest population and healthier farm ecosystem.

Pest Management with Rotations

Planting different crops in successive seasons or years can disrupt the reproductive cycles of many weed, insect, plant pathogen, or other pest species. This in turn can lower their potential for reducing crop yields or quality and diminish pest-control costs. Organic crop production systems rely almost entirely on rotations plus genetic resistance to manage crop pests without chemical pesticide applications. Problems with soil-borne pathogens such as bacteria, fungi, and nematodes can be reduced by crop rotations of different species. For some soil biotic pests, rotation is the only economical way to control them. A number

Table 2 Mechanisms of crop rotation contributions to plant protection against weeds, insects, pathogens, nematodes, and other biotic problems

Mechanism or process	*Contribution to plant protection*
Cereal–legume sequences	Favors different weed spp., facilitates mechanical control
Unlike crop sequences	Reduces soil-borne pathogens
Maize–soybean sequence	Controls maize stem borer, maize rootworm
Vegetable–cereal sequence	Controls nematodes
Potato–legume sequence	Controls wireworms
Continuous plant cover	Provides competition with weeds, reduces erosion
Legume cover crops	Reduces soil erosion, competition with undesirable weeds
Alfalfa in crop sequence	Control of annual weeds, increased predators and parasites
Strip-cutting alfalfa	Increases insect movement, increases natural enemies
Diverse crops in sequence	Soil biology diversity and suppression of soil pathogens
Cropping diversity over time	Promotes soil microorganism biodiversity

of mechanisms or methods for pest control are listed in Table 2.

Longer rotations that include annuals and perennials, or crops and pastures, provide both soil-fertility and pest-protection advantages. Annual crops allow frequent cultivation that will often suppress perennial weeds. Perennial crops may provide year-long competition that shades out annual weed species. Crop/animal mixed farming and a sequence of several years of annual crops followed by perennial grass/legume pasture mixes and grazing help to diversify income sources, as well as control many weeds and other pests that become problems in continuous cultivation of annual crops. Rotations of different animal species in the pastures – ruminants, pigs, poultry – make use of their different preferences and make more complete use of available forage, as well as reducing animal diseases that become problems in confined raising of large animal numbers in a small space. Goats will eat many weed species that are passed by other grazing ruminants, while chickens eat insect larvae from the manure deposits of larger grazers. The 4-year annual crop–4-year pasture rotation is a common system in Argentina.

Economic Diversity

Table 3 Mechanisms of crop rotation contributions to economic and social consequences of farming systems

Mechanism or process	*Contribution to economics or social viability*
Rotation of fields	Places uniform areas in all crops each year
Diverse crops in rotations	Provides forages, cereals, cash crops, soil building
Sequences of pesticides	Avoids buildup of resistance of insects, pathogens, weeds
Different tillage depths	Avoids buildup of plow-pan or compaction layer
Range in crop-planting dates	Spreads labor needs, reduces equipment costs
Range in harvest dates	Reduces peak labor demand, reduces equipment costs
Diverse array of crops/products	Buffers price changes in marketplace, gives direct sales
Different skills, labor needed	Increases interaction with local community for labor

In addition to these biological advantages of crop rotations, there are positive economic and social aspects that result from diverse plant combinations on the farm. Table 3 lists a number of direct economic benefits or those that will develop as a result of biological and other soil changes in fields over time. Either rotation of large fields or within-field diversity in strip-cropping annual systems will diversify the income stream and protect against a weak market for a single commodity. Planting summer and winter crops, or annual and perennial crops, will spread the demand for labor and equipment and allow farming of the same area with smaller equipment. Such rotations provide income at different times of the year. The diversification into more intensive crops and products provides the opportunity for using labor from other farms or nearby communities, perhaps opening new markets for direct sale as well.

Economic diversity can be enhanced by both crop rotations and the diverse products that come from the farm. If value-added activities on the farm or in the local community can make effective use of available labor when field operations are less intense, this strategy can increase farm income and return to investment. Crop rotations and diversity of products open new economic opportunities for on-farm sale or cooperation with neighbors in other direct sale options. A diverse farm is also attractive to visitors for educational programs, farm stays, hunting, and other creative ways to add value to the rural landscape and its natural resources.

Future Perspectives

Although economic rewards, consolidation, and specialization drove many farmers away from rotations in the latter years of the twentieth century, it is likely that cropping diversity and rotations will become more important in the future. Federal and regional subsidy programs that link payments to conservation of soil, water, and other natural resources often stipulate field and farm diversity as conditions for farmers to continue to qualify for payments. Organic farming regulations and subsidies in some countries specify the types of rotations that must be used for farmers to qualify. There is often a rule against continuous cropping of the same crop species and a requisite to include legumes in at least half of the years in a crop sequence.

With increased costs of energy, the nutrients from green manure and animal wastes become more attractive and cost-effective. When animal manure from concentrated confinement operations becomes an excessive liability, due to water-quality dangers, there will be greater pressure from society and from regulators to rescue this resource and cycle it back into the crop production process. Diverse crop rotations provide multiple opportunities through the year to apply manure and compost back in production fields. A growing appreciation of how agroecosystems can be designed to mimic the diversity and function of natural ecosystems provides the motivation to seek innovative and diverse crop rotations that maximize productivity and minimize negative impacts on the environment. Crop rotations provide multiple benefits to farmers and to society, and will be increasingly important in tomorrow's agriculture.

Further Reading

Altieri MA (1994) *Biodiversity and Pest Management in Agroecosystems*. New York: Food Products Press, Haworth Press.

Francis CA and Clegg MD (1990) Crop rotations in sustainable production systems. In: Edwards CA, Lal R, Madden P, Miller RH, and House G (eds) *Sustainable Agricultural Systems*, pp. 107–122. Ankeney, Iowa: Soil & Water Conservation Society.

Francis CA, Flora CB, and King LD (eds) (1990) *Sustainable Agriculture in Temperate Zones*. New York: John Wiley.

Gustafson AF (1941) *Soils and Soil Management*. New York: McGraw-Hill.

Karlen DL, Varvel GE, Bullock DG, and Cruse RM (1994) Crop rotations for the 21st century. *Advances in Agronomy* 53: 1–45.

Lal R and Pierce FJ (1991) *Soil Management for Sustainability*. Madison, WI: Soil Science Society of America.

Olson R, Francis C, and Kaffka S (eds) (1995) *Exploring the Role of Diversity in Sustainable Agriculture*. Madison, WI: American Society of Agronomy.

Thurston HD (1992) *Sustainable Practices for Plant Disease Management in Traditional Farming Systems*. Boulder, CO: Westview Press.

CROP WATER REQUIREMENTS

L S Pereira and I Alves, International Commission on Agricultural Engineering (CIGR), Lisbon, Portugal

Introduction

Crop water requirements (CWR) are defined as the depth of water (millimeters) needed to meet the water consumed through evapotranspiration (ET_c) by a disease-free crop, growing in large fields under nonrestricting soil conditions, including soil water and fertility, and achieving full production potential under the given growing environment. Defining 'crop evapotranspiration' (ET_c) as the rate of evapotranspiration (millimeters per day) of a given crop as influenced by its growth stages, environmental conditions, and crop management to achieve the potential crop production, then the CWR is the sum of ET_c for the entire crop growth period. When management or environmental conditions deviate from the optimal, then that rate of evapotranspiration has to be adjusted to the prevailing conditions and is called actual crop evapotranspiration (ET_a). Both CWR and ET_c concepts apply to either irrigated or rainfed crops.

For irrigated crops, the concept of CWR has to be complemented by that of irrigation water requirement (IWR), which is the net depth of water (millimeters) that is required to be applied to a crop to satisfy fully its specific crop water requirement. The IWR is the fraction of CWR not satisfied by rainfall, soil-water storage, and groundwater contribution. When it is necessary to add a leaching fraction to assure appropriate leaching of salts in the soil profile, this depth of water is also included in IWR. In practice, IWR has to be converted into gross irrigation requirements to take into consideration the efficiency of the irrigation systems utilized.

Crop Evapotranspiration

The rate of evapotranspiration (ET) can be computed with the Penman–Monteith (PM) equation:

$$\mathrm{ET} = \frac{1}{\lambda}\frac{\Delta(R_\mathrm{n} - G) + \rho c_\mathrm{p}(e_\mathrm{s} - e_\mathrm{a})/r_\mathrm{a}}{\Delta + \gamma(1 + r_\mathrm{s}/r_\mathrm{a})} \qquad [1]$$

where λ is the latent heat of vaporization (kilojoules per kilogram), $R_\mathrm{n}-G$ is the net balance of energy available at the surface (kilojoules per square meter per second), $(e_\mathrm{s}-e_\mathrm{a})$ represents the vapor pressure deficit (VPD) of air at the reference (weather measurement) height (kilopascals), ρ represents mean air density (kilograms per cubic meter), c_p represents specific heat of air at constant pressure (kilojoules per kilogram per degrees Celsius), Δ represents the slope of the saturation vapor pressure–temperature relationship at mean air temperature (kilopascals per degrees Celsius), γ is the psychometric constant (kilopascals per degrees Celsius), r_s is the bulk surface resistance (seconds per meter), and r_a is the aerodynamic resistance (seconds per meter).

The PM eqn [1] can be utilized for the direct calculation of ET_c because the surface and aerodynamic resistances are crop-specific. However, data for these crop characteristics are scarce for most crops.

The transfer of heat and vapor from the evaporative surface into the air in the turbulent layer above a canopy is determined by the aerodynamic resistance r_a (seconds per meter) between the surface and the reference level above the canopy:

$$r_\mathrm{a} = \frac{\ln\left(\frac{z_\mathrm{m} - d}{z_\mathrm{om}}\right)\ln\left(\frac{z_\mathrm{h} - d}{z_\mathrm{oh}}\right)}{k^2 u_z} \qquad [2]$$

where z_m is the height of wind velocity measurements (meters), z_h is the height of air temperature and humidity measurements (meters), d is the zero-plane displacement height (meters), z_om is the roughness length relative to momentum transfer (meters), z_oh is the roughness length relative to heat and vapor transfer (meters), u_z is the wind velocity at height z_m (meters per second), and k is the von Karman constant (0.41).

Equation [2] assumes that the evaporative surface may be represented as a 'big leaf' inside the canopy. However, exchanges in the top layer of the canopy between heights $d + z_\mathrm{om}$ and the crop height h (meters) are important as sources of vapor fluxes. Adopting $d + z_\mathrm{om}$ as the level of the evaporative surface can lead to overestimation of r_a and underestimation of r_s. Thus, an alternative r_a can be computed from the top of the canopy:

$$r_\mathrm{a} = \frac{\ln\left(\frac{z_\mathrm{m} - d}{z_\mathrm{om}}\right)\ln\left(\frac{z_\mathrm{h} - d}{h - d}\right)}{k^2 u_z} \qquad [3]$$

Both parameters d and z_om depend upon the crop height, h, and canopy architecture. Information exists relating d and z_om to h. Most of these relationships are crop-specific. More general functions also consider the leaf area index (LAI) or the plant area index (Table 1).

The height z_oh is estimated as a fraction of z_om, commonly $z_\mathrm{oh} = 0.1\, z_\mathrm{om}$ for short and fully developed canopies. The factor 0.2 is often preferred for tall and partial-cover crops. However, there is a relatively small impact on ET calculations from selecting a z_oh:z_om ratio between 0.1 and 0.2.

The surface resistance, r_s (seconds per meter), for full-cover canopies is often expressed by:

$$r_\mathrm{s} = r_\mathrm{l}/\mathrm{LAI}_\mathrm{eff} \qquad [4]$$

where r_l is the bulk stomatal resistance of a well-illuminated leaf (seconds per meter), and $\mathrm{LAI}_\mathrm{eff}$ is the effective leaf area index (dimensionless), usually taken as 0.5 LAI. r_l usually increases as a crop matures and begins to ripen. Typical values for r_l and r_s are listed in Table 2. The use of these equations for prediction of crop water requirements is difficult due to differences among varieties and crop-management practices. Information on stomatal conductance or stomatal resistance available in the literature is mainly

Table 1 Relationships of d and z_om with crop height (h) and leaf area index (LAI)

d	z_{om}	*Comment*
0.7 h	0.29 h	Theoretical study
0.6–0.7 h	0.12–0.14 h	
0.7 h	0.1 h	Complete cover crops
0.79 h		Cotton
(0.78 ± 0.04) h	(0.041 ± 0.002) h	Potato
0.75 h	0.065 h	Beans
0.644 h	0.034–0.092 h	Peas
(0.66 ± 0.04) h	(0.077 ± 0.04) h	Beans
0.4–0.63 h	0.03–0.13 h	Soybean
0.54–0.604 h	0.076–0.0964 h	Sorghum
0.56 h	0.06 h	Soybean
0.56 h	0.05 h	Wheat
0.56 h	0.11 h	Grass
1.04 $h^{0.88}$	0.062 $h^{1.108}$	
1.04 $h^{0.88}$	0.062 $h^{1.08}$	Rice, maize
	0.025 $h^{1.1}$	Maize, sugarcane
	0.105 h	Barley, potato, alfalfa
0.70 $h^{0.979}$	0.131 $h^{0.997}$	
$h\left[1 - 2/\mathrm{LAI}\left(1 - e^{-\mathrm{LAI}/2}\right)\right]$	$he_{-\mathrm{LAI}/2}\left(1 - e^{-\mathrm{LAI}/2}\right)$	

Adapted from Alves I (1995) *Modelação da Evapotranspiração Cultural. Resistências Aerodinâmica e do Coberto*. [Modeling crop evapotranspiration. Aerodynamic and surface resistances.] PhD thesis. Lisbon: instituto Superior de Agronomia.

Table 2 Typical values of the leaf resistance per unit leaf area (r_l) and bulk stomatal resistance (r_s) for several canopy types. Parameters $r_{l_{min}}$ and $r_{s_{min}}$ are minimum daytime values when all environmental variables are optimum

Canopy type	r_l *(s m^{-1})*	$r_{l_{min}}$ *(s m^{-1})*	r_s *(s m^{-1})*	$r_{s_{min}}$ *(s m^{-1})*
Tropical forest	–	–	125–150	50
Deciduous forest	–	45–150	70–160	50–60
Aspen	400	–	–	–
Eucalyptus	200–400	–	–	–
Maple	400–700	250	–	–
Coniferous	–	–	70–150	30–60
Crops, general	50–320	–	30–130	20–150
Grain sorghum	200	150–200	100–140[a]	–
Snapbeans	–	130	–	–
Soybean	120	–	–	50
Maize	160	70	80	25–40
Barley	150–250	–	45–70	40
Wheat	–	50	–	30
Alfalfa	80	50	40	–
Sugar beet	100	50	–	–
Clipped grass (0.15 m)	100–150	–	80–120	70
Clipped and irrigated grass (0.10–0.12 m)	75	40	40–60	25–30

[a] LAI = 2.

Adapted from Allen RG, Pereira LS, Raes D, and Smith M (1998) *Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper 56. Rome, Italy: Food and Agriculture Organization of the UN.

oriented to physiological or ecophysiological studies, rather than practical agricultural management. Information on bulk stomatal resistances is scarce.

Resistances r_l and r_s are influenced by climate and water availability: r_s increases when soil water availability limits ET, the VPD increases, and r_a is high. r_s decreases when energy available at the surface increases. In general, r_s varies according to:

$$r_s = r_a\left(\frac{\Delta}{\gamma}\beta - 1\right) + (1+\beta)\frac{\rho c_p \mathrm{VPD}}{\gamma(R_n - G)} \qquad [5]$$

where β is the Bowen ratio (the ratio between the sensible and latent heat fluxes). In eqn [5], β plays the role of a water-stress indicator. This equation shows that weather variables interact and their influences are interdependent, thus adding to the difficulties in appropriately selecting r_s.

These difficulties create challenges in applying the PM equation or other 'multilayer' resistance equations to estimate ET from agricultural crop canopies. Current research work is focused on improving our ability to apply the PM equation or multilayer ET models to specific agricultural crops; this work often utilizes relatively complex computer models. Meanwhile, the PM equation is used to compute the reference evapotranspiration and to determine $\mathrm{ET_c}$ with crop coefficients.

Crop Coefficients

$\mathrm{ET_c}$ can be calculated by multiplying the reference evapotranspiration, $\mathrm{ET_o}$ (millimeters per day), by a dimensionless crop coefficient, K_c:

$$\mathrm{ET_c} = K_c \mathrm{ET_o}$$

The reference crop is a hypothetical crop with an assumed height of 0.12 m, having a surface resistance of 70 s m^{-1} and an albedo of 0.23, closely resembling an extensive surface of green grass of uniform height, actively growing and adequately watered. $\mathrm{ET_o}$ can then be computed easily with the PM eqn [1], since the aerodynamic and surface-resistance terms can be parameterized, resulting in the United Nations Food and Agriculture Organization (FAO)–Penman–Monteith (FAO-PM) equation:

$$\mathrm{ET_o} = \frac{0.408\Delta(R_n - G) + \gamma\dfrac{900}{T+273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \qquad [7]$$

where, in addition to variables defined for eqn [1], T is mean daily air temperature (degrees Celsius) and u_2 is wind speed (meters per second), both at 2 m height. In this equation $\mathrm{ET_o}$ is in mm per day and $R_n - G$ in mJ m^{-2} day^{-1}.

The reference crop – corresponding to a living, agricultural crop (i.e., a cold-season clipped grass) – incorporates the majority of the weather effects into $\mathrm{ET_o}$ estimates. Therefore, since $\mathrm{ET_o}$ represents an index of climatic demand on evaporation, the K_c varies predominantly with the specific crop characteristics and little with climate. This enables the transfer of standard values for K_c between locations and climates.

K_c represents an integration of the effects of three primary characteristics that distinguish the crop from the reference: crop height (affecting roughness and aerodynamic resistance); crop–soil surface resistance (related to leaf area, fraction of ground covered by vegetation, leaf age and condition, degree of stomatal control, and soil surface wetness); and albedo of the crop–soil surface (influenced by the fraction of ground covered by vegetation and soil surface wetness).

Two K_c approaches are considered. The first uses a time-averaged K_c to include multiday effects of evaporation from the soil. The second concerns the basal crop coefficient and a separate calculation of evaporation from the soil.

The crop coefficient curve represents the changes in K_c over the length of the growing season (Figure 1). Its shape relates to changes in the vegetation and ground cover during plant development and maturation that affect the ratio ET_c:ET_o. Shortly after planting of annuals, or after the initiation of new leaves for perennials, the value for K_c is often small. The K_c increases from that initial value, $K_{c_{ini}}$, at the beginning of rapid plant development and reaches a maximum, $K_{c_{mid}}$, at the time of maximum or near-maximum plant development, the midseason period. During the late-season period, as leaves begin to senesce, the K_c begins to decrease until it reaches a lower value, $K_{c_{end}}$, at the end of the growing period.

The form for the equation used in the dual K_c approach is:

$$K_c = K_s K_{cb} + K_e \qquad [8]$$

where K_s is the stress-reduction coefficient (0–1), K_{cb} is the basal crop coefficient (0–~1.4), and K_e is the soil-water evaporation coefficient (0–~1.4). K_{cb} represents the ratio ET_c:ET_o when the soil surface layer is dry but the average soil-water content of the root zone is adequate to sustain full plant transpiration, thus representing the baseline potential K_c in the absence of evaporation from the soil (Figure 2). K_s reduces the value of K_{cb} when the soil-water content is not adequate.

Because eqn [8] requires the calculation of a daily soil-water balance for the surface soil layer, a simplification is required for routine application. The time-averaged K_c is then adopted:

$$K_c = (K_{cb} + K_e) \qquad [9]$$

where ($K_{cb} + K_e$) represents the sum of the basal K_{cb} and time-averaged effects of evaporation from the soil, K_e. Typical shapes for the K_{cb}, K_e, and $K_{cb} + K_e$ curves are shown in Figure 2. When summed, the values for K_{cb} and for K_e represent the total crop coefficient, K_c. The time-averaged K_c is used for planning, irrigation system design, and typical irrigation management. The dual K_c is best where effects of day-to-day variation in soil surface wetness are important to estimate the resulting impacts on daily ET_c, soil moisture profile, and deep percolation.

The Single-Crop Coefficient Approach

A simple procedure may be used to construct the K_c curve (Figure 3):

1. Divide the growing period into four general growth stages that describe crop phenology or development, and determine the lengths (days) of these stages. The four crop growth periods are:
   a. Initial: for annual crops, duration is from planting date to approximately 10% ground

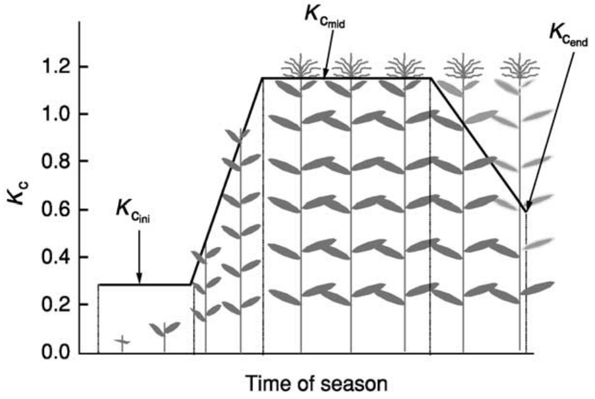


Figure 1 Generalized crop coefficient curve (K_c) during a growing season. Reproduced from Allen RG, Pereira LS, Raes D, and Smith M (1998) *Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper 56. Rome, Italy: Food and Agriculture Organization of the UN.

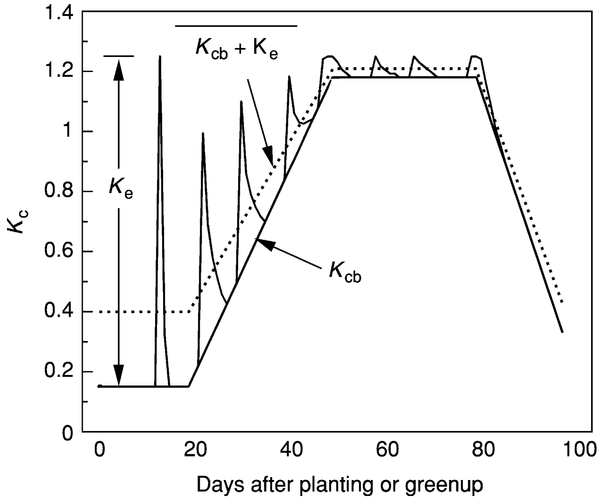


Figure 2 Crop coefficient (K_c) definitions showing the basal K_{cb}, soil evaporation K_c, and time-averaged $K_{cb} + K_e$ values. Reproduced from Allen RG, Pereira LS, Raes D, and Smith M (1998) *Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper 56. Rome, Italy: Food and Agriculture Organization of the UN.

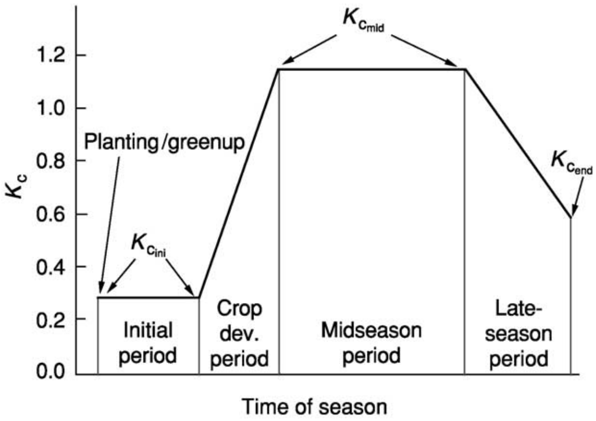


Figure 3 Crop coefficient curve (K_c) and stage definitions. Reproduced from Allen RG, Pereira LS, Raes D, and Smith M (1998) *Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper 56. Rome, Italy: Food and Agriculture Organization of the UN.

    cover. For perennials, the planting date is replaced by the 'greenup' date, when initiation of new leaves occurs;

   b. Crop development: from 10% ground cover to effective full cover, which often occurs at the initiation of flowering or when LAI reaches 3;
   c. Midseason: from effective cover to start of maturity, which is often indicated by the beginning of the aging, yellowing or senescence of leaves, leaf drop, or the browning of fruit;
   d. Late season: from start of maturity to harvest or full senescence. For some perennial vegetation in frost-free climates, crops may grow year-round so that the date of termination may be taken as the same as the date of 'planting';

2. Identify the three K_c values that correspond to $K_{c_{ini}}$, $K_{c_{mid}}$, and $K_{c_{end}}$;
3. Connect straight-line segments through each of the four growth-stage periods.

The length of crop growth stages is crop-specific and changes duration with crop variety, planting date, cultivation practices, and weather conditions, mainly air temperature. The length of crop growth stages may be predicted using cumulative, degree-based equations or plant growth models.

The lengths of the initial and development periods may be relatively short for deciduous trees and shrubs that develop new leaves in the spring at relatively fast rates. The $K_{c_{ini}}$ should then reflect the ground condition prior to leaf initiation, including the amount of grass or weed cover, soil wetness, tree density, and mulch density. The length of the late-season period may be relatively short for vegetation killed by frost or for crops harvested before senescence. The value for $K_{c_{end}}$ should reflect the soil surface condition and that of the vegetation following plant death or harvest. Indicative lengths of growth stages are given in FAO guides. However, local observations or information should be used to incorporate effects of plant variety, climate, and cultural practices.

Values for $K_{c_{ini}}$, $K_{c_{mid}}$, and $K_{c_{end}}$ are listed in Table 3 for various agricultural crops. Usually there is close similarity in K_c within the same crop group, since the plant height, leaf area, ground coverage, and water management are usually similar.

The K_c values in Table 3 represent potential water use by healthy, disease-free, and densely planted stands of vegetation, with adequate levels of soil water. When stand density, height, and leaf area are less than that attained under perfect or normal conditions, K_c should be reduced by as much as 0.3–0.5 for poor stands, according to the amount of effective leaf area relative to healthy vegetation with normal planting densities.

The $K_{c_{ini}}$ values in Table 3 are only approximate, because they vary widely with soil wetting conditions and because ET during the initial stage for annual crops is predominantly in the form of evaporation from the soil. Therefore, estimates for $K_{c_{ini}}$ must consider the frequency of irrigation and rainfall that wet the soil surface.

Evaporation from bare soil, E_s (millimeters per day), can be characterized as occurring in two stages (Figure 4). During stage 1, termed the 'energy-limited' stage and having a duration t_1 (days), moisture is transported to the soil surface at a rate sufficient to supply the potential rate of evaporation, E_{so} (millimeters per day), which is governed by energy availability at the soil surface. E_{so} can be estimated from:

$$E_{so} = 1.15\mathrm{ET_o} \qquad [10]$$

where $\mathrm{ET_o}$ is averaged for the initial period.

Stage 2 is termed the 'soil water-limited' stage, where hydraulic transport of subsurface water to the soil surface is smaller, thus making $E_s < E_{so}$. Some of the evaporation occurs from below the soil surface, and energy is supplied by transport of heat into the soil profile. E_s decreases as soil moisture decreases and can be assumed to be linearly proportional to the depth of water remaining in the evaporation layer.

When the time interval (days) between two successive wettings is $t_w > t_1$, $K_{c_{ini}}$ is approached as:

$$K_{c_{ini}} = \frac{\mathrm{TEW} - (\mathrm{TEW} - \mathrm{REW})\exp\left(\frac{-(t_w - t_1)E_{so}\left[1 + \frac{\mathrm{REW}}{\mathrm{TEW}-\mathrm{REW}}\right]}{\mathrm{TEW}}\right)}{t_w \mathrm{ET_o}} \qquad [11]$$

Table 3 Single (time-averaged) crop coefficients (K_c) and basal crop coefficient (K_{cb}), for nonstressed, well-managed crops in subhumid climates ($RH_{min} \approx 45\%$, $u_2 \approx 2\ m\ s^{-1}$) for use with the FAO Penman–Monteith ET_o, and indicative mean maximum plant heights (h), maximum root depths ($z_{r_{max}}$), and depletion fractions for no stress (p)

Crop	$K_{c_{ini}}$ [a]	$K_{c_{mid}}$	$K_{c_{end}}$	$K_{cb_{ini}}$	$K_{cb_{mid}}$	$K_{cb_{end}}$	*Maximum crop height* h *(m)*	*Maximum root depth*[b] $Z_{r_{max}}$ *(m)*	*Depletion fraction*[c] *(for ET 5 mm day^{-1})*p
Vegetables	0.7	1.05	0.95	0.15	0.95	0.85			
Cabbage		1.05	0.95		0.95	0.85	0.4	0.5–0.8	0.45
Carrots		1.05	0.95		0.95	0.85	0.3	0.5–1.0	0.35
Lettuce		1.00	0.95		0.90	0.90	0.3	0.3–0.5	0.30
Onions, dry		1.05	0.75		0.95	0.65	0.4	0.3–0.6	0.30
Onions, green		1.00	1.00		0.90	0.90	0.3	0.3–0.6	0.30
Tomato		1.15[d]	0.70–0.90		1.10[d]	0.60–0.80	0.6	0.7–1.5	0.40
Cucumber, fresh market	0.6	1.00[d]	0.75		0.95[d]	0.70	0.3	0.7–1.2	0.50
Pumpkin, winter squash		1.00	0.80		0.95	0.70	0.4	1.0–1.5	0.35
Watermelon	0.4	1.00	0.75		0.95	0.70	0.4	0.8–1.5	0.40
Roots and tubers	0.5	1.10	0.95	0.15	1.00	0.85			
Cassava, year 1	0.3	0.80	0.30		0.70	0.20	1.0	0.5–0.8	0.35
Cassava, year 2	0.3	1.10	0.50		1.00	0.45	1.5	0.7–1.0	0.40
Potato		1.15	0.75–0.40		1.10	0.65–0.30	0.6	0.4–0.6	0.35
Sugar beet	0.35	1.20	0.70[e]		1.15	0.50[e]	0.5	0.7–1.2	0.55
Legumes (Leguminosae)	0.4	1.15	0.55	0.15	1.10	0.50			
Beans, green	0.5	1.05[d]	0.90		1.00[d]	0.80	0.4	0.5–0.7	0.45
Beans, dry and pulses	0.4	1.15[d]	0.35		1.10[d]	0.25	0.4	0.6–0.9	0.45
Garbanzo	0.4	1.15	0.35		1.05	0.25	0.8	0.6–1.0	0.45
Soybeans		1.15	0.50		1.10	0.30	0.5–1.0	0.6–1.3	0.50
Perennial vegetables (with winter dormancy and initially bare or mulched soil)	0.5	1.00	0.80						
Asparagus	0.5	0.95	0.30	0.15	0.90	0.20	0.2–0.8	1.2–1.8	0.45
Strawberries	0.40	0.85	0.75	0.30	0.80	0.70	0.2	0.2–0.3	0.20
Fiber crops	0.35			0.15					
Cotton		1.15–1.20	0.70–0.50		1.10–1.15	0.50–0.40	1.2–1.5	1.0–1.7	0.65
Oil crops	0.35	1.15	0.35	0.15	1.10	0.25			
Safflower		1.0–1.15[f]	0.25		0.95–1.10[f]	0.20	0.8	1.0–2.0	0.60
Sunflower		1.0–1.15[f]	0.35		0.95–1.10[f]	0.25	2.0	0.8–1.5	0.45
Cereals	0.3	1.15	0.4	0.15	1.10	0.25			
Spring wheat		1.15	0.25–0.40		1.10	0.15–0.30	1	1.0–1.5	0.55
Winter wheat, frozen soils	0.4	1.15	0.25–0.40	0.15–0.50	1.10	0.15–0.30	1	1.5–1.8	0.55
Winter wheat, nonfrozen soils	0.7	1.15	0.25–0.40						
Maize, field (grain; field corn)		1.20	0.60,0.35	0.15	1.15	0.50,0.15	2	1.0–1.7	0.55
Rice	1.05	1.20	0.90–0.60	1.00	1.15	0.70–0.45	1	0.5–1.0	0.20[g]

(*Continued*)

Table 3 (*Continued*)

Crop	$K_{c_{ini}}$ [a]	$K_{c_{mid}}$	$K_{c_{end}}$	$K_{cb_{ini}}$	$K_{cb_{mid}}$	$K_{cb_{end}}$	*Maximum crop height* h *(m)*	*Maximum root depth*[b] $Z_{r_{max}}$ *(m)*	*Depletion fraction*[c] *(for ET 5 mm day*$^{-1}$*)*p
Forages									
Alfalfa hay, averaged cutting effects	0.40	0.95	0.90				0.7		
Alfalfa hay, individual cutting periods	0.40[h]	1.20[h]	1.15[h]	0.30[h]	1.15[h]	1.10[h]	0.7	1.0–2.0	0.55
Grazing pasture, rotated grazing	0.40	0.85–1.05	0.85	0.30	0.80–1.00	0.80	0.15–0.30	0.5–1.5	0.60
Grazing pasture, extensive grazing	0.30	0.75	0.75	0.30	0.70	0.70	0.10	0.5–1.5	0.60
Turf grass, cool season	0.90	0.95	0.95	0.85	0.90	0.90	0.10	0.5–1.0	0.40
Turf grass, warm season	0.80	0.85	0.85	0.75	0.80	0.80	0.10	0.5–1.0	0.50
Sugar cane	0.40	1.25	0.75	0.15	1.20	0.70	3	1.2–2.0	0.65
Tropical fruits and trees									
Banana, first year	0.50	1.10	1.00	0.15	1.05	0.90	3	0.5–0.9	0.35
Banana, second year	1.00	1.20	1.10	0.60	1.10	1.05	4	0.5–0.9	0.35
Date palms	0.90	0.95	0.95	0.80	0.85	0.85	8	1.5–2.5	0.50
Pineapple, bare soil	0.50	0.30	0.30	0.15	0.25	0.25	0.6–1.2	0.3–0.6	0.50
Pineapple, with grass cover	0.50	0.50	0.50	0.30	0.45	0.45	0.6–1.2		
Grapes and berries									
Berries (bushes)	0.30	1.05	0.50	0.20	1.00	0.40	1.5	0.6–1.2	0.50
Grapes, table or raisin	0.30	0.85	0.45	0.15	0.80	0.40	2	1.0–2.0	0.35
Grapes, wine	0.30	0.70	0.45	0.15	0.65	0.40	1.5–2	1.0–2.0	0.45
Fruit trees									
Apples, cherries, pears								1.0–2.0	0.50
No ground cover, no frosts	0.60	0.95	0.75[i]	0.50	0.90	0.70[i]	4		
Active ground cover, no frosts	0.80	1.20	0.85[i]	0.75	1.15	0.80[i]	4		
Apricots, peaches, stone fruit								1.0–2.0	0.50
No ground cover, no frosts	0.55	0.90	0.65[i]	0.45	0.85	0.60[i]	3		
Active ground cover, no frosts	0.80	1.15	0.85[i]	0.75	1.10	0.80[i]	3		
Citrus, no ground cover									
70% canopy	0.70	0.65	0.70	0.65	0.60	0.65	4	1.2–1.5	0.50
50% canopy	0.65	0.60	0.65	0.60	0.55	0.60	3	1.1–1.5	0.50
20% canopy	0.50	0.45	0.55	0.45	0.40	0.50	2	0.8–1.1	0.50
Olives (40–60% ground coverage by canopy)[j]	0.65	0.70–0.45	0.70–0.65	0.55	0.65–0.40	0.65–0.60	3–5	1.2–1.7	0.65

Reproduced from Allen RG, Pereira LS, Raes D, and Smith M (1998) *Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper 56. Rome, Italy: Food and Agriculture Organization of the UN.

[a] $K_{c_{ini}}$ in the table is only indicative. The procedure relative to eqns [11] and [12] should be used.

[b] The larger values for $z_{r_{max}}$ correspond to rainfed conditions.

[c] The tabulated value for p must be corrected when $ET_c \neq 5\,\text{mm day}^{-1}$ using $p = p_{table} + 0.04\,(5 - ET_c)$.

[d] For crops grown on stalks, $K_{c_{mid}}$ and $K_{cb_{mid}}$ have to be increased by 0.05 to 0.10 and h should be increased too.

[e] The $K_{c_{end}}$ and $K_{cb_{end}}$ are increased by 0.1 to 0.3 when irrigation or significant rain occurs during the last month.

[f] The smaller value is for rainfed conditions when plant density is smaller than under irrigation.

[g] This p-value represents the depletion fraction below soil moisture at saturation.

[h] These K_c coefficients for hay crops represent immediately after cutting, at full cover, and immediately before cutting, respectively. The growing season is described as a series of individual cutting periods.

[i] These $K_{c_{end}}$ (and $K_{cb_{end}}$) values represent K_c prior to leaf drop. After leaf drop, $K_{c_{end}} \approx 0.20$ (0.15) for bare, dry soil or dead ground cover and $K_{c_{end}} \approx 0.50$–0.80 (0.45–0.75) for actively growing ground cover.

[j] The larger values correspond to well-watered conditions and the smaller concern controlled, stressed conditions.

where REW is the readily evaporable water, corresponding to the depth of evaporation when stage 1 drying is complete (millimeters), TEW is the total evaporable water, i.e., the maximum evaporation depth when soil evaporation effectively ceases. From the concept of stage 1 drying results $t_1 = \mathrm{REW}/E_{\mathrm{so}}$.

When $t_{\mathrm{w}} < t_1$, the entire process resides within stage 1, then:

$$K_{\mathrm{c_{ini}}} = E_{\mathrm{so}}/\mathrm{ET_o} \qquad [12]$$

Where furrow or trickle irrigation is practiced, and only a portion of the soil surface is wetted, $K_{\mathrm{c_{ini}}}$ in eqns [11] and [12] should be reduced in proportion to the average fraction of wetted soil surface, f_{w} (ranging from 0, when no rain or irrigation occurs, to 1). Indicative values for f_{w} are shown in Table 4. The infiltration depth from irrigation, I_{w} (millimeters), should also be adjusted:

$$I_{\mathrm{w}} = I/f_{\mathrm{w}} \qquad [13]$$

where I is the total irrigation depth (millimeters).

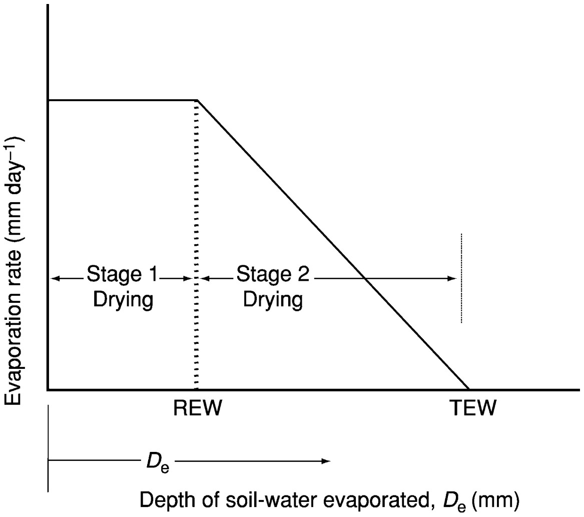


Figure 4 Two-stage model for soil evaporation. REW, readily evaporable water; TEW, total evaporable water. Reproduced from Allen RG, Pereira LS, Raes D, and Smith M (1998) *Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper 56. Rome, Italy: Food and Agriculture Organization of the UN.

Table 4 Indicative values of average fraction of wetted soil surface, f_w

Irrigation method	f_w
Rain, sprinkling, basin, and border irrigation	1.0
Furrow irrigation	0.4–0.6
Irrigation with alternate furrows	0.3–0.4
Trickle irrigation	0.2–0.5

REW is higher for medium-textured soils and is lower for coarse soils. Maximum values for REW ($\mathrm{REW_{max}}$) may be predicted according to soil texture:

$$\mathrm{REW_{max}} = 20 - 0.15(\mathrm{Sa}) \quad \text{for Sa} > 80\%$$

$$\mathrm{REW_{max}} = 11 - 0.06(\mathrm{Cl}) \quad \text{for Cl} > 50\%$$

$$\mathrm{REW_{max}} = 8 + 0.08(\mathrm{Cl}) \quad \text{for Sa} < 80\%, \mathrm{Cl} > 50\% \qquad [14]$$

where Sa and Cl are the fractions of sand and clay in the soil (percentage).

The TEW value is governed by the depth of soil contributing to evaporation, z_{e} (100–150 mm). The soil water-holding properties within this evaporative layer, the presence of a hydraulically limiting layer beneath it, the unsaturated hydraulic conductivity, the conduction of sensible heat into the soil, and any root extraction of water from the evaporative layer all influence TEW. An approximation to $\mathrm{TEW_{max}}$ is:

$$\mathrm{TEW_{max}} = z_{\mathrm{e}}(\theta_{\mathrm{FC}} - 0.5\theta_{\mathrm{WP}}) \quad (\mathrm{ET_o} \geq 5\,\mathrm{mm\,day^{-1}}) \qquad [15]$$

$$\mathrm{TEW_{max}} = z_{\mathrm{e}}(\theta_{\mathrm{FC}} - 0.5\theta_{\mathrm{WP}})\sqrt{\frac{\mathrm{ET_o}}{5}} \quad (\mathrm{ET_o} < 5\,\mathrm{mm\,day^{-1}}) \qquad [16]$$

where θ_{FC} and θ_{WP} are the soil-water content at field capacity and wilting point (millimeters per millimeter). Typical values for θ_{FC} and θ_{WP} are given in Table 5.

The average total water available for evaporation, D_{a} (millimeters), during each drying cycle is computed from the average depth added to the evaporative layer at each wetting:

$$D_{\mathrm{a}} = P_{\mathrm{mean}} + W_{\mathrm{ini}}/n_{\mathrm{w}} \qquad [17]$$

where W_{ini} is the available soil water (millimeters) in the evaporation layer at the time of planting, n_{w} is the number of wetting events, and P_{mean} is the average depth (millimeters) of water added to the evaporating layer at each wetting event. P_{mean} can be obtained with:

$$P_{\mathrm{mean}} = \left(\sum P_{\mathrm{n}} + \sum I_{\mathrm{w}}\right)/n_{\mathrm{w}} \qquad [18]$$

but where each value of P_{n} and I_{w} must be limited to $P_{\mathrm{n}} \leq \mathrm{TEW_{max}}$ and $I_{\mathrm{w}} \leq \mathrm{TEW_{max}}$. The values for TEW and REW in eqn [11] are calculated from $\mathrm{TEW_{max}}$ and $\mathrm{REW_{max}}$ as:

$$\mathrm{TEW} = \min(\mathrm{TEW_{max}}, D_{\mathrm{a}}) \qquad [19]$$

and:

Table 5 Typical soil water characteristics for different soil types

Soil type (USA soil texture classification)	*Soil water characteristics*			*Amount of water that can be depleted by evaporation for ze = 0.10 m*	
	θ_{FC} m^3/m^3	θ_{WP} m^3/m^3	$(\theta_{FC}-\theta_{WP})$ m^3/m^3	*Stage 1 REW mm*	*Stages 1 and 2 TEW mm*
Sand	0.07–0.17	0.02–0.07	0.05–0.11	2–7	6–12
Loamy sand	0.11–0.19	0.03–0.10	0.06–0.12	4–8	9–14
Sandy loam	0.18–0.28	0.06–0.16	0.11–0.15	6–10	15–20
Loam	0.20–0.30	0.07–0.17	0.13–0.18	8–10	16–22
Silt loam	0.22–0.36	0.09–0.21	0.13–0.19	8–11	18–25
Silt	0.28–0.36	0.12–0.22	0.16–0.20	8–11	22–26
Silt clay loam	0.30–0.37	0.17–0.24	0.13–0.18	8–11	22–27
Silty clay	0.30–0.42	0.17–0.29	0.13–0.19	8–12	22–28
Clay	0.32–0.40	0.20–0.24	0.12–0.20	8–12	22–29

(TEW = q FC − 0.5 qWP) ze.

Reproduced from Allen RG, Pereira LS, Raes D, and Smith M (1998) *Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements.* FAO Irrigation and Drainage Paper 56. Rome, Italy: Food and Agriculture Organization of the UN.

$$\mathrm{REW} = \mathrm{REW}_{\max}\left[\min\left(\frac{D_a}{\mathrm{TEW}_{\max}}, 1\right)\right] \qquad [20]$$

When eqns 14–16 are used, appropriate checking of results is required.

K_c Adjustment for Climate

The $K_{c_{\mathrm{mid}}}$ and $K_{c_{\mathrm{end}}}$ values in Table 3 represent $K_{cb} + K_e$ for irrigation management and precipitation frequencies typical of a subhumid climate where $\mathrm{RH}_{\min} = 45\%$ and $u_2 = 2\,\mathrm{m\,s^{-1}}$.

Under humid and calm conditions, the K_c for 'full-cover' agricultural crops generally do not exceed 1.0 by more than about 0.05, because 'full-cover' agricultural crops and the reference crop behave similarly regarding absorption of short-wave radiation, the primary energy source for evaporation under humid and calm conditions. Because the VPD is small under humid conditions, differences in ET caused by differences in r_a between the agricultural and the reference crop are also small, especially with low-to-moderate wind speeds. Thus the values of K_c are less dependent on differences between the aerodynamic components of ET_c and ET_o. On the contrary, under arid conditions, the effect of differences in r_a between the agricultural and the reference crop on ET_c become more pronounced because the VPD is often large. Hence, K_c will be larger under arid conditions, mainly for tall crops. Because the $K_{c_{\mathrm{mid}}}$ and $K_{c_{\mathrm{end}}}$ in Table 3 represent conditions where $\mathrm{RH}_{\min} \approx 45\%$ and $u_2 \approx 2\,\mathrm{m\,s^{-1}}$, when climatic conditions deviate from these values, the tabled values need to be adjusted:

$$K_c = (K_c)_{\mathrm{tab}} + [0.04(u_2 - 2) - 0.004(\mathrm{RH}_{\min} - 45)]\left(\frac{h}{3}\right)^{0.3} \qquad [21]$$

where $(K_c)_{\mathrm{tab}}$ represents the $K_{c_{\mathrm{mid}}}$ or $K_{c_{\mathrm{end}}}$ taken from Table 3, u_2 is the average daily wind speed at 2 m height (meters per second), $\mathrm{RH}_{\min}$ is the average daily minimum relative humidity (percentage), and h is the average plant height (meters), all averages referring to the midseason or the late-season period. Indicative values for h are listed in Table 3, although data from field observations are more accurate. When crops are allowed to senesce and dry in the field ($K_{c_{\mathrm{end}}} < 0.45$), no adjustment is necessary.

K_c Adjustment for Nonpristine Conditions

The values of K_c in Table 3 reflect typical crop- and water-management practices. When local water management and harvest timing deviate from those that are typical, adjustments are made to $K_{c_{\mathrm{mid}}}$ and $K_{c_{\mathrm{end}}}$.

When stand density, height, or leaf area of the crop are less than those attained under appropriate crop and irrigation management conditions, the value for K_c is reduced by 0.1–0.5, according to the amount of effective (green) leaf area relative to that of healthy vegetation having normal plant density:

$$K_c = K_{c_{\mathrm{table}}} - A_{\mathrm{cm}} \qquad [22]$$

where A_{cm} is the adjustment factor (0–0.5) that can be approximated through a green-cover ratio of the type:

$$A_{\mathrm{cm}} = 1 - \mathrm{LAI}_{\mathrm{actual}}/\mathrm{LAI}_{\mathrm{normal}} \qquad [23]$$

where the LAI refers to the midseason period. Other procedures, including remote sensing, may be used to estimate the K_c values for nonpristine conditions.

The Dual-Crop Coefficient Approach

The basal crop coefficient, K_{cb} (eqn [8]), represents primarily the transpiration component of ET. Its use provides for separate adjustment for evaporation from wet soil immediately following rain or irrigation events. This results in more accurate estimates of $\mathrm{ET_c}$ when computed on a daily basis. Recommended values for K_{cb} are listed in Table 3: these must be adjusted for climate using a similar equation to eqn [21].

The computation of the soil-water evaporation coefficient K_{e} is based on the fact that evaporation from the soil is governed by the amount of energy available at the soil surface, which depends, in turn, on the portion of total energy that has been consumed by plant transpiration. K_{e} decays after a wetting depending on the cumulative amount of water evaporated from the surface soil layer. Thus K_{e} can be calculated from:

$$K_{\mathrm{e}} = K_{\mathrm{r}}(K_{\mathrm{c_{max}}} - K_{\mathrm{cb}}) \qquad [24]$$

where K_{r} is the evaporation-reduction coefficient (0–1) and $K_{\mathrm{c_{max}}}$ is the maximum value for K_{c} after rain or irrigation. However, K_{e} is limited by the fraction of wetted soil exposed to sunlight, f_{ew} (0.01–1):

$$K_{\mathrm{e}} \leq f_{\mathrm{ew}} K_{\mathrm{c_{max}}} \qquad [25]$$

$K_{\mathrm{c_{max}}}$ represents an upper limit on ET from any cropped surface (1.05–1.35). It changes with climate similarly to K_{c} (eqn [21]), thus:

$$K_{\mathrm{c_{max}}} = \max\rangle\Big\{1.2 + [0.04(u_2 - 2) - 0.004(\mathrm{RH_{min}} - 45)](h/3)^{0.3}\Big\}, K_{\mathrm{cb}} + 0.05\langle \qquad [26]$$

u_2, $\mathrm{RH_{min}}$, and h may refer to the midseason period or, when more detailed computations are applied, be averaged for shorter periods (e.g., 5 days). h for the initial period can be considered the same as for the grass reference crop ($h = 0.12$ m).

The method used to estimate evaporation from soil is similar to the one used to compute $K_{\mathrm{c_{ini}}}$, where the evaporation rate is at the maximum rate until the depth of water evaporated, D_{e} (millimeters), equals REW (Figure 4). When $D_{\mathrm{e}} > \mathrm{REW}$, the evaporation process is at stage 2, and its rate decreases in proportion to the remaining water. Therefore, K_{r} (eqn [24]) can be calculated as:

$$K_{\mathrm{r}} = 1 \quad \text{for } D_{\mathrm{e}} \leq \mathrm{REW} \qquad [27a]$$

$$K_{\mathrm{r}} = \frac{\mathrm{TEW} - D_{\mathrm{e}}}{\mathrm{TEW} - \mathrm{REW}} \quad \text{for } D_{\mathrm{e}} > \mathrm{REW} \qquad [27b]$$

REW and TEW can be estimated with eqns [14]–[20]. D_{e}, the current depth of water depleted from the f_{ew} fraction of wetted soil exposed to sunlight, is computed from the daily water balance of the upper 100–200 mm of soil:

$$D_{\mathrm{e_i}} = D_{\mathrm{e_{i-1}}} - (P_{\mathrm{i}} - \mathrm{RO_i}) - \frac{I_{\mathrm{i}}}{f_{\mathrm{w}}} + \left(\frac{K_{\mathrm{e}}\mathrm{ET_o}}{f_{\mathrm{ew}}}\right)_i + T_{\mathrm{s_i}} \qquad [28]$$

limited to ($0 \leq D_{\mathrm{e_i}} \leq \mathrm{TEW}$), where the subscript i refers to the day of estimation, P_{i} is the precipitation (millimeters), $\mathrm{RO_i}$ is runoff (millimeters; $0 \leq \mathrm{RO_i} \leq P_{\mathrm{i}}$), I_{i} is the net irrigation depth (millimeters) that infiltrates the soil (eqn [13]), $(K_{\mathrm{e}}\ \mathrm{ET_o}/f_{\mathrm{ew}})_{\mathrm{i}}$ is the evaporation from the f_{ew} fraction of the exposed soil surface (millimeters), and $T_{\mathrm{s_i}}$ is the transpiration from the f_{w} fraction of the evaporating soil layer (millimeters). To initiate the water balance, $\mathrm{D_{e_i}} = 0$ immediately following heavy rain or irrigation, or $\mathrm{D_{e_i}} = \mathrm{TEW}$ if a long time has passed since the last wetting. When $P_{\mathrm{i}} < 0.2\ \mathrm{ET_o}$, it can be ignored. For most applications, $\mathrm{RO_i} = 0$ and, for the majority of crops, except for very shallow-rooted crops, $T_{\mathrm{s_i}}$ can be neglected.

When the complete soil surface is fully wetted (e.g., by precipitation or sprinkler irrigation), $f_{\mathrm{ew}} = (1 - f_{\mathrm{c}})$, where f_{c} is the average fraction of ground covered by vegetation (0–0.99). For irrigation systems where only a fraction of the soil surface is wetted, f_{ew} is calculated as:

$$f_{\mathrm{ew}} = \min(1 - f_{\mathrm{c}}, f_{\mathrm{w}}) \qquad [29]$$

When not observed, f_{c} can be estimated daily from:

$$f_{\mathrm{c}} = \left(\frac{K_{\mathrm{cb}} - K_{\mathrm{c_{min}}}}{K_{\mathrm{c_{max}}} - K_{\mathrm{c_{min}}}}\right)^{1+0.5h} \qquad [30]$$

where $K_{\mathrm{c_{min}}}$ is the minimum K_{c} for dry, bare soil (0.15–0.20). The exponent $1 + 0.5h$ represents the effect of plant height on soil-shading and in increasing the K_{cb} given a specific value for f_{c}. $(K_{\mathrm{cb}} - K_{\mathrm{c_{min}}}) \geq 0.01$ for numerical stability.

K_{cb} values are reduced when the soil water content in the root zone is too low to sustain transpiration at potential levels. The reduction is made through the water stress coefficient, K_{s} (0–1):

$$K_{\mathrm{s}} = \frac{\theta - \theta_{\mathrm{WP}}}{\theta_{\mathrm{p}} - \theta_{\mathrm{WP}}} \quad (\theta < \theta_{\mathrm{p}}) \qquad [31]$$

where θ is average soil-water content in the root zone (millimeters per millimeter) and θ_{p} is the threshold θ below which transpiration is decreased due to water stress (millimeters per millimeter). By definition, $K_{\mathrm{s}} = 1.0$ for $\theta > \theta_{\mathrm{p}}$. The threshold θ_{p} is:

$$\theta_{\mathrm{p}} = (1 - p)(\theta_{\mathrm{FC}} - \theta_{\mathrm{WP}}) \qquad [32]$$

where p is the depletion fraction for no stress (0–1). Indicative values can be found in Table 3. The determination of K_{s} requires a daily balance of soil-water content.

K_c for Nonpristine and Unknown Conditions

For vegetation where the K_c is not known, but where estimates of the fraction of ground surface covered by vegetation can be made, $K_{cb_{mid}}$ can be approximated as:

$$K_{cb_{mid}} = K_{c_{min}} + \left(K_{cb_{full}} - K_{c_{min}}\right) f_{c_{eff}}^{1/1+h} \qquad [33]$$

where $f_{c_{eff}}$ is the effective fraction of ground covered by vegetation (0.01–1), and $K_{cb_{full}}$ is the maximum value for K_{cb} for vegetation having complete ground cover:

$$K_{cb_{full}} = \min[(1.0 + 0.1h), 1.2] + 0.04(u_2 - 2) - 0.004(\mathrm{RH}_{min} - 45)\left(\frac{h}{3}\right)^{0.3} \qquad [34]$$

For small, isolated stand sizes, $K_{cb_{full}}$ may need to be increased beyond the value given by eqn [34]. $K_{cb_{full}}$ may be reduced for vegetation that has a high degree of stomatal control.

Irrigation Water Requirements

The soil-water balance is calculated for the effective rooting depth as:

$$\theta_i = \theta_{i-1} + \frac{(P_i - \mathrm{RO}_i) + I_{w_i} - \mathrm{ET}_{c_i} - \mathrm{DP}_i + \mathrm{GW}_i}{1000 z_{r_i}} \qquad [35]$$

where, in addition to the symbols used previously, DP_i represents deep percolation (millimeters), GW_i is groundwater contribution (millimeters), and z_{r_i} is the rooting depth (meters), all referred to day i. DP is often estimated as $\mathrm{DP}_i = 0$ when $\theta_i \leq \theta_{FC}$ and $\mathrm{DP}_i = 1000 (\theta_i - \theta_{FC}) z_{r_i}$ otherwise. GW is estimated from soil hydraulic properties and the water table depth. z_{r_i} can be predicted assuming a linear variation from planting to maximum rooting. Maximum root depths for most common crops are presented in Table 3.

The latest date for scheduling irrigation to avoid water stress is when $\theta_i = \theta_p$ (eqn [32]). However, irrigation is often scheduled when the 'management-allowed depletion' (MAD) is attained. Generally, MAD $< p$ when there is risk aversion or uncertainty, and MAD $> p$ when plant water stress is intentional. Then:

$$\theta_i = \theta_{MAD} = (1 - \mathrm{MAD})(\theta_{FC} - \theta_{WP}) + \theta_{WP} \qquad [36]$$

The net irrigation depth to be applied will be:

$$I_{w_i} = 1000\, z_{r_i} (\theta_{FC} - \theta_i) \qquad [37]$$

which, summed for the entire season, leads to the IWR:

$$\mathrm{IWR} = \frac{\mathrm{ET}_c - P_e - \mathrm{GW} - \Delta S}{1 - \mathrm{LR}} \qquad [38]$$

where P_e is the effective precipitation (gross precipitation less all runoff and deep percolation), GW is groundwater contribution, ΔS is the change in soil-water storage in the root zone between planting and harvesting, and LR is the leaching requirement (the percentage of irrigation water that must pass through the root zone to keep the salinity of the soil below a specified value). The soil water balance is currently computed through crop-water simulation models, which allow the selection of the best irrigation scheduling alternatives.

The gross irrigation water requirement (GIWR) is computed as:

$$\mathrm{GIWR} = \frac{\mathrm{IWR}}{\mathrm{Eff}} \qquad [39]$$

where Eff is the efficiency of the irrigation system. Indicative values of the efficiencies are presented in Table 6.

Table 6 Indicative values of irrigation efficiencies

System	*Efficiency (%)*
Irrigation methods	
Surface irrigation, precision leveling	
Furrow	65–85
Border	70–85
Basin	70–90
Surface, traditional	
Furrow	40–70
Border	45–70
Basin	45–70
Basin, rice fields	25–50
Sprinkler	
Solid set	65–85
Hand-move lateral	65–80
Side-roll wheel move	65–80
Traveler sprinkler	55–70
Lateral move systems, center pivot	65–85
Microirrigation	
Trickle, $\geq$3 emitters per plant	85–95
Trickle, $<$3 emitters per plant	80–90
Bubblers and sprayers	85–95
Line-source emitters	70–90
Distribution and transport systems	
Pipe	95–100
Lined canals	60–90
Nonlined canals	55–85

Summary

This article summarizes the essential definitions and methodologies for estimating crop water and irrigation requirements. The concept of reference evapotranspiration is assumed relative to a crop canopy such as grass but with constant crop characteristics. The hypotheses on which this approach is based are discussed relative to crop surface and aerodynamic resistances to heat and vapor fluxes. The crop evapotranspiration is defined using crop coefficients applied to the reference evapotranspiration, which reflect the canopy differences between the crop and the reference crop. Both time-averaged and dual-crop coefficients are explained, the first when the coefficients relative to crop transpiration and evaporation from the soil are summed and averaged for the crop-stage periods, the latter when a daily calculation of transpiration and evaporation coefficients is adopted. Finally, essential information on the soil water balance to estimate crop irrigation requirements is provided.

List of Technical Nomenclature

β	Bowen ratio (dimensionless)
γ	Psychrometric constant (kPa°C^{-1})
Δ	Slope of the saturation vapor pressure–temperature function (kPa°C^{-1})
ΔS	Change in soil-water storage (mm)
θ	Soil-water content (mm mm^{-1})
θ_{FC}	Soil-water content at field capacity (mm mm^{-1})
θ_{WP}	Soil-water content at wilting point (mm mm^{-1})
λ	Latent heat of vaporization (kJ kg^{-1})
ρ	Air density (kg m^{-3})
c_p	Specific heat of dry air (kJ kg^{-1} °C^{-1})
D_a	Average total water available for evaporation (mm)
D_e	Evaporated water (mm)
DP	Deep percolation (mm)
d	Zero-plane displacement height (m)
E_s	Evaporation from bare soil (mm day^{-1})
E_{so}	Potential rate of soil evaporation (mm day^{-1})
ET	Evapotranspiration (mm s^{-1})
ET$_c$	Crop evapotranspiration (mm day^{-1})
ET$_o$	Reference evapotranspiration (mm day^{-1})
$e_s - e_a$	Vapor pressure deficit (VPD) (kPa)
f_c	Fraction of ground covered by vegetation (dimensionless)
f_{ew}	Fraction of exposed wetted soil (dimensionless)
f_w	Fraction of wetted soil fraction (dimensionless)
GIWR	Gross irrigation water requirement (mm)
GW	Groundwater contribution (mm)
h	Crop height (m)
I	Total irrigation depth (mm)
IWR	Irrigation water requirement (mm)
I_w	Infiltration depth from irrigation (mm)
K_c	Crop coefficient (dimensionless)
K_{cb}	Basal crop coefficient (dimensionless)
K_e	Soil-water evaporation coefficient (dimensionless)
K_r	Evaporation reduction coefficient (dimensionless)
K_s	Stress-reduction coefficient (dimensionless)
k	Von Karman constant (dimensionless)
LAI	Leaf area index (dimensionless)
LR	Leaching requirement (mm)
MAD	Management-allowed depletion (dimensionless)
n_w	Number of wetting events (dimensionless)
P	Precipitation (mm)
P_n	Net precipitation (mm)
p	Depletion fraction for no stress (dimensionless)
REW	Readily evaporable water (mm)
RH	Relative humidity (%)
RO	Runoff (mm)
$R_n - G$	Net balance of energy at the surface (kJ m^{-2} s^{-1})
r_a	Aerodynamic resistance (s m^{-1})
r_l	Bulk stomatal resistance (s m^{-1})
r_s	Bulk surface resistance (s m^{-1})
TEW	Total evaporable water (mm)
T	Mean daily temperature (°C)

u	Wind velocity ($\mathrm{m\,s^{-1}}$)
W	Soil water (mm)
z_e	Depth of soil contributing to evaporation (m)
z_h	Height of air temperature measurement (m)
z_m	Height of wind velocity measurement (m)
z_{oh}	Roughness length for heat (m)
z_{om}	Roughness length for momentum (m)
z_r	Rooting depth (m)

See also: **Evapotranspiration**; **Irrigation:** Methods; **Plant–Soil–Water Relations**

Further Reading

Allen RG, Pereira LS, Raes D, and Smith M (1998) *Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper 56. Rome, Italy: Food and Agriculture Organization of the United Nations.

Pereira LS, van den Broek B, Kabat P, and Allen RG (eds) (1995) *Crop-Water Simulation Models in Practice*. Wageningen, Germany: Wageningen Press.

Tiercelin JR (ed.) (1998) *Traité d'Irrigation*. Paris, France: Lavoisier, Technique & Documentation.

van Lier HN, Pereira LS, and Steiner FR (eds) (1999) *CIGR Handbook of Agricultural Engineering, vol. I, Land and Water Engineering*. St. Joseph, MI: ASAE.

Wooton TP, Cecilio CB, Fowler LC, Hui SL, and Heggen RJ (eds) (1996) *Hydrology Handbook*. ASCE Manuals and Reports on Engineering Practice 28. New York: ASCE.

CROP-RESIDUE MANAGEMENT

D C Reicosky and A R Wilts, USDA Agricultural Research Service, Morris, MN, USA

Published by Elsevier Ltd.

Introduction

Soils, along with water, air and sun, are the major resources that sustain our food supply and terrestrial ecosystems. Soil organic matter is one of the primary contributors to soil quality. Crop residues are precursors to soil organic matter (SOM). The stems, leaves, chaff and husks that remain in the fields after crops are harvested for grain, seed or fiber play a critical role in soil quality and environmental issues since they are primary inputs of elemental carbon (C) into soil systems (Figure 1). Crop residues have been referred to as 'wastes' but as a natural and valuable resource are also considered to be 'potential black gold.' The negative connotation of 'residues' may refer to the remains after a part is taken or something leftover or useless. On the contrary, crop residues offer a large, but finite potential mechanism for C sequestration and nutrient cycling.

Crop residue management (CRM) is a widely-used cropland conservation practice for wind and water erosion control. Crop residue provides significant quantities of nutrients for crop production. In addition to affecting soil physical, chemical and biological functions and properties, crop residues can also affect water movement, infiltration, runoff and quality. However, the decomposition of crop residues can have both positive and negative effects on crop production and the environment. The presence of crop residues on the surface generally results in wetter and cooler conditions, thus favoring disease and pests, and also provides pathogens with an additional source of energy to multiply. Agricultural managers aim to increase the positive environmental effects of CRM since each practice has drawbacks. Ideally, improved management practices should enhance crop yields and have minimal adverse effects on the environment.

Crop management recommendations for maximum residue production require basic scientific research information regarding site-specific soils, crops and climate. Soil conservation and CRM research covers many aspects including the factors affecting residue decomposition, effects on erosion control, nutrient cycling and plant availability, disease control problems, weed control problems, alternate uses of excess residue, selection of plant varieties for conservation tillage systems, machinery requirements and control of the soil–water–temperature regime.

Conservation, Carbon Cycle, Soil Organic Matter and Carbon Sinks

Conservation of soil resources requires proper management of crop residues. The primary limiting factor for microbial growth in most soils is the C energy source. An abundance of C and other nutrients are returned to the soil through decomposition of crop

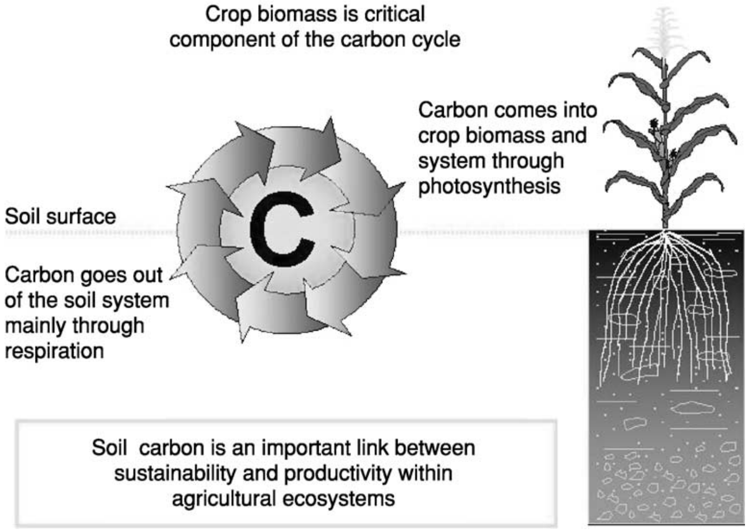

Crop biomass is critical component of the carbon cycle
Carbon comes into crop biomass and system through photosynthesis
Soil surface
C
Carbon goes out of the soil system mainly through respiration
Soil carbon is an important link between sustainability and productivity within agricultural ecosystems

Figure 1 Schematic representing the role of crop biomass in the agricultural ecosystems carbon cycle.

residues and biological nutrient cycling. Since organic matter (OM) is known to maintain soil aggregate stability, the addition of crop residues often improves soil structure and aggregation. Crop residues and tillage management can also affect the leaching of the nutrients, which may pollute the groundwater or surface waters. Crop residue influences soil temperature primarily by insulating the soil surface from the sun's radiant energy. Studies have shown that soils high in SOM retain more moisture, especially when residues are retained on the soil surface as compared to when they are incorporated.

The recent interest in global climate change has prompted many to value C sources and sinks in agroecosystems. Soil C pools and fluxes in agroecosystems are influenced by a number of factors including amount and type of plant residue, temperature and precipitation and soil texture, pH and drainage. Carbon sequestration or storage in terrestrial ecosystems can be defined as the net removal of carbon dioxide (CO_2) from the atmosphere by crop photosynthesis into stable, long-lived pools of C. The soil organic carbon (SOC) pool is estimated to compose about two-thirds of the terrestrial biosphere C pool. Soil organic C storage in cropland soils is determined by tillage systems and the amount and placement of the crop residue. As grain and biomass yields increase and less intensive tillage systems are employed, farmers should gradually develop an enhanced C sink.

Crop Residues and Nutrient Cycling

The annual cycling of plant nutrients in the plant–soil ecosystem is essential to maintaining a productive agricultural system. The management of crop residues has important implications for the total amounts of nutrients removed from and returned to the soil. Soil organic matter also improves the dynamics and bio-availability of main plant nutrient elements. The soil, water, and air, which are in contact with plants, contain various inorganic chemicals necessary for plant growth. Soil organic matter is the main determinant of biological activity because it is the primary energy source. The amount, diversity and activity of soil fauna and microorganisms are directly related to SOM content and quality.

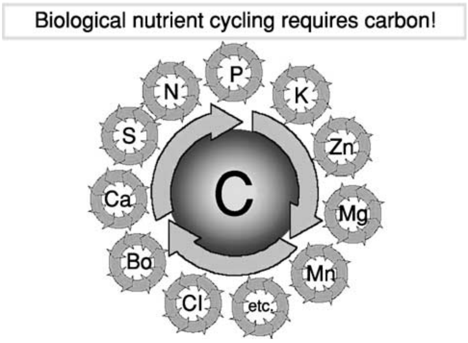


Figure 2 Soil carbon plays a critical role in biological nutrient cycling.

Organic matter and the biological activity that it generates have a major influence on the physical and chemical properties of the soils. Each plant nutrient has its own C-dependent cycle (Figure 2) that controls its availability to the next generation of plants. Carbon compounds in the residue are the 'fuel' or energy sources for the soil microbes and fauna responsible for biological recycling of these inorganic plant nutrients. During microbial decomposition of crop residues, chemical elements are released into the immediate environment that may be utilized by living plants or other organisms. This constitutes the basic framework of biological nutrient cycling in agricultural production systems. Carbon-enriched crop biomass becomes the primary food source for soil microorganisms and fauna and as a result ‘nurtures’ nutrient cycling.

Plant availability of nutrients (nitrogen (N), phosphorus (P), and potassium (K)) in crop residues is regulated largely by soil water, soil temperature, other soil physical and chemical properties, and soil and crop management practices. For N, activity of soil microorganisms is usually most important in determining the cycling and potential availability from crop residues; for P, both microbial activity and soil mineralogy are involved; and for K, mineralogy and soil water movement are important parameters. Management practices such as fertilization and the amounts of residue remaining after harvest determine the extent of cycling and plant availability of nutrients from crop residues. The shift from conventional to conservation tillage systems necessitates new research to determine the rate of cycling and plant nutrient availability.

Crop residues can be managed for increased OM levels, thereby sequestering C. The position and quantity of crop residue as well as N fertilization have variable influences on SOM. Several studies have indicated that when more crop residues were on or near the surface, as in no-till or reduced tillage systems, SOM content near the surface was increased, but when incorporated to depth by moldboard plow tillage, the quantity of crop residue had little or no influence on SOM. The physical incorporation and mixing of residue maximize soil-residue contact and result in rapid decomposition and C loss as CO_2.

The C:N ratio of the residue, an important key in soil management, also varies. Crop biomass is generally 40–50% C, but the nitrogen (N) content varies considerably within and among species. Thus, an adequate supply of N may be required to build SOM for crops with a high C:N ratio since C and N and their proportionality (i.e., C:N ratio) is relatively constant across a range of agricultural soils at about 10:1. For example, wheat straw has a C:N ratio around 90:1 and will require N addition or C loss to decompose to the soil equilibrium value of 10:1. Thus, soil C sequestration may be reduced when C and N inputs of the residue are out of balance.

Factors Controlling Residue Decomposition and Soil Quality

Soil microorganisms play a major role in the synthesis and degradation of crop residues into SOM. The decomposition rates of OM and crop residue depend primarily on chemical composition and on factors that affect the soil environment. Factors having the greatest effect on microbial growth and activity will have the greatest potential for altering the rate of residue decomposition in soil.

Soil factors that typically influence residue decomposition most include water, temperature, pH, aeration or oxygen supply and available nutrients. Residue factors include chemical composition of the residue, C:N ratio, age of plant material, particle size and the indigenous microflora. Additional factors that must be considered are the residue water content and the method of residue application to the soil (i.e., mixed with the soil, layered or banded in soil or left on the soil surface). Many of these factors are not independent and a change in one may affect a change in others.

Plant residue decomposition rates generally increase when residues are accessible to soil microbes. Many studies indicate that burying residues in soil increases the decomposition rate compared to placing residues on the soil surface. The effect of placement decreases with time. Thus, tillage maximizes residue–soil contact and enhances decomposition. Soil enzyme activities also respond to tillage practices. Crop residues and tillage have been reported to significantly and rapidly alter the composition, distribution and activity of the soil enzymes. Although straw amendments also contain some enzymes, the increase in activity in soils with organic residues most likely results from the stimulation of microbial and fungal activity rather than the direct addition of enzymes found in organic residues. Soil macroorganisms, such as earthworms, also play a role in crop residue decomposition. The nutritional quality and C:N ratios of plant material appear to be important factors that influence earthworm population.

Soil organic matter and humic substances exert physical, chemical and biological effects on soil quality by serving as soil conditioners, nutrient sources and substrates for microorganisms. Humic substances contribute to soil structure by acting as binding agents in the formation of soil aggregates. As a result, a stable soil structure ensures satisfactory drainage and aeration, provides protection against erosion and enhances soil properties. Humic substances are sources of nutrients that are essential for plant growth. These substances, which can be derived from crop biomass, impact soil quality and fertility and contribute to the vital role of SOM.

Crop Residues: Social and Environmental Benefits

Many of the social and environmental benefits of carbon cycled from crop residues are subtle and often difficult to detect. Integrating the numerous

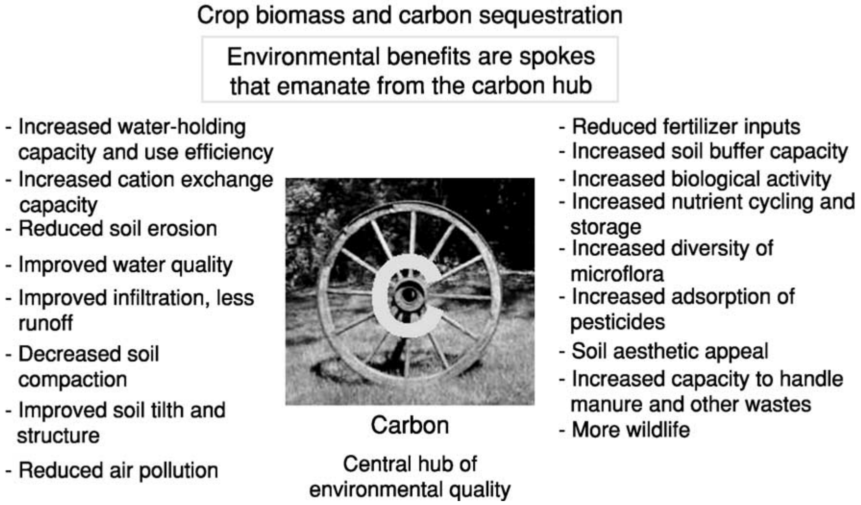

Crop biomass and carbon sequestration
Environmental benefits are spokes that emanate from the carbon hub
- Increased water-holding capacity and use efficiency
- Increased cation exchange capacity
- Reduced soil erosion
- Improved water quality
- Improved infiltration, less runoff
- Decreased soil compaction
- Improved soil tilth and structure
- Reduced air pollution
Carbon
Central hub of environmental quality
- Reduced fertilizer inputs
- Increased soil buffer capacity
- Increased biological activity
- Increased nutrient cycling and storage
- Increased diversity of microflora
- Increased adsorption of pesticides
- Soil aesthetic appeal
- Increased capacity to handle manure and other wastes
- More wildlife

Figure 3 An environmental quality wheel illustrates results from numerous environmental benefits emanating from the carbon hub.

'small' benefits from C generated from crop residues yields important environmental benefits. Soil C from crop residues may be considered analogous to the central hub of a wheel (Figure 3). The spokes of the wheel represent incremental links to or benefits from soil C that lead to environmental improvement. Each of the secondary benefits that emanate from soil C contributes to environmental enhancement through improved soil C management. Crop biomass provides the C that becomes the supporting spokes of the environmental sustainability wheel.

Social benefits of agricultural residue management may include many off-site consequences of adopting new farm technologies and improving cultural practices. Reductions in runoff and soil erosion from cropland and rangeland enhance the functioning of streams, rivers and lakes through reduced flooding and sedimentation. The useful life of public facilities, a benefit to society, may be extended through improved water quality. Other important off-site benefits related to wind erosion have a direct impact on air quality. Air quality is primarily stressed in areas of industrial pollution and concentrated populations, but is often overlooked in rural areas as soil erosion by wind-driven particles. Additional benefits of CRM include providing both protective cover and a source of food for wild game.

Crop Residues, Research and Global Change

Agricultural crop residues and their proper management can also play an important role in helping society cope with increased greenhouse gas emissions from the burning of fossil fuels. The agricultural sector has a large capacity for removing CO_2 from the atmosphere and abating the emission of its own greenhouse gases. Croplands have the potential to offset a very significant portion of greenhouse gas emissions, but questions about climate change impacts on crop residue decomposition research need to be addressed. Specifically: (1) To what extent does climate change affect diversity of plant species and soil biota and residue decomposition processes? (2) What methods of decomposition management should be utilized to control C sequestration and CO_2 emissions? (3) What are the tillage methods and residue interactions important in carbon cycling for nutrient-use efficiency? Within a given ecosystem, the soils have a finite capacity to store carbon limited by natural soil formation factors. As a result, agriculture's contribution to these larger global climate change issues will likely be for the short term (25–50 years). Nevertheless, agriculture can help society buy time to develop new technologies and cleaner burning fuels.

Summary

Crop residue management through conservation agriculture can improve soil productivity and crop production by maintaining SOM levels. Two significant advantages of surface-residue management are increased OM near the soil surface and enhanced nutrient cycling and retention. Greater microbial biomass and activity near the soil surface acts as a reservoir for nutrients needed in crop production and increases structural stability for increased infiltration. In addition to the altered nutrient distribution within the soil profile, changes also occur in the chemical and physical properties of the soil. Improved soil C sequestration through enhanced CRM is a cost-effective option for minimizing agriculture's impact on the environment.

Ideally, CRM practices should be selected to optimize crop yields with minimal adverse effects on the environment. Results from many experiments have indicated differing effects of CRM practices on harvested yield. Conflicting results occur due to the large number of complex interactions associated with residue quality, soil-related factors, health of the previous crop, potential susceptibility of the next crop and management options such as cultivar selection, crop rotation and planting date. Results suggest that no one CRM system is superior under all conditions. Thus, farmers have a responsibility in making management decisions that will enable them to optimize crop yields and minimize environmental impacts. Multidisciplinary and integrated efforts by a wide variety of scientists are required to design the best site-specific systems for CRM practices to enhance agricultural productivity and sustainability while minimizing environmental impacts.

Crop residues of common agricultural crops are important resources, not only as sources of nutrients for succeeding crops and hence agricultural productivity, but also for improved soil, water and air quality. The development of effective CRM systems depends on a thorough understanding of factors that control residue decomposition and their careful application within a specific crop production system. Maintaining and managing crop residues in agriculture can be economically beneficial to many producers and more importantly to society. Improved residue management and reduced tillage practices should be encouraged because of their beneficial role in reducing soil degradation and increasing soil productivity. Soil C sequestration contributes to these benefits and can play a significant role in mitigating global climate change. Food security and environmental improvement depend on soil C, a valuable resource, that can be sustainable in agroecosystems through improved, cost-effective CRM.

See also: **Carbon Cycle in Soils:** Dynamics and Management; **Carbon Emissions and Sequestration**; **Cover Crops**; **Cultivation and Tillage**; **Mulches**; **Nitrogen in Soils:** Cycle; **Nutrient Availability**; **Organic Matter:** Principles and Processes; Genesis and Formation; **Organic Residues, Decomposition**

Further reading

Allmaras RR, Schomberg HH, Douglas CL Jr, and Dao TH (2000) Soil organic carbon sequestration potential of adopting conservation tillage in U.S. croplands. *Journal of Soil and Water Conservation* 57(3): 365–373.

Cadisch G and Giller KE (1997) *Driven by Nature: Plant Litter Quality and Decomposition*. Cambridge: University Press, Wallingford: CAB International.

Crovetto Lamarca C (1996) *Stubble over the Soil: The Vital Role of Plant Residue in Soil Management to Improve Soil Quality*. Madison, WI: ASA.

CTIC (1995) *Survey Guide; National Crop Residue Management Guide*. West Lafayette, IN: Conservation Technology Information Center.

Guérif J, Richard G, Dürr B, Machet JM, Recous S, and Roger-Estrade J (2001) A review of tillage effects on crop residue management, seedbed conditions and seeding establishment. *Soil and Tillage Research* 61: 13–32.

Kumar K and Goh KM (2000) Crop residues and management practices: effects on soil quality, soil nitrogen dynamics, crop yield, and nitrogen recovery. *Advances in Agronomy* 68: 197–319.

Lal R (1997) Residue management, conservation tillage and soil restoration for mitigating greenhouse effect by CO_2-enrichment. *Soil and Tillage Research* 43: 81–107.

Lal R, Kimble JM, Follett RF, and Stewart BA (1997) *Management of Carbon Sequestration in Soil*. Boca Raton, FL: CRC Press, Lewis Publishers.

Lal R, Kimble JM, Follet RF, and Cole V (1998) *Potential of U.S. Cropland for Carbon Sequestration and Greenhouse Effect Mitigation*. USDA–NRCS Washington, DC. Chelsea, MI: Ann Arbor Press.

Oschwald WR (ed.) *Crop Residue Management Systems*. ASA Special Publication 31. Madison, WI: ASA.

Paul E, Paustian K, Elliott ET, and Cole CV (1997) *Soil Organic Matter in Temperate Agro-ecosystems: Long-term Experiments in North America*. Boca Raton, FL: CRC Press.

Reeder R (2000) Conservation tillage systems and management. In: *Crop Residue Management with No-till, Ridge-till, Mulch-till and Strip-till*. MWPS-45, 2nd edn. Ames, IA: MidWest Plan Service, Iowa State University.

Reicosky DC (1994) Crop residue management: soil, crop, climate interaction. In: Hatfield JL and Stewart BA (eds) *Advances in Soil Science, Crop Residue Management*, pp. 191–214. Special Series. Chelsea, MI: Lewis Publishers.

Unger P (1994) *Managing Agricultural Residues*. Boca Raton, FL: CRC Press.

Unger PW, Langdale GW, and Papendick RI (1988) Role of crop residues – improving water conservation and use. In: Hargrove WL (ed.) *Cropping Strategies for Efficient Use of Water and Nitrogen*, pp. 69–100. ASA Special Publication 51. Madison, WI: ASA.

CRUSTS

Contents

Biological
Structural

Biological

J Belnap, USGS Southwest Biological Science Center, Moab, UT, USA

Published by Elsevier Ltd.

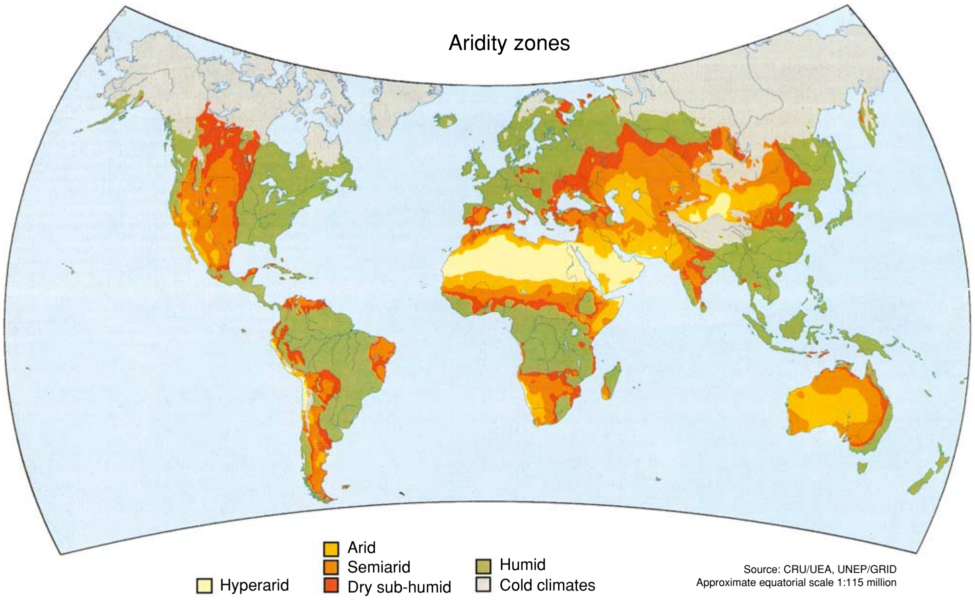
Introduction

Biological soil crusts occur in a wide variety of climate regimes and vegetative communities around the globe (Figure 1). Wherever vascular plant cover is sparse enough for light to reach the soil surface, it will be colonized by some form of biological soil crust organisms. This is true even if the exposure of the soil surface is temporary, such as occurs when trees fall, or when volcanic activity or fire removes the vegetative cover. Soil crusts reach their greatest expression in arid and semiarid environments where sparse vegetation leaves large expanses of soils available for colonization (Figure 2). In the USA, this is mostly in western states (Figure 3). They are also found in other surprising places such as pine barrens, where infertile soils restrict vascular plant growth; subhumid grasslands where limited rainfall leaves open spaces between the plants; and tundra and alpine areas, where harsh conditions preclude high vascular plant cover.

Biological soil crusts (also referred to as cryptogamic, cryptobiotic, microbiotic, or microphytic soil crusts) consist of an interwoven community of cyanobacteria, green algae, microfungi, bacteria, lichens, and mosses (Figures 4 and 5). Cyanobacterial and microfungal filaments weave throughout the top few millimeters of soil, gluing loose soil particles together and forming a coherent crust that stabilizes and protects soil surfaces from erosive forces. These crusts have only recently been recognized as having a major influence on the functioning of terrestrial ecosystems.

Species Composition and Growth Forms

Globally, biological soil crusts have many similarities in species composition, despite occurring in



Source: CRU/UEA, UNEP/GRID
Approximate equatorial scale 1:115 million

Figure 1 Aridity zones of the world. Shaded zones are regions where biological crusts have the potential to play pivotal roles in ecosystem functioning. Courtesy of United Nations Environment Program. Middleton N and Thomas D (eds) (1997) *World Atlas of Desertification*, 2nd edn. London: Arnold.

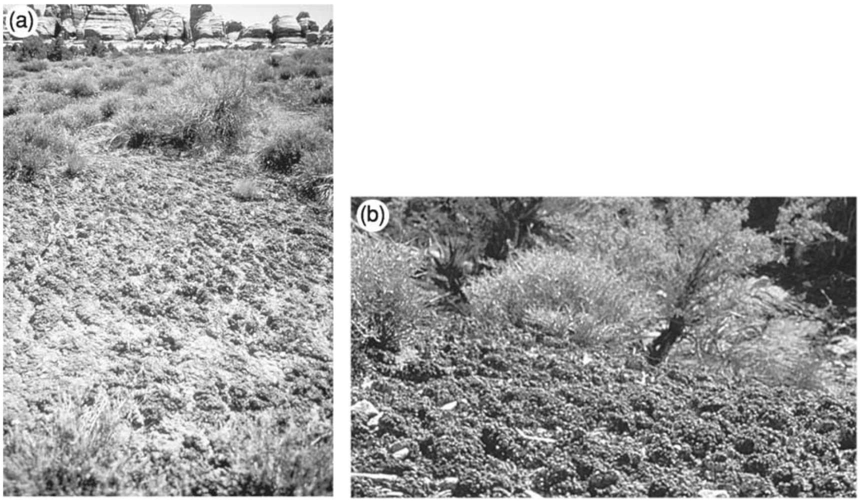


Figure 2 Biological soil crusts, completely covering the interspaces between plants, on the Colorado Plateau (northern Arizona, southern Utah, western Colorado).

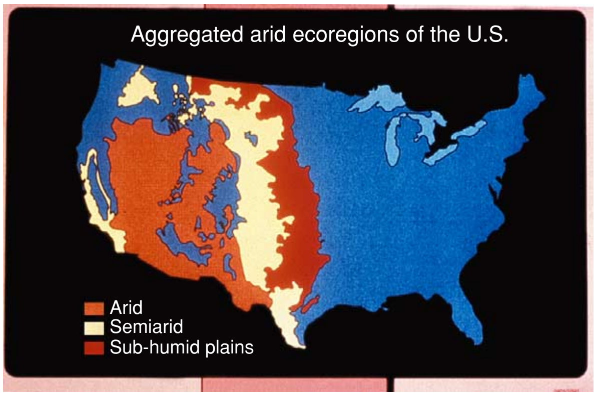


Figure 3 Arid, semiarid, and subhumid plains of the USA. When relatively undisturbed, soils across much of these regions are covered with biological soil crusts.

unconnected and seemingly dissimilar environments. Many of the dominant cyanobacteria, lichens, and moss species and genera found in soil crusts have a cosmopolitan distribution (Table 1). The proportional amount of the different species, however, varies with climate (Figure 6). In addition, climate influences what species are present. For instance, large filamentous species such as *Microcoleus* dominate the cyanobacterial flora in deserts where most rain falls during cool seasons. In contrast, the cyanobacterial flora in hot deserts with predominantly summer rainfall is often dominated by smaller genera such as *Scytonema*, *Nostoc,* and *Schizothrix*. Whereas hot desert soil crusts are dominated by cyanobacteria, cool desert soil crusts are dominated by lichens and mosses. Common lichens found in deserts throughout the world include *Fulgensia*, *Diploschistes*, *Psora*, *Placidium*, and *Collema*. Common mosses include *Tortula*, *Bryum*, and *Grimmia*.

Biological soil crusts have four general growth forms (Figure 7). Hot deserts that lack frost-heaving are generally characterized by smooth cyanobacterial crusts or rugose lichen-moss crusts. In cool deserts where frost-heaving is present, soil crusts with moderate (about 40%) lichen-moss cover are often pinnacled, due to frost-heaving upwards and differential erosion downwards. Crusts in cool deserts with a heavy lichen-moss cover are generally rolling, as frost-heaving and erosion are mitigated by the extensive lichen-moss cover.

Because the dominant components of biological soils crusts are photosynthetic organisms, they require sunlight. When soils are dry, the bulk of the crustal biomass is up to 0.5 mm below the soil surface, with some individuals found down to 4 mm. While mosses and lichens have ultraviolet (UV)-protective pigments or heavy coloration to protect them from UV radiation, only some cyanobacteria have such protection. The smaller, less motile cyanobacteria such as *Scytonema* and *Nostoc* manufacture UV-screening pigments, and are generally found on the soil surface. In contrast, the large, more motile filamentous species such as *Microcoleus*, *Lyngbya*, *Phoridium*, and *Oscillatoria* do not have UV-protective pigments, and are seldom found on the soil surface except on cloudy days when soils are moistened. Instead, these larger species are found tucked underneath the pigmented species, using them as a sunscreen. Despite sunscreening efforts by all species, mortality can be high in the summer, when UV radiation is high and moisture is limited.

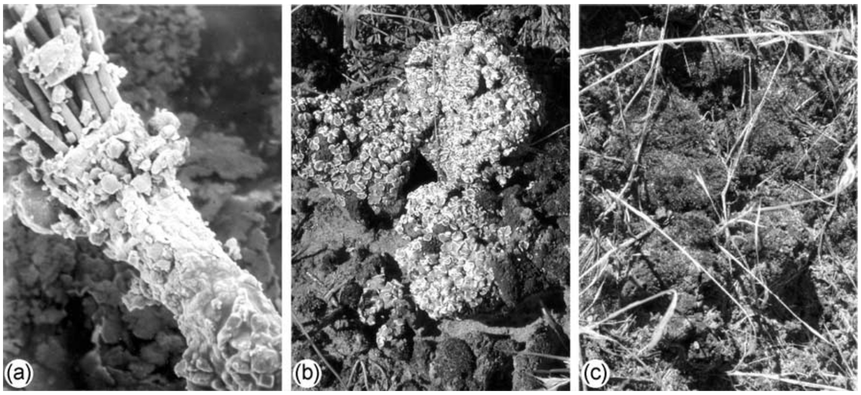


Figure 4 Close-up photos of (a) the cyanobacterium *Microcoleus vaginatus*, (b) the lichens *Psora decipiens* and *Collema coccophorum*, and (c) the moss *Syntrichia caninervis*.

Figure 5 Close-up of a biological crust community.

Table 1 Genera and species of lichens and cyanobacteria that have a cosmopolitan distribution

Lichen	*Cyanobacteria*
Fulgensia fulgens	*Microcoleus vaginatus*
Psora decipiens	*M. paludosus*
Squamarina spp.	*M. chtonoplastes*
Toninia sedifolia	*M. sociatus*
Catapyrenium spp.	*Nostoc commune*
Diploschistes spp.	*Calothrix* spp.
Endocarpon spp.	*Lyngbya* spp.
Collema spp.	*Oscillatoria* spp.
	Phormidium spp.
	Scytonema spp.
	Tolypothrix spp.

Ecological Roles

Carbon Fixation

Biological soil crusts are an important source of fixed carbon in the sparsely vegetated areas commonly found throughout the world. While vascular plants provide organic matter to the soils directly underneath them, large interspaces between plants have little opportunity to receive such input. Where biological soil crusts are present, they contribute carbon and help keep plant interspaces fertile. They do this both by leaking carbon-rich compounds into surrounding soils while alive, and contributing body carbon upon their death. These inputs help support carbon-limited microbial populations that are essential in the decomposition of plant materials.

Nitrogen Fixation

Nitrogen levels are low in desert ecosystems relative to other ecosystems, and many deserts have few nitrogen-fixing plants. Since nitrogen can limit plant productivity, the maintenance of normal nitrogen cycles is critical to maintaining the fertility of semiarid soils. Most soil crusts in deserts are dominated by complexes of organisms capable of fixing nitrogen, including *Scytonema*, *Nostoc* and the common soil lichens *Collema* and *Peltula*. In desert areas, rainfall events that are too small to promote plant growth can often stimulate crustal activity; thus, the overall time of soil crust activity can actually be fairly high. Soil crusts can be the dominant source of nitrogen for desert shrub and grassland communities. Input estimates range from 1 to 10 kg ha^{-1} annually. Nitrogen inputs are highly dependent on temperature, moisture, and crustal species composition; thus, climatic regimes and the timing, extent, and type of past disturbance which affects species composition are critical in determining fixation rates.

As with carbon, crusts contribute nitrogen to soils both underneath plants and in plant interspaces, counteracting the tendency of nutrients to concentrate around plants. Five to 88% of nitrogen fixed by crusts has been shown to leak into the surrounding soils. Nitrogen leaked from these organisms is

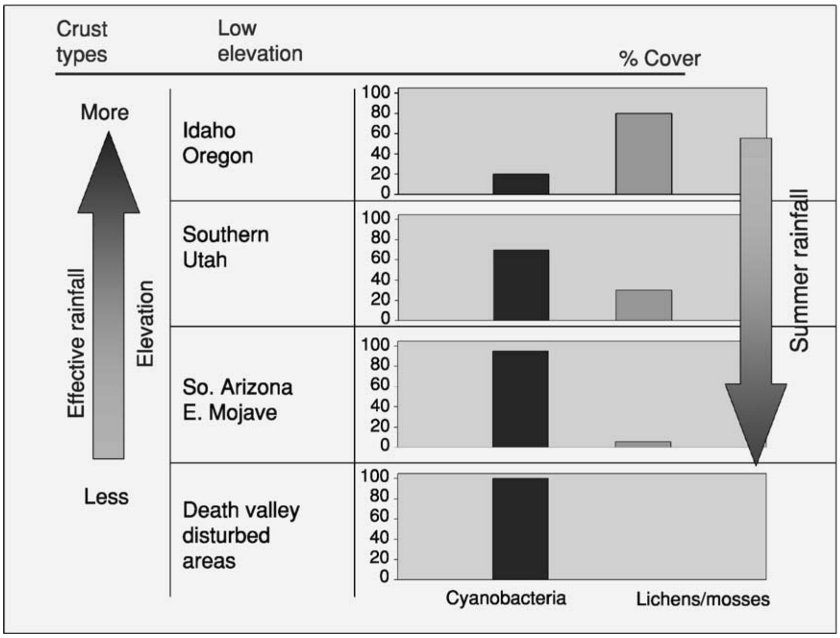


Figure 6 The cover of lichens and mosses, relative to that of cyanobacteria, increases with higher precipitation. Consequently, cyanobacteria dominate biological soil crusts in hotter deserts, whereas mosses and lichens dominate biological soil crusts in cooler deserts.

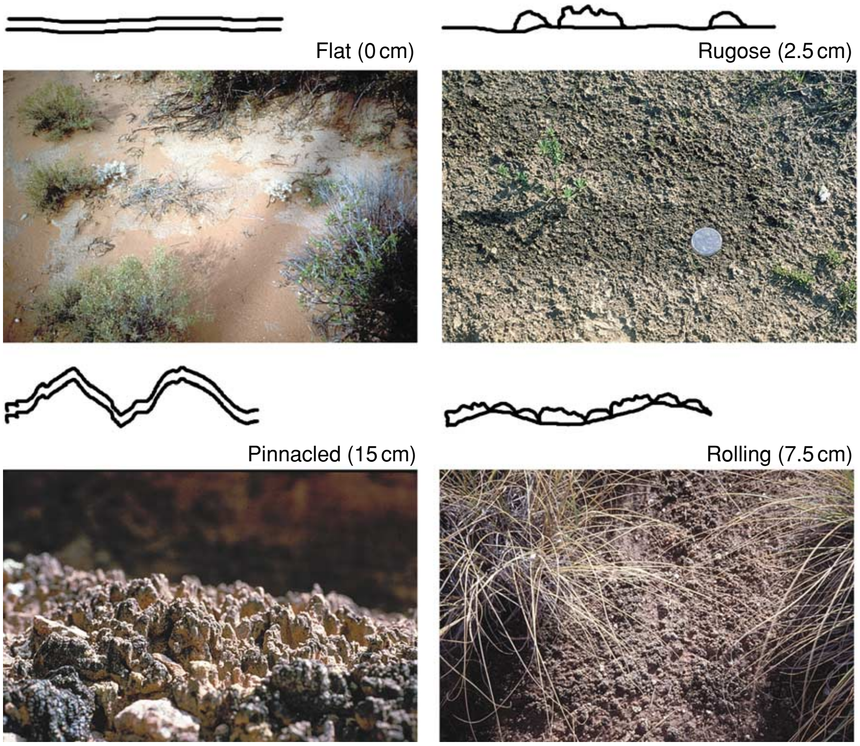


Figure 7 There are four distinct forms of soil crusts. Flat and rugose crusts are found in regions where soils do not freeze. Pinnacled and rolling crusts are found where soils do freeze. Flat crusts are also formed when any of the crust types are severely disturbed.

available to nearby vascular plants and microbial communities. Vascular plants growing in crusted areas show higher leaf concentrations of nitrogen when compared to plants growing in uncrusted soils. Leaked nitrogen has also been found in associated fungi, actinomycetes, and bacteria.

Dust Trapping

Dust can be an essential component of desert soil fertility, and soil crusts are effective in capturing eolian (wind-flown) dust deposits. Recent work in ecosystems throughout the world shows that dust inputs significantly increase levels of all major and minor soil nutrients in the tested soils. Nutrients are especially enriched in biologically crusted soils.

Effects on Vascular Plants

Soil crusts can influence the location of safe sites for seeds, as well as the germination and establishment of vascular plants. In hot deserts with flat cyanobacterial crusts, seeds generally skid off the smoothed surfaces and lodge under the nearest obstacle. The slight soil surface roughening of rugose crusts provides for some on-site seed retention. In contrast, the highly roughened soil surface created by soil crusts in cool and cold deserts results in very high retention of seeds and organic matter.

Crusts can influence seed germination, as they affect soil moisture, temperature, and stability. The increased soil moisture and temperature found in well-developed crusts generally favor germination. In deserts, low air humidity means most seeds need some form of cover in order to stay hydrated long enough for germination. Small seeds utilize small cracks for cover, while most large-seeded plants need a cover of soil or litter to germinate. Native seeds have self-burial mechanisms (such as hygroscopic awns) or are cached by rodents. However, there are nonnative seeds that lack such adaptations, and germination of these species can be inhibited by crusts.

Once seeds germinate, no studies have found that crusts affect root penetration or plant establishment. Survival of vascular plants is generally much higher, or unaffected, when crusted areas are compared with uncrusted ones. No studies have shown that biological crusts decrease vascular plant survival.

Many studies have attempted to correlate crust and vascular plant cover and results show negative, positive, and no relationship, depending on other site characteristics. At more arid sites, the correlation is generally positive, suggesting plants aid in the survival of crustal components, perhaps by providing shade. At higher elevations and/or areas with more plant cover, plants appear to inhibit crust cover by restricting the amount of light reaching the soil surface. No study has demonstrated a negative influence of crusts on overall plant cover.

Nutrient Levels in Vascular Plants

Plants growing on crusted soil generally show higher concentrations and/or greater total accumulation of various bioessential nutrients (including nitrogen, potassium, sodium, calcium, iron, and magnesium) when compared to plants growing in adjacent, uncrusted soils. Dry weight of plants in pots with cyanobacteria is up to four times greater than in pots without cyanobacteria, and dry weight of plants in untrampled areas can be twice that of trampled areas.

Several mechanisms have been postulated to explain the effect of soil crusts on vascular plants. Soil crusts contribute nitrogen and carbon, which directly increases soil fertility and also probably decomposition rates. Dust capture increases soil fertility as well. Cyanobacterial sheath material is negatively charged, binding positively charged macronutrients and thus preventing their leaching. Cyanobacteria secrete chelators that keep iron, copper, molybdenum, zinc, cobalt, and manganese more available in high-pH soils. In addition, dark-colored crusts increase soil temperature and therefore nutrient uptake rates.

Soil Hydrology and Stabilization

The effect of biological soil crusts on soil-water relations is heavily influenced by soil texture, soil structure, and the growth form of the crusts. In hot deserts, the presence of the mucilaginous cyanobacteria smooths the soil surface, decreasing soil permeability, and thus water infiltration. In contrast, the increased surface roughness that accompanies soil crust development in cool deserts increases water pooling and residence time, thus increasing the amount and depth of rainfall infiltration.

Soils in arid regions are often highly erodible, and soil formation is extremely slow, taking 5000 to 10 000 years or more. Consequently, reducing soil loss is very important in these regions. Crusts have been shown to reduce soil loss caused by wind and water erosion in all types of deserts. Polysaccharides extruded by the cyanobacteria and green algae, in combination with lichen and moss rhizines, entrap and bind soil particles together (Figure 8). These larger soil aggregates are heavier, have a greater surface area, and are more difficult to move by wind or water, thus reducing soil loss (Figure 9). When wetted, cyanobacterial sheath material swells and covers the soil surface even more extensively

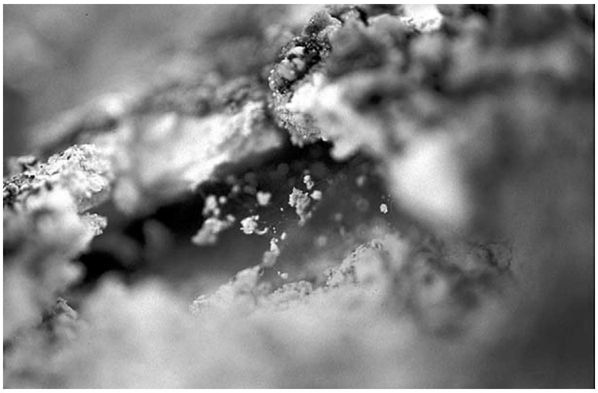


Figure 8 Clumps of soil held in place by cyanobacterial filaments. Reproduced from Belnap J and Gardner JS (1993) Soil microstructure in soils of the Colorado Plateau: the role of the cyanobacterium *Microcoleus vaginatus. Great Basin Naturalist* 53: 40–47.

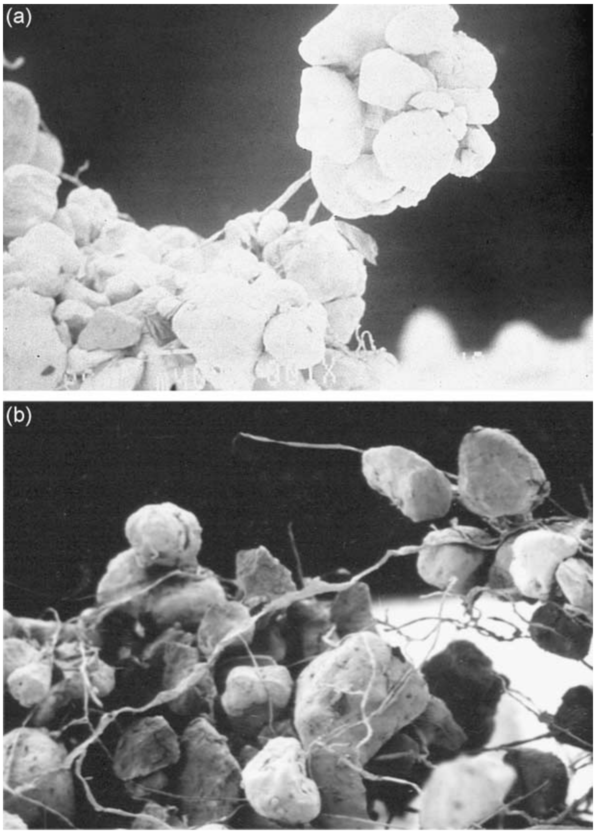


Figure 9 Scanning electron micrographs of biological soil crusts. Filaments of *Microcoleus vaginatus* can be seen connecting sand grains together and forming large soil aggregates. Top panel ×90; bottom panel ×100. Reproduced with permission from Belnap J and Gardner JS (1993) Soil microstructure in soils of the Colorado Plateau: the role of the cyanobacterium *Microcoleus vaginatus. Great Basin Naturalist* 53: 40–47.

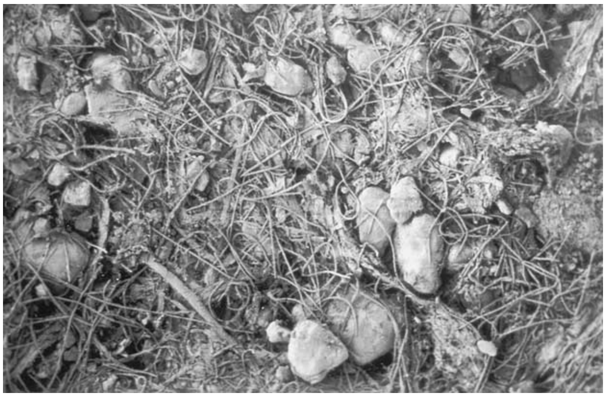


Figure 10 When wetted, *Microcoleus vaginatus* swells, casting a net over soil surface particles (×70).

than when dry, protecting soils from both raindrop erosion and overland water flow during rainstorms (Figure 10). Resistance to wind erosion increases with biological crust development, with well-developed soil crusts able to withstand all recorded ground wind speeds.

Effects of Disturbance

Species Composition

Trampling of crusted surfaces generally reduces the cover, biomass, and species richness of soil crusts. Whereas a well-developed crust can have up to 70 species of cyanobacteria and green algae and 10 or more species of soil lichens and mosses, severely trampled areas generally contain only a few species of cyanobacteria.

Water Erosion

As crustal components are brittle and easily crushed when dry, the soil aggregates formed by crust organisms are disrupted when trampled. When the roughened microtopography of undisturbed cool desert crusts is flattened, the velocity of surface water flows is increased and surfaces are subjected to sheet erosion. Seed and dust retention is reduced. Surface disturbance also crushes the buried cyanobacterial sheath material which binds soil particles, adsorbs nutrients, and increases soil moisture retention, despite lacking living filaments. Damage to this abandoned material is irreparable, since living cyanobacteria are no longer present at these depths to regenerate sheath materials. Consequently, trampling can greatly accelerate desertification processes through increased soil loss and water runoff.

Wind Erosion

Wind is a major erosive force in deserts, where there is little soil surface protection by organic matter or vegetative cover, and where soil formation is slow.

Experiments have demonstrated that, while well-developed, undisturbed crusts protect soil surfaces from wind erosion, any compressional disturbances to these crusts leave surfaces vulnerable. A decrease in soil wind resistance is directly associated with increased sediment movement. In addition, nearby biological soil crusts can be buried by blowing sediment, resulting in the death of the photosynthetic organisms. Because most photosynthetic productivity occurs in organisms in the top 1 mm of soil, very small losses can dramatically reduce site fertility and further reduce soil surface stability.

Nutrient Cycles

Nitrogen fixation in crusts shows long-term reductions in response to all types of experimentally applied disturbance, including human feet, mountainbikes, four-wheel-drive trucks, tracked vehicles (tanks), and shallow and deep raking. Consequently, disturbance can result in large decreases in soil nitrogen through a combination of reduced biological nitrogen input and elevated gaseous loss of nitrogen and soil loss. Short-term reductions in nitrogen fixation range up to 100%. Long-term studies show a 42% decrease in soil nitrogen 25 years following disturbance.

Albedo

Trampled surfaces, when compared to untrampled crusted surfaces, show up to a 50% increase in reflectance of wavelengths ranging from 0.25 to 2.5 μm. This represents a change in the surface energy flux of approximately $40\,\mathrm{W\,m^{-2}}$. Large acreages of trampled areas, combined with lack of urban areas to offset this energy loss, may lead to changes in regional climate patterns in many semiarid regions. Increased albedo can radically decrease soil surface temperatures.

Surface temperatures can regulate many ecosystem functions. Lower temperature decreases carbon and nitrogen fixation rates, microbial activity, plant nutrient uptake rates, and plant growth rates. It can also delay seed germination. Because timing of these events is often critical in deserts, relatively small delays can reduce species fitness and seedling establishment, which may eventually affect community structure. Foraging times are often partitioned among ants, arthropods and small mammals based on surface temperature. Many small desert animals are weak burrowers, and soil surface microclimates are of great importance to their survival. Consequently, altering surface temperatures can affect nutrient availability and community structure for many desert organisms, thus increasing susceptibility to desertification.

Recovery from Disturbance

Natural Recovery Rates

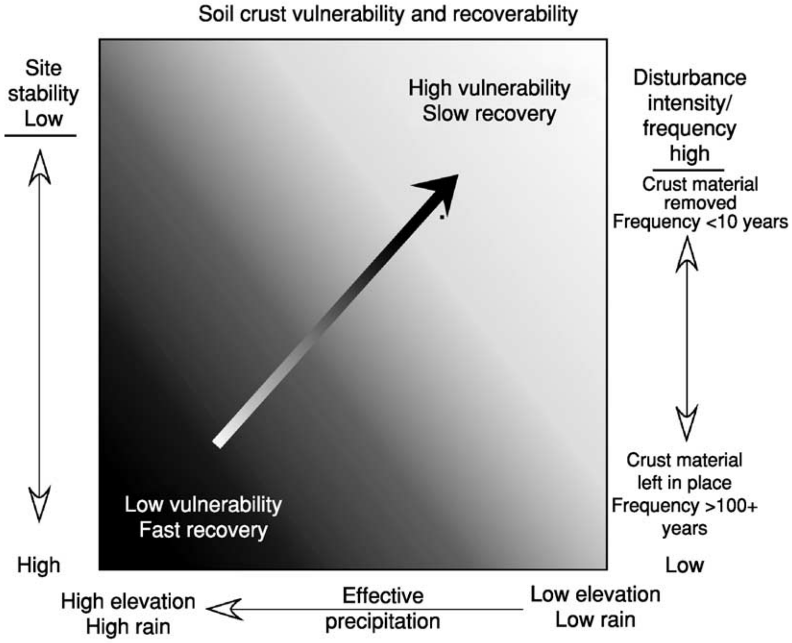
Recovery rates are related to the type of soil crust present and the evolutionary history of the site; the type, timing, and intensity of disturbances; climate; and soil characteristics (Figure 11). In addition, 'recovery' must be defined. Visual recovery, which is not species-dependent, will occur more quickly than nitrogen flxation, which requires specific species. Nitrogen fixation in turn, recovers more rapidly than soil stability, as it depends on the recovery of all species. Faster recovery is found where: (1) the disturbance type has been present in evolutionary time and the flora is pre-adapted to disturbance; (2) disturbance is light enough to crush crust material in place, as opposed to removing it, and is infrequent; (3) surface-to-volume ratios are low, so colonizing organisms can reach the site more quickly; (4) sites are relatively stable (fine-textured soils, low or no slope as opposed to coarse soils and steep slopes); and (5) effective rainfall is high. Recovery times are also very species-dependent, as cyanobacteria are both more resistant to disturbance, and faster to recover, than lichens or mosses (Figure 12). Estimates of recovery times are shown in Table 2.

Assisted Recovery

The use of inoculants can substantially hasten recovery of soil crusts. However, development of inoculant in the laboratory has not been widely successful.

Conclusion

Unfortunately, the increasing activities of humans in desert areas are often incompatible with the well-being of biological soil crusts. The cyanobacterial fibers that confer such tensile strength to these crusts are no match for the compressional stresses placed on them by vehicles or trampling. Crushed crusts contribute less nitrogen and organic matter to the ecosystem. Impacted soils are left highly susceptible to both wind and water erosion. Raindrop erosion is increased, and overland water flows carry detached material away. Relatively undisturbed biological soil crusts can contribute a great deal of stability to otherwise highly erodible soils. Unlike vascular plant cover, crustal cover is not reduced in drought, and unlike physical crusts, these organic crusts are present year-round. Consequently, biological crusts offer stability over time and under adverse conditions, features often lacking in other soil surface protectors. Unfortunately, ever-increasing recreational and commercial uses of semiarid and arid areas are resulting in a rapid loss of well-developed biological crusts around the world.



Factors determining site stability

	Low stability	High stability
Soil texture	Coarse	Fine
Rock/gravel cover	Unstable	Stable
Soil depth	Deep	Shallow
Plant spacing	Sparse	Dense

Figure 11 Summary chart of factors that influence susceptibility and recoverability of biological soil crusts to disturbance. Reprinted from Belnap J and Lange O (eds) *Biological Soil Crusts: Structure, Function, and Management*, p. 375. Berlin: Springer Verlag.

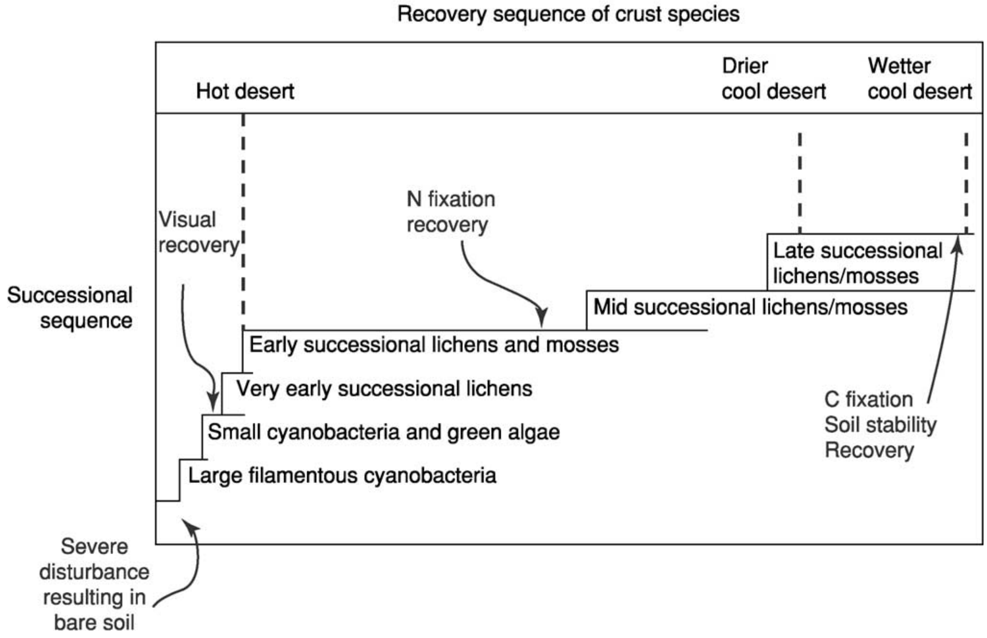

Recovery sequence of crust species
Hot desert
Drier cool desert
Wetter cool desert
Visual recovery
N fixation recovery
Successional sequence
Late successional lichens/mosses
Mid successional lichens/mosses
Early successional lichens and mosses
Very early successional lichens
Small cyanobacteria and green algae
Large filamentous cyanobacteria
C fixation
Soil stability
Recovery
Severe disturbance resulting in bare soil

Figure 12 Colonization sequence of individual components of biological soil crusts after severe disturbance. This chart also represents the vulnerability of these components to disturbance: large filamentous cyanobacteria are the most resistant to disturbance, whereas late successional mosses and lichens are the most vulnerable to disturbance. Reprinted with permission from Belnap J and Lange O (eds) *Biological Soil Crusts: Structure, Function, and Management*. Berlin: Springer Verlag.

Table 2 Estimates of recovery time for biological soil crusts on sandy soils after severe disturbance that resulted in the removal of all crust material. Estimates are based on linear extrapolations of data collected from plots disturbed 20–80 years ago

	Years to recovery						
Desert type	*Very early successional*		*Early successional*		*Mid successional*		*Late successional*
Lower mojave (<1500 m, 100 mm)	1200	⟶	3800				
High mojave (1500 m, 200 mm)	200	⟶	800		NK		NK
Colorado plateau (1500 m, 200 mm)	50	⟶	500		NK		NK
Northern great basin (1000 m, 350 mm)	20	⟶	25	⟶	60	⟶	125

Reproduced from Belnap J and Lange O (eds) (2001) *Biological Soil Crusts: Structure, Function, and Management*. Berlin: Springer Verlag.

Further Reading

Belnap J (1995) Surface disturbances: their role in accelerating desertification. *Environmental Monitoring and Assessment* 37: 39–57.

Belnap J (2001) Factors influencing nitrogen fixation and nitrogen release in biological soil crusts. In: Belnap J and Lange OL (eds) *Biological Soil Crusts: Structure, Function and Management*, pp. 241–261. Berlin: Springer-Verlag.

Belnap J and Eldridge D (2001) Disturbance and recovery of biological soil crusts. In: Belnap J and Lange OL (eds) *Biological Soil Crusts: Structure, Function and Management*, pp. 363–383. Berlin: Springer-Verlag.

Belnap J and Gillette DA (1998) Vulnerability of desert biological soil crusts to wind erosion: the influences of crust development, soil texture, and disturbance. *Journal of Arid Environments* 39: 133–142.

Belnap J, Büdel B, and Lange OL (2001) Biological soil crusts: characteristics and distribution. In: Belnap J and Lange OL (eds) *Biological Soil Crusts: Structure, Function and Management*, pp. 3–30. Berlin: Springer-Verlag.

Crawford CS (1991) The community ecology of macroarthropod detritivores. In: Polis G (ed.) *Ecology of Desert Communities*. Tucson: University of Arizona Press.

Dregne HE (1983) *Desertification of Arid Lands*. New York: Harwood Academic Publishers.

Harper KT and Marble JR (1988) A role for nonvascular plants in management of arid and semiarid rangeland. In: Tueller PT (ed.) *Vegetation Science Applications for Rangeland Analysis and Management,* pp. 135–169. Dordrecht: Kluwer Academic Publishers.

McKenna-Neuman C, Maxwell CD, and Boulton JW (1996) Wind transport of sand surfaces crusted with photoautotrophic microorganisms. *Catena* 27: 229–247.

Sagan C, Toon OB, and Pollack JB (1979) Anthropogenic albedo changes and the earth's climate. *Science* 206: 1363–1368.

Wallwork JA (1982) *Desert Soil Fauna*. London: Praeger Scientific Publishers.

Williams JD, Dobrowolski JP, West NE, and Gillette DA (1995) Microphytic crust influences on wind erosion. *Transactions of the American Society of Agricultural Engineers* 38: 131–137.

Structural

R L Baumhardt and R C Schwartz, USDA Agricultural Research Service, Bushland, TX, USA

Published by Elsevier Ltd.

Introduction

Many of the processes that define soil productivity are governed at the soil–atmosphere interface. That is, soil otherwise capable of supplying nutrient and water needs of a crop may become deficient because of physical or transport limitations within a thin surface layer overlying the bulk soil. Frequent and often extensive changes in the physical conditions of the soil surface affect various fundamental hydrologic and biological processes, further emphasizing the importance of this very dynamic interfacial layer. This modified surface layer is called a 'crust' when dry or a 'seal' when wet.

Crusts occur as biological or physical surface layers that modify plant growth and various exchange processes in virtually all soils. Biotic, i.e., biological soil crusts are 10–100-mm-thick microbial plant communities that usually develop on semiarid and arid soils, which remain relatively undisturbed for decades. In contrast, physical soil crusts develop in response to a combination of raindrop impact and aggregate dispersion to form rapidly, in minutes, a 1–10-mm surface layer that is less porous and less conductive than the underlying bulk soil. Soil crusts typically reduce infiltration, thus causing increased runoff and, hence, water erosion. Additionally, physical soil crusts can impede emerging crop seedlings and reduce plant stand density and seedling vigor. The impact of this thin surface layer on hydrologic processes and agronomic practices has stimulated investigation of physical soil crusts. These investigations link soil mineralogy, physical properties, and the interacting soil and water chemical conditions to crust formation and the development of practical management practices.

Crust Types and Morphology

Biological crusts found in arid or semiarid regions are formed from microbial plant communities of cyanobacteria, green algae, lichens, mosses, and microfungi that colonize the near-surface soil during an establishment period of many years. (*See* **Crusts:** Biological.) This network of bacterial filaments, fungal hyphae, and moss rhizines binds an unconsolidated, usually sandy, soil matrix into cryptobiotic, cryptogamic, microbiotic, or microphytic soil crusts that range in thickness from 10 to 100 mm. Biotic crusts can constitute up to 70% of the soil cover in nutrient-poor areas between localized growths of vascular plants, thus modifying albedo and, subsequently, carbon and nitrogen fixation. Other functions performed by biotic crusts include stabilizing the soil surface against wind or water erosion, and promoting seedling establishment and plant growth by concentrating water and nutrients in microdepressions near crust fractures. The worldwide distribution of biotic crusts in desert-like environments highlights the importance of their complex ecologic role in promoting the growth of vascular plants, which in turn benefits various animal communities, including ungulates such as deer and elk. The primary emphasis of this article, however, will be about physical crusts.

Physical soil crusts, in contrast to biotic crusts, are thin, 1–10-mm-deep, layers that, compared with the underlying bulk soil, are less conductive, denser, and often cemented between soil aggregates and primary soil particles. Physical soil crusts may be divided into depositional or structural categories depending on the primary process or agent of formation. Depositional soil crusts are formed primarily in association with ponded water, e.g., flood and furrow irrigation conditions. Water slakes aggregates and disperses the soil, which is subsequently deposited in sedimentation layers. Surge irrigation makes use of depositional crusts formed by periodic water application to depress in-furrow infiltration and modify irrigation advance rates to distribute water applications more evenly (*see* **Irrigation:** Methods).

In contrast, structural crusts form during rain or irrigation wetting processes that initially weaken and then slake or disperse soil aggregates. Impacting raindrops simultaneously fracture aggregates and detach soil particles. Soil aggregates, primary particles, and colloids are carried, subsequently, into and through the bulk soil matrix by infiltrating water. As the water infiltrates, soil particles and aggregates precipitate and occlude the near-surface conducting pores of the bulk soil, thus forming a ‘washed-in layer,’ as shown in Figure 1. This gradual

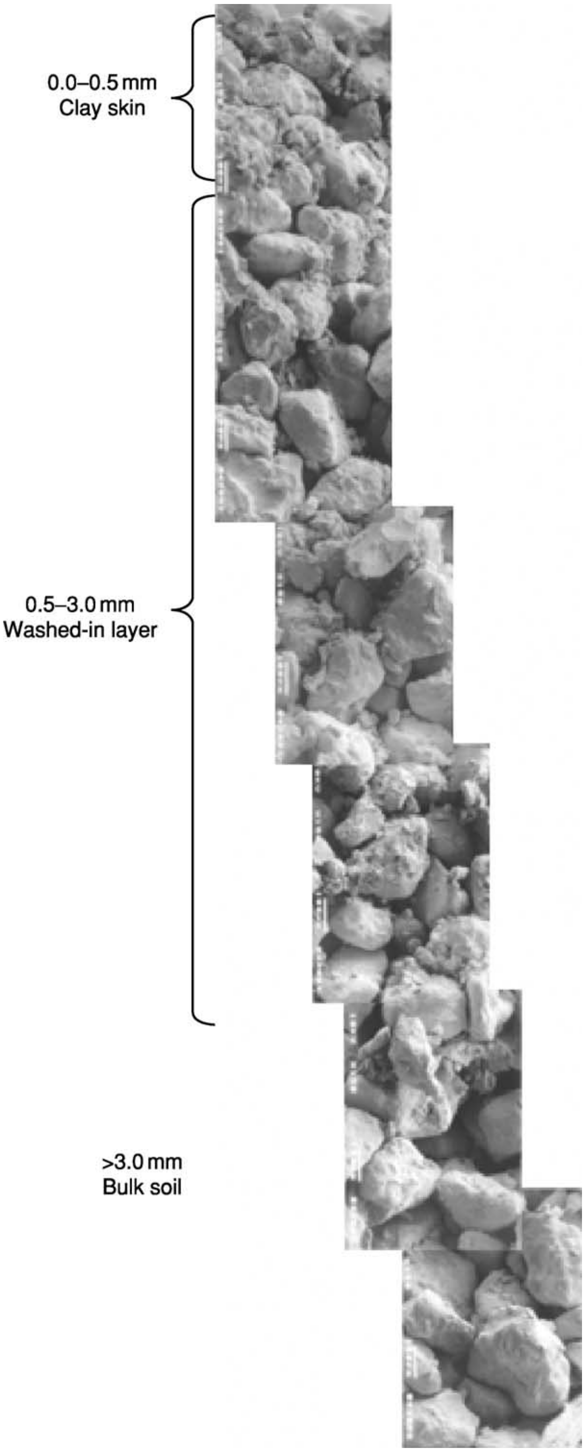


Figure 1 Scanning electron micrographs of a rain-formed structural crust, showing the typical morphology of skin and washed-in layers overlying the bulk soil. Adapted from Wakindiki IIC and Ben-Hur M (2002) Soil mineralogy and texture effects on crust micromorphology, infiltration, and erosion. *Soil Science Society of America Journal* 66: 897–905.

occlusion of conducting pores further reduces infiltration sufficiently to cause surface ponding that further accelerates aggregate slaking and dispersion. The resulting deposition of fine soil material at the soil surface forms a thin ‘skin’ layer (approximately 0.1 mm) over the remaining conducting pores and completes soil crust formation. The soil pore micromorphology of depositional crusts often reflects sand or silt bedding planes and occasional vesicular structure in contrast with the more complex changes found in structural crusts that include pore-size reduction or partial blocking of conducting pores.

Factors Governing Crust Formation

Rainstorm Characteristics

Virtually all soil surfaces are subject to mechanical changes attributed to either dispersion or the cumulative beating action of raindrop impact that rearranges aggregates and particles into a less-conductive, compacted crust layer. Intercepting drop impact with an energy-absorbing material such as crop residues appreciably reduces crust formation. For example, Figure 2 illustrates the ability of raindrop impact on a bare soil to form a smooth, crusted surface in contrast with the case where raindrop energy is intercepted before reaching the soil (no drop impact). Cumulative rainstorm impact energy calculated as the product of rain intensity and the kinetic energy density (drop size) is frequently used, as a first approximation, to describe the potential for physical changes in the surface soil matrix during crust formation. One such relationship between increasing cumulative rainstorm kinetic energy and the decreasing hydraulic conductance through the crust or seal is shown in Figure 3. Crust formation is delayed during initial wetting and softening of the surface soil aggregates that absorb drop impact without any corresponding change in the surface conductance. Continued drop impact on the wetted soil surface, however, fractures aggregates and releases primary soil particles into the infiltrating water, resulting in pore occlusion of the developing crust or seal and a rapid decrease in the crust conductance to a final value. Increasing storm energy density independent of rain intensity accelerates crust or seal formation and, consequently, depresses the infiltration at an accelerated rate (Figure 4). In this example, the final infiltration rate converged to approximately the same value for all nonzero drop impact conditions.

Alternatively, when storm energy density remains constant, increasing rain intensity increases kinetic energy accumulation and similarly accelerates crust or seal formation, which results in a steep decline in infiltration rates. Higher rainfall intensities also shorten the time to ponding independent of energy density. Infiltration rate plotted with time in Figure 5 decreased more rapidly with increasing rain intensity; however, the final infiltration rate was greater with the higher rain intensity. Comparison of rainstorm intensity and kinetic energy density effects on both final infiltration rate and the corresponding hydraulic conductance through the crust or seal are shown in

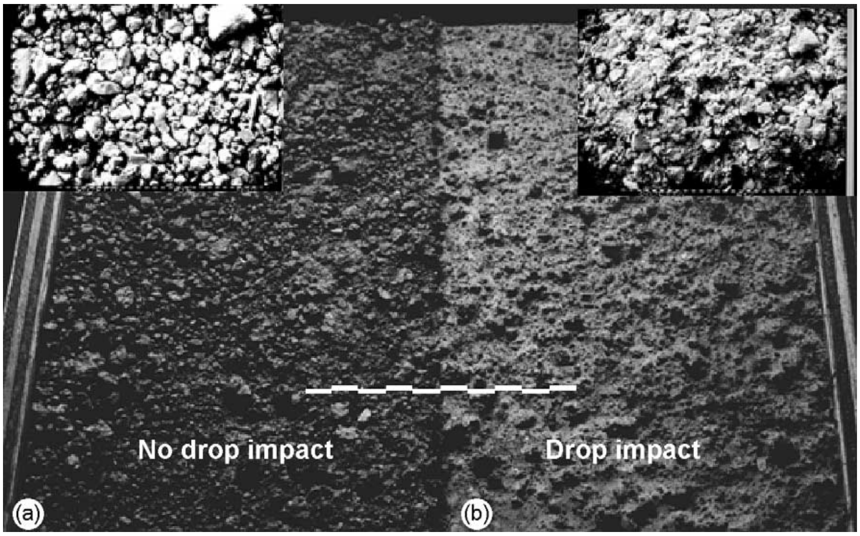


Figure 2 The surface of a silt loam soil exposed to simulated rain with no drop impact (a) has negligible crust formation in contrast to normal drop impact conditions (b). Scale 10 cm. Fractured soil aggregates provide primary particles and microaggregates needed to develop a crust and render a smooth surface (inset, b) compared with the practically undisturbed soil aggregates not exposed to raindrop impact (inset, a). Reproduced with permission from Gantzer CJ (1980) *Physical and Morphological Investigations of Surface Seal Formation on Selected Soils*. PhD Thesis. Saint Paul, MN: University of Minnesota Press.

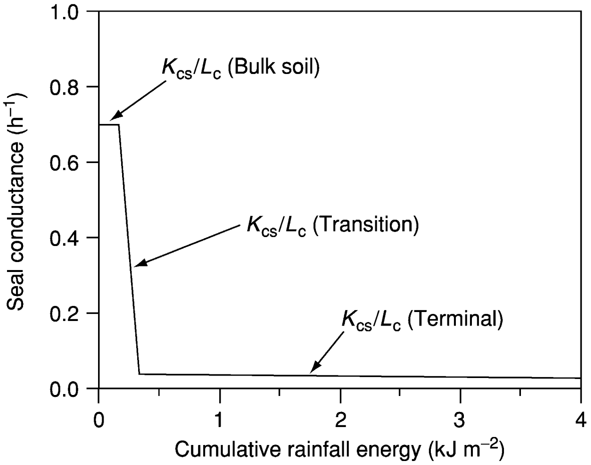


Figure 3 Seal or crust conductance through a silty clay loam crust calculated as the ratio of the crust's saturated hydraulic conductivity, K_{cs}, divided by crust thickness, L_c, and plotted as a function of cumulative rainfall energy. Initially, crust conductance is equivalent to the bulk soil; however, surface wetting and the raindrop impact modifies the soil surface matrix, resulting in a rapidly decreasing crust conductance that changes to a terminal value. Adapted from Baumhardt RL, Römkens MJM, Whisler FD, and Parlange J-Y (1990) Modeling infiltration into a sealing soil. *Water Resources Research* 26: 2497–2505.

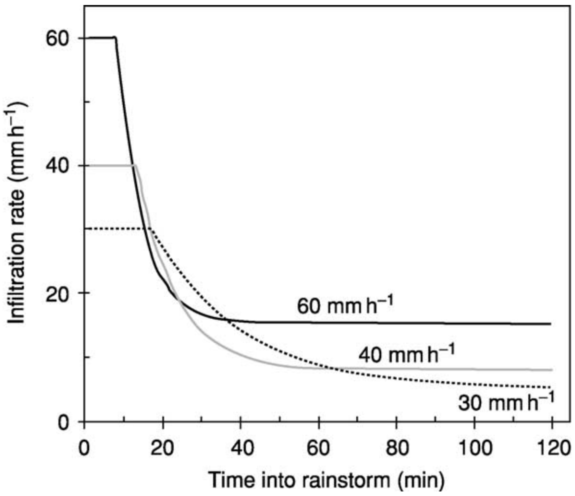


Figure 5 Modeled infiltration rate (millimeters per hour) during a rainstorm with a kinetic energy density of 0.0275 kJ m^{-2} is plotted against time for application intensities of 30, 40, and 60 mm h^{-1}. As rain intensity increases, observed infiltration decreases more rapidly; however, infiltration rate after 120 min is greater with higher rain intensity that erodes the seal or crust and increases its conductance. Adapted from Baumhardt RL, Römkens MJM, Whisler FD, and Parlange J-Y (1990) Modeling infiltration into a sealing soil. *Water Resources Research* 26: 2497–2505.

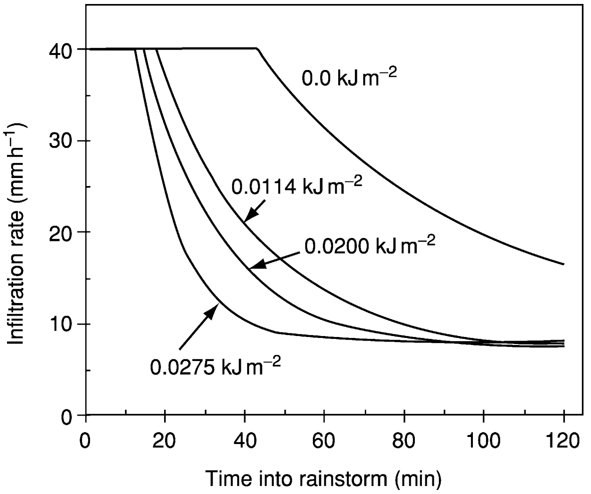


Figure 4 Modeled infiltration rate (millimeters per hour) is plotted with time into rainstorm for intercepted raindrop impact 0.0 kJ m^{-2} compared with storms having progressively higher kinetic energy densities, 0.0114–0.0275 kJ m^{-2} (larger drop sizes). Increasing rainstorm energy density progressively decreases infiltration more rapidly because of accelerated crust formation and decreased crust conductance. Intercepting drop impact delays time of ponding and increases infiltration rate. Adapted from Baumhardt RL, Römkens MJM, Whisler FD, and Parlange J-Y (1990) Modeling infiltration into a sealing soil. *Water Resources Research* 26: 2497–2505.

Table 1. These data show that both final infiltration rate and the corresponding crust hydraulic conductance increased with increasing rainstorm intensity for different energy densities. Possible explanations for these and similar observations of rain-intensity effects include formation of a thinner crust, erosion of the developing crust, and the formation of a water film that absorbs drop impact energy. Higher rain intensities cause more rapid crust formation, thus preserving a higher gradient to drive infiltration, and may not independently modify the surface hydraulic conductivity. Because rainstorm energy density does not vary widely within common rain intensities, crust properties may be estimated as a function of the cumulative rainfall or infiltration rather than from more cumbersome rainstorm intensity and cumulative energy relationships.

The mechanical crusting process is accompanied by a chemical interaction between rainwater and the soil surface that contributes to soil flocculation or dispersion and subsequent aggregate formation or slaking. Water and soil chemical properties interacting to flocculate soil delay and diminish crust formation, while monovalent cations such as sodium that disperse clays on wetting collapse soil aggregation and increase crusting. For example, applying distilled water as simulated rain to a sandy loam soil with an exchangeable sodium percentage (ESP) of 1.0 produces a 7.5-mm h^{-1} final infiltration rate. Increasing the soil ESP from 1.0 to 2.2 and 4.6, a more dispersive condition, results in the formation of crusts that depress the final infiltration rate to 2.3 and 0.7 mm h^{-1}, respectively. Water with a high polyvalent electrolyte

Table 1 Measured final infiltration rate and crust conductance for three kinetic energy densities (drop sizes) and increasing simulated rainstorm application intensities

	Kinetic energy (kJ m^{-2} per mm rain)					
Application intensity	*Final infiltration rate (mm h^{-1})*			*Final conductance (h^{-1})*		
(mm h^{-1})	*0.0275*	*0.0200*	*0.114*	*0.0275*	*0.0200*	*0.114*
30	5.4	6.2	6.9	0.025	0.030	0.035
40	8.1	7.7	7.4	0.035	0.035	0.040
60	15.3	14.0	9.7	0.070	0.050	0.060
90	19.3	16.6		0.110	0.080	

Adapted from Baumhardt RL, Römkens MJM, Whisler FD, and Parlange J-Y (1990) Modeling infiltration into a sealing soil. *Water Resources Research* 26: 2497–2505.

concentration (EC) flocculates the soil and consequently stabilizes soil aggregates, reducing crust or seal formation. Infiltration rates increase from 1.2 to 7.5 mm h^{-1} as the EC of water applied as simulated rain to the sandy loam soil (ESP = 1.0) increases from 0.0 to 5.6 dS m^{-1}. Understanding the dispersive effects of low EC water in rainfall allows researchers to develop soil-amendment applications of powdered phosphogypsum, which enters solution rapidly enough to increase the rainwater EC sufficiently to reduce soil dispersion and crust formation.

Soil Properties

Crust or seal formation is governed by interacting chemical, biological, and physical properties that affect soil susceptibility to mechanical changes in surface aggregation, pore structure, and consequently depressed final infiltration rate. Physical soil properties affecting crust or seal formation include dynamic conditions and essentially fixed features, e.g., antecedent water content (dynamic) compared with soil texture (fixed). The final infiltration rate of two silt loam soils wetted for 24 h under 0.5-kPa suction is depressed approximately 60% compared with dry soil exposed to identical crusting conditions. Although soil crust formation is observed over various textures, including coarse-textured, sandy soils, those soil textures with greater silt concentrations are the most subject to crusting. For example, as soil silt content increases from 51 to 84%, infiltration decreases as much as 300%, with a corresponding 700% increase in crust strength. This may be due, in part, to similarities in the size of the conducting pores and soil particles forming the washed-in layer. Sandy soil readily forms weak structural crusts in contrast to soil having clays that provide bonding surfaces, which stabilize aggregates and reduce soil susceptibility to crust formation.

Soil chemical properties interact with rainwater to disperse and slake aggregates during structural crust formation, but aggregate stabilization with natural and synthetic flocculating or bonding agents delays formation and reduces the strength of dry crusts. For example, the soil amendment polyacrylamide (PAM) is a potent flocculating agent that stabilizes aggregates by offsetting the dispersive effects of water on primary particles. This amendment is frequently applied in irrigation water to reduce soil crust formation while increasing infiltration and irrigation efficiency. Animal manure and other composted organic materials increase nutrient content and bioavailability, which stimulates microbial activity and promotes the production of complex organic polymers and chemical metabolites that serve as bridges and bonding agents for cementing flocculated clay particles into aggregates (*See* **Aggregation:** Microbial Aspects). Soil aggregates are consequently stabilized naturally by the accumulation of organic matter produced by microorganisms such as fungi, whose hyphae hold soil particles together and generate a glycoprotein (glomalin) cementing agent that helps bond primary soil particles. Other microbial by-products and secretions of complex organic polymers or chemical metabolites similarly strengthen soil aggregates. Polymer soil amendments such as polyvinyl alcohol are applied to cement soil particles synthetically and strengthen aggregates, thus increasing soil aggregate stability and reducing crust formation.

Hydrologic Impact of Crusts

Soil crusts determine field infiltration, runoff, and erosion processes on practically all landscapes, but the dynamic processes involved in surface crusting vary spatially and often result in crusts with different characteristics on the same landscape. The degree to which crusting modifies field hydrology is influenced largely by tillage and cropping practices. For example, the interception of raindrop impact by surface residues and plant canopies can prevent or at least delay the formation of surface crusts. Desirable tillage practices retain sufficient residue to protect approximately 40–60% of the soil surface area by intercepting raindrop impact, thus, maintaining an aggregated surface soil with high infiltration rates. With sufficient biomass production, residue-conserving, no-tillage practices frequently increase saturated conductivities through improved aggregate stability imparted by greater soil organic carbon content.

In semiarid regions, crop residue production is often limited and, generally, provides insufficient cover (less than 40%) to intercept raindrop impact and protect the soil surface from crusting regardless of conservation or no-tillage practices. Soil crusts commonly develop in fields under no-tillage residue management during infrequent but intense rainstorms. Thus, crust development progresses rapidly in soils with dispersible clays and low residue cover and consequently infiltration rates of no-tillage soils are frequently similar to tilled soils after 1 or 2 h of rainfall. A spatially clumped residue distribution may promote higher, effective one-dimensional infiltration rates than fields with evenly distributed residues.

Infiltration – Developed Crusts

Improved understanding of the relationship between crust formation and infiltration processes has permitted the projection of crust impact on field hydrology through the use of progressively more sophisticated models. Empirical models that predicted infiltration as an exponential decay function, such as the Kostiakov- or Horton-type equations, have been replaced largely by mechanistic approaches to calculating infiltration fluxes (*See* **Infiltration**). Infiltration of water through soil with a developed surface crust is calculated as a special, simplest, case of flow through a two-layered system. As steady-state flow is approached, the flux through the crust governs the flux into the more rapidly draining, bulk-soil transmission zone, with a water potential gradient approaching unity. Darcy flux through the crust layer can be described in the expression:

$$q = K_c \frac{h_0 - h_c + L_c}{L_c} \qquad [1]$$

where q is infiltration flux (meters per second), K_c is the flow-averaged conductivity across the crust (meters per second), h_0 is the positive hydraulic pressure head (meters) imposed at the surface and usually taken as zero, h_c is the water potential (meters) at the crust interface with bulk-soil, and L_c is crust thickness (meters). Using measurements of the water potential at or just below the crust and the infiltration flux during steady flow, we can rewrite Eqn [1] to estimate crust conductance C (per second) or hydraulic resistance r_c (seconds) as:

$$C = \frac{1}{r_c} = \frac{q}{h_0 - h_c + L_c} \qquad [2]$$

Water flow through crusted soils, therefore, becomes a self-regulating system whereby the physical properties of the crust and bulk soil control the infiltration rate and developing water potentials within the transmission zone. Drainage through the underlying bulk-soil transmission zone under a unit gradient causes the water potential beneath the crust to decrease as the hydraulic conductance through the crust decreases.

Infiltration – Developing Crusts

In contrast to steady-state infiltration through developed soil crusts with fixed conductive properties, forming soil crusts present complex challenges in quantifying infiltration and consequently field hydrology. Approximations of infiltration into a crusting soil use simplified, quasianalytical methods based on variants of the Green–Ampt equation derived to account for the presence of a hydrologically developed surface crust. Assuming a constant water potential at the wetting front (h_f), a uniformly wetted transmission zone to a depth, L_f, with a constant water content $\theta(h_f)$ and conductivity $K(h_f)$, and a soil profile with an initially uniform water content (θ_i), one can approximate the rate of infiltration into a crusted or sealed soil as:

$$\frac{dI}{dt} = \frac{K(h_f)[h_0 - h_f + L_f]}{L_f + K(h_f)r_c} \qquad [3]$$

where I is cumulative infiltration. Setting I equal to $L_f \cdot [\theta(h_f) - \theta_i]$ permits a time-dependent solution for infiltration from known crust resistance and the $K(\theta)$ and $h(\theta)$ functions. However, initial infiltration is overestimated with Eqn [3] because it implicitly assumes a constant water potential h_f at the crust–soil interface, when, in fact, it increases asymptotically toward a steady-state water potential. Piecewise implementations of Green–Ampt-type equations (and other derivations) have been developed to address the dynamics of potential and conductivity changes in the transmission zone. One such formulation is:

$$\frac{dI}{dt} = K(h_f)\frac{\int_{h_f}^{h_i} K(h)/K(h_f)dh + SL_f}{SL_f} \qquad [4]$$

where S is a shape factor that ranges from 1.0 to 1.2. To solve for I and h_f with time, Eqn [4] is combined with the surface boundary condition:

$$q = -K(h_f)\left[\frac{\partial h}{\partial z} - 1\right] = \frac{-h_f + L_c}{r_c} \qquad [5]$$

Although infiltration fluxes calculated using Eqn [4] with an r_c measured after 2 h achieve satisfactory agreement with two-layer solutions of the Richards equation (Figure 6), the applied r_c was estimated before steady-state flow conditions. Because the r_c calculated with Eqn [2] decreases asymptotically over time for a stable crust, crust resistances measured during near steady-state flow conditions (e.g., after 20 h) are underestimated. Consequently, use of the steady-state r_c leads to serious overestimation of earlier infiltration fluxes with the Green–Ampt approach (Figure 6). This estimation error may be attributed to the effects of rapidly increasing crust resistances as the soil surface matrix changes and because of the concomitant changes in the unsaturated conductivity with increasing water content of the crusting or sealing surface.

Field-scale models of infiltration into crusted landscapes often apply a Green–Ampt approach. The decrease in the soil surface conductance through a developing crust is typically described with an exponential decay function that incorporates the cumulative kinetic rainfall energy since the last tillage event and a soil stability factor. Not surprisingly, raindrop intercepting residue cover greatly influences the Green–Ampt parameters describing crust effects on infiltration. This approach has been found sufficiently sensitive to simulate the effects of tillage, residue cover, and crusting when projecting infiltration for field-scale applications. Added precision in calculating field infiltration requires more complex techniques to account for the changing hydraulic properties of a dynamic crust. This is further complicated by the spatial uncertainty in describing the crust-formation process.

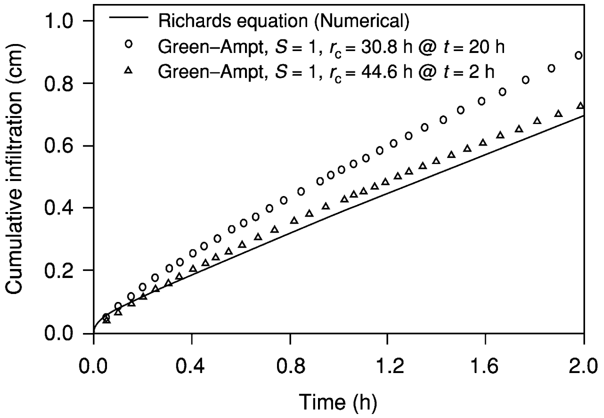


Figure 6 Cumulative infiltration for a fine-textured soil calculated using a two-layer numerical solution to the Richards equation with a 5-mm seal and using the Green–Ampt formulation (Eqns [4] and [5]). Crust resistances for the Green–Ampt solutions were calculated from the soil–crust interface water potentials and surface fluxes obtained from the numerical solution at $t = 20$ h (near steady state) and $t = 2$ h.

Numerical Infiltration Models for Developing Crusts

Crust formation begins when the energy imparted by the first falling raindrops transforms the bulk soil matrix into a more dense and less-conductive surface layer and continues beyond incipient ponding. Simultaneously, raindrops wet the soil and increase the unsaturated hydraulic conductivity of the developing crust. The earlier relationship between the decreasing crust hydraulic conductance and cumulative rainstorm kinetic energy is revisited in Figure 7; however, the unsaturated hydraulic conductivity of the crust (dashed line) initially increases as the soil wets before being reversed as the crusted soil matrix saturates. These dynamic surface boundaries are suited to numerical infiltration models.

Finite-difference solutions of the Richards equation are typically used to obtain numerical approximations of infiltration into crusted soils. In these solutions, the crust layer is treated either as a separate layer with distinct hydraulic properties or as a hypothetical membrane with a hydraulic resistance. The latter method is identical to the boundary condition, Eqn [5], and implicitly assumes that the crust layer becomes rapidly saturated. Because the crust or seal–soil interface may never become saturated, r_c does not represent an independent crust or seal hydraulic property, and imposing it under differing flow conditions can result in errors of undetermined magnitude.

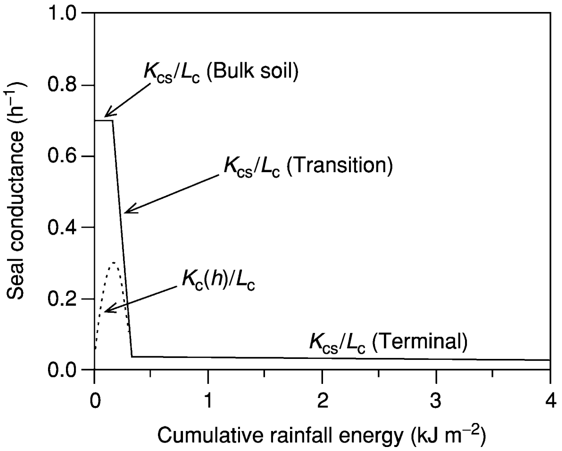


Figure 7 Model-calculated unsaturated hydraulic conductivity, $K_c(h)$, during crust formation increases as the soil is wetted (dashed line); however, as the surface soil matrix crusts, K_{cs} rapidly decreases with increasing cumulative rainfall energy and causes a corresponding reduction in $K_c(h)$ that eventually follows the previously plotted conductance through a silty clay loam crust. Adapted from Baumhardt RL, Römkens MJM, Whisler FD, and Parlange J-Y (1990) Modeling infiltration into a sealing soil. *Water Resources Research* 26: 2497–2505.

Treating the crust as a thin layer with dynamic bulk density and $K(h)$ and $\theta(h)$ functions that are independent of the underlying soil has stimulated the use of numerical models to estimate infiltration into crusting soils. In these models, the hydraulic properties of the crust change in response to the cumulative kinetic energy, rainfall intensity, and soil-rainwater chemical properties. For example, the changes in crust hydraulic conductance are functionally related to rainstorm cumulative kinetic energy, thus providing an estimate of the saturated crust or surface soil matrix hydraulic conductivity K_{sc} (Figure 7). Corresponding adjustments to crust porosity can be calculated from K_{sc} using the Kozeny–Carman equation. Assuming that decreasing porosity results from isometric reduction in pore radius, concomitant adjustments to the air-entry value may be calculated using the Poiseuille equation in conjunction with the capillary equation. An alternative approach describes the increased bulk density of the surface soil matrix changes resulting from raindrop impact and pore occlusion with sediments carried by infiltrating water to varying depths. The $K(h)$ and $\theta(h)$ functions are related to the crust bulk density and therefore are time-dependent.

A fundamental difficulty with describing changes in the crust matrix is that the physical properties of crusts developing over time generally cannot be directly measured. Crust properties are typically estimated using inverse methods, whereby parameters describing the matrix are adjusted so that calculated infiltration rates match observations. Calculating infiltration using numerical methods that adjust crust or seal conductivity in response to wetting and soil matrix changes must rely on researcher insight to quantify the effects of the crust-formation process.

Agronomic Importance of Crusts

Reduced infiltration through a less-conductive surface crust is accompanied by a corresponding reduction in gas and vapor transport; however, crusts have a complex effect on the soil surface energy balance. For example, the denser-crusted soil surface can increase its thermal conductivity and potential to increase the energy driving evaporation. Conversely, the lower exposed area and higher albedo of the crusted surfaces absorb less irradiant energy, thus reducing the potential evaporation. While tillage to fracture soil surface crusts may reduce albedo and increase infiltration, no-tillage residue management practices retain surface residue to intercept both raindrop impact and irradiant energy. No-tillage residues minimize crust formation and increase storage of precipitation as plant-available soil water.

Soil crusting reduces infiltration and increases the potential for water-driven soil erosion. As the amount of runoff increases, the potential entrainment and transport of sediment increases, which is manifested as increased soil erosion. Alternatively, crusted soil is 40–70 times less erodible to wind than the corresponding unconsolidated material. This benefit is, generally, offset by the increased surface wind speeds and the loss of large-aggregate shelter angle effects with a smooth crust surface. In the absence of standing residue to reduce surface wind speed and provide a nonerodible barrier against saltating sand grains, the distribution of fine soil particles within a crust exposes sand particles at the surface that may move with the wind and further contribute to saltation or surface-creep wind erosion. Soil erosion increases as the abrasive action of the moving sand grains on the crust provides more material from the soil surface, thus escalating wind erosion. Tillage to fracture the soil crust is required to disrupt sand saltation and abrasion of the eroding soil. The 'blowing sand' observed from eroding crusted soils produces severe collateral crop damage to tender seedling tissues.

Reduced and delayed seedling emergence resulting in poor crop establishment is a common agronomic production hazard caused by soil crusts (*See* **Germination and Seedling Establishment**). The cemented soil crust also acts as a barrier to an emerging crop (Figure 8) which must either grow through a fracture or exert sufficient force to lift or penetrate the crust. Soil crust strength varies with formation conditions. For example, the strength of crusts formed by overhead sprinkler irrigation with lower

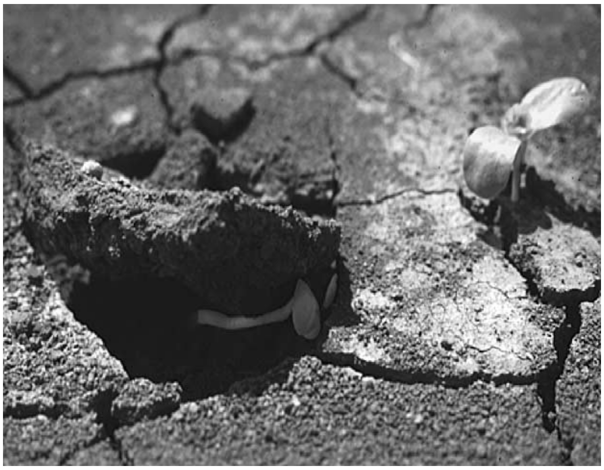


Figure 8 The reduction of seedling emergence is, possibly, the most visible and the most significant agronomic impact by a soil crust. Emergency tillage operations are performed to fracture the crust and, when ineffective, are frequently followed by replanting decisions.

drop impact energy and less-dispersive well water will have a lower penetration resistance (approximately 0.5 MPa using a fine-probe penetrometer) compared with crusts formed during intense natural rainstorms (approaching 1.2 MPa). Likewise, the crust impact on seedling emergence varies with the crop. The force exerted by a seedling varies from 0.15 N for alfalfa (crust susceptible) to more than 4.0 N for corn (crust tolerant) depending on water imbibition and growth-limiting factors such as temperature and seed mass. For cotton seedlings exerting a maximum emergence force of 3.02–4.63 N, the corresponding measured axial pressure (or maximum penetrable crust strength) ranges from 1.25 to 1.90 MPa. As the crust penetrometer resistance increases from 0 to 1.0 MPa, cotton seedling emergence decreases from 78 to 21% 2 days after planting, thus illustrating the impact of soil crusts to injure sensitive crops.

Agronomic practices to improve seedling emergence through a crust rely on tillage to fracture the crust or small irrigation applications (when available) to wet and soften the crust. Directed tillage to fracture soil crusts improves seedling emergence, but also imposes injury risks to emerged seedlings. Irrigation to soften the crust may not be available or practical because of the resulting depression in seed-zone soil temperature. Some cotton producers in California (USA) and elsewhere control severe crusting effects on seedling emergence by removing previously formed soil caps above the seed (Figure 9). The cap may also reduce the amount of force required to penetrate and/or fracture the crust by providing a fracture zone along the cap peak or reducing the force required to move the crust.

Agronomic management paradigms achieving the greatest success in minimizing the impact of soil crusts attempt to reduce crust formation. Residue-retaining tillage practices reduce soil crust formation by intercepting raindrop impact and ameliorate crust impact on seedling emergence by delaying surface drying. Likewise, agronomic cultural practices that retain residues on the soil surface can decrease runoff and surface wind speeds attributed to soil crust formation and, consequently, decrease water and wind erosion while increasing plant-available soil water. The continued study of soil crusts will continue to result in agronomic management practices to reduce crust formation and minimize crust impact.

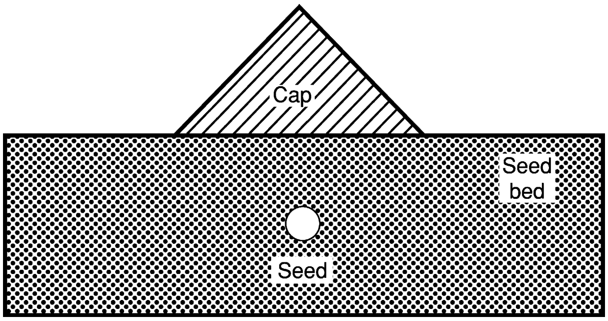


Figure 9 Diagram of seed and seedbed with a soilcap (small hill). Soil caps can be removed by tillage to encourage seedling emergence under crusting conditions.

See also: **Aggregation:** Microbial Aspects; **Compaction**; **Crusts:** Biological; **Flocculation and Dispersion**; **Germination and Seedling Establishment**; **Infiltration**; **Irrigation:** Methods

Further Reading

Ahuja LR (1983) Modeling infiltration into crusted soils by the Green–Ampt approach. *Soil Science Society of America Journal* 47: 412–418.

Assouline S and Mualem Y (2000) Modeling the dynamics of soil seal formation: analysis of the effect of soil and rainfall properties. *Water Resources Research* 36: 2341–2349.

Baumhardt RL, Römkens MJM, Whisler FD, and Parlange J-Y (1990) Modeling infiltration into a sealing soil. *Water Resources Research* 26: 2497–2505.

Belnap J, Kaltenecker JH, Rosentreter R, Williams J, Leonard S, and Eldridge D (2001) *Biological Soil Crusts: Ecology and Management*. Technical Reference 1730. Denver, CO: US. Department of the Interior.

Cary JW and Evans DD (1974) *Soil Crusts*. Technical Bulletin 214. Tucson, AZ: University of Arizona Agricultural Experiment Station.

Duley FL (1939) Surface factors affecting the rate of intake of water by soils. *Soil Science Society of America Proceedings* 4: 60–64.

Gantzer CJ (1980) *Physical and Morphological Investigations of Surface Seal Formation on Selected Soils*. PhD Thesis. Saint Paul, MN: University of Minnesota Press.

Hillel D and Gardner WR (1969) Steady infiltration into crust-topped profiles. *Soil Science* 108: 137–142.

Hillel D and Gardner WR (1970) Transient infiltration into crust-topped profiles. *Soil Science* 109: 69–76.

Mannering JV (1967) *The Relationship of Some Physical and Chemical Properties of Soils to Surface Sealing*. PhD Thesis. West Lafayette, In: Purdue University Press.

McIntyre DS (1958) Permeability measurements of soil crusts formed by raindrop impact. *Soil Science* 85: 185–189.

Poesen JWA and Nearing MA (1993) *Soil Surface Sealing and Crusting*. Cremlingen, Germany: Catena-Verlag.

Römkens MJM, Baumhardt RL, Parlange MB, Whisler FD, Parlange J-Y, and Prasad SN (1985) Rain-induced surface seals: their effect on ponding and infiltration. *Annales Geophysicae* 4: 417–424.

Shainberg I (1985) The effect of exchangeable sodium and electrolyte concentration on crust formation. In: Stewart BA (ed.) *Advances in Soil Science*, pp. 101–122. New York: Springer-Verlag.

Sumner ME and Stewart BA (1992) *Soil Crusting: Chemical and Physical Processes*. Boca Raton, FL: Lewis Publishers.

Tackett JL and Pearson RW (1965) Some characteristics of soil surface seals formed by simulated rainfall. *Soil Science* 99: 407–412.

Wakindiki IIC and Ben-Hur M (2002) Soil mineralogy and texture effects on crust micromorphology, infiltration, and erosion. *Soil Science Society of America Journal* 66: 897–905.

CULTIVATION AND TILLAGE

M R Carter, Agriculture and Agri-Food Canada, Charlottetown, PE, Canada
E McKyes, McGill University, Ste-Anne de Bellevue, PQ, Canada

Introduction

Cultivation and tillage are activities related to soil manipulation and movement. Soil movement can occur due to natural (e.g., freeze–thaw processes) and biotic (e.g., earthworm movement) processes; however, both ‘cultivation’ and ‘tillage’ generally refer to the direct intervention of a ‘soil manager,’ who uses tillage tools to modify, manipulate, or ameliorate a soil condition. Although both terms are used interchangeably, ‘cultivation’ is usually restricted to soil manipulation to create a seedbed suitable for crop seeding and establishment. In comparison, the term ‘tillage’ serves a wider function than cultivation. As a subsystem of a crop-production system, tillage is employed to facilitate crop establishment, modify soil structure, incorporate fertilizer and soil amendments (e.g., lime and manure), control plant residues and weeds, and alleviate both climatic and soil constraints. Over time, basic tillage tools have been developed to perform specific soil-moving operations. A combination of tillage operations and their timing results in the development of a tillage system to provide specific functions in given situations.

Goals of Soil Tillage

The traditional aims of tillage are to improve soil structure for crop growth, incorporate organic amendments into the soil, and to control weeds. The last goal can often be met by use of herbicides, and this has led to the development of no-tillage systems, where tillage is confined to soil disturbance associated with crop seeding or planting. Soil tillage is also involved in soil-water conservation and regulation, and in soil-erosion control. Table 1 outlines some of the various goals of tillage in agricultural systems.

The role of tillage in soil structure improvement has various facets related to crop growth and productivity. Tillage is conducted to improve soil functions such as water and air regulation and flow, to enhance the water-storage capacity of the soil, and to create a desirable aggregation size distribution conducive for crop seed–soil contact. Tillage is often needed to increase soil-water infiltration and thus to improve soil drainage. Soil loosening to ameliorate soil

Table 1 Specific goals of soil tillage in agricultural systems

Reasons for use of soil tillage	*Critical processes and characteristics*
Preparation of a seedbed	Creation of a soil aggregate size distribution to suit specific seed size
Weed control	Removal of unwanted plant species that compete with the crop for soil resources. Use of herbicides can remove the requirement for tillage
Soil erosion or degradation	Excess tillage can create a potential for soil erosion. Reduced tillage systems can retain crop residues that can help decrease the potential for both water and wind erosion
Conserve or regulate soil water	Tillage can conserve soil water content by manipulation of both soil structure and crop-residue retention. Excess soil water can also be removed by deep tillage
Improve plant root penetration and growth	Shallow tillage can remove soil-surface crusts that impede plant germination, while deep tillage can remove compact, impervious soil layers. Tillage can improve soil aeration
Incorporate crop residue, fertilizer and lime, and organic amendments	Tillage can bury or mix crop residues and incorporate fertilizer additions

compaction is another role of tillage. Tillage is also used to manipulate crop residues, either by burying and mixing them into the soil, or by retaining them at the soil surface to serve as mulch for soil protection or as a barrier to reduce soil and water movement.

The amount and degree of tillage needed are not the same for all soil types and cropping situations. Some soils, due to climate (e.g., cool, wet soils) or type (e.g., high clay or sand content, or poor permeability), have a relatively high 'tillage requirement.' A combination of easily compactable soils and excessive vehicular traffic can lead to compacted subsoils or soil layers that present the need for deep tillage. In contrast, some soils have a relatively low tillage requirement, as they are able to regenerate soil structure under natural processes associated with soil wetting and drying, and freeze–thaw processes. To meet the goals of soil tillage, agricultural engineers have developed various tools and implements which have certain advantages and limitations.

Main Types of Tillage Tools

Moldboard Plow

The term 'moldboard plow' describes an implement that cuts soil, lifts it, and turns it at least partly upside down by means of a curved plate, or moldboard (Figure 1). The concept of the moldboard plow is quite ancient. Wooden plows have been in existence in Asia and Africa for more than 5000 years and adapted versions were in use in Europe 500 years ago, featuring a drawbar for animals, wheels, a leading cutting-coulter, the soil-cutting blade, and a moldboard.

The precursor to the modern moldboard plow was invented by Thomas Jefferson, of Virginia, USA, around 1790, after he had observed European plow designs. He first designed and tested a wooden moldboard that could be duplicated easily. By 1814 he had them cast in iron, and they soon became known throughout North America and Western Europe.

In 1837 John Deere, of Vermont, USA, invented the modern moldboard plow, in Grand Detour, Illinois, using smooth, self-cleaning steel for the moldboard rather than cast iron. By 1847 his company was manufacturing more than 1000 plows per year, and his Moline Plow Works factory was producing 75 000 per year by 1875. This basic design is still being manufactured today, although with numerous improvements.

Applications The moldboard plow performs the following tillage operations and soil-conditioning functions:

- It cuts, lifts, breaks up, and loosens soil that has been compacted through machinery traffic or natural causes to a depth of usually 100–200 mm below the soil surface. Plowed soil is easier to form into a loose seedbed, drains water better, and warms up more quickly after winter;
- It overturns vegetation, buries it, and exposes topsoil. This is useful when developing new arable fields in grasslands or burying crop residues, weeds, and insect pests;
- It has a flat bottom that covers the entire width of the tillage zone, and thus all soil down to the depth of tillage is cut, lifted, and loosened.

Disadvantages and limitations The action of burying most surface vegetation leaves the soil surface much more vulnerable to erosion by wind and water, and leads to an increased loss of soil and fertilizer resources. Furthermore the eroded soil material

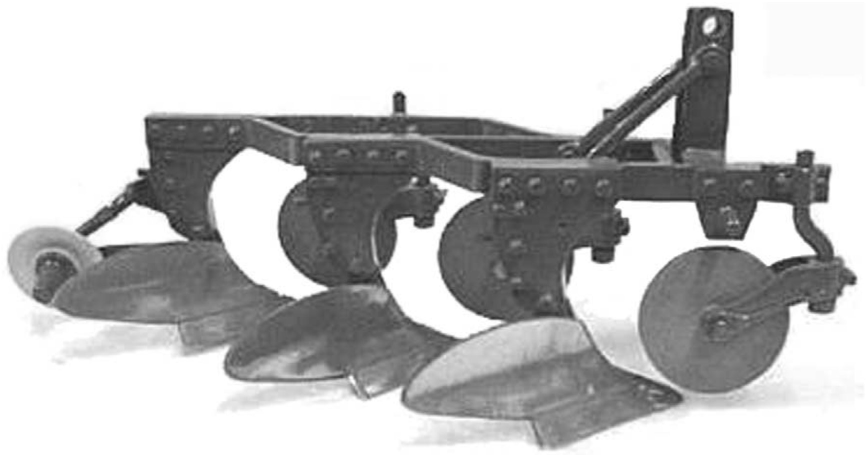


Figure 1 A three-bottom moldboard plow.

and nutrients cause pollution in their recipient surface watercourses. The sealing of the soil surface by impacting rainfall and the resulting reduced water infiltration can be a problem. In addition, continuous plowing can destroy the soil aggregate structure considerably, leading to reduced water and air movement, root growth, and crop yields. The moldboard plow requires the most energy to operate per unit field area of all tillage implements, with the exception of the deep subsoiler.

As a result of these disadvantages, the regular use of the moldboard has been reduced considerably in most of the world over the past 20 years. For example, the US Department of Agriculture (USDA) Agricultural Research Service has estimated that the use of the moldboard plow by American farmers had decreased from 75–85% in 1980 to less than 10% in 1993. Problems of weed and insect pests are offset in the absence of this plow by the use of pesticides or cover crops. Abandoning the moldboard plow has also been reported to result in an increase in the accumulation of organic matter in soils. This implement remains in use, however, in northern cool and humid climates, where the enhanced soil drainage and spring warming outweigh the disadvantages.

Chisel Plow

A basic chisel plow (Figure 2) was used for agricultural tillage at least 5000 years ago in Northern Africa and the Middle East. Wooden chisel ploughs were replaced by iron ones in the 1700s, and by smooth steel in the 1800s. Most chisel ploughs have a shank or leg width of 63.5 mm and an angle at the bottom tip of 40–60° to the horizontal.

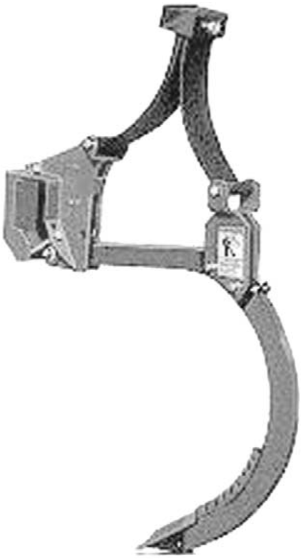


Figure 2 A chisel plow.

Applications The chisel plow also cuts, lifts, and loosens soil prior to the preparation of a suitable seedbed, but differs from the moldboard plow in the following ways:

- Individual chisels are usually spaced at approximately 300 mm laterally, and their width is approximately 63.5 mm, so not all the soil to the tillage depth is loosened. The cross-section of soil loosened is V-shaped, with greater width at the soil surface; therefore there is a resulting lack of homogeneity in the tilled soil profile;
- The chisel plow leaves much more crop residue or stubble on the soil surface, usually four or five times more depending on the crop. This has the advantage of reducing soil erosion from wind and water by 40–80%, decreasing the sealing of the soil surface by rainfall, decreasing the amount of soil moisture that can be lost in a dry spring season, and increasing the organic matter content of the soil surface layer.

Disk Harrow

The disk harrow (Figure 3), consisting of a set of curved, circular rotating disks attached to a frame, has been in use for over 100 years. This implement can be used for secondary tillage, following moldboard or chisel plowing, to smooth the soil surface and reduce the size of aggregates in preparation for planting. It can also be used as the sole tillage implement when the goal is to chop crop residues to facilitate their decomposition and the subsequent planting of crops.

Because of its reversed angle of attack on the soil, each disk relies on its weight to penetrate vegetation and soil. As a result, disks do not lift and loosen soil

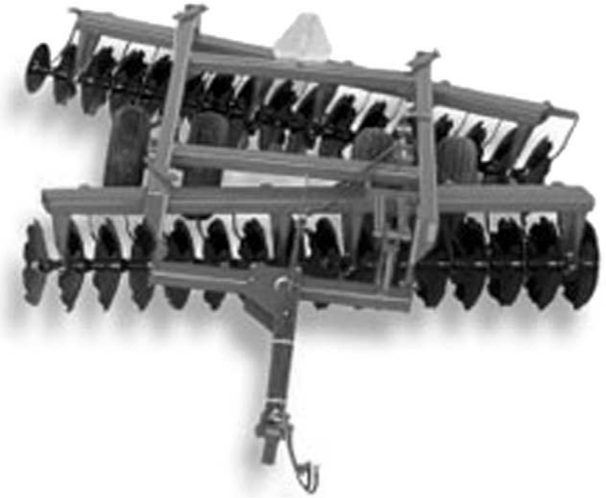


Figure 3 An offset disk harrow. (Photo courtesy of Case IH Corporation.)

much, and, in soils that are subject to compaction, the disk when used alone can produce an excessively dense topsoil layer, with resulting problems in infiltration, drainage, air movement, and root growth.

Toothed Cultivator

The toothed cultivator shank or leg, usually made from flexible spring steel, has the same shape as the chisel plow, except that it is smaller and operates at less depth, typically 20–50 mm (Figure 4). In most cases a shovel-shaped point is fixed to the bottom of each shank, with a width of 50–100 mm and a low attack angle of 20–30° to the horizontal. The cultivator removes weeds from between rows of crops and loosens the top layer of soil.

Subsoiler

The subsoiler is so named because it cuts and loosens soil below the normal tillage depth of 100–200 mm. Its shape is similar to the chisel plow except that it is made with a stronger shank or leg in order to resist the higher force required to till soil at greater depth. The bottom cutting tip can be narrow, 25–50 mm, or a low-angle sweep or foot can be attached as shown in Figure 5. The sweep has the advantage of loosening a greater volume of soil at the expense of only a small increase in the required pulling force or draft.

Use of a subsoiler is indicated when there is a dense and/or hard layer of soil below the normal tillage depth. Such a layer can result from physical causes such as machinery traffic or from physicochemical sources that produce cementation among soil aggregates. Such a hard layer, sometimes called a plow pan, can lead to poor water infiltration and drainage, reduced soil aeration, and diminished growth of crop roots.

No-tillage

No-till or no-tillage agriculture is conducted without the use of any of the tillage implements described above in most years, although occasional tillage may be required to control weeds that are tolerant to

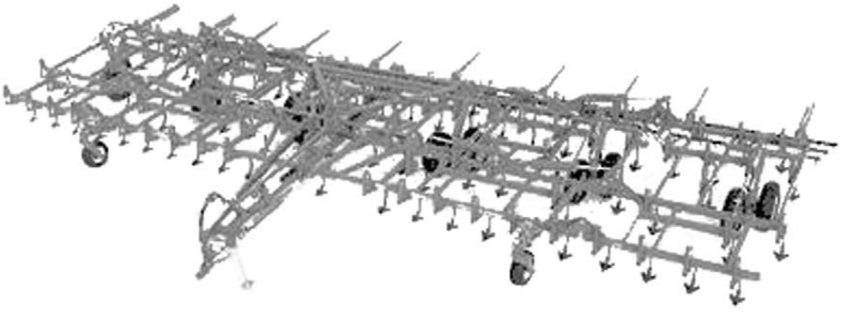


Figure 4 A spring-toothed cultivator.

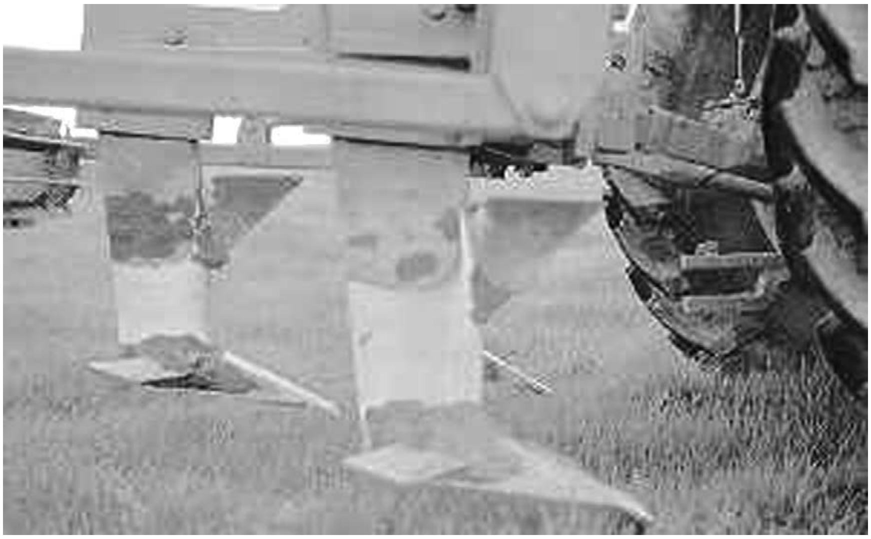


Figure 5 A three-shanked subsoiler. (Photo by K C Cameron, Lincoln University, Canterbury, New Zealand.)

herbicides. Crop seeds are planted directly into the stubble and residue of the previous year's crop. A special planter is generally required to penetrate residual vegetation and a stronger soil surface, and many manufacturers have developed effective no-till planters in the past 30 years. This production system has gained favor on many thousands of farms, especially in the plains of North America, where temperatures and rainfall are moderate.

No-till agriculture can have the following advantageous results:

- Less compression and breaking down of soil aggregates;
- Greater amounts of organic residue on and near the soil surface, resulting in less surface-sealing by rainfall, reduced soil erosion, enhanced moisture retention, and more organic matter accumulation;
- Similar crop yields to those of conventional-till and reduced-till systems, and even higher yields in relatively dry regions such as midwestern North America.

Basic Tillage Operations

Traditionally, most tillage operations consist of primary tillage followed by secondary tillage. This traditional combination is referred to as 'conventional tillage,' although this term is also applied to any tillage system that has been adopted for a period of time. Primary tillage usually consists of relatively deep (15–25 cm) tillage, using a moldboard or chisel plow. In contrast, secondary tillage is relatively shallow (10–15 cm) and used to pulverize and consolidate the soil to form a suitable seedbed. Both primary and secondary tillage are usually conducted uniformly across a field. The last few decades of the twentieth century have seen concerted efforts to reduce the amount and degree of tillage by adoption of reduced, minimum, or shallow tillage. Reduced tillage allows a reduction in depth, degree, and frequency of tillage, while 'minimum tillage' refers to the minimum soil manipulation necessary for crop production or other soil tillage requirement. Shallow tillage, sometimes called 'noninversion tillage,' involves restricting tillage to a shallow (less than 15 cm) soil depth. Usually, the soil is not turned over or inverted, thus the tillage operation is often termed 'ploughless tillage.' Overall, the main impetus behind such developments is to reduce the degree of soil degradation often associated with excess tillage, and also to reduce tillage costs in commercial agriculture.

In many cases, uniform tillage across a field may not be required or desirable. For row crops such as maize (*Zea mays* L.) and potato (*Solanum tuberosum* L.), tillage can be restricted to zones across a field. Partial-width tillage is based on the concept that seedbed conditions can be confined to the planting row, while interrow areas receive less or a different type of tillage treatment. For example, in semiarid climates, for maize production, tillage can be used to create a 'planting zone,' characterized by optimum seedbed conditions conducive to seed germination and establishment, while in the interrow zone a coarse soil structure is maintained to allow optimum water intake. Further to this, crop residue cover can be maintained in the interrow zone to facilitate water and soil conservation.

Basic soil tillage operations consist of inverting, loosening, mixing, and pulverizing, or clod-breaking (Table 2). Specific tillage tools operating at a range of soil depths are used to achieve these operations. Soil inversion is best accomplished using a moldboard plow, while soil-loosening can be achieved using several tillage tools. Rotary cultivators are well suited for mixing soil and thus are often used to mix soil fertilizer and organic amendments into the soil. Plant residues can often pose a problem for tillage tools in regard to interference with the tillage operation. Under such conditions, coulters or disks are often used in combination with the tillage tool to cut or divert crop residue away from the immediate tillage zone.

Types of Tillage Systems

In many cropping situations and climates, innovations in tillage practices lead to the development of conservation tillage and other closely related systems. 'Conservation tillage' is a generic term used to describe tillage systems that have the potential to

Table 2 Main role and function of tillage implements and their efficiency

Tillage tool	*Tillage depth (cm)*	*Inverting*	*Loosening*	*Mixing*	*Clod breaking*
Moldboard plow	15–25	Good	Good	Partial	None
Chisel plow	10–20	Partial	Good	Partial	Partial
Disk plow	10	Partial	Good	Partial	Partial
Rotary cultivator	10	None	Partial	Good	Good
No tillage	5	None	None	None	None

conserve soil and water by reducing their loss relative to some form of conventional tillage. Generally, there are four main types of conservation tillage: mulch tillage, ridge tillage, strip tillage, and no-tillage. Precise definitions of conservation tillage are only possible within the context of known crop species, soil types and conditions, and climates. A well-accepted operational definition of conservation tillage is tillage or a tillage and planting combination which retains 30% or greater cover of crop residue on the soil surface. Combinations of tillage tools can be developed to meet the various requirements posed by both soil and crop needs, and the constraints related to farming systems and climate.

List of Technical Nomenclature

Conventional tillage Combined primary and secondary tillage operations normally performed in preparing a seedbed for a given crop and area

Deep tillage A primary tillage operation that manipulates soil to a greater depth than normal plowing (i.e., more than 100–200 mm; e.g., heavy-duty deep moldboard plow, deep disk plow, heavy chisel, or subsoiler)

Inversion tillage Primary tillage that 'inverts' or 'turns over' soil, causing much soil mixing. The main tillage implement used for inversion tillage is the moldboard plow

Land-forming tillage Tillage operations that move soil in order to create the desired configurations, including level ground, slopes for irrigation, contouring and terracing, small-scale ridging, or pitting

Minimum tillage The minimum use of primary and secondary tillage necessary to meet crop-production requirements under existing soil and climatic conditions, usually resulting in fewer tillage operations than for conventional tillage

Primary tillage The tillage that constitutes the first major soil-working operations, normally designed to reduce soil strength, cover plant and insect materials, and rearrange soil aggregates (e.g., by moldboard plow, chisel plow, and/or disk plow)

Reduced tillage Reduction in total number of primary and secondary tillage operations usually used in conventional tillage, to prepare a soil for crop establishment; combining the primary tillage operation with special planting operations in order to reduce or eliminate secondary tillage operations

Secondary tillage Any of a group of different tillage operations, following primary tillage, that are designed to create refined soil conditions before seed planting. Examples include disk harrow, cultivator chisels or sweeps, and roller harrow

Shallow or non-inversion tillage Primary tillage confined to shallow soil depth (less than 15 cm); absence of soil inversion tillage

Soil cultivation Shallow tillage operations to create improved soil aeration, water infiltration, moisture conservation, a level surface, or to control weeds to promote the growth of crop plants

See also: **Carbon Emissions and Sequestration**; **Compaction**; **Crop-Residue Management**; **Drainage, Surface and Subsurface**; **Germination and Seedling Establishment**; **Shifting Cultivation**; **Structure**; **Tilth**; **Weed Management**; **Zone Tillage**

Further Reading

Cannell RQ and Hawes JD (1994) Trends in tillage practices in relation to sustainable crop production with special reference to temperate climates. *Soil Tillage Research* 30: 245–282.

Carter MR (ed.) (1994) *Conservation Tillage in Temperate Agroecosystems*. Boca Raton, FL: CRC Press.

Gill WR and Vanden Berg GE (1968) *Soil Dynamics in Tillage and Traction*. Washington, DC: US Government Printing Office.

Gregorich EG, Turchenek LW, Carter MR, and Angers DA (2001) *Soil and Environmental Science Dictionary*. Boca Raton, FL: CRC Press.

Hillel D (1980) *Applications of Soil Physics*. New York: Academic Press.

Koolen AJ and Kuipers H (1983) *Agricultural Soil Mechanics*. Berlin, Germany: Springer-Verlag.

Kuipers H (1963) The objectives of soil tillage. *Netherlands Journal Agricultural Science* 11: 91–96.

Lal R (1991) Tillage and agricultural sustainability. *Soil Tillage Research* 20: 133–146.

Larson WE (1964) Soil parameters for evaluating tillage needs and operations. *Soil Science Society of American Proceedings* 28: 119–122.

McKyes E (1985) *Soil Cutting and Tillage*. Amsterdam, The Netherlands: Elsevier.

Spoor G (1975) Fundamental aspects of cultivations. In: *Soil Physical Conditions and Crop Production*, pp. 128–144. London, UK: HMSO.

Unger PW and Van Doren DM Jr (eds) (1982) *Predicting Tillage Effects on Soil Physical Properties and Processes*. ASA Special Publication 44. Madison, WI: American Society of Agronomy.

D

DARCY'S LAW

D Swartzendruber, University of Nebraska-Lincoln, Lincoln, NE, USA

Introduction

In a porous medium consisting of interconnected pores amid the solid particles, one of the foremost physical properties is the liquid content within the network of pores. If the porous medium is soil and the liquid is water, the implications are crucial not only for the growth of plants, but also for the land phase of the hydrologic cycle (e.g., water infiltration, drainage, and water movement to plant roots) and many other aspects of soil physics. The soil water is also a liquid harbor for dissolved or suspended substances, thus enabling pollutants to move through or to accumulate in the soil. In still another context, the water content can have a major influence on the strength of the soil mass in its ability to resist compaction in agricultural soils and to provide foundational support for buildings, bridges, and earthworks such as highways and dams, these being of prime concern in the field of soil mechanics.

For the quantitative study of the soil water content as it changes within the soil mass as a function of position and time, to have basic flow relationships in mathematical form of reasonable simplicity is of great utility for analysis and measurement. The simple example is provided by a water-saturated porous medium, and the more complicated example by an unsaturated porous medium – in which the soil pores are not completely filled with water. This is the more common condition for soils generally, except for wetlands and some lowland soils subject to high water tables or frequent flooding.

Water-Saturated Soil Conditions

The mathematical–physical foundation for analytical descriptions of liquid flow through a porous medium was laid by Henry P.G. Darcy in 1856, in an appendix to a major and famous treatise setting forth a comprehensive design for the municipal water system of the city of Dijon, France. The porous medium was sand, the liquid was water from the system of a local hospital, and the experiments were conducted by two engineers, Ritter and Baumgarten. Darcy was a highly regarded engineer and an esteemed public-spirited citizen. He died in 1858, not long after the publication of his book.

In simplest terms, Darcy's classic law, or equation, states that the water flux q in one-dimensional flow is directly proportional to the driving hydraulic gradient i, or mathematically in scalar form:

$$q = Ki \qquad [1]$$

where the proportionality constant K is called the hydraulic conductivity of the porous medium or soil and is recognized as a composite property of both the porous medium (soil) and the flowing liquid (water). Referring to Figure 1, where A is the constant bulk cross-sectional area of the soil column and Q is the volumetric time rate of water flow perpendicular to A, then $q = Q/A$. Let h_1 and h_2, both being measured from the same but arbitrary datum plane, be the respective hydraulic heads (in equivalent height of

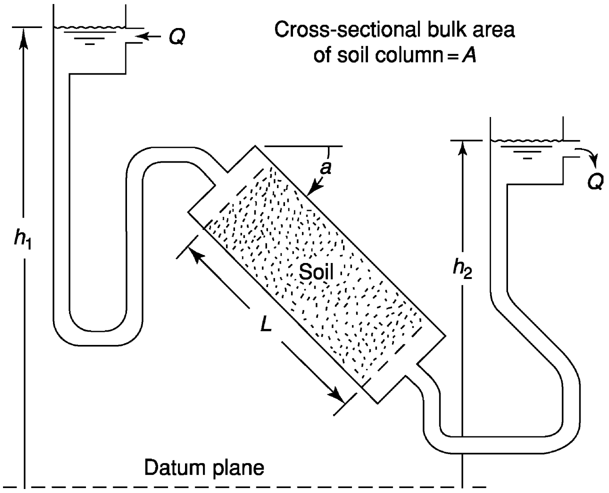


Figure 1 A water-saturated soil column to illustrate Darcy's equation for one-dimensional saturated flow.

water column) at the inlet and outlet ends of the soil column of constant bulk length L, assuming no loss of head in the tubings connecting the inlet and outlet ends to the reservoirs. Then, $i = (h_1 - h_2)/L = -\Delta h/L$. Substituting these expressions for q and i back into eqn [1] and rearranging yields:

$$Q = -KA\Delta h/L \qquad [2]$$

which is essentially the form employed by Darcy. The definition of $\Delta h = h_2 - h_1$ is the standard one of mathematics.

Some textbooks have presented separate diagrams of each of three special angular orientations of the flow column ($a = \pi/2$, 0, and $-\pi/2$ in Figure 1), with eqn [2] written for each orientation, but with h_1 and h_2 expressed as sums of components (i.e., pressure head and elevation head). Unfortunately, these diagrams may foster an illusory impression of Q being affected by the direction of flow. We therefore examine these three angular orientations as follows. Assume in Figure 1 that there is sufficient length and flexibility of the connecting tubing from the inlet and outlet water-level reservoirs, so that Δh can remain fixed (in sign and magnitude) for any angle a between $-\pi/2$ and $\pi/2$. Then, for the first extreme of $a = \pi/2$, the direction of water flow through the now-vertical soil column is vertically upward, maximally counterdirectional to gravity. For the intermediate $a = 0$, the water-flow direction through the now-horizontal soil column is horizontal in the positive direction (left to right), independent of gravity. For the second extreme of $a = -\pi/2$, the water-flow direction through the once-again-vertical soil column is vertically downward, maximally codirectional with gravity, and was the column orientation used in Darcy's experiments. Noting Q for each of these three a values with Δh unchanged, observe from eqn [2] that Q will be the same for all three orientations, because K, A, Δh, and L remain the same. This same Q would also hold for all values of a such that $-\pi/2 < a < \pi/2$. Thus, for a given soil column (fixed K, A, and L), Q only changes with Δh, so that Δh is the sole determiner of Q regardless of flow direction due to soil-column orientation. Hence, there is no need to present the three separate orientation diagrams and their corresponding equations for Q, because Figure 1 and eqn [2] encompass them all.

On the basis of viscous fluid flow through small-bore cylindrical tubes or between flat parallel plates, the flux of fluid is ideally proportional to the density ρ of the fluid and inversely proportional to its absolute viscosity η. This suggests the possibility of partitioning the K of eqns [1] and [2] by:

$$K = k\rho g/\eta \qquad [3]$$

where g is the acceleration of gravity, and k is a constant independent of the flowing fluid and hence dependent only on the pore system of the soil. The name for k is either 'intrinsic permeability' or merely 'permeability' if there is no likelihood of confusion with K. In principle, the permeability k should be independent of the fluid used to determine it, provided that no interaction occurs between the fluid and the soil particles to change the internal geometry of the soil-pore system. In practice, this seems to hold reasonably well for sand and polar (e.g., water) or nonpolar fluids and air. But for soil containing swelling colloids or clays, with the fluid being water alone or water with dissolved electrolytes, the strong fluid–soil interaction can alter the internal geometry of the soil-pore system, so that k will not be independent of the flowing fluid.

Validity of Darcy's Equation

Darcy himself was well aware that eqn [1] would not hold for values of i and q large enough to create and be affected by inertia and turbulence effects in the flowing fluid. Such effects would cause the flux–gradient curve to depart from proportionality by bending toward the gradient axis, meaning that the flux would increase less than proportionally with gradient. But unless the soil particles were unusually large (e.g., gravel or very coarse sand), it was generally assumed that such less-than-proportional effects would seldom be encountered.

As steadily increasing attention was being focused on Darcy's equation in soil physics and hydrology up through the first half of the twentieth century, there seemed to have been but little inhibition against accepting an extension of Darcian validity from sand to soil. During the decades of 1952–1972, however, several extensive sets of experimental data were published that did not adhere to Darcian proportionality. Moreover, these flux–gradient departures from proportionality were the very opposite of the less-than-proportional behavior at large gradients as just discussed, and occurred in clays and clay-bearing porous sandstones. As shown in Figure 2 for such a sandstone, the experimental points of q versus i curved upward in a more-than-proportional manner, as described rather precisely by:

$$q = B\{i - J[1 - exp\,(-Ci)]\} \qquad [4]$$

where B, J, and C are the characterizing constants. As i becomes large, the term exp $(-Ci)$ vanishes and eqn [4] reduces to the linear form:

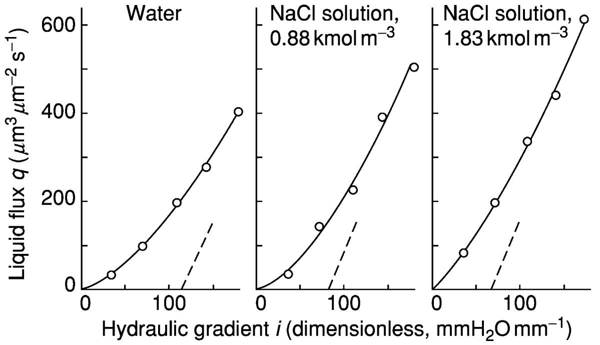


Figure 2 Flux–gradient values and relationships for clay-bearing sandstone sample No. 15L of Von Engelhardt and Tunn, published in 1954/1955. Each solid-line curve is eqn [4] fitted to its data set, with each short, broken straight line being the initial portion of the linear asymptote, eqn [5], for each solid-line curve.

$$q = B\{i - J\} \qquad [5]$$

which is the linear asymptote approached at large i by the generally curvilinear eqn [4], where J is the i-intercept when the linear asymptote is extended back to the i-axis. The broken straight lines in Figure 2 are the linear asymptotes in the vicinity of the i-intercepts (or J-values). Note also that the flow behavior at large gradient, although linear, is still not Darcian, because eqn [5] does not pass through the origin as does eqn [1].

If $J = 0$ or $C = 0$, then eqn [4] reduces to $q = Bi$, which is simply Darcy's equation with B replacing K in eqn [1]. Hence, B has the character of hydraulic conductivity and could be partitioned in the manner of eqn [3] if desired, using the constant b to replace k. Again, b would have the character of permeability, and would equal k for Darcian flow ($J = 0$ or $C = 0$). A measure of non-Darcian behavior, N_d, for J and C generally, has been defined by:

$$N_d = J^2C \qquad [6]$$

in which $N_d = 0$ for Darcian behavior ($J = 0$ or $C = 0$). The values of N_d for the three subgraphs in Figure 2 are 104.6, 73.3, and 35.8 mmH$_2$O mm^{-1}, in progressing from water (zero electrolyte) through the two successively higher concentrations of salt (NaCl). Clearly, N_d is maximal for water alone and decreases steadily with increasing salt concentration, although Darcian behavior is not completely restored ($N_d = 0$) even at the largest concentration of salt.

The most extreme expression of non-Darcian behavior would occur if there were an initial gradient range over which $q = 0$, starting at the origin and extending to some finite, nonzero value i_0 called the threshold gradient, after which $q > 0$ for $i > i_0$. This would imply the soil water initially to behave like a solid until the threshold gradient were exceeded. This drastic idea has not been well received by soil physicists and soil hydrologists. An i_0 value as large as 65.2 mmH$_2$O mm^{-1} has been reported for water in a sodium-saturated montmorillonite clay. Note that i_0 must not be confused with the gradient intercept J of eqn [5].

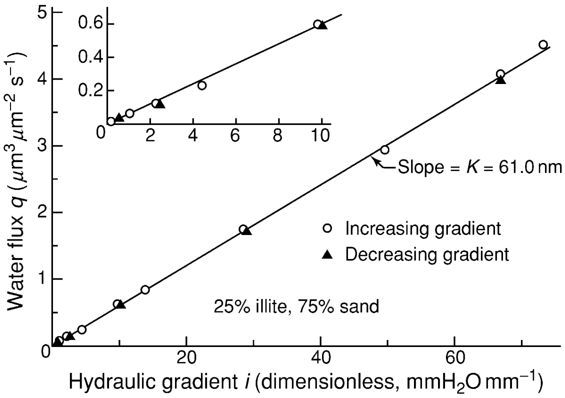


Figure 3 Experimental values of water flux versus hydraulic gradient for a mixture of one part illite in three parts fine quartz sand. Reproduced from Russell DA and Swartzendruber D (1971) *Soil Science Society of America Proceedings* 35: 25 with permission.

To account for the curvature of the graphs of Figure 2, the original postulate was that the pore water in the vicinity of the clay surfaces was changed from a Newtonian to a non-Newtonian viscosity by forces emanating from these clay surfaces. This would also have explained the effect of salt, inasmuch as increasing salt concentration is known to break down the water structure.

To obtain further and refined experimental data, a special technique, with highly sensitive flowmeters and differential manometers, yielded very rapid, simultaneous measurements of q and Δh (and hence i) for mixtures of fine quartz sand, quartz silt, kaolinite, illite, and montmorillonite. For the various mixtures of sand, silt, and kaolinite or illite, the results are typified in Figure 3, which is unequivocally Darcian for both the main graph and the small, expanded-scale inset graph. Also, the decreasing-gradient points, obtained immediately after those for increasing gradient, replicated very well. Neither is there any hint of a threshold gradient, nor of curvature in the flux–gradient relationship.

In contrast, however, if the mixture of sand, silt, and clay is 5% montmorillonite, the results are rather different, as displayed in Figure 4. With generally good agreement between increasing and decreasing gradients, the non-Darcian behavior of a sigmoidal character is obviously more complex than the monotonic curves of Figure 2; the modified-water postulate

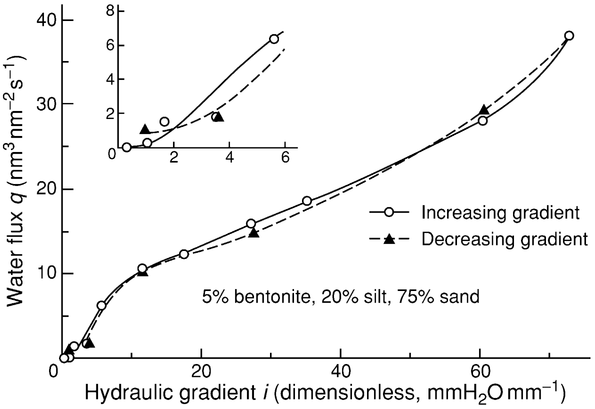


Figure 4 Experimental values of water flux versus hydraulic gradient for a mixture of 5% montmorillonite (bentonite) with 20% quartz silt and 75% fine quartz sand. Reproduced from Russell DA and Swartzendruber D (1971) *Soil Science Society of America Proceedings* 35: 25 with permission.

could therefore not apply. Lastly, in the inset graph (Figure 4), the first point is suggestive of a threshold gradient.

The refined experimentation indicates Darcy's equation to be applicable with confidence to sands, silts, nonswelling clays, and mixtures of these, but that Darcian behavior in the presence of swelling clays is much more problematic. Admittedly, a straight line through the origin could be drawn through the experimental points of Figure 4 as a first approximation, but doing this would seem less applicable to the data in Figure 2. Ultimately, however, in a pragmatic or field setting, the determining factor would become the error attendant to measuring K. If such error is much greater than the errors associated with approximating the data (Figures 2 and 4) with proportional lines, then it would be unreasonable to deny the use of Darcy's equation in such a case. Conversely, nonetheless, the soil physicist or hydrologist should not lose sight of the non-Darcian manifestations observed in swelling clays.

Solving Saturated-Flow Problems

To describe flow in the three cartesian dimensions, the components of water flux in the mutually perpendicular x, y, and z directions are employed. For the x direction:

$$q_x = -K\partial h/\partial x \qquad [7]$$

with K considered constant. The equations for q_y and q_z are just like eqn [7], but with x replaced by y and z, respectively. The use of the same K for all three dimensions (x, y, and z) means that the soil is assumed to be isotropic, in distinction from the less-simple anisotropic case, wherein K is not the same for the three dimensions.

The conservation of mass of an incompressible fluid, also called the equation of continuity, is:

$$\frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z} = -\frac{\partial f}{\partial t} \qquad [8]$$

where f is the total porosity of the soil. Taking the soil to be incompressible yields $\partial f/\partial t = 0$, and introducing the q_x from eqn [7] along with the similar equations for q_y and q_z into eqn [8] yields finally:

$$\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} + \frac{\partial^2 h}{\partial z^2} = 0 \qquad [9]$$

which is the classic Laplace equation, for which the shortened 'del squared' operator notation, $\nabla^2(\bullet)$, is used to write eqn [9] as:

$$\nabla^2 h = 0 \qquad [10]$$

There are an infinite number of solutions to Laplace's equation, but only when a given solution also satisfies the boundary conditions can it be claimed as the unique solution to the problem at hand. Also, the particular form of $\nabla^2(\bullet)$ depends on the coordinate system.

The simple, proportional form of Darcy's equation takes its place alongside Ohm's law of electrical flow, the Newton–Fourier heat-flux equation, and Ficks' first law of diffusion. Solutions from these other fields are thus available for scrutiny and potential use or modification in saturated water-flow problems. Historically, this has been an advantage and motivation for preferring Darcy's equation. Clearly, if one attempted to use eqn [4] in eqn [8], the result would be far more complicated and difficult to use than eqn [9] and would probably still remain difficult even with modern computer capabilities.

Water in Unsaturated Soils

When the total porosity f of the soil is no longer completely occupied with water, the soil is said to be unsaturated. This means that part of the pore space is then occupied by air and, consequently, that liquid water flow can only take place through the remaining porosity that contains liquid water. We therefore expect the K of the unsaturated soil to be less than the saturated K. Defining θ to be the volumetric water content of the soil, we formulate this first aspect of unsaturation by taking K to be a function of θ alone, $K = K(\theta)$, rather than a constant as in eqn [7]. Note that if $\theta = f$, $K = K(f)$ is simply the saturated K once again.

The second aspect of unsaturation is that the reduction of θ from f creates an attraction of the unsaturated soil for water, akin to a blotter that 'sucks up' ink. This means that the hydraulic head h is also affected by θ being less than f. Therefore, we take $h = z - \tau(\theta)$, where the position coordinate z is called the elevation head as measured from the datum plane placed at the xy plane, and $\tau = \tau(\theta)$ is called the soil-water suction-head function. Because $\tau(\theta)$ is different for wetting of soil than for drying of soil – a phenomenon known as hysteresis – the use of a given $\tau(\theta)$ must be restricted to monotonic changes in water content. In principle, $K(\theta)$ may also be envisaged as subject to hysteresis, but its effect is usually negligible in practice. Finally, the functions $K(\theta)$ and $\tau(\theta)$ have been called the Buckingham functions in honor of their originator. Buckingham's discussion and explanations of almost a century ago, which we have followed here, are so instructive and apt that they can scarcely be improved upon today, and thus stand as a remarkable and striking example of prescience.

Incorporation of $K = K(\theta)$ and $h = z - \tau(\theta)$ into eqn [7] yields:

$$q_x = K(\theta)\frac{\partial\tau}{\partial x} \qquad [11]$$

where $\partial\tau/\partial x$ is the suction-head gradient in the x direction. Expanding $\partial\tau/\partial x$ by making use of $\tau(\theta)$, eqn [11] can be written for homogeneous soil as:

$$q_x = -K(\theta)\left[-\frac{d\tau}{d\theta}\right]\frac{\partial\theta}{\partial x} \qquad [12]$$

Now $d\tau/d\theta$ is also a function of θ, since τ is a function of θ. Therefore, since $K(\theta)$ is obviously a function of θ, the composite multiplier of $\partial\theta/\partial x$ in eqn [12] is also a function only of θ, which we label $D(\theta)$, setting $D(\theta) = K(\theta)[-d\tau/d\theta]$, so that eqn [12] becomes:

$$q_x = -D(\theta)\frac{\partial\theta}{\partial x} \qquad [13]$$

$D(\theta)$ is called the soil-water diffusivity function, and eqn [13] expresses the horizontal water flux in terms of a response to a water-content gradient, $-\partial\theta/\partial x$. The physical process, however, is not one of diffusion, since nothing has been done to alter the hydrodynamic basis of eqn [7], from which the derivation began. With the same analysis conducted for q_y, the results will be the same at every stage corresponding to eqns [11–13], except that y replaces x. For q_z, however, the results are different, because $\partial z/\partial z = 1$ and not zero. The counterpart of eqn [11] is:

$$q_z = K(\theta)\frac{\partial\tau}{\partial z} - K(\theta) \qquad [14]$$

and the counterpart of eqn [12] also has $K(\theta)$ subtracted from the right-hand side. The counterpart of eqn [13] is:

$$q_z = -D(\theta)\frac{\partial\theta}{\partial z} - K(\theta) \qquad [15]$$

Eqns [11–15] are all forms of the Buckingham–Darcy flux-gradient equation, which fills the same role for unsaturated soil as does Darcy's equation for saturated soil. For horizontal flux, only one soil-characterizing function is needed, either $K(\theta)$ of eqn [11] or $D(\theta)$ of eqn [13]. For vertical flux, the single soil-characterizing function $K(\theta)$ employed in eqn [14] is sufficient, whereas if $D(\theta)$ is employed in eqn [15] it is still necessary to use $K(\theta)$ as well. Experimentally, however, it may be easier to measure θ and its gradient than τ and its gradient, especially at low values of θ, when τ values are very large.

Early experimental tests that successfully validated the Buckingham–Darcy equation were carried out, ironically enough, on sandy soils or materials, reminiscent of Darcy-equation verification on sands rather than soils. In the early 1960s, however, some very precise gamma-ray measurements of horizontally transient θ values were reported for a nonswelling silty clay loam. These data were amenable for a test of eqn [13]. To visualize this test, note in eqn [13] that, at a fixed value of θ, D is also fixed, so that a plot of q_x versus $(-\partial\theta/\partial x)$ is asserted to be a straight line through the origin with a slope equal to the fixed D. But, instead of just a single proportional line as in the saturated case illustrated in **Figure 3**, there will now be a family of proportional lines that constitute eqn [13], one proportional line for each value of fixed $D(\theta)$. The steepest slope will be for the largest (wettest) θ, with slopes becoming progressively less steep as the fixed θ is closer to air dryness. For the first three largest water contents in the silty clay loam, $\theta = 0.45$, 0.40, and 0.35 $\text{mm}^3\,\text{mm}^{-3}$, the experimental points fell on straight lines through the origin and with progressively decreasing slope, entirely in accord with eqn [13]. At $\theta = 0.30\,\text{mm}^3\,\text{mm}^{-3}$, however, the experimental points curved upward in more-than-proportional fashion similar to the curves of **Figure 2**, and this behavior accentuated progressively as θ decreased to $0.15\,\text{mm}^3\,\text{mm}^{-3}$. Hence, eqn [13] was obeyed for the wetter but not the drier regions of the soil.

An alternative explanation for these deviations was developed from the time-dependent premise that $K = K(\theta, t)$, $\tau = \tau(\theta, t)$, and $D = D(\theta, t)$, rather than $K = K(\theta)$, $\tau = \tau(\theta)$, and $D(\theta)$, as used here in deriving eqns [11–15]. This makes it less compelling to invoke failure of the Buckingham–Darcy equation. Admittedly also, precise experiments in unsaturated flow are even more difficult than in saturated flow. Once again, then, general errors of measurement may sufficiently exceed those due to departures from the Buckingham–Darcy equation, so that its use in this pragmatic or practical sense is still permissible.

Solving Unsaturated-Flow Problems

The cartesian equation of continuity for unsaturated soil conditions is the same as eqn [8] except that the right-hand side is changed to $-\partial\theta/\partial t$. Using q_x from eqn [13], $q_y = -D(\theta)\partial\theta/\partial y$, and q_z from eqn [15], eqn [8] becomes:

$$\frac{\partial}{\partial x}\left[D(\theta)\frac{\partial\theta}{\partial x}\right] + \frac{\partial}{\partial y}\left[D(\theta)\frac{\partial\theta}{\partial y}\right] + \frac{\partial}{\partial z}\left[D(\theta)\frac{\partial\theta}{\partial z}\right] + \frac{\partial K(\theta)}{\partial z} = \frac{\partial\theta}{\partial t} \qquad [16]$$

An alternative form emerges if we note that, since $\tau = \tau(\theta)$, then conversely $\theta = \theta(\tau)$, whereupon $K(\theta) = K[\theta(\tau)]$ so that we can replace $K(\theta)$ with $K(\tau)$. Also, differentiation of $\theta = \theta(\tau)$ yields $\partial\theta/\partial t = (\mathrm{d}\theta/\mathrm{d}\tau)(\partial\tau/\partial t)$, and $(\mathrm{d}\theta/\mathrm{d}\tau)$ is a function of τ because of $\theta = \theta(\tau)$. Making use of these relationships along with q_x from eqn [11], $q_y = K(\tau)\partial\tau/\partial y$, and q_z from eqn [14], eqn [8] becomes:

$$\frac{\partial}{\partial x}\left[K(\tau)\frac{\partial\tau}{\partial x}\right] + \frac{\partial}{\partial y}\left[K(\tau)\frac{\partial\tau}{\partial y}\right] + \frac{\partial}{\partial z}\left[K(\tau)\frac{\partial\tau}{\partial z}\right] - \frac{\partial K(\tau)}{\partial z} = \frac{\mathrm{d}\theta}{\mathrm{d}\tau}\frac{\partial\tau}{\partial t} \qquad [17]$$

As before, the use of a single $K(\tau)$ in eqn [17] or of a single $D(\theta)$ in eqn [16] means that the unsaturated soil is taken to be isotropic in its water-transport properties.

Eqns [16] and [17] are both forms of the Richards equation, so named in honor of the originator of eqn [17]. Including eqn [16] is appropriate, however, because Richards explicitly mentioned that the choice of either $\tau = \tau(\theta)$ or $\theta = \theta(\tau)$ was a matter of mathematical expediency. The Richards equation, along with the necessary boundary and initial conditions, fills the same role as a problem-solving framework for unsaturated flow as does the Laplace equation (eqn [9]) for saturated flow. But the solutions to Richards' equation are much more difficult – recourse to numerical and computer solutions has been needed almost from the beginning.

List of Technical Nomenclature

η Absolute viscosity of flowing water or fluid (eqn [3]) (mN s m^{-2} = g m^{-1} s^{-1})

θ Volume of water in a soil sample divided by the sample bulk volume, 'volumetric water content' (mm^3 mm^{-3} = 1)

π 3.1416 (radian)

ρ Mass density of flowing water or fluid (kg m^{-3})

$\tau(\theta)$ The suction head created by the attraction of unsaturated soil for water and which is a function of water content θ (mmH$_2$O)

$\nabla^2(\bullet)$ Cartesian Laplacian operator (left-hand side of eqn [9]) (mmH$_2$O mm^{-2})

A Bulk cross-sectional area of soil column (mm^2)

a Angle of soil column orientation (Figure 1) (radian)

B Non-Darcian hydraulic conductivity (eqn [4]) (μm s^{-1})

b Non-Darcian permeability of a soil (μm^2)

C Non-Darcian parameter (eqn [4]) (mm mmH$_2$O^{-1})

$D(\theta)$ Soil-water diffusivity function (eqns [13] and [15]) (mm^2 s^{-1})

f Soil porosity: volume of pores in a sample of soil divided by the bulk volume of the sample (mm^3 mm^{-3} = 1)

g Acceleration due to gravity (m s^{-2})

h Height of water column sustained by the soil water at any point in the soil, measured above an arbitrary datum plane and called hydraulic head (mmH$_2$O)

h_1, h_2 Inlet and outlet hydraulic heads, respectively, at the ends of the soil column (Figure 1) (mmH$_2$O)

i Reduction in hydraulic head per unit distance along the path of water flow, e.g., $-(h_2 - h_1)/L$, called 'hydraulic gradient' (mmH$_2$O mm^{-1})

i_0 Nonzero hydraulic gradient at and below which no detectable water flux (flow) occurs (mmH$_2$O mm^{-1})

J Non-Darcian parameter (eqn [4]) (mmH$_2$O mm^{-1})

K Darcian hydraulic conductivity (eqn [1]) (μm s^{-1})

k Darcian permeability of a soil (eqn [3]) (μm^2)

$K(\theta)$ Darcian hydraulic conductivity of an unsaturated soil as a function of water content (μm s^{-1})

L Bulk length of the soil column (Figure 1) (mm)

N_d A measure of non-Darcian behavior (eqn [6]) (mmH$_2$O mm^{-1})

Q Volume of water (or fluid) flowing per unit time through the soil column (Figure 1) (mm^3 s^{-1})

q Volume of water (or fluid) flowing per unit time through unit bulk cross-sectional area, Q/A, called water (or fluid) flux (mm^3 mm^{-2} s^{-1} = mm s^{-1})

q_x, q_y, q_z Components of water (or fluid) flux in the three cartesian directions x, y, and z, respectively (mm s^{-1})

t Time after the beginning of water or fluid flow (s)

x, y, z Cartesian position coordinates, respectively (mm)

See also: **Capillarity**; **Evaporation of Water from Bare Soil**; **Infiltration**; **Isotropy and Anisotropy**; **Porosity and Pore-Size Distribution**; **Thermodynamics of Soil Water**; **Vadose Zone:** Hydrologic Processes; **Water Potential**; **Water Retention and Characteristic Curve**; **Water, Properties**

Further Reading

Buckingham E (1907) *Studies on the Movement of Soil Moisture. US Department of Agriculture Bureau of Soils Bulletin 38*. Washington, DC: Government Publishing Office.

Darcy H (1856) *Les Fontaines Publique de la Ville de Dijon*. Paris, France: Victor Dalmont.

Guerrini IA and Swartzendruber D (1992) Soil water diffusivity as explicitly dependent on both time and water content. *Soil Science Society of America Journal* 56: 335–340.

Kutilek M (1965) Influence de l'interface sur la filtration de l'eau dans les sols. *Science du Sol* 1: 3–14.

Lutz JF and Kemper WD (1959) Intrinsic permeability of clay as affected by clay–water interaction. *Soil Science* 88: 83–90.

Miller RJ and Low PF (1963) Threshold gradient for water flow in clay systems. *Soil Science Society of America Proceedings* 27: 605–609.

Olson TC and Swartzendruber D (1968) Velocity–gradient relationships for steady-state unsaturated flow of water in nonswelling artificial soils. *Soil Science Society of America Proceedings* 32: 457–462.

Philip JR (1995) Desperately seeking Darcy in Dijon. *Soil Science Society of America Journal* 59: 319–324.

Richards LA (1931) Capillary conduction of liquids through porous mediums. *Physics (NY)* 1: 318–333.

Russell DA and Swartzendruber D (1971) Flux–gradient relationships for saturated flow of water through mixtures of sand, silt, and clay. *Soil Science Society of America Proceedings* 35: 21–26.

Swartzendruber D (1962a) Modification of Darcy's law for the flow of water in soils. *Soil Science* 93: 22–29.

Swartzendruber D (1962b) Non-Darcy flow behavior in liquid-saturated porous media. *Journal of Geophysical Research* 67: 5205–5213.

Swartzendruber D (1963) Non-Darcy behavior and the flow of water in unsaturated soils. *Soil Science Society of America Proceedings* 27: 491–495.

Swartzendruber D (1968) The applicability of Darcy's law. *Soil Science Society of America Proceedings* 32: 11–18.

Swartzendruber D (1969) The flow of water in unsaturated soils. In: de Wiest RJM (ed.) *Flow Through Porous Media*, pp. 215–292. New York: Academic Press.

Von Engelhardt W and Tunn WLM (1955) *The Flow of Fluids Through Sandstones* (translated by Witherspoon PA from *Heidelberger Beträge zur Mineralogie und Petrographie* 2: 2–25, 1954). *Illinois State Geological Survey* Circular 194.

DEGRADATION

C J Ritsema, Alterra, Wageningen, The Netherlands
G W J van Lynden, ISRIC, Wageningen, The Netherlands
V G Jetten and S M de Jong, Utrecht University, Utrecht, The Netherlands

Introduction

Soil is under increasing threat from a wide range of human activities that are undermining its long-term availability and viability. One third of the world's agricultural soils, or approximately 2 billion hectares of land are affected by soil degradation. Water and wind erosion account for most of the observed damage, while other forms such as physical and chemical degradation are responsible for the rest. Appropriate soil and water conservation strategies are needed to prevent and combat the effects of soil degradation in the field and at the planning level.

Soil degradation is "a process that describes human-induced phenomena which lower the current and/or future capacity of the soil to support human life." In a general sense, soil degradation could be described as the deterioration of soil quality, or in other words: the partial or entire loss of one or more functions of the soil. Quality should be assessed in terms of the different potential functions of the soil.

Land degradation is the reduction in the capability of the land to produce benefits from a particular land use under a specified form of land management. Seven main groups of land-degradation processes are normally distinguished: (1) mass movement (such as debris flows and avalanches), (2) water erosion (sheet, rill, gully erosion), (3) wind erosion, (4) excess of salts (salinization, sodification), (5) chemical degradation (acidification, contamination, toxicity), (6) physical degradation (crusting, compaction, oxidation), and (7) biological degradation (loss of soil biodiversity).

An important aspect of many soil and land degradation processes are the so called off-site effects; for example, dust storms or eroded sediment cause problems such as damage by mudflows, siltation of dams, or pollution of drinking water in downwind or downstream areas.

Factors and Processes Affecting Degradation of Soils

Various types of human activities may lead to soil degradation. Although some degradation processes may also occur naturally, many degradation types are the result of human disturbance of either a natural or anthropogenic state of equilibrium. Some of these are:

Agricultural causes: Defined as the improper management of cultivated arable land. It includes a wide variety of practices, such as insufficient or excessive use of fertilizers, shortening of the fallow period in shifting cultivation, use of poor quality irrigation water, absence or bad maintenance of erosion-control measures, improper use of heavy machinery, etc. Degradation types commonly linked to this causative factor are erosion (water or wind), compaction, loss of nutrients, salinization, and pollution (by pesticides or fertilizers).

Deforestation or removal of natural vegetation: Defined as the near complete removal of natural vegetation (usually primary or secondary forest) from large stretches of land, for example by converting forest into agricultural land (hence sometimes followed by agricultural mismanagement), large-scale commercial forestry, road construction, urban development, etc. Deforestation often causes erosion and loss of nutrients.

Overexploitation of vegetation for domestic use: Contrary to 'deforestation or removal of natural vegetation,' this causative factor does not necessarily involve the (near) complete removal of the 'natural' vegetation, but rather a degeneration of the remaining vegetation, thus offering insufficient protection against erosion. It includes activities such as excessive gathering of fuel wood, fodder, (local) timber, etc.

Overexploitation of natural water resources: This leads to water shortages for the natural ecosystem and in the long term to the removal of the natural vegetation cover. The result is an increased vulnerability of the land for surface runoff, soil erosion, and soil surface crusting. As soon as the process of vegetation deterioration starts, it normally has a self-enhancing effect which is difficult to stop or to reverse.

Overgrazing: Besides actual overgrazing of the vegetation by livestock, other phenomena of excessive livestock amounts are also considered here, such as trampling. The effect of overgrazing usually is soil compaction and/or a decrease in plant cover, both of which may in turn give rise to water or wind erosion.

Industrial activities: All human activities of an industrial or bioindustrial nature are included: industries, power generation, infrastructure and urbanization, waste handling, traffic, etc. It is most often linked to pollution of different kinds (either point source or diffuse) and loss of productive function.

Types of Soil Degradation

The type of soil degradation refers to the nature of the degradation process. Soil particles may be displaced by the action of water or wind (erosion and sedimentation), which may cause damage to crops, infrastructure, buildings, and the environment in general. Erosion can be linear, i.e., concentrated along certain channels (rill or gully erosion and mass wasting such as landslides), sometimes creating very deep scars in the landscape (Figure 1). Less conspicuous, but often even more detrimental to crops is the gradual removal of the topsoil layer (sheet erosion). Off-site effects of erosion may consist of siltation of reservoirs and river beds and/or flooding, or dune formation and 'overblowing' in the case of wind erosion. Degradation *in situ*, i.e., without movement of soil particles, can be chemical (soil pollution by chemical wastes or excessive fertilization; fertility decline due to nutrients being removed by harvesting, erosion and leaching; salinization due to irrigation with saline groundwater and/or without proper drainage in semiarid and arid areas, acidification due to pH-lowering additions to the soil from fertilizers or from the atmosphere), or physical (compaction due to the use of heavy machinery; deteriorating soil structure such as crusting of the soil surface; waterlogging due to increased water table or its opposite, aridification).

Assessment of Degradation

Approaches

The status of soil degradation can be assessed in a qualitatively broad manner or in a more detailed quantitative manner. The former generic approach is better suited for small-scale assessments, such as for entire countries, continents, or global overviews. A quantitative approach is required for more specific and detailed assessments, e.g., to determine the erosion status for a watershed or the pollution status for a province. Qualitative assessments are based on expert judgement and hence more liable to subjectivity than quantitative methods. A method does not have to be fully qualitative or quantitative, mixtures may occur. Some frequently used methods or tools are:

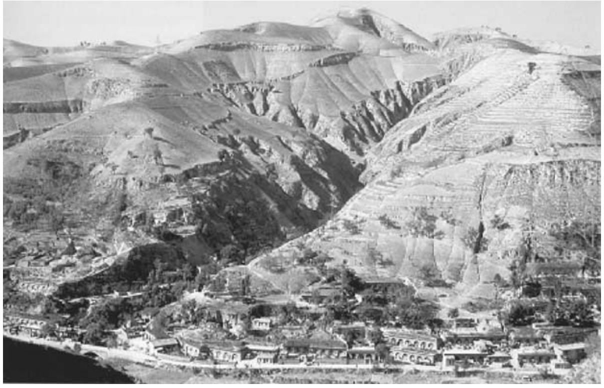


Figure 1 Severely degraded soils on the Loess Plateau of China.

1. Expert opinion: Qualitative assessment on a controlled mapping base and semiquantitative definitions, as employed for instance in the Global Assessment of Human-induced Soil Degradation (GLASOD) survey. GLASOD and related methods are based on an assessment of land suitability by national experts that use defined, semiquantitative class limits on a given mapping base. Its major disadvantage is the inevitable degree of subjectivity. Its major advantage is its capacity to produce results, such as achieving complete world coverage (Figure 2), in a short time and on a small budget. Costs per unit area are relatively low. In Figure 3 an integrated global soil degradation severity map is shown, indicating areas with different degradation rates;
2. Remote sensing: Analysis of low- and high-resolution satellite data and airborne imagery (e.g., analysis of composite indices such as the Normalized Difference Vegetation Index (NDVI)). Remote sensing always includes linkages with ground observations. The basis of this method is comparison of remotely sensed imagery of different dates, for regional coverage, mainly low-resolution imagery; and, specifically, comparison of the NDVI, derived from imagery collected by the sensor aboard the National Oceanographic and Atmospheric Administration (NOAA) satellite, and more detailed imagery. This method was tested amongst others in Saudi Arabia and shows areas where vegetation response to rainfall is decreasing (degradation of resources) or increasing (rehabilitation of resources). It has been applied particularly to early warning systems. For longer-term comparisons, some form of calibration for preceding rainfall is needed. Costs are relatively low. It is recognized that remote sensing cannot be used alone.

Spectral mixture analysis (SMA): Since 1985 hyperspectral remote sensing has been developed, opening new methods to survey and assess degradational state of the soil surface. Hyperspectral remote sensing refers to the collection of images in the solar spectrum, with many narrow spectral bands allowing the collection of very accurate spectra of objects and the earth surface and identification of absorption features of plants and of soil minerals in these spectra. SMA is a technique to unravel the spectral information in the remote-sensing images by assuming that the spectral variation is caused by a limited number of surface material (green vegetation, senescent vegetation, a

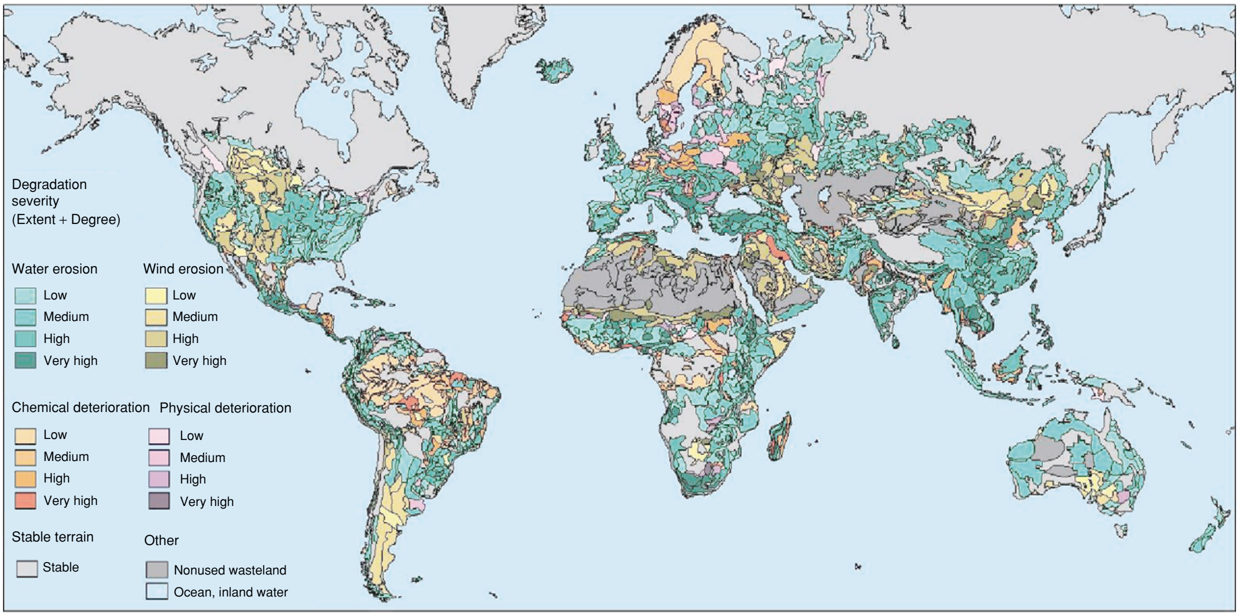


Figure 2 Global assessment (in 1990) of the status of human-induced soil degradation. (Reproduced with permission from Oldeman LR, Sombroek WG, and Hakkeling R (1991) *World Map on the Current Status of Human-Induced Soil Degradation. An Explanatory Note*, 2nd edn. Wageningen, the Netherlands: ISRIC/Nairobi, Kenya: UNEP.)

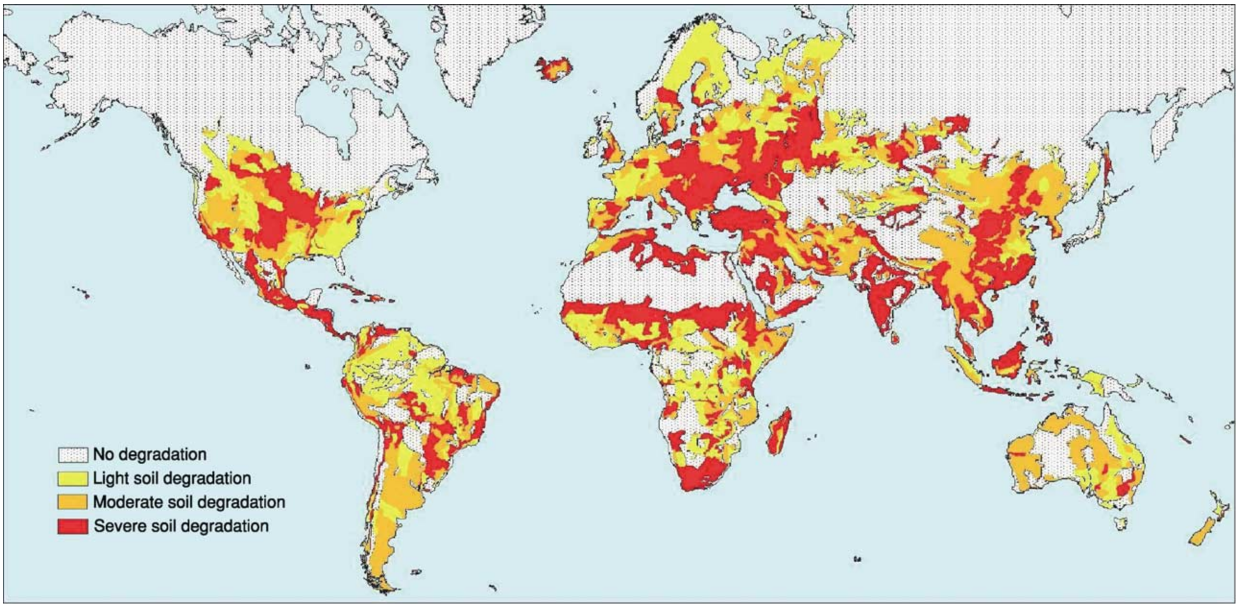


Figure 3 Global soil degradation severity map as produced by the GLASOD initiative.

number of soil types, and water). A reference library of these surface materials collected in the field or in the laboratory yields the basis for SMA of the remote-sensing images. This approach has been applied successfully in a number of case studies to survey soil conditions and to identify classes of degradation. The SMA approach normally improves on results using the NDVI but requires more spectral bands: SMA is successfully applied to separate, in images, bare soil surfaces from senescent vegetation and yellow vegetation from green vegetation. These three factors are important inputs in soil-erosion models because they act differently with respect to raindrop interception.

3. Field monitoring: Stratified soil sampling and analysis, and field observation of vegetation and biodiversity under certain land-use or management practices and climate variability. To date, soil monitoring has been applied mainly in developed countries, and tests are needed of its cost-effectiveness in developing countries. In areas where baseline studies have

been established, monitoring of changes will be undertaken; in other areas, establishment of a baseline will be a priority. Stratified soil-sampling with analysis, and/or benchmark sites, repeated over 5- to 10-year intervals, has been advocated as a basic activity for national soil survey organizations. Examples of application to date (2003) are few, but successful: the method has been applied to 20 000 sites over a 25-year period in Japan; is currently being used for a national 16-km-grid in France; and has been started in Denmark and Switzerland. The same approach has been applied to field observations of vegetation, along transects or in sampling plots, and to biodiversity. Costs per unit area are relatively high, but could be reduced by application to priority areas only, on a stratified sampling basis.

4. Productivity changes: Observation of changes in crop yields, biomass production, and livestock output, which directly apply to the definition of land degradation in terms of lowered productivity, although they are influenced by many other factors. There is a range of possibilities: At national level, use might be made of national yield statistics (of which the reliability is still under debate), adjusted for fertilizer use and climate. At local level, yield monitoring is possible by comparisons with a standard crop, either without fertilizer or with standard fertilizer and management. Substantial problems arise in that productivity decline could be due to factors other than land degradation, e.g., removal of fertilizer subsidy or civil strife. The same cost constraints apply as for soil monitoring.

5. Sample studies at farm level, based on field criteria and the expert opinion of land users. Even at national level, such detailed studies are essential on a sample basis, to obtain grass roots views both of the severity of degradation and its causes, together with practicable remedies (Stocking and Murnaghan, 2001). Field indicators of soil degradation were developed about 20 years ago, and could be extended to condition of vegetation. Talking with farmers means getting the views of farmers, and other land users, on whether things have got worse – which are of course, subjective and perhaps systematically biased, but still essential to get grass roots view at local level. The method is clearly applicable only at a local scale, and thus on a selective sampling basis. Observations of the state of the land can be combined with assessment of driving factors and impacts.

6. Modeling: Based on data obtained by other methods, modeling can be used in many ways, such as: (1) prediction of degradation hazard; (2) operational definition of degradation in terms of unfavorable changes in plant productivity, soil properties, and hydrology; (3) design of conservation measures using climatic data with a specific return period (worst-case scenario modeling); (4) extending the range of applicability of results; (5) integrating biophysical with socioeconomic factors. Much research has been put into devising models for the prediction of soil-erosion hazard. There are established methods for the modeling of both water and wind erosion, which have been widely applied, in part because it is vastly cheaper than any form of field observation. The modeling approach is mainly relevant to degradation hazard, but can be applied to actual degradation first, as a means of calibration of the model to the specific requirements of an area, optimizing sampling design, or to extrapolate the applicability of results obtained on a sampling basis. Risk reflects a potential development in the future, while status reflects the development to date. Models vary widely in complexity and data requirements, depending on the type of degradation they are addressing and the size of the area under investigation. Models are useful to learn and understand degradation processes, but both the model and input data are a simplification of reality, hence extrapolation of models should be done with care. Very often models are developed for experimental plots or pilot zones of a restricted size and under more or less controlled conditions, which should be taken into account when applying the model elsewhere. The data requirements and structure of a model, and the type of processes included in it, depend on many things: (1) the temporal scale of the research objectives: Is an annual, daily, or event-based result required? (2) the spatial scale: Are predictions needed for a single plot, a field, complex spatial catchment, or an entire region? (3) Is the emphasis on the on-site effects of land degradation (e.g., soil erosion or crop yield changes) or on the off-site effects such as water sediment levels and pollution? Spatial and temporal scales are often linked, as, for example, is the case for physically based spatial erosion models (Figure 4) that simulate single events for first-order catchment with a high level of detail. They can be used to answer subtle questions about the effects of specific land-use changes or soil and water conservation measures in the catchment upon reducing runoff and erosion (Figure 5). On the other hand there are less-complex empirical models that can simulate continuous periods mostly for fields or hillslopes, but they can only show the change in annual erosion or soil loss.

Potentials and Limitations

It is useful to emphasize some potentials and limitations of land degradation assessments. It is obvious that an assessment at a small scale (e.g., 1:1 M) does not have a direct value for activities at the field level, but can be highly useful (if well done) to planners, government agencies, legislative bodies, educational institutions, nongovernment organizations (NGOs), and the general public in highlighting (potential) problem areas and decision-making for further action.

Besides geographic coverage and scale, another factor that determines the usefulness of an assessment methodology is the range of degradation issues the assessment tries to cover. Land degradation is a very broad issue, covering a wide range of degradation issues, which makes its assessment a rather generic or alternatively highly unwieldy exercise. It is already quite complicated to assess one specific type of soil pollution, not to mention the various other types of soil degradation, which in itself is just one aspect of land degradation. This also means that the frequently observed desire to have 'simple' assessment methods is not entirely realistic, if this is supposed to be anything more than just a general awareness-raising tool.

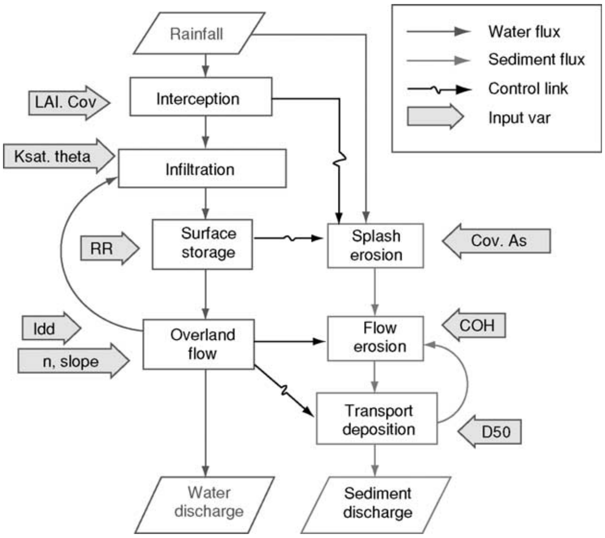


Figure 4 Model structure of a physically based spatial hydrologic and soil-erosion model, in which water and sediment are routed to the outlet of a catchment and produced as discharge. Input var, input variable; LAI, leaf area index; Cov, soil cover; Ksat, hydraulic conductivity; theta, moisture content; RR, surface roughness; Idd, runoff network; n, flow resistance; slope, terrain slope; As, aggregate stability; COH, Cohesion; D50, median grain size of suspended sediment.

Two types of assessments have been identified one is 'backward-looking' and determines the result of degradation over a recent past period. The other approach is forward-looking in the sense that it makes predictions for the future based on models and scenarios. Although the backward-looking approach considers the current status, it does not necessarily reflect the result of the degradation process over that period, but the net result of a number of acting and counteracting factors. Degradation is one of these, but remedial activities compensating the degradation effect to some extent is another one.

Though the wish to have 'simple' degradation assessment methods is often expressed, it should be realised that soil degradation is a complex process, determined by a range of factors of a natural and socioeconomic character. Hence a simple method will tend to correspond less with reality than a more complex and comprehensive one.

Degree and Impact of Degradation

Degree of Degradation

Degree is defined as the intensity of the soil degradation process, e.g., in the case of erosion, the amount

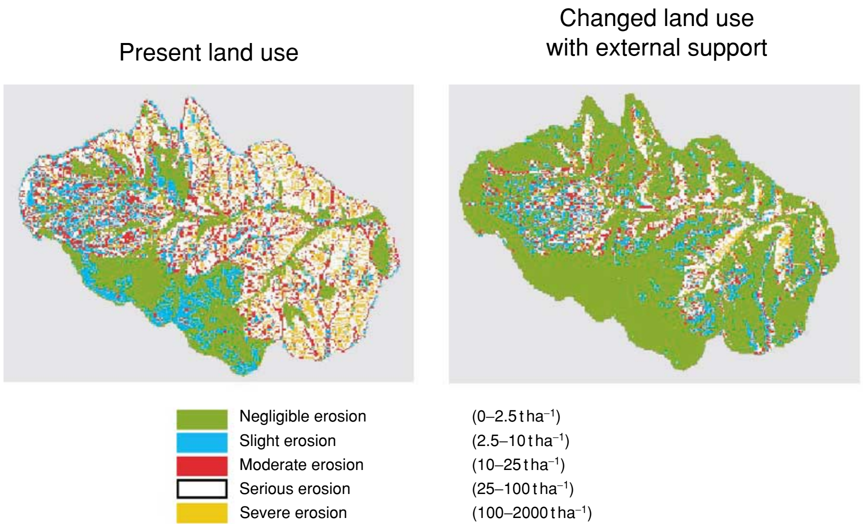

Present land use
Changed land use with external support
Negligible erosion (0–2.5 t ha⁻¹)
Slight erosion (2.5–10 t ha⁻¹)
Moderate erosion (10–25 t ha⁻¹)
Serious erosion (25–100 t ha⁻¹)
Severe erosion (100–2000 t ha⁻¹)

Figure 5 Computed soil losses in a first-order watershed for the current land use and management conditions and for an alternatively defined land-use distribution.

of soil washed or blown away. The FAO has proposed values for maximum acceptable limits of soil loss by erosion with respect to decreased agricultural productivity. Four classes are distinguished, ranging from no loss of productivity to severe loss of productivity. The classes are <12, 12–25, 25–50, and >50 t ha^{-1} year^{-1}. Relative changes of the soil properties are other good indicators of soil degradation: the percentage of the total topsoil lost, the percentage of total nutrients and organic matter lost, the relative decrease in soil moisture-holding capacity, changes in buffering capacity, etc. However, although such data may exist for experimental plots and pilot areas, precise and actual information is often lacking at a regional scale.

Rate of Soil Degradation

The recent rate of degradation relates to the rapidity of degradation over the past 5–10 years or, in other words, the trend of degradation. A severely degraded area may be quite stable at present (i.e., low rate, hence no trend toward further degradation), while other areas that are now only slightly degraded may show a high rate, hence a trend toward rapid further deterioration. From a purely physical point of view, the latter area would have a higher conservation priority than the former. Areas where the situation is improving (through soil conservation measures, for example), can also be identified. A comparison of the actual situation with that of the preceding decade may suffice, but often it is preferable to examine the average development over the last 5–10 years to level out irregularities. Whereas the degree of degradation only indicates the current, static situation (measured by decreased or increased productivity compared with some 10–15 years ago) the rate indicates the dynamic situation of soil degradation, namely the change in degree over time.

Impact of Degradation

Impact refers to the effects of soil degradation on the various soil functions. Changes in soil and terrain properties (e.g., loss of topsoil, development of rills and gullies, exposure of hardpans in the case of erosion) may reflect the occurrence and intensity of soil degradation but not necessarily the seriousness of its impact. Removal of a 5-cm layer of soil may have a greater impact on a poor shallow soil than on a deep fertile soil. The impact depends on the function and/or use of the soil: a heavily compacted soil is unsuitable for agriculture, but may be an appropriate basis for road construction.

Whereas the degree of degradation mainly refers to the degradation process, the impact of degradation can be manifold, depending on the current function (or use) of the soil. In many cases, the impact of degradation types will be on its biotic functions, or more specifically on its productivity. A significant complication in indicating productivity losses caused by soil degradation is the variety of reasons that may contribute to yield decline. Falling productivity may be caused by a wide range of factors such as erosion, fertility decline, improper management, drought, or waterlogging, quality of inputs (seeds, fertilizer), pests, and plagues, often in combination with each other. However, if one considers a medium- to long-term period (e.g., 25 years), large aberrations resulting from fluctuations in the weather pattern or pests should be leveled out.

The effects of soil degradation can be partially hidden by various management measures such as soil conservation, use of improved varieties, fertilizers, and pesticides. Some of these inputs are used to compensate for the productivity loss caused by soil degradation, for example application of fertilizers to compensate for lost nutrients. In other words, yields could have been much higher in the absence of soil degradation (and/or costs could have been reduced). Therefore, productivity changes should be seen in relation to the degree of input or level of management. The latter may include use of fertilizers, biocides, improved varieties, mechanization, various soil conservation measures, and other important changes in the farming system.

Changes in productivity should be expressed in relative terms, i.e., the current average productivity compared with the average productivity in the non-degraded situation and in relation to inputs. For instance, if previously an average yield of 2 t of wheat ha^{-1} was attained while at present only 1.5 t is realized in spite of high(er) inputs – and all other factors being equal – this would be an indication of strong soil degradation. Sometimes the impact may be ranked as negligible, even when degradation occurs, because of the capacity of the soil to resist a certain amount of degradation. Although for most degradation types the dominant impact is on productivity, some types (pollution in particular) may have additional or different impacts, e.g., on human or animal health or on entire ecosystems.

Preventing and Combating Degradation

There are a wide variety of measures to prevent or combat land degradation. These measures are generally known as soil conservation or soil and water conservation (SWC), especially when related to aspects like erosion, soil-moisture problems and soil fertility. More broadly applicable are names such as land husbandry or sustainable land management.

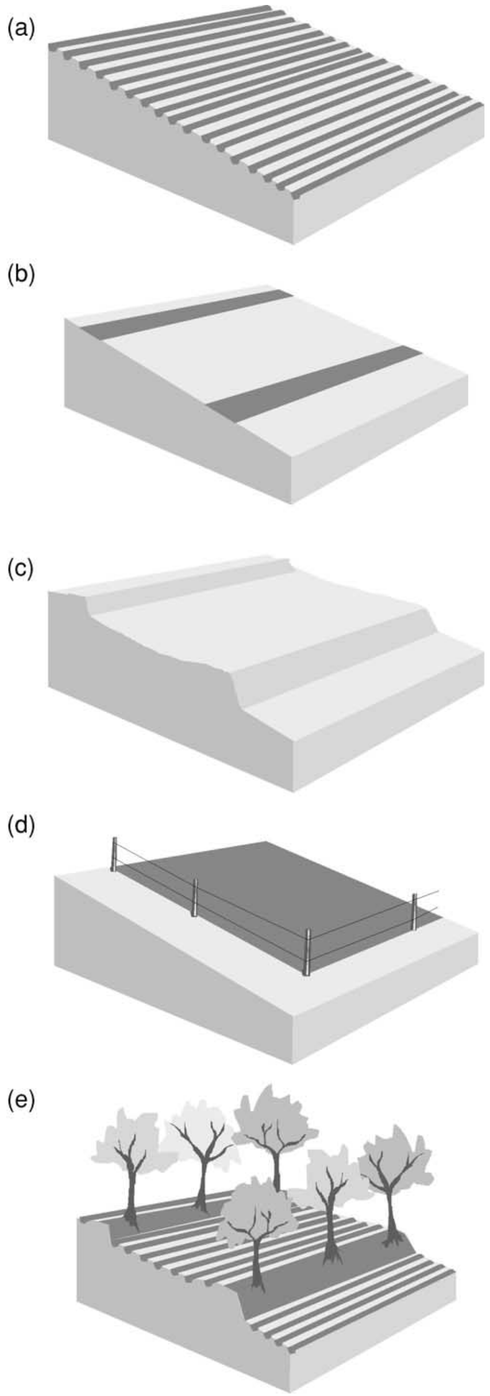


Figure 6 Categorization of soil and water conservation measures according to the World Overview of Conservation Approaches and Technologies (WOCAT) initiative: (a) agronomic measures such as mixed cropping, contour cultivation, mulching, etc. which: (1) are usually associated with annual crops, (2) are repeated routinely each season or in a rotational sequence, (3) are of short duration and not permanent, (4) do not lead to changes in slope profile, (5) are normally not zoned, (6) are normally independent of slope; (b) vegetative measures such as grass strips, hedge barriers, windbreaks, etc. which: (1) involve the use of perennial grasses, shrubs, or trees, (2) are of long duration, (3) often lead to a change in slope profile, (4) are often zoned on the contour or at right angles to wind, (5) direction, (6) are often spaced according to slope; (c) structural measures such as terraces, banks, bunds, constructions, palisades, etc. which: (1) often lead to a change in slope profile, (2) are of long duration or permanent, (3) are carried out primarily to control runoff, wind velocity, and erosion, (4) require substantial inputs of labour or money when first installed, (5) are often zoned on the contour/against wind direction, (6) are often spaced according to slope, (7) involve major earth movements and/or construction with wood, stone, concrete, etc.; (d) management measures such as land use change, area closure, rotational grazing, etc. which: (1) involve a fundamental change in land use, (2) involve no agronomic and structural measures, (3) often result in improved vegetative cover, (4) often reduce the intensity of use; (e) combinations in conditions where they are complementary and thus enhancing each other. Any combinations of the above measures are possible, e.g.: structural: terrace, with vegetative: grass and trees, with agronomic: ridges.

The WOCAT (World Overview of Conservation Approaches and Technologies) network, which constitutes an international consortium of institutions and individuals from all over the world, provides an evaluation tool for SWC activities, an information-management system designed to collect, analyze, present, and disseminate SWC knowledge and a decision-support system designed to assist in the search for SWC options appropriate to the prevailing biophysical and socioeconomic settings. WOCAT was initated in 1992 and has developed a common framework and methodology, consisting of three comprehensive questionnaires (in English, French, and Spanish) for the documentation and evaluation of SWC.

The WOCAT methodology consists of three major modules:

1. Questionnaire and database on SWC technologies;
2. Questionnaire and database on SWC approaches;
3. Questionnaire and database on the geographic distribution of SWC (mapping).

The first two modules aim at a comprehensive and detailed description of specific technologies, i.e., agronomic, vegetative, structural, and/or management measures used in the field (Figure 6), and the ways and means used to implement an SWC technology on the ground. The mapping module is more or less similar to the qualitative methodology for degradation assessment described earlier. In this approach, information is collected for individual units of a (physiographic or other) base map on the following items:

- Land use: type, extent, trend in area, trend in intensity;
- (Per land use type) degradation, as above, but only for water and wind erosion and fertility decline in erosion-prone areas;

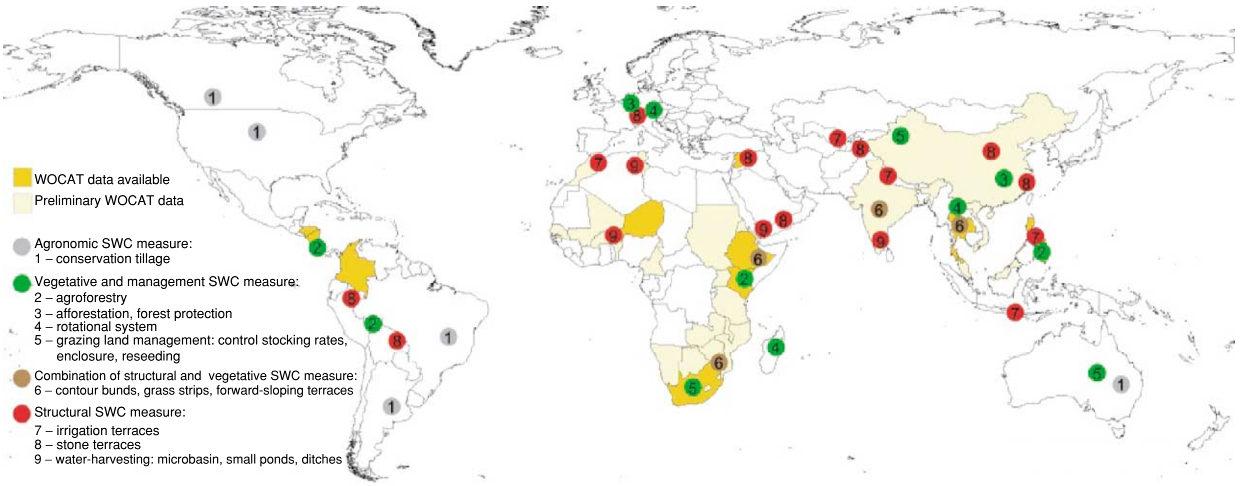

WOCAT data available
Preliminary WOCAT data
Agronomic SWC measure:
1 – conservation tillage
Vegetative and management SWC measure:
2 – agroforestry
3 – afforestation, forest protection
4 – rotational system
5 – grazing land management: control stocking rates, enclosure, reseeding
Combination of structural and vegetative SWC measure:
6 – contour bunds, grass strips, forward-sloping terraces
Structural SWC measure:
7 – irrigation terraces
8 – stone terraces
9 – water-harvesting: microbasin, small ponds, ditches

Figure 7 Reported soil conservation measures as recently compiled by the WOCAT initiative.

- (Per land use type) conservation: type, extent, period of implementation, effectiveness, trend in effectiveness, and reference to a corresponding questionnaire in the technology database for more detailed information;
- (Per land use type) productivity: trend, contribution of SWC or degradation to this trend, average production value, average input value.

An overview of soil and water conservation measures applied in different geographic regions of the world is shown in Figure 7.

See also: **Desertification**; **Erosion:** Water-Induced; Wind-Induced; **Salination Processes**

Further Reading

Blaikie PM and Brookfield HC (1987) *Land Degradation and Society*. London, UK: Methuen.

Blum WEH (1988) *Problems of Soil Conservation*. Nature and Environment Series 39. Strasbourg, France: Council of Europe.

Crosson P (1997) The on-farm economic costs of erosion. In: Lal R, Blum WEH, Valentin C, and Stewart BA (eds) *Methods for Assessment of Soil Degradation*. Boca Raton, FL: CRC Press.

Dregne HE (1997) Desertification assessment. In: Lal R, Blum WH, Valentin C, and Stewart BA (eds) *Methods for Assessment of Soil Degradation*. Boca Raton, FL: CRC Press.

Lal R, Blum WH, Valentine C, and Stewart BA (eds) (1997) *Methods for Assessment of Soil Degradation*. Boca Raton, FL: CRC Press.

Liniger HP, van Lynden GWJ, and Schwilch G (2002) Documenting field knowledge for better land management decisions – Experiences with WOCAT tools in local, national and global programs. *Proceedings of ISCO Conference 2002*, vol. I, pp. 259–267, Beijing, China: Tsinghua University Press.

Oldeman LR, Sombroek WG, and Hakkeling R (1991) *World Map on the Current Status of Human Induced Soil Degradation. An Explanatory Note*, 2nd edn. Wageningen, the Netherlands: ISRIC/Nairobi, Kenya: UNEP.

Ritsema CJ (ed.) (2003) Soil erosion and participatory land use planning on the Loess Plateau in China. (Special issue.) *Catena* 754.

Stocking M and Murnaghan N (2001) *Handbook for the Field Assessment of Land Degradation*. Sterling, VA: Earthscan Publications.

Van der Meer FD and de Jong SM (eds) (2001) *Imaging Spectrometry: Basic Principles and Prospective Applications*. Bookseries Remote Sensing and Digital Image Processing, vol. 4. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Van Lynden GWJ (1995) *European soil resources. Current status of soil degradation, causes, impacts and need for action*. Nature and Environment Series 71. Strasbourg, France: Council of Europe.

DENITRIFICATION

D A Martens, USDA Agricultural Research Service, Tucson, AZ, USA

Introduction

Denitrification is the major biological process through which the soil N available to plants is returned to the nonavailable atmospheric N pool as various N oxides (NO, N_2O) and dinitrogen (N_2). Despite the central role of microbial denitrification for N-cycle regulation of plant-available N, this process remains one of the least quantified mechanisms of soil N transformation. Rates of N loss from fertilizer applied to agricultural soils through denitrification can vary tremendously, ranging from 0 to 70% of applied N. In the USA, N_2O from the use of N fertilizers in 1998 accounted for 45% of the total N_2O emission budget.

Definitions and Pathways

Denitrification is defined as the "microbial reduction of nitrate or nitrite coupled to electron transport phosphorylation resulting in gaseous N either as molecular N_2 or as an oxide of N." The key to denitrification as defined is the availability of the N oxides, nitrite (NO_2^-) or nitrate (NO_3^-), which are formed from the autotrophic nitrification pathway substrate, ammonia (NH_3), which is derived from ammonium (NH_4^+). Chemical fertilizer inputs and soil organic-matter mineralization are the main sources of NH_4^+ in the environment. Nitrous oxide production from soils has been linked to two biological processes. The first is during the process of nitrification of NH_4^+ under aerobic conditions and the second is the coupled nitrification/denitrification (denitrification) pathway that occurs under anaerobic conditions. The pathway proposed for nitrifying organisms to release N_2O during the nitrification process has been defined as nitrifier denitrification.

The denitrification pathway is prevalent once NO_3^- is formed and the correct environmental conditions (low- or no-O_2 concentrations, and high soluble C content) are imposed with microbial reduction of the N oxides. The term 'denitrification' or 'respiratory denitrification' has been defined as a bacterial respiratory process and as such, a clear distinction between the denitrification pathway and the nitrifier denitrification pathway is necessary, because the relative proportion of N_2O production from the pathways is affected by different environmental conditions. A third pathway involving the chemical decomposition of NO_2^- has also been found in soils and can be prevalent in low-pH environments. The nonbiological pathways, or chemodenitrification, are so closely linked with nitrification that is often difficult to determine whether the nitric oxide (NO) and N_2O produced are formed through nitrification or chemodenitrification. Early research on denitrification was the result of the inability to mass balance total inputs of N and outputs of N in agricultural systems. The large portion of N unaccounted for was determined to be lost as gaseous N, which agronomically, resulted in decreased N fertilizer efficiency.

Research into global-change processes has found another impact of the N gas loss during the nitrification–denitrification process. Nitrous oxide is now known to be a potent greenhouse gas with important impacts on our environment. Nitrous oxide has a global-warming potential about 320 times as strong as carbon dioxide mainly due to an atmospheric lifetime of about 120 years. As the amounts of industrial or biologically fixed N_2 used for crop production increase, the production of N_2O due to nitrification–denitrification processes will also increase and potentially cause significant depletion of the Earth's stratospheric ozone layer and contribute to a warming of the Earth's surface by influencing the radiative budget of the troposphere.

Organisms and Substrates

Denitrification

Denitrification is the stepwise reduction of N oxides with gaseous products such as N_2O or N_2 under conditions of limited O_2 (Figure 1). The process results from the use of N oxide as a terminal electron acceptor instead of molecular O_2 by bacteria and is irreversible once NO is formed. Thus, a source of organic C is required for bacteria metabolism and sufficient NO_3^- to act as an electron acceptor must be available in an environment that has limited O_2. The requirement of all three factors, C source, low O_2, and sufficient NO_3^- must be present for the occurrence of denitrification. The large coefficient of

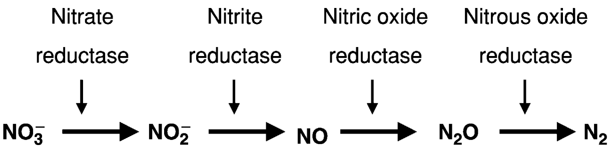


Figure 1 Pathway and enzymes involved in denitrification.

variation for measurement of gaseous N loss from field soils has been accredited to the formation of soil ‘hotspots’ that exhibit increased soil respiration. Spatial variability in N_2O fluxes typically causes coefficients of variation from 30 to 100% to be measured at a spatial scale of several meters with field plots. Hotspots of organic debris can stimulate denitrification due to rapid microbial mineralization and release of NH_4^+ and NO_3^- formation with concomitant depletion of O_2 concentrations. The best field indicator for denitrification has been found to be drainage class, which determines the potential for the soil becoming waterlogged at times of high precipitation. Seasonal differences generally noted for denitrification (lower in spring compared with higher in autumn) have also been attributed to low spring NO_3^- concentrations in temperate soils. Microbial activity has also been found to result in increased denitrification at the soil surface compared with deeper subsoils due to the greater organic inputs to the soil surface.

The genera of denitrifying bacteria are diverse. The capacity to denitrify is present in about 23 genera of bacteria. The list of denitrifying genera in Table 1 includes 13 genera (considered facultative aerobes) for which there is confirmed or multiple documentation of denitrifying activity. Although the distribution of denitrifiers is ubiquitous, the activity or measurement of denitrifier biomass and synthesis of denitrifying enzyme activity from soil field cores is poorly correlated with measured N gas loss. Recent laboratory research has also found that yeast and filamentous fungi are also capable of gaseous N production, but field confirmation of importance is lacking. Fungal N_2O production may be of great significance, because fungi often dominate the microbial biomass of temperate soils, yet fungi are not tolerant of waterlogged conditions. Since denitrifiers are not fermentative (facultative anaerobes), growth under O_2-limited conditions is solely dependent on the presence of N oxides. Enzymes involved in the reactions are nitrate reductase, nitrite reductase, nitric oxide reductase, and nitrous oxide reductase (Figure 1). During this process, N_2O is an intermediate and so the ratio of N_2O to N_2 fluxes can be affected by various environmental factors. The proportion of N_2O compared with N_2 is increased if the soil pH is decreased, if the amount of NO_3^- is increased, and if lower O_2 concentrations are present. In each case, the environmental factors have an influence on the enzyme nitrous oxide reductase. To summarize, the release of N_2O as an intermediate of denitrification can be an important N loss mechanism in low-O_2 environments with sufficient NO_3^- coupled with rapid mineralization of C.

Table 1 Genera of documented denitrifying bacteria

Genus	*Characteristics*
Alcaligenes	Common soil isolate
Agrobacterium	Some species are plant pathogens
Azospirillum	Associated with grass species N_2 fixation
Bacillus	Thermophilic denitrifiers reported
Flavobacterium	Recent denitrifier identification
Halobacterium	Requires high salt concentrations for growth
Hyphomicrobium	Grows on one-carbon substrates
Paracoccus	Capable of lithotrophic and heterotrophic growth
Propionibacterium	Fermenters capable of denitrification
Pseudomonas	Common soil isolates
Rhizobium	Capable of N_2 fixation in symbiosis with legumes
Rhodopseudomonas	Photosynthetic bacteria
Thiobacillus	Generally grow as chemoautotrophs

Nitrifier Denitrification

Nitrification is the bacterial oxidation of NH_4^+ or NH_3 via NO_2^- to NO_3^-. This process is carried out by two groups of autotrophic bacteria. The first step, from NH_3 to NO_2^-, is catalyzed by NH_3 oxidizers, and the second step, from NO_2^- to NO_3^-, is carried out by NO_2^- oxidizers. The best-studied member of the first group is *Nitrosomonas europaea*, and *Nitrobacter winogradskyi* is a representative of the NO_2^- oxidizers. Use of molecular probes specific for genes encoding for heme-type dissimilatory nitrite and nitrous oxide reductase gene homologies derived from *Nitrosomonas europaea* gene sequences has found that prominent nitrifiers of the genera *Nitrosospiria* and *Nitrosolobus* may differ from *Nitrosomonas europaea* in the mechanism and capacity for denitrification. Nitrous oxide can be released during the synthesis of the nitrification intermediates (Figure 2). The first possible step can form hydroxylamine (NH_2OH) as a potentially unstable intermediate, and N_2O can be released at this step. The formation of NO_2^- can serve as a second point where N_2O can

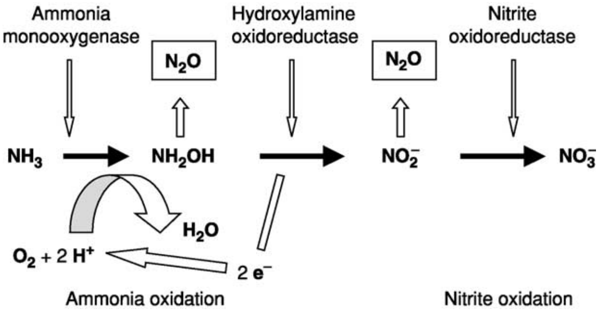


Figure 2 Pathway and enzymes involved in nitrification.

be released. The NO_2^- pathway to N_2O via nitrifier denitrification has been observed to be consistent with the kinetics exhibited by the denitrification pathway, but can occur in an environment with higher O_2 concentrations. In contrast to the traditional pathway of denitrification previously outlined, autotrophic and heterotrophic microorganisms are often able to release N_2O under aerobic conditions, because the intermediate NH_2OH can act as an electron donor. The relative proportion of nitrifier denitrification to the N_2O budget ranges from 0 to 30% of the total N_2O released.

Nitrification has also been found to occur in heterotrophic fungi and bacteria that utilize the same substrate, intermediates, and products as autotrophic nitrification, and again in contrast to denitrifiers, heterotrophic nitrifying bacteria are often able to form N_2O under aerobic conditions. Release of N_2O by heterotrophic nitrification is considered a minor contribution to N_2O budget.

Environmental Factors

Denitrification supports microbial life through respiration, because mineralization of a reduced substrate (mainly organic C, but reduced iron or sulfur may also donate electrons) donates electrons via carriers to an oxidized N oxide. The decline in free energy is coupled with the generation of adenosine triphosphate (ATP). In the presence of O_2, the ATP yield per number of electrons transported is about 1.7 times greater than if NO_3^- is the terminal electron acceptor, with the final efficiency apparently being species-dependent. Thus the rate of denitrification processes depends on the interactions and availability of reduced substrates, limited O_2 potential and presence of N oxides to function as terminal electron acceptors.

Organic C Availability

The availability of a suitable reduced C source is important because of the strong relationship between denitrification and soil organic matter. More important, the water-soluble portion of the soil organic matter has been found to be a good indicator for potential denitrification. Recent work has also found that organic litter with low C to N ratios has a greater potential to promote higher denitrification activity than litter with higher C to N ratios. Animal manures with a readily decomposable C source and a high N content can greatly enhance soil denitrification rates.

The presence or absence of plants has also been found to affect denitrification rates. Higher denitrification rates have been noted in the plant rhizosphere as compared with bulk soil several millimeters away from the roots if NO_3^- was sufficient; but if NO_3^- availability was limited, the presence of plants decreased denitrification activity. Plant roots have the potential to impact denitrification activity by providing C to support microbial populations and supply electrons for NO_3^- reduction. Roots can also create anaerobic zones through respiratory O_2 consumption when O_2 supply is limited by water conditions. The rooting environment can create a drier soil through evapotranspiration processes and result in increasing O_2 diffusion in the drier soil. Roots can also assimilate NO_4^- and NO_3^- from the soil solution reducing the availability of substrate to the denitrification pathways.

Due to a typical decrease in available soil C with depth, denitrification activity has also been found to decrease with soil depth and is potentially limited by a lack of available C below 60–75 cm. Other soil characteristics can also change with depth concomitant with organic C content such as porosity/permeability, pH, temperature, and depth to water table that may also have an impact on denitrification rates.

Controls on O_2 Availability

Reduced availability or absence of O_2 is required for both the synthesis and activity of enzymes involved in denitrification. Soil parameters can control overall O_2 availability and influence denitrification rates. The universally recognized inverse relationship between water content and denitrification is due to the slower diffusion of O_2 to metabolically active sites with the greater amounts of water present in the soil matrix. Bulk soil parameters that have been used to describe O_2 control of denitrification include soil moisture content, soil water potentials, partial pressure of O_2, O_2 concentration in solution, percentage air-filled porosities, and redox potentials. Bulk measurements have shown promise for predicting denitrification activity in laboratory studies, but, in field measurements, spatial variability due to hot spots can mask the predictive nature of O_2 diffusion and/or C mineralization rates.

Soil texture is an important factor for determining critical air-filled porosity, as a finer-textured soil (less sand content) will have a higher critical air-filled porosity value than found for coarser-textured soils. O_2 concentrations in soil atmospheres at which denitrification has been observed for different soils types range from 4 to 17%. Although the bulk soil O_2 concentrations are important, it has been noted that the O_2 concentration at the microsite level is more important for the determination of denitrification activity.

The occurrence of anaerobic microsites can explain how an O_2-sensitive process is possible in a well-structured soil whose bulk O_2 measurements show aerobic conditions. The microsite concept proposes that anaerobic sites occur within centers of water-saturated soil aggregates because of limited diffusion of O_2 and consumption of O_2 by microbial metabolism. Research has shown that the occurrence of anaerobic microsites is largely a function of localized C availability and is possible within aggregates exceeding 9 mm. The microsites are more likely to occur in the rhizosphere and with particles of decomposing plant litter in zones where water can limit O_2 diffusion. Again, denitrification activity is a function of available C, limited O_2 potentials, and a source of N oxides.

Nitrate Supply

Research has indicated that the rate of denitrification activity in pure-culture studies was independent of NO_3^- concentration. Recent research has reported first-order kinetics with NO_3^- concentrations of less than 40 mg l^{-1}, and the kinetics changed from first-order to zero-order for NO_3^- concentrations of more than 100 mg. Microbial NO_3^- reduction is an enzyme-catalyzed reaction, and use of Michaelis–Menten kinetics has found K_m values for NO_3^- reduction to be less than 15 μmol l^{-1} NO_3^- or approx. 93 μg NO_3^- l^{-1} for several bacterial isolates. The use of K_m values determines that concentrations required for NO_3^- reduction are higher for soils than those found for pure culture or enzyme studies. The rate kinetics also change if the soil has recently received chemical fertilizers. However, Michaelis–Menten kinetics has not been found to describe NO_3^- reduction dependence accurately on available C and NO_3^- concentrations in soil. Several studies have found that kinetic analysis may not reflect denitrification rates, but the rate of NO_3^- diffusion from solution into the actively denitrifying soil aggregate. The evidence indicates that even very low NO_3^- concentrations present in soil solutions are in excess of enzyme and microbial needs for terminal electron acceptors during O_2 stress, but denitrifying activity may actually be limited by the diffusion of NO_3^- into the soil microsites.

Temperature

Denitrification, because of enzyme dependence on temperature for the activity rate, would be expected to increase exponentially with increased temperature within the range of enzyme activity. The temperature effect is impacted by the interaction of temperature on O_2 solubility and O_2 consumption as well as O_2 diffusion. Decreasing temperature will increase O_2 solubility and decrease O_2 consumption. Research has found a minimum soil temperature range of 2.7–10°C for denitrifying activity and a maximum temperature of approx. 50°C. Above 50°C, chemical decomposition reactions of NO_2^- can complicate the denitrification–temperature relationship. Pure culture experiments with soil isolates found 30°C to be the optimum temperature for growth, but evidence suggests that denitrifiers are adapted to soils and growth conditions of the specific soil. Bacterial isolates from tropical soils (33 in total) show no growth when incubated at 10°C, whereas 68% of temperate soil isolates (95 in total) grow at 10°C.

pH

The pH requirement for denitrifying organisms appears to be similar to heterotrophic organisms. In the neutral pH range from 6 to 8, there is a limited effect of pH on the activity of denitrification, but denitrification rates can be limited in acid pH ranges. Adjusting a naturally acid soil (pH 3.5) with reduced microbial activity to pH 6.5 rapidly stimulates a strong increase in denitrification rates and suggests that a general population with neutral growth optimum is present in a low-pH environment. Research has suggested that the decrease in denitrification activity in a moderately acid environment favors NO_3^- reduction to NO_2^- and that increased soil acidity may favor NO_3^- reduction rather than denitrification. Isotopic evidence has suggested that, although a moderate change in pH (pH 6.7 to 5.2) has little effect on the total rate of N_2 and N_2O production, the pH adjustment results in a rapid change from predominantly N_2 production at pH 6.7 to predominantly N_2O at pH 5.2. Due to the ratio change noted with change in soil pH value, the enzyme N_2O reductase may be extremely sensitive to pH.

Conclusion

Three possible soil-based mechanisms can give rise to gaseous loss of N. Denitrification and nitrifier denitrification are biological pathways that are very closely linked by the production of NO_3^- in soils and sediments. The chemodenitrification pathway is also linked to the biological processes, but occurs due to NO_2^- instability in an acid environment. The requirements for denitrification activity include a reduced C source, a NO_3^- pool, and an O_2-limited environment. The interactions with the three factors can result in different proportions of N_2O and N_2 being produced. The N_2O to N_2 ratio produced by the denitrification pathway can be increased by increased NO_2^- and NO_3^- concentration. The ratio can also be increased by decreased O_2 concentration, pH values, and increasing the C availability. The spatial variation of soil C and N and low O_2 tensions required for denitrification during field quantification and the importance of nitrifier denitrification processes are crucial areas of research that need to be understood before management of gaseous N loss can be achieved.

***See also*:** **Fertilizers and Fertilization**; **Greenhouse Gas Emissions**; **Microbial Processes:** Kinetics; **Nitrogen in Soils:** Cycle; Nitrification; **Organic Matter:** Principles and Processes; **Organic Residues, Decomposition**

Further Reading

Arah JRM (1997) Apportioning nitrous oxide fluxes between nitrification and denitrification using gas phase mass spectrometry. *Soil Biology and Biochemistry* 29: 1295–1299.

Firestone MK (1982) Biological denitrification. In: Stevenson FJ (ed.) *Nitrogen in Agricultural Soils*, pp. 289–326. Agronomy Monograph no. 22, Madison, WI: American Society of Agronomy.

Focht DD (1982) Denitrification. In: Burns RG and Slater JH (eds) *Experimental Microbial Ecology*, pp. 194–211. Oxford: Blackwell Scientific Publications.

Hochstein LI and Tomlinson GA (1988) Enzymes associated with denitrification. *Annual Review of Microbiology* 42: 231–261.

Laughlin RJ and Stevens RJ (2002) Evidence for fungal dominance of denitrification and codenitrification in a grassland soil. *Soil Science Society of America Journal* 66: 1540–1548.

Parkin TB (1990) Characterizing the variability of soil denitrification. *FEMS Symposium of the Federal European Microbiological Society* 56: 213–228.

Parkin TB and Meisinger JJ (1989) Denitrification below the crop rooting zone as influenced by surface tillage. *Journal of Environmental Quality* 18: 12–16.

Poth M and Focht DD (1986) ^{15}N kinetic analysis of N_2O production by *Nitrosomonas europaea*: an examination of nitrifier denitrification. *Applied and Environmental Microbiology* 49: 1134–1141.

Tiedje JM (1982) Denitrification. In: Page AL, Miller RH, and Keeney DR (eds) *Methods of Soil Analysis, part 2, Chemical and Microbiological Properties*, 2nd edn, pp. 1011–1026. Madison, WI: American Society of Agronomy.

Wrage N, Velthof GL, van Beusichem ML, and Oenema O (2001) Role of nitrifier denitrification in the production of nitrous oxide. *Soil Biology and Biochemistry* 33: 1723–1732.

DESERTIFICATION

D Hillel, Columbia University, New York, NY, USA
C Rosenzweig, NASA Goddard Institute for Space Studies, New York, NY, USA

Introduction

Ecosystems in semiarid and arid regions around the world appear to be undergoing various processes of degradation commonly described as ‘desertification.’ According to the United Nations Environmental Program (UNEP), all regions in which the ratio of total annual precipitation to potential evapotranspiration (P/ET) ranges from 0.05 to 0.65 should be considered vulnerable to desertification. Such regions constitute some 40% of the global terrestrial area. They include northern Africa, southwestern Africa, southwestern Asia, central Asia, northwestern India and Pakistan, southwestern USA and Mexico, western South America, and much of Australia, and are home to an estimated sixth of the world’s population.

‘Desertification’ is a single word used to cover a wide variety of interactive phenomena – both natural and anthropogenic – affecting the actual and potential biological and agricultural productivity of ecosystems in semiarid and arid regions. It is an emotive term, conjuring up the specter of a tide of sand swallowing fertile farmland and pastures. Apparently with this somewhat simplistic image in mind, UNEP sponsored projects in the early 1980s to plant trees along the edge of the Sahara, with the aim of warding off the invading sands. While there are places where the edge of the desert can be seen encroaching on fertile land, the more pressing problem is the deterioration of the land due to human abuse in regions well outside the desert. The latter problem emanates not from the expansion of the desert *per se* but from the centers of population outside the desert, owing to human mismanagement of the land. A vicious cycle has begun in many areas: as the land degrades through misuse, it is worked or grazed ever more intensively, so its degradation is exacerbated; and as the returns from ‘old’ land diminish, ‘new’ land is brought under cultivation or under grazing in marginal or even submarginal areas.

As defined in recent dictionaries, desertification is the process by which an area becomes (or is made to become) desert-like. The word ‘desert’ itself is derived from the Latin *desertus*, being the past participle of *deserere*, meaning 'to desert,' 'to abandon.' The clear implication is that a desert is an area too barren and desolate to support human life. An area that was not originally desert (e.g., a steppe or savanna) may come to resemble a desert if it loses so much of its usable resources that it can no longer provide adequate subsistence to a given number of humans. This is a very qualitative definition, since not all deserts are the same. An area's resemblance to a desert does not make it a permanent desert if it can recover from its damaged state, and, in any case, the modes of human subsistence and levels of consumption differ greatly from place to place.

In recent decades, the very term 'desertification' has been called into question as being too vague, and the processes it purports to describe too ill-defined. Some critics have even suggested abandoning the term, in favor of what they consider to be a more precisely definable term, namely 'land degradation.' However, 'desertification' has already entered into such common usage that it can no longer be revoked or ignored. It must therefore be clarified and qualified so that its usage is less ambiguous.

'Land degradation' itself is a vague term, since the land may be degraded with respect to one function and not necessarily with respect to another. For example, a tract of land may continue to function hydrologically – to regulate infiltration, runoff generation, and groundwater recharge – even if its vegetative cover is changed artificially from a natural species-diverse community to a monoculture and its other ecologic functions are interrupted. Perhaps better than 'land degradation' is the term 'semiarid ecosystem degradation.' A semiarid ecosystem encompasses the diverse biotic community sharing the domain. Included in this community is the host of plants, animals, and microorganisms that interact with one another through such modes as competition or symbiosis, predation, and parasitism. It also includes the complex physical and chemical factors that condition the lives of those organisms and are in turn influenced by them. Each ecosystem performs a multiplicity of ecologic functions. Included among these are photosynthesis, absorption of atmospheric carbon and its incorporation into biomass and the soil, emission of oxygen, and regulation of temperature and the water cycle, as well as the decomposition of waste products and their transmutation into nutrients for the perpetuation of diverse, interdependent forms of life.

A semiarid ecosystem may be more or less natural, relatively undisturbed by humans, or it may be artificially managed, such as an agroecosystem. An agroecosystem is a part of the landscape that is managed for the economic purpose of agricultural production. The transformation of a natural ecosystem into an agroecosystem is not necessarily destructive if the latter is managed sustainably and productively, and if it coexists harmoniously alongside natural ecosystems that continue to maintain biodiversity and to perform vital ecologic functions. In too many cases, however, the requirements of sustainability fail, especially where agricultural systems expand progressively at the expense of the remaining, more-or-less natural ecosystems. The appropriation of ever-greater sections of the remaining native habitats, impelled by the increase in population as well as by the deterioration of farmland or rangeland due to overcultivation or overgrazing, decimates those habitats and imperils their ecologic functions. In the initial stages of degradation, the deteriorating productivity of an agroecosystem can be masked by increasing the inputs of fertilizers, pesticides, water, and tillage. Sooner or later, however, if such destructive effects as organic matter loss, erosion, leaching of nutrients, and salination continue, the degradation is likely to reach a point at which its effects are difficult to overcome either ecologically or economically.

Key processes related to desertification include drought, primary production and carrying capacity, soil degradation, and use of water resources. The role of social factors is also important.

The Role of Drought

A typical feature of arid regions is that the mode (the most probable) amount of annual rainfall is generally less than the mean; i.e., there tend to be more years with below-average rainfall than years in which the rainfall is above average, simply because a few unusually rainy years can skew the statistical mean well above realistic expectations for rainfall in most years. The variability in biologically effective rainfall is yet more pronounced, as years with less rain are usually characterized by greater evaporative demand, so the moisture deficit is greater than that indicated by the reduction of rainfall alone. Timing and distribution of rainfall also play crucial roles. Below-average rainfall, if well distributed, may produce adequate crop yields, whereas average or even above-average rainfall may fail to produce adequate yields if the rain occurs in just a few ill-timed storms with long dry periods between them.

In semiarid agricultural regions, 'drought,' like desertification, is a broad, somewhat subjective term that designates years in which cultivation becomes an unproductive activity, crops fail, and the productivity of pastures is significantly diminished. Drought is a constant menace, a fact of life with which rural dwellers in arid regions must cope if they are to survive. The occurrence of drought is a certainty, sooner or later; only its timing, duration, and severity are ever in doubt. And it is during a drought that ecosystem degradation in the form of denudation and soil erosion occur at an accelerated pace, as people try to survive in a parched habitat by cutting the trees for fuel and browse, and by animals overgrazing the wilted grass. The topsoil, laid bare and pulverized by tillage or the trampling of animals, is then exposed to a greatly increased risk of wind erosion. When the coveted rains recur, they tend to scour the erodible soil.

Any management system that ignores the eventuality of drought and fails to provide for it ahead of time is doomed to fail in the long run. That provision may take the form of grain or feed storage (as in the biblical story of Joseph in Egypt), or of pasture tracts kept in reserve for grazing when the regular pasture is played out, or of the controlled migration of pastoralists to other regions able to accommodate them for the period of the drought.

A much-debated question is whether the frequency, duration, and severity of droughts have been increasing in recent decades. One possibility is that the process of desertification, once begun, produces a feedback effect that is self-exacerbating. Some have hypothesized that the increase in atmospheric dust from denuded and wind-eroded drylands (the so-called dust-bowl effect), as well as from air pollution (as denudation of an area's vegetation is associated with biomass burning, which releases smoke into the air), may have changed the patterns of air mass movement and hence of precipitation. Another hypothesis is that droughts can be worsened by the increased reflectivity of the bared surface to incoming sunlight. That reflectivity, called 'albedo,' may rise from approximately 20% for a well-vegetated area to perhaps 35% or more for an exposed, bright, sandy soil. As a larger proportion of the incoming sunlight is reflected skyward rather than absorbed, the surface becomes cooler than it would be otherwise, and so the air in contact with the surface has less of a tendency to rise and condense its moisture so as to yield rainfall.

An additional effect of denudation is to decrease the interception of rainfall by vegetation and the infiltration rate, while increasing surface runoff, thereby reducing the amount of soil moisture available for evapotranspiration. Crops and grasses, which have shallower roots than trees and in any case transpire less than the natural mixed vegetation of the savanna, transpire even less when deprived of moisture during a drought. The 'biophysical feedback' hypothesis is that such changes may reduce regional precipitation. Lower rainfall leads in turn to more overgrazing and less regrowth of biomass, and to further reduction in reevaporated rain owing to the decline in soil moisture. Thus, the feedback hypothesis offers its own explanation as to why the drought in the African Sahel, for example, has tended to persist.

Primary Production and Carrying Capacity

The biological productivity of any ecosystem is due to its primary producers (known as autotrophs), which are the green plants growing in it. They alone are able to create living matter from inorganic materials. They do so by combining atmospheric carbon dioxide with soil-derived water, thus converting radiant energy from the sun into chemical energy in the process of photosynthesis. Green plants also respire, which is the reverse of photosynthesis, and in so doing they utilize part of the energy to power their own growth. The net primary production then becomes available for the myriad of heterotrophs, which subsist by consuming (directly or indirectly) the products of photosynthesis. A stable ecosystem is one in which production and consumption, synthesis and decomposition, are in balance over an extended period of time.

When humans enter into an ecosystem and appropriate some of its products for themselves, they do so in competition with other potential consumers. As populations increase, the tendency is to intensify the use of resources by promoting the production of desired goods while suppressing the species that do not serve that end. In the process, the ecosystem's biodiversity and natural productivity are profoundly affected. Especially affected are areas within the semiarid and arid regions, which, because of the paucity of water and the fragility of the soil (typically deficient in organic matter, structurally unstable, and highly erodible) are most vulnerable and least resilient.

The term 'carrying capacity' has been used to characterize an area's productivity in terms of the number of people or grazing animals it can support per unit area on a sustainable basis. However, the productive yield obtainable from an area depends on how the area is being used. Under the hunter-gatherer mode of subsistence, an area may be able to carry only, for example, 1 person per square kilometer, whereas under shifting cultivation it may carry 10, and under intensive agriculture perhaps 100. The intensive forms of utilization also involve inputs of capital, energy, and materials, such as fertilizers and pesticides, which are brought in from the outside to enhance an area's productivity. As the usable productivity is strongly affected by the supply of water (i.e., by seasonal rainfall), it varies from year to year and from decade to decade, so a long-term average is difficult to determine, especially given the prospect of climate change. It is therefore doubtful if any given area can be assigned an intrinsic and objectively quantifiable carrying capacity. By whatever measure, however, the capacity of an area to support a given population is clearly diminished by human mismanagement.

Wherever human pressure on the land ceases or is diminished, even a severely eroded soil may recover gradually. However, on the time-scale of years to a few decades, especially if overgrazing and overcultivation continue, soil erosion may become, in effect, irreversible.

Soil Degradation and Rehabilitation

An important criterion of soil degradation is the loss of soil organic matter. Compared with soils in more humid regions, those in warm arid regions tend to be inherently poor in organic matter from the outset, owing to the relatively sparse natural vegetative cover and to the rapid rate of decomposition. The organic matter present is, however, vitally important to soil productivity. Plant residues over the surface protect the soil from the direct impact of raindrops and from deflation by wind, and help to conserve soil moisture by minimizing evaporation. Plant and animal residues that are partially decomposed and that are naturally incorporated into the topsoil help to stabilize its structural aggregates, which in turn enhance infiltrability, reduce water loss by runoff, and enable seed germination and root growth. The organic matter present also contributes to soil fertility by the gradual release of nutrients.

When the natural vegetative cover is removed, and especially when the soil is tilled and/or trampled repeatedly, there follows a rapid process of organic matter decomposition and depletion. Accelerated erosion also removes the layer of topsoil that is richest in organic matter. Consequently, the destabilized soil tends to form a surface crust that further inhibits infiltration. Water losses by both runoff and evaporation increase, and the soil loses an important source of nutrients. These destructive processes can be countered or ameliorated by methods of conservation management, including minimum or zero tillage, maintenance of crop residues, the periodic inclusion of green manures in the crop rotation, and agroforestry.

The destructive processes induced by soil mismanagement, and, in contrast, the constructive processes induced by conservation management, though seemingly local, may have – when practiced on a regional scale – an impact on climate charge. Soils subject to accelerated decomposition of organic matter tend to release carbon dioxide and thus contribute to the enhanced greenhouse effect. Conversely, soils that are being revegetated and enriched with organic matter can absorb and sequester quantities of carbon that are extracted from the atmosphere in photosynthesis.

Potentialities and Problems of Irrigation

Where fresh water resources (from riverine or underground sources) are available and can be utilized economically, irrigation can be an effective way to intensify and stabilize production in semiarid or arid regions, and thus to relieve the pressure on extensive areas of rain-fed land that are most vulnerable to degradation. Irrigation is the deliberate supply of water to agricultural crops, designed to permit farming in arid regions and to offset drought in semiarid regions. Even in areas where total seasonal rainfall is adequate on average, it may be poorly distributed during the year and variable from year to year.

Wherever traditional rain-fed farming is a high-risk enterprise owing to scarce or uncertain precipitation, irrigation can help to ensure stable production. It not only raises the yields of specific crops but also prolongs the effective crop-growing period in areas with dry seasons, thus permitting multiple cropping (two, three, or even four crops per year) where only a single crop could be grown otherwise. With the security provided by irrigation, additional inputs needed to intensify production further (pest control, fertilizers, improved varieties, etc.) become economically feasible. Irrigation reduces the risk of these expensive inputs being wasted by crop failure resulting from lack of water. Although irrigated land amounts to only some 17% of the world's total cropland, it contributes well over 30% of the total agricultural production. That vital contribution is most important in arid regions, where the supply of water by rainfall is least, even as the demand for water imposed by the bright sun and the dry air is greatest.

Irrigation, however, is not without its own problems. From its inception in the Fertile Crescent some six or more millennia ago, irrigated agriculture, especially in poorly drained river valleys, has brought about processes of degradation that have threatened its sustainability. The application of water to the land tends to raise the water table, which in turn induces the self-destructive twin scourges of waterlogging and salination.

Some investigators include the degradation of irrigated lands in the category of desertification. Though the processes taking place differ fundamentally from those in rain-fed lands, the damage done to injudiciously irrigated lands is indeed in the category of ecosystem degradation. Processes occurring off-site (upstream as well as downstream of the irrigated area) strongly affect the sustainability of irrigation. For example, denudation of upland watersheds by forest clearing, cultivation, and overgrazing induces erosion and the subsequent silting of reservoirs and canals, thereby reducing the water supply. The construction of reservoirs often causes the submergence of natural habitats as well as of valuable scenic and cultural sites. Concurrently, the downstream disposal of drainage from irrigated land tends to pollute aquifers, streams, estuaries, and lakes with salts, nutrients, and pesticide residues. Finally, the irrigation system itself (its reservoirs, canals, and fields) may harbor and spread water-borne diseases, thus endangering public health. Thus the very future of irrigated agriculture has been called into question.

Experience shows that irrigation can be sustained, but at a cost. The primary cost is the necessary investment in systems of efficient irrigation (avoiding excessive application of water such as causes watertable rise, waterlogging, and salination), as well as in the timely provision of effective land drainage and the safe disposal of its salt-laden effluent.

Social Factors

Social factors are necessarily involved in both semiarid ecosystem conservation and its inverse, which is ecosystem degradation. Farmers who do not have tenure to the land are not likely to invest in its conservation or improvement. Neither are communities that lack stable institutional structure likely to establish and maintain essential infrastructure and services that enable, encourage, and coordinate farmers' efforts to implement land improvement and conservation measures. And no effective action at all may be possible in the absence of a proactive governmental policy, including the provision of credit or subsidies, professional guidance and training, as well as the preparation and implementation of national and regional drought contingency plans for both farmers and herders. The conservation of soil, water, and biotic resources is a collective societal concern, and an intergenerational one, not merely a private concern of the people utilizing the land directly at any particular time.

Finally, there looms the most difficult, yet inescapable, problem of population numbers. No system of management, however efficient it may be, can be sustained if the population continues to grow without limit. A crucial aspect of population control is the empowerment of women, through education and equal rights, as full participants in the management of their societies' physical, biological, and human resources. The issue is extremely sensitive inasmuch as it carries cultural and traditional, as well as social and economic, implications.

Monitoring Desertification

The techniques of remote sensing have made possible the monitoring of changes to ecosystems on a regional scale. Studies based on remote sensing of the African Sahel have shown that, contrary to many alarmist reportings, there has been no progressive change of the Saharan desert boundary. Rather, there has been a back-and-forth shifting of vegetative density during alternating spells (sometimes lasting several years) of below-average and of above-average rainfall. In principle, however, statistical criteria designed to test the probability levels of differences (between sites or between successive measurements on the same site) should not be used to 'prove' the opposite, namely that there are no differences. Measurements made at various times on large areas may obscure subtle local changes that could have taken place on a smaller scale.

One of the main indexes in use at present is the so-called normalized difference vegetation index (NDVI). It is obtained from the ratio between the red and near-red infrared reflectance bands, obtained from high-resolution radiometer data generated by the polar-orbiting satellite of the US National Oceanic and Atmospheric Administration (NOAA). In arid and semiarid regions, NDVI evidently correlates with the density of the vegetative cover and its biomass, as well as with its 'leaf area index' (LAI) and its photosynthetic activity. Care is needed, however, in applying NDVI to the assessment of net primary production, since the measurement of NDVI is oblivious to the amount of vegetation harvested by humans and by their animals prior to the time of the measurement.

Taken to be a general indicator of the 'greenness' of an area, NDVI has also been conjectured to correlate with biological productivity, but that correlation may not necessarily hold. In principle, the amount of vegetation present per unit area should depend on the amount produced *in situ* minus the amount removed from it. Therefore, the relation between an area's productivity and its vegetative biomass at any time must depend on whether the vegetation has been or is being 'harvested' (grazed by livestock or cut and carried away by humans). An area could be quite productive, yet relatively bare, if it had been harvested just prior to the NDVI measurement. Even if there is no discernible change in the density of an area's overall vegetative cover, there might well be a considerable change in the composition of the vegetation (i.e., its biodiversity, ecologic function, and feed value). For example, an overgrazed area may exhibit a proliferation of less nutritious plants at the same time that it loses the most palatable species of grasses and legumes that had contributed to the area's original carrying capacity.

Clearly, the most decisive factor affecting the overall density of vegetative cover in an arid region is the fluctuation of rainfall amounts. Taking the African Sahel as an example once again, we see that the annual precipitation has fluctuated widely over the decades. The amounts (as seen in Figure 1) for the last three decades of the twentieth century appear to be generally lower than those in the preceding decades. Although an analysis based on any particular short period may be misleading, the question does arise as to the possible effects of global climate change.

The Role of Climate Change

Ecosystems in arid and semiarid regions are likely to be increasingly influenced by global climate change. Emissions of radiatively active gases and aerosols due to human activity are altering the Earth's radiation balance and hence the temperature of the lower atmosphere. One of the manifestations of the change may be an increase in climate instability. In a warmer world, climatic phenomena are likely to intensify. Thus, episodes or seasons of anomalously wet conditions (violent rainstorms of great erosive power) may alternate with severe droughts, in an irregular and unpredictable pattern.

A more unstable climatic regime will make it harder to devise and more expensive to implement optimal land use and agricultural production practices, including drought-contingency provisions. Failure to prepare for such contingencies may exacerbate the consequences of extreme events such as floods and droughts, to the effect of worsening land degradation and periods of severe food shortages. Especially vulnerable is the continent of Africa, where the issues of greatest concern pertain to human health and food security, water resources, natural ecosystems and biodiversity, and, not least, land degradation or desertification (Table 1).

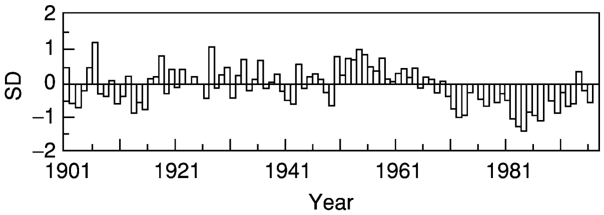


Figure 1 Rainfall fluctuations in the African Sahel during the period 1901–1998, expressed as a regionally averaged standard deviation, SD (departure from the long-term mean divided by the standard deviation). Reproduced from Densanker P and Magadza C (2001). Africa in climate change 2001: impacts, adaptation and vulnerability. In: McCarthy JJ, Canzania OF, Leary NA, Dokker DD, and White KS (eds) *International Panel of Climate Change*.

Climate change appears likely to cause further semiarid ecosystem degradation through alteration of spatial and temporal patterns in temperature, rainfall, solar insolation, winds, and humidity. Analyses based on global and regional climate models suggest that droughts may become more frequent, severe, and prolonged. It is still impossible, however, to ascribe exact probabilities to any of the various climate change scenarios, owing to uncertainties regarding future emissions of the radiatively active trace gases and tropospheric aerosols, and the potential responses of the climate system to those changes.

The question is sometimes posed: is it human exploitation of the land or is it overall climate change that constitutes the predominant cause of desertification? The answer is that the two sets of factors or processes interact and may well have become mutually reinforcing. Ultimately, both are impelled by human intervention and therefore can only be redressed by coordinated actions at the local, regional, and global levels.

Overview

The pressures generated by growing populations, intensified land use, and overall environmental change are evidently causing a progressive degradation of natural and managed ecosystems, especially in arid and semiarid regions. To define and quantify the nature, degree, and extent of the degradation, national and international agencies are working to implement consistent monitoring programs. These consist not only of remote sensing from above but also of ground-based observations.

Table 1 Sectors vulnerable to climate change in Africa

Sector	*Projected impacts*
Water resources	Dominant impact is predicted to be a reduction in soil moisture in the subhumid zones and a reduction in runoff
Food security	There is wide consensus that climate change, through increased extremes, will worsen food security in Africa
Natural resources and biodiversity	Climate change is projected to exacerbate risks to already threatened plant and animal species, and fuelwood
Human health	Vector-borne and water-borne diseases are likely to increase, especially in areas with inadequate health infrastructure
Desertification	Changes in rainfall, increased evaporation, and intensified land use may put additional stresses on arid, semiarid, and dry subhumid ecosystems

Reproduced from Densanker P and Magadza C (2001) Africa in climate change 2001: impacts, adaptation and vulnerability. In: McCarthy JJ, Canzania OF, Leary NA, Dokker DD, and White KS (eds) *International Panel of Climate Change*.

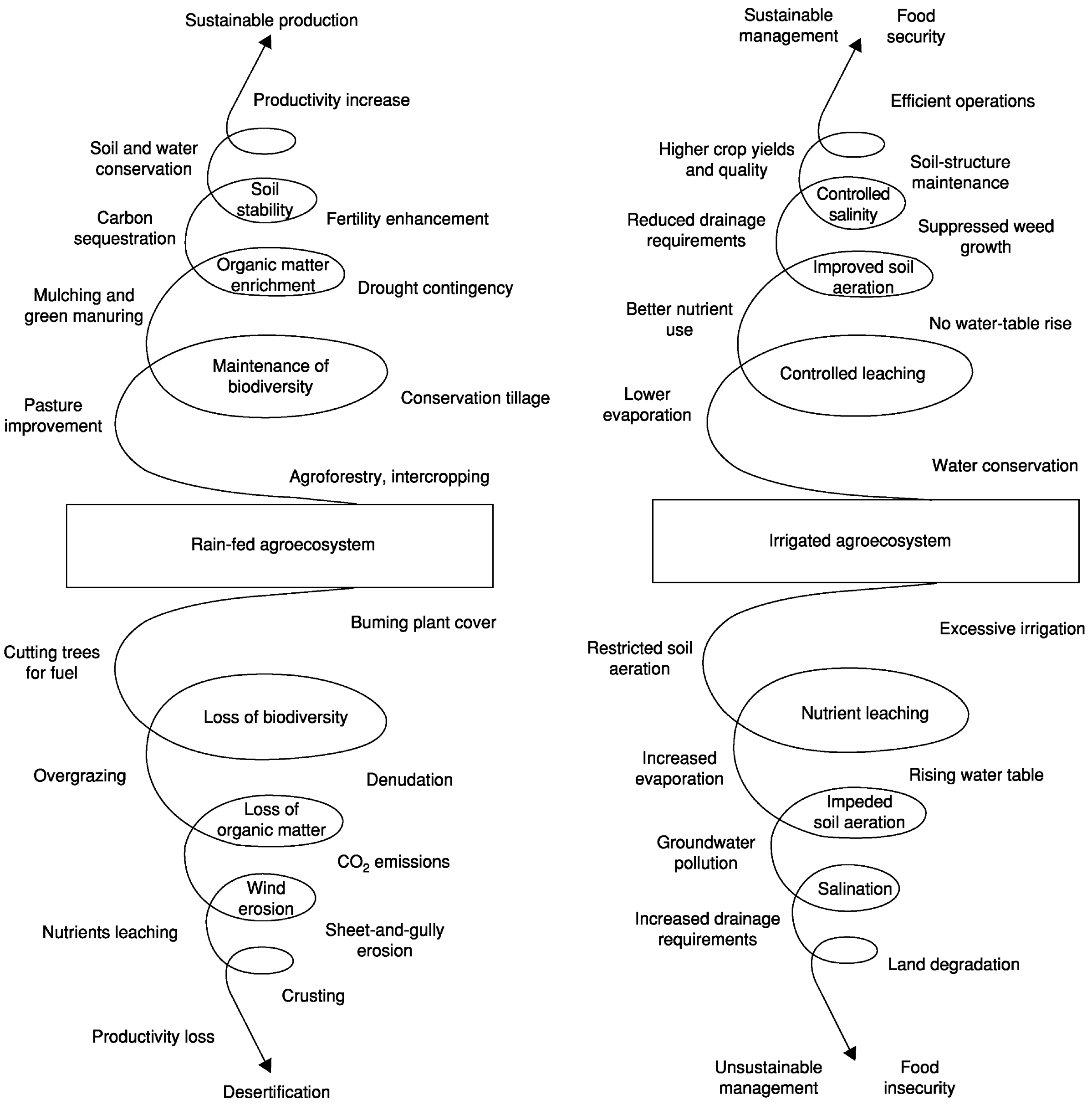


Figure 2 Upward and downward spirals of sustainable versus unsustainable patterns of management in rain-fed and irrigated agriculture in arid regions. Reproduced with permission from Hillel D and Rosenzweig C (2002) Desertification in relation to climate variability and change. *Advances in Agronomy* 77: 1–38.

To redress or rehabilitate degraded ecosystems, vulnerable countries are beginning to institute appropriate policies and programs. These include keeping reserve areas to protect biodiversity, avoiding overgrazing on managed lands, reseeding of pastures, implementing soil and water conservation measures, and – in the social arena – land tenure, family planning, and contingencies for droughts. Determining the availability of fresh water resources (surface water, renewable groundwater, and, in some cases, nonrenewable groundwater as well) and planning their careful utilization are important components of such programs. Inappropriate patterns of management may lead to a downward spiral of ecosystem degradation, whereas appropriate measures of conservation and sustainable use hold the promise of sustainable development in the context of both rain-fed and irrigated agriculture (Figure 2).

See also: **Degradation**; **Erosion:** Water-Induced; Wind-Induced

Further Reading

Bouwman AF (ed.) (1992) *Soil and the Greenhouse Effect.* Chichester, UK: John Wiley.

Charney JG (1975) Dynamics of deserts and drought in the Sahel. *Quarterly Journal of the Royal Society* 101: 193–202.

Dregne HE (1994) Land degradation in the world's arid zones. In: Baker RS, Gee GW, and Rosenzweig C (eds) *Soil and Water: Keys to Understanding Our Global Environment*. SSSA Special Publication No. 41. Madison, WI: Soil Science Society of America.

Glantz MH and Orlovsky N (1983) Desertification: a review of the concept. *Desert Control Bulletin* 9: 15–22.

Hillel D and Rosenzweig C (2000) Desertification in relation to climate variability and change. *Advances in Agronomy* 27: 1–38.

Justice CD (ed.) (1986) *Monitoring the Grassland of Semi-arid Africa Using NOAA–AVHRR Data*. London, UK: Taylor & Francis.

Lal R (2001) Potential of desertification control to sequester carabon and mitigate the greenhouse effect. *Climatic Change* 51: 35–72.

Le Houreou HN (1977) Biological recovery versus desertization. *Economic Geography* 63: 413–420.

Mainguet M (1994) *Desertification: Natural Background and Human Mismanagement*. Berlin, Germany: Springer-Verlag.

McCarthy JJ, Canziani OF, Leary NA, Dokken DJ, and White KS (eds) (2001) *Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press.

Nicholson S (2000) Land surface processes and Sahel climate. *Reviews of Geophysics* 38: 117–139.

Rind D, Rosenzweig C, and Peteet D (1989) African drought: history, possible causes, prognosis. In: Lemma A and Malaska P (eds) *Africa Beyond Famine*. London, UK: Tycooly.

Tucker CJ, Dregne HE, and Newcomb WW (1991) Expansion and contraction of the Sahara desert from 1980 to 1990. *Science* 253: 299–301.

UNEP (1992) *World Atlas of Desertification*. Seven Oaks, UK: Edward Arnold.

DIFFUSION

T M Addiscott, Rothamsted Research, Harpenden, UK
P Leeds-Harrison, Cranfield University, Silsoe, Bedford, UK

Introduction

In soil science the diffusion of gases and solutes needs to be understood. A gas in a closed vessel distributes itself such that its pressure is the same at all points. If there is a mixture of gases, each exerts its own 'partial pressure' and distributes itself such that its partial pressure is the same throughout the vessel. If you change the partial pressure of one gas in any part of the vessel there will be a gradient in its partial pressure, and the resulting flow of that gas within the other gases will tend to equalize its partial pressure throughout the volume. A solute in a solution behaves in several ways like a gas in a closed vessel. It has a chemical potential, which depends on its concentration and is conceptually related to the partial pressure of a gas through the thermodynamic concept of free energy. The solute distributes itself so that its concentration and chemical potential are the same throughout the solution, as a gas does with its partial pressure in a closed vessel. Changing the concentration in any part of the solution leads to a gradient in chemical potential which causes a flow of solute that tends to equalize the concentration. You can readily watch this happening by placing a crystal of potassium permanganate at the bottom of a beaker of water. There is a gradient of chemical potential between the solution adjacent to the crystal and the rest of the water, and the rich purple color of the permanganate gradually spreads out into the water. This process, which is essentially the same for gases and solutes, is diffusion, which involves gases that diffuse in the soil atmosphere, that is, within other gases, and solutes that diffuse in the soil solution, often interacting with soil solids. Although diffusion in solution is caused in principle by gradients in chemical potential, the topic is usually treated in terms of the concentration.

How Diffusion Occurs

Diffusion occurs because of Brownian motion, the random movement of ions or molecules in a state of thermal motion. Although the molecules move randomly, the probability is that gas molecules will move from high-pressure to low-pressure zones, and solute molecules from zones of high concentration to zones of lower concentration in a solution. Because gradients in partial pressure and concentration are gradually lessened by diffusion but never eliminated, partial pressures and concentrations of solute become uniform only at infinite time (although in practice many diffusive flows become too small to measure after a relatively short time). This is why many diffusion equations include the concentration at infinite time C_∞ or the mass diffusing in infinite time M_∞.

Fick's Laws

The first mathematical description of diffusion was given by Fick in 1885. He based his ideas on the mathematical treatment of the conduction of heat given by Fourier in 1822. Fick's equations are similar to other flux equations in which the flux is proportional to the gradient in some property of the system. The first law of Fick is stated in mathematical form as:

$$F = -D\frac{\partial C}{\partial x} \qquad [1]$$

where F is the amount of diffusing substance passing perpendicularly through a reference surface in unit time per unit cross-sectional area, C is the concentration, x is the space coordinate, and D is the diffusion coefficient, which has dimensions of L^2T^{-1} (a surface area per unit time). The value of D depends on the units of distance and time, but is independent of the units in which C is measured. In eqn [1], D is often assumed constant. However, in some circumstances it depends on the concentration of diffusing substance in the medium.

A quantitative prediction of the rate at which a diffusion process occurs can be obtained from eqn [1] when the diffusion coefficient is known. The equation expressing the change in concentration with time is obtained by combining the concept of diffusion with that of continuity:

$$\frac{\partial C}{\partial t} = -\frac{\partial F}{\partial x} \qquad [2]$$

to give the transient state equation in one dimension:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left[D\frac{\partial C}{\partial x}\right] \qquad [3]$$

When the diffusion coefficient D is constant, eqn [3] becomes the second law of Ficks:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \qquad [4]$$

Initial and Boundary Conditions

To solve eqns [1–4] for real systems, initial and boundary conditions must be known for the system under study. Very often we set our initial conditions to those of uniform concentration throughout the system or within a subsystem. A perturbation at a boundary at time zero then sets up concentration gradients and fluxes of diffusing material, and concentrations within the system can be found at subsequent times.

The boundary conditions of the region of interest are of critical importance. The boundary conditions most usually encountered are:

1. The Dirichlet or concentration condition. Here the concentration is maintained constant at the boundary of the system. As an example, when considering oxygen diffusion into soil, such a boundary occurs at the soil surface, where the concentration of oxygen in the air is constant and equal to that of the atmosphere;

2. The Neuman or flux condition. For the case where a flux density, $F(t)$, is imposed for $t \geq 0$ on a boundary, the Neuman condition is written (after eqn [1]):

$$-D\frac{\partial C}{\partial x} = F(t)$$

When the flux at a surface is zero, as occurs at impermeable barriers, the Neuman boundary condition is:

$$\frac{\partial C}{\partial x} = 0$$

3. The Robin or radiation condition. For the case where the flux across the surface is proportional to the difference between the surface concentration, C_s, and the surrounding medium concentration, C_0, the boundary condition is:

$$-D\frac{\partial C}{\partial x} = \beta(C_s - C_0)$$

where β is a proportionality constant.

Often the complexity of real soil systems means that some simplifying assumptions are made about the nature of the boundaries so that analytical solutions approximating to the real situation can be found. These may include the simplification of the surface geometry, for instance, assuming aggregates of soil to be spherical, or assuming the solute concentration in macropores surrounding an aggregate to be constant. Interpretation of the results from experiments on diffusion in soils within the framework of theoretical analysis always needs to take account of the assumptions that have been made about initial and boundary conditions.

Components of the Diffusion Coefficient in Soil

The diffusion coefficient can in principle be measured in aqueous solution, although the practice is not always simple. Soil introduces further complications, particularly if the diffusing substance is a solute which is sorbed by the soil. Nye and his colleagues developed a theory for assessing the components of the diffusion coefficient of such a solute. They started with the assumption that both the solid and liquid phases contributed to the diffusion coefficient, but experiments with chloride, phosphate, sodium, and strontium ions suggested that the contribution by the solid phase was not important. There was a suggestion that exchangeable (in contrast to solution) sodium made a contribution to the diffusion constant for sodium but only at the lowest water contents used. Exchangeable strontium seemed to make no contribution, and neither did sorbed phosphate. This suggested that for most situations likely to occur in practice all diffusive flow occurred in the solution phase.

The relation between the diffusion constant in soil D and that in free solution D_l has to take account of three soil characteristics:

1. The fraction of the area perpendicular to the direction of flow in which diffusive flow occurs. This is the same as the fraction of the soil volume occupied by the solution v_l and it is dimensionless;
2. The impedance factor f_l which allows for two factors, the tortuous path that diffusive flow has to follow round soil particles and the probability that some pathways are 'dead ends.' For anions, both factors are likely to be accentuated by anion exclusion (the repulsion of anions from the negatively charged surfaces that predominate in nonacid soils). The impedance factor is also dimensionless;
3. The slope of the isotherm relating the solute concentration C_l in solution to the concentration in the whole soil C. The isotherm is usually taken to be linear.

Combining these factors gives the following:

$$D = D_l v_l f_l (C_l/C) \qquad [5]$$

Of these components of the diffusion coefficient, D_l will usually be known and v_l and (C_l/C) can usually be measured. The impedance factor f_l cannot usually be measured independently and often has to be obtained by measuring D_l when the other components are known. It might be possible to estimate it from acoustic or conductivity measurements.

Diffusion and Geometry

The nature and mathematical treatment of diffusion depend on the shape of the volume into or from which the diffusion occurs. Analytical solutions of eqn [4] have been developed for various geometric shapes, which are represented by the appropriate boundary conditions. Although these solutions can be applied only to simple geometries and constant diffusion coefficients, they have proved widely useful. The mathematical treatment for a sphere is presented below. That for a cylinder is somewhat similar.

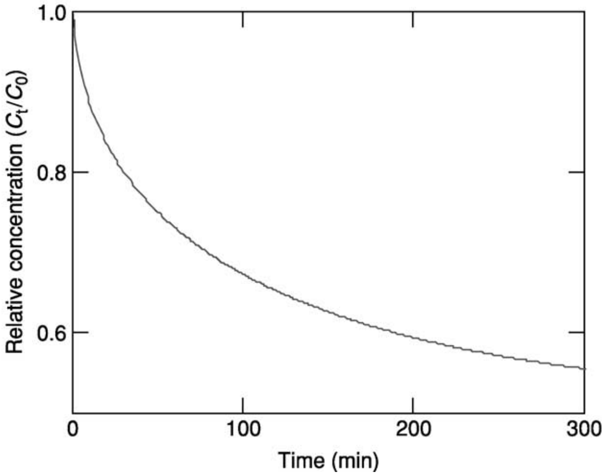


Figure 1 Change in concentration of solute within an aggregate as a function of time.

Diffusion in a sphere needs to be described in spherical coordinates r, θ, ϕ:

$$\frac{\partial C}{\partial t} = \frac{D}{r^2}\left\{\frac{\partial}{\partial r}\left(r^2\frac{\partial C}{\partial r}\right) + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial C}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 C}{\partial\phi^2}\right\} \qquad [6]$$

Here we assume that the concentration of solute within the sphere is C_0 and the solution is free from solute. If the solution is well stirred, the concentration of the solution depends only on time, while the total amount of solute in the solution and sphere remains constant.

The time rate of change of solute concentration, C, within the sphere (Figure 1) is described by eqn [5], and, since there is only variation of concentration radially, we have:

$$\frac{\partial C}{\partial t} = D_b\left(\frac{\partial^2 C}{\partial r^2} + \frac{2}{r}\frac{\partial C}{\partial r}\right) \qquad [7]$$

where r is the radial coordinate in the sphere, since the concentration depends on the radius only and is independent of both angles, θ and ϕ.

Some Useful Analytical Solutions for Various Geometries

Diffusion from a Semiinfinite Medium

A semiinfinite medium is notionally one which is large enough for diffusion to occur from it for infinite time without becoming limited by the length dimension. It is semiinfinite because if it was completely infinite it would have no face across which diffusion could occur. Neither C_∞ nor M_∞ appears, and the solution for the mass M_t diffusing from the medium in time t is given by Crank as:

$$M_t = 2C\sqrt{\frac{Dt}{\pi}} \qquad [8]$$

where C is concentration of the diffusing substance within the medium, and D is again the diffusion coefficient. Eqn [8] provides an example of the general rule that all diffusion processes are characterized by a relation to the square root of time.

Diffusion in a Cylinder

Only radial diffusion is considered. The cylinder is assumed to be long enough for diffusion along it to be ignored. In the solution supplied by Crank, the mass M_t diffusing in time t from a cylinder of diameter a is given by the series:

$$\frac{M_t}{M_\infty} = \left(\frac{4}{\pi^{1/2}}\right)\left(\frac{Dt}{a^2}\right)^{1/2} - \frac{Dt}{a^2} - \left(\frac{1}{3}\pi^{1/2}\right)\left(\frac{Dt}{a^2}\right)^{3/2} + \ldots \qquad [9]$$

where D is the diffusion coefficient. If the series converges sufficiently rapidly for the third and higher terms to be ignored:

$$\left(\frac{1}{t}\right)\frac{M_t}{M_\infty} = \left(\frac{4}{\pi^{1/2}}\right)\left(\frac{D}{a^2 t}\right)^{1/2} - \frac{D}{a^2} \qquad [9a]$$

Plotting $(1/t)(M_t / M_\infty)$ against $t^{-1/2}$ should enable D and a to be estimated from the slope and intercept of the resulting relation.

Diffusion in a Sphere

Crank has supplied an analytical solution to eqn [6] in which the mass of solute diffusing from a sphere into a well-stirred solution of limited volume at a constant concentration is:

$$\frac{M(t)}{M_\infty} = 1 - \sum_{n=1}^{\infty} \frac{6}{\pi^2 n^2} \exp\left(\frac{-9n^2\pi^2 Dt}{a^2}\right) \qquad [10]$$

If we apply the following initial and boundary conditions to eqn [6]:

$$C = C_0 \qquad 0 \le r \le a, t = 0$$

$$C_M(t) = 0 \qquad r = a, t \ge 0$$

$$C = C_M(t) \qquad r = a, t \ge 0$$

where $C_M(t)$ is the concentration in the solution external to the sphere at time t and a is its radius. An analytical solution for the above diffusion equation for uniform size spheres is given by Crank:

$$\bar{C}(t) = C_\infty + (C_0 - C_\infty) \sum_{n=1}^{\infty} \frac{6\alpha(1+\alpha)\exp(-D_b q_n^2 t/a^2)}{9 + 9\alpha + q_n^2\alpha^2} \qquad [11]$$

where C_∞ is the equilibrium concentration at infinite time and q_n is the positive root of:

$$\tan q_n = \frac{3q_n}{3 + \alpha q_n^2} \qquad [11a]$$

and $\alpha = V_M/V_A$, the ratio of the volume of external solution, V_M, to that of the sphere, V_A.

Simple Model for Various Geometries

The analytical approaches outlined above are limited in their applicability by their inability to deal with externally imposed changes in concentration. For example, diffusion from spherical aggregates into the interaggregate solution is likely to be interrupted in the real world by rainfall, which changes the concentration of the solution. The theory cannot allow for this. Furthermore, few soils comprise uniformly sized aggregates. Most aggregate sizes are log-normally distributed. Analytical approaches typically deal with this problem by using a single aggregate size computed on a volume-weighted basis. This, however, is not really adequate because of the interaction between aggregates of different sizes. Small aggregates lose, for example, 50% of their solute very much more rapidly than larger aggregates. This is taken into account in the volume-weighting process, but what is not accommodated is that the solute lost rapidly from the smaller aggregates decreases the concentration gradient determining diffusion from the larger ones.

A simple nonanalytical model was developed for cubes and other regularly shaped aggregates. This is based on a cube which is divided into concentric equal volumes. The model uses the first law of Ficks to compute the flux between adjacent volumes and between the outermost volume and the external solution. The model can be used without amendment for diffusion in a sphere and with minor amendments for a regular tetrahedron or a rhombic octahedron. Shapes with larger numbers of faces are unlikely to be more relevant than a sphere for simulating soil aggregates. This model can accommodate changes imposed from outside the system in the concentration of the external solution – or changes in its volume. It can also cope with normal or log-normal distributions of aggregate sizes. Its main disadvantage is that it is less exact than the analytical solutions.

Applications in Soil Science

Plant Nutrition

Many studies have been made on the mechanisms by which plant nutrients reach roots. Three mechanisms are postulated: diffusion, convection (mass flow), and root extension. The dominant mechanism depends on the nutrient studied and the experimental conditions. The nutrients with the largest concentrations in the soil solution, often calcium and nitrate, tend to be carried to the roots in sufficient quantities by convection. Where convective flow fully supplies the needs of the plant, no appreciable concentration gradient builds up and no diffusion occurs. Nutrients such as potassium and phosphate strongly held by the soil and therefore at low concentration in the soil solution are less likely to be supplied adequately by convection. This can result in a low concentration at the root surface and the development of a concentration gradient which makes diffusion the main transport mechanism. Diffusion to a root is often studied using the theory for diffusion in a cylinder and considers a cylinder of soil around the root. This is satisfactory for isolated roots, but roots rarely exist in isolation from other roots, and dealing with interacting cylinders is far more difficult. This is part of a general problem: diffusion to roots is difficult to measure other than in specially designed systems that are somewhat artificial. The studies have generally suggested that root extension is not important by comparison with the other mechanisms, but this result could reflect the fact that they were made at a fairly small scale, usually in pots in a glasshouse. Root extension could have been more important at the field scale.

Kinetics of Potassium Release From Clays

The release of nonexchangeable potassium from illites and related clays has been treated as very slow diffusion in a cylinder, with a diffusion coefficient of the order of 10^{-18}–$10^{-20}\,\text{cm}^2\,\text{s}^{-1}$. Eqns [9] and [9a] are used for this purpose.

Gaseous Diffusion

Air occupies approximately one-third of the volume of the soil but is not very free to move. There is therefore little convective movement of gases in the soil. Because of the consumption of oxygen by aerobic microbes and the generation of carbon dioxide by most microbes, there can be substantial gradients in the partial pressures of these gases. Diffusion therefore plays an important role in the transport of these gases and also of dinitrogen and nitrous oxide when denitrification occurs. We are often concerned with diffusion through the whole soil or, more correctly, through the larger pores of the soil. But diffusion of oxygen within aggregates is also important, because it determines whether or not anoxic zones develop at the centers of the aggregates, with the consequent risk of denitrification and other anoxic processes. This oxygen diffusion is usually treated from a theoretical standpoint as diffusion within spheres.

Leaching

The nature of solute leaching depends greatly on the texture and structure of the soil. In soils in which there is appreciable aggregation, water moves to a large extent around the aggregates and solutes within the aggregates enjoy temporary protection from leaching. This only lasts until they diffuse out of the aggregates into the water flowing round them. Leaching and intraaggregate diffusion therefore need to be simulated together. This has been done using the theory for diffusion in spherical aggregates and also the simple model for various geometries. As noted above, the latter is better able to cope with a distribution of aggregate sizes but is less exact.

Of particular interest in leaching is the behavior of pesticides. Many pesticides partition between the solid (soil particles, particularly organic matter and clays) and the liquid in the pores. To describe the movement of reactive solute, a retardation factor, R, is introduced which accounts for the partitioning of the solute. It is defined as:

$$R = 1 + \frac{\rho K_\mathrm{d}}{\theta}$$

where K_d is the partition coefficient describing the ratio of solute sorbed to solid particles to the solute concentration in solution at equilibrium, ρ is the bulk density of the soil, and θ is the water content. The effective diffusion coefficient, D_eff, is then defined as $D_\mathrm{eff} = (D/R)$.

List of Technical Nomenclature

Diffusion coefficient	Proportionality coefficient required by Fick's law. Dimensions L^2T^{-1}
Impedance	Factor allowing for tortuous path of diffusion, including dead-end pores. Dimensionless

Reference area Area perpendicular to direction of flow in which diffusion occurs. For solutes this is multiplied by the fraction of the soil volume occupied by the soil solution to give the area across which diffusion occurs, and for gases by the fraction occupied by air. Dimensions L^2

Further Reading

Addiscott TM (1982) Simulating diffusion within soil aggregates: a simple model for cubic and other regularly shaped aggregates. *Journal of Soil Science* 33: 37–45.

Addiscott TM, Armstrong AC, and Leeds-Harrison PB (1998) Modeling the interaction between leaching and intraped diffusion. In: Selim HM and Ma L (eds) *Physical Nonequilibrium in Soils. Modeling and Application*, pp. 223–241. Chelsea, MI: Ann Arbor Press.

Crank J (1975) *The Mathematics of Diffusion*, 2nd edn. Oxford, UK: Oxford Scientific Publications.

Currie JA (1961) Gaseous diffusion in the aeration of aggregated soils. *Soil Science* 92: 40–45.

Hartley GS and Graham-Bryce IJ (1980) *Physical Principles of Pesticide Behaviour*, vol. 1, *The Dynamics of Applied Pesticides in the Local Environment in Relation to the Biological Response*, ch. 3. London, UK: Academic Press.

Nye PH and Tinker PB (1977) *Solute Movement in the Soil–Root System*. Oxford, UK: Blackwell.

Olsen SR and Watanabe FS (1963) Diffusion of phosphorus as related to soil texture and plant uptake. *Soil Science Society of America Proceedings* 27: 648–653.

Rao PSC, Jessup RE, Rolston DE, Davidson JM, and Kilcrease DP (1980) Experimental and mathematical description of nonadsorbed transport by diffusion in spherical aggregates. *Soil Science Society of America Journal* 44: 684–688.

Rowell DL, Martin MW, and Nye PH (1967) The measurement and mechanism of ion diffusion in soils. III. The effect of moisture content and soil solution concentration on the self-diffusion of ions in soils. *Journal of Soil Science* 18: 204–222.

Smith KA (1980) A model of the extent of anaerobic zones in aggregated soils, and its potential application to estimates of denitrification. *Journal of Soil Science* 31: 263–277.

DISINFESTATION

A Gamliel, The Volcani Center, Bet Dagan, Israel
J Katan, The Hebrew University of Jerusalem, Rehovot, Israel

Introduction

Soil disinfestation is one of the most effective means of controlling soilborne pests and improving plant health. Soil disinfestation is a drastic means applied to soil before planting in order to reduce or eliminate soilborne pests. Effective soil disinfestation aims to promote healthy and productive crops. All crops are sensitive to one or more harmful biotic or abiotic soilborne agents that affect plant health and productivity. The incidence of these agents increases with frequent cropping, especially under monoculture. The biotic agents consist of major and minor pathogens, including fungi, bacteria, nematodes, soil viruses, arthropods, parasitic plants, and weeds. These are referred to here as 'pests.' The abiotic agents include, among others, the accumulation of harmful chemicals from various sources, deficiencies in essential mineral nutrients, and deterioration in physical status of the soil. In addition, the population of beneficial microorganisms such as mycorrhizae and plant growth-promoting rhizobacteria may also decrease with continuous cropping. These harmful effects are reflected in either typical disease symptoms and eventually plant death or in poor plant health and growth retardation. The latter phenomena (which are especially connected with continuous cropping of the same crop in the same plot) are also referred to as soil 'fatigue,' soil 'sickness,' and 'replant disease' (the latter typical in tree plantations). Soil disinfestation, crop rotation, and specific treatments are potential tools for controlling these phenomena, improving plant health, and resuming soil productivity.

The increase in the incidence of harmful biotic and abiotic agents with a long history of cropping stems from the fact that frequent planting of the same crop causes the enrichment and consequently rapid buildup of detrimental biological, chemical, and physical agents in the soil. Soil disinfestation aims to control these agents and thereby improve plant health, and reestablish high and sustainable levels of yield. These goals are especially important in regions where both crop options and agricultural land are limited. Therefore soil disinfestation can be regarded as soil reclamation.

Disinfestation in its present form was established at the end of the nineteenth century, in the early stages of the establishment of crop-protection sciences. The spread of the soil pest *Phylloxera* in vineyard soils in France during that period led to the introduction of carbon disulfide (CS_2), the first soil fumigant, for soil disinfestation. A physical means of soil disinfestation, steaming, was also developed. Approximately 100 years later, a third approach for soil disinfestation, 'soil solarization,' was developed.

Principles of Soil Disinfestation

The main goal of soil disinfestation is to eradicate, usually before planting, harmful, soilborne biotic agents, uniformly to the desired depth, with minimal disturbance of the biological equilibrium and with minimal effect on chemical or physical soil properties. This has to be achieved using an effective, feasible, economic and environmentally acceptable technology. With respect to the spectrum of pest control, we face a dilemma: wide-spectrum pest control is economically and practically desirable, but usually involves the use of means that can be biologically destructive. This is the case when soil disinfestation is carried out using drastic physical or chemical means to ensure the control of the harmful biotic agents in the deeper soil layers, as well as in all other soil niches. Therefore, soil disinfestation has to be carried out before planting to avoid harm to the crop. Also, the control agent, or its decomposition product, has to dissipate before planting to avoid residual harmful effects. Soil disinfestation was initially developed for the control of harmful biotic agents, although possible effects on abiotic soil components, e.g., improvement of mineral nutrition status of the plant, should not be excluded.

The basic principles of soil disinfestation are:

1. Treating the soil with a wide-spectrum control agent in order to eliminate (or strongly reduce) populations of a variety of soilborne pests;
2. Causing minimal disturbance to the biotic (especially beneficial microorganisms) and abiotic components of the soil;
3. Effectively reaching and controlling the pests with the control agent to the desired soil depth, usually to 30–50 cm, or even deeper in some cases. In soil-less culture, this layer is shallow;
4. Using most of the existing technologies, treating the soil before planting;
5. Avoiding contamination from outside sources. soil disinfestation can only control the existing populations of the soil pests.

Methods of Soil Disinfestation

Physical

Steaming, aerated steam, and hot-water treatments are used in greenhouses, especially with container (growth) media. Steam has been utilized for soil disinfestation for more than a century. Resting structures (i.e., cells or structures of microorganisms that can survive for a long period in soil) of plant pathogens and other pests, e.g., chlamydospores, sclerotia, oospores, and bacterial spores, are eliminated by steaming if heated to lethal temperatures, even in cases of heavy soil contamination. Moreover, steaming frequently has a growth-stimulating effect on the subsequent crop.

Careful soil preparation is essential for good steam penetration. The soil should be tilled as deeply as possible and then left to dry completely before steaming. It is important to reduce the amount of plant debris. Good preparation permits good steam penetration and enables pest control even in heavy soils. The steaming of aerated growth substrates is usually effective, but peat poses difficulties owing to its high water content.

The soil is steamed by either 'passive' or 'active' techniques. In passive steaming, the steam is blown to the surface, under a covering sheet, and left to heat the upper layer. Lower layers are then heated by heat transmission. This process continues until 100°C is reached at a depth of 10 cm. Disinfestation of deeper layers, especially in sandy soil, may only be partial.

Active steaming can be performed using either positive or negative pressure. Both techniques employ drainage systems, based on pipes laid at 50–70 cm depth, approx. 80 cm apart. With the positive-pressure technique, the steam is blown through holes located along the pipes. The negative pressure involves an improved technique, utilizing the advantages of the other two application methods. As in passive steaming, the steam is released over the treated area under plastic sheeting, ensuring rapid and even distribution over the entire surface of the plot, followed by active suction into the deeper layers of the soil, achieved by negative pressure applied through the drainage system. This technique, widely used in the Netherlands, is cheaper than the other two, due to the energy savings incurred by the faster heat transfer. Nevertheless, steaming treatments are expensive and are feasible mainly in places where there are heating systems (for heating the greenhouse during the cold season) or if applied by contractors. Steaming, however, can be useful and economic for the disinfestation of shallow layers of growth media, such as are usually found in nurseries.

Chemical

Soil fumigation is achieved by applying toxic pesticides to the soil by various means, and these fumigants move down and across the soil profile to reach the target organisms, either directly or by very efficient secondary distribution due to their relatively high vapor pressure.

Methyl bromide The most powerful soil fumigant available is methyl bromide (MBr), with a very broad spectrum of activity. Many soilborne fungi, nematodes, and weeds are sensitive to MBr; in contrast, some soilborne bacteria such as *Clavibacter michiganensis* sp. *michiganensis*, are not satisfactorily controlled at regular (commercial) rates of application. The duration of the application depends on soil temperature (1–2 days at 15°C, in the 0- to 20-cm soil layer, 3 days at 10–15°C, and more than 4 days at 8–10°C at the same depth). Possible problems due to the toxicological hazard of MBr are related mainly to health hazards for the applicators and to the increase in inorganic bromine residues in edible plant products. In a few cases, MBr has been found in water near greenhouses in the Netherlands, where PVC water pipes were improperly placed only 10 cm deep in the ground. Concerns regarding the possible role of MBr in ozone depletion and the forthcoming phase-out of MBr have triggered research efforts to develop alternative chemical and nonchemical methods for soil disinfestation.

Metham sodium Effective against several soil-borne pathogens in both covered and open outdoor cultivation, metham sodium (MS, sodium methyldithiocarbamate) is a fumigant that generates methylisothiocyanate (MITC). In water solutions, MS rapidly changes to MITC. The spectrum of control includes fungi, free-living nematodes, and some weeds. Since most pest resting structures are present in the upper 40 cm of the soil profile, and since MS is 100% water-soluble, it is most effective when applied via the irrigation system (chemigation). However, special care should be taken to avoid contamination of the main irrigation system with MS. The chemical is used at various doses according to the target pathogen and/or the soil type to be treated. Soil temperature is also a critical factor in the effective application of the chemical: a range of 18–30°C at a soil depth of 10 cm is best.

Dazomet (3,5-dimethyltetrahydro-1,3,5,(2H)-thiodiazinothione) A second fumigant that generates MITC is dazomet, a product formulated as either a powder or granules. The chemical is gradually hydrolyzed to at least four subproducts, MITC being the main one. Dazomet is effective against various fungal pathogens at a rate of 400–600 kg a.i. ha^{-1}. This fumigant can be used for the control of several diseases in seed beds and greenhouses, or in field-grown vegetables, cotton, tobacco, and ornamentals. It is applied to the soil by spreading or irrigating followed by mechanical mixing (such as rotovator cultivation or shovel plough) into the soil. The chemical, which is not applicable at temperatures lower than 8°C, is also partially effective against insects, various nematodes, and weed seeds. One of the disadvantages of dazomet is the long period (3 weeks) needed after application of the chemical before planting or sowing is permissible.

Other fumigants and mixtures Fumigants other than those listed, having a narrower range, are registered and used in various cropping systems. These include nematicides such as 1,3-dichloropropene (1,3-D), fungicides such as chloropicrin (CP), and bactericides such as formlin. These are still in use on relatively small scales. Recently, the potential of iodomethane as soil fumigant has also been examined. It is clear that, with the currently available fumigants, there is no satisfactory replacement for MBr. The use of other fumigants involves identification of the causal agent and, in many cases, the use of a mixture of two or more chemicals, to control a wider range of disease agents, pests, and weeds in the treated plots. Methylisothiocyanate 20 + dichloropropane-dichloropropene 80 may serve as an example of this type of product, as this pesticide was formulated to control both pests controlled by MS and the root-knot nematode. Furthermore, data regarding residual effects of these alternative fumigants before planting are needed and their environmental impact is not yet fully clear.

Application Soil fumigants should be applied to well-prepared soil before planting. Most chemicals are injected into the soil in a liquid form to the desired depth with special application machinery. The chemical vaporizes in the soil and diffuses to reach every niche in which pests exist. Application of a fumigant through an irrigation system with delivery via the water to deep soil layers is also common. Such application, however, requires special formulation of the fumigants.

A fumigated field is usually covered with plastic mulch following fumigation to minimize gas escape. Standard polyethylene films are permeable to fumigants and the fumigants dissipate quickly by escaping through the film shortly after application. The permeability of fumigants such as MBr, MITC, CP, and 1,3-D through impermeable film is only

0.001–0.0001 g m^{-2} per hour, depending on the barrier formula, compared with an emission of 5 g m^{-2} per hour for regular, low-density polyethylene. Pest control is determined by the factors of pesticide concentration (C) and exposure time (T). Thus, extending fumigant retention in the soil under impermeable films for a longer period allows the use of reduced fumigant dosages with the same CT values, without reducing control efficacy. Further reduction is possible by burying the film edges more deeply in the soil and by continuous mulching. Machinery for the continuous mulching of plastic film over a large area is in commercial use.

Soil Solarization

The basic idea of soil solarization is to heat the soil by means of solar energy, e.g., by mulching it with transparent polyethylene under the appropriate climatic conditions, thereby killing soilborne pests, improving plant health, and consequently increasing yields. Under the appropriate conditions, the results obtained by soil solarization can be comparable with those obtained by the widely used chemical disinfestation. The first publication on the method of soil solarization was in 1976, although ideas on using solar heating in crop protection have been known for centuries.

The term 'solarization' has several meanings. "To solarize" is defined in *Webster's Third New International Dictionary* as "to expose to sunlight; to affect or alter in some way by the action of the sun's rays." Although many terms have been used to describe this process since 1976 – solar heating of the soil, polyethylene or plastic mulching (tarping), solar pasteurization, and solar disinfestation – here the term 'soil solarization,' is preferred (which was introduced by American plant pathologists), because it is both widely accepted and concise. In addition, 'soil solarization' implies an active process of solar heating of the soil, rather than the usual passive heating of a soil exposed to sunlight.

Principles of soil solarization The principles of soil solarization are summarized as follows:

1. Solarization heats the soil through repeated daily cycles. At increasing soil depths, maximal temperatures decrease, are reached later in the day, and are maintained for longer periods (Figure 1);
2. The best time for soil mulching, i.e., when climatic conditions are most favorable, can be determined experimentally by tarping the soil and measuring the temperatures. Meteorological data from previous years and predictive models further aid in this task;

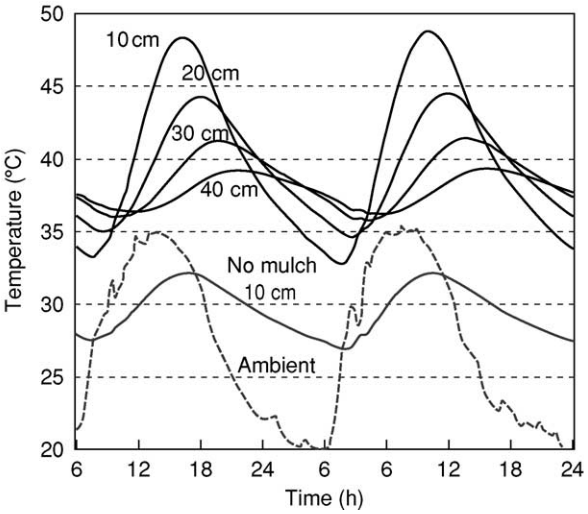


Figure 1 The daily course of soil heating by polyethylene at three soil depths, as compared to nonsolarized (no mulch) soil at a depth of 10 cm. Typical results obtained during July–August in Rehovot, Israel.

3. Adequate soil moisture during solarization is crucial to increase the thermal sensitivity of the target organisms, improve heat conduction in the soil, and enable biological activity during solarization. The soil can be moistened by a single irrigation shortly before tarping. Additional irrigation during solarization via drip system or furrow irrigation is usually not necessary, except for very light soils; in addition to which it may reduce soil temperatures unless carried out during the night;
4. Proper preparation of a soil ready for planting is essential. This is the case because, after plastic removal, the soil should be disturbed as little as possible to avoid recontamination;
5. The soil is mulched with thin, transparent polyethylene sheets or other plastic material. Another method of solarization involves a closed glasshouse (or plastic house), provided climatic conditions are suitable and the soil is kept wet. Novel technologies such as the use of sprayable plastics can replace plastic mulching of the soil;
6. Successful pathogen control in various regions of the world is usually obtained within 20–60 days of solarization. Extending the solarization period enables control in deeper soil layers, as well as of pathogens that are less sensitive to heat;
7. Solarization causes chemical, physical, and biological changes in the soil that affect pest control, plant growth, and yield.

Although both solarization and artificial soil heating involve soil heating, there are important biological and technical differences between these two methods of soil disinfestation. With soil solarization, there is no need to transfer the heat from its source to the field. It can therefore be carried out directly in the open field or in the greenhouse. Solar heating is carried out at relatively mild temperatures (Figure 1), as compared to artificial heating, which is usually carried out at 70–100°C; thus, the former's effects on living and nonliving soil components are likely to be less drastic. Indeed, negative side-effects observed with soil steaming in certain cases, e.g., phytotoxicity due to the release of manganese or other toxic products and rapid soil reinfestation due to the creation of a 'biological vacuum,' have rarely been reported with solarization. This term refers to a situation where microbial populations are much reduced, resulting in an unbalanced population of soil microflora. Nevertheless, the possibility of the occurrence of such negative side-effects should not be excluded *a priori*. Under appropriate conditions, many soilborne pathogens such as fungi (e.g., *Verticillium, Fusarium, Phytophthora*, *Pythium*, *Pyrenochaeta*), nematodes (e.g., *Pratylenchus* and *Ditylenchus*), and bacteria (e.g., *C. michiganensis*), as well as a variety of weeds, especially annuals, are controlled by soil disinfestation and consequently yields are increased.

As with any soil disinfestation method, soil solarization has advantages and limitations. It is a nonchemical method with less drastic effects on the biotic and abiotic components of the soil; it is simple (and is therefore suitable for both developing and developed countries); and it is frequently less expensive than chemical soil disinfestation. The limitations of this method stem from its dependence on climate and it can therefore be used only in certain climatic regions and during limited periods of the year. In addition, during solarization, the soil remains without a crop for several weeks. Nevertheless, this method has attracted many researchers in more than 60 countries and it is used by farmers, especially in combination with other methods.

Soil heating: simulation models for the prediction of soil temperatures The principles of soil heating and the energy balance of bare and mulched soils have been described in various publications. The main factors involved in soil heating are climatic (e.g., solar radiation, air temperature, air humidity, and wind speed), soil properties, and photometric and physical characteristics of the mulch. The models described below have been developed by Mahrer (The Hebrew University of Jerusalem) and his coinvestigators. The fluxes of energy which have to be considered are:

- R_g, global radiation (waveband of approximately 0.3–4 μm);
- R_L, atmospheric (long-wave) radiation (waveband of approximately 4–80 μm);
- S, conduction of heat in the soil (soil heat flux);
- H, vertical heat exchanges with the air enclosed between mulch and soil by conduction and with the surrounding air by convection (sensible heat flux);
- E, condensation and evaporation of water (latent heat flux).

The two basic equations describing the energy balance of bare and mulched soils, respectively, are:

$$R_{sn} + R_{Ln} - H - E - S = 0 \qquad [1]$$

$$R_{snm} + R_{Lnm} - H_m - E_m - S_m = 0 \qquad [2]$$

where subscript m stands for mulched soil, and R_{sn} and R_{Ln} are the net fluxes of short- and long-wave radiation at the bare soil surface, respectively; S, H, E are soil heat flux, vertical heat exchanges, and condensation and evaporation of water, respectively.

Using these equations, a one-dimensional numerical model to predict the diurnal cycle of soil temperature in mulched and bare soils was developed by Mahrer, with good agreement between observed and predicted soil temperatures. Calculations show that increased soil temperature in wet mulched soil is mainly due to the reduction of heat loss through sensible and latent heat fluxes during the day, and partially due to the greenhouse effect of the wet cover (owing to the formation of small water droplets on its inner surface). A two-dimensional model has been used to study the spatial temperature regime in the soil. Results show that soil heating at the edge of the mulch is lower than at its center. It has also been found that a narrow strip of mulch is less effective in heating the soil than a wide one. Predicted results agree well with observations.

Combining Disinfestation Methods

Soil fumigation with chemicals may have negative effects on the environment, could be extremely dangerous to humans, and may leave toxic residues in plant products. Thus, innovative approaches are urgently needed by farmers and consumers. This can be achieved by combining fumigants with pesticides, at reduced dosages, or with nonchemical methods, e.g., solarization or alternative methods.

Combining solarization with chemicals at reduced dosages or other measures, e.g., biocontrol agents, can reduce the limitations of solarization. The control efficacy may be increased owing to additive effects or to a synergistic effect caused by the hotter environment, which increases vapor pressure and chemical activity of the added pesticide. Another reason for the improved activity of the pesticide is the weakening of the pathogenic resting structure by the heat. Solarization combined with fumigants could reduce the required duration of solarization, thus making the method more acceptable by the farmers. Furthermore, sublethal fumigation in combination with solarization is especially useful for areas that are marginal for the application of solarization.

Benefits and Limitations

Plant growth in disinfested soil is, as expected, enhanced as compared to that in untreated, infested soil, as a result of pest control. Less expected is a phenomenon of plant growth enhancement in disinfested soils in the absence of known pests, discovered at the end of the nineteenth century, which has since been repeatedly reported with all disinfestation methods, including solarization. It is possible that some cases of increased plant growth response (IGR) might be found to have resulted from the control of hitherto unknown pests, which could not be identified by the procedures available at the time. Nevertheless, different mechanisms, not related to pathogen control, have been suggested to explain IGR in disinfested soils: increased micro- and macroelements in the soil solution; elimination of minor pathogens or parasites; destruction of phytotoxic substances in the soil; and release of growth regulator-like substances, including soluble organic matter and humic substances. Stimulation of mycorrhizae, fluorescent pseudomonads, or other beneficial microorganisms has been frequently observed in solarized soils under greenhouse conditions, and to a lesser extent under field conditions. Soil solarization has been found to result in increased electrical conductivity (EC) in many treated soils, while having little or no effect on others. In saline soils, however, where the level of saline groundwater is relatively close to the soil surface, soil solarization has resulted in a significant decrease in the EC of the surface-soil extract. This is attributed to the prevention of evaporation in mulched (solarized) soil, thereby eliminating salt transport toward the soil surface and its eventual deposition.

Microbial changes take place in the soil during and after solarization. These have been studied specifically in relation to the biological control of pathogens stimulated by solarization, beyond the killing by heat. In drastically disinfested soils, a biological vacuum usually occurs. This can lead to reinfestation by pathogens, which can occur at a faster rate than in nondisinfested soils.

Thus, IGR has very important economic implications that should be taken into account when considering the use of soil disinfestation. The major difficulty is that we cannot predict whether or to what extent a soil will respond with an IGR.

The MBr Crisis and Its Implications

Soil fumigation with MBr has become the major method used for controlling soilborne pests in intensive agriculture worldwide. However, since MBr was listed under the Montreal Protocol in 1992 as an ozone-depleting substance, regulations on its use and consumption have been imposed. In developed countries, the reduction of MBr consumption was started in 1999 and, except for certain exemptions for critical uses, MBr is to be phased out by the year 2005 (or earlier in certain countries). This is being done to protect the vital ozone layer in the stratosphere from depletion. This is a crucial issue, since the ozone layer has already been depleted by a variety of substances such as chlorofluorocarbons. The impending phaseout of MBr poses new and unprecedented challenges for the agricultural research community and the authorities, since many major crops, especially in intensive agriculture, have become totally dependent on MBr use. To avoid an economic and social crisis, alternatives have to be developed relatively quickly.

During a period of more than 100 years of accelerated development in crop-protection sciences, involving the development of hundreds of highly effective pesticides, only three approaches for soil disinfestation have been developed. In practice, only a small number of fumigants have been used in any given period. We need to understand the reasons for this situation in order to avoid additional crises in the future.

There are lessons to be learnt from the MBr crisis. First, we have to avoid dependence on a single soil disinfestation method. Second, we have to combine and alternate methods of control in order to improve them, reduce pesticide dosage, and minimize-side effects. Last, but not least, disinfested soils should be continuously monitored to detect, at the earliest possible stages, negative effects. Soil disinfestation is an expensive, but highly effective, method of control and should be practiced in the best way possible from both plant protection and environmental points of view.

Further Reading

Bell CH, Price N, and Chakrabarti (1996) *The Methyl Bromide Issue*. Chichester, UK: John Wiley.

Gamliel A, Grinstein A, and Katan J (1997) Improved technologies to reduce emissions of methyl bromide from soil fumigation. In: Grinstein A, Ascher KRS, Mathews G, Katan J, and Gamliel A (eds) Improved application technology for reduction of pesticide dosage and environmental pollution. *Phytoparasitica* 25: 21–30.

Gullino A, Katan J, Garibaldi A, and Matta (eds) (2002) Chemical and non-chemical soil disinfestation. *Acta Horticulturae* 532: 9–256.

Katan J (1981) Solar heating (solarization) of soil for control of soilborne pests. *Annual Review of Phytopathology* 19: 211–236.

Katan J (1984) The role of soil disinfestation in achieving high production in horticultural crops. *British Crop Protection Conference*, pp. 1189–1196. Croydon, UK: British Crop Protection Council.

Katan J (1999) The methyl bromide issue: problems and potential solutions. *Journal of Plant Pathology* 81: 153–159.

Katan J and DeVay JE (1991) *Soil Solarization*. Boca Raton, FL: CRC Press.

Ristaino JB and Thomas W (1997) Agriculture, methyl bromide and the environment. *Plant Disease* 81: 964–978.

Runia WT (2000) Steaming methods for soil and substrates. *Acta Horticulturae* 532: 115–123.

Stapleton JJ (2000) Soil solarization in various agricultural production systems. *Crop Protection* 19: 837–841.

Wilhelm S (1966) Chemical treatments and inoculum potential of soil. *Annual Review of Phytopathology* 4: 53–78.

Dispersion *See* **Flocculation and Dispersion**

DISSOLUTION PROCESSES, KINETICS

K G Scheckel and C A Impellitteri, USEPA, Cincinnati, OH, USA

Introduction

Chemistry by its very nature is concerned with change. There are simple but significant interactions between air, water, and minerals that impact our natural environment. Minerals with well-defined structure are converted by various environmental chemical reactions into their elemental building blocks with, perhaps, differing chemical properties relative to the original crystal configuration. The influence of natural phenomena may cause minerals to dissolve in aqueous solutions, thus a solution of atoms is eventually formed, succumbing to the fundamental natural law of element-cycling. Further, these reactions are rarely at equilibrium. They proceed at varying kinetic rates as any dissolved material in solution may be removed (e.g., mineral recrystallization) or added by further mineral dissolution. Recent studies have shown that mineral-like, divalent metal surface precipitates exhibit similar dissolution behaviors to clay and oxide minerals. The examination of mineral dissolution can become even more complex when interactions of surfaces with microorganisms and charged compounds are considered that may induce reductive dissolution of redox-sensitive minerals.

The study of kinetics can be defined broadly as the rate of change of concentrations of reactants in a chemical reaction. The rates are affected by both physical (e.g., diffusion of reacting species) and chemical processes. The kinetics involved in dissolution are often ignored in environmental studies by eliminating time as a variable and assuming a state of equilibrium or ‘pseudo’ equilibrium. For example, the myriad chemical extractions utilized to assess the potential mobility of metals do so in a set time period. Generally, the rigorousness of a particular extraction acts as a surrogate for time. However, the assumptions employed may be more suitable for some soils than others because of the high variation in dissolution kinetics as a function of the minerals and compounds present. In the short term, (e.g., time of a typical soil chemical extraction) a soil containing relatively small concentrations of weakly sorbed Pb may be assessed as a higher risk than a soil containing a high concentration of galena (PbS). In the long term, as the kinetics of dissolution approach equilibrium, the actual risk may be much higher for the galena soil. Thus, understanding the fundamental rates of dissolution is necessary when pondering the comprehensive health and sustainability of the natural environment.

Rates and Limits of Element-Cycling

Mineral formation and dissolution are two major processes influencing the overall cycle of elements within the natural environment. Dissolution of minerals involves several key steps which are either rate-limiting or not, depending on whether the dissolution is transport (diffusion)- or surface-controlled (Figure 1). A transport-controlled, rate-determining step involves the movement of a reactant or a weathering product through a diffusion layer on the surface of the mineral that often results in a buildup of concentration at the surface interface greater than concentrations found in the bulk solution at distances from the surface; this is often best described by the parabolic kinetic rate law. Likewise, if surface-controlled processes are in command of the dissolution reactions, the concentrations adjacent to the surface build up to values essentially identical to those in the surrounding bulk solution and may approach steady-state conditions conforming to zero-order kinetics. Typically, surface-controlled dissolution mechanisms dictate the kinetics of mineral dissolution. The possible rate-limiting steps for dissolution are: (1) mass transfer of the reactants in the bulk solution to the surface, (2) adsorption of the reactants, (3) interlattice transfer of reacting species, (4) chemical reactions at the surface, (5) movement of the reaction products away from the surface, and (6) mass transfer of products and excess reactants into the bulk solution. Under normal system conditions, steps 1 and 6 (transport controlled) are not usually rate-limiting, while steps 3, 4, and 5 (surface controlled) are often rate-limiting. Figure 1 shows a comparison of concentration in solution as a function of distance from the surface for transport- and surface-controlled dissolution mechanisms.

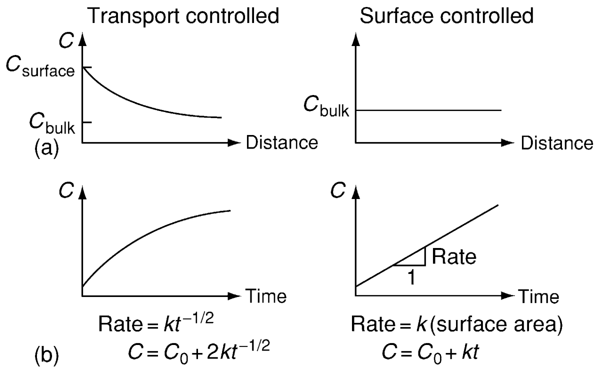


Figure 1 Transport- versus surface-controlled dissolution. Schematic representation of concentration in solution, C, with respect to surface concentration (C_{surface}) and bulk concentration (C_{bulk}), as a function of (a) distance from the surface of the dissolving mineral; and (b) of time, t. Kinetic rate of dissolution is a function of the rate coefficient (k) and time for transport controlled or surface area for surface controlled dissolution. (Reproduced with permission from Stumm W (1992) *Chemistry of the Solid–Water Interface*. New York: John Wiley & Sons, Inc.)

Dissolution is known to occur by interactions of surfaces with ligands, protons, water, and metals. The surface protonation of O and OH lattice sites governs the dissolution of silicates. The dissolution rate increases with decreasing pH. For example, the points of attack of the protons in layered phyllosilicates are the O atoms that interlink the Al-oxide groups with the Si-oxide structures. The protonation slowly releases Al from the phyllosilicate surface followed by the subsequent detachment of $Si(OH)_4$ species. Ligands such as organic acids from biological decomposition and/or root exudates are known to promote the dissolution of phyllosilicates. Dissolution of phyllosilicates and other minerals has been widely studied on a macroscopic scale: kaolinite, muscovite, δ-Al_2O_3, and α-FeOOH as influenced by pH, orthosilicates by divalent metal–oxygen bonds, Fe(III) (hydr)oxides via reductive dissolution, albite promoted by temperature and pressure, and biotite with acids to name a few.

Kinetics of Proton- and Ligand-Promoted Mineral Dissolution

All environmental dissolution processes are time-dependent to varying degrees, thus, in order to comprehend the functioning interactions of solid minerals with respect to their fate with time, an understanding of the kinetics of mineral dissolution is important. Key reasons for examining the kinetics of environmental processes include determination of reaction rates, assessment of time needed to attain equilibrium, and determination of possible reaction mechanisms. A basic approach for describing dissolution kinetics is the development of rate laws based on differential equations that establish the premise that the rate of the reaction is proportional to some power of the concentrations of reactants or intermediate species in the system. For example, the proton-promoted dissolution reaction of gibbsite with a known surface area in an acidic solution can be represented by the following:

$$\gamma\text{-Al(OH)}_3 + 3\mathrm{H}^+ \underset{k'_{-\mathrm{H}}}{\overset{k'_{\mathrm{H}}}{}} \mathrm{Al}^{3+} + 3\mathrm{H_2O} \qquad [1]$$

for which the kinetic rate expression for eqn [1] can be written as:

$$\mathrm{d}[\mathrm{Al}^{3+}]_{\mathrm{aq}}/\mathrm{d}t = -\mathrm{d}[\gamma\text{-Al(OH)}_3]_{\mathrm{ss}}/\mathrm{d}t \qquad [2]$$

meaning that the change in solution concentration of Al^{3+} with time is directly related to the change in concentration of available reactive surface sites (ss; in this case for protons) on the gibbsite surface.

The forward reaction rate law can be written as:

$$\mathrm{d}[\mathrm{Al}^{3+}]_{\mathrm{aq}}/\mathrm{d}t = k_{\mathrm{H}}[\gamma\text{-}\mathrm{Al(OH)}_3]_{\mathrm{ss}}[\mathrm{H}^+]^3 = k_{\mathrm{H}}(\equiv\mathrm{Al(OH)}_2^+)^3 = k_{\mathrm{H}}(C^{\mathrm{s}}_{\mathrm{H}})^3 \quad [3]$$

where k_{H} is the proton-promoted forward rate constant, $[\mathrm{Al}^{3+}]_{\mathrm{aq}}$ is the dissolved Al concentration in solution as a result of gibbsite dissolution by H^+, $[\gamma\text{-}\mathrm{Al(OH)}_3]_{\mathrm{ss}}$ is the amount of gibbsite in the system, often expressed in terms of reactive surface area, and $[\mathrm{H}^+]$ is the proton concentration in solution directly related to pH. This relationship is further defined as $\equiv\mathrm{Al(OH)}_2^+$, which conforms to surface protonation leading to polarization of the lattice sites in the proximity of the surface metal (Al) center and is proportional to the power of the surface protonation rate, in this case $3([\mathrm{H}^+]^3)$, and equates to $C^{\mathrm{s}}_{\mathrm{H}}$, the concentration of the surface proton complex (in moles per square meter).

The reverse reaction rate law for eqn [1] is:

$$\mathrm{d}[\mathrm{Al}^{3+}]_{\mathrm{aq}}/\mathrm{d}t = -k_{-\mathrm{H}}[\mathrm{Al}^{3+}]_{\mathrm{aq}} \quad [4]$$

where $k_{-\mathrm{H}}$ is the proton-promoted reverse rate constant.

Individually, eqns [3] and [4] are only true far from equilibrium where the influences of back reactions are inconsequential; however, if both reactions occur near equilibrium then eqns [3] and [4] must be combined in order to describe the reaction as the difference between the sums of all forward and reverse reaction rates (eqn [5]):

$$\mathrm{d}[\mathrm{Al}^{3+}]_{\mathrm{aq}}/\mathrm{d}t = k_{\mathrm{H}}[\gamma\text{-}\mathrm{Al(OH)}_3]_{\mathrm{ss}}[\mathrm{H}^+]^{3+} + - k_{-\mathrm{H}}[\mathrm{Al}^{3+}]_{\mathrm{aq}} \quad [5]$$

Designing experiments that measure only initial rates of reactions can ensure that back reactions are not significant. By employing an initial rate method for eqn [5], one may plot the concentration of Al^{3+} against time over a short reaction period, during which the forward reaction (described by eqn [3]) prevails and the concentration of gibbsite changes very little, resulting in no alteration of the initial rate.

The proton-promoted dissolution rate, R_{H}, can be expressed as:

$$R_{\mathrm{H}} = k'_{\mathrm{H}}(\equiv\mathrm{MOH}_2^+)^j = k'_{\mathrm{H}}(C^{\mathrm{s}}_{\mathrm{H}})^j \quad [6]$$

where k'_{H} is the rate constant for proton-promoted dissolution, $\equiv\mathrm{MOH}_2^+$ represents the metal–proton complex, $C^{\mathrm{s}}_{\mathrm{H}}$ is the concentration of the surface proton complex, and j is the oxidation state of the metal ion in the surface structure. Drawing upon the previous example of proton-promoted dissolution of gibbsite, the relationship:

$$R_{\mathrm{H}} = \mathrm{d}[\mathrm{Al}^{3+}]_{\mathrm{aq}}/\mathrm{d}t = k_{\mathrm{H}}(\equiv\mathrm{AlOH}_2^+)^3 = k_{\mathrm{H}}(C^{\mathrm{s}}_{\mathrm{H}})^3 \quad [7]$$

can be linearized to a $y = b + mx$ equation where:

$$\log R_{\mathrm{H}} = \log k_{\mathrm{H}} + 3\log(C^{\mathrm{s}}_{\mathrm{H}}) \quad [8]$$

for which a plot of $\log(C^{\mathrm{s}}_{\mathrm{H}})$ versus $\log R_{\mathrm{H}}$ should yield a straight line with a slope of 3 (oxidation state of Al^{3+}) and a y-intercept of k_{H} (as $C^{\mathrm{s}}_{\mathrm{H}}) \rightarrow 1, \log(C^{\mathrm{s}}_{\mathrm{H}} \rightarrow 0)$.

One can express the rate of ligand-promoted dissolution, R_{L}, as:

$$R_{\mathrm{L}} = k'_{\mathrm{L}}(\equiv\mathrm{ML}) = k'_{\mathrm{L}}C^{\mathrm{s}}_{\mathrm{L}} \quad [9]$$

where k'_{L} is the rate constant for ligand-promoted dissolution (per time), $\equiv\mathrm{ML}$ represents the metal–ligand complex, and $C^{\mathrm{s}}_{\mathrm{L}}$ is the surface concentration of the ligand complex (moles per square meter).

The overall rate of dissolution is the sum of ligand-promoted, proton-promoted, deprotonation-promoted, and pH-independent dissolution, which is expressed as:

$$R = R_{\mathrm{L}} + R_{\mathrm{H}} + R_{\mathrm{OH}} + R_{\mathrm{H_2O}} \quad [10]$$

where R_{OH} is the deprotonation-promoted dissolution rate and $R_{\mathrm{H_2O}}$ is defined as the pH-independent dissolution rate.

The overall rate of dissolution can be expressed as:

$$R = k'_{\mathrm{L}}C^{\mathrm{s}}_{\mathrm{L}} + k'_{\mathrm{H}}(C^{\mathrm{s}}_{\mathrm{H}})^j + k'_{\mathrm{OH}}(C^{\mathrm{s}}_{\mathrm{OH}})^i + k'_{\mathrm{H_2O}} \quad [11]$$

where $\mathrm{k}'_{\mathrm{OH}}$ is the rate constant for deprotonation-promoted dissolution, $C^{\mathrm{s}}_{\mathrm{OH}}$ is the concentration of the surface deprotonation sites, and i is the oxidation state of the metal ion in the surface structure. The pH-independent dissolution rate is represented by $k'_{\mathrm{H_2O}}$. It should be noted that temperature plays a crucial role in the kinetic rate of reactions; maintaining a constant temperature during kinetic experiments is vital. Furthermore, the employment of environmentally reasonable temperatures helps to ensure that kinetic rate data are relevant to natural environmental conditions.

The rate of ligand-promoted and proton-promoted dissolution of δ-$\mathrm{Al_2O_3}$ has been investigated. The ligand-promoted dissolution of δ-$\mathrm{Al_2O_3}$ by the aliphatic ligands oxalate, malonate, and succinate (**Figure 2**) follows a linear relationship between the ligand-promoted dissolution rate, R_{L}, and the surface concentration of the ligand complexes, $C^{\mathrm{s}}_{\mathrm{L}}$. While all the ligands examined in **Figure 2** are bidentate (possessing two donor atoms, oxygen, and able to occupy two sites in a coordination sphere) chelating chemicals, the overall size of the molecules plays an important role in the kinetic rate of δ-$\mathrm{Al_2O_3}$ dissolution. As molecular-ring size increases (oxalate < malonate < succinate), the rate of δ-$\mathrm{Al_2O_3}$ dissolution decreases accordingly. Likewise, as the number

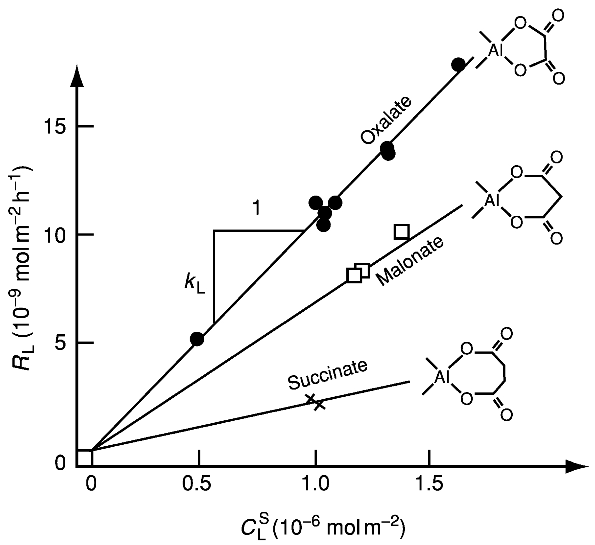


Figure 2 The rate of ligand-catalyzed dissolution of δ-Al_2O_3 by the aliphatic ligands oxalate, malonate, and succinate, R_L (nanomoles per square meter per hour), can be interpreted as a linear dependence on the surface concentration of chelate complexes, C_L^s (moles per square meter). In each case the individual values for C_L^s were determined experimentally. (Reproduced with permission from Furrer G and Stumm W (1986) The coordination chemistry of weathering. I. Dissolution kinetics of δ-Al_2O_3 and BeO. *Geochimica et Cosmochimica Acta* 50: 1847–1860.)

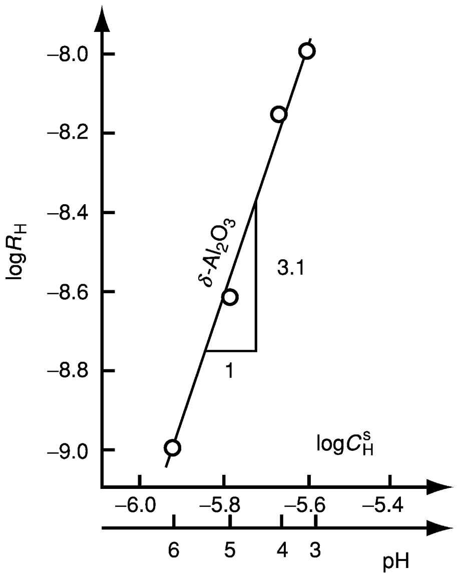


Figure 3 Dependence of the rate of proton-promoted dissolution of δ-Al_2O_3, R_H (nanomoles per square meter per hour), on the surface concentration of the proton complexes, C_H^s (moles per square meter). (Reproduced with permission from Furrer G and Stumm W (1986) The coordination chemistry of weathering. I. Dissolution kinetics of δ-Al_2O_3 and BeO. *Geochimica et Cosmochimica Acta* 50: 1847–1860.)

of donor atoms within a chelating agent increases, the rate of dissolution typically increases (i.e., ethylenediaminetetraacetic acid (EDTA) > oxalate). This phenomenon is known as the 'chelate effect.' Thermochemical studies of complex formation in aqueous solution show that in nearly all cases the chelate effect is due to a more favorable entropy change for complex formation involving multidenate ligands. Proton-promoted dissolution of δ-Al_2O_3 (Figure 3) exhibits a linear dependence with the slope (j) equal to the oxidation state of Al(III) when plotted as logR_H (rate of proton-promoted dissolution) versus logC_H^s, the surface concentration of the proton complexes (eqn [8]).

Figure 4 shows the enhanced and inhibited influence of protons, ligands, cations, and oxyanions on dissolution. Enhancement of dissolution is evident from surface protonation and deprotonation reactions as well as surface complexation with multidentate ligands that are mononuclear in structure. Surface-associated reactions with bi- or multinuclear complexes and some metals result in the inhibition of dissolution due to blockage of surface site reactivity. Hydrophobic compounds can also inhibit dissolution by obstructing surface groups. In studies on the reactivity of Fe(III) (hydr)oxides, dissolution is severely inhibited by phosphate, arsenate, and selenite at near-neutral pH values; however, at pH <5, dissolution is accelerated by the presence of phosphate, arsenate, and selenite.

Metal-Promoted Mineral Dissolution

While some studies have shown that the steric effects of metals can inhibit dissolution, metal-promoted dissolution mechanisms have also been proposed. Sorbing metal cations may induce the dissolution of cations held within the sorbent lattice structure and the released cations then become incorporated in multinuclear surface precipitates with the sorbing metal cations in the bulk solution. Figure 5 shows the kinetics of Ni sorption on pyrophyllite and congruent Si release during Ni sorption as a function of time, as well as Si release from pyrophyllite under identical sorption conditions without Ni present. During Ni sorption in Figure 5, the curve of Ni uptake is mimicked closely by Si release, and that Si release is distinctly different from Si dissolution from pyrophyllite when Ni is not present in solution. It is speculated that the Ni sorption-promoted dissolution of Si and Al from the pyrophyllite structure follows a similar mechanism of proton-promoted dissolution where Ni binding to surface oxide ions causes the lattice bonds to weaken, enhancing the detachment of lattice metal species into solution. Also, X-ray absorption fine-structure (XAFS) spectroscopy has determined

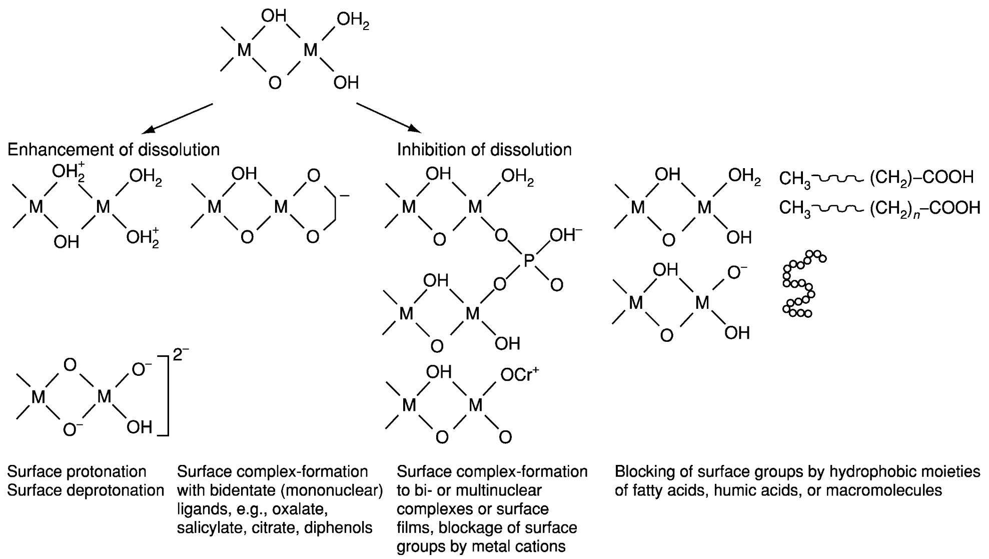

Enhancement of dissolution
Inhibition of dissolution
M
OH
OH_2
O
OH_2^+
O^-
OH^-
P
OCr^+
2−
CH_3-(CH$_2$)-COOH
CH_3-(CH$_2$)$_n$-COOH
Surface protonation
Surface deprotonation
Surface complex-formation with bidentate (mononuclear) ligands, e.g., oxalate, salicylate, citrate, diphenols
Surface complex-formation to bi- or multinuclear complexes or surface films, blockage of surface groups by metal cations
Blocking of surface groups by hydrophobic moieties of fatty acids, humic acids, or macromolecules

Figure 4 The dependence of surface reactivity and kinetic mechanisms on the coordinative environment of the surface groups. (Reproduced with permission from Stumm W and Wollast R (1990) Coordination chemistry of weathering: kinetics of the surface-controlled dissolution of oxide minerals. *Reviews of Geophysics* 28: 53–69.)

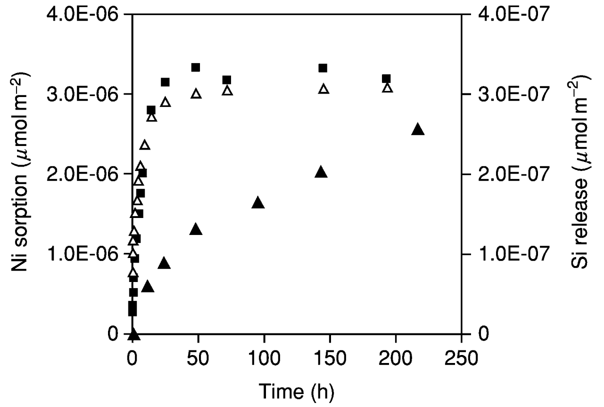


Figure 5 The kinetics of Ni sorption on pyrophyllite from a 3 × 10^{-3} M Ni solution at pH 7.5. Squares, the amount of sorbed Ni (micromoles per square meter), and empty triangles, the amount of simultaneously dissolved Si (micromoles per square meter). The dissolution of untreated pyrophyllite at pH 7.5 is shown for comparison (filled triangles). (Reproduced with permission from Scheidegger AM, Lamble GM, and Sparks DL (1997) Spectroscopic evidence for the formation of mixed-cation hydroxide phases upon metal sorption on clays and aluminum oxides. *Journal of Colloid Interface Science* 186: 118–128.)

that the released Al and Si are incorporated by the newly formed Ni surface precipitates into the octahedral oxide layer and interlayer, respectively, forming a mixed Ni–Al layered double hydroxide (LDH) precipitate.

Reductive Dissolution of Minerals

Changes in the oxidation state of metals have a dramatic impact on the solubility of metal oxide minerals. The oxide minerals of iron and manganese are more soluble and dissolve more quickly when the surface metal centers are reduced by naturally occurring or anthropogenic reductants such as organic chemicals from microorganisms or industrial pollution. The mode of action for the reductive dissolution of metal hydroxide minerals involves a suite of surface reactions that contribute to the overall dissolution rate, including precursor complex formation with the reductant, electron transfer, release of oxidized organic product, and release of the reduced metal ion (Figure 6). This dissolution mechanism assumes that, prior to the detachment step, the reduction of the metal ion and the protonation of the nearest-neighbor oxide or hydroxide must take place.

Dissolution of Metal Surface Precipitates

In recent years, studies have shown that sorption of metals onto natural materials results in the formation of new, mineral-like precipitate phases. Formation of these precipitate phases can reduce metal concentration in soil and sediment solutions. These three-dimensional structures can co-occur with adsorption processes, and may, in some instances, be the product or extension of sorption reactions. However, the stability of the precipitates and the potential long-term release of the metal back into the soil solution have not been extensively examined. While it is evident that surface precipitates often form on mineral surfaces, investigations on the dissolution of surface precipitates are not common. In light of the knowledge that formation of polynuclear metal complexes on

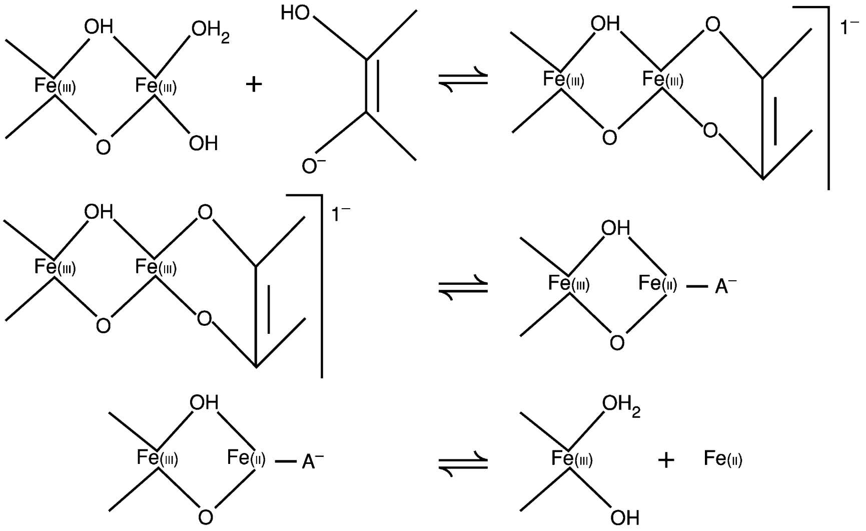


Figure 6 Schematic representation of the dependence of surface reactivity and the kinetic mechanisms on the reductive dissolution surface metal sites.

natural materials is common, there is a need to better understand the degree and mechanism(s) of metal dissolution from surface precipitates. Such information is vital to the better assimilation of the fate of metals in the subsurface environment.

Sorption reactions at the mineral–water interface largely determine the mobility and bioavailability of metals in soils and sediments. Spectroscopic and microscopic studies since the 1990s have consistently shown the importance of metal hydroxide precipitate formation on a variety of clay mineral and oxide surfaces reacted with Ni, Co, and Zn. In spite of the difficulties studying trace amounts of low-crystalline surface precipitates, substantial advances could be achieved in characterizing their chemical composition and structure. Where the sorbent phase releases Al from the lattice structure during reaction with the metal solutions, the precipitates are predominantly Al-containing, layered double hydroxide phases. In the case of Al-free or inert sorbents, however, metal sorption results in α-type metal hydroxide precipitates. Both types of precipitate consist of brucite-like metal hydroxide layers. In contrast to the highly crystalline hydroxide minerals, however, the layers are separated from each other by water and anions, leading to a turbostratic structure, and metal–metal distances within the layer are significantly reduced with respect to the well crystalline metal hydroxides.

The formation of Ni–Al LDH precipitates, as well as the dissolution of these polynuclear Ni(II) surface complexes, from pyrophyllite via proton-promoted dissolution with HNO_3, has been investigated. Nickel detachment from surface complexes is rapid initially at both pH values (with less than 10% of total Ni released) and is attributable to desorption of specifically adsorbed, mononuclear-bound Ni. Dissolution then slows tremendously, primarily owing to the gradual dissolution of the multinuclear surface precipitates. A reference compound, crystalline $Ni(OH)_2$, has also been examined for its dissolution potential. The replenishment method has been employed to simulate steady-state flow and a conventional batch method is also used to compare the influence of reaction products present in solution after dissolution. The replenishment method is more effective in removing Ni from the surface precipitates (approximately 12% at pH 6 and approximately 48% at pH 4) due to removal to reaction products to keep the system from equilibrium. Compared with the dissolution of crystalline $Ni(OH)_2$ (approximately 96% dissolved), Ni release from pyrophyllite is extremely slow.

Studies on the dissolution of Ni–Al LDH surface precipitates on pyrophyllite as a function of residence (aging) time have shown that detachable Ni drastically decreases when the age of the precipitate increases from 1 h to 1 year (Figure 7). By employing high-resolution thermogravimetric analysis, which is sensitive to changes in the interlayer composition of LDH, and by paralleling the results of the surface precipitates with those of reference compounds, a substantial part of the aging effect is shown to be due to replacement of interlayer nitrate by silicate, which transforms the initial Ni–Al LDH into a Ni–Al phyllosilicate precursor (Figure 8). The source of the silicate is the dissolving surface of the pyrophyllite. Studies have also investigated the dissolution of Ni–Al LDH phases on pyrophyllite and gibbsite, and α-$Ni(OH)_2$ precipitates on talc and a mixture of

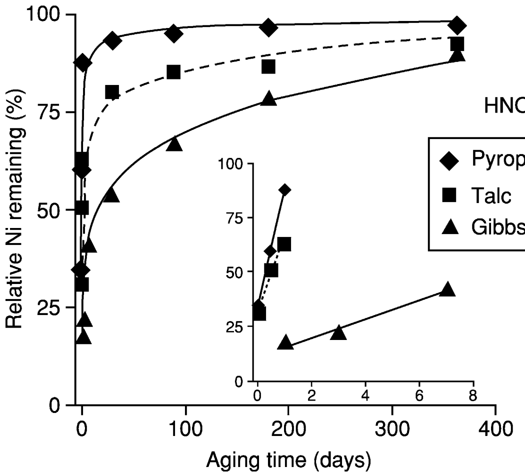


Figure 7 Ni remaining on solids after 10 replenishment steps as a function of aging of the Ni-reacted clay minerals. The inset shows the early aging times. (Reproduced with permission from Scheckel KG, Scheinost AC, Ford RG, and Sparks DL (2000) Stability of Ni hydroxide surface precipitates – a dissolution kinetics study. *Geochimica et Cosmochimica Acta* 64: 2727–2735.)

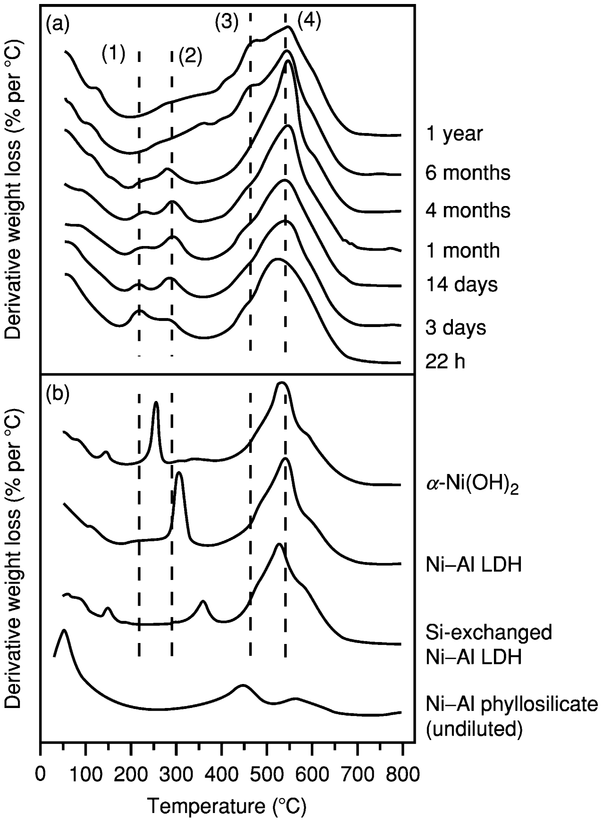


Figure 8 Changes in the thermal stability of the Ni surface precipitate with aging. The derivative of the weight loss curve is shown for (a) aged Ni-pyrophyllite samples and (b) reference precipitates physically diluted with pyrophyllite to match surface loading in sorption samples (2% w/w). Weight-loss events: (1) expulsion of H_2O and nitrate from the layered double hydroxide (LDH) interlayer, (2) dehydroxylation of nitrate-bearing LDH, (3) decomposition of the precursor Ni–Al phyllosilicate, and (4) dehydroxylation of pyrophyllite. (Reproduced with permission from Ford RG, Scheinost AC, Scheckel KG, and Sparks DL (1999) The link between clay mineral weathering and the stabilization of Ni surface precipitates. *Environmental Science and Technology* 33: 3140–3144.)

gibbsite–amorphous silica (gibbsite–silica), employing EDTA (pH 7.5) and nitric acid (pH 4.0) for sorption aging times that range from 1 h to 1 year. In these studies, an array of analytic techniques, including diffuse reflectance spectroscopy (DRS) (Figure 9) and XAFS (Figure 10), have been applied to examine the dissolution of Ni surface precipitates formed on pyrophyllite, talc, gibbsite, and the gibbsite–silica mixture. The differences in stability due to aging of the surface precipitates are speculated to be a combination of Al-for-Ni substitution in the hydroxide layers (for pyrophyllite and gibbsite) and silicate-for-nitrate substitution in the interlayer (for pyrophyllite, talc, and gibbsite–silica mixture).

Macroscopic dissolution studies have demonstrated increased stability in Ni surface precipitates with aging and, to understand this phenomenon better, dissolution of synthetic Ni precipitates of varying composition have been examined (Figure 11). Ni–Al LDH phases with nitrate interlayers are more stable than α-Ni(OH)$_2$ precipitates with nitrate interlayers. Upon changing the interlayer ion from nitrate to silica, there is a dramatic increase in stability for both Ni–Al LDH and α-Ni(OH)$_2$ precipitates, so much so that the Si-containing interlayer α-Ni(OH)$_2$ is more stable than the nitrate interlayer Ni–Al LDH precipitates. Macroscopic and spectroscopic data also show that Al-containing sorbents yield Ni–Al LDH phases and Si-containing sorbents lead to silica-for-nitrate exchanged interlayers as aging time increases. Therefore, the increase in stability with residence time is attributed to three mechanisms: (1) Al-for-Ni substitution in the octahedral sheets of the brucite-like hydroxide layers, (2) Si-for-nitrate exchange in the interlayers of the precipitates, and (3) Ostwald-ripening (size increase) of the precipitate phases. It appears that the second factor, Si-for-nitrate exchange in the interlayers, contributes largely to the increase in stability (Figure 11). The Ni–Al LDH precipitates on pyrophyllite, which possess all three aging factors, result in the most stable complexes. However, Ni–Al LDH on gibbsite, which cannot undergo Si-for-nitrate exchange in the interlayers, are the least stable precipitates, but do increase slightly in stability with aging time due to Ostwald-ripening. The α-Ni(OH)$_2$ precipitates on a gibbsite–silica mixture, amorphous silica, and talc fall between pyrophyllite and gibbsite in regard to stability and are probably a result of the degree of interlayer silication. These results show that increased aging (or residence) times

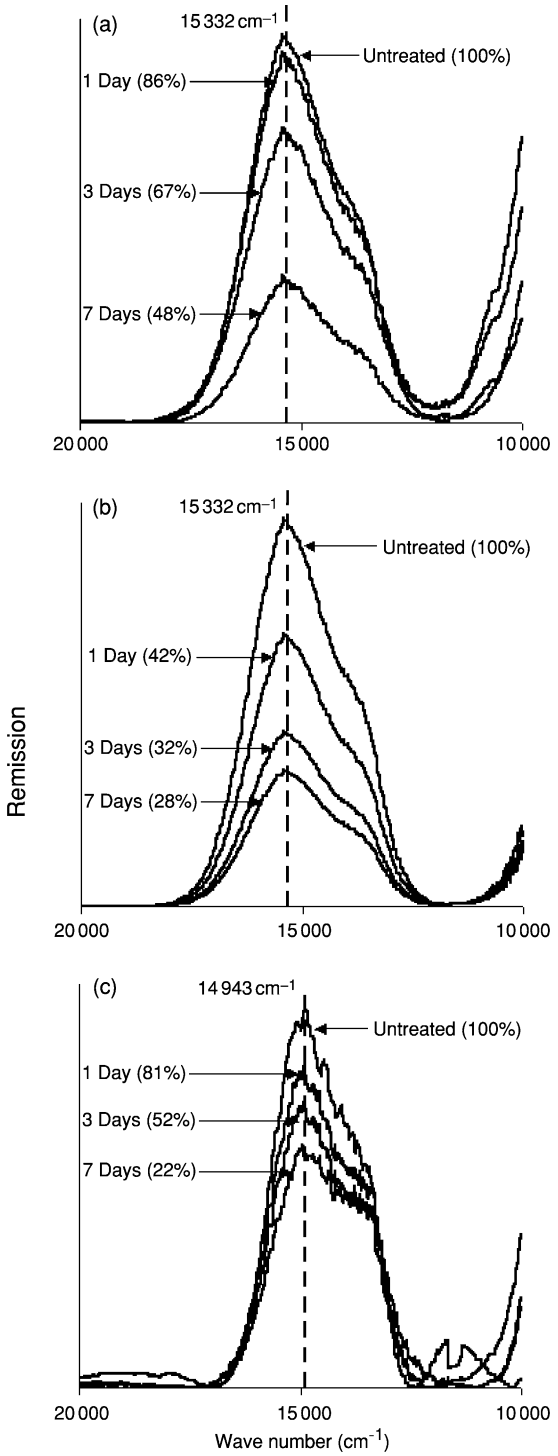


Figure 9 The diffuse reflectance spectroscopy (DRS) $\nu 2$ absorption band for the Ni surface precipitates on (a) pyrophyllite, (b) gibbsite, and (c) talc aged 1 month ('untreated') and subsequently extracted with EDTA (pH 7.5) for 1, 3, and 7 days. The relative amount of Ni remaining on the clay mineral is also given. (Reproduced with permission from Scheckel KG, Scheinost AC, Ford RG, and Sparks DL (2000) Stability of Ni hydroxide surface precipitates – a dissolution kinetics study. *Geochimica et Cosmochimica Acta* 64: 2727–2735.)

under optimal conditions lead to progressively more stable surface precipitates. When environmental conditions change toward those more favorable for dissolution, the rate of dissolution is generally much slower for the precipitates that age for longer periods of time.

The Good and Bad of Dissolution Kinetics

The rate of dissolution of solids in the environment influences a great number of processes. For soil fertility, the rate at which some minerals break down in the presence of plant root exudates, such as oxalate, has a significant impact on plant survival. Some soils are naturally sustainable because of this phenomenon. The mineral components for soil fertility in these soils are slowly extracted over time, requiring little need for anthropogenic inputs of fertilizer; usually nitrogen is the exception. Other soils containing very stable mineral components in climates that are not conducive to dissolution may need regular supplements of required trace elements and nutrients to support crops.

Another issue of recent concern is the dissolution of minerals associated with high amounts of arsenic (As), particularly in parts of Bangladesh and Vietnam. This problem seems to be coupled with the stability of Fe-containing minerals (e.g., pyrite) and Fe oxides. The rate of As recharge into water wells suggests that the dissolution rate is rapid.

Acid mine-drainage directly results from the oxidation of sulfide-containing minerals. While the reaction rate is generally mediated by microorganisms, the protons released in the process can attack other minerals, including nonsulfur-bearing minerals, by proton-promoted dissolution. Thus the production of acid mine-drainage increases pollutants directly by the oxidation of sulfides and indirectly by providing high concentrations of protons that can accelerate the dissolution of minerals.

Recent remediation attempts to immobilize heavy metals in precipitated forms as LDHs or pyromorphite precipitates would not benefit from rapid dissolution kinetics. The objectives of such work are to extensively sequester the metals in solid phases that possess little chance of releasing the metal back into the natural environment where fate, transport, and bioavailability entities may be affected. If the metal can be transformed into a precipitate form with an extremely slow kinetic dissolution rate, then serious issues pertaining to risk assessment and biological availability considerably diminish.

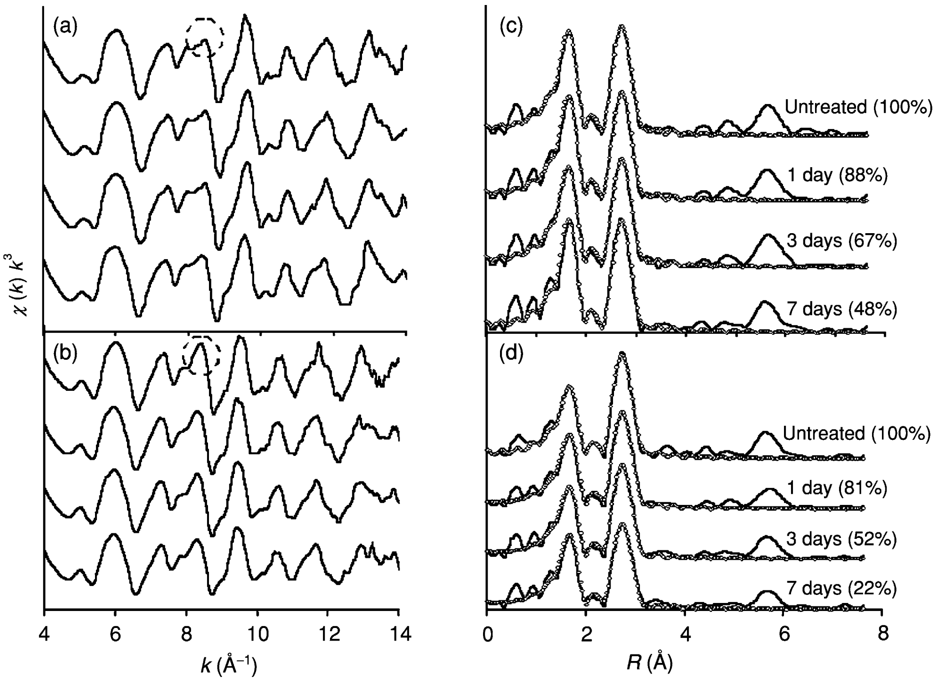


Figure 10 Ni-K_α X-ray absorption fine-structure spectroscopy (XAFS) spectra of pyrophyllite (a and c) and talc (b and d) aged with Ni for 1 month and subsequently treated with EDTA (pH 7.5) for 1, 3, and 7 days. The k^3-weighted χ functions are shown on the left side, and the measured (solid lines) and fitted (dotted lines) radial structure functions are shown on the right side (uncorrected for phase shifts). The circle shows a key identification for Ni–Al LDH versus α-Ni hydroxide. (Reproduced with permission from Scheckel KG, Scheinost AC, Ford RG, and Sparks DL (2000) Stability of Ni hydroxide surface precipitates – a dissolution kinetics study. *Geochimica et Cosmochimica Acta* 64: 2727–2735.)

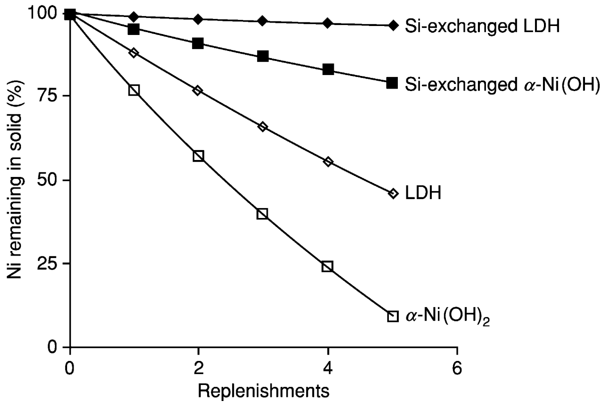


Figure 11 Ni release by HNO_3 at pH 4.0 from homogeneous synthetic Ni-Al layered double hydroxide (LDH) and α-Ni hydroxide, both with either predominantly nitrate or silicate in the interlayer.

List of Technical Nomenclature

$[Al^{3+}]_{aq}$	Dissolved Al concentration in solution
$[\gamma\text{-}Al(OH)_3]$	Amount of gibbsite in the system
$[H^+]$	Proton concentration in solution
=ML	Metal–ligand complex
$=MOH_2^+$	Metal–proton complex
Al	Aluminum
Co	Cobalt
C^S_H	Concentration of the surface proton complex (mol m^{-2})
C^s_L	Concentration of the surface ligand complex (mol m^{-2})
C^s_{OH}	Concentration of the surface deprotonation complex (mol m^{-2})
DRS	Diffuse reflectance spectroscopy
EDTA	Ethylenediaminetetraacetic acid
Fe	Iron
HNO_3	Nitric acid
HRTGA	High-resolution thermogravimetric analysis
i	Oxidation state of the metal ion in the surface structure during deprotonation-promoted dissolution
j	Oxidation state of the metal ion in the surface structure during proton-promoted dissolution
k_H	Proton-promoted forward rate constant
k_{-H}	Proton-promoted reverse rate constant

k_{H_2O}	pH independent-promoted forward rate constant
k_{-H_2O}	pH independent-promoted reverse rate constant
k_L	Ligand-promoted forward rate constant
k_{-L}	Ligand-promoted reverse rate constant
k_{OH}	Deprotonation-promoted forward rate constant
k_{-OH}	Deprotonation-promoted reverse rate constant
LDH	Layered double hydroxide
mol m^{-2}	Moles per meter squared
Ni	Nickel
$Ni(OH)_2$	Nickel hydroxide
pH	Negative log of the proton concentration ($-\log[H^+]$)
R_H	Proton-promoted dissolution rate
R_{H_2O}	pH independent-promoted dissolution rate
R_L	Ligand-promoted dissolution rate
R_{OH}	Deprotonation-promoted dissolution rate
Si	Silicon
ss	Surface sites
XAFS	X-ray absorption fine-structure spectroscopy
Zn	Zinc

See also: **Chemical Equilibria**; **Heavy Metals**; **Kinetic Models**; **Metal Oxides**; **Minerals, Primary**; **Precipitation–Dissolution Processes**; **Redox Reactions, Kinetics**; **Sorption:** Metals; Oxyanions; **Sorption-Desorption, Kinetics**

Further Reading

Ford RG, Scheinost AC, Scheckel KG, and Sparks DL (1999) The link between clay mineral weathering and the stabilization of Ni surface precipitates. *Environmental Science and Technology* 33: 3140–3144.

Furrer G and Stumm W (1986) The coordination chemistry of weathering. I. Dissolution kinetics of δ-Al_2O_3 and BeO. *Geochimica et Cosmochimica Acta* 50: 1847–1860.

Laidler KJ (1965) *Chemical Kinetics*, 2nd edn. New York: McGraw-Hill.

Lasaga AC (1998) *Kinetic Theory in the Earth Sciences*. Princeton, NJ: Princeton University Press.

Scheckel KG, Scheinost AC, Ford RG, and Sparks DL (2000) Stability of Ni hydroxide surface precipitates – a dissolution kinetics study. *Geochimica et Cosmochimica Acta* 64: 2727–2735.

Scheidegger AM, Lamble GM, and Sparks DL (1997) Spectroscopic evidence for the formation of mixed-cation hydroxide phases upon metal sorption on clays and aluminum oxides. *Journal of Colloid Interface Science* 186: 118–128.

Sparks DL (1989) *Kinetics of Soil Chemical Processes*. San Diego, CA: Academic Press.

Sparks DL and Suarez DL (eds) (1991) *Rates of Soil Chemical Processes*. SSSA Special Publication 27. Madison, WI: Soil Science Society of America, Inc.

Sposito G (1994) *Chemical Equilibria and Kinetics in Soils*. New York: John Wiley & Sons, Inc.

Stumm W (ed.) (1990) *Aquatic Chemical Kinetics*. New York: John Wiley & Sons, Inc.

Stumm W (1992) *Chemistry of the Solid-Water Interface*. New York: John Wiley & Sons, Inc.

Stumm W and Morgan JJ (1996) *Aquatic Chemistry*. New York: John Wiley & Sons, Inc.

Stumm W and Wollast R (1990) Coordination chemistry of weathering: kinetics of the surface-controlled dissolution of oxide minerals. *Reviews of Geophysics* 28: 53–69.

DRAINAGE, SURFACE AND SUBSURFACE

N R Fausey, USDA Agricultural Research Service, Columbus, OH, USA

Introduction

Soil drainage is a natural process by which water moves across, through, and out of the soil as a result of the force of gravity. Drainage is a component of the global hydrologic cycle; and streams and rivers are the naturally developed drainage conduits through which some of the water arrives at the land surface as precipitation is transported across the landscape and eventually to the oceans. This natural process also provides the water that supports seeps, springs, stream baseflow, and aquifer recharge. As water leaves the soil, air moves into the space previously occupied by the water; this process is called aeration.

Adequate soil aeration is vital for maintaining healthy plant roots and the many beneficial organisms that live in the soil and require oxygen for respiration. As the proportion of water and air in the soil changes as a result of drainage, the ability of the soil to provide support and traction for animals and vehicles (trafficability) is altered as the strength of the soil changes with water content.

The natural drainage of the soil may limit human use of the resource. Poor drainage has social and economic impacts. The drainage of the soil can be accelerated by the use of surface and subsurface drainage practices. Surface drainage diverts excess water from the soil surface directly to streams, thereby reducing the amount of water that will move into and possibly through the soil. Subsurface drainage, provided by ditches and drainpipes, collects and diverts water from within the soil directly to streams.

Hydrology and Drainage

Precipitation, as well as irrigation, delivers water to the land surface. The type, rate, and amount of precipitation and the properties of the soil determine the relative amounts of the water that will infiltrate, runoff, or remain on the surface temporarily as surface storage. The water that runs off does so by the natural process of surface drainage, sometimes called runoff or overland flow. The water that remains temporarily stored on the surface will eventually either evaporate or infiltrate. Water that infiltrates into the soil increases the soil water content and may ultimately be evaporated back to the atmosphere directly, be taken up by plants, or move deeper into the soil.

The forces of adhesion, cohesion, and gravity control redistribution of the water that infiltrates into the soil. Water infiltrates into the pores between soil particles and adheres to these particles. Additional soil water accumulates as water coheres to water already in contact with the soil particles. Within small pores, enough water can be held against the force of gravity to fill the pores completely. Pores that hold water against the force of gravity are called capillary-size pores. In larger pores, all of the water that enters cannot be held against the force of gravity and some of the water moves downward through these pores. This process is called percolation; percolation is a natural drainage process. Water that percolates deeper into the soil may: enter capillary-size pores at deeper depths and increase the soil-water content; encounter restrictive features that block its descent and cause accumulation of water in the form of a perched water table; reach and add to an existing water table; and reach and replenish an aquifer.

Accelerated Drainage

The history of human adoption of accelerated drainage is not well documented. However, there is evidence from archeological investigations that indicates early adoption and use of manmade drainage facilities. Some examples are raised bed and furrow systems used by the Mayan culture and hillside terraces used by ancient Asian cultures. Although the latter example causes decelerated drainage, it is clearly an effort to alter the natural drainage process for the benefit of humans. There is written record from the Roman era indicating early use of subsurface drainage practices, including instructions for drain installation for agricultural areas.

Surface Drainage Principles and Practices

Surface drainage systems typically consist of an outlet channel (existing natural stream or constructed channel emptying to a natural stream), lateral ditches, and field ditches. Such systems are used primarily in flat areas having poor natural drainage to remove water that collects on the land surface when the rainfall rate exceeds the infiltration capacity. Additional improvement for surface drainage may include land smoothing or land grading to fill in shallow depressions and to assure a continuous slope in the field toward the field ditches. A primary goal in the design and construction of surface drainage systems is to remove the water from the surface as quickly as possible while avoiding soil erosion that can occur when the water moves too rapidly. To avoid soil erosion during surface runoff (drainage), designers look for ways to 'walk the water' off the surface.

Because steep-sided ditches impede the movement of the modern, large machines used in farming, many of these ditches have been replaced with grassed waterways to convey the concentrated runoff flows. This type of channel can be easily crossed by machines and can follow the natural contour of the land surface, thereby minimizing the amount of land taken out of production and the expense involved in construction of the surface runoff collectors. Drop structures are then often needed to dissipate the energy in the runoff water as it drops from the shallow grassed waterways into deeper, open-ditch field drains.

In flat, poorly drained landscapes, surface drains are typically spaced approximately 100 m apart, as shown schematically in Figure 1. The field drains are perpendicular to the prevailing land slope and empty into lateral drains that carry the runoff to larger collectors and main ditches or outlet ditches. Smoothing or grading is used between the field drains to minimize surface storage and ponding. Typical

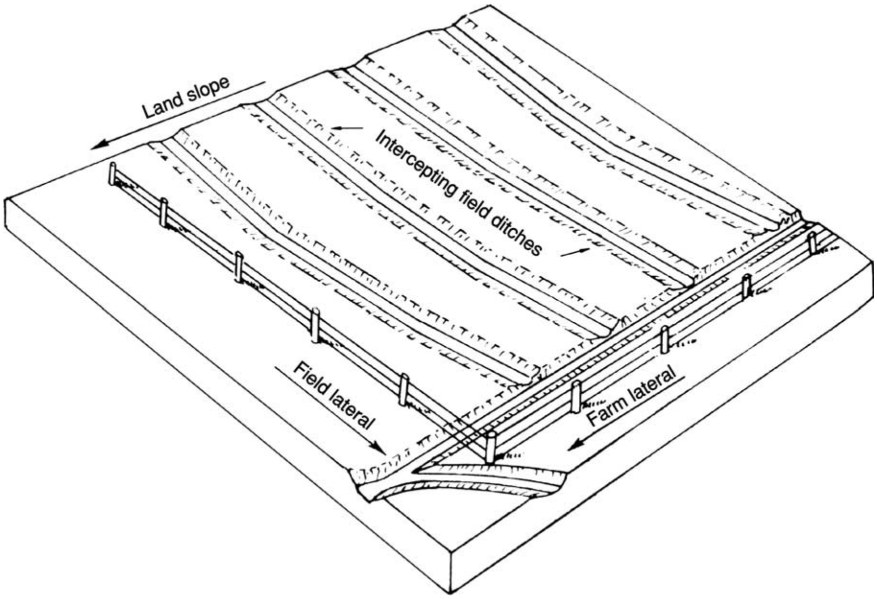


Figure 1 Schematic view of a field surface drainage system with field drains and lateral drains. Reprinted from Pavelis GA (ed.) (1987) *Farm Drainage in the United States: History, Status, and Prospects*. Miscellaneous Publication No. 1455. Economic Research Service. US Department of Agriculture. Washington, DC: US Government Printing Office.

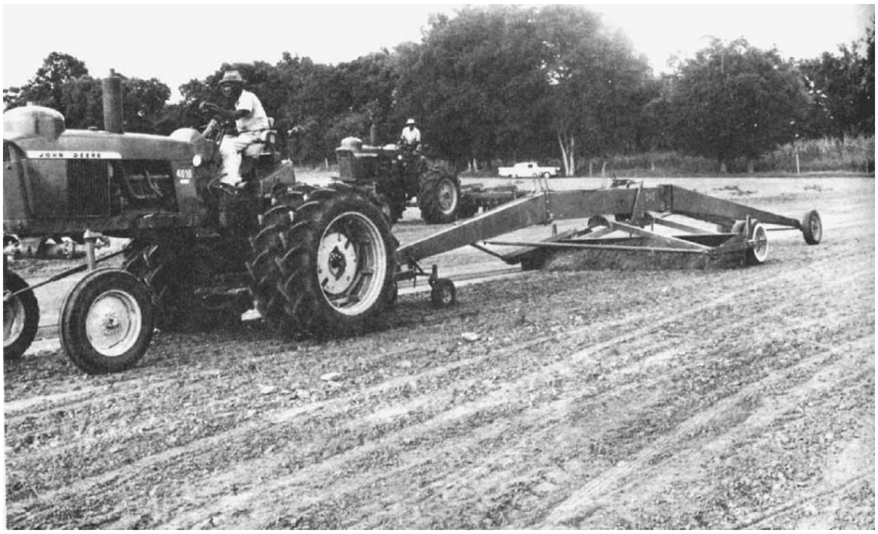


Figure 2 Landplane used for smoothing the soil between surface drains. Reprinted from Pavelis GA (ed.) (1987) *Farm Drainage in the United States: History, Status, and Prospects*. Miscellaneous Publication No. 1455. Economic Research Service. US Department of Agriculture. Washington, DC: US Government Printing Office.

equipment used for this type of work is shown in Figure 2.

Minimizing surface storage is an important aspect of surface drainage. Filling and/or providing outlets for depressions located above and between the grassed waterways and lateral ditches can accomplish this. Land smoothing is typically accomplished by using a land plane to drag small amounts of soil from around the edges of depressions into the depression to fill it in. Land grading involves greater amounts of soil disturbance to construct connections between deeper depressions and to fill up some of the depression storage. This is done with earthmoving equipment.

Subsurface Drainage Principles and Practices

Subsurface drainage describes the process of removal of that water which has infiltrated into the soil in excess of the amount that can be held by capillary forces against the force of gravity. Soils that require accelerated subsurface drainage typically have some impermeable or slowly permeable feature below the surface that prevents water that has entered the soil from moving deeper into the soil and underlying materials at a rate that allows agricultural production to be economically viable. Other criteria may involve the stability of roads and building sites. The obstruction to rapid percolation of water through the soil may be shallow bedrock, highly dense glacial till, depositional clay layers, and other similar causes. In other words, there is no natural outlet for the water, and the soil becomes saturated by the accumulated infiltration of water.

A primary goal in the design and construction of subsurface drainage systems is to remove noncapillary water from the upper layers of the soil profile as quickly as possible to ensure an adequately aerated root zone and trafficability for critical field operations such as planting and harvesting. An illustration of how subsurface drains lower the water table in the soil is given in Figure 3. The depth and spacing of subsurface drains are dependent upon many factors, including especially the availability of an outlet, the soil texture, and the crops to be grown.

The basic equation describing movement of water in the soil was derived by Darcy in 1856. His equation describing the flow rate of water through a soil column is:

$$Q = KA(H_1 - H_2)/L$$

where Q is the flow rate, A is the cross-sectional area of the column, L is the column length in the direction of flow, H_1 and H_2 are the hydraulic heads at the ends of the column, and K is the hydraulic conductivity of the soil. Following this approach, others contributed to the development of an equation to calculate the design spacing between subsurface drains. Subsurface drain spacing varies by soil type and ranges from approximately 10 m for heavy clay soils to as much as 50 m for highly permeable soils.

The path of water flow through the soil is tortuous, around and between soil particles and aggregates, especially in the horizontal direction, where there are no residual channels from roots or insects or animals. Thus, although at the bank of a stream water may drain out of the soil and into the stream due to the force of gravity, this will not continue for very much distance horizontally from the streambank. The stream only provides an outlet for water from a narrow strip of soil near the streambank. There is no outlet for the water in the soil further from the stream. A ditch dug away from the stream into the soil can provide an outlet for the water in a narrow strip of soil along the ditch. The ditch lies at an elevation below the surface of the soil and provides an outlet that allows drainage of the soil above the bottom of the ditch and a short distance away from the side of the ditch. Thus, the ditch is also a subsurface drain. Of course a conduit that admits water through its walls can be placed into the ditch, the ditch can be filled in above the conduit, and the conduit will continue to function as a subsurface drain. And so, by connecting ditches and buried conduits, subsurface drainage systems can be created to remove noncapillary water from broad areas of the soil.

Early subsurface drains were probably always open ditches, but eventually various types of conduits were placed in the bottoms of the ditches and they were covered over. Tree branches tied in bundles, stones, fired clay products in various shapes, extruded clay and concrete pipes, and eventually smooth- and corrugated-wall plastic pipes were used to create subsurface drainage conduits. Drainage ditches were at first dug by hand, but this process has been mechanized greatly over the years. Large machines, like those shown in Figure 4, with automated control of the depth and slope of the ditch bottom or conduit, are now in routine use around the world for installing subsurface drainage systems.

Recent Innovations in Drainage Practice

Traditionally, drainage systems have been installed and operated with open unmanaged outlets that allow free and unrestricted discharge of any water that reaches the drainage conduit or channel. Recent concerns about the delivery of nonpoint source pollutants to streams have encouraged the practice of closing drainage outlets during times of the year when trafficability

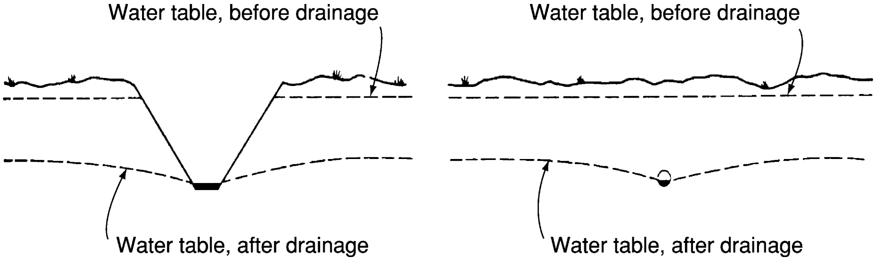

Water table, before drainage
Water table, before drainage
Water table, after drainage
Water table, after drainage

Figure 3 Water table drawdown by ditch and pipe subsurface drains. Reprinted from Pavelis GA (ed.) (1987) *Farm Drainage in the United States: History, Status, and Prospects*. Miscellaneous Publication No. 1455. Economic Research Service. US Department of Agriculture. Washington, DC: US Government Printing Office.

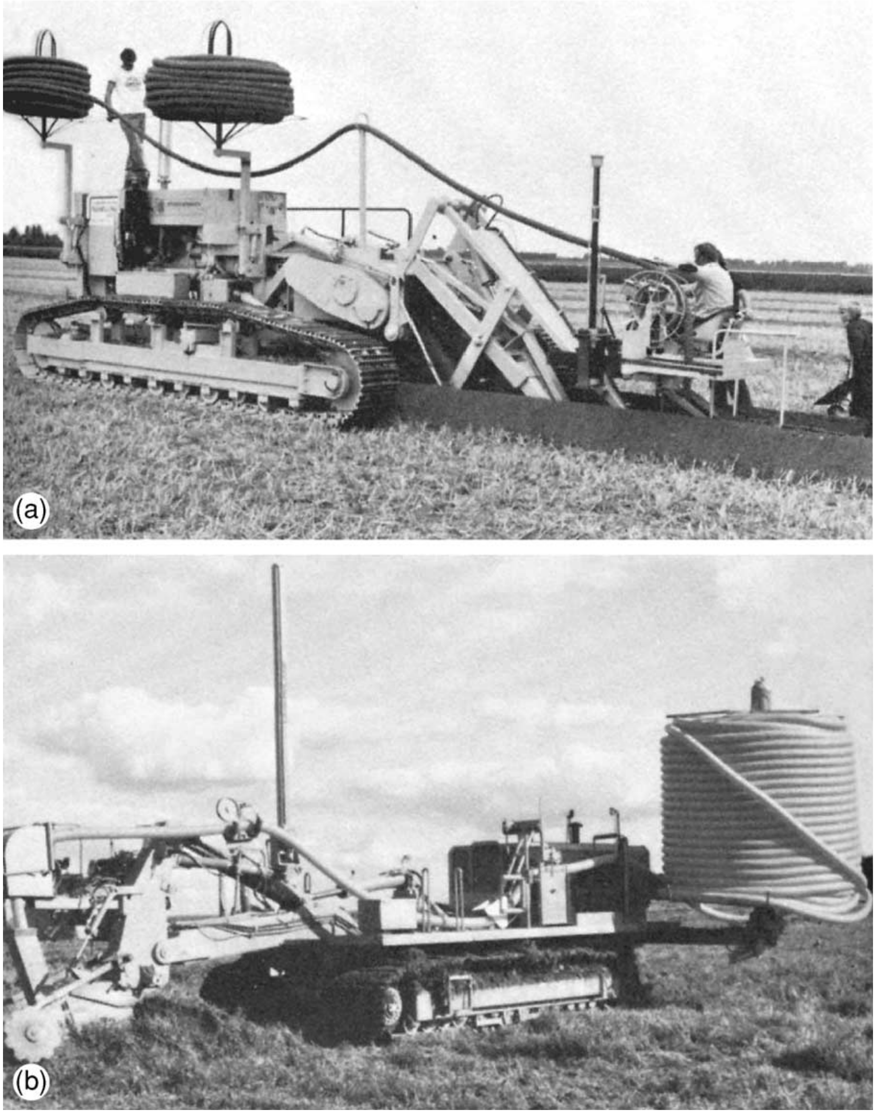


Figure 4 (a) Trench- and (b) plow-type machines for installing subsurface drainage. Reprinted from Pavelis GA (ed.) (1987) *Farm Drainage in the United States: History, Status, and Prospects*. Miscellaneous Publication No. 1455. Economic Research Service. US Department of Agriculture. Washington, DC: US Government Printing Office.

and aeration of crop roots are not required. The practice is known as controlled drainage and involves using a control mechanism at the outlet of surface and subsurface drainage systems to close and open the drains as needed. In some cases when control structures are in place, subirrigation water can be added back into the soil during the growing season through the same conduits that are used to drain the soil. Such systems have been shown to reduce the delivery of pollutants to streams and to increase crop yields. However, a much greater level of management input is required to operate controlled drainage systems.

See also: **Childs, Ernest Carr**; **Hooghoudt, Symen Barend**; **Irrigation:** Environmental Effects; Methods

Further Reading

Pavelis GA (ed.) (1987) *Farm Drainage in the United States: History, Status, and Prospects*. Miscellaneous Publication No. 1455. Economic Research Service. US Department of Agriculture. Washington, DC: US Government Printing Office.

Skaggs RW and van Schilfgaarde J (eds) (1999) *Agricultural Drainage*. Monograph No. 38. Madison, WI: American Society of Agronomy.

van Schilfgaarde J (1974) Non-steady flow to drains. In: *Drainage for Agriculture*, pp. 245–270. Monograph No. 17. Madison, WI: American Society of Agronomy.

Weaver MM (1964) *History of Tile Drainage (In America Prior to 1900)*. Waterloo, NY: M.M. Weaver.

DRYLAND FARMING

G A Peterson, Colorado State University, Fort Collins, CO, USA

Introduction

Dryland farming can be described as crop production in semiarid regions of the world with no irrigation. Historically these regions have provided much of the world's grain supply. The world population today still depends heavily on dryland regions for much of their wheat, barley, millet, sorghum, and pulse supply. Judicious management is critical to the sustainability of dryland farming.

Definition

Dryland farming is frequently defined as crop production in areas with less than 500 mm of annual precipitation, but this definition omits a critical component of the equation, evaporation potential. Operatively, dryland farming is practiced where annual potential water evaporation exceeds annual precipitation. The example for the Central Great Plains of the USA in Figure 1 illustrates this. Note that the deficit between precipitation and potential evaporation is large and is at its peak in the middle of the summer crop growing season. As the water deficit increases (i.e., the difference between annual precipitation and potential evaporation becomes more negative), the difficulty of producing crops increases proportionally. Worldwide dryland farming areas are characterized by deficits between precipitation and potential evaporation, but differ in the size of the deficit and the time of the year it occurs. For example, in Morocco in northwestern Africa the deficits are so large in the summer that no dryland crop production can occur. Overall dryland farming productivity is inversely related to the size of the deficit between annual precipitation and annual potential evaporation. A large deficit indicates more plant stress and lower yields.

World Scope of Dryland Farming

Dryland farming is an ancient practice with a fascinating historical record. Today, one can find dryland farming practices that are basically unchanged since ancient times. Practices used since 500 BC stand in sharp contrast to dryland farming practices involving the latest discoveries of chemistry and plant genetics. In many instances, the productivity of the soils under the older management techniques has decreased, especially in situations where population pressures on food supply have multiplied.

All continents, with the exception of Antarctica, have substantial amounts of land where dryland farming is practiced. The climates and soils where dryland farming is practiced vary widely from continent to continent and even within continents. A primary factor for dryland farmers is when they receive their annual precipitation and how that fits with the growing season of the particular crops they produce. There are many combinations of precipitation pattern and crop production. Figure 2 illustrates two very different precipitation environments, both of which are used for dryland production. The US Central Great Plains has a summer rainfall pattern, while

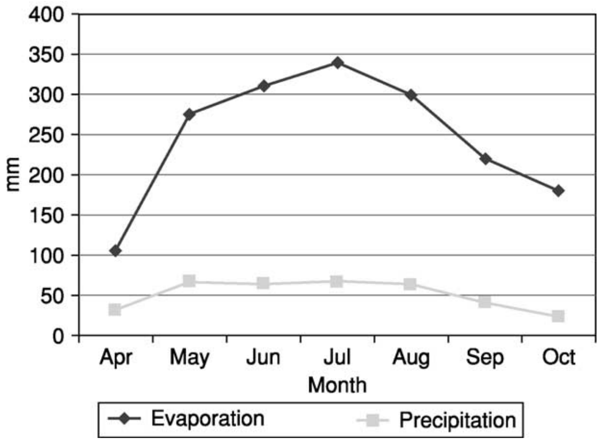


Figure 1 Evaporation and evaporation potential in the US Central Great Plains.

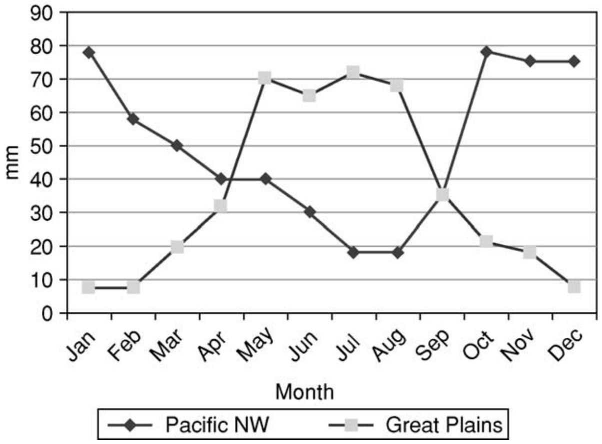


Figure 2 Rainfall patterns in the US Pacific Northwest and the Great Plains.

the US Pacific Northwest has a winter precipitation pattern. Winter precipitation patterns are found in both cold and warm climates, greatly altering management. For example, in the Pacific Northwest the winter precipitation is primarily snow, whereas the winter precipitation pattern in Northwest Africa or Southern Spain is all rainfall with no snow.

In areas with a winter precipitation pattern, all precipitation is received in the winter, and the water is conserved in the soil for the crop that follows. In summer precipitation pattern areas, most of the precipitation is received during the crop growing season, but is accompanied by high temperatures that encourage evaporation. These examples illustrate the complexity of the management issues where dryland farming is practiced.

A wide diversity in soils exists in dryland farming areas but, in general, soils are neutral to basic in pH, low in organic matter, and low in both nitrogen and phosphorus supply capacity. They are relatively unweathered because of the climatic conditions under which they have been formed, and thus have adequate amounts of basic cations such as calcium, magnesium, and potassium. Physically, these soils have weak structure, which is easily destroyed by tillage or natural events such as raindrop impact. The low organic matter contents contribute to the weak structure.

Major Dryland Farming Issues

Despite the wide climate and soil variations where dryland farming is practiced, there are similarities that create identifiable management issues. The two major commonalities that exist for dryland farming areas are: (1) annual crop production is always limited by water supply; and (2) soil erosion hazards are high, both from water and wind.

Lack of water results in relatively limited yields and general lack of vegetative cover, which in turn creates soil conditions that are highly vulnerable to erosion. Thus, dryland farming has often been associated with ‘dust bowls’ and ecosystem degradation. Susceptibility to erosion by water also remains high because many dryland farming areas receive short bursts of high-intensity rainfall, which can cause immediate soil erosion.

The primary means of improving the stability of dryland farming is to maintain cover on the soil surface for as much of the year as possible. Maintaining cover on the land, either as vegetative canopy or crop residue from previous production, improves water conservation both between crop seasons and during the crop growing season. Improved water conservation in turn improves plant productivity and increases the opportunity to keep the soil from eroding, which is a positive feedback to the agroecosystem.

Dryland Farming Management Techniques

Fallowing

Fallowing has been the primary yield stabilizing tool used by farmers in the majority of the dryland areas worldwide. Simply stated, this means leaving the land idle as frequently as every other year to store water and to allow nitrate-nitrogen to accumulate in the soil. The stored water plus the nitrate-nitrogen accumulation increase the probability of producing an adequate crop in the year after fallow. Fallows are generally managed as weed-free as possible if water conservation is the primary objective, but in some areas of the world weeds are allowed to grow during the fallow year and are grazed by livestock. In the latter situation, little soil water is stored because the weeds transpire the water. However, some organic residues are returned to the soil, and the farmer has a source of forage for animals. The practice of ‘weedy fallow’ is most commonly found in subsistence farming areas.

Unfortunately, fallowing in any form is highly inefficient in terms of water storage. Depending on atmospheric evaporative demand and cover management, the proportion of precipitation stored in the soil during fallow ranges from near 0 to 50%. Most often, however, less than 20% of the precipitation is stored in the soil even if weeds are controlled during the fallow period. Despite this inefficient precipitation management, fallowing has remained popular because it ensures some measure of crop yield, which can be critical to farmer survival. Introduction of no-till management of fallow has improved water storage potential, but in most cases precipitation storage is still less than 50% of the precipitation, meaning that more than 50% of the precipitation is lost to evaporation, weed growth, or water runoff. No-till is a farming system that maximizes water conservation and minimizes soil erosion. Herbicidal weed control is used, no tillage events of any kind are allowed, and thus residues from the previous crop remain as protection on the soil surface.

Sustainability of dryland farming in the long term requires improved precipitation-use efficiency. Present research efforts focus on three aspects of precipitation management: (1) precipitation capture in the soil, whether from rainfall or from snow melt; (2) water retention in the soil after capture, which involves weed control and minimization of evaporation; and (3) efficient water use by the crop plants, which involves agronomic practices associated with soil fertility, variety choice, etc.

Water Capture Management

Capturing the incident precipitation, whether it is rainfall or snow melt, is the first water conservation step in dryland farming. It is critical to system sustainability that capture be maximized within the economic constraints of the particular situation. Water capture is governed by soil properties like soil texture, aggregation, pore size, and pore-size distribution. Large amounts of surface soil macroporosity are needed to maximize water infiltration. This soil characteristic is highly governed by soil texture; fine-textured soils have less macropore space than coarse-textured soils, and subsequently lower infiltration rates than coarse-textured soils. Degree of aggregation within a given soil texture also governs water capture. Well-aggregated soils have larger amounts of macropore space, and thus improved water capture relative to poorly aggregated soils of the same texture. Degree of soil aggregation is possible to manage, while differences due to texture are not feasibly altered.

Structural stability of the aggregates also affects water capture. Soils with weak soil structure quickly lose their ability to absorb water as the surface aggregates disintegrate upon wetting and the surface pore spaces become smaller (e.g., soil crusting). Soils in dryland regions tend to have low organic matter content, which contributes to weak aggregation.

Soil water content at the beginning of a precipitation event also modifies the effects of other properties. As the water content of a soil increases, the water infiltration rate decreases because the surface soil pores are full of water and there is little space for water entry. The maximum water-intake rate is obtained at the beginning of a rainfall event and then decreases rapidly as water fills the surface pore space. If the soil is dry, there is a large storage capacity, relative to the same soil in a moist condition; hence water will flow rapidly into the soil storage reservoir, thus accounting for a higher infiltration rate in dry soil.

Maintaining cover on the soil surface is a key element for maximization of water capture in dryland systems. Cover, whether plant canopy during the growing season or crop residue after harvest, protects the soil aggregates from raindrop forces and thus soil pores remain open for water capture. Cover also slows the running water after a storm and increases opportunity time for water absorption by the soil. Historically, dryland farmers could not maintain cover on the soil between cropping seasons because tillage for either weed control or seedbed preparation incorporated the remaining crop residues into the soil. Cultivation for weed control during the cropping season further depleted any residue cover. New scenarios for cover management have evolved with the advent of no-till management. Herbicidal weed control permits residue retention between seasons and during the growing season, and modern no-till planting equipment incorporates only small amounts of residue (<5%), leaving most of the residue on the soil surface. Thus the soil has maximum protection year-round.

Tillage has a secondary negative effect on water capture because it physically grinds aggregates and reduces their overall size distribution, which slows water absorption. It also increases soil organic matter oxidation rates because of aggregate destruction and subsequent increased exposure of organic compounds to soil organisms, which further weakens aggregates. Aggregate size distributions shift, such that microporosity increases at the expense of macroporosity, which results in decreased capability to absorb water. The degree to which tillage affects water capture is governed by complex interactions of tillage type and time, with an array of soil characteristics such as texture, original structure, and organic matter content. Overall, long-term tillage of any soil decreases aggregate resistance to physical disruptions such as raindrop impact and tillage operations of all sorts.

Water Retention Management

Weed control during time periods between crop growing seasons and weed control within the growing season are critical to retaining water for the cropping system. Tillage and herbicidal weed control are two extremes of a spectrum. Both can be effective in controlling weeds and preventing water use by weeds. They differ greatly, however, in terms of their effect on water evaporation from the soil surface. Herbicidal weed control minimizes water evaporation, while tillage maximizes evaporative losses: the reason being that all tillage events expose moist soil to the atmosphere, while herbicides kill weeds with no soil disturbance. Thus, water retention is maximized if the soil is not stirred in most environments.

Residue cover also reduces water evaporation rates because it reflects and absorbs heat energy, preventing it from warming the soil beneath. Note that the residue cover slows evaporation rate, but does not stop evaporation. Eventually the water in the surface soil beneath a residue cover will dry out if no rainfall occurs, but at a much slower rate than if no cover is present. Thus in climates where rain events occur frequently, the residue is more beneficial for water retention than in geographic regions where weeks and months pass with no rainfall. For the climates illustrated in Figure 2, the summer precipitation regions benefit from residue cover in terms of evaporation control, whereas the winter precipitation areas with dry, hot, summers benefit less from residue cover. Residue cover is still valuable in winter precipitation areas for other reasons such as improved water capture and erosion control.

Efficient Water Use by Crop Plants

Once the water is captured and retained in the soil, the farmer needs to be sure that the best agronomic practices are used so that healthy plants are available to use the water. These practices include crop variety choice, proper planting dates, and adequate fertilization. Weak, stressed plants cannot develop root systems that will maximize crop yield response to the available soil water and precipitation that is received during the growing season.

Conclusion

Managing cover, the primary dryland-farming management technique, affects water capture, water retention, and even efficiency of water use. Cover absorbs raindrop impact energy, catches snow in winter seasons, slows runoff water, decreases water evaporation, and improves nutrient management for the plants. Furthermore, it provides protection against erosion by wind and water. None of these roles is independent of the others. Dryland farming will always play a key role in feeding the world's population, and as that population grows, the need for improved management techniques will also increase.

Further Reading

Brengle KG (1982) *Principles and Practices of Dryland Farming*. Boulder, CO: Colorado Associated University Press.

Dregne HE and Willis WO (1983) *Dryland Agriculture*. Agronomy No. 23. Madison, WI: American Society of Agronomy.

Farahani HJ, Peterson GA, and Westfall DG (1998) Dryland cropping intensification: a fundamental solution to efficient use of precipitation. *Advances in Agronomy* 64: 197–223.

Skujiņš J (1991) *Semiarid Lands and Deserts: Soil Resource and Reclamation*. New York, NY: Marcel Dekker.

E

Earthworms *See* **Fauna**

EDAPHOLOGY

A L Ulery, New Mexico State University, Las Cruces, NM, USA

Introduction

Edaphology is the science or study of soil, especially with respect to plant growth. The root of the word is 'édaphos,' Greek for 'foundation,' 'soil,' 'ground,' or 'land.' 'Soil science' is the term more commonly used today and includes the study of soil as a natural body in the landscape (pedology), as well as a medium to be managed for optimum crop and rangeland productivity, environmental waste disposal, and construction. Pedology focuses on the description and classification of soil and how it is formed. While both pedology and edaphology refer to the study of soil, the latter deals specifically with soil as a medium for plant growth. (*See* **Pedology:** Basic Principles.)

Edaphologists study the soil by examining a variety of factors or soil properties that govern plant (flora) and animal (fauna) life in the soil. Edaphic factors have been described as those soil conditions that affect the organisms living in a particular area. These factors are distinguished from climatic or physiographic factors that can also affect plant growth largely through the influence of water and temperature by concentrating on the root system and soil. The study of edaphic factors is an important part of ecology.

Ecosystems are thrown out of balance when major disturbances occur such as the overgrazing of rangeland or the conversion of native forests or grassland to agriculture. Such ecosystems are no longer in equilibrium and cannot be adequately described by traditional theories of plant succession or climax models. An edaphic climax is a localized vegetative community that differs from the surrounding vegetation because of different soil types and properties, including water-holding capacity, drainage class, soil depth, and soil fertility. New state and transition models for the study of nonequilibrium ecology and vegetation dynamics recognize that the interactions between the soil resource and the associated vegetative community determine the functional status of ecologic processes. This theory also states that the condition of the soil is directly connected to aboveground vegetation.

The organisms that live in a soil influence soil formation and are in turn affected by the edaphic properties that govern water, air, and nutrient supply. Thus, a continual feedback relationship is established between the soil flora and fauna, and the physical and chemical characteristics of the soil. Edaphology is the study of that relationship and how it affects plant growth. One approach to edaphology is agronomic and involves cultivating the soil and adding water and nutrients for maximum plant yield. Irrigation and soil-fertility management are examples of this approach. Ecological studies that correlate plants to specific soil types or properties in addition to landscape position and climate information are another example of edaphology. Evaluating or predicting the effect of management on soil–plant interactions and edaphic factors is often the objective of studies on ecosystem health, soil quality, and land degradation.

Edaphic Properties

Edaphic factors include those physical, chemical, and biological properties of a soil that influence plant growth. The various properties that are important for plant growth are interrelated but can be distilled to a few simple requirements. Essentially, what plants need from the soil are water, air, nutrients, and physical support. Sunlight is required for photosynthesis but is provided aboveground and may be affected by aspect, latitude, and other physiographic and climatic factors. Some of the soil or edaphic properties that affect plant requirements are listed in Tables 1–3.

Edaphic factors often resemble soil-quality indicators. 'Soil quality' is a somewhat contentious term in the soil science community, but it refers to some very useful concepts concerning the ability of a soil to sustain biological production and diversity, regulate and partition water, filter and buffer contaminants, store and cycle nutrients, and provide structural support. Soils of high quality are fertile and productive, resist erosion, and help maintain good air and water quality. They are also resilient to natural and anthropogenic stresses and are therefore resistant to degradation. Soil-quality indicators are those physical, chemical, and biological properties that measure a soil attribute or function and are sensitive to changes in environment and management (Tables 1–3).

Table 1 Some physical edaphic properties and soil-quality indicators that affect plant growth by influencing water and air movement, or rooting depth in soil

Soil property	*Water and air movement*	*Root support or impedance*
Texture	X	X
Structure	X	X
Soil or solum depth	X	X
Depth to impermeable layer	X	X
Infiltration	X	
Hydraulic conductivity	X	
Porosity and macropores	X	X
Soil-water potential	X	
Water-holding capacity	X	
Drainage class	X	X
Bulk density	X	X

Table 2 Some chemical edaphic properties and soil-quality indicators that affect plant growth by influencing water and air movement, nutrients, or rooting depth in soil

Chemical property	*Water and air movement*	*Nutrient availability*	*Root support or impedance*
pH		X	
Cation exchange capacity		X	
Organic matter content	X	X	X
Nitrate nitrogen		X	
Organic nitrogen		X	
Total or soluble phosphorus		X	
Total or soluble potassium		X	
Micronutrients		X	
Clay mineralogy	X	X	X
Salinity or electrical conductivity	X	X	
Toxic ions or heavy metals		X	

Table 3 Some biological edaphic properties or soil-quality indicators that affect plant growth by influencing water and air movement, nutrients, or rooting depth in soil

Biological property	*Water and air movement*	*Nutrient availability*	*Root support or impedance*
Earthworm number	X	X	X
Respiration rate (measure of microbial population and activity)		X	
Microbial diversity		X	
Composition of organic matter	X	X	X
Plant root depth	X	X	X

Edaphic Factors Affecting Water and Air

In general, soil physical properties have the greatest effect on water and air movement into and through the soil. The role of water is so important in plant growth that some definitions of edaphology refer to it as the science of physics applied to soil and water. Water and air are considered together because they use the same pathways to enter and move through the soil. The water-holding capacity of a soil refers to the amount of water contained in a soil and is determined largely by texture, structure, porosity, soil depth, and organic matter content.

'Texture' refers to the particle-size distribution of the soil material. Particles range from sand size (2 mm) to clay size, which is less than 2 μm (1000 times smaller). Finer-textured or clayey soils hold more total water than sandy soils, but all of that water is not available to the plants, because it is held so tightly to the clay particles. Sandy soils have large, interconnected pores that fill and drain easily. While the texture refers specifically to the distribution of particles less than 2 mm in diameter, another valuable piece of information is the percentage of gravel or amount of particles greater than 2 mm. This factor affects soil packing, porosity, bulk density, nutrient status, and root growth.

Porosity results from the arrangement of soil particles into aggregates with inter- and intraaggregate pores, as well as from root growth and faunal activity. The pores within and around the aggregates conduct water and air through the soil. Macropores can form after roots decay or as a result of burrowing soil fauna. One of the most dominant groups of soil animals are earthworms, which may number from approximately 50 to 300 worms m^{-2} in agricultural soils and more than 500 worms m^{-2} in forest and grassland soils. Earthworms shred and consume organic matter and excrete wastes as casts, which help create stable soil aggregates and enhance soil porosity. Their burrows are persistent over time and provide a major conduit for drainage and preferential water movement through the soil, but also minimize surface water erosion by increasing infiltration. The presence and abundance of earthworms often indicate healthy, organic-rich soils.

Bulk density is the dry soil weight per volume of soil and is a good measure of porosity and compaction. Pore space contributes to the volume of a soil and thus, as porosity increases, the dry weight decreases, resulting in a lower bulk density. Compacted soils have higher bulk densities and tend to restrict water and air flow as well as root growth. In addition, texture and organic matter both affect bulk density. Sandy soils have a higher bulk density than clayey soils because of lower porosity. Organic matter weighs very little, takes up a large volume, and so creates lower bulk density. Some minerals have a high particle density and can contribute to a higher soil bulk density if they are dominant in the soil. Soils with a lower bulk density contain more water than soils with a high bulk density, but they tend to drain too quickly and become droughty.

Organic matter contributes to water-holding capacity in several ways. It aggregates the soil and produces a crumb-like structure at the surface that facilitates infiltration, percolation, storage, and diffusion of water and air. Sticky substances such as waxes and tannins left behind when soil organisms decompose organic residues help bind soil particles together into porous clusters that hold and release water, nutrients, and humus. Soil organic matter or humus holds as many as 20 times its weight in water and has many beneficial physical, chemical, and biological properties. The O and A horizons in soil profiles have the highest amounts of organic matter and are often measured in field studies to evaluate their influence on nutrient supply or relationship to vegetation. Organic carbon or total organic matter is also commonly measured in edaphic studies.

The depth of the soil and the water-holding capacity also directly determine the amount of water in a soil. Many trees and desert shrubs have root systems that exploit soil water and nutrients to depths below 2 m, but most agronomic plant roots are concentrated in the upper meter of the soil. The bottom of a soil profile may be an impermeable layer or hardpan, bedrock, or a shallow groundwater table. A change in texture – for example, a sandy layer under a clay loam, or vice versa – can restrict water flow through the soil. Shallow, eroded soils hold less water than deep, uniformly textured soils. Gleying and mottles are evidence of standing water and poor aeration or anaerobic conditions.

Mineralogy, specifically the kind and amount of certain clay minerals, may also affect water and air movement through the soil. Some soil minerals such as montmorillonite, a member of the smectite family, absorb water and swell to fill up cracks and macropores, resulting in restricted hydraulic conductivity. This is particularly problematic in sodic soils or alkali-affected areas. Sodium is a cation with low charge density that tends to neutralize and disperse negatively charged clay and the aggregates formed from clay and organic matter. When aggregates disperse they form an impermeable seal at the soil surface and along cracks that limits the movement of water and air into the soil. The sodium adsorption ratio or exchangeable sodium percentage are two measures of the sodium content affecting plant growth, especially in arid regions where the rainfall is insufficient to leach salts out of the soil profile. Sodic soils are also a problem in closed basins with shallow groundwater that is often saline and detrimental to plant growth.

Edaphic Factors Affecting Plant Nutrients

For healthy growth, plants require many inorganic nutrients in varying amounts and forms. Macronutrients have concentrations of at least 500 mg kg^{-1} in plants, while micronutrients are required in lower amounts, usually less than 100 mg kg^{-1}. Nitrogen (N), phosphorus (P), and potassium (K) are the most commonly measured and applied macronutrients by farmers and gardeners. Inorganic fertilizers are labeled with their N–P–K amounts and should be applied to the soil only when soil test results indicate that the nutrients are below optimum levels. Specific edaphic factors that may be measured to monitor soil fertility include nitrate-N, available P, and soluble K (Table 2). Other commonly measured nutrients include calcium (Ca), magnesium (Mg), sulfate (SO_4^{2-}), and iron (Fe). Sometimes it is not just the total amount of each nutrient or compound that affects plant growth and distribution on a landscape, but the relative proportions of nutrients as well. For example, the calcium-to-magnesium ratio may be used to explain the occurrence of some endemic plants growing on serpentine soils in California, USA.

The soil reaction or pH is related to nutrient availability and optimum plant growth. This is one of the most common soil chemical properties measured. Some plants prefer acidic soils, while others are best suited for basic or alkaline soils. There are few plants that will survive in extremely acid or alkaline soils, however, as most flora and fauna prefer conditions between pH 5.5 and 8.5.

Soil pH also affects the microbial population of the soil, which is necessary to convert soil organic matter to inorganic nutrient forms. Most microbes prefer warm, moist, near-neutral conditions to mineralize organic matter and wastes. Soil microbes are responsible for decomposing plant residues and wastes and converting them to their final inorganic products of carbon dioxide, water, and nutrients. Where inorganic fertilizers are unavailable or too expensive, or when slow release of nutrients is desired, organic materials such as manure, compost, and plant residues are the best nutrient source for plants. In addition to carbon, these materials contain nitrogen, phosphorus, sulfur, and other nutrients. Soils rich in organic matter are darker and tend to warm up faster in the spring. The combined physical, chemical, and biological benefits of organic matter or humus make it one of the most important edaphic factors affecting plant growth.

Soil organic matter and clay minerals are usually negatively charged because of their composition. This diffuse negative charge allows them to attract and retain positively charged cations from the soil solution. The ability of a soil to adsorb or hold cations is called the cation exchange capacity (CEC). This is an important feature of soil materials because so many plant nutrients are cations (e.g., K^+, Ca^{2+}, Mg^{2+}, ammonium (NH_4^+) and many micronutrients). Cations are held loosely and temporarily by negatively charged exchange sites in and around soil particles so that they are not leached away with every rainfall or irrigation event. However, their attraction is not permanent and the cations are easily available to plants and microbes as they are needed. Soils that have abundant humus and smectite or vermiculite clays are higher in CEC than sandy soils low in organic matter. Organic matter is also valuable in complexing or chelating some metals to make them more soluble and plant-available. In addition, soil organic matter has been shown to negate the toxicity of aluminum (Al) in some soils.

In addition to aluminum, several other elements are toxic to plants and may require characterization to determine whether they are affecting plant growth and distribution on the landscape (Table 4). Nickel (Ni), manganese (Mn), lead (Pb), copper (Cu), chromium (Cr), and cobalt (Co) are among the metals that have been analyzed and correlated with vegetation growth around mines or smelters. Some elements are plant micronutrients at low concentrations, but become toxic at high concentrations in the soil or solution. Examples of these elements are boron (B), copper, and zinc (Zn).

Soil mineralogy is also an important edaphic factor, influencing the availability of some important nutrients such as phosphate and sulfate. Allophane or clay materials derived from volcanic glass can fix or permanently adsorb phosphate and sulfate anions. Iron and aluminum oxides are prevalent in leached, acidic soils of humid zones and can also combine with phosphate to make insoluble phosphate minerals that are unavailable to plants. In arid regions, where calcareous soils are common, the calcium combines with and binds the phosphate. Phosphorus is generally most plant-available at near-neutral pH and in soils not dominated by allophane, iron or aluminum oxides, or calcium carbonate. Available or extractable phosphorus is measured to distinguish it from organic or mineral forms that are not accessible to plants.

Potassium and ammonium also become fixed or unavailable in soils high in vermiculite. The size of these ions and their charge density allow them to become permanently sandwiched in the interlayer region of high-charge vermiculites that collapse around them in response to opposite electrical charges. Thus, soluble potassium and/or clay mineralogy is measured to determine the amount of potassium available for plant uptake.

Table 4 Toxic elements that may affect plant growth and soil quality

Trace element	*Maximum concentration allowed in irrigation water*[a] *(mg $^{-1}$)*	*Essential to plants at low concentrations?*[b]
Aluminum (Al)	5.0	Possibly
Arsenic (As)	0.10	Possibly
Beryllium (Be)	0.10	No
Bismuth (Bi)	NA	No
Boron (B)	NA	Yes
Cadmium (Cd)	0.01	No
Chromium (Cr)	0.10	Possibly
Cobalt (Co)	0.05	Possibly
Copper (Cu)	0.20	Yes
Fluorine (F)	1.0	Possibly
Iron (Fe)	5.0	Yes
Lead (Pb)	5.0	Possibly
Lithium (Li)	2.5	Possibly
Manganese (Mn)	0.20	Yes
Molybdenum (Mo)	0.01	Yes
Nickel (Ni)	0.20	Yes
Selenium (Se)	0.02	Possibly
Silver (Ag)	NA	No
Tin (Sn)	NA	No
Vanadium (V)	0.10	Possibly
Zinc (Zn)	2.0	Yes

NA, no maximum concentration has been designated.

[a] Criteria established by the Food and Agricultural Organization (FAO) of the United Nations.

[b] Some elements are required by plants at low concentrations but are toxic at higher levels (therefore 'Yes'), and some elements are still under investigation or only partially accepted as being essential, thus the term 'Possibly.'

Salinity or electrical conductivity (EC) is often measured in arid regions or areas under irrigation. EC increases as the salt content increases. Salinity can be a problem in closed basins with shallow water, tables. Salts can affect plant growth in two ways, either directly as a toxic specific ion or indirectly by lowering the osmotic potential of the soil water, which hinders the plants' ability to uptake water. Increased salts lower the osmotic potential, which in turn lowers the total soil-water potential, making plants work harder (expend more energy) to extract water from the soil. Specific ion toxicity can be a problem when boron, chloride (Cl^-), selenium (Se), or sodium (Na) concentrations are excessive. Toxicities of certain ions are plant-dependent; however, high sodium concentrations may also disperse the soil particles and create physical impediments to water and air flow. Inorganic fertilizers are also salts, designed to dissolve easily in water. Excess fertilizer may contribute to soil EC and saline stress.

Edaphic Factors Affecting Root Growth

Many of the factors already discussed also affect plant root growth and development. Compaction and high bulk density limit the root system if the roots cannot force their way through the soil. However, good soil–root contact is important for nutrient and water uptake, so moderate compaction may actually enhance plant growth. Repeated cultivation by heavy machinery can cause a plow pan or hard layer to develop that can impede root growth and water movement. This layer is usually found at the bottom of the Ap horizon, and the thickness of the Ap horizon is thus a useful edaphic factor to measure (Figure 1).

Other horizons of interest in a soil profile include the B horizon, a zone of accumulation of clay (Bt), organic matter (Bh), calcium carbonate (Bk), iron oxides (Bs), or salts (Bz) (Figure 1). These horizons may affect rooting depth and plant growth either physically by impeding root growth and water flow or chemically by fixing nutrients or supplying toxic concentrations of salts and other compounds. Bt horizons can also enhance the water-holding capacity of a soil. The solum depth (the combined thickness of the A and B horizons) is sometimes analogous to soil depth, when the C horizon is considered to be unweathered or little affected by pedogenic processes (Figure 2).

Shallow soils often result from erosion, which is a major problem in many areas of the world. Good plant cover is one of the best ways to minimize soil loss, whether from wind or water erosion, and thus is an important component in soil conservation and management. Physiographic factors such as slope percentage, aspect, and location on the landscape are often recorded along with edaphic factors to evaluate plant–soil–landscape relations. The depth of the soil profile and the thicknesses of various horizons are pedological as well as edaphic factors; this is an example of where the distinction between edaphology and pedology becomes blurred.

Temperature is an important edaphic factor, influencing plant growth and root development. Soil temperature is affected by several physical factors, including porosity, water content, soil color, organic matter content, bulk density, and soil depth. It is also a function of landscape placement and physiographic features such as aspect, elevation, and latitude. The temperature may be measured directly in the soil, or

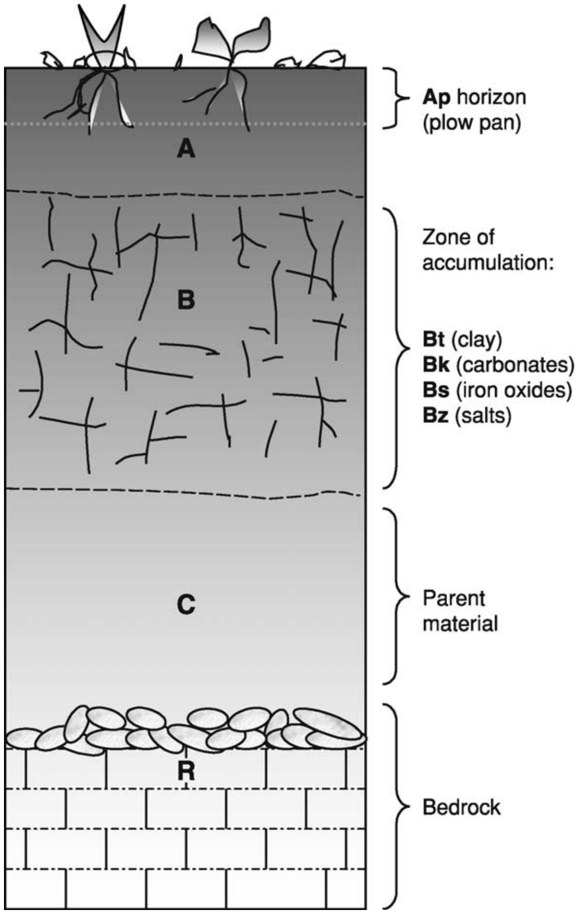


Figure 1 A soil profile, a vertical section of the soil through all its horizons (layers), extending into the C horizon (parent material). Every soil is individual and has its own unique characteristics and properties. Some of the master horizons and just a few of the many possible subhorizons that may form in a soil as it develops from the parent material in response to climate, vegetation, biota, topography, and time are shown. These include the Ap and various B horizons that may inhibit root growth and water movement through the profile if they are dense or cemented. Alternatively, the accumulation of clays and organic matter in the B horizon can enhance water-holding capacity and nutrient availability, and thus plant growth.

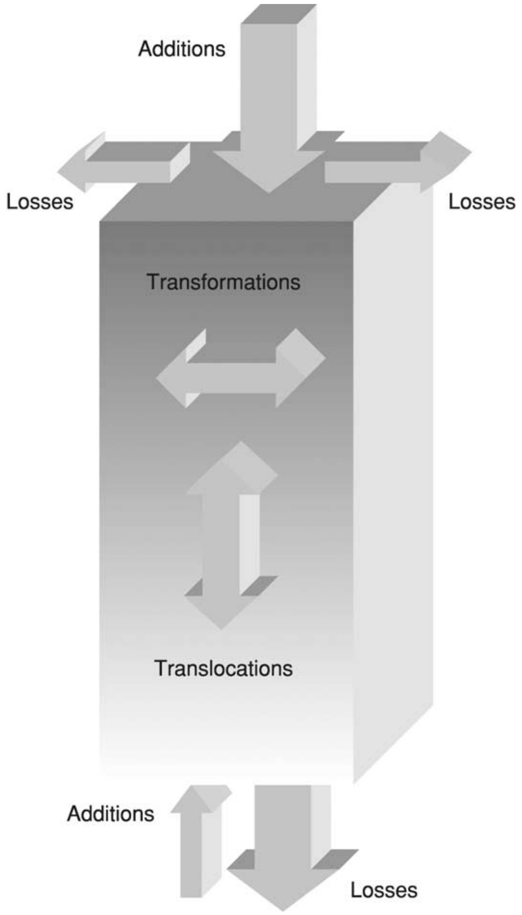


Figure 2 Pedogenic processes include the addition and removal of matter into and out of the profile as well as the chemical alteration (transformation) and movement of matter within the profile (translocation).

indirectly as a function of climatic factors such as air temperature and solar radiation, as well as the edaphic features mentioned above.

Research

In addition to the use of the word 'edaphic' instead of 'soil' to describe any condition or property pertaining to the soil, there are specific studies conducted by ecologists, plant scientists, agronomists, and soil scientists (or edaphologists) that are designed to measure, characterize, monitor, and assess the health of the soil as it serves as a medium for plant growth. The literature on edaphology research is largely divided into ecologic and agronomic studies. In the science of ecology, the spatial variation or distribution of vegetation as a function of soil type or edaphic property is important to characterize an ecosystem. The information gleaned from these studies provides a baseline of native vegetation in the landscape and helps identify changes due to anthropogenic and climatic effects. The long-term goal of this kind of study is to manage or protect native or desired vegetation by understanding the relationships among plant communities, the underlying soil, disturbance regimes, and physiographic features.

The other area of edaphology research is concerned with sustainable agriculture and long-term cropping practices to maintain soil fertility and productivity. The field of agroecology deals with the management and cultivation of agronomic ecosystems (or agroecosystems). Intensive conventional agriculture can deplete a soil and lower its quality, reducing the soil's capacity to function in an ecosystem or to resist degradation. By monitoring soil-quality indicators and various edaphic factors, scientists can assess the sustainability of adopted management practices. The two broad areas of ecological and agricultural edaphology have also been combined to allow comparison of intensively managed agroecosystems with virgin soils and native vegetation.

Soil is dynamic, teeming with organisms, and is an integral, interactive part of the environment. It is the foundation upon which our civilization stands and it must be used wisely and conserved if we are to continue benefiting from its numerous functions, including food and fiber production, environmental protection, water and nutrient storage, and engineering materials.

List of Technical Nomenclature

Agroecology The science of applying ecological concepts and principles to the design and management of sustainable agroecosystems

Ca^{2+} Calcium ion

CEC Cation exchange capacity

Climax The most advanced successional community of plants capable of development under, and in dynamic equilibrium with, the prevailing environment

EC Electrical conductivity, related to the concentration of salts in solution

Ecology The science of the relationship between organisms and their environment

Edaphology The science that deals with the influence of soils on living things; particularly plants, including human use of the land for plant growth

Gleyed/gleying A condition resulting from prolonged soil saturation and reducing conditions, indicated by the presence of bluish or greenish colors from ferrous iron in the soil mass or as mottles

Humus	Organic soil compounds exclusive of undecayed plant and animal tissues, their 'partial decomposition' products, and the soil biomass. The term is often used synonymously with soil organic matter
K^+	Potassium ion
Macronutrient	A plant nutrient found at relatively high concentrations (greater than 500 mg kg^{-1}) in plants; usually refers to nitrogen, phosphorus, and potassium, but may also include calcium, magnesium, and sulfur
Mg^{2+}	Magnesium ion
$mg\,kg^{-1}$	Milligrams per kilogram of dry soil or plant material. Similar to parts per million (ppm)
Micronutrient	A plant nutrient found in relatively small amounts (less than 100 mg kg^{-1}) in plants. These include boron, chloride, copper, iron, manganese, molybdenum, nickel, cobalt, and zinc
mm	Millimeters or 10^{-3} meters. Used to define the length or diameter of soil particles, roots, etc
Mottles	Spots or blotches of different color or shades of color interspersed with the dominant soil color
NH_4^+	Ammonium ion
Sodium adsorption ratio (SAR)	A relation between soluble sodium and soluble divalent cations that can be used to predict the exchangeable sodium fraction of soil equilibrated with a given solution. It is defined as the concentration of sodium divided by the square root of the sum of calcium and magnesium concentrations
Worms m^{-2}	The number of earthworms per square meter of soil

See also: **Pedology:** Basic Principles; **Quality of Soil**

Further Reading

Brady NC and Weil RR (2002) *The Nature and Properties of Soils*, 13th edn. Upper Saddle River, NJ: Prentice-Hall.

Doran JW, Coleman DC, Bezdicek DF, and Stewart BA (eds) (1994) *Defining Soil Quality for a Sustainable Environment*. SSSA Special Publication No. 35. Madison, WI: Soil Science Society of America.

Gershuny G and Smillie J (1999) *The Soul of the Soil: A Guide to Ecological Soil Management*. White River Junction, VT: Chelsea Green.

Gliessman SR (1998) *Agroecology: Ecological Processes in Sustainable Agriculture*. Chelsea, MI: Ann Arbor Press.

Hillel D (1998) *Environmental Soil Physics*. San Diego, CA: Academic Press.

Miller RW and Gardiner DT (2001) *Soils in Our Environment*, 9th edn. Upper Saddle River, NJ: Prentice-Hall.

Plaisance G and Cailleux A (1981) *Dictionary of Soils*. Paris, France: La Maison Raustique Publication (Translated from French and published for the USDA and NSF: New Delhi, India: Amerind Publishing Company, PVT, Ltd).

Porazinska DL and Wall DH (2001) Soil conservation. In: Levin SA (ed.) *Encyclopedia of Biodiversity,* vol. 5, pp. 315–326. San Diego, CA: Academic Press.

Singer MJ and Munns DN (2002) *Soils: An Introduction*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Sparks DL (2003) *Environmental Soil Chemistry*, 2nd edn. San Diego, CA: Academic Press.

Tugel AJ and Lewandowski AM (eds) (1999) *Soil Biology Primer.* Ames, IA: NRCS Soil Quality Institute.

Electron Paramagnetic Resonance *See* Electron-Spin Resonance Spectroscopy

ELECTRON-SPIN RESONANCE SPECTROSCOPY

N Senesi and G S Senesi, Università di Bari, Bari, Italy

Introduction

Electron-spin resonance (ESR) spectroscopy, otherwise known as electron paramagnetic resonance (EPR) spectroscopy, is a nondestructive, noninvasive, highly sensitive and accurate analytical technique that can detect and characterize chemical species possessing unpaired electrons, i.e., paramagnetic. Soil species able to produce ESR spectra include organic free-radical moieties in humic substances (HS) and organic- and mineral-associated paramagnetic metal ions. Further, ESR labeling, trapping, and metal-probing methods are used to study the dynamics of macromolecules in soil solution, HS complexation chemistry, and ion sorption on mineral surfaces.

Basic Principles

The Resonance Phenomenon

The basic physical phenomenon underlying ESR spectroscopy is referred to as the 'Zeeman effect,' which consists of the interaction between the magnetic moment of an unpaired electron and an external magnetic field that produces the splitting of the energy levels of the unpaired electron. If a static magnetic field of strength H is applied in the z-direction, i.e., parallel to the z-axis, the Zeeman interaction leads to an energy, E, for an unpaired electron, given by:

$$E = -\mu_z H = -g\beta H M_z \qquad [1]$$

where μ_z is the electron magnetic moment in the direction of the conventional z-axis; g is a dimensionless constant called the spectroscopic electronic splitting factor or g-value or g-tensor; β is the electron Bohr magneton; and M_z is the component of the electron-spin angular momentum in the direction of the applied magnetic field H (z-axis). The values that M_z may assume are $+1/2$ and $-1/2$ depending on the alignment of the electron spin, S, either parallel (high-energy) or antiparallel (low-energy) to the magnetic field direction. Thus, two energy levels exist with an energy difference that increases linearly with the magnitude of H (Figure 1).

In a sample containing unpaired electrons in the thermodynamic equilibrium in a magnetic field of value H_0, a population difference exists between the two energy levels, with an excess population in the lower level. If an incident electromagnetic radiation is supplied to the sample, e.g., by applying an alternating magnetic field of frequency v_0 perpendicular to the static magnetic field H_0, an absorption of energy, ΔE, occurs provided that v_0 satisfies the following equation:

$$h v_0 = \Delta E = g\beta H_0 \qquad [2]$$

where h is the Planck constant. This is known as the 'resonance condition.' The measurement of this energy absorption, recorded as its first derivative, is the ESR signal (Figure 1), i.e., the basis of ESR spectroscopy.

Spectral Parameters

The position and the shape of the ESR signal depend on the environmental conditions in the vicinity of the electron, which may result in spectral patterns more complicated (Figure 1b) than the single-line spectrum (Figure 1a). The most important effects that influence the position and pattern of the ESR spectrum are the electron Zeeman, nuclear hyperfine, ligand superhyperfine, and nuclear quadrupole interactions. These effects are related to spectral parameters that include the g-factor and nuclear hyperfine and superhyperfine coupling constants.

At a given magnetic field and for a particular microwave frequency, the electron Zeeman effect shifts the resonance position of the free electron from a value of $g = 2.00232$ to a g-value (eqn [2]) that is dependent on the molecular properties of the paramagnetic species. The g-value of organic free radicals is generally close to 2.00 and does not distinguish well between different radicals, whereas for paramagnetic metal ions it is often typical of a particular ion and its valence state. For most paramagnetic ions, the g-value is anisotropic, i.e., dependent on the orientation of the molecule relative to the external magnetic field, and is completely described by the three components g_{xx}, g_{yy}, and g_{zz}, along the three axes x, y, z (x-, y-, z-system) that are generally coincident with molecular symmetry axes. These three components differ from each other ($g_{xx} \neq g_{yy} \neq g_{zz}$) for paramagnetic species that have no principal axis of symmetry, whereas a single isotropic g-factor, g_{iso} or $g_0 = 1/3$ ($g_{xx} = g_{yy} = g_{zz}$) is exhibited by octahedral, tetrahedral, or cubic symmetries. Species with axial symmetry, such as Cu^{2+} and V^{4+}, have one principal or threefold axis of symmetry, conventionally the z-axis, and equivalent x- and y-axes.

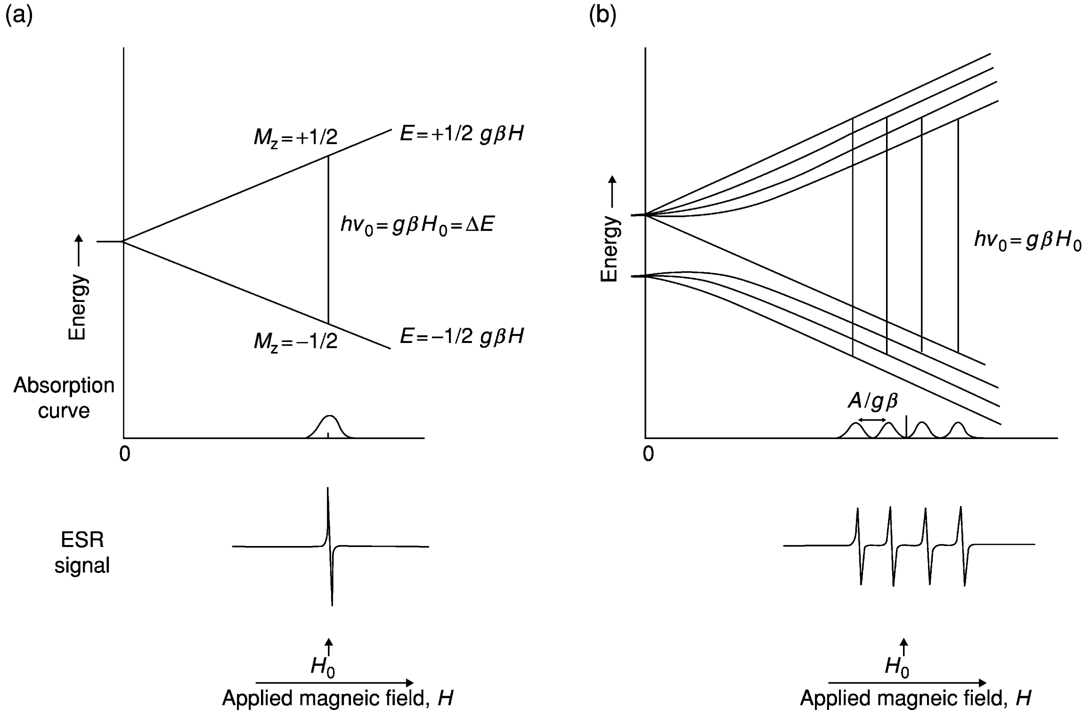


Figure 1 Effect of an applied magnetic field, H, on the energy levels (E) of the two spin states of an electron ($M_z = \pm 1/2$). Electron-spin resonance (ESR) transition(s) at $\nu_0 = g\beta\ H_0/h$, actual absorption curve, and experimental first-derivative ESR spectrum for the cases of no nuclear interaction (a) and interaction with a nucleus having $I = 3/2$ (b), e.g., Cu^{2+}. The nuclear hyperfine splitting is given by $A/g\beta$.

These species feature so-called rigid-limit spectra and exhibit two g-values, usually labeled $g_{||}$ (= g_{zz}, the g-value parallel to the symmetry axis or z-axis), and $g\perp$ (= g_{xx} = g_{yy}, the g-value perpendicular to the z-axis in the x,y-plane). The g-values can provide information on the nature of the paramagnetic species and the symmetry of the metal ion in the sample matrix.

The ‘hyperfine splitting or structure’ arises from the ‘nuclear hyperfine’ interaction of the unpaired electron with its nucleus if it features a nonzero spin ($I \neq 0$), such as Cu (I = 3/2), Mn (I = 5/2), or V (I = 7/2), and is independent of the applied field H_0. Permanent local fields arising from magnetic nuclei split each electron spin level into $2I + 1$ components, which results in a set of $2I + 1$ equally spaced hyperfine-structure lines replacing the single line resonance in the ESR spectrum. For example, four lines appear for Cu (Figure 1b), six for Mn, and eight for V. The hyperfine splitting is generally approximated by $A/g\beta$, where A is the magnitude of the nuclear hyperfine interaction, the so-called hyperfine coupling constant. The parameter A, like g, may also exhibit an orientation-dependence (i.e., anisotropy), thus it may provide useful information on the nature and molecular symmetry of paramagnetic species possessing magnetic nuclei.

The ‘superhyperfine splitting or structure’ arises when an interaction occurs between the unpaired electron and ligand nuclei having nonzero nuclear spin in paramagnetic metal–ligand complexes. The most common nucleus giving rise to superhyperfine structures in ESR spectra is ^{14}N, which has a nuclear spin of $I = 1$. Thus, each hyperfine line is split into three approximately equally spaced components of equal intensity. Very complex spectra are obtained, however, when more than one N-ligand, especially if not equivalent, is involved in metal complexation.

The ‘nuclear quadrupole’ interaction arises from the nuclear quadrupole moment (for example, of the Mn nucleus, $I = 5/2$) and the electric field gradient. The interaction, besides modifying the hyperfine splitting, also ‘mixes’ the hyperfine levels, so that ‘forbidden’ transitions can occur. Generally, quadrupole effects are very small and rarely observed in powder and frozen solution spectra.

Instrumentation and Methodology

The Instrument and the Experiment

The basic components of a typical ESR spectrometer operating in the X-band (microwave) frequency region of the electromagnetic spectrum are (Figure 2): (1) an electromagnet supplying a static, or direct current (DC), continuous magnetic field that can be varied in strength, usually in the range 0–1 tesla ($1\,\mathrm{T} = 10^4$ gauss); (2) the source of microwaves, usually a vacuum tube oscillator named 'klystron' that provides a monochromatic coherent source of electromagnetic radiation with a frequency typically around 9.5 GHz that is held constant to within 1 part in 10^6; (3) a resonant or 'microwave' cavity, which is a hollow-conducting box generally cuboid where the sample is placed and irradiated; (4) a waveguide that couples the klystron to the resonant cavity that is locked electronically to the resonance frequency; (5) a circulator incorporated between the klystron and the cavity, which ensures optimum transfer of the microwave power from the klystron to the cavity and from the cavity to the detector; (6) a small iris diaphragm that interfaces the cavity to the waveguide and whose effective diameter can be varied until all the incident microwaves are absorbed by the cavity so that it is said to be 'critically coupled'; (7) a magnetic field modulator that increases the intensity of the reflected microwave at low frequency (usually 100 kHz) by superimposing upon the external magnetic field H_0 a second magnetic field of a few tens of millitesla; and (8) a phase-sensitive detector locked to the magnetic field modulation frequency.

In the practical ESR experiment, the microwave frequency is usually fixed and the external applied magnetic field is continuously swept until the resonance condition (eqn [2]) is met. Most ESR spectrometers allow for repetitive sweeping of the magnetic field through the resonance and employ a computer to average transients, thus improving sensitivity. The output of the detector–amplifier system, that is, the ESR spectrum, is displayed as the first derivative of the absorption peak and recorded on a chart recorder as a function of the applied external magnetic field, H (Figure 1). Occasionally, the second derivative is plotted to improve the resolution of closely lying lines.

Sample Preparation

The ESR technique can analyze samples in any form, that is, liquid, solid powder, amorphous polymer, frozen glass, or single crystal. Generally, solid powders are packed into an ESR tube with an inside diameter of a few millimeters. Aqueous solutions or suspensions are placed in capillary tubes or specially designed flat cells to reduce dielectric absorption of microwave radiation by water molecules. Tubes or cells are made of highly pure quartz with no defects, which gives no ESR signal and has a very small dielectric loss. Clay films can also be used mounted on flat quartz 'tissue cells.'

Removal of paramagnetic O_2 molecules that can broaden ESR spectra by dipole–dipole interaction is sometimes required before ESR analysis. If the concentration of paramagnetic species in a sample is too large, the sample must be magnetically diluted in a diamagnetic matrix to minimize spin–spin coupling that may broaden the spectrum.

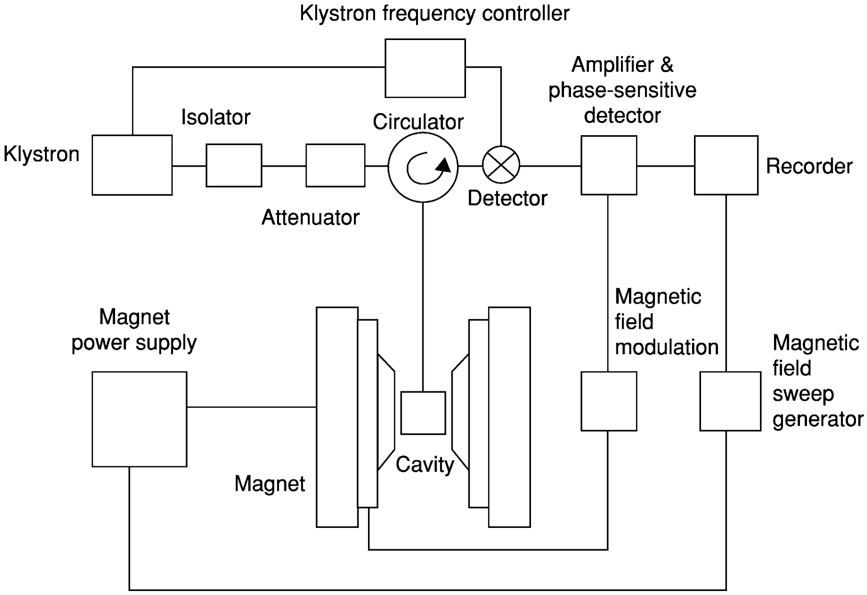

Klystron frequency controller
Amplifier & phase-sensitive detector
Isolator
Circulator
Klystron
Recorder
Detector
Attenuator
Magnetic field modulation
Magnet power supply
Magnetic field sweep generator
Cavity
Magnet

Figure 2 Block diagram of a typical X-band ESR spectrometer.

Sensitivity and Resolution

The sensitivity of the technique depends primarily upon the number of spins present in the sample. Typically, a modern X-band spectrometer can detect 10^{15}–10^{16} spins, which implies a sensitivity of 10^{-7}–10^{-8} moles. The sample must be correctly placed in the cavity that generally has a detection region about 2 cm high, with the greatest sensitivity at its center. Sensitivity is generally improved by setting the modulation amplitude at larger values that, however, should not exceed a fraction of the peak-to-peak resonance linewidth, otherwise distortion of the spectral line shape can occur. According to the Curie law, maximum sensitivity for paramagnetic species is attained at the smallest possible sample temperature. Although organic free-radical species can generally be measured easily at room temperature (RT), for most paramagnetic metal ions cooling at liquid N_2 (bp 77 K) or liquid He (bp 4.2 K), a higher temperature is often required. The extremely large dielectric loss of water at microwave frequencies prevents working at RT with aqueous systems and requires ESR measurements made at liquid N_2 or liquid He temperature on frozen aqueous solutions. However, large metal–organic molecules cannot tumble rapidly in solution, thus the ESR spectrum at RT is often similar to that observed for powders or frozen solutions.

The major limitation of an ESR experiment is the resolution of the signal linewidths that may overlap to such an extent that information is lost. The resonance linewidth is affected by two mechanisms, namely, homogeneous and inhomogeneous broadening. The most important homogeneous broadening effects are: (1) 'spin-lattice relaxation,' by which an electron at a higher energy level returns to the ground state by losing energy to its environment (i.e., the 'lattice'), either in a short time, thus producing linewidth broadening, or in a long time, which results in narrower linewidths; and (2) microwave-power saturation, which may arise if a sufficiently large microwave power is applied which tends to equalize the electron populations of the two levels, thus decreasing signal intensity and increasing linewidth. In the absence of saturation, the ESR signal intensity should increase as the square root of the microwave power. Inhomogeneous broadening arises from nonuniformities in the magnetic field throughout the sample resulting from other neighboring paramagnetic species or from neighboring magnetic nuclei, or from dipolar interactions between unlike spins. These effects, often referred to as spin–spin interactions, are random in direction and result in a merging of individual resonant lines or spin packets into a single overall line or envelope, with loss of line resolution and related information.

Two techniques are potentially useful to overcome this limitation: (1) the electron nuclear double magnetic resonance (ENDOR) spectroscopy, which is a combination of ESR and nuclear magnetic resonance (NMR) techniques; and (2) the electron spin echo (ESE) spectroscopy, which is a time-domain electron magnetic resonance technique. ENDOR and ESE have not yet been applied to soil chemistry, thus major scientific activity is expected to occur in this area.

Determination and Interpretation of Spectral Parameters

Once the ESR spectrum is measured, values of spectral parameters can be determined accurately and rigorously by comparing it with a computer-simulated spectrum for the species being studied, which is calculated with the use of trial parameters and the expressions for the magnetic fields at which transitions occur. Then, the procedure is repeated until satisfactory agreement is found between the two spectra. In practice, the elaborate computer simulation procedure may be avoided by the relatively simple, even though not rigorous, computation of spectral parameters directly from the experimental ESR spectrum and from spectrometer measurement settings using the equations described in detail below.

Once the ESR parameters are obtained, they can be related to the nature of paramagnetic species, metal oxidation state, and type and site symmetry of metal binding by using rigorous physicochemical approaches and complex mathematical elaborations. This procedure can be avoided by empirical comparison and correlation of experimentally obtained ESR spectral parameters with ESR data available on similar model, synthetic, or natural systems, which can provide the information of chemical and structural interest to the soil scientist.

Whole Soil

The ESR spectra of whole-soil samples commonly show a very broad and intense ferromagnetic resonance peak centered at approximately $g = 2.00$, which can be attributed to strongly magnetic materials consisting of randomly oriented crystals and/or polycrystalline powdered particles of different shapes and random orientation, such as iron oxyihydroxides and minerals rich in Mn, Ti, and other transition metals. High porosity and crystal defects and/or impurities can further increase ESR signal linewidth. As a result of these effects, ESR of whole soil can provide very limited chemical and structural information, thus soil components, i.e., HS and their most important fractions, humic acid (HA) and fulvic acid (FA), metal–HS complexes, clay minerals and metal–clay associations, must be isolated from the whole soil and studied separately.

Organic Free Radicals in Soil Humic Substances

ESR Spectra and Parameters

Humic substances are the most abundant and most chemically and biologically active fraction of the nonliving organic matter in soil, whose macromolecules typically contain indigenous free radical moieties. The ESR spectrum of HS free radicals is usually obtained at RT on solid or water-dissolved samples of unfractionated HS, and HA and FA fractions and subfractions. The magnetic field is swept over a relatively narrow scan range (generally 5–10 mT) through the field at which the free electron resonates until the resonance condition is met, which generally occurs at a field of about 340 mT, which corresponds to a g-value close to that of the free electron. Typical spectrometer settings and operating conditions used in the measurement of ESR spectra of HS free radicals are: microwave frequency close to 9.5 GHz; microwave attenuation, 13 dB, corresponding to a microwave power of about 10 mW, at which the signal is generally least saturated; and modulation amplitude, 0.63 mT. A typical ESR spectrum of HA and FA free radicals features a single-line resonance, devoid of any structure (**Figure 3a**), while a partially resolved hyperfine structure is rarely observed (**Figure 3b**).

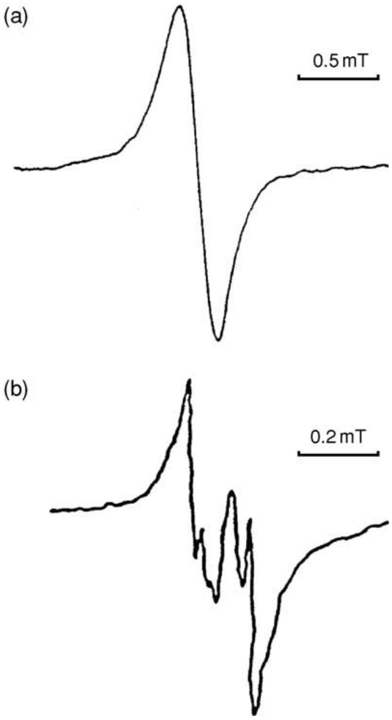


Figure 3 Typical single-line electron-spin resonance (ESR) spectrum of organic free radicals in soil humic acids and fulvic acids (a) and four-line ESR spectrum observed for some acid-boiled humic acids (b). (Adapted with permission from Atherton NM, Cranwell PA, Floyd AJ, and Haworth RD (1967) Humic acid. I. ESR spectra of humic acid. *Tetrahedron* 23: 1653–1667.)

At most, three spectral parameters of importance for the characterization of the nature, origin, and properties of HS free radicals can be derived from ESR spectra, which are the g-value, the width of the absorption line, i.e., the linewidth, and, rarely, the hyperfine structure. The g-value can be accurately approximated from the magnitudes of the magnetic field at which the resonance occurs for the sample and for a standard of known g-value, usually N,N-diphenylpycrylhydrazil (DPPH) diluted in powdered KCl ($g_{\mathrm{DPPH}} = 2.0036$). A small amount of standard contained in a capillary tube can be taped to the sample tube and the signal of the standard is used as a field 'marker.' Since υ_0 is identical for both, the sample and standard, g is readily calculated from the ratio of field positions using the relationship derived from eqn [2]:

$$g_{\mathrm{u}} = g_{\mathrm{k}} H_{\mathrm{k}}/H_{\mathrm{u}} \qquad [3]$$

where 'u' (unknown) and 'k' (known) refer to the sample and standard, respectively, and H_{k} can be calculated from eqn [2] using υ_0 and g_{k}. The width of the resonance line, or linewidth, ΔH, is generally measured in tesla or in gauss (1 gauss = 10^{-4} T), as the peak-to-peak separation of the first derivative ESR signal. The hyperfine splitting is measured as the separation (in gauss or in tesla) between the hyperfine lines.

The concentration of unpaired electrons, i.e., of organic free radicals in HS can be determined by comparing its signal area with that of a standard chart containing a known content of paramagnetic centers, e.g., the 'strong pitch' supplied by the manufacturer, or DPPH. Rigorously, a double integration of the first derivative curves should be performed to obtain the corresponding areas. However, since the line shape of the sample and standard are generally both of the Lorentzian type, the double integration can be avoided by simply calculating each area as the product of the height (h) and square of the width (w) of the first derivative signal. If hyperfine splitting of the sample occurs, the area under the hyperfine lines must be summed to obtain the total area to be considered. The spin concentration, expressed in spins per gram, can then be calculated by:

$$(\mathrm{spin\,g^{-1}})_{\mathrm{u}} = [(\mathrm{spin\,g^{-1}})_{\mathrm{k}}\,(hw^2)_{\mathrm{u}}\,(\mathrm{gain})_{\mathrm{k}}]/[(hw^2)_{\mathrm{k}}\,(\mathrm{gain})_{\mathrm{u}}\,q_{\mathrm{u}}] \qquad [4]$$

where ‘u’ and ‘k’ are as above (eqn [3]), gain is the signal amplifier gain used, and q_u is the weight (in grams, moisture- and ash-free) of the sample.

Since this procedure requires that the sample and standard be measured in an identical environment, i.e., the same matrix, geometry, volume, temperature, and spectrometer settings, either a double-cavity instrument or concentric sample tubes should be used. However, more often, separate tubes for the sample and standard are used, placing them alternatively in exactly the same position in the resonant cavity.

Interpretation of ESR Parameters

The *g*-values commonly measured for organic free radicals in soil HAs and FAs range between 2.0030 and 2.0050, which are consistent with indigenous semiquinone radical moieties conjugated to aromatic rings (e.g., $g = 2.0041$ for 9,10-anthraquinone), although a contribution from methoxybenzene radicals (*g*-values range, 2.0035–2.0040) and N- or S-associated radicals ($g = 2.0031$–2.0037) cannot be excluded.

The linewidths of the ESR signal of HA and FA free radicals generally range between 3.5 and 7.5×10^{-4} T for solid samples and between 2.0 and 3.0×10^{-4} T for the corresponding water solutions. ESR linewidths do not show any particular dependence on the nature and origin of FAs and HAs, but they are generally slightly greater for FA than for HA from the same source. Factors affecting ESR signal linewidth are free-radical concentration and aggregation state of the sample, interactions with solvent and/or metal ions, power-saturation effects, and temperature.

The ESR spectra of HA and FA free radicals rarely show a hyperfine structure. For example, some acid-boiled HAs feature a four-line hyperfine structure (**Figure 3b**) attributed to the interaction of the unpaired electron with two nonequivalent H nuclei, whereas an oxidized soil FA exhibits a three-line spectrum ascribed to the interaction of the unpaired electron of a semiquinone O atom with two adjacent equivalent H nuclei.

The concentration of organic free radicals in soil HAs and FAs depends on their nature and origin, and generally ranges between 10^{16} and 10^{18} spins g^{-1}. FAs usually show one-third to one-fifth the spin content of HAs from the same source. The spin concentration of FAs measured in solution at neutral and alkaline pHs is always greater than that measured in the solid state. Several factors affect the spin content measurement, including power-saturation effects, the presence of pronounced shoulders associated with a broad signal, and, in the case of solution samples, solvent, pH, and time. Caution should thus be used in evaluating and comparing the spin contents of HS. A reasonable reliability and comparability can be expected, however, in the evaluation of changes in spin concentrations occurring in a certain sample subjected to variations of various physical and chemical conditions.

Factors Affecting ESR Parameters of Humic Substances

The changes of concentration, *g*-value, and linewidth of free radicals in HS subjected to variations of factors such as pH, ionic strength, state of aggregation, hydrolysis, alkylation, redox potential, irradiation, temperature, and humidity can provide valuable insights into their molecular features.

Increase in pH, chemical reduction, acid hydrolysis, or visible- and ultraviolet (UV)-light irradiation enhances the free radical concentration of HS, while leaving almost unaltered the *g*-value and the signal linewidth. However, in all cases except acid hydrolysis, the spin content increase is not sustained in time, i.e., it returns gradually, and almost reversibly, to a value similar to that before the treatment. These results suggest that short-lived, ‘transient’ free radicals chemically similar to stable native radicals are produced in HS upon any of the mentioned treatments.

Increase in temperature up to 450°C or exposure to high-energy irradiation (gamma rays) causes a marked increase in spin content, broadening of linewidth, and decrease in *g*-value in solid HAs. These changes, which are accompanied by loss of O and increase in C content, can be ascribed to homolytic bond-cleavage and subsequent delocalization and stabilization of newly produced free electrons from semiquinonic O atoms to aromatic C atoms.

Chemical or electrochemical oxidation generally produces a time- and pH-dependent decrease in the free radical content of water-dissolved HAs and FAs, which is reversed by a subsequent reduction treatment. A decrease in spin content is also observed upon increase of neutral electrolyte concentration in FA and HA solutions at pH close to neutrality. However, no changes in *g*-values or linewidths are observed upon any of these treatments. A decrease in both organic free-radical concentration and ESR signal linewidth is observed with increasing humidity of solid HA samples, which is almost reversible upon redrying of the samples. Contrasting results are obtained for the effect of methylation on the free radical content of HS, i.e., it decreases dramatically for podzolic soil HAs and FAs, whereas it increases for several other soil HAs and FAs.

Structural Implications

The free radical concentration is correlated to other compositional and structural parameters of HS, either positively, e.g., to the E_4/E_6 ratio (ratio of absorbance at 465 nm and 665 nm), absorbance at 465 nm, atomic C/H and O/C ratios, O percentage, phenolic OH content, and aromaticity, or negatively, e.g., to H percentage and aliphaticity. These results imply that the free radical content is directly related to the dark color, aromatic or aliphatic character, molecular complexity, and particle size of HS. A good correlation also exists between the free radical concentration and the degree of humification of HS in peat soils and in organic-amended soils.

Accumulated ESR evidence strongly supports an indigenous quinone (Q)-hydroquinone (H_2Q) system as the principal responsible for the generation and stabilization of semiquinone radicals ($HQ^\bullet$) and radical anions ($Q^{\bullet-}$) in HS macromolecules, which is based on the simple electron donor–acceptor (or charge-transfer) reversible mechanism:

$$Q + H_2Q \rightleftarrows 2HQ^\bullet \underset{2OH^-}{\overset{2H^+}{\rightleftarrows}} 2Q^{\bullet-} \qquad [5]$$

The large decrease in spin content observed in podzolic HAs and FAs as a consequence of selective blocking of phenolic OH groups by methylation confirms these groups as the most important electron donors responsible for the formation and existence of organic free radicals in HS. A quinone content of $0.5\,\text{mmol}\,g^{-1}$, typical of most HAs, would theoretically yield up to $6 \times 10^{20}\,\text{spins}\,g^{-1}$ of semiquinone anion at alkaline pH and in the presence of a donor group such as hydroquinone. This effect can increase from two to five orders of magnitude the amount of stable free radicals in HS, with relevant implications in their chemical and biochemical reactivity. Further, semiquinone radicals can be generated by reduction or photoirradiation of quinones in solution in the presence of electron donors.

Interactions of Humic Free Radicals with Organic Chemicals and Metal Ions

The ESR technique can usefully be applied for evaluating the role of organic free radicals in the interaction of HS with organic chemicals, such as pesticides. For example, the increase in free radical concentration measured in the interaction products between HAs and *s*-triazine or substituted urea herbicides provides evidence of the occurrence of an electron donor–acceptor mechanism, with formation of charge-transfer complexes. Differently, the products of interaction of chlorophenoxyalkanoic herbicides with HA show a considerable decrease in free radical concentration, which suggests the occurrence of homolytic cross-coupling reactions between HA free radicals and free radical intermediates generated in the preliminary chemical, photochemical, and/or biological degradation of the herbicide.

The ESR spectroscopy can also be used to study the effect of some metal ions on HS free radicals. In general, addition of diamagnetic metal ions such as Na^+, Zn^{2+}, and Al^{3+} to HA or FA solutions does not affect their free radical content, whereas addition of paramagnetic metal ions such as Fe^{3+}, Cu^{2+}, Mn^{2+}, VO^{2+}, Ni^{2+}, and Co^{2+} causes a marked decrease in HS free radical content as a function of the metal species and HS origin. In studies of podzolization processes, ESR data suggest that recently formed, low-molecular-weight HS cannot undergo further polymerization because their free radicals combine with Fe^{3+} and move to the Bh horizon. In another study, ESR spectroscopy reveals the formation of semiquinone radicals by single electron transfers occurring at goethite and manganese oxide surfaces during the oxidation of hydroquinone. Further, ESR spectroscopy can be used to characterize the nature of HS formed by oxidative polymerization of phenolic and other organic compounds in the presence of natural clays, soils, soil oxides, and other metal-ion catalysts.

Paramagnetic Transition Metal Ions in Soil Constituents

The ESR technique can be successfully applied to study natural complexes of paramagnetic metal ions of importance in soil, such as Fe^{3+}, Cu^{2+}, Mn^{2+}, and V^{4+}, with HAs, FAs, and plant litters, and structural paramagnetic metal ions such as Fe^{3+} in soil minerals. In particular, ESR analysis can provide unique information on the identity and oxidation states of metal ions, metal binding site(s) including ligand types, coordination and symmetry, and stability of metal–organic complexes.

ESR Spectra and ESR Parameters

The ESR spectra of paramagnetic metals in soil organic and mineral constituents are usually obtained at either RT on powder samples or at liquid N_2 temperature (77 K) on powders or frozen (77 K) solution samples. Usually, the magnetic field is initially scanned tentatively over a wide range (from 0 to 1 T), then an enlarged spectrum is recorded over a narrower field (0.2 or 0.1 T) comprising the region where resonances appear to allow for their detailed analysis and interpretation. Measurement conditions commonly employed are: microwave frequency, approximately 9.2 GHz (spectra at 77 K) or 9.8 GHz (spectra at RT); microwave attenuation, 13 dB (microwave power oscillates between 9.0 and 9.2 mW); and modulation amplitude, between 0.63 and 4 mT,

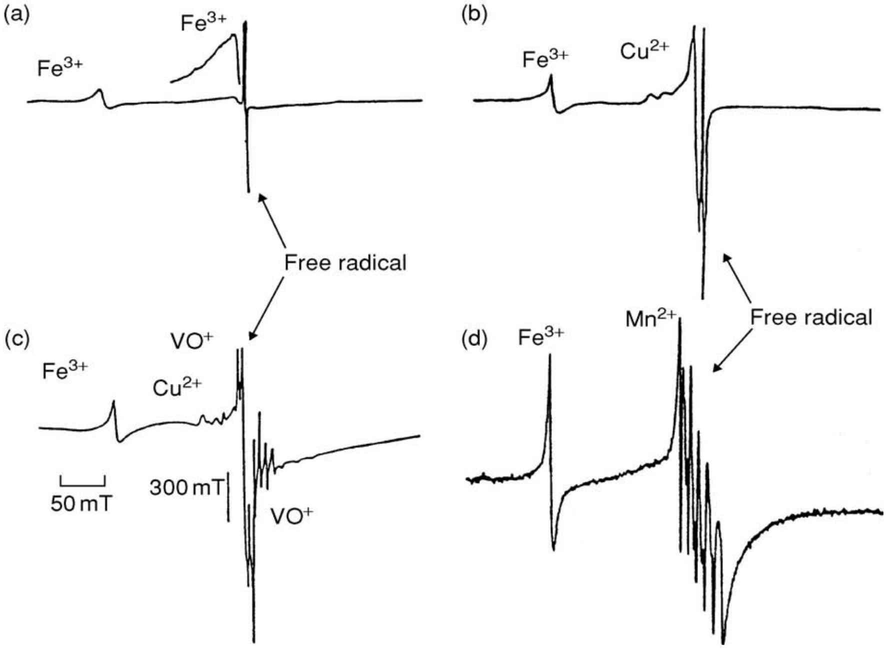


Figure 4 Representative wide scan range (800 mT) electron-spin resonance spectra at 77 K of a Mollisol humic acid from the IHSS Reference and Standard collection (a) (inset, higher gain); a Paleosol humic acid (b); a loam soil humic acid (c); and an aqueous extract of decomposing leaf litter from a forest soil (d). (Reproduced with permission from: Senesi N (1996) Electron spin (or paramagnetic) resonance spectroscopy. In: Sparks DL (ed.) *Methods of Soil Analysis*: *Chemical Methods*, pp. 323–356. Book series no. 5. Madison, WI: Soil Science Society of America/American Society of Agronomy.

according to the metal ion being measured. Four examples of typical ESR spectra of metal–organic complexes naturally occurring in isolated soil organic fractions are shown in Figures 4 and 5.

Three types of spectral parameters can be obtained from ESR spectra of metal–organic complexes: (1) the *g*-value(s) of the metal(s) present in the sample; (2) the hyperfine coupling constant(s), *A*, if the metal nucleus has a nonzero spin; and (3) the ligand superhyperfine splitting(s) if the unpaired electron of the metal ion is delocalized on to magnetic nuclei of surrounding ligands. Rarely, resonance lines due to 'forbidden transitions' are observed, e.g., for Mn^{2+}.

The *g*-values and hyperfine and superhyperfine constants, *A*, can be calculated from experimental spectral data and spectrometer setting values used according to the standard equations:

$$g = h\nu_0/\beta H_0 = 0.714484\,\nu_0/H_0 \qquad [6]$$

$$A(\mathrm{cm}^{-1}) = A(\mathrm{MHz})/c = 2.80247\,(a\,g)/(c\,g_e) = 0.469766\ 10^{-4}\,a\,g \qquad [7]$$

where ν_0 (megahertz) is the microwave frequency value at the resonance condition; H_0 is the value of the magnetic field at which the resonance is centered (on calibrated chart paper); g_e (2.00232) is the *g*-value of the free electron; *g* is calculated by eqn [6]; *a* is the hyperfine splitting measured as the peak-to-peak separation (in 10^{-4} T) between the hyperfine lines in the experimental spectrum; and *c* is the speed of light in a vacuum. Accuracy of magnetic field calibration of chart paper is checked by using a suitable standard.

Ferric iron The ESR spectra of HAs, FAs, and decomposing litter layers generally exhibit an asymmetrical isotropic resonance line at about $g = 4.2$, which is consistent with high-spin (five unpaired d-electrons) Fe^{3+} ions held in tetrahedral or octahedral sites of low-symmetry (rhombic) ligand field possibly by carboxylic and/or phenolic hydroxyl groups (Figure 4). This form of Fe exhibits considerable resistance to proton and metal exchange, and to chemical reduction, which suggests that Fe^{3+} is strongly bound and protected in inner-sphere complexes in HS.

A very broad signal near $g = 2$ is often exhibited by HS (Figure 4a, inset), which possibly arises from extended spin–spin coupling of various neighboring Fe^{3+} ions. Iron in such sites is easily reduced by chemical agents and easily extracted by complexing agents, thus suggesting that it is weakly bound on external surfaces of HS. Two weak resonances near $g = 9$ and $g = 6$ are sometimes observed in ESR spectra of HS, which possibly arise, respectively, from Fe^{3+} in sites with near orthorhombic symmetry and from high-spin Fe^{3+} in largely distorted, axially symmetric crystal fields.

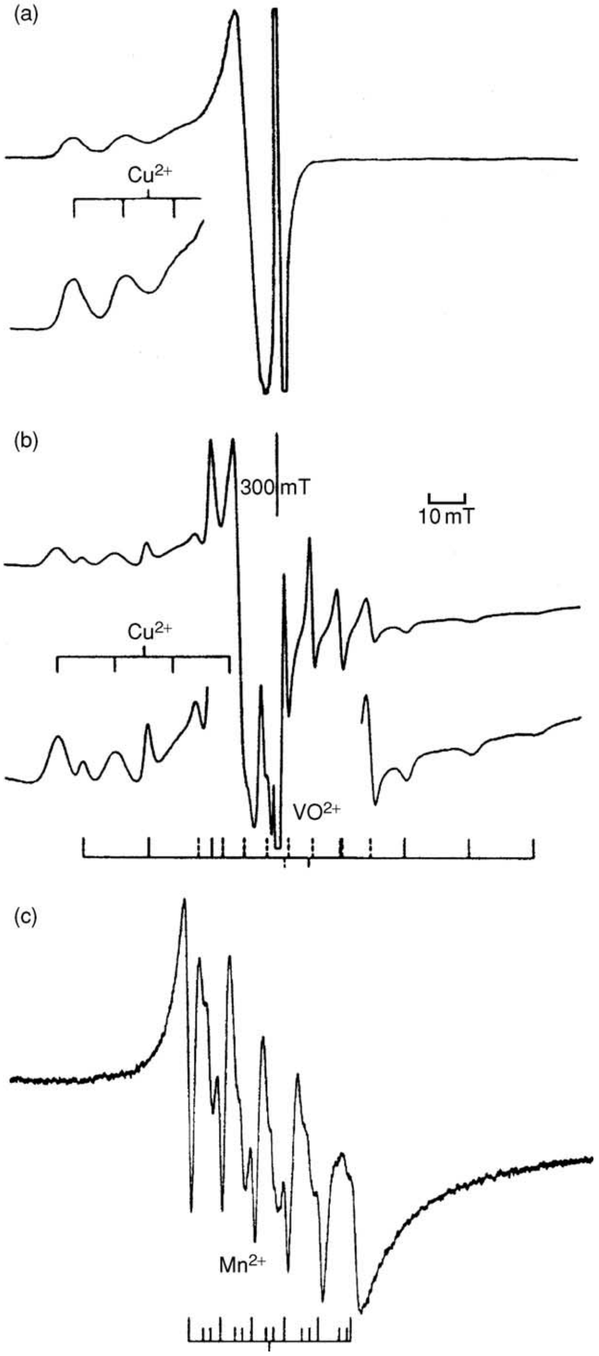


Figure 5 (a)–(c) Same electron-spin resonance spectra as in Figure 4b,c,d, respectively, but recorded on an enlarged scan range (200 mT). (Reproduced with permission from: Senesi N (1996) Electron spin (or paramagnetic) resonance spectroscopy. In: Sparks DL (ed.) *Methods of Soil Analysis*: *Chemical Methods*, pp. 323–356. Book series no. 5. Madison, WI: Soil Science Society of America/American Society of Agronomy.

An ESR signal at $g = 4.3$ can be observed in layer silicates such as kaolinites, vermiculites, micas, and smectites, which are generally composed of an anisotropic and an isotropic g-factor, and possibly arise from structural octahedral $FeO_4(OH)_2$ groups. Soil clays such as vermiculites and weathered phlogopite often exhibit a strong ferrimagnetic resonance at $g = 2$ that can be attributed to ferrimagnetic clusters of structural Fe^{3+} and/or ferric oxides and tends to obscure other spectral details. No direct ESR evidence of Fe^{2+} species is obtained in soil mineral and organic components.

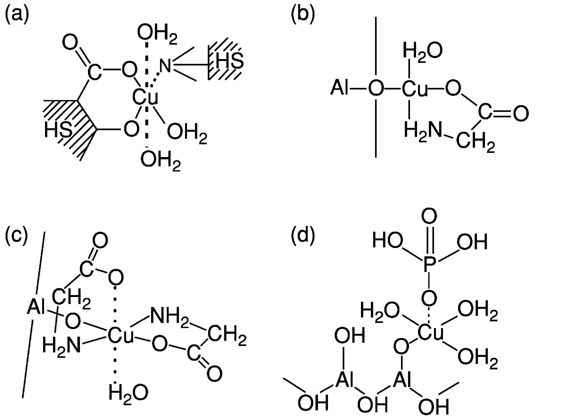


Figure 6 Models for Cu^{2+} complexes in humic substances (a), on surfaces of gibbsite (b), and boehmite (c) in the presence of glycin, and on aluminum oxides and allophanes in the presence of phosphate (d).

Divalent copper Soil HS often exhibits an anisotropic rigid-limit spectrum of the axial type in the $g = 2$ region (Figures 4b,c, and 5a,b), which is ascribed to Cu^{2+} ions. The nuclear spin of Cu is $I = 3/2$, thus its ESR spectrum should be split into four (i.e., $2I + 1$) features at both $g_{\|}$ and $g_{\perp}$. However, only the component at $g_{\|}$ is generally resolved partially into a quadruplet (Figure 5a (and inset), b (and inset)), while the splitting of the $g_{\perp}$ component is not resolved. The ESR parameters of this spectrum are consistent with a $d_{x^2-y^2}$ groundstate for Cu^{2+} ions held in square planar (distorted octahedral) coordination sites (tetragonal symmetry) as inner-sphere complexes with either only O ligands, such as carboxyls, phenolic hydroxyls, carbonyls, and, often, water molecules, or both O and N ligands (Figure 6a), or even only N ligands (i.e., a tetraporphyrin site). The resolved pattern observed in some cases at $g_{\perp}$ is ascribed to superhyperfine coupling of the Cu^{2+} unpaired electron to N ligand nuclei ($I = 1$).

The values of spectral parameters also provide evidence of a variable covalent bond contribution (i.e., delocalization of the unpaired electron toward the ligands) for Cu^{2+} in HS. At small Cu^{2+} loading, covalent bonding is favored and complexation to amine-N groups is preferred to O ligands; whereas large Cu^{2+} loading in HS determines the formation of small covalency binding of Cu^{2+}, largely to

O-containing ligands. The accurate analysis of Cu^{2+} ESR patterns at $g_{\|}$, possibly with the aid of a computer-simulated spectrum, often reveals the presence of two (or more) superimposed quadruplets, which proves the existence of different binding sites for Cu^{2+} in HS.

Vanadyl ion Soil HAs and FAs may feature a richly structured but relatively well-resolved pattern at about $g = 2$ (Figures 4c and 5b), comprising two distinct, overlapping rigid-limit spectra of the axial type. The one is the typical pattern of Cu^{2+} described previously, and the other consists of two superimposed hyperfine octuplets corresponding to the $g_{\|}$ and $g_{\perp}$ components of VO^{2+} ions (nuclear spin of V, $I = 7/2$) rigidly bound in square planar coordination sites as inner-sphere complexes with phenolate and possibly N ligands (large covalency, tightly bound forms) and/or surface carboxylate groups and water molecules (small covalency, relatively labile and exchangeable forms).

Divalent manganese The ESR spectra obtained for decomposing leaf litters (Figures 4d and 5c) and some soil HAs and FAs feature a well-resolved isotropic pattern in the region $g = 2$, which consists of six almost equally spaced principal lines and, possibly, 10 secondary lines (corresponding to forbidden transitions) of lesser intensity. The ESR parameters of such spectra are consistent with high-spin hexahydrated Mn^{2+} ($I = 5/2$) ions in outer-sphere complexes, bound by electrostatic forces to six O atoms of carboxylate and phenolate groups in a distorted octahedral environment.

Spin-Derivatization Studies

Spin-Labeling and Spin-Trapping

The spin-labeling technique, which consists in the attachment of a simple and stable paramagnetic species, such as the nitroxide radical, to the compound of interest, is a powerful ESR method for evaluating the dynamics of HS macromolecules in solution, and, in turn, for providing information on the aggregation state, molecular conformation, and micellar character of HS. Spin labels can also be used to investigate anisotropic molecular motion on mineral surfaces, which provides information on the nature of adsorption processes not obtainable by other methods.

An interesting illustration of this technique is the use of two organic anionic nitroxide spin labels, provided with either a carboxylate or an organophosphate functional group, to study the nature of bonding to noncrystalline alumina, boehmite, and gibbsite in aqueous suspensions. The ESR analysis reveals that both species are rapidly adsorbed on to the large surface area of alumina and boehmite, whereas only the organophosphate is adsorbed on gibbsite, with a loss in rotational motion, especially of the carboxylate. The values of motional restriction of organic radical anions on surfaces suggest that boehmite adsorbs carboxylate largely by ligand exchange with a single-surface OH and organophosphate by bidentate binding, whereas noncrystalline alumina binds weakly carboxylate by nonspecific electrostatic forces in addition to ligand exchange of surface OH.

The ESR spectra of the spin-label probe 5-SASL (stearic acid spin-label with nitroxide free radical in position 5 of the hydrocarbon chain) in HA suspensions suggest its bonding with surface hydrophobic sites of HA below pH 5, whereas these sites appear not available for bonding above pH 5. Kinetics of release of the nitroxide spin probes, neutral Tempol and cationic Tempamine, from Ca-hectorite aggregates and/or pastes measured by ESR suggest that the probes are not durably sequestered. The use of spin-labeled xenobiotics, e.g., pesticides, may represent a promising application of ESR spectroscopy to study their distribution, fate, and availability in soils.

The technique of spin-trapping involves the reaction of the ‘spin trap,’ such as nitrosobenzene, 2-methyl-nitrosopropene, 5,5-dimethylpyrroline 1-oxide (DMPO), or phenyl-*N*-*t*-butylnitrone (PBN), with a short-lived radical, e.g., produced in HS by light irradiation, reduction, or raising the pH, to yield a relatively stable radical having an ESR spectrum that allows identification and quantification of the short-lived radical trapped.

Metal Spin Probes

Complexation chemistry of humic substances The ESR analysis of synthetic complexes obtained by reaction of a paramagnetic metal ‘probe’ such as Cu^{2+}, Mn^{2+}, or VO^{2+}, with natural soil HS can provide useful information on the chemistry of ‘residual’ complexing capacity of HS toward metal ions and on the stability of formed complexes. For example, when an FA is doped with 5.5–50.1% Fe^{3+}, the $g = 2$ signal is enhanced relative to the $g = 4$ signal, which suggests that Fe^{3+} is preferentially bound in easily exchangeable forms to surface sites of FA. Evidence of inner-sphere complexes of Mn^{2+} coordinated octahedrally, possibly with carboxyl, phenolic hydroxyl, and/or a carbonyl group, is obtained for some soil HAs doped with Mn^{2+} at high pH (>8) or high temperature (>50°C). Rotational correlation times obtained by ESR analysis combined with gel-filtration chromatography data allow measurement of the dynamics of motion and stoichiometry of FA–VO^{2+} complexes in solution, and give information on their molecular conformation and aggregation properties. Quantitative ESR spectroscopy can also be used as a more sensitive, convenient, and faster technique to determine weighted-average equilibrium constants (K_c) for water-soluble Mn^{2+}–FA complexes, which are in excellent agreement with K_c values determined by ion-exchange methods.

Ion exchange on layer silicates ESR spectroscopy confirms that metal ions such as Cu^{2+}, Mn^{2+}, VO^{2+}, and Cr^{3+} retain their inner hydration sphere and a great degree of rotational mobility on exchange sites of layer silicate clays. For example, ESR spectra of exchangeable Cu^{2+} and Mn^{2+} on kaolinite indicate the presence of planar $Cu(H_2O)_4^{2+}$ ions oriented parallel to the surface and possessing a great degree of mobility. In contrast, an orientation-dependent, rigid-limit ESR spectrum of Cu^{2+} is obtained for hydrated Cu^{2+} ions adsorbed on hectorite when the interlamellar spacing on the clay is limited to the equivalent of one or two molecular layers of water, which suggests a great degree of motional restriction for Cu^{2+} ions.

Chemisorption on mineral surfaces ESR analysis proves that paramagnetic metal ions at trace levels are bound rigidly in chemisorbed, nonexchangeable forms at isolated surface sites of soil oxides, hydroxides, and aluminosilicates, whereas they are sorbed in exchangeable forms to layer silicate clays. For example, Cu^{2+} is proven to chemisorb at isolated sites on noncrystalline $Al(OH)_3$ and microcrystalline AlOOH, possibly by formation of one or two direct bonds between surface Al-O groups and Cu^{2+}, which is favored at high pH. At low pH, gibbsite is able to chemisorb small amounts ($<0.5\,\text{mmol}\,100\,\text{g}^{-1}$) of monomeric Cu^{2+} that is oriented with its z-axis perpendicular to the (001) planes of the mineral. At pH >5, the appearance of a broad, featureless resonance and the reduction of the rigid-limit spectrum of Cu^{2+} suggest the presence of different, possibly hydrolyzed and polymerized forms of Cu^{2+} on the surface. The large decrease in g-values and increase in $|A_{\|}|$-values of the low-pH, gibbsite-chemisorbed Cu^{2+} upon exposure to NH_3 vapors suggest that Cu^{2+} may form a complex with NH_3 while remaining rigidly bound to the oxide surface. ESR data indicate that Cu^{2+} ions in the presence of the chelating ligand glycine absorb on microcrystalline gibbsite and boehmite in the form of ternary complexes in which Cu^{2+} coordinates simultaneously with one surface hydroxyl and one (on gibbsite) or two (on boehmite) glycine molecules, with the orientation of the Cu^{2+} z-axis normal to the (001) sheets of the mineral (Figure 6b,c).

ESR studies reveal that Cu^{2+} can be adsorbed by a direct bond (chemisorption) in nonexchangeable forms on allophanes and imogolite both by a preferential binuclear mechanism involving two adjacent AlOH groups and on a weaker type of binding site probably involving isolated AlOH or SiOH groups. Here also NH_3 is able to readily displace H_2O and OH-ligands from chemisorbed Cu^{2+}, leading to the formation of ternary Cu^{2+}–ammonia–surface complexes. At low pH, Cu^{2+} is also shown to chemisorb strongly to bidentate sites of titanium dioxide, whereas at higher pH a weaker complex is formed involving single Ti-OH groups.

The ESR technique shows that Cu^{2+} is adsorbed on aluminum oxides and allophanes in the presence of phosphate, forming a ternary complex with the phosphate coordinated to the axial position of a surface-bound Cu^{2+} (Figure 6d). Large amounts of phosphate suppress Cu^{2+} adsorption, apparently by blocking the coordination of Cu^{2+} to surface AlOH groups. ESR evidence is also provided that VO^{2+} can be coadsorbed with phosphate on boehmite and aluminosilicates as an inner-sphere complex. Finally, ESR proves that Cr^{3+} in the presence of selenite, phosphate, and fluoride is coadsorbed on hectorite and montmorillonite, with formation of ternary surface complexes.

Thermodynamic constants from ESR parameters Metal spin probes can be used to estimate thermodynamic stability constants of metal–surface complexes from the ESR parameter $g_{\|}$. For example, a linear relationship is found between the $g_{\|}$-value and the corresponding formation constants of square planar Cu^{2+} complexes with hydrous Al_2O_3, TiO_2, and some silicas in the presence of bidendate ligands. A decrease in the $g_{\|}$ value by 0.1 units corresponds to an increase in thermodynamic stability of about eight orders of magnitude. On these bases, a major revision of current concepts of cation adsorption on layer silicates is possible.

List of Technical Nomenclature

β	Electron Bohr magneton
ΔE	Absorption of energy by the electron
ΔH	Width of the resonance line, or linewidth
μ_z	Electron magnetic moment
A	Hyperfine coupling constant
DC	Direct current
DMPO	5,5-Dimethylpyrroline-1-oxide
DPPH	N,N-Diphenylpycrylhydrazil

E	Energy for the unpaired electron
ENDOR	Electron nuclear double resonance
EPR	Electron paramagnetic resonance
ESE	Electron spin echo
ESR	Electron-spin resonance
FA	Fulvic acid
g	Electronic-splitting factor, or g-value, or g-tensor
$g_{\parallel}$	Component of the g-tensor parallel to the symmetry axis
$g_{\perp}$	Component of the g-tensor perpendicular to the z-axis
g_{iso} or g_0	Isotropic g-factor
g_{xx}, g_{yy}, and g_{zz}	Components of the g-tensor along the axes x, y, z
h	Planck constant
HA	Humic acid
H, H_o	Strength of the static magnetic field
H_2Q	Hydroquinone
$HQ^{\bullet}$	Semiquinone radical
HS	Humic substances
I	Nuclear spin
K_c	Weighted-average equilibrium constant
M_z	Electron-spin angular momentum
NMR	Nuclear magnetic resonance
PBN	Phenyl-*N*-*t*-butylnitrone
Q	Quinone
$Q^{\bullet -}$	Semiquinone radical anion
q_u	Weight of the sample
RT	Room temperature
S	Electron spin
5-SASL	Stearic acid spin-label with nitroxide free radical in position 5 of the hydrocarbon chain
T	tesla
v_0	Frequency of the alternating magnetic field

See also: **Clay Minerals**; **Fourier Transform Infrared Spectroscopy**; **Heavy Metals**; **Organic Matter:** Principles and Processes; Interactions with Metals; **Pollutants:** Persistent Organic (POPs); **Sorption:** Metals

Further Reading

Abragam A and Bleaney B (1970) *Electron Paramagnetic Resonance of Transition Ions*. Oxford, UK: Clarendon Press.

Berliner LJ (1976) *Spin Labelling: Theory and Applications*. New York: Academic Press.

Gordy W (1980) *Theory and Application of Electron Spin Resonance*. New York: Wiley-Interscience.

McBride MB (1986) Magnetic methods. In: Klute A (ed.) *Methods of Soil Analysis*, 2nd edn, part 1, pp. 219–270. Book Series No. 5. Madison, WI: Soil Science Society of America/American Society of Agronomy.

Poole CP (1983) *Electron Spin Resonance: Comprehensive Treatise on Experimental Techniques*, 2nd edn. New York: John Wiley.

Senesi N (1990) Application of electron spin resonance (ESR) spectroscopy in soil chemistry. In: Stewart BA (ed.) *Advances in Soil Science,* vol. 14, pp. 77–130. New York: Springer-Verlag.

Senesi N (1990) Molecular and quantitative aspects of the chemistry of fulvic acid and its interactions with metal ions and organic chemicals, part I, The electron spin resonance approach. *Analitica Chimica Acta* 232: 51–75.

Senesi N (1996) Electron spin (or paramagnetic) resonance spectroscopy. In: Sparks DL (ed.) *Methods of Soil Analysis: Chemical Methods*, pp. 323–356. Book Series no. 5. Madison, WI: Soil Science Society of America/American Society of Agronomy.

Senesi N and Steelink C (1989) Application of ESR spectroscopy to the study of humic substances. In: Hayes MHB, MacCarthy P, Malcolm RL, and Swift RS (eds) *Humic Substances. II. In Search of Structure,* vol. 2, pp. 373–407. Chichester, UK: John Wiley.

Thomson AJ (1990) Electron paramagnetic resonance and electron nuclear double resonance spectroscopy. In: Andrews DL (ed.) *Perspectives in Modern Chemical Spectroscopy*, pp. 295–320. Berlin, Germany: Springer-Verlag.

Weil JA, Bolton JR, and Wertz JE (1994) *Electron Paramagnetic Resonance: Elementary Theory and Practical Applications*. New York: John Wiley.

Electrostatic Double-Layer *See* **Cation Exchange**

ENERGY BALANCE

M Fuchs, Agricultural Research Organization, Bet Dagan, Israel

Introduction

The genesis of soils is the transformation of the lithosphere in contact with the atmosphere. It involves physical, chemical, and biological processes that depend on meteorological conditions. For this reason, the geographic distribution of soil types is related to the distribution of climatic zones. The variables characterizing the climate, temperature, humidity, precipitation, wind, and radiation take on the values that equilibrate the energy balance of the surface. The driving energy source is solar radiation. The atmosphere, water in its three phases, and life determine how components of the energy balance derive from the solar energy input. Temporal variation of incoming solar energy causes heating or cooling of the atmosphere, heating or cooling of the ground, evaporation or condensation of water, thawing or freezing of water, and photochemical reduction or oxidation. Spatial variation of the sun's energy input drives the active atmospheric circulation.

The Energy Input From the Sun

The Earth's surface intercepts permanently from the Sun a mean radiant energy flux density of $342\,\mathrm{W\,m^{-2}}$. This power input undergoes a predictable annual fluctuation of ±3.3%, because the Earth revolves around the Sun along an elliptic track. It reaches its maximum by 3 January and its minimum on 5 July. Unpredictable changes of solar activity cause additional variations amounting to less than ±1%.

The annual energy density reaching the top of the atmosphere has a latitudinal distribution. It decreases from $13.17\,\mathrm{GJ\,m^{-2}}$ (mean flux density $= 434\,\mathrm{W\,m^{-2}}$) at the equator down to $5.45\,\mathrm{GJ\,m^{-2}}$ (mean flux density $= 173\,\mathrm{W\,m^{-2}}$) at the poles. The latitudinal distribution of the daily amount varies in the course of the year, because the rotation axis of the Earth has a fixed declination of 23.5° on the plane formed by the Earth's track around the Sun. The instantaneous flux density fluctuates in the course of a day because of the Earth's rotation. The range of variation is between the extraterrestrial mean maximum of $1368\,\mathrm{W\,m^{-2}}$ and zero.

Energy Balance of a Planet Without Atmosphere

To illustrate the impact of the atmosphere on the energy budget of the Earth's surface, it is instructive to consider the energy balance of a planet on the same track around the Sun, but deprived of its atmosphere. In this case the only energy-exchange processes are by radiation:

$$(1 - r_\mathrm{S})R_\mathrm{S} - \epsilon_G \sigma T_\mathrm{G}^4 = 0 \qquad [1]$$

where r_S is the mean albedo or solar radiation reflection coefficient of the planet ground surface, R_S is the incoming radiation originating from the Sun ($342\,\mathrm{W\,m^{-2}}$ in the wavelength band from 0.2 to $3\,\mu\mathrm{m}$), σ is the Stefan–Boltzmann constant ($5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$), and ϵ_G is the emissivity of the planet's surface ($\cong 0.98$). The radiation emitted by the planet $\epsilon_\mathrm{G}\sigma T_\mathrm{G}^4$ is in the wavelength band between 5 and $100\,\mu\mathrm{m}$. It balances the solar radiation absorption of the surface. Setting the albedo of the planet's surface to be equal to that of the Earth as viewed from space, $r_\mathrm{S} = 0.37$, the planetary radiation loss of the surface $\epsilon_\mathrm{G}\sigma T_\mathrm{G}^4 = (1 - r_\mathrm{S})\ R_\mathrm{S} = 216\,\mathrm{W\,m^{-2}}$. The resulting mean Kelvin surface temperature T_G of the planet is 249.6 K or −23.6°C.

An increase of $1\,\mathrm{W\,m^{-2}}$ in solar activity believed to have occurred over the past 300 years would have produced a warming of 0.2°C in a planet without atmosphere. This is the probable temperature rise of the Earth that occurred during the 200 years preceding the acceleration of the warming trend observed since the early 1900s.

Energy Balance of the Earth's Surface

The presence of an atmosphere transforms the energy balance, because the molecules of gases and the solid particles in suspension absorb and scatter some of the incoming solar radiation. The absorbed energy heats gases and particles. In turn, the heated gases and particles emit long-wave radiation toward the surface and to outer space. According to Kirchhoff's law, the emissivity is equal to the absorptivity of radiation at the same wavelength, so the atmosphere absorbs a fraction of the radiation emitted by the Earth's surface, thereby preventing its escape to outer space. The absorbed radiation is partly radiated back to the ground. Turbulent convection carries some of the heat absorbed by the ground to the atmosphere and contributes to the temperature elevation of the atmospheric gases and of the particles held in suspension. Part of this heat is included in the back-radiation. The remainder is lost by radiation into space. Heat transport from the ground surface to the atmosphere implies that the ground is warmer than the atmosphere. As radiative emission is proportional to the fourth power of the Kelvin temperature, radiation from the atmosphere to outer space is less than the heat emitted from the ground. To compensate for this decreased global emission, the ground elevates its surface temperature to emit more radiation.

Early information on the energy balance of the Earth was derived from the analysis of meteorological and hydrological ground observations. Satellite measurements of the radiative components of the atmosphere, modeling of the radiative transfer through gases, and of atmospheric circulation are the sources of the current state of knowledge.

Figure 1 shows annual mean values of the main components of the Earth's energy balance. Of the 342 W m^{-2} of solar radiation reaching the top of the atmosphere, 205 W m^{-2} meets clouds that transmit 88 W m^{-2} to the Earth's surface and 77 W m^{-2} is reflected into space. One hundred and thirty-seven watts per square meter impinge on clear atmosphere and 110 W m^{-2} is transmitted to the surface. The mean atmospheric transmissivity τ_A for solar radiation is 0.58. The Earth's surface reflects 30 W m^{-2} into space; thus the net absorption of solar radiation is 168 W m^{-2}. The spectrum of solar radiation reaching the Earth's surface is referred to as short-wave radiation and includes ultraviolet, visible, and near-infrared radiation.

The presence of water on the ground surface and in the atmosphere has a major impact on the energy balance of the Earth. The 60% mean cloud cover enveloping the Earth reflects back to space 22% of the incoming solar radiation. Water vapor in the cloudless part of the atmosphere absorbs two-thirds of the 27 W m^{-2} of solar radiation absorbed by atmospheric gases; clouds absorb 40 W m^{-2}.

Terrestrial matter emits radiation in the spectral waveband from 5 to 100 μm. It is referred to as 'long-wave radiation' or 'far-infrared radiation.' The radiative emission of the atmosphere including clouds is 324 W m^{-2}. The Earth's surface absorbs α_G $R_A = 318$ W m^{-2} of this flux density (where $\alpha_G = 0.98$ is the long-wave absorptivity of the Earth surface) and reflects 6 W m^{-2}. The surface emission is

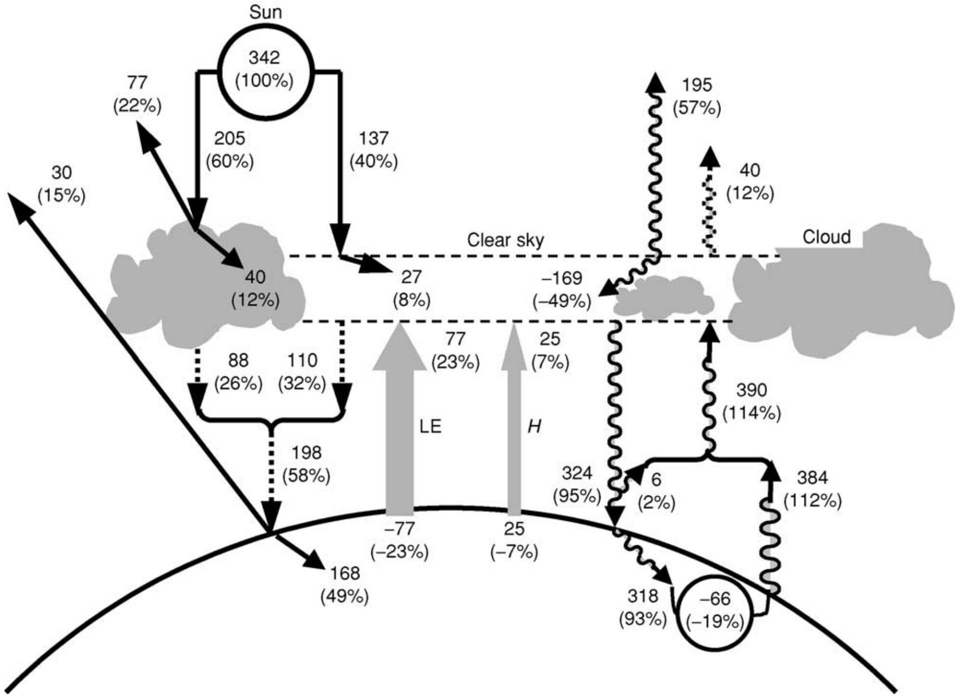


Figure 1 Mean annual values of the Earth's energy balance (main terms expressed in W m^{-2}). The numbers enclosed in parentheses are percentages of 342 W m^{-2} extraterrestrial solar radiation as 100. The straight-tailed arrows symbolize flux densities in the solar wave band (0.2–3 μm). Arrows with a wavy tail represent radiation emitted by the Earth. The wavelength range of this terrestrial radiation is from 5 to 100 μm. Vertical arrows show downward flux densities. Slanted upward arrows represent reflected flux densities. Slanted downward arrows indicate absorbed flux densities. Dotted-tailed arrows mark flux densities transmitted either upward or downward according to the direction of the arrowhead. The gray arrows indicate latent heat for evapotranspiration (LE) from water surfaces or water held in the soil, and *H* is the sensible heat convected from the surface to the atmosphere.

$\epsilon_G \sigma T_G^4 = 384\,\mathrm{W\,m^{-2}}$, leading to a mean temperature of the Earth: $T_G = 288.3$ K or. 15.1°C. The total long-wave radiation loss from the surface is $390\,\mathrm{W\,m^{-2}}$. The resulting net long-wave radiative loss of the Earth surface is $-66\,\mathrm{W\,m^{-2}}$.

As Kirchhoff's law states $\alpha_G = \epsilon_G = 0.98$, the summarized format of the Earth's surface radiation heat balance R_n reduces to:

$$R_n = \tau_A R_S + \epsilon_G (R_A - \sigma T_G^4) \qquad [2]$$

Comparison between T_G of the Earth (15.1°C) with T_G (−23.7°C) of the planet without an atmosphere gives an estimate of the global temperature elevation due to the radiative shielding by the atmosphere: gases and suspended liquid – solid particles. The radiative blanketing effect is best illustrated by comparing the net long-wave loss of $-216\,\mathrm{W\,m^{-2}}$ from the planet without atmosphere with the net long-wave loss of $-66\,\mathrm{W\,m^{-2}}$ from the Earth's surface. This radiative heat gain is commonly referred to as the 'greenhouse effect.' The wording is misleading, because the heating of a greenhouse exposed to solar radiation results from preventing turbulent convective heat transport and not from radiation trapping.

Gases and suspended particles emit $195\,\mathrm{W\,m^{-2}}$ of long-wave radiation to space. As the atmosphere receives more radiation from the Earth surface than it emits to space, its net long-wave radiation deficit is only $-169\,\mathrm{W\,m^{-2}}$. An additional $40\,\mathrm{W\,m^{-2}}$ is lost to space by transmission through the radiative window of the atmosphere. The release of radiatively active gases such as CO_2 and aerosols into the atmosphere diminishes this transmission. Yet, for a proper perspective on the significance of the possible effects, one should keep in mind that water vapor absorbs 60% of long-wave radiation emission blocked by the atmosphere as compared with only 26% absorbed by CO_2.

The mean value of R_n is $102\,\mathrm{W\,m^{-2}}$, but consensus among scientists on its value has not yet been reached. Most publications agree within a few percent on the solar radiation term $\tau_A R_S$, but estimates of the net long-wave radiative loss vary between −72 and $-40\,\mathrm{W\,m^{-2}}$.

The net radiation dissipates as turbulent sensible heat convection into the air (H) and latent heat for evapotranspiration (LE):

$$R_n + H + \mathrm{LE} = 0 \qquad [3]$$

The partition between H and LE depends on the availability of water at the ground surface. As oceans cover more than 70% of the Earth, water is readily available for evaporation.

The instantaneous energy balance for a local site also involves conductive heat flow G through the surface into or out of the ground subsurface. The heat flow is inward when the surface is warmer than the subsurface and outward for the opposite condition. Its magnitude depends on the thermal conductivity and the volumetric heat capacity of the ground. The depth of the subsurface layer affected by the diurnal change of the solar radiation on land is approximately 1 m. The corresponding annual variation penetrates to a depth of 20 m ($\approx$1 m $\times \sqrt{365}$). The heat penetration in oceans is at least one order of magnitude larger than on continental areas. As during the annual cycle heating and cooling of the Earth's surface have symmetrical amplitudes, the term G vanishes in eqn [3] applied to a local site. The term G also vanishes in eqn [3] when it is applied to the entire planet, because the distribution of heating and cooling on the rotating spherical Earth is symmetrical.

Based on observations of the hydrologic balance (worldwide distribution of precipitation and runoff from rivers flowing into the oceans) current estimates set mean evapotranspiration expressed as latent heat flux density to $-77\,\mathrm{W\,m^{-2}}$. Sensible heat flux density of $-25\,\mathrm{W\,m^{-2}}$ is obtained as the residual term of eqn [3] (**Table 1**). Clearly, evapotranspiration is an important constituent of the surface energy balance that cools the Earth's surface by dissipating a large proportion of the absorbed radiative heat.

Table 1 shows that the net radiation of oceans $R_n = 118\,\mathrm{W\,m^{-2}}$ is more than that of the Earth. This is because the albedo of water is only 0.06, whereas the mean Earth-surface albedo is 0.15. Mean evaporation from oceans amounts to a latent heat flux density of $-94\,\mathrm{W\,m^{-2}}$. The theoretical equilibrium latent heat flux density into a vapor-saturated atmosphere is $-73\,\mathrm{W\,m^{-2}}$. The higher evaporation is an indication that the average water vapor content of the atmosphere above the oceans is below saturation, vapor being removed by condensation and advection to the continental areas. As latent heat release equivalent of the rainfall values over oceans is only $86\,\mathrm{W\,m^{-2}}$, some of the evaporation from the oceans condenses and precipitates on continents. Indeed, the hydrologic balance of continents shows that mean

Table 1 Partition of the global surface energy balance ($\mathrm{W\,m^{-2}}$)

	R_n	*LE*	H
Oceans	118	−94	−24
Land	64	−37	−27
Earth	102	−77	−25

R_n, the Earth's surface radiation balance; LE, latent heat for evapotranspiration; H, heat convection.

precipitation over land releases 56 W m^{-2} as latent heat into the atmosphere and exceeds the 37 W m^{-2} of mean latent heat flux density for evapotranspiration from land locked water. The balance is closed by runoff of water from the rivers flowing into the oceans. Continental evapotranspiration includes direct evaporation of water from the soil and transpiration of the soil water taken up by vegetation. Desiccation of soil surfaces forms a barrier that retards direct evaporation of water held in the deeper layers. Plant roots can extract water held in the soil below the surface and transport it to the leaves where evaporation occurs. For this reason, the presence and vitality of a vegetation cover have a major role on latent heat dissipation from land areas.

Eqn [3] also summarizes the energy balance of the atmosphere. Its net radiation R_n is the sum of the total solar radiation absorption, 67 W m^{-2} (40 + 27) and the net long-wave radiation loss −168 W m^{-2} amounting to −102 W m^{-2}. This flux density is balanced by the 77 W m^{-2} of latent heat released by the condensation of the water vapor produced by evapotranspiration and the 25 W m^{-2} of direct convective heating.

Anthropogenic Change of the Energy Balance

The previous sections show that the long-wave radiative effects of clouds, aerosols, and of gases elevate the Earth's surface temperature. The −66 Wm^{-2} long-wave radiation loss by the Earth surface are partly reabsorbed in the atmosphere, but 40 Wm^{-2} escape through a radiative window of the atmosphere to space. Gases with long-wave absorption bands such as CO_2, CH_4, and NO_X diminish the transparency of this radiative window. The anthropogenic release of these gases elevates their presence in the atmosphere and increases long-wave radiation absorption. Part of this absorbed radiation is remitted toward space, but part augments the long-wave radiative load of the ground surface. As 5.5 Wm^{-2} is required to raise the temperature by 1 K, the obstruction of the radiative window could increase the surface temperature by several degrees.

Air pollution is also reducing the transparency of the atmosphere for solar radiation on a global scale. A recent study shows that solar radiation reaching the Earth's surface has decreased by 20 Wm^{-2} over the past 30 years. The study does not show how much of the missing solar radiation is absorbed or reflected to outer space. Yet, the lowering of the primary energy input reaching surface should cool the Earth's surface or at least retard the heating trend due to long-wave absorbing gases.

Further Reading

Budyko MI (1958) *The Heat Balance of the Earth's Surface*. (translated by Stepanova NA). US Department of Commerce. Washington, DC: Government Publishing Office.

Budyko MI (1982) *The Earth's Climate: Past and Future*. London, UK: Academic Press.

Chahine MT (1992) The hydrological cycle and its influence on climate. *Nature* 359: 373–380.

Kiehl JT and Trenberth KE (1997) Earth's annual global mean energy budget. *Bulletin of the American Meteorological Society* 78: 197–208.

Paltridge GW and Platt CMR (1976) *Radiative Processes in Meteorology and Climatology*. Amsterdam, the Netherlands: Elsevier Scientific.

Peixoto JP and Oort AH (1992) *Physics of Climate*. New York: American Institute of Physics.

Rind D (2002) The Sun's role in climate variations. *Science* 296: 673–677.

Stanhill G and Cohen S (2001) Global dimming: a review of the evidence for a widespread and significant reduction in global radiation with discussion of its probable causes and possible agricultural consequences. *Agricultural and Forest Meteorology* 107: 255–278.

ENVIRONMENTAL MONITORING

P J Loveland and P H Bellamy, Cranfield University, Silsoe, UK

> The pains that come from the necessities of nature, are monitors to us to beware of greater mischiefs, which they are the forerunners of; and therefore they must not be wholly neglected, nor strain'd too far. (John Locke, *Some Thoughts Concerning Education* §107, 1693)

Introduction

Soil monitoring (SM) is often driven by the desire to understand soil quality and changes in it. This usually requires the establishment and population of soil-quality indicators (*See* **Quality of Soil**). This can, and does, have a profound influence on what SM is carried out, how a program is designed and executed, and what it costs. Effective SM implies long-term commitment and very tight control of potential sources of variability. Decisions taken at the planning stage are of extreme importance, and careful attention to them can save much money and trouble. Two essential points are:

- There is no universal system of soil monitoring; the means must fit the purpose and *vice versa*.
- There is a clear distinction between a soil-inventory program and a soil-monitoring program.

The first point might seem obvious, but is often overlooked. What is appropriate for an industrial site and its surroundings, where potential hazards to human health might be the driver, can be wholly inappropriate in an area where maintenance of a particular soil ecosystem is the target.

The second point is that an inventory is a collection of information about a place. For soils, it might be the features observed at the location of a soil profile and the properties of the profile horizons. 'Monitoring' implies temporal change, i.e., observation, sampling, and measurement repeated over time. There is thus a different intent and commitment in soil monitoring compared with an inventory, although the latter might be the starting point (baseline) of a soil-monitoring program. It is stressed that a single set of measurements in time does not constitute soil monitoring.

Most commonly, soil-monitoring targets change in a specific property or properties, e.g., pH, metal content, biomass. However, it can be equally important to study a process, e.g., soil erosion. Increasingly, there is a demand to understand potential soil functions. Linking soil monitoring to other environmental monitoring programs can maximize long-term benefits from the data collected. Planning SM needs to take these different objectives into account.

Finally, although many of the points discussed below might also apply to small parcels of land, this article is mostly concerned with the concept of soil monitoring applied at the regional or national scale.

Why Monitor Soils?

In recent years, there has been growing acceptance that not enough is known about soils and the changes that they are, or might be, undergoing, especially under increased pressure from man. This contrasts with the effort that has gone into the assessment of air and water. There are concerns that pressures on soils, usually driven by economic needs, will lead to irreversible decline in their ability to maintain ecosystem diversity, food, fiber and timber production and this will lead, indirectly, to serious problems with water and air quality. Thus, the whole framework of sustainable development would be compromised. There have been many statements over the last few decades pointing to the need to consider soils in terms of the:

- establishment of a set of principles for the rational use of soils and protection of them against irreversible degradation;
- pursuit of programs of soil conservation and reclamation and appropriate use of soils;
- recognition of soils as a fragile and essentially irreplaceable resource;
- importance of specific soil properties as the determiners of best use and land management practices;
- need for inventories and monitoring programs to establish baseline soil properties and changes in these over time.

Against this background, soil monitoring can be instigated for a number of reasons:

A. Basic science: are soil properties changing and, if so, at what rate and in what direction, and what does this tell us about our understanding of the world? For example, establishment of the magnitude of change in soil carbon content could improve inputs into models of the global carbon cycle;

B. Political: the need for information from which to formulate, test or improve policy, e.g., conventions on transboundary pollution require reductions in the acid load to soils; this could be demonstrated by showing that soil pH is remaining above a given value over a given time;

C. Legal: whether change has occurred to a soil that is in conflict with legislative obligations, e.g., demonstrate that loadings of so-called heavy metals are within statutory limits;

D. Financial: are the properties of a soil changing (or likely to change) over a given time-frame such that the soil is, or will become, more or less valuable in a given context, e.g., is there a potential pollution problem with otherwise very valuable development land?

Much of this thinking is increasingly considered in terms of the ability of soils to perform several functions, often simultaneously, i.e., food and fiber production, filtering, buffering, acting as a pool of genetic resource, support for diversity of above- and belowground ecosystems, keeper of the archeological record, support for the built environment. Thus SM is under pressure to provide information on the performance (or potential performance) of these functions and, often, to guide choices as to which function

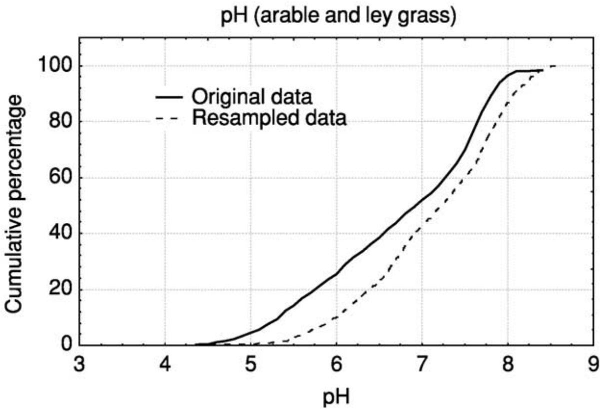


Figure 1 Change in topsoil pH between 1980 and 1995 under an arable/ley grassland rotation in England and Wales ($n=900$). (Loveland PJ (1990) The National Soil Inventory: survey design and sampling strategies. In: Lieth H and Markert B (eds) *Element Concentration Cadasters in Ecosystems*, pp. 73–80. Weinheim, Germany: VCH Verlagsgesellschaft.) (Data from the National Soil Inventory of England and Wales.)

should be 'allocated' to a parcel of land in preference to another, e.g., build here but not there. Thus, simple measurement of simple properties is often no longer the absolute *raison d'être* of soil monitoring, although it is often the baseline. Figure 1 shows that after 15 years the pH of topsoils (0–15 cm) has increased under arable/ley grass rotation in England and Wales, i.e., the soils are less acid. This finding has been interpreted as a mixture of a reduction in acid deposition (policy response), better targeting of the use of agricultural lime (improved agronomic practice), and deep plowing into less acid subsoil because of larger, more powerful farm machinery (economic driver). All these reasons can be viewed as useful outcomes of SM and could lead to more focused targeting of future environmental research.

The Starting Point

Whatever the driver for SM, some fundamental questions need addressing:

A. What information is already available about soil type(s), properties, scale of observations, etc.?
B. Is this information internally consistent, e.g., are, or were, all the data obtained in the same way, or in ways comparable enough for the purpose intended?
C. What information is not available and does this matter?
D. Is further information required at the same spatial scale as existing data, at a smaller scale (fewer points), or a larger scale (more points)?
E. Is information required simply as values of a property, or in ways that describe soil functions; if so, which function is of greatest interest?
F. What is the target, e.g., all soils in an area, the common soils, rare soils, a specific ecosystem, a specific soil process?
G. What magnitude of change is of interest, over what time scale, and is this a realistic target?
H. Is the property measurable or is further research needed?
I. Is it necessary to collect more information, or can the need be met from surrogate measurements, pedotransfer functions, extrapolation and interpolation?

Organizational and practical questions arise:

1. How can soil monitoring yield the greatest amount of information for the greatest number of people for the least cost?
2. At what time interval is the monitoring to take place?
3. Is this interval suitable for all soil properties of interest?
4. Can the needs of the users be satisfied with one soil-monitoring design, i.e., number and layout of sites; if not, what modifications are needed?
5. Will one organization do everything, or is it to be a consortium (if so, who leads?)
6. Does the lead organization have the ability, infrastructure and will to keep that role for a sufficient period? It is no use setting up a 10-year program if those involved drop out after 5 years;
7. Are there agreed protocols for the design of the monitoring network, for the location, layout, and description of sites, for the soil sampling, for the preparation and storage of samples, for laboratory work, for quality control; if not, who will orchestrate this?
8. Who will keep the samples, the raw data, the processed data, and be responsible for them into the future?
9. Who 'owns' the data, i.e., are there issues of copyright, intellectual property rights, confidentiality, and controls on the release of data?
10. Is there an agreed publication policy for the output from the soil monitoring, given that many years might pass before sufficient data are available to allow robust statements to be made about change?
11. What will the various options cost, bearing in mind that the costs of very long term site maintenance, quality control, and storage can be substantial.

Many of these points might seem trivial. However, the time scale over which SM must run before large amounts of meaningful data are acquired has to be seen in decades. Thus the objectives, responsibilities, targets, and costs need to be thought through with great care and appropriate commitments obtained.

Soil-Monitoring Networks

Many countries have conventional soil survey maps showing the spatial distribution of soil types. Although the scale of the map controls the smallest area of any soil type that can be shown, even small-scale maps, e.g., 1:250 000 scale, can indicate the relative abundance of soil types. At a very early stage, it has to be decided whether the network is to be capable of sampling rare soils, i.e., those occupying only a small proportion of the landscape. As a guide, a network that is expected to 'capture' a soil occupying 5% of the landscape with 95% confidence would need several thousand sampling locations (Table 1). (*See* **Spatial Variation, Soil Properties.**) An example is the National Soil Inventory (NSI) for England and Wales (UK; approx. 150 000 km^2), which sampled almost 5700 points, but only captured about 400 of the 700 soil series that are known to occur in the two countries. If no soil maps exist, then the problem becomes even more pressing.

Numerous designs have been proposed for national or regional soil-monitoring networks. The NSI network is based on sampling at the intersects of a 5 × 5-km-square grid. The Forest Soil Monitoring Network in Europe is based on a regular 16 × 16-km grid, and there are many other examples of the use of square grids. The most efficient regular geometric grid is triangular, but the square grid has the advantage that it can be co-located easily with a national geographic grid, thus making it easy to find points on the ground using conventional maps. This proviso may become less important with the increasing use of global positioning systems (GPS) (but these currently do not work equally well everywhere). One caution often expressed about square grids is that the periodicity of the grid might match that of some periodicity of the soil itself. However, whilst this might be true at the field scale, e.g., rows of crops in orchards and olive groves, 'tramlines' in arable crops, etc., no evidence has yet been presented that such periodicity exists for very large areas of land. On a practical note, if a square grid is used, it is best to locate its origin such that the sampling points do not fall on the edges or corners of map sheets; a 1-km offset is usually adequate. Alternative networks can be designed based on a simple random selection of points, stratified random selections, and a stratified, systematic unaligned selection of points (*See* **Spatial Variation, Soil Properties**). An alternative approach is weighted or purposive sampling, i.e., the sampling points are located on the basis of, for example, an existing network of sites such as experimental farms, sites for other forms of monitoring, selected soil–landscape combinations or, possibly, perceived soil functions. The principal problems with this approach are:

- It might not yield enough points to allow the spatial structure of the data to be estimated reliably by geostatistical methods – too much of the landscape is omitted;
- There is a high risk of introducing bias into the design.

However, such schemes can yield reliable baseline data for particular soil properties, e.g., lead content, so long as the limitations are appreciated. Well-conducted monitoring networks of this kind are being run, for example, in Switzerland, Austria, and the Netherlands among others (see also the Quality Control section, below).

Table 1 Size of sample required (*n*) to estimate property (*P*) within limits $P \pm 0.1P$ for three levels of confidence, calculated from the normal approximation

		n		
P *or 1*−P	*Confidence limits*	*80%*	*90%*	*95%*
0.5	±0.05	164	269	384
0.4	±0.04	245	331	576
0.3	±0.03	382	627	896
0.2	±0.02	655	1075	1536
0.1	±0.01	1474	2420	3456
0.05	±0.005	3112	5109	7296

Adapted from Webster R and Oliver MA (1990) *Statistical Methods in Soil and Land Resource Survey*. New York: Oxford University Press, with permission.

Protocols

Whatever is done and by whom, it is absolutely essential to have protocols in place for all aspects of the work: site selection, accuracy of location, sampling strategy at each site, sampling, sample treatment, analytical procedures, data storage and handling, and data access. Uncontrolled sources of potential variability are the greatest enemy of long-term monitoring. Peoples' memories are not reliable over long timescales, nor can it be relied on that those people will be available some years later to answer any questions. Information must be written down, multiple copies kept, photographs taken, and proper recording and training undertaken. Above all, it is vital that information is shared between investigators.

Monitoring Sites

Methods for establishing all of these types of networks are described in detail in the literature.

Whatever is chosen, it is extremely important to avoid bias in the location of sites, i.e., the person on the ground should not be allowed to pick and choose where to locate a site and how to sample. It is also strongly recommended that the network should have sufficient points to be robust, i.e., to buffer against loss or major change at some of the sites. Whilst it is obvious that the larger the number of sites (*n*), the more one is likely to capture the range of values of a variable, it is often forgotten that *n* should be sufficiently large to allow adequate resampling to establish a given level of change. For example, in England and Wales, the 0- to 15-cm layer of *c*. 2500 arable/ley grassland sites (this counts as a single land use in the UK) was sampled in 1980, and these samples had a mean organic carbon content of *c*. 3.2%. In 1995, these sites were resampled in sufficient numbers to be able to detect a change of 0.2% (absolute) in this mean at 95% (this level of change represents just under 4 t organic C ha^{-1}). This was achieved by sampling *c*. 900 of the original 2500 sites at random (Table 2). It is worth using existing data, or reasonable synthetic data, however limited, to test out some of these sampling strategies at different degrees of detection of change as part of the exercise in determining how many sampling points might be required.

Table 2 Size of sample required to detect different levels of change in mean organic carbon content in arable/ley grassland sites (at 95% confidence level)

Change in % mean organic carbon	*Sample size*
1.0	48
0.5	180
0.2	820
0.1	1668

Adapted from National Soil Inventory of England and Wales.

Table 3 shows the effect on the data collected for zinc during the NSI program if smaller numbers of sites are visited at regular intervals across the original square grid or by selecting similar numbers of points randomly from that grid. One obvious result of taking fewer samples is that some of the extreme (high or low) values are missed, so the observed range becomes smaller. Hence the spatial variability of values in the region may appear less than it is.

Although it might be possible to control activities at a small number of key sites, it is uncommon to be able to tie land owners or managers to particular land-use agreements for decades. Thus, sites can be built upon, become dumps for waste, become polluted and thus unacceptably dangerous to work on, or the land use might change so drastically that the site is no longer fit for monitoring purposes. The latter is always difficult to assess. Some would argue that if a soil is converted from arable to woodland monitoring the progress of that change is a matter of considerable interest. Others might argue differently because they want to know how cultivated land behaves in the long term. Whatever is decided, the situation whereby the loss of a relatively small number of sites compromises the whole network needs to be avoided. Again, to use the example from England and Wales (UK), approximately 3% of the *c*. 900 sites under arable/ley grassland in 1980 had been lost to agriculture by 1995. This change was interesting in itself, but the 'loss' of these sites did not seriously alter the conclusions to be drawn from the data for the remaining sites (approximately 870).

The simplest site is a sampling point, a spot on the ground located to a given accuracy. Points need to be located geographically according to a protocol that fits the purpose of the network. Experience has shown that people with some background of field work, using good topographic maps

Table 3 The effect of reducing the sample size on parameters for total zinc (mg kg^{-1})

No. of samples	*Mean*	*Median*	*Minimum*	*Maximum*	*Lower quartile*	*Upper quartile*	*Quartile range*	*SD*
On regular 5 × 5-km grid								
5676	96.78	82.0	5.0	3648.0	59.0	108.0	49.0	108.85
1434	94.95	83.0	7.0	1985.0	59.0	108.0	49.0	89.40
363	92.66	81.0	8.0	765.0	57.0	108.0	51.0	69.85
94	95.56	83.5	14.0	434.0	61.0	110.0	49.0	65.10
15	74.94	72.0	28.0	126.0	52.0	95.0	43.0	26.05
Selected randomly								
5676	96.78	82.0	5.0	3648.0	59.0	108.0	49.0	108.85
1438	92.78	81.0	6.0	2017.0	58.0	107.0	49.0	82.50
367	90.30	81.0	10.0	753.0	56.0	103.0	47.0	66.39
97	95.26	85.0	10.0	753.0	54.0	110.0	56.0	85.26
17	105.76	87.0	37.0	270.0	67.0	118.0	51.0	62.77

Adapted from National Soil Inventory of England and Wales.

(1:10 000 or 1:25 000 scale) and/or aerial photographs, can easily relocate a point to within 10 m of a target point. With GPS, relocation to within 2–3 m is possible, although nearby buildings, trees, and hilly ground can seriously distort signals. It can be helpful to mark these points to aid future relocation. A vandal-proof approach is the burial of either a metal plate (a sheet of aluminum alloy approximately 25 cm square and 5 mm thick) or of a small but powerful permanent magnet (wrapped in a plastic bag) at approximately 50 cm depth, i.e., well below plow depth. These can easily be found again after many years by appropriate hand-held, electromagnetic detectors.

There are many alternative arrangements of kinds of sites. Two of the most common are:

- Area-based sites, in which it is decided that it is insufficient for the purposes of the monitoring that a given area of land shall be sampled in a consistent manner. This approach is quite common in what might be termed 'agricultural monitoring,' where the argument is that fields or other parcels of land are subject to such intense management that the whole parcel becomes the site;
- Key sites or control sites that are specifically managed so as to be available for revisiting in the long term. An area of relatively uniform soil is identified (often of a hectare or more) and sampling plots are laid out across it in sufficient number to allow resampling according to a strict protocol that avoids resampling of previous plots for many years. These kinds of sites can be extended to allow investigation of large-scale processes such as soil erosion or the effects of afforestation, etc., but it is axiomatic that large sites, with their (partial) dedication to the long-term study of specific soil processes are expensive to run. In most countries, therefore, that have followed this thinking, the number of such sites is small (10–15 is not uncommon).

There is a growing belief that a small number of such key sites should be a component of any soil-monitoring network as they can form a framework by which the quality control of a much larger network can be achieved.

Soil Sampling

There are two key questions to be asked when considering soil sampling:

- Which soil layers are to be sampled, e.g., soil horizons (the people doing the sampling need to be able to recognize these) or fixed-depth sampling, and are litter layers to be included?
- Is the sample to be taken at a point, e.g., from the face of a soil pit, or is it to be a bulked sample taken using the geographic point as a centroid?

In practical terms, often a combination of these approaches is used. Surface layers are readily sampled around a centroid, e.g., 25 subsamples at 4-m intervals on a 20 × 20-m grid, a 'random' walk across a parcel of stated extent, subsampling along the radii of a circle, or a numbered layout of sufficient sampling plots to allow a considerable number of resamplings without the risk of revisiting disturbed soil. There is much to be said for this 'bulked' approach as at least a partial answer to dealing with the short-range variability of soils. Others argue for a formal, geostatistical approach to each site, so that its variability is known and an appropriate sampling scheme drawn up. This can, however, have large cost implications if it is to be applied to hundreds of sites. Whatever is done should not compromise future sampling – the temptation to remove the whole of the topsoil layer at the centroid should be resisted; similarly, soil pits should be small.

Finally, the sample must be big enough, because there are usually several people who want some of it. At least 2 kg on an air-dried, less than 2 mm basis is a useful target. This has implications for the volume of organic (peat) soils sampled. Sample storage needs considerable prior discussion with all who might have an interest, e.g., the chemists might be happy with dried material, but what do the biologists want? Long-term 'fresh' sample storage implies access to specialized facilities – Do they exist? Can they handle the number of samples (this year, next year, every *X* years, etc.)? Storage of frozen samples can pose similar problems. Air-dried material can be stored in screw-cap glass bottles with Teflon lid liners (but these are not cheap). The amount of shelf space required for the next 10, 20, 50, etc. years needs to be estimated, as well as the cumulative weight of the material. An intelligible sample-identification system is required so that samples can be retrieved, and a mechanism is necessary that ensures no one arbitrarily moves the samples or throws them away.

If the physical properties of soils are of interest, e.g., bulk density, water-retention characteristics, and hydraulic conductivity, then other sampling techniques are required (e.g., undisturbed cores or samples of known volume) and are best discussed with the appropriate specialists. It might be necessary to designate a separate area of the site for this kind of work, as considerable disturbance can be involved. Above all, it is necessary to allow for sufficient sampling space to accommodate all likely measurements without compromising any of the others.

Quality Control

This is not just a laboratory matter, but applies equally well to field operations. Whichever is under consideration, protocols, testing, and training are paramount; slipshod procedures at any stage can undermine the whole enterprise.

In the field, there is much to be said for having a formal recording system – the site location, local land characteristics, aspect, slope, etc. being recorded in standard ways, and (permitted) deviations from these should be recorded also. For example, it is common to find that the specific geographic point is not accessible (e.g., because it is covered by a road or building). Rather than abandon the site, limited, structured deviation is permitted, e.g., move 100 m north, then east, then south, etc., the details of the deviation being recorded. Such information can be recorded on proformas on portable computers or on printed proformas (synthetic papers are now widely available that will withstand any kind of weather).

Long-term quality control requires that a proportion of the samples are analyzed in duplicate or triplicate during the current program and during every subsequent resampling. This is to detect long-term drift. The proportion is a matter for discussion, but, as the monitoring program continues, the amount of ‘repeat’ analysis becomes the largest component of the laboratory work, with implications for costs. If samples are to be sent to distant laboratories, a system has to be established for this. For example: Are field refrigerators or ‘cool boxes’ required? How quickly should the samples be sent (overnight for biological monitoring)? Does the laboratory know what to do when the samples arrive? Protocols have to be set up for all these matters and tested before the monitoring commences, e.g., can the protocol deal with samples arriving at the start of a public holiday?

Ideally, all the determinations should be made in one laboratory but, even where this is not possible, thought should be given to dividing the work into blocks, e.g., all the metals are determined in one laboratory, all the organics in another, all the biology at a third. Are the chosen laboratories up to standard and are they familiar with soils? Do they have formal QA procedures in place and are they accredited and audited? Do they use widely recognized standard methods (International (ISO) standards should be used wherever possible)? Are they viable in the long term? Cheapest is not necessarily best.

Similar approaches are necessary with regard to output from the soil-monitoring program. The data need to be held securely in a system that has a good prospect of long-term stability in terms of software and staffing. Data ownership, intellectual property, and related issues need to be clarified at an early stage, as does a policy over publication of results.

Summary

Soil-monitoring programs need to be fit for purpose. At the very start, a clear answer to the question ‘Why are we doing this?’ is needed. Considerable planning is required to achieve success. A large number of issues need to be considered and decided at an early stage and written into robust protocols. Above all, soil monitoring has considerable spatial, temporal, and cost implications, and there has to be a commitment at the decadal scale.

See also: **Acid Rain and Soil Acidification**; **Aggregation:** Physical Aspects; **Carbon Cycle in Soils:** Dynamics and Management; **pH**; **Quality of Soil**; **Spatial Variation, Soil Properties**

Further Reading

Acton DF and Gregorich LJ (eds) (1995) *The Health of Our Soils: Towards Sustainable Agriculture in Canada*. Ottawa, Canada: Agriculture and Agri-food.

Bak J, Jensen J, Larsen MM, Pritzl G, and Scott-Fordsmand J (1997) A heavy metal monitoring programme in Denmark. *The Science of the Total Environment* 207: 179–186.

Blum WEH, Brandstetter A, Riedler C, and Wenzel WW (1996) *Bodendauerbeobachtung: Empfehlung für eine einheitliche Vorgangsweise in Österreich*. Wien, Austria: Bundesministerium für Umwelt, Jugend und Familie.

Council of Europe (1992) *Soil Protection*. Recommendation to the Council of Ministers R(92)8, 18 May 1992. Strasbourg, France: Council of Europe Publications.

Davidson DA (2000) Soil quality assessment: recent advances and controversies. *Progress in Environmental Science* 2: 342–350.

Dumanski J (ed.) (2000) Indicators of land quality and sustainable land management. *Agriculture, Ecosystems and Environment* 81 (Special Issue).

Eriksson J, Andersson A, and Andersson R (1997) *Tillståndet i svensk åkermark (Current Status of Swedish Arable Soils)*. Rapport 4778. Stockholm, Sweden: Naturvårdsverket. (In Swedish with extensive English Summary.)

European Commission (2002) *Soil Protection for Sustainable Development: The Soil Protection Communication*. Environment Directorate General. Brussels, Belgium: European Commission Publications.

FAO (1994) *Harmonisation of Soil Conservation Monitoring Systems*. Proceedings of an International Workshop, 14–17 September 1993. Research Institute for Soil Science and Agrochemistry. Budapest, Hungary. Geneva, Switzerland: FAO.

Federal Republic of Germany (1998) *Bundesbodenschutzgesetz*. (Federal Soil Protection Act), 17 March 1998. Bonn, Germany: (Available in English translation.)

Huber S, Freudenschuss A, and Stärk U (eds) (2001) *European Soil Monitoring and Assessment Framework*. Technical Report No. 67. Copenhagen, Denmark: European Environment Agency.

Lorenz M, Becher G, Fischer R, and Seidling (2000) *Forest Condition in Europe – 2000: Technical Report*. Geneva, Switzerland: UNECE.

Loveland PJ (1990) The National Soil Inventory: survey design and sampling strategies. In: Lieth H and Markert B (eds) *Element Concentration Cadasters in Ecosystems*, pp. 73–80. Weinheim, Germany: VCH Verlagsgesellschaft.

McGrath SP and Loveland PJ (1992) *The Soil Geochemical Atlas of England and Wales*. Glasgow, Scotland: Blackie Academic.

Office for Economic Co-operation and Development (1998) *Agriculture and the Environment: Issues and Policies*. Paris, France: OECD Publications.

Schipper LA and Sparling GP (2000) Performance of soil condition indicators across taxonomic groups and land. *Soil Science Society of America Journal* 64: 300–311.

Schulin R, Desaules A, Webster R, and von Steiger B (eds) (1993) *Soil Monitoring: Early Detection of Soil Contamination and Degradation*. Basel, Switzerland: Birkhäuser-Verlag.

Sykes JM and Lane AJM (eds) (1996) *The United Kingdom Environmental Change Network: Protocols for Standard Measurements at Terrestrial Sites*. London, UK: The Stationery Office.

Wagner G, Quevauviller P, Desaules A, Muntau H, and Theocharopoulos S (2001) Comparative evaluation of European methods for sampling and sample preparation of soils. *The Science of the Total Environment* 264(1–2): 3–204.

Webster R and Oliver MA (1990) *Statistical Methods in Soil and Land Resource Survey*. New York: Oxford University Press.

ENZYMES IN SOILS

R P Dick, Ohio State University, Columbus, OH, USA
E Kandeler, University of Hohenheim, Stuttgart, Germany

Introduction

Biological and biochemically mediated processes in soils are fundamental to terrestrial ecosystem function. Ultimately, members of all trophic levels in ecosystems are dependent on the soil as a source of nutrients and energy, and for degradation and cycling of complex organic compounds. Soil enzymes are central to these processes by catalyzing innumerable reactions in soils that have global biogeochemical significance.

Enzymes are proteins that act as catalysts by accelerating rates of reaction without undergoing permanent change. Enzymes are specific activators because they combine with their substrates in stereospecific fashion that decreases the stability of certain susceptible bonds (i.e., changes electronic configuration), which reduces the energy of activation of reactions (the amount of energy required for a reaction to proceed).

In soils, enzymes can exist intracellularly (inside the cytoplasmic membrane), which is of course important in cellular life processes. In addition, enzymes can exist outside the cytoplasmic membrane, in the periplasmic space or cell surface, and as extracellullar enzymes in soil solution or stabilized in the soil matrix. The latter two categories are known as the abiontic form. Both intracellular and extracellular enzymes may be involved in biogeochemical processes, as discussed below, but it is difficult to quantify the contribution of each group of enzymes in performing a given reaction.

It is generally assumed that soil enzymes are largely of microbial origin, but it is also possible that animals and plants may contribute enzymes to soils. However, studies on the fate of root enzymes have shown that they are rapidly degraded and/or denatured, suggesting they do not contribute much to the overall enzyme activity of soils. It is difficult to conclusively discriminate between sources of enzymes in soils, and thus evidence for the primary role of microbes as a source of soil enzymes is from indirect evidence such as sterilization of soils using radiation, which kills the organisms but leaves the enzymes catalytic.

Soil fauna are probably a limited source of enzyme activity in soils but earthworm (*Lumbricus terrestris*) casts stimulate enzyme activities on localized basis. There is increased enzymatic activity in rhizosphere soil compared with nonrhizosphere soil for many enzymes because roots can excrete extracellular enzymes and there is increased microbial activity due to root exudates and sloughing of cellular debris. This elevated activity in the rhizosphere is probably due to roots stimulating microbial activity than from enzymes exuded by roots.

The activity of approximately 100 enzymes has been identified in soils. Undoubtedly the number in soils is far greater, but techniques for determining the presence or activity of other enzymes have not yet been developed for soils. The soil enzymes most often studied are oxidoreductases, transferases, and hydrolases. The oxidoreductase dehydrogenase has been widely studied in soils partly because of its apparent role in the oxidation of organic matter, where it transfers hydrogen from substrates to acceptors. Catalase activity is based on the rate of release of oxygen from added hydrogen peroxide or the amount of recovered hydrogen peroxide. Some hydrolases and transferases have been extensively studied because of their role in decomposition of various organic compounds and thus are important in nutrient cycling and formation of soil organic matter. These include enzymes involved in: the C cycle, i.e., amylase, cellulase, xylanase, glucosidase, and invertase; the N cycle, i.e., protease, amidase, urease, and deaminase; the P cycle, i.e., phosphatase; and the S cycle, i.e., arylsulfatase. Lyase activity has been found in soils but relatively few studies have been conducted on this group of enzymes.

The functional role of abiontic soil enzymes has yet to be conclusively shown experimentally. First there has been limited success in extracting enzymes from soils. This is because the enzymes complexed in the soil matrix often lose their integrity during the extraction process. As a result soil enzymes have largely been studied by measuring activities, which makes separation of abiontic enzymatic activities indistinguishable from activities associated with the living organisms.

Nonetheless, it has been hypothesized that abiotic enzymes may have important relationships with soil organisms. For some microorganisms, substrates or their breakdown products may be useful for the organism but, due to substrate size or insolubility, unavailable for direct microbial uptake. It may be advantageous for a microbial cell to be located on the surface of a humic colloid containing a number of enzyme molecules. Indeed, for some species their successful survival in a hostile soil environment may depend on an association with humus–enzyme complexes.

Spatial Distribution of Enzymes in Soils

Extracellular Enzyme Stabilization in Soil Matrix

Enzymes have been categorized according to their location in the soil. Three enzyme categories (termed 'biotic enzymes') are associated with viable proliferating cells located: (1) intracellularly in cell cytoplasm; (2) in the periplasmic space; and (3) at the outer cell surfaces. The remaining categories are broadly characterized as abiontic, from the Greek words '*a*,' meaning 'removal or absence of a quality,' and '*bios*,' 'having a form of life.' Abiontic enzymes are those exclusive of live cells that include enzymes: excreted by living cells during cell growth and division; attached to cell debris and dead cells; and leaked into soil solution from extant cells or lysed cells, but whose original functional location was on or within the cell. Additionally, abiontic enzymes can exist as stabilized enzymes in two locations: adsorbed to internal or external clay surfaces; and complexed with humic colloids through adsorption, entrapment, or copolymerization during humic matter genesis.

Microscale Distribution

The functionality and activity of enzymes in soils are closely controlled by their location with enzyme–substrate interactions occurring at pico- and nanoscales. It is difficult to study extracellular enzymes directly, as the recovery of an enzyme is generally less than 20% of the total present in soil. In addition, humic–enzyme complexes may be modified during the extraction procedure. Microscale investigations have concentrated on different size aggregates after physical separation or by studying microhabitats of the high-turnover, particulate organic matter (mainly undecomposed plant material or microbial debris).

Generally, soil organic matter chemistry ranges from minimally decomposed and course plant debris with a high C/N ratio to highly processed humic substances as organomineral complexes with a narrow C/N ratio associated with clay fractions. These spatial and physical size distribution patterns of C sources control the location of microbial community members and associated enzymes. This variability can be related to microhabitat hot spots of microbial activity on organic material associated with the rhizospheres (roots), drilospheres, and detritus (decaying organic matter). Characterization of rhizosphere soil with a root mat–soil interface technique has shown that the abundance of microorganisms and their activities decreases to levels similar to bulk soil within a distance of several millimeters from root surfaces. This spatial relationship is closely related to the levels in the rhizosphere of easily degradable root exudates, and mass flow and diffusion distance of dissolved organic substrates used by soil microorganisms. The soil–litter interface distance of influence is similar to root surfaces with enzyme activities, being greater up to 1.3 mm from the litter interface.

The spatial distribution of various microbial groups or species, besides substrate availability, is related to protection from predation from larger organisms. Commonly, much of the soil microbial biomass and enzyme activities are associated with the smaller-sized fractions (fine silt and clay). Studies have found that the clay fraction is dominated by bacterial biomass, with invertase, phosphatase, and protease being greater in the fine-textured fraction, but fungi and xylanase activity dominating in the coarse sand fraction that contains particulate organic matter, suggesting xylanase is of fungal origin. Enzyme activities of the coarse-sized particle fractions are strongly influenced by tillage practices; the enzyme activities of these particle fractions are therefore a valuable indicator for organic matter input and management changes.

Macroscale Distribution

Enzyme activity declines with soil depth and generally is correlated with organic C and microbial biomass distribution in the soil profile. This has been shown in many studies and is to be expected, as most of the biological activity and organic matter is in the surface soil horizons.

Plot-scale investigations have determined that organic matter turnover rates, microbial biomass, and enzyme activities of soil samples vary with vegetation and soil type. Relatively little is known about the topographic, pedogenic, soil mineralogy, and other properties that control microbial and enzyme distribution at landscape levels. The characterization of these interactions is essential to achieve a better understanding of complex ecosystem processes. This latter scale is of particular importance for developing a soil-quality indicator, because enzyme activities can vary more as a function of soil type than the differences caused by soil management.

Fuzzy operations (which use logic inference procedures by allowing individuals, or soil properties in this case, to be assigned into continuous class membership values instead of exact hard classes in order to make interpretations based on several or many soil properties) and multivariate analysis (factor analysis of many soil properties simultaneously; in this case to see which ones are most important in explaining soil variation in space or between management systems) have shown that land use is the strongest factor and soil contamination the weakest factor governing the level of soil enzyme activities at the ecosystem level in Central Europe. Soil type is an important site factor, as it summarizes climatic, topographic, and geologic conditions, acidification, and vegetation influence on soil biology and enzyme activity.

Methods of Studying Enzyme Activities in Soils

The discussions above indicate that, for most soil enzymes studied, enzymes exist as extracellular or abiontic enzymes that can retain their activities outside living cells and in viable soil organisms. The proportion of total activity coming from extracellular activity for a given enzyme is generally unknown and probably varies from enzyme to enzyme. This offers two ways to measure enzymes in soils; direct extraction of pure enzyme protein or measurement of enzyme activity.

Direct extraction and purification of enzymes in soils has been largely unsuccessful. This is evident when total enzyme activity of the soil is used to estimate how much enzyme protein is in the soil, and the amount extracted is always much lower than the estimate based on activity. The protein cannot be separated easily from mineral or organic colloids or it is denatured during the extraction process. This difficulty with enzyme extraction from soils means that soil enzymes are studied by measuring enzyme activity. A large number of assays have been developed to measure enzyme activities. For the assays to be effective, it is imperative that the reaction can be stopped completely after a specified incubation period and that the product is quantitatively extracted from the soil.

Because measuring activity is the primary means of studying enzymes in soils, there are a number of implications and limitations for understanding the spatial distribution, biochemistry, microbial ecology, and other factors relative to soil enzymes. The first consideration is the assay itself. Enzyme assays are done at substrate concentrations exceeding enzyme saturation on a known amount of soil and under a strict set of conditions that include temperature, buffer pH, and ionic strength. Thus, the results are operationally defined and any change in these conditions will change the measured activity. This means that enzyme assays measure the potential activity under optimal conditions and not the *in situ* activity, because the assay provides an environment quite different from the original soil.

A second implication of soil enzyme assays is that enzymatic activity of the abiontic or extracellular enzymes cannot be separated from that of living cells. Enzyme assays, particularly those with long incubations, generally include antiseptics such as toluene to inhibit growth and metabolism during the assay (shorter assays, 1–2 h, may not need growth inhibitors because the time is too short for significant microbial growth). As a result enzyme activity is not necessarily related to microbial activity (although many assays can be highly correlated with microbial biomass or respiration, but this relationship for the same enzyme can vary widely due to soil type, soil management, chemical and physical manipulations of soils, and climatic factors).

A majority of enzymes assays developed for soils are hydrolases, because there are a wide number of hydrolases in soils that are important in soil functions. The activity of most hydrolases is investigated using artificial substrates that do not occur naturally in soil. One of the most common types are simple esters that combine the functional group of the substrate, e.g., phosphate (for phosphatase) or glucose (for β-glucosidase) with a chromophore, such as *p*-nitrophenol (PNP), which is easily extractable from soils (except very high organic matter soils where there may be some adsorption of PNP). After a specified incubation period, a $CaCl_2$-alkaline solution is added, which stops the reaction, enables quantitative extraction of PNP, and causes PNP to turn yellow for colorimetric determination of the PNP product.

Other hydrolytic enzymes that work on major polymeric substrates include cellulases, chitinases, and lipases. Important enzymes in the N cycle are those that hydrolyze proteins, and peptides, and release NH_4 from amino acids. These enzymes are largely extracellular, because the substrates have a large molecular weight and/or are insoluble, making it difficult or impossible for their direct uptake across microbial membranes.

Oxidoreductase enzymes that have been detected in soils are dehydrogenase (which is actually a number of different intracellular enzymes that can perform the same reaction), glucose oxidase, catalase (peroxidase), and polyphenol oxidases. These latter two enzymes are important in lignin degradation. Dehydrogenase activity has been of particular interest because it is a component of the electron transport system of oxygen metabolism and requires intracellular environment of viable cells to express its activity. Therefore, it is not likely to exist in an extracellular form and thus should be a good indicator of physiologically active microorganisms, unlike nearly all other enzymes. However, it has turned out to be a poor indicator of microbial activity. This may be due to unsuitable assay conditions or the presence of extracellular phenol oxidases in the soil that cause the same reaction, or there can be common soil constituents that act as electron acceptors (e.g., nitrate or humic acids), all of which cause an overestimation of dehydrogenase activity (which is one reason why this assay may not correlate with microbial activity). Furthermore, elevated levels of Cu can interfere with the assay procedure. Other oxidoreductases of note are the polyphenol oxidases because of their role in humification of soil organic matter.

Other assays that have been developed include lysases (e.g., glutamate decarboxylase, which hydrolyzes apartic acid, and L-histidine ammonia and tyrosine decarboxylase, which are both involved in N mineralization) and transferases (e.g., dextransucrase, which hydrolyzes sucrose, releasing glucose and fructose and thiosulfate S-transferase, which oxidizes elemental S). Fluorescein diacetate hydrolysis is a broad-spectrum hydrolytic enzyme assay; this reaction can be carried out by proteases, lipases, and esterases.

Many of the enzyme assays utilize spectrophotometric analysis of reaction products. These methods are advantageous because they are, for the most part, rapid and accurate, and they have low equipment costs. Titrimetric methods are commonly used for N-cycling enzymes, where the hydrolysis product from organic N compounds is NH_4. Other methods include fluorimetric techniques, which can be very sensitive, radioisotopic detection of products (rarely used), manometric methods, typically used for oxidases and decarboxylation assays, ion-specific electrodes, and various chromatographic methods (e.g., high-performance liquid and gas chromatography), available for measuring a wide array of enzyme products.

Enzyme Kinetics

Enzyme reaction kinetics can be described by various equations. Enzyme-catalyzed reaction velocity is generally described by a rectangular hyperbola where enzyme concentration is held constant and substrate concentration is varied over a wide range. At low substrate concentrations, this reaction is a first-order reaction, which shifts to a zero-order reaction at high concentrations. The classic mathematical derivation for the rate equation was first developed by Michaelis and Menten in 1913 as follows:

$$S + E \underset{k_2}{\overset{k}{\rightleftharpoons}} ES \overset{k_3}{\rightleftharpoons} E + P$$

when S is substrate, E is enzyme, ES is enzyme–substrate complex and P is product. The reaction rate (v) at any moment is equal to k_3 and when substrate concentration reaches infinity the rate of reaction approaches a maximum ($V_{\max}$). For this to hold there are a number of assumptions that include: the substrate concentration remains nearly constant during the time of measurement; and if substrate concentration is much greater than enzyme concentration, only a negligible amount of substrate can accumulate in the ES intermediate complex. The K_m value is the Michaelis–Menten constant where substrate concentration is at one-half of the maximum velocity (moles per liter). The Michaelis–Menten constants (V_{max} and K_m) are specific properties of individual enzymes.

Traditionally the Michaelis–Menten constants were estimated by doing various linear transformations. The constants can then be determined by the slopes and intercepts based on these transformations of the data. With the advent of widely available, high-speed computers, it is now preferable to use nonlinear fitting of model equations to estimate K_m and V_{max}, because these have the least-biased fit of the data.

The Michaelis–Menten constants are used to characterize a particular enzyme. The V_{max} can be used to determine the concentration of the enzyme if the molecular weight is known; however, because it is difficult to extract enzymes from soils, the molecular weight is not normally known. Therefore, V_{max} can be considered as an index of the amount of enzyme in the soil but its unit of measurement is activity.

The K_m constant is an index of the affinity of the enzyme for the substrate (the smaller the K_m the greater the affinity). In soils this is referred to as the 'apparent' K_m, because there are a range of biological, chemical, and physical factors that can affect enzyme affinity. First, in all likelihood a soil has many isoenzymes that can act on a particular substrate, because most enzymes come from a variety of soil organisms or even from plant roots. Secondly, some of the activity for most enzymes originates from a component stabilized on mineral surfaces or complexed in soil organic matter, and there are cofactors in soils or other environmental conditions that can affect the reactivity of the enzyme. These latter physical and chemical factors may change the conformation of the enzyme or some other enzymatic property, which will affect affinity.

Michaelis–Menten constants can be used to provide insights into how soil management affects enzymatic properties. Long-term cultivation of soils causes the V_{max} to decrease and K_m to increase, the latter indicating that there was less affinity of the enzyme for the substrate. Differences in K_m values between these systems suggest that there is a different suite of isoenzymes and/or differences in abiontic forms of the enzymes in soils.

Soil-Quality Indicators and Technologies

Sensitivity to Soil Management and Ecosystem Stress

Soil is the final arbitrator of the impact of human activity on the ecosystem's health and function. Consequently, there has been interest in developing indicators of soil quality to reflect the impacts of pollution, and agricultural and forestry activity on the ability of soils to perform important biogeochemical ecosystem processes.

The rationale for soil enzyme activity as a soil-quality indicator is that enzyme activities:

1. Are often closely related to important soil-quality parameters such as organic matter, soil physical properties, and microbial activity or biomass;
2. Can begin to change much sooner (1–2 years) than other properties (e.g., soil organic C), thus providing an early indication of the trajectory of soil quality with changes in soil management;
3. Can be an integrative soil biological index of past soil management;
4. Utilize procedures that are relatively simple compared with other important soil-quality properties (e.g., physical and some biological measurements) and therefore have the potential to be done routinely by soil-testing or environmental-analysis laboratories.

A conceptual model for enzyme assays as integrative biological indexes is based on the observation that activity originates both from viable cells and abiontic components. The abiontic component provides an indicator of semipermanent changes in soil quality, because such enzymes are probably complexed and protected in the soil humic- or clay-complexes and change over a period of years. The interpretation of this characteristic from a soil quality perspective is that soil management that promotes stabilization of organic matter and associated structural properties (e.g., aggregation and porosity) would also promote stabilization of enzymes in the soil matrix.

Field-scale studies have shown that enzymes are diffenentially affected on a temporal basis due to soil-management activities. Table 1 summarizes activities of enzymes involved in carbon-, nitrogen-, phosphorus- and sulfur-cycling. These studies have shown that soil enzyme activities vary seasonally and over the long term, which has improved the understanding of the functioning and dynamics of soil. In many cases, enzyme activities are early predictors of the effects of soil management on soil quality and how rapidly these changes are expected to occur.

Numerous studies have shown, in side-by-side comparisons, that enzyme activities are sensitive to soil-management effects such as crop rotation, tillage versus no-tillage, and forest management. A few studies have shown that enzyme activities can be temporally sensitive within the first 2 years of changes in soil management, long before there are measurable differences in soil organic matter. Figure 1 shows such an early effect, with a distinct separation after 3 years between soils managed with or without winter cover crops without measurable changes in soil organic matter. However, not all enzymes can detect soil-management effects, and activities naturally vary as a function of soil type.

A few enzymes have potential as indicators of viable soil microbial biomass or activity. Most notable is dehydrogenase, but for the problems mentioned in the previous section it has not worked well as an indicator of microbial activity. Furthermore, enzymes that correlate closely with microbial activity may be less suited to predict long-term changes or trajectory in soil quality because they reflect recent management or seasonal (climatic) effects that may be transitory.

The effect of air-drying on enzyme activity varies with the enzyme and can cause an increase in activity, but for most enzymes activity is reduced by 40–60%. Screening of a range of enzymes has shown that, although activity may decrease with air-drying, the relative change (or rank according to field management) in activity for some enzymes within the same

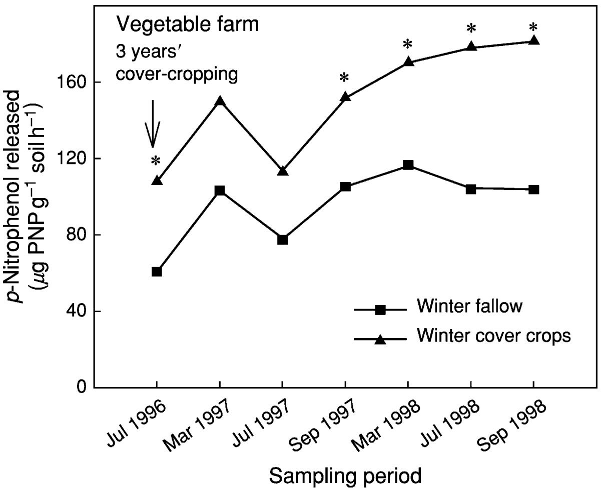


Figure 1 β-Glucosidase activity after 3 years of winter-fallowing or cover-cropping following summer vegetable crops in western Oregon, USA. (Adapted from Ndiaye EL, Sandeno JM, McGrath D, and Dick RP (2000) Integrative biological indicators for detecting change in soil quality. *American Journal of Alternative Agriculture* 15: 26–36. CABI Publishing, with permission.)

Table 1 The response of enzyme activities to the type of vegetation and soil

Soil enzyme activity	*Range of activities*	*Vegetation/soil type*
Xylanase activity (mg glucose g^{-1} per 24 h)	13–24	Spruce forest/n.d.
	0.28–8.0	Beech forest/n.d.
	1.8–3.0	Grassland/orthic luvisol
	0.24–1.83	Agricultural land/haplic luvisol, entisol
β-Glucosidase (μg *p*-nitrophenol g^{-1} h^{-1})	20–55	Grassland/pachic arguistoll
	36–160	Forest/Haplohumult
	130–310	Crop rotation/hapludalf
	71–86	Crop rotation/pachic ultic argixerolls
	41–253	Crop, manured soil, pasture/typic haploxeroll
Protease activity (μg tyrosine g^{-1} per 2 h)	150–520	Agricultural land/haplic chernozem
	224–514	Pasture/typic dystrochrept
	120–430	Wheat seeds/loamy sand
	198–288	Crop rotation/haplic luvisol
Arginine deaminase activity (μg N g^{-1} h^{-1})	2.5–5.0	Grassland/pachic arguistoll
	1.7–2.0	Crop rotation/phaeozem, lithosol, cambisol
	4.0–11.0	Forest/sandy soils
	0.1–1.3	Crop rotation/fluventic ustochrept
Arylsulfatase activity (μg *p*-nitrophenol g^{-1} h^{-1})	30–50	Grassland/pachic arguistoll
	115–340	Agricultural land/hapludoll
	6.9–213	Pasture/typic dystrochrept
	21–49	Forest/podzol
	12–58	Crop, manured soil, pasture/typic haploxeroll
Alkaline phosphatase (μg *p*-nitrophenol g^{-1} h^{-1})	40–80	Grassland/pachic arguistoll
	40–790	Agricultural land/aeric vertic epiaqualfs
	100–500	Crop rotation/hapludalf
	181–225	Crop rotation/ustochrept
Dehydrogenase (μg TPF g^{-1} 24 h)	114–155	Crop rotation/haplumbrepts, hapludalfs
	0.6–0.9	Crop rotation/fluvisol
	68–97	Crop rotation/ustochrept
	148–207	Crop rotation/fluventic xerochrept

Adapted from Kandeler E, Tscherko D, Stemmer M, Schwarz S, and Gerzabek MH (2001) Organic matter and soil microorganisms – investigations from the micro- to the macro-scale. *Die Bodenkultur* 52: 117–131.
n.k., soil type not known.

soil type remains the same. Air-drying stabilizes enzyme activity, greatly facilitates sample handling, and allows for timely analysis (unlike most other microbial measures that must be done as soon as possible after sampling). Combining this with the relative simplicity of many enzyme assays makes it possible to run a large number of samples on a routine basis and enables adoption by commercial soil-testing laboratories.

In general, hydrolytic enzymes are good choices as soil-quality indexes because it is likely that organic residue-decomposing organisms are the major contributors to soil enzyme activity. However, there is evidence that some of these enzymes can be confounded by long-term applications of fertilizers or liming. Phosphatase activity can be depressed by phosphate fertilizers and is also affected by pH, which is independent of other factors of soil quality such as organic matter content. Also, there is limited evidence that N fertilizers can have a similar effect on certain enzymes involved in the N cycle (e.g., urease and amidase).

Enzymes involved in the C cycle are thought to be better choices as a soil quality indicator than enzymes involved in cycling of nutrients that are heavily fertilized in agricultural systems. It seems likely C-cycling enzymes are more closely related to organic matter inputs, soil organic matter, and disturbance, all of which are related to soil quality. One such enzyme that that has been sensitive to soil management in many studies in β-glucosidase, which releases C for energy. Also, arylsulfatase has been an effective discriminator of soil-management effects and has been correlated with fungal biomass.

Enzyme assays have been used to evaluate the degree of highly degraded soils and polluted soil. Highly degraded soils can result from excessive erosion in agricultural or forest soils or from strip-mining activities. Enzyme expression has been closely related to the degree of recovery and corresponding microbial life and plant productivity in highly disturbed soils in the first years after remediation is initiated. Carbon hydrolytic enzymes appear to be well correlated with plant productivity in early stages of soil recovery, which is probably related to decomposition and organic matter accumulation being closely coupled with soil quality.

Enzyme assays on soils polluted with heavy or crude oil fractions have shown that relatively high rates of oil must be applied before enzyme activity decreases, whereas lighter petroleum products do not generally inhibit enzyme activity. Consequently, enzyme activity does not appear to be an appropriate technology for characterizing hydrocarbon-polluted soils. Pesticides have had no effect, stimulatory, or only short-term effects on soil enzyme activities even at pesticides rates far in excess of recommended rates for pathogen or weed control. However, in the case of heavy metal-contaminated soils, soil enzyme assays have potential for practical applications to assess bioavailability of metals. However, each metal often may require a specific enzyme assay to reflect the bioavailability of metals to soil microorganisms or plant availability. These assays are advantageous over total or even extractable metal content, because they better reflect toxicity to biological organisms, which for metals varies widely as a function of soil characteristics such as organic matter content/ chemistry, textural distribution, and mineralogy.

It is important to recognize that soil enzyme activity is operationally defined. If the conditions of the assay (e.g., temperature, buffer pH, buffer type, ionic strength) are altered, results will also change. Therefore, to make meaningful comparisons among studies or over different time periods, it is important that the exact same protocol be followed for each enzyme assay.

Systematic studies across soil types, environments, and soil-management systems are still needed to fully determine the potential of soil enzyme activity to characterize soil quality and develop calibration data to interpret enzyme activities. Therefore, enzyme activities should be interpreted with caution and be measured along with other soil properties to assess soil quality.

Ecological Dose Value

The ecological dose value (ED_{50}) is analogous to LD_{50} (lethal dose at 50% kill rate) used for assessing the toxicity of substances on animal and human life. Applying this to enzyme activity means that an ED_{50} value would be the pollutant (inhibitor) concentration required to cause a 50% inhibition of enzyme activity in soil. Two general models have been used: the sigmoidal dose–response curve; and the Michaelis–Menten kinetic model.

Thus far, this approach has only been used to assess heavy metal pollution of soils; but presumably this could be used on a relative basis to quantify the impact of other soil pollutants or of highly disturbed landscapes (e.g., strip mines) on soil enzyme activities.

Detoxification of Polluted Soils

Enzyme technologies are emerging as a means to remediate polluted soils. Common organic contaminants include pesticides, volatile hydrocarbons from industrial and automobile-related compounds (e.g., benzene, toluene, trichloroethylene), polycyclic aromatic hydrocarbons from fossil fuel wastes, polychlorinated biphenyls (PCBs) from electrical insulation, and chloroderivatives, chlorophenols, and chlorobenzene from paper mill effluents. A range of metals also commonly contaminates soils from industrial activity and mismanagement of irrigated agricultural land.

The goal of remediation is to transform pollutants into innocuous products. There are a number of potential ways that enzyme remediation processes can be invoked through microbial stimulation or introduction of a specific organism(s) capable of degrading a particular pollutant. Another approach is to add an enzyme or a suite of enzymes appropriate to degrade and detoxify a specific pollutant or to add substrates to soil that upon enzymatic hydrolysis release products that remediate soils.

Enzymes added to soils in a soluble form have been less effective, because the enzyme in this state is more susceptible to degradation by microorganisms, adsorption on mineral surfaces, and complexation in organic matter, and is not likely to be reusable. Consequently, the preferred method has been to stabilize the enzyme on a solid support by various chemical and physical immobilization techniques such as entrapment, encapsulation, covalent bonding, adsorption, and cross-linking or cocross-linking with bifunctional agents. Although immobilization typically reduces activity, the net effect is greatly increased potential for detoxification of soils by increasing stability and reusability of the enzyme. Supports have included clays (surface adsorption) and entrapment in gels, porous glass, or silica beads.

Summary

Soil enzymes are central to ecosystem processes because they catalyze innumerable reactions in soils that have biogeochemical significance. Catalytic soil enzymes can exist internally or on surface membranes of viable cells, be excreted into soil solution, or be complexed in the soil matrix or microbial debris. Extracellular enzymes may play an ecologic role for some microbial community members by hydrolyzing substrates that are too large or insoluble for direct absorption by microbial cells. More than 100 enzymes have been characterized in soils. With the exception of dehydrogenase and possibly a few other enzymes, which only exist in viable cells, nearly all other enzymes exist in both viable and complexed forms, independent of viable cells, stabilized in the soil matrix.

Research on soil enzymes provides insights into biogeochemical cycling of C and other nutrients and on microbial community functions in space and time. A relatively small amount of any given enzyme can be directly extracted from soil; therefore enzymes are mainly studied by measuring activity. The activity of enzymes varies temporally (seasonal), which often corresponds to microbial community responses to the environment, vertically (decreasing from the surface), at microscales, according to microbial community distribution, and at landscape level, where soil type is a major controlling factor (particularly textural distribution and organic matter).

Soil enzyme assays are emerging as technological tools for various applications in environmental and ecosystems management. Several enzymes have shown sensitivity in reflecting early changes (1–3 years) in soil quality due to soil management long before there are measurable changes in total organic C levels. This holds potential to guide ecosystem management for long-term sustainability. Enzyme assays can detect the level of degradation and recovery of soils in highly disturbed landscapes such as reclaimed strip-mine landscapes. The bioavailability of certain heavy metals in soils can be reflected with enzyme-activity measurements. Enzymes stabilized on colloid surfaces and incorporated into soils have been shown to degrade certain contaminants in soils.

There is still considerably more information needed to understand the ecology and function of extracellular enzymes in soils because of the diversity and complexity of the soil physical and chemical environment and microbial communities. Partly because of this, utilization of enzyme technologies requires careful consideration for interpretation and application. This is particularly true for assessing soil quality, where soil enzyme activities should be used in conjunction with other key soil measurements. Further research is needed to develop mechanisms for calibrating and interpreting soil enzyme technologies that are independent of soil type.

See also: **Pollutants:** Persistent Organic (POPs)

Further Reading

Burns RG and Dick RP (eds) (2002) *Enzymes in the Environment: Activity, Ecology, and Applications*. New York: Marcel Dekker, Inc.

Dick RP, Breakwill D, and Turco R (1996) Soil enzyme activities and biodiversity measurements as integrating biological indicators. Doran JW, Jones AJ *et al.* (eds) Handbook of Methods for Assessment of Soil Quality, pp. 247–272. SSSA Special Publication 49. Madison, WI: Soil Science Society of America, Inc.

Kandeler E, Tscherko D, Stemmer M, Schwarz S, and Gerzabek MH (2001) Organic matter and soil microorganisms – investigations from the micro- to the macro-scale. *Die Bodenkultur* 52: 117–131.

Kiss I, Dragan-Bularda M, and Radulescu D (1975) Biological significance of enzymes in soil. *Advances in Agronomy* 27: 25–87.

Ladd JN (1978) Origin and range of enzymes in soil. In: Burns RG (ed.) *Soil Enzymes*, pp. 51–96. London, UK: Academic Press.

Ndiaye EL, Sandeno JM, McGrath D, and Dick RP (2000) Integrative biological indicators for detecting change in soil quality. *American Journal of Alternative Agriculture* 15: 26–36.

Skujinš J (1978) History of abiontic soil enzyme research. In: Burns RG (ed.) *Soil Enzymes*, pp. 1–49. London, UK: Academic Press.

Tabatabai MA (1994) Soil enzymes. In: Weaver RW, Angle S, Bottomley P *et al.* (eds) *Methods of Soil Analysis,* part 2, *Microbiological and Biochemical Properties*, No. 5. Madison, WI: Soil Science Society of America, Inc.

EROSION

Contents

Irrigation-Induced

Water-Induced

Wind-Induced

Irrigation-Induced

G A Lehrsch, D L Bjorneberg, and R E Sojka, USDA Agricultural Research Service, Kimberly, ID, USA

Published by Elsevier Ltd.

Introduction

Soil erosion is caused by wind, tillage, precipitation, or irrigation. Erosion caused by irrigation, usually termed 'irrigation-induced erosion,' can be the most damaging because it affects many of the most productive soils in the world. These are the soils of arid irrigated regions, which typically have thin A horizons, little organic matter, and weak structure, making them highly erodible. Moreover, these soils, once degraded, recover very slowly. Irrigation-induced erosion occurs as an unintended consequence of irrigation for improved crop production.

To produce food and fiber worldwide, irrigation is vital. Irrigation enables crops to be produced in many areas where they could not otherwise be grown. In other drought-prone areas, irrigation on average doubles crop yield and nearly triples crop value, while improving production reliability and commodity quality. According to the Food and Agriculture Organization (FAO) of the United Nations, irrigation is practiced on only approximately 5% of the world's food-producing land, which includes rangeland and permanent cropland. That irrigated land, however, produces approximately 30% of the world's food. Similarly in the USA, only 15% of harvested cropland is irrigated, yet that land produces 40% of the nation's total crop value.

Three basic types of irrigation are drip, surface, and sprinkler. Drip irrigation supplies water to growing plants at very small rates, wetting relatively small soil volumes either at or below the soil surface. Properly designed and operated drip systems produce neither erosion nor runoff. In contrast, surface (or gravity-flow) irrigation requires water flow across the soil surface and is often designed to produce runoff to improve irrigation uniformity. With overland flow, however, comes erosion. In surface irrigation, the soil surface is the conduit used to deliver and distribute water. Surface irrigation that occurs (1) on sloping areas includes graded furrows (small ditches parallel to crop rows) and border strips, and (2) on relatively flat areas includes level or contour basins, terraces, and wild flooding. Sprinkler irrigation practices, too, can produce both runoff and erosion if not designed and managed properly. In sprinkler irrigation, water droplets are distributed through the air to the soil. Sprinkler irrigation includes: (1) moving lateral systems, including center-pivot, lateral-move, and big-gun systems; and (2) stationary systems, including solid-set and side-roll systems.

Irrigation-induced erosion from sprinkler irrigation resembles that from rainfall in many ways. In both cases, water droplet impact can deteriorate surface soil structure by fracturing soil aggregates, thereby producing aggregate fragments, primary particles, or both that can obstruct surface pores leading to surface sealing and increased runoff. Water that does not infiltrate into the profile accumulates on the surface and, once surface depression storage is satisfied, runs off, often transporting detached soil downslope or off-site. Water droplet impact not only detaches soil but also increases turbulence in shallow flow, increasing the amount of sediment the flow can transport.

There are, however, notable differences between erosion from rainfall and from sprinkler irrigation. For sprinkler irrigation: (1) only a portion of the field receives water at any given time, (2) water droplet characteristics vary from system to system, and (3) irrigation is controlled and managed to apply water only when the growing crop needs more soil water or in preparation for planting, tillage, or harvest. An area's rainfall is usually very low in total dissolved solids (TDS) and its chemical composition changes little. In contrast, irrigation water contains TDS and can vary chemically as a function of water source.

The differences between erosion from surface irrigation and from rainfall are even more distinct. The key difference is surface irrigation's lack of water droplet kinetic energy, which affects surface soil structure and thus infiltration, runoff, and erosion. Also absent is the additional turbulence in overland and rill flow caused by droplet impact. In furrow irrigation, water is applied to only a small portion of the soil surface. Erosion from surface irrigation most often occurs during a number of small events rather than one or two large events, characteristic of erosion from precipitation. Water temperature, affecting water viscosity, is more likely to change during a 12- or 24-h irrigation under cloudless skies than during a rainstorm. The hydraulics of rill flow from rain also differ from those from surface irrigation. In rainfall rills, flow volume increases as water accumulates downslope. In furrow irrigation, the flow rate and volume decrease with distance down the furrow but increase with time as the soil's infiltration rate decreases. These processes gradually change the furrow stream's sediment detachment and transport capacities with both time and distance from the furrow inlet. As the irrigation proceeds, upper furrow ends often become deeper and narrower owing to detachment and transport from relatively large inflows, while the lower furrow reaches become shallower and wider owing to deposition from reduced flow. The duration of inflow, often 12 h or more, is much longer than the runoff from most rainfall events.

Sediment concentration in runoff tends to decrease with time during a furrow irrigation, but not necessarily during a rainstorm. In a furrow during irrigation, many factors change, which, in combination, may explain this phenomenon. Loose soil, frequently positioned in the furrow by recent tillage or cultivation, is often flushed from the furrow early in the irrigation. At the furrow head, coarser, more erosion-resistant fragments may armor the furrow bottom. As soil in the furrow becomes wetter, there is less tendency for the rapid aggregate disintegration that is common during the initial wetting of hot, dry soil. In the lower furrow reaches, deposition can cause the furrow to widen, thereby decreasing its flow depth and reducing shear.

The chemical composition of irrigation water affects irrigation-induced erosion, whether from sprinkler or surface irrigation. High sodium concentrations or sodium adsorption ratios (SAR) and low electrical conductivity (EC) in irrigation water allow the diffuse double layers of 2:1 clay domains to thicken, dispersing clays and weakening or fracturing aggregates. Primary particles, released from aggregates as clay disperses, and aggregate subunits obstruct surface pores, increasing both runoff and soil loss. In addition, small aggregates or fragments, rather than large ones, are more easily transported in overland flow, once they are detached. Moreover, irrigation-water chemistry can change markedly with water sources and sometimes through the irrigation season, as water sources change or as upstream return flow is mixed in changing proportions with surface water.

Significance

Furrow irrigation is an inherently erosive process. It is exacerbated by the need for long fields to increase farming efficiency and for clean tillage to ensure uniform and steady flow of water down the furrow. Soil erosion from irrigation occurs across entire fields as a consequence of overland flow and, from sprinkler irrigation, droplet impact. Soil or sediment loss, in contrast, is a measure of the sediment entrained in runoff that leaves a furrow or field at its outlet. Measured soil loss is often much less than the field total of eroded soil, predominantly from upper furrow reaches, because much sediment is redistributed and, as flow rates decrease, often deposited on to lower furrow reaches before it can leave the field in runoff. Annual soil loss from surface-irrigated fields can vary from less than 1 Mg ha^{-1} to more than 100 Mg ha^{-1}, depending on crop type, field slope, soil properties, and water management, particularly flow rate. A single 24-h furrow irrigation of erodible soil on slopes of more than 2% has caused more than 50 Mg ha^{-1} of soil loss in runoff. Little erosion occurs from level fields, surface-irrigated pastures, or fields producing forage. In contrast, much erosion occurs from row crops grown on fields with steeper slopes, generally those exceeding 2%.

The magnitude of sprinkler irrigation-induced erosion is not well documented for at least two reasons. First, it is difficult to measure, particularly so because it varies widely across time and space. Second, it tends to be an on-field problem occurring only in the area being irrigated at that time. Although sprinkler irrigation is normally regarded as a less-erosive alternative to surface irrigation, problems sometimes occur, particularly where systems are improperly designed or poorly operated. Farmers may irrigate excessively steep slopes with sprinklers, creating erosion problems because they have exceeded their irrigation system's design limits. Where center pivots with high-volume end guns are placed on rolling topography, the combination of high application rates, variable sloping land, and tower-wheel tracks can produce severe erosion in a single irrigation or in one season.

Erosion, whether occurring from sprinkler or surface irrigation, is caused by humans. Consequently, with an understanding of the processes involved, properly designed irrigation systems, and enlightened, skillfull management, irrigation-induced erosion can be nearly eliminated in many cases or at least adequately controlled.

Erosion Under Sprinkler Irrigation

Processes Causing Soil Loss

Soil erosion from water, whether caused by precipitation or irrigation, can be described in terms of three components or processes: detachment, transport, and deposition. Detachment is the release of soil aggregates, aggregate fragments, or primary particles from the soil surface as a consequence of energy input, usually from droplet impact or shear from runoff flow. Transport occurs as detached soil, that is, sediment or bedload, is splashed about and carried downslope in overland flow. Deposition occurs as sediment settles out of the flow as the water's carrying capacity for sediment is exceeded. Depending upon flow hydraulics, deposition may occur within a few meters of the detachment point or may not occur until the sediment is transported off-site.

When properly designed and carefully operated, stationary sprinkler systems, especially solid-set systems with a grid of simultaneously operating sprinklers, apply water for lengthy periods at a relatively low rate (e.g., $3\,\mathrm{mm\,h^{-1}}$). The soil's infiltration rate is seldom exceeded, so little (if any) runoff or erosion occurs. In contrast, center-pivot systems, with a moving lateral that pivots around a fixed point, apply water at higher rates (e.g., $80\,\mathrm{mm\,h^{-1}}$) to smaller areas (e.g., 5 to 20-m-wide strips) than solid-set systems. With center-pivot irrigation, the irrigated area per unit length of lateral must increase with distance from the pivot point. Consequently, the outer spans of pivots have relatively high discharge rates per unit lateral length (e.g., $15\,\mathrm{l\,min^{-1}\,m^{-1}}$) and high instantaneous application rates per unit wetted area. This greatly increases the potential to exceed a soil's infiltration rate, causing runoff and erosion.

Soil erosion from sprinkler irrigation is directly proportional to the application rate in the wetted area which, in turn, is affected by sprinkler type. Low-pressure-type spray heads, which are relatively economical to operate and thus have become popular, have reduced pattern widths and increased application rates relative to other sprinkler types. Again, high application rates can lead to erosion, runoff, and soil loss.

Water-drop impact, or more specifically droplet kinetic energy, detaches surface soil particles and splashes the detached soil in all directions. Some of the soil entrained in the infiltrating water obstructs surface pores. Droplet energy also compacts surface soil. The increased bulk density and obstructed pores reduce infiltration. Droplet kinetic energy also causes turbulence in shallow surface flow, increasing the flow's carrying capacity for sediment. An irrigation's total kinetic energy is a function of its droplet size distribution; the larger the droplet, the greater the kinetic energy. Droplet size distributions can be altered within limits by modifying nozzle pressure, nozzle size, and spray-head deflector plate. Sprinkler irrigation system designers must often balance desired design parameters with environmental and economic constraints.

Slope and topography also affect erosion processes from moving lateral sprinkler systems, particularly center-pivot systems. Depending upon the slope and the pivot's direction of travel, runoff can move on to dry soil, with relatively large infiltration rates, or previously wetted soil, with much smaller infiltration rates. In the first case, runoff rates decrease rapidly, fortunately because the dry soil is easily eroded. In the second, runoff accumulates and concentrates in rills or larger, ephemeral gullies, increasing in rate, erosivity, and sediment-carrying capacity. In the special case where the pivot lateral is parallel to the slope direction, the effective wetted slope length is long and erosion can be particularly severe. Under both center-pivot and lateral-move systems, the tower-wheel tracks are relatively large flow paths 40–50 m apart, with smeared and sealed surfaces underlain by compacted soil. Runoff is common in wheel tracks where they are parallel to the slope direction. Even where the tower-wheel tracks cross the slope, the tracks cause problems, because they collect and concentrate overland flow.

Practices Controlling Soil Loss

Irrigation practices New irrigation systems must be properly designed. Central to the design is an accurate estimate, preferably based upon measurements, of the soil's infiltration characteristics, particularly the infiltration decrease with time. Both new and existing systems must be operated in accordance with (1) design parameters such as nozzle diameter and pressure, and (2) operational guidelines such as set times and travel speed.

To control irrigation-induced erosion, one must minimize runoff. Without runoff, there will be no sediment transport apart from splash at the point of detachment. To minimize runoff, irrigators should schedule irrigations using scientific techniques and apply no more water than is needed for maximum economic yield. From an erosion-control standpoint, no runoff should be the goal.

Modifying the sprinkler type, nozzle pressure, and nozzle diameter alters both the application rate and wetted area. Changes that decrease the sprinkler flow rate, decrease the application rate, or increase the wetted area minimize erosion, runoff, and soil loss. For spray heads, changing the nozzle and deflector plate changes the drop size distribution. Shifting the distribution to smaller and fewer large droplets reduces total droplet kinetic energy striking the soil, thus reducing detachment. Disadvantages of such a size distribution change are that smaller droplets evaporate more readily and are more susceptible to wind drift, which distorts the spray pattern, decreasing both irrigation uniformity and efficiency. Another disadvantage of small droplets is that they travel relatively short distances, giving the spray head a small wetted diameter and high application rate.

A goal of irrigation system design and operation is to match the system's application rate to the soil's infiltration rate (to minimize runoff), both spatially and temporally. This goal is difficult to achieve, however. One relatively new technique with promise for moving lateral systems is to use variable-rate sprinklers that can be programmed to operate on a site-specific basis. Appropriately programmed, the sprinklers could change their discharge rate, depending upon field slope, soil-infiltration differences, presence of rock outcrops, or other factors.

Soil and crop management practices One effective way to help reduce erosion caused by early-season irrigations is to eliminate unneeded seedbed-preparing tillage. In the spring, surface soil aggregates of many soils are structurally weak and susceptible to breakdown from tillage or droplet impact. Unnecessary springtime tillage weakens or breaks particle-to-particle bonds within aggregates, often fracturing them. Aggregate fragments and primary particles are more easily transported than are larger, intact aggregates. Moreover, such tillage buries crop residue and indirectly destroys soil organic matter, further weakening aggregates.

Some tillage practices, on the other hand, instead of contributing to soil erosion can help control it. One such practice, paratilling, uses broad, angled subsoiling shanks to partially lift and laterally shatter soil, increasing the tilled soil's infiltration rate, often substantially, thereby decreasing runoff and soil loss. Another tillage practice that decreases runoff is reservoir tillage. In this postplant operation, small water-storage basins or pits are formed at intervals across a field's surface. Those basins increase surface-depression storage by collecting and temporarily holding water, allowing the water to infiltrate rather than run off. Reservoir tillage reduces runoff, even when an irrigation system's application rate somewhat exceeds the soil's infiltration rate. This practice is particularly effective where performed under the outer spans of center pivots, where application rates often exceed soil infiltration rates.

No-till and conservation tillage are other tillage practices that reduce irrigation-induced erosion. These practices leave crop residues on the soil surface as mulch. Surface mulch absorbs droplet kinetic energy, protects soil structure, and maintains surface roughness, thereby minimizing the decrease in the soil's infiltration rate with time. No-till or conservation tillage also keep soil surfaces rougher, increasing both depression storage and the tilled soil's initial infiltration rate. Within limits, crops in a rotation can be sequenced to produce crop residue regularly throughout a multiyear rotation. A canopy of growing vegetation also absorbs droplet energy, reducing energy input directly to surface soil. Production practices that hasten canopy coverage can reduce erosion from droplet impact, and may reduce erosion from overland flow by shading surface aggregates and keeping them moist and less susceptible to slaking. Vegetation on the soil surface also slows runoff and absorbs overland flow shear.

Another management practice that helps to control runoff, thus soil loss, on slightly sloping surfaces is to till or plant so that the final tillage or planting marks are perpendicular, rather than parallel, to the slope direction. On rolling topography, one should practice contour tillage, in which both tillage and planting operations are performed on the contour, as much as possible. These practices slow runoff, allowing more time for water to infiltrate into the soil.

Erosion with Surface Irrigation

Processes Causing Soil Loss

In surface irrigation, as water flows across a soil's surface or advances down a furrow, it quickly wets relatively dry aggregates or clods in its path. As a consequence of the small matric potential in the dry soil, water will quickly enter the aggregate from all directions, causing 2:1 clay domains to swell, displacing O_2 and N_2 from particle surfaces, and often compressing those gases and air within the aggregate. As the compressed air finally escapes, the force it exerts often fractures interparticle bonds within the aggregate, or the aggregate itself, liberating aggregate fragments and primary particles. This process, in which an air-dry aggregate breaks into subunits or fragments when quickly wetted or immersed in low-electrolyte water, is termed 'slaking.' It contributes substantial amounts of soil for transport in the furrow stream, accounting in large part for the relatively great sediment concentrations often observed early in an irrigation.

Water must flow across the soil during surface irrigation. This flowing water exerts shear along the wetted perimeter, detaching soil once the imposed shear exceeds a threshold, termed the 'critical shear stress.' In a furrow, this critical shear varies both spatially and temporally. In addition to detaching soil, the flowing water transports detached soil downslope, further contributing to the erosion process. Level-basin irrigation systems may have no runoff, thus no soil loss from the basin. Other surface systems on sloping fields, in contrast, have runoff. To ensure adequate wetting of the soil near their field or furrow outlet, those surface irrigation systems are designed and operated so that 20–40% of the added water runs off. Thus, without proper precautions and management, soil loss will occur from many surface-irrigated areas.

Competing processes affect the erosivity and hydraulics of the flowing irrigation water. Infiltration through the wetted perimeter reduces the furrow flow rate with distance from the furrow inlet. This decrease in flow rate with distance reduces the furrow stream's shear and carrying capacity, at times leading to sediment deposition. As time passes, however, the soil's infiltration rate decreases and, with no change in the inflow rate, the furrow flow rate increases. Increasing the flow rate increases the shear and carrying capacity. Also, as much of the slaked and easily eroded soil is flushed from the furrow early in the irrigation, the sediment concentration in the furrow stream often decreases. This decreasing sediment concentration with time (and with increasing flow rate) increases the furrow stream's transport capacity.

Practices Controlling Soil Loss

Irrigation practices One of the best ways to control erosion of surface-irrigated land is to convert to a well-designed sprinkler irrigation system, with its higher efficiency, better application uniformity, minimal runoff, and often reduced labor needs. Sprinkler irrigation does require, however, more energy, a larger capital investment, and a greater level of management than surface irrigation. Sprinkler irrigation can also encourage disease and may not meet peak crop water demand. Thus, such conversion is not possible or practical in every situation, and other practices must be used to control erosion under surface irrigation.

As mentioned above, one must minimize runoff to minimize soil loss from irrigated fields. With surface irrigation, this goal is more difficult to achieve, because runoff is usually necessary to assure adequate application uniformity. None the less, irrigation should still be performed to produce no more runoff than is needed. Scientific irrigation scheduling, good water control, and close monitoring of ongoing irrigations help to minimize both runoff and soil loss.

In some areas, irrigators may be able to shorten furrow lengths. This reduces erosion, because the inflow rate can be reduced yet still allow the furrow stream to advance to the outlet in a reasonable length of time, termed 'advance time,' usually 25–40% of the total set time. Reducing inflow rates is desirable because much detachment and transport occurs near furrow inlets, where furrow flow rates are highest. On some fields, furrow length can be halved by adding a midfield gated pipe to supply the needed inflow. Shortening furrow lengths, however, may increase runoff and soil loss from the entire field (because twice as many furrows are producing runoff) and always requires more labor. For example, adding a midfield pipe doubles the number of furrows that need to be set and the pipe itself must be moved when performing field operations. If field size is reduced to shorten furrow lengths, then more time will be required to plant, till, and harvest those smaller fields. In many areas, furrow lengths cannot be shortened due to existing return-flow channels.

In some situations, furrows may be oriented to cross the slope slightly, rather than run parallel to the slope direction. This repositioning reduces the furrow's slope, reducing the flowing water's shear on the soil along the wetted perimeter, thus reducing both sediment detachment and transport capacity. Repositioning furrows may lead, however, to increased erosion of the now-steeper tailwater collection ditch.

Another means of reducing erosion is to manage furrow inflow rates and advance times appropriately. Inflow rates must be large enough for the furrow stream to reach the outlet, but, once runoff begins, the inflow rate can be reduced ('cut back') to minimize erosion near the furrow inlet as well as runoff at the furrow outlet. Monitoring is required, however, because if a furrow's inflow is reduced too much, its outflow may cease, greatly reducing the uniformity of water application in that furrow. Also, to minimize differences in intake opportunity time from furrow inlet to outlet, irrigators would like advance times to be relatively small. However, a tradeoff must be made, since smaller advance times require greater inflow rates, yet those greater rates increase erosion near furrow inlets.

Some producers use surge irrigation to improve application uniformity. Surge irrigation is a technique wherein flow is applied intermittently ('surged') during a single irrigation set to overcome initially high infiltration rates near furrow inlets. While surge irrigation helps infiltration to be more uniform from furrow inlet to outlet, it must be used carefully or erosion near furrow inlets can be greater with its intermittent inflow than with continuous inflow if inflow rates are higher when surging than when not.

Irrigation water quality can be changed to reduce soil loss. In some areas, one may be able to mix water sources or otherwise add electrolytes to alter inflow water chemistry, principally by increasing the water's Ca^{2+} concentration. Increasing the concentration of divalent cations in the irrigation water reduces the thickness of 2:1 clay domains' diffuse double layer. This minimizes clay dispersion and enables aggregates to remain intact, less susceptible to transport downslope in the furrow stream. Since the divalent cations stabilize soil structure along furrow-wetted perimeters, they also lessen infiltration decreases with time that make furrow-irrigation management difficult. Gypsum is commonly added to water with very low EC or high SAR to improve its suitability for irrigation.

Runoff management practices Runoff can also be managed to minimize, or at least control, soil loss under surface irrigation. One technique is to use pump-back runoff reuse systems, in which all runoff and sediment are collected in a reservoir at the field end, then pumped back to the inlet, where the runoff is reintroduced during the same irrigation as inflow to the field. While incurring energy and equipment costs for pumping, pump-back return systems offer many benefits. Reintroduced inflow that contains some sediment reduces furrow-stream sediment-carrying capacity. Depending upon flow hydraulics, sediment eroded from the field may be redeposited on to the field near its origin. Where irrigation return-flow water-quality regulations are stringent, irrigators with pump-back systems will have no off-farm (or off-site) discharge of sediment, fertilizer, pesticides, weed seeds, or microbes.

To collect or retain soil eroded from irrigated fields, settling basins varying in size and shape may be constructed along runoff collection channels, often at field ends. These basins collect much of the runoff and, under quiescent conditions, allow soil particles from the runoff to settle. Some basins are large (for collecting runoff from 20 ha or more); some are small (for runoff from only a few furrows). After draining the basins at the season's end, the collected sediment can be returned to the field. While offering this advantage, settling basins suffer from many disadvantages. Erosion still occurs in the field. Clay-sized soil, containing most of the P, other plant nutrients, and agricultural chemicals, does not fully settle out but is largely lost in the basin's outflow during the irrigation season. A settling basin's sediment collection efficiency declines as it fills with sediment, reducing residence time in the basin. Land area is taken out of production. Settling basins also require weed control, can be safety hazards, and can be the source of flying insect pests. Energy, time, and, for bigger basins, heavy equipment not common on farms are required to remove sediment from the basins and redistribute the sediment on to the field or another area. In spite of these disadvantages, settling basins have their place, particularly when used in combination with other erosion-control practices.

Buried drains with standpipes are a special type of settling basin. In a field's tail ditch, plastic, corrugated pipe is placed in a trench as a drain. Standpipes that extend vertically from the drain to just above the soil surface are installed every 5–10 furrows along the drain's length, then the trench is backfilled. Earthen dams are then constructed across the tail ditch, just downstream of each standpipe's inlet, thus forming a small basin at each standpipe. In operation, each dam forces runoff to pond, allowing some sediment to settle, before the runoff enters the standpipe's inlet and drains from the field. With appropriate management, this special drainage system can eliminate excessive erosion that often occurs at field ends where furrow slope increases sharply as runoff drains into a deep tail ditch. These drainage systems: (1) increase yields from field ends, (2) bring additional land into production, (3) ease farm equipment's ingress and egress across the lower field boundary, and (4) reduce weed problems common in and near wet tail ditches. Unfortunately, buried drains do not control erosion from the bulk of the field, and still allow some sediment to enter the drain and be transported from the field in the drainage water.

Soil and crop-management practices Placing mulch or maintaining crop residues in irrigation furrows effectively reduces both erosion and soil loss. Mulch in the furrow absorbs shear and slows furrow-stream velocity, thus reducing both sediment detachment and transport. By reducing flow velocity, the mulch can reduce overland flow by allowing the added water more time to infiltrate. If available, previous crop residues should be used as mulch, but in some rotations and areas, straw from off-site works well. While effective at controlling furrow erosion and often increasing crop yields, if not properly managed, the mulch tends to float downstream and obstruct the channel, damming the water which breaks over into adjacent furrows, increasing their flow while reducing the flow in the obstructed furrow. By increasing infiltration, mulch can increase erosion from upper furrow reaches if the mulched furrows require greater inflows. Mulch placed in level basins can hinder the even spreading of water, at times channeling it to erode some areas and underirrigate others. Instead of placing mulch in an irrigation furrow, one can establish semipermanent vegetation, such as turf, along the furrow's wetted perimeter, much like a grassed waterway. Turf, once established, nearly eliminates furrow erosion but complicates field management and can reduce crop yield. Turf-covered furrows are a viable practice only for rare cropping patterns and on very steep slopes.

Narrow or twin-row plantings also reduce erosion. By positioning crop rows on bed shoulders, close to an intervening irrigated furrow, the plant root systems stabilize soil along the furrow's wetted perimeter, while overhanging vegetation drooping into the furrow and plant debris reduces furrow-stream velocity, minimizing both detachment and transport.

Filter strips, often seeded to small grain or forage, can also be placed perpendicular to the furrow direction at the downstream end of row-crop fields to trap sediment that would otherwise leave the field in the furrow outflow. By slowing and spreading the flow as it progresses through the 3- to 6-m-wide strip, the furrow stream's carrying capacity is greatly decreased, with much sediment being deposited within the strip. Filter strips do not, however, prevent erosion from occurring upslope, nor do they produce much marketable yield from the crop seeded in the strip.

A recently developed, highly effective erosion-control practice is the adding of certain types of synthetic organic polymers to surface irrigation water. These polymers, high-molecular-weight, moderately anionic polyacrylamides (PAM), are added to inflow water to be present at dilute concentrations of approximately $10\,\mathrm{mg\,l^{-1}}$. When evenly distributed throughout the inflow early in an irrigation, PAM stabilizes soil along furrow-wetted perimeters and flocculates sediment that may be present in the flow. PAM-treated water also reduces seal formation in the furrow, thus slowing the decrease in the soil's infiltration rate with time. All told, their use reduces furrow soil loss by approximately 95%, economically (e.g., less than US\$40 $\mathrm{ha^{-1}}$) and with minimal additional management. PAM also reduces erosion and increases infiltration under sprinkler irrigation, but its use there requires specialized equipment and is not yet user-friendly.

Effective furrow-erosion control is also possible using whey, a natural organic by-product of cheese manufacture, at times viewed as a food-processing waste. When added without running off to newly formed furrows early in an irrigation season, it too stabilizes soil along wetted perimeters, in part owing to greatly enhanced microbial activity that leads to aggregate formation and stabilization at and below the wetted perimeter. Soil loss from subsequent irrigations of whey-treated furrows is reduced by 50–98% and infiltration increased by 50–60%.

A combination of practices can be particularly effective. PAM and/or conservation tillage can be used to reduce on-field erosion, while filter strips and small settling basins remove additional sediment before the runoff leaves the field. Larger settling basins and wetlands in return-flow streams can further reduce the runoff's sediment load before the runoff reaches receiving waters.

Summary

Controlling erosion on and soil loss from irrigated lands is critical to sustain agricultural production. Protecting and stabilizing the soil surface will minimize sediment detachment; slowing or reducing overland flow will minimize sediment transport. Reducing or managing runoff is the key to controlling soil loss wherever sprinkler irrigation or surface irrigation is practiced. Erosion caused by sprinkler irrigation is similar to that caused by rainfall, with many erosion-control practices effective for both. Techniques that protect the soil surface from raindrop or sprinkler-drop impact are effective in maintaining infiltration rates, reducing overland flow, and controlling both detachment and transport. Erosion processes with surface irrigation are quite different from those with rainfall, due to the absence of droplet kinetic energy input to the soil surface, and thus require different control strategies. Controlling erosion from surface irrigation is a challenge, due to the requirement for overland flow and runoff, and to varying flow regimes and soil infiltration rates.

For both sprinkler and surface irrigation, off-site soil loss is often least where combinations of control practices are employed. For surface irrigation, PAM use is not only economical but probably offers the most promise for effective erosion control for most furrow-irrigated production systems.

Prospects for Future Control

In the USA, surface irrigation is practiced on about 50% of the irrigated land; worldwide, however, more than 95% is surface-irrigated. Wherever surface irrigation is practiced, improved irrigation scheduling and better water control can reduce erosion and soil loss while minimizing off-site environmental damage. In furrow-irrigated areas where labor is available and relatively inexpensive, changing management practices to reduce runoff by shortening furrow lengths, reorienting furrows to reduce furrow slopes, and/or managing inflows will help reduce on-field erosion and off-site soil loss. In more industrialized areas, with established surface water quality standards, pump-back return systems offer the most comprehensive control of both runoff and soil loss. Filter strips and buried drains with standpipes can minimize future off-site soil loss. Without doubt, though, the use of PAM in surface irrigation holds the greatest potential for cost-effective erosion control.

Effective sprinkler erosion-control techniques already exist and more are on the horizon. Variable-rate sprinklers on center pivots will probably prove cost-effective for site-specific soil and water management to increase yields, improve water-use efficiency, and decrease water requirements while simultaneously reducing runoff and attendant soil loss. Engineering hindrances to PAM use in center pivots will probably be overcome, enabling PAM's erosion-controlling and infiltration-enhancing benefits to be extended to sprinkler-irrigated lands also. PAM's other environmental benefits, such as minimizing off-site discharge of sediment, weed seeds, plant disease agents, and microbes (including possible human pathogens), will become more important with stricter environmental regulations, spurring ever greater PAM use under irrigated conditions.

See also: **Erosion:** Water-Induced; **Irrigation:** Environmental Effects; **Overland Flow**

Further Reading

Bjorneberg DL, Kincaid DC, Lentz RD, Sojka RE, and Trout TJ (2000) Unique aspects of modeling irrigation-induced soil erosion. *International Journal of Sediment Research* 15: 245–252.

Carter DL (1990) Soil erosion on irrigated lands. In: Stewart BA and Nielsen DR (eds) *Irrigation of Agricultural Crops*, pp. 1143–1171. Agronomy Monograph 30. Madison, WI: American Society of Agronomy.

Kincaid DC and Lehrsch GA (2001) The WEPP model for runoff and erosion prediction under center pivot irrigation. In: Ascough JC and Flanagan DC (eds) *Soil Erosion Research for the 21st Century*, pp. 115–118. St. Joseph, MI: American Society of Agricultural Engineers.

Lehrsch GA and Robbins CW (1996) Cheese whey effects on surface soil hydraulic properties. *Soil Use and Management* 12: 205–208.

Lehrsch GA, Sojka RE, and Westermann DT (2001) Furrow irrigation and N management strategies to protect water quality. *Communications in Soil Science and Plant Analysis* 32: 1029–1050.

Sojka RE (1998) Understanding and managing irrigation-induced erosion. In: Pierce FJ and Frye WW (eds) *Advances in Soil and Water Conservation*, pp. 21–37. Chelsea, MI: Ann Arbor Press.

Sojka RE and Bjorneberg DL (2002) Erosion, controlling irrigation-induced. In: Lal R (ed.) *Encyclopedia of Soil Science*, pp. 411–414. New York: Marcel Dekker.

Trout TJ, Sojka RE, and Okafor LI (1990) Soil management. In: Hoffman GJ, Howell TA, and Solomon KH (eds) *Management of Farm Irrigation Systems*, pp. 872–896. St. Joseph, MI: American Society of Agricultural Engineers.

Withers B and Vipond S (1980) *Irrigation: Design and Practice*, 2nd edn. Ithaca, NY: Cornell University Press.

Water-Induced

J E Gilley, USDA Agricultural Research Service, Lincoln, NE, USA

Introduction

Water erosion is caused by the detachment and transport of soil by rainfall, runoff, melting snow or ice, and irrigation. Excessive erosion can threaten the production of agricultural and forest products. Erosion may also impact water conveyance and storage structures, and contribute to pollution from land surfaces. Water erosion may occur within rills, interrill areas (the regions between rills), gullies, ephemeral gullies, stream channels, forest areas, and construction sites. Rainfall characteristics, soil factors, topography, climate, and land use are important elements affecting soil erosion. Conservation measures that have been effectively used to reduce soil erosion on agricultural areas include contouring, strip cropping, conservation tillage, terraces, buffer strips, and use of polyacrylamide on irrigated areas. Specialized erosion control practices have been developed for use within stream channels, forest areas, and construction sites. One of the most effective means of reducing erosion is to maintain a vegetative or residue cover on the soil surface.

Impacts of Erosion

Erosion is a natural process. Topographic features such as canyons, stream channels, and valleys are created by long-term geologic erosion. Geologic erosion influences soil formation and distribution. Accelerated erosion results from the removal of natural vegetation by human activities such as farming, ranching, forestry, and construction.

The production of agricultural and forest products can be affected by excessive erosion. Erosion causes a breakdown of soil aggregates and accelerates the removal of organic and mineral materials. The loss of surface soil is critical because the exposed subsoil remaining following excessive erosion usually has reduced infiltration capacity, water storage, and nutrient characteristics. The exposed subsoil is usually finer-textured, making seedbed preparation and crop production more difficult. Smaller-size particles are more easily detached and transported by overland flow. Thus, sandy soils are made even coarser by erosion.

Sedimentation resulting from erosion may significantly reduce the effectiveness of water-conveyance and storage structures. The capacity and functional life of lakes, reservoirs, and streams can decrease as a result of sedimentation. The suitability of streams and rivers as a biological habitat and effective water supply can be affected by excessive amounts of sediment. The existence of sediment can also impair the use of streams and rivers as fish-spawning areas.

Nutrients, pesticides, and pathogens transported by sediment can contribute to pollution of streams and lakes, thus reducing their suitability for aquatic organisms and their use as water supplies. A large nutrient concentration in streams and lakes can also cause excessive vegetative growth, resulting in seasonal oxygen deficiencies. The type of fertilizer that is used, application rate, and nutrient content of the soil influence nutrient transport.

Types of Erosion

Water erosion can be separated into individual categories, each with distinct characteristics. Rills are small channels that form as runoff rate increases. The regions between rills are defined as interrill areas. Gully erosion occurs when concentrated flow is large enough to form large channels that cannot be crossed during normal tillage operations. Ephemeral gullies appear at the same position on the landscape each year, but they are small enough to be filled in by tillage operations. Stream-channel erosion may take place within a water-course that usually has continuous flow. Each of the erosion types may occur on croplands, rangelands and pastures, forest areas, and construction sites.

Rill Erosion

As overland flow moves downslope, it concentrates due to surface microtopography. Small channels or rills may form as the runoff velocity of overland flow increases. Rills often occur between crop rows or along tillage marks. The hydraulic shear of flowing water and soil properties influence rill erosion. Normal tillage operations usually remove rills. The soil materials detached within rills and sediment delivered from interrill areas are transported by rill flow. Once rills have formed, substantial amounts of erosion may occur, resulting in a loss of soil productivity.

Interrill Erosion

Raindrops impacting the soil surface detach soil particles on interrill areas. The detached soil particles may then be transported to rills by shallow overland flow. Soil properties, rainfall intensity, and slope all influence interrill erosion. Interrill erosion is often most apparent on the light-colored upper portions of convex slopes where tillage mixes surface soil and subsoil.

Gully Erosion

Deep channels larger than rills that cannot be removed by tillage are classified as gullies. Gullies usually form near the upper portion of intermittent streams or where trails, paths, or roads cause runoff to concentrate. In tropical areas, gullies may develop following deforestation and cultivation. The runoff-generating characteristics of the watershed influence gully formation and development. Once they form, gullies become a permanent part of the landscape. Gullies that separate portions of fields or pastures are a nuisance.

A gully continues to develop and move upslope where there is a water overfall. As water moves through the channel during large runoff events, a gully may rapidly expand and deepen. Large runoff events may also remove sections of the exposed banks that were previously undercut and had fallen into the channel. Gullies are described as active if their walls are free of vegetation and as inactive when they are stabilized by vegetation.

Ephemeral Gully Erosion

The topography of many fields causes runoff to collect and concentrate in a few natural waterways before leaving the field. These channels or ephemeral gullies serve as the primary drainage system for a field and form in the same locations each year. However, ephemeral gullies are transitory rather than permanent since they are small enough to be filled in by tillage operations. Ephemeral gully erosion occurs when soil particles are detached and transported from the concentrated flow channel.

Stream Channel Erosion

The removal of soil from the banks or beds of streams results in stream channel erosion. Only relatively small areas are affected by stream channel erosion, but the impacts on highly productive soils can be severe. Stream channels erode as water flows over the side of the stream bank or scours below the water surface, especially during severe floods. Stream channel erosion is usually greatest on the outside of bends. Meandering may also cause erosion along stream banks. A reduction in sediment delivery from upland areas through control practices or the capture of sediment in water-storage facilities may increase stream channel erosion.

Factors Affecting Erosion

Erosion is influenced by a variety of elements including rainfall characteristics, soil factors, topography, climate, and land use. Erosion models can be used to estimate the interrelated effects of these factors on erosion and to predict the impact of various land uses on runoff and soil loss.

Rainfall Characteristics

For erosion to occur, runoff must first be present. Runoff is defined as that portion of rainfall that does not infiltrate nor accumulate on the soil surface but moves downslope. Rainfall rate and duration both influence runoff and erosion. Runoff occurs only when rainfall intensity exceeds soil infiltration rate. Infiltration will decrease with time during the initial stages of a storm. Thus, no runoff may occur from a storm of short duration, while a storm of the same intensity but of longer duration may result in substantial runoff.

Rainfall intensity influences both the rate and volume of runoff. During a high-intensity storm, infiltration capacity is exceeded by a greater margin than during a less intense rainfall event. Thus, even though the precipitation amount may be similar for two events, a high-intensity storm will produce a greater volume of runoff.

Soil Factors

The susceptibility of soil particles and aggregates to detachment is influenced by soil characteristics. Soil texture, organic matter content, structure, and permeability have been shown to influence soil erodibility. Because they lack cohesiveness, sand particles are relatively easy to detach but they are difficult to transport by overland flow because of their large mass. In contrast, clay particles are difficult to detach since they readily bond together but are easily transported once they are separated from the soil mass. Silty soils are usually well-aggregated, but the aggregates break down rapidly when wetted, allowing nonaggregated soil particles to be easily transported. Soils containing large stable aggregates are difficult to detach and transport and usually have greater infiltration rates.

As soil organic matter content increases, individual aggregates become more stable, soil structure improves, and infiltration rate becomes greater. Cultivated soils recently removed from native vegetation, pasture, or meadow usually have excellent structure and stable aggregates. The incorporation of organic materials into the soil profile helps to increase aggregate stability. Cultivation without the addition of organic materials causes a reduction in aggregate stability and organic matter content, and an increase in soil erodibility.

More runoff usually occurs from fine-textured than from sandy soils because of differences in infiltration. Maintaining high infiltration rates is one of the most effective means of reducing erosion. Soil surface sealing caused by aggregate destruction and plugging of pores with soil particles may substantially reduce infiltration. High infiltration rates can be maintained when vegetative material protects the soil surface from sealing, when soil structure is preserved, and soil compaction is minimized.

Topography

Erosion is influenced by slope gradient, length of slope, and size and shape of the watershed. The velocity of flowing water becomes greater as slope gradient increases. A larger runoff velocity allows flowing water to detach and transport additional soil material. An increased accumulation of overland flow on longer slopes results in greater rill erosion. Convex slopes with a larger gradient at the bottom of a hillslope are more erosive than concave land surfaces. On concave surfaces, deposition frequently occurs at the bottom of the slope because of reduced transport capacity of flow. Crops growing in productive areas at the bottom of a hillslope may be covered with water and sediment during extreme precipitation events.

Climate

The total amount and intensity of rainfall influence the quantity of erosion within a region. The dense vegetation found on areas that receive substantial rainfall reduces erosion potential. Significant erosion may occur in regions with low rainfall and limited vegetation when high-intensity rainstorms do occur. Maintaining high infiltration rates not only reduces runoff and erosion, but also helps to maintain soil water supplies needed by vegetation.

Runoff from melted snow and ice moving over a thin layer of freshly thawed soil can cause substantial erosion. In colder climates, more erosion may result from snowmelt than from rainfall events. Frozen soil is not subject to erosion during several months of the year. However, significant runoff may result from the rapid melting that occurs when air temperatures rapidly rise or rain falls upon a snow-covered surface.

Land Use

Land use is the only factor affecting erosion that can be modified to reduce soil loss potential. Erosion results from different land-use conditions on cropland areas, rangelands and pastures, and forest areas.

Cropland areas Soil erosion potential can substantially increase as natural forest or rangeland is converted to cropland. Cropland areas least susceptible to erosion have a complete ground cover throughout the year. The amount of surface cover maintained on a particular site is influenced by cropping and management conditions. The greatest erosion potential on croplands exists after planting when residue cover is usually at a minimum and high-intensity rains frequently occur.

Cultivated land left fallow with no vegetative cover is particularly vulnerable to runoff and erosion (Table 1). Areas with steep slopes on which row crops such as corn and cotton are grown continuously may also be of concern. The dense surface cover resulting when row crops are planted in rotation with grasses and legumes may substantially reduce erosion potential.

Table 1 Mean annual runoff and erosion under different land uses with mean annual rainfall of 1400 mm

Land use	*Runoff (mm)*	*Soil loss (mg ha^{-1})*
Rotation; coastal Bermuda grass and crimson clover after corn	70	5
Corn grown continuously	180	27
Cotton grown continuously	250	49
Fallow; cultivated with no vegetative cover	470	135

After Carreker JR, Wilkinson SR, Barnett AP, and Box JE (1978) *Soil and Water Systems for Sloping Lands.* USDA, ARS-S-160.

On areas that receive sufficient precipitation, interseeding row crops with a legume can be an effective conservation measure. The legume provides a protective cover during the critical planting period. Following the cropping season, the legume may also serve as a supplemental nitrogen source.

Irrigation is used on some agricultural areas. Both irrigation and natural precipitation events may result in erosion. Because of the increased quantities of water introduced through irrigation, the potential for runoff may be greater on irrigated areas.

Rangelands and pastures In humid areas, the dense sod found in pastures reduces erosion. When adequate surface cover is maintained on rangelands and pastures, erosion is usually minimal. In regions with limited rainfall, severe erosion may still occur from the exposed areas between bunches of grasses during intense storms in regions with limited rainfall. Bare soils lack a protective surface cover and root mass and are subject to physical abrasion from livestock hooves. The reduction in vegetative cover caused by excessive grazing may increase the potential for both water and wind erosion.

As soils erode, they become less resistant to further degradation. Soil aggregation, water infiltration, and water retention are diminished as surface soil is eroded and subjected to trampling. The proportion of grazing-resistant species that have shorter and shallower roots often increases on eroded sites. Grazing management systems should be selected to provide a productive balance of plant species that maintain root mass.

The areas adjacent to stream banks often contain productive soils and have greater amounts of soil water because they receive runoff from upland regions. Because of the relatively large amounts of vegetation on these areas, they are prime sites for grazing. Preventing cattle from grazing next to stream banks has been found to result in a substantial reduction in stream-bank erosion and sediment delivery to streams.

Gravel and cobble materials are found throughout the soil profile on some rangeland areas. Since shallow overland flow cannot easily transport gravel and cobble materials, they often remain on the surface of eroded soils. Thus, a surface armoring process is established that reduces further soil loss from the eroded rangeland sites.

Forest areas The vegetative cover continuously maintained on the surface of undisturbed forests substantially reduces soil loss. On an undisturbed forest soil, most erosion occurs from channel banks and adjacent steep slopes. Road construction, timber harvesting, or fires may substantially increase erosion rates. The potential for erosion may be severe on forest areas when a fire destroys plant material and the litter layer. A condition called hydrophobicity occurs when soils repel water due to the intense heat from a fire. Erosion potential is influenced by the intensity of the fire and the extent of the burned area within a watershed.

Construction sites Improper construction practices can accelerate soil erosion. On-site damage in urban areas may result from the loss of topsoil by heavy equipment during construction activities or by erosion. If disturbed sites are not properly protected, runoff and erosion can adversely affect the surrounding environment. Although the area that is disturbed may be relatively small, the amount of soil eroded from construction sites can be substantial. Preventing sediment from leaving a construction site can reduce serious off-site impacts.

Erosion Control Measures

A variety of measures are available for controlling erosion on agricultural areas including contouring, strip cropping, conservation tillage, terraces, buffer strips, grassed waterways, and the use of polyacrylamide on irrigated areas. Specialized control measures are also available for reducing erosion from forest areas, construction sites, and stream channels. Depending upon the severity of the problem, it may be necessary to use a combination of control measures to reduce erosion to reasonable limits.

Erosion Control Measures on Agricultural Areas

Contouring Performing tillage and planting crops along the contour of the land can be an effective conservation measure. Rill development is reduced when surface runoff is impounded in small depressions. Contour farming not only minimizes erosion but also reduces runoff by storing rainfall behind ridges. The storage capacity of furrows is significantly increased with ridge tillage systems. Row crops are planted on top of the same furrow each year to maintain furrow storage capacity.

The effectiveness of ridges in trapping runoff and reducing soil loss decreases as slope gradient becomes greater. When contouring is performed on steep slopes, furrow storage capacity may be exceeded during high-intensity rainfall events. Water previously stored in the furrow is then released. As the volume of water increases with each succeeding row, small gullies may form. To prevent runoff from large precipitation events from overtopping ridges, a small slope gradient along the row is desirable. Field boundaries should be located on the contour or moved to eliminate odd-shaped fields with short rows.

Strip cropping Under strip-cropping conditions, alternate parcels of different crops are grown on the same field. The strips with the greatest surface vegetative cover capture soil eroded from upslope areas. Strip widths are dictated by farm implement requirements. To improve erosion control, the strips are usually planted on the contour in a rotation that shifts crops annually from one strip to the next. The most effective strip-cropping rotations include perennial grasses and legumes that alternate with grain and row crops. In arid and semiarid regions, strips may be placed perpendicular to the prevailing wind direction for wind erosion control.

Conservation tillage On cropland areas, erosion potential is substantially reduced when residue mulch from the previous crop is left on the soil surface (Figure 1). Residue mulch serves to protect the soil surface by adsorbing and dissipating raindrop energy. The percentage of residue cover maintained on the soil surface influences soil loss potential. Substantial reductions in erosion can result from small amounts of residue cover. Any tillage or planting system that leaves at least 30% of the soil surface covered with residue after planting has been defined as conservation tillage. It can be seen from Figure 1 that a 30% cover of wheat or corn residue can reduce soil loss by approximately 62% and 97%, respectively. The type of residue material influences the amount of erosion protection provided for a given surface cover. During runoff events, small impoundments may form above residue elements. The impoundments found above the larger-diameter residue elements such as corn have greater volumes and are therefore more effective in trapping sediment and reducing soil loss.

Excessive tillage can destroy soil structure, resulting in surface sealing and decreased infiltration. With the increased availability of herbicides, the use of tillage for weed control has diminished. When tillage is performed, existing crop residues are maintained by using special implements that cause only minimal disturbance to the soil surface. To maintain sufficient residue cover to control erosion, no-tillage is used before planting for some row crops such as soybeans.

Terraces On steep land, terraces or broad channels are built perpendicular to the slope to reduce rill erosion by decreasing overland flow length. Sediment settles from overland flow as runoff travels at relatively low velocities along the gentle grades used

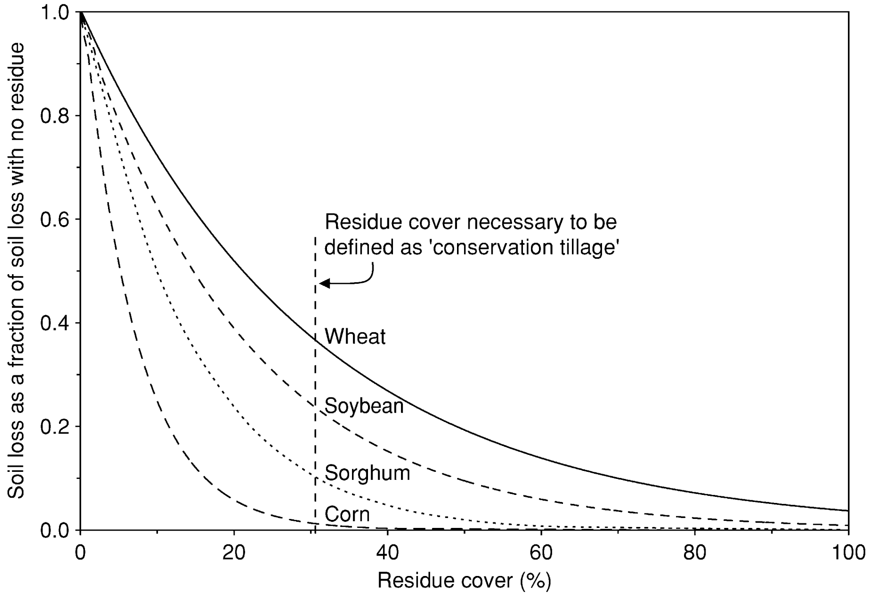


Figure 1 Ratio of soil loss for a given residue cover to soil loss with no cover. After Colvin TS and Gilley JE (1987) Crop residue–soil erosion combatant. *Crops and Soils* 39(7): 7–9, with permission.

in terraces. Therefore, the quality of surface water leaving terraced fields is improved. Gully formation is prevented since the terraces usually empty on to grassed waterways or into underground pipes. Since terraces retain runoff they also increase the amount of soil water available for crop production.

Important considerations in terrace design include soil characteristics, cropping and management practices, and climatic conditions. Contouring is included as a conservation practice on terraced fields since the crop rows are usually planted parallel to the terrace channel. Since they are expensive to construct, cause some inconvenience to the farming operation, and require periodic maintenance, terraces should only be used when other erosion-control measures do not provide adequate protection.

Buffer strips Buffer strips are designed to intercept runoff using permanent vegetation. Other erosion-control practices are usually employed in association with buffer strips. As an integral part of a planned conservation system, buffer strips may be located at a variety of locations within a landscape. To maintain buffer-strip performance, periodic maintenance is required. Contour buffer strips, filter strips, and grassed waterways are frequently used types of buffer strips.

Perennial grasses are usually planted along steep slopes within contour buffer strips. Sediment is removed as overland flow approaches and enters the grass strips. Site-specific conditions dictate the types of vegetation and spacing of contour buffer strips. As a result of sedimentation, a narrow terrace may form along the upslope portion of the grass strip. Contour buffer strips are much less expensive to establish than terraces.

Filter strips provide increased infiltration and remove sediment from overland flow. However, they do not interfere significantly with normal farming operations since they are located at the edge of fields or adjacent to streams, ponds, or wetlands. Areas with gentle slopes where rilling is not a problem are the best locations for filter strips.

Grassed waterways Runoff from terraces or other concentrated flow areas can be conveyed using grassed waterways, thus preventing channel erosion and gully formation. Costly downstream sedimentation is reduced because the sediment transported by overland flow is deposited in the grassed waterways. A stable outlet located below the grassed waterway serves to disperse the flow before it enters a vegetative filter.

Grassed waterways reduce peak runoff rates and provide a stable channel that can easily handle the flow that remains. Channel stabilization is provided by modifying the cross-section and slope of the waterway to limit flow velocity and by establishing vegetative protection. The types of vegetation used in the channel are dictated by local soil and climatic conditions. In addition to grassed waterways, permanent gully control structures may be needed on areas with relatively large runoff volumes or steep slopes.

To prevent failure, the waterway should not be used as a road, stock trail, or pasture, especially during wet conditions. Care should also be taken when farm machinery crosses the waterway. The waterway should be managed to stimulate new growth and control weeds, and an annual application of fertilizer is recommended.

Polyacrylamide Erosion on irrigated areas has been reduced by the use of polyacrylamide. Polyacrylamide is a long-chain synthetic polymer that serves as a strengthening agent to bind soil particles together. Larger, heavier particles are more difficult to detach and transport. Polyacrylamide is added to the irrigation flow at rates dictated by the irrigation system, soil type, and water source. Increased infiltration and reduced transport of pollutants have also been reported for soils on which polyacrylamide has been applied.

Erosion Control Measures on Forest Areas

Proper road and drainage design, and effective slope stabilization can reduce erosion potential during road construction on forest areas. Detailed planning of the timber-harvesting operation can minimize runoff and soil loss. Compaction of the soil surface occurs when logs are moved across the land surface by tractors or skidders. Approximately 25–35% of the harvested area is disturbed during the tractor-logging operation. In comparison, the highlead system where the ends of the logs are raised off the ground by a cable system, and the skyline system where the disturbed logs are entirely lifted off the ground reduce the disturbed harvested area to approximately 15% and 12%, respectively. The risk of erosion is reduced when damaged trees are felled to reduce the velocity of overland flow. Check dams can be established in drainages using straw bales. Straw can also be spread on burned areas to protect the soil and stabilize reseeded areas.

Erosion Control Measures on Construction Sites

The construction project should be conducted in phases so that only those sites under active development are exposed and those areas should be kept as small as possible. Management practices should be implemented to reduce the volume and velocity of runoff, and to retain sediment within the construction site. Small sediment basins, perimeter sediment fences, and straw bale or fabric check dams can be used to reduce off-site sediment transport. Runoff from adjacent areas should not be allowed to enter the disturbed site. Finally, a vegetative cover should be established as soon as possible to provide permanent protection for the construction site.

Erosion Control Measures in Stream Channels

Stream channel erosion can be reduced by the use of vegetation, mechanical, or a combination of vegetation and mechanical means. Grading of the stream bank to a less severe slope may be necessary, depending upon the size of the upstream drainage area. Grasses, shrubs, and trees can be successfully used to stabilize stream channels. Fast-flowing water can be diverted away from stream banks by dikes made of loose stone or rock piles placed within the stream channel. Materials such as a mechanical cover of stone or rocks may also serve as a protective cover. A mechanical cover can be employed to protect areas with the greatest erosion hazard such as the bottom of a stream, while vegetation is usually used to stabilize the upper portion of the stream banks.

See also: **Erosion:** Irrigation-Induced; Wind-Induced

Further Reading

Carreker JR, Wilkinson SR, Barnett AP, and Box JE (1978) *Soil and Water Systems for Sloping Lands*. USDA, ARS-S-160.

Haan CT, Barfield BJ, and Hayes JC (1994) *Design Hydrology and Sedimentology for Small Catchments*. San Diego, CA: Academic Press.

Hudson N (1995) *Soil Conservation*. Ames, IA: Iowa State University Press.

Laflen JM, Tian J, and Huang C (2000) *Soil Erosion and Dryland Farming*. Boca Raton, FL: CRC Press.

Lal R (1999) *Soil Quality and Soil Erosion*. Boca Raton, FL: CRC Press.

Lal R, Blum WH, Valentine C, and Stewart BA (1998) *Methods for Assessment of Soil Erosion*. Boca Raton, FL: CRC Press.

Napier TL, Napier SM, and Turdon J (2000) *Soil and Water Conservation Policies and Programs: Successes and Failures*. Boca Raton, FL: CRC Press.

Pierce FJ and Frye WW (1998) *Advances in Soil and Water Conservation*. Chelsa: Sleeping Bear Press.

Schwab GO, Fangmeier DD, Elliot WJ, and Frevert RK (1993) *Soil and Water Conservation Engineering*, 4th edn. Singapore: John Wiley.

Toy TJ, Foster GR, and Renard KG (2002) *Soil Erosion: Processes, Prediction, Measurement and Control*. New York: John Wiley.

Troeh FR, Hobbs JA, and Donahue RL (1999) *Soil and Water Conservation: Productivity and Environmental Protection*, 3rd edn. Upper Saddle River, NJ: Prentice-Hall.

Ward AD and Elliot WJ (1995) *Environmental Hydrology*. Boca Raton, FL: CRC Press.

Wind-Induced

T M Zobeck, USDA Agricultural Research Service, Lubbock, TX, USA
R S Van Pelt, USDA Agricultural Research Service, Big Spring, TX, USA

Introduction

Wind erosion is a dynamic physical process leading to soil degradation that occurs when strong winds blow on loose, dry, bare soils. Fine, fertile soil particles are often removed during wind erosion (wind-induced particle movement), reducing soil productivity and causing significant on-site and off-site problems. The widespread social and economic hardships that occurred in the USA during the disastrous Dust Bowl days of the 1930s were caused primarily by wind erosion on cropland. Much progress has been made in reducing the effects of wind erosion in the USA through soil conservation efforts made by individual land-owners with the technical assistance of the US Department of Agriculture, Natural Resource Conservation Service (USDA, NRCS). However, wind erosion continues as a national and international problem. Dust clouds may still be seen in many parts of the USA (Figure 1). Studies by the NRCS in 1997 show that wind causes about 44% of the 1.9 billion tons per year of soil lost from US cropland by water and wind erosion. In this article we explore the causes, effects, and control of wind erosion.

Wind Profile

Wind erosion is a process that results from the interaction of the wind and the soil surface. The *Encyclopedia of Climatology* defines wind as "a stream of air flowing relative to the earth's surface, usually more or less parallel to the ground." The Earth's surface exerts a drag on the wind, resulting in a vertical profile of wind speed that is described by a semilogarithmic equation:

$$u(z) = \frac{u^*}{k}\ln\left(\frac{z}{z_0}\right) \qquad [1]$$

where $u(z)$ (m s^{-1}) is the wind speed at height z (m); u^* (m s^{-1}) is the friction velocity, which is indicative of the amount of atmospheric turbulence (its value is independent of height for a given wind profile); k is the von Karman constant (0.4), a dimensionless number; and z_0 (m) is the aerodynamic roughness height, which is a measure of the roughness of the ground surface. From eqn [1], it is evident that there is a sharp decrease of mean horizontal wind speed as the surface is approached. This gradient in wind speed produces a shearing force as the result of surface roughness causing a drag on the airflow. The shearing force produces a tangential stress on the surface, called the shear stress, that is equal to the product of the square of the friction velocity and fluid density. The shear stress provides the interchange of momentum necessary for erosion to occur.

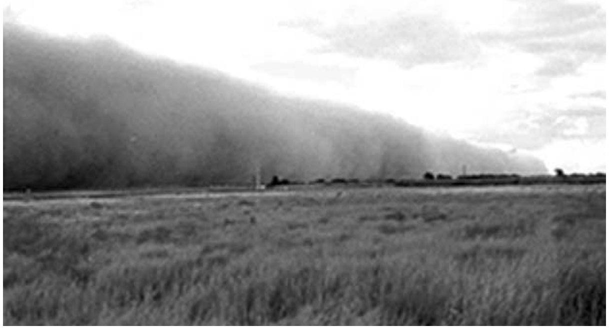


Figure 1 This dust storm was advancing in a thunderstorm outflow in west Texas in 1997. Reproduced with permission from Chen W and Fryrear DW (2002). Sedimentary characteristics of a haboob dust storm. *Atmospheric Research* 61: 75–85, Elsevier.

Modes of Transport

The minimum wind velocity initiating particle movement is known as the threshold velocity. Particle susceptibility to wind erosion is affected by size, density, and shape. The quartz sand particle diameter most susceptible to wind erosion is approximately 100 μm; this size particle begins to move with a wind velocity of approximately 14 m s^{-1} measured at a height of 2 m. Particles larger and smaller than this size require greater wind velocities to initiate movement from rest.

Wind-blown materials move in three modes: creep, saltation, and suspension. In general, the largest soil particles (1–2 mm) will roll or slide along the surface in the creep mode because they are too massive to leave the soil surface. Particles between 100 μm and 1 mm in size tend to move in saltation (bouncing) mode. The third mode of particulate transport, suspension, involves soil particles less than 100 μm in diameter. These materials may travel great distances before returning to Earth.

Fine materials carried in suspension are less susceptible to direct entrainment by the wind than are fine sand particles. Saltating particles return to the soil surface with a force that is a function of their mass and speed, resulting in the disruption of soil aggregates and surface crusts and the release of finer particles. As this fine particulate is produced, turbulent eddies carry it higher and higher into the air. In order for a particle to be lifted to any height, the upward velocity of the eddy must be greater than the terminal velocity of the particle. Sand-sized quartz grains have significantly higher terminal velocities than silt- and clay-sized particles and soil particulate organic matter, and are rarely lifted to heights greater than a few meters. While in most cases sand-size particles are redeposited in the source field or nearby, in extreme events fine sands may enter true suspension mode and be transported great distances.

As saltating particles are ejected from the surface, they initially rise at an angle of about 25–50° and the force of the wind accelerates them constantly until gravity returns them to the surface. The force these returning particles exerts on the soil surface often results in the ejection of several sand grains from the surface, so the number of grains in saltation often increases exponentially in a short distance in sandy soils. This phenomenon has been called the avalanching effect. There is a limit to the number of grains that can be entrained for any given wind speed, however, as the accelerating sand grains exert a drag on the passing wind that results in transfer of the wind energy into the energy of the sand particles and heat of friction. There is an equilibrium known as the transport capacity where as much saltating material is being deposited as is eroded. This is generally not true for particles moving in the suspension mode, as the amount of suspended particles continues to increase in the atmosphere with increasing fetch across the eroding surface.

Soil Surface Conditions

Although wind drives the wind-erosion process, the condition of the soil surface often controls whether or not a soil is transported by the wind. Any factor that reduces the impact of the wind on the soil surface reduces wind erosion.

The surface soil texture (the amount of sand, silt, and clay) is a primary factor affecting the erodibility of a site. Sandy soils tend to have very low stability and are easily moved by wind. The NRCS has related soil texture to wind erodibility, classifying the soil into wind erodibility groups according to soil texture and carbonate content (Table 1). Calcareous soils (soils containing enough calcium carbonate that they effervesce in the presence of dilute acid) tend to be more erodible than similar noncalcareous soils. The surface soil texture and carbonate content are difficult to change and are generally considered intrinsic or static soil properties. Other surface soil properties are temporal in nature and may change annually, seasonally, or even daily. Thus, the effects of these properties on wind erosion are temporal in nature and may be subject to management practices.

Table 1 Relation of soil texture and soil erodibility

Soil texture[a]	*Predominant soil texture class of surface layer*	*Wind erodibility group (WEG)*	*Soil erodibility index (I) (mg ha^{-1} year^{-1})*[b]
C	Very fine sand, fine sand, sand, or coarse sand	1	694[c] 560 493 403 358
C	Loamy very fine sand, loamy fine sand, loamy sand, loamy coarse sand, or sapric organic soil materials	2	300
C	Very fine sandy loam, fine sandy loam, sandy loam, or coarse sandy loam	3	193
F	Clay, silty clay, noncalcareous clay loam, or silty clay loam with more than 35% clay	4	193
M	Calcareous loam and silt loam or calcareous clay loam and silty clay loam	4L	193
M	Noncalcareous loam and silt loam with less than 20% clay, or sandy clay loam, sandy clay, and hemic organic soil materials	5	125
M	Noncalcareous loam and silt loam with more than 20% clay, or noncalcareous clay loam with less than 35% clay	6	108
M	Silt, noncalcareous silty clay loam with less than 35% clay, and fibric organic soil material	7	85
–	Soils not susceptible to wind erosion due to coarse surface fragments or wetness	8	–

[a] C, coarse; M, medium; F, fine.

[b] The soil erodibility index is based on the relationship of dry soil aggregates greater than 0.84 mm to potential soil erosion.

[c] The *I* factors for WEG 1 vary from 160 for coarse sands to 310 for very fine sands. For coarse sand with gravel, use a low figure. For very fine sand without gravel, use a higher value. When unsure, use an *I* value of 220 as an average figure.

Adapted from the USDA, NRCS (1999) *National Agronomy Manual*. Title 190 Part 502. Wind Erosion, USDA Natural Resources Conservation Service.

Soil surface roughness, also called microrelief or microtopography, can modify the effect of the wind on the surface and physically protect part of the surface from abrasion caused by blowing particles. In tilled soils, the tillage tool creates an oriented roughness, or ridges, in the direction of tillage and random roughness due to the random orientation of soil aggregates and clods. The effects of tillage-induced roughness on wind erosion depends upon wind direction when ridges are present. Winds in the direction perpendicular to tillage are affected by both ridges and clods while wind in the direction parallel to tillage is only affected by the random roughness. The physical protection of the soil surface by ridges and clods is illustrated in Figure 2. More of the soil surface is protected from abrasion when the wind blows perpendicular to the ridges.

In agricultural fields, tillage not only creates surface roughness but also disrupts the soil and creates an unconsolidated surface layer of soil aggregates of various sizes. Some of the aggregates and clods are small enough to be eroded by wind. Aggregates of mineral soils <0.84 mm diameter are generally considered erodible. Larger particles of organic soils blow because they have much lower density than mineral soils due to the high content of organic material present. The estimated potential wind erosion is related to the amount of nonerodible aggregates, as indicated in Table 2. Practices that encourage the development of nonerodible aggregates reduce the amount of wind erosion in tilled fields.

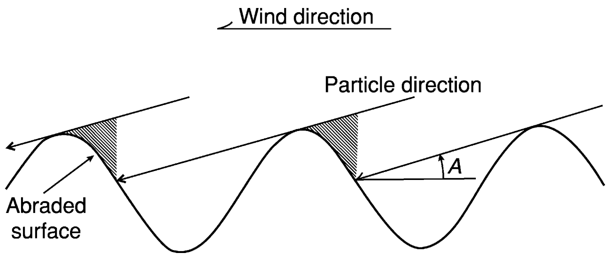


Figure 2 Schematic representation of a ridged field. Part of the surface is protected from abrasion. Adapted with permission from Zobeck TM (1991) Soil properties affecting wind erosion. *Journal of Soil and Water Conservation* 46: 112–118.

The ability of soil aggregates to resist breakdown due to abrasion caused by blowing sand grains or other applied forces is called aggregate stability. Fragile soils have low aggregate stability and are easily disrupted during wind erosion events, causing additional fine material to be added to the wind stream. Aggregate stability is quantified by measuring the amount of energy per unit mass needed to crush aggregates of a given size.

Rainfall often impacts the soil surface with enough force to disrupt aggregates and rearrange the particles on the surface of unconsolidated tilled soils to form a relatively thin consolidated layer, called a crust. The crust is more dense and resistant to abrading particles than the unconsolidated soil immediately below it. Very sandy deposits form very weak crusts that are easily destroyed by blowing sand particles. However, soils with even a small amount of clay and silt will bind together to form a layer that is 40–70 times as resistant to erosion as entirely erodible soil of the same texture.

In sandy soils, a thin layer of loose, highly erodible sand may form on the surface of a crust. This sand is highly susceptible to wind and starts blowing at relatively low wind speeds. The amount and distribution of this loose, erodible sand on the soil surface are related to the potential wind erosion of a site. If little

Table 2 Soil erodibility index (I) in units of mg ha^{-1} as determined by percentage of nonerodible soil

Aggregates >0.84 mm diameter (%)										
	Units (mg ha^{-1})									
Tens[a]	0	1	2	3	4	5	6	7	8	9
0	–	694	560	493	437	403	381	358	336	314
10	300	293	287	280	271	262	253	244	237	228
20	220	213	206	202	197	193	186	181	177	170
30	166	161	159	155	150	146	141	139	134	130
40	125	121	116	114	112	108	105	101	96	92
50	85	81	74	69	65	60	56	54	52	49
60	47	45	43	40	38	36	36	34	31	29
70	27	25	22	18	16	13	9	7	7	4
80	4	–	–	–	–	–	–	–	–	–

[a] The rows and columns represent the amount of nonerodible soil. For example, if nonerodible soil is 25%, first find the 20 in the Tens row and then the 5 in the Units column to find 193 mg ha^{-1}.

Adapted from the USDA, NRCS (1999) *National Agronomy Manual.* Title 190 Part 502. Wind Erosion, USDA Natural Resources Conservation Service.

or no loose, erodible sand is present, erosion generally does not occur.

Vegetation

Vegetation, whether growing or dead, standing or flat, reduces the susceptibility of the soil surface to erosion. Standing vegetative cover effectively raises the roughness height (z_0) and reduces the shear stress acting on the surface behind the barrier. Movement of the standing biomass absorbs some of the wind energy and eddies created in the wake of the standing biomass dissipate much of the rest. The effectiveness of standing biomass at reducing erosion increases with the height, density, and stiffness of the plants. Flat plant residue, while not greatly affecting the value of z_0, also serves to protect the soil surface from erosion. Dense flat residue and low-growing vegetation effectively form the surface against which the wind exerts its force, and thus protect the underlying soil surface. Where the flat residue is less dense, the residue acts much like random soil roughness elements. If saltation is initiated in a localized area of the field, the saltating grain may strike a relatively nonerodible plant part and settle to the ground, effectively breaking the chain reaction of saltation. Where adequate water from rainfall or irrigation is available, winter-cover crops are often planted as a preferred wind-erosion control technique.

Wind Erosion Modeling

Several wind erosion models have been used or proposed. The model most commonly used currently in the USA is the Wind Erosion Equation (WEQ), developed by the USDA. The WEQ was developed to predict long-term average annual soil loss from farm fields. The WEQ has been used by the USDA, NRCS to determine soil loss due to wind erosion for the US Natural Resource Inventory. The general form of the equation, $E = f(IKCLV)$, expresses soil loss (E) as functionally related to soil erodibility (I), ridge roughness (K), a climatic index (C), field length (L), and a vegetative cover factor (V). In practice, the I value is determined from a table similar to Tables 1 or 2 and a series of charts and nomographs are used to apply the other factors in the equation. A computer spreadsheet program is now often used to calculate soil loss using WEQ.

A modified version of the WEQ called the revised wind erosion equation (RWEQ) was recently released by USDA. The RWEQ uses some of the same concepts of the WEQ but incorporates new variables in the following equation: $Q = 109.8\ (WF \times EF \times SCF \times K' \times COG)$. The maximum soil loss determined by the field conditions, Q, is the product of several variables. The wind is described by a weather factor, WF, that accounts for the wind velocity, soil wetness, snow, air density, and gravity. The erodible fraction of soil (EF) is estimated based on intrinsic soil properties. The surface crust factor, SCF, reduces erosion based on clay and organic matter content. Soil roughness, K', is related to oriented and random soil roughness. And a combined crop factor, COG, accounts for the effects of crop canopy and flat and standing residue. RWEQ estimates soil loss due primarily to saltation, with no direct estimate of suspended load. In practice, soil and site factors and crop management are entered into a computer program that computes erosion for periods defined by the user.

A wind erosion model called the Wind Erosion Stochastic Simulator (WESS) is a single-event wind erosion submodel of the Environmental Policy Integrated Climate (EPIC) model of USDA. This model is similar to RWEQ and uses the following basic equation:

$$YW = (FI \times FR \times FV \times FD)\int_0^t \frac{YWR}{WL}\,\mathrm{d}t \qquad [2]$$

where YW is the wind erosion estimate, FI is a soil erodibility factor, FR is surface roughness, FV is vegetative cover, FD is a field length factor based upon the unsheltered field length WL, and YWR is the erosion rate based upon wind energy and soil properties.

A much more detailed Wind Erosion Prediction System (WEPS) is under development by USDA to make daily or shorter-time-interval estimates of wind erosion. WEPS is a modular computer-based prediction system that also requires significant soils, site, and crop management input. The model uses a series of complex submodels that account for the effects of weather, crop growth and decomposition, soil surface properties, soil hydrology, tillage, and erosion. In general, the model simulates the weather, grows a specified crop and determines when erosion occurs. The output includes estimates of saltation and suspended sediment.

Recent concern in global climate change has created interest in using models to estimate atmospheric dusts, produced primarily by deserts or semiarid ecosystems. Advances in understanding of atmospheric transport of soil-derived dusts have been made by laboratories in France and Australia. Scientists at the Laboratoire Interuniversitaire des Systèmes Atmosphériques, Creteil, France have developed a model considering the size distribution of erodible particles and surface roughness as controls on dust emissions. Scientists with Commonwealth Scientific and Industrial Research Organization (CSIRO) in Australia have developed a model accounting for the effects of soil, climate, and vegetation to estimate sand drift intensity and dust emission. The model has been coupled with a geographic information system to estimate dust emission over a large area of Australia.

Wind Erosion Sampling

Measurement of wind erosion is necessary to determine the effects of control measures, validate models, and assess the intensity of erosive forces. A wide variety of devices has been developed for this purpose. The ideal sampler has a high sampling efficiency by collecting sediment without disturbing the wind stream or developing suction or back-pressure. Since about 50% of the blowing sediment is transported within 2 cm of the soil's surface, sampling close to the surface is desirable to make accurate estimates of transported material.

A simple method to measure creep material is to bury a jar level with the soil surface and measure the sediment that rolls into it. The center portions should be covered to exclude saltation material. A combination creep/saltation sampler has been developed by the USDA to sample very near the soil surface at heights of 0–3, 3–9, and 9–20 mm (Figure 3). The sampler rotates into the wind and each inlet has a separate collection pan. This sampler is used in combination with other samplers to analyze the entire saltation zone.

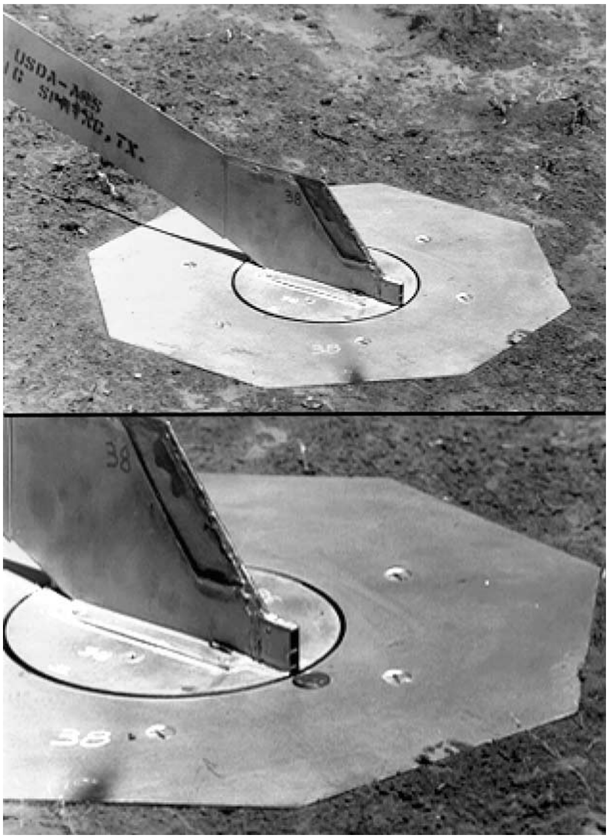


Figure 3 Creep/saltation sampler. Reproduced with permission from Zobeck TM (2002) Field measurement of erosion by wind. *Encyclopedia of Soil Science*, Marcel Dekker: New York.

Several types of samplers have been developed to sample the saltation zone, which is about 1–2 m above the soil surface. Big Spring Number Eight (BSNE) samplers are wedge-shaped samplers with 60-mesh screens on the top or side to relieve back-pressure and 2-cm-wide and 5-cm-high inlets (Figure 4). Several BSNE samplers are mounted on a pole to sample several heights. The samplers have wind vanes to ensure they are oriented into the wind. A similar sampling scheme (Wilson-Cooke-type dust samplers) employs bottles mounted on a pole with glass tubes as inlets (Figure 5).

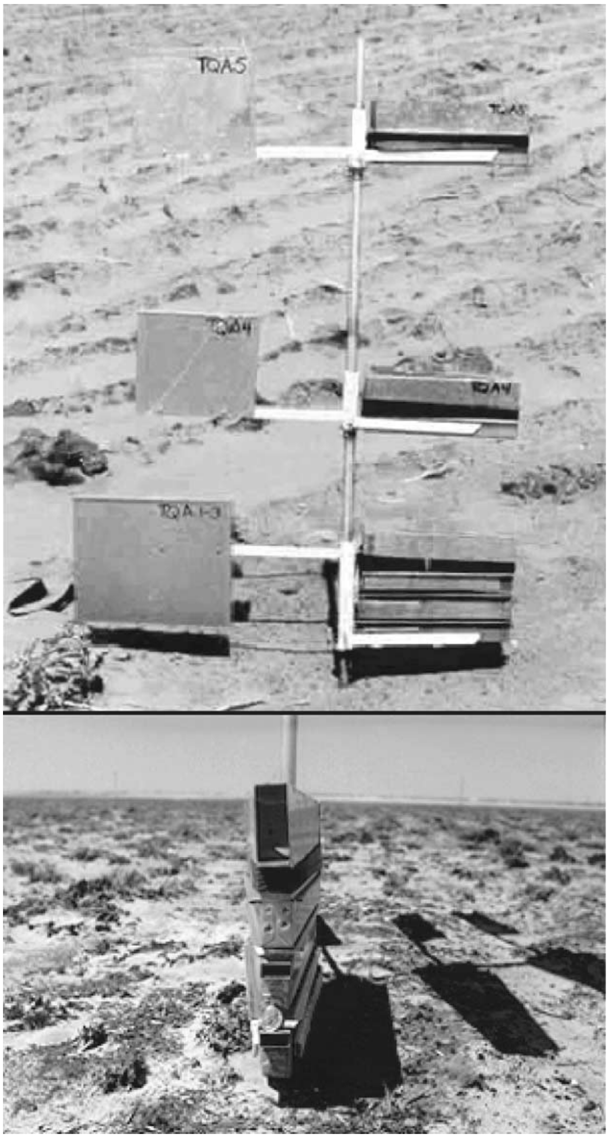


Figure 4 BSNE dust sampler. Reproduced with permission from Zobeck TM (2002) Field measurement of erosion by wind. *Encyclopedia of Soil Science*, Marcel Dekker: New York.

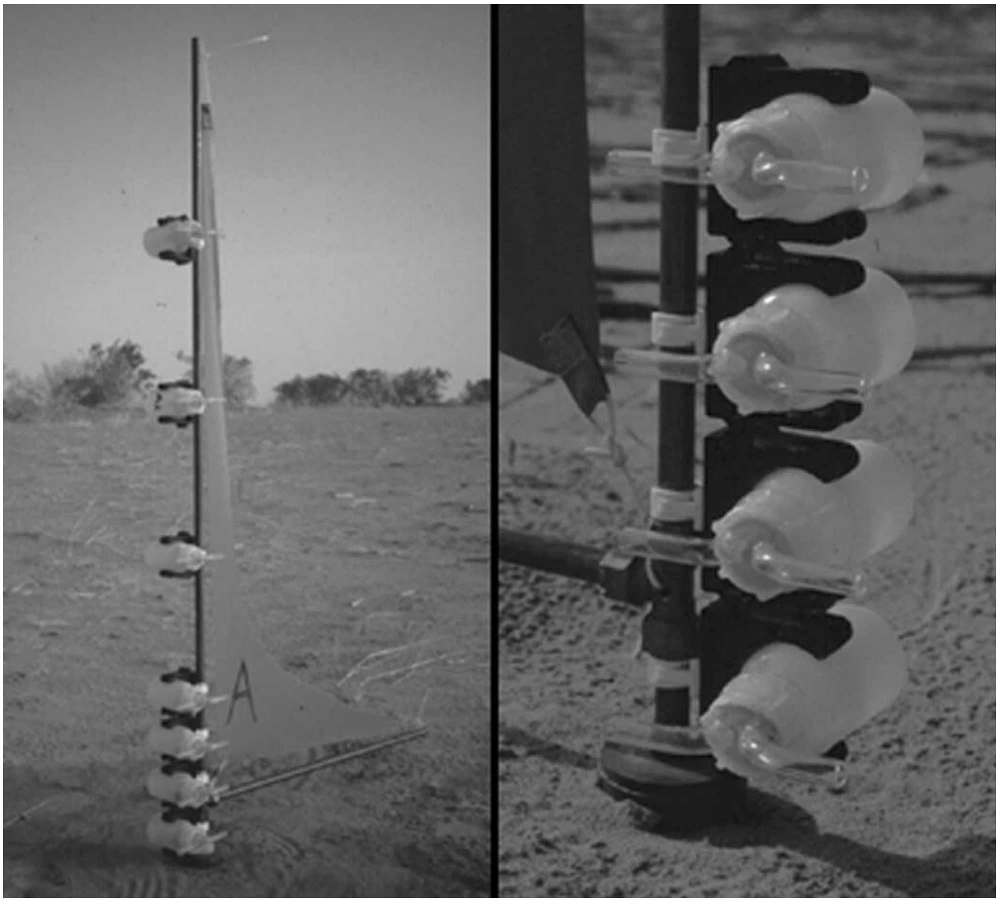


Figure 5 Wilson-Cooke-type dust sampler. Reproduced with permission from Zobeck TM (2002) Field measurement of erosion by wind. *Encyclopedia of Soil Science*, Marcel Dekker: New York.

Use of Anthropogenic Radioisotope Tracers

While wind-erosion measurement devices and samplers give good estimates of wind erosion for a given event, they are rarely employed for periods exceeding weeks or months. Wind erosion may occur at almost any time of the year and varies in intensity from year to year. From the time a field is put into cultivation, it is susceptible to erosion, and the effects of wind erosion are cumulative. Until the last 40 years, there has been no reliable way to estimate long-term wind erosion rates. Atmospheric testing of nuclear weapons in the 1950s and 1960s resulted in the release of radioactive isotopes that were distributed worldwide and were deposited, primarily by rainfall, on the soil surface. Many of these isotopes, including ^{90}strontium, ^{137}cesium, and $^{239/240}$plutonium, are strongly adsorbed to the surfaces of soil clays and organic matter. Strong pulses in 1959 and 1963–1964 created a very sharp rise in soilborne radioisotope tracers. Transport of these soilborne tracers is almost exclusively by actual physical movement of the particles to which they are adsorbed.

Radioisotope tracers, particularly ^{137}cesium, have been used for the last four decades to study soil redistribution from water erosion and for the last two decades to study soil redistribution from wind erosion. The greatest advantage to the use of radioisotope tracers is the ability to estimate long-term (since 1963) rates of wind erosion. Radioisotope activity of soil cores sampled in an undisturbed reference site such as native grassland is compared to the activity of soil cores sampled in a field or site for which the long-term estimate is desired. The loss or gain of radioisotope activity relative to the reference site has been shown to be proportional to the erosion or deposition of soil at the site of interest.

Effects of Wind Erosion

The on-site effects of wind erosion are generally considered to be those in or very near the eroding field. The on-site effects of wind erosion may be further divided into long-term cumulative effects and immediate effects. The effects of wind erosion on the soil are cumulative and, where wind erosion is a problem, soil losses often exceed the rates of soil formation dramatically. In extreme situations, wind may erode the entire topsoil horizon and reveal the underlying, less fertile horizons. Wind erosion is a winnowing process that, while it moves all the surface soil in a field, tends to remove the finer, more fertile parts of the soil and transport these materials away from the field. The result is a soil with a lower percentage of silt and clay, lower cation exchange capacity, lower organic matter, and reduced water-holding capacity. All of these factors serve to reduce the productivity of the soil and to diminish crop yields.

Immediate on-site effects of wind erosion also result in economic losses to agriculture. During windstorms, sandblast injury from saltating sand grains striking and abrading plants damages crop stands, especially seedling stands. Depending on the intensity of the event, seedling stands may be retarded or stunted in later growth and development, damaged to the point that replanting is warranted, or completely destroyed. As the seedlings start to produce true leaves, and especially during exponential growth phases, tissue conversion to woody matter reduces the susceptibility of the plant to damage and the increase in vegetative ground cover tends to reduce wind erosion frequency and intensity. In general, small grains are less susceptible to sandblast injury than corn, soybeans, and sunflowers, which are more tolerant than cotton, cabbage-related crops, and legumes. The least sandblast-injury-tolerant crops are garden vegetables and flowers.

Studies by the USDA, Economic Research Service suggest that the off-site impacts from wind erosion may be much larger than the on-site impacts. The estimated off-site costs of wind erosion in the western USA range from $4 to $12 billion per year. Off-site impacts from wind erosion are caused primarily by the release of fugitive dust, which may travel long distances and which constitutes most of the airborne dust that settles in homes and on automobiles, imposing costs for cleaning, reduced recreational opportunities, and impaired health. Estimates of the annual flux of airborne dust deposition over land areas range from less than 10 to about 200 t km^{-2} year^{-1}. As expected, these estimates vary with respect to distance and direction from the major source regions. The source areas of suspended dust are primarily deserts, agricultural operations in arid and semiarid environments, unpaved roads, dry lake beds, braided streams, deltas, and glacial outwash. These areas can be located on any mid-latitude continent, and the literature is composed of studies of dust source regions on all continents.

Fugitive dust obscures visibility and negatively affects air quality. Reduced visibility results in the cancellation of flights and numerous automotive accidents annually. Reduced visibility also depresses tourism and recreation revenues. Fugitive dust fouls machinery and adversely impacts environmental and human health. Pathogens, both plant and animal, may be transported thousands of kilometers, even across oceans, on dust particles.

Wind Erosion Control

Any treatment that reduces the wind at the soil surface to below the threshold of particle movement will control wind erosion. A wide variety of practices have been developed.

Perhaps the best control is to establish and maintain adequate ground cover (Figure 6). The ground cover may be growing vegetation of some type of mulch material such as cotton-gin trash. The effect of such cover is shown in Figure 7.

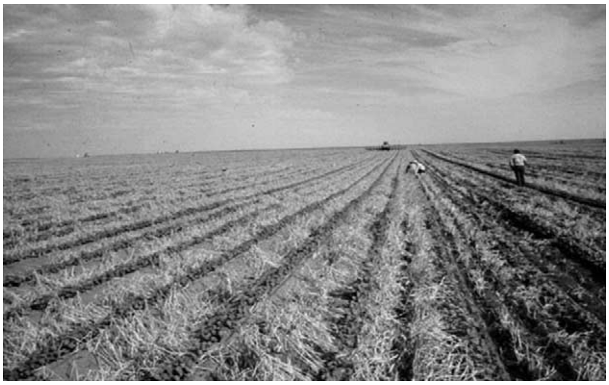


Figure 6 This wheat cover crop protects the cotton from blowing dust in Texas. Courtesy of USDA Natural Resources Conservation Service.

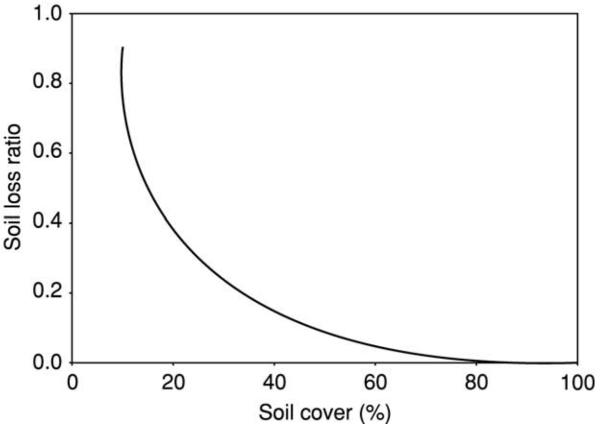


Figure 7 The effect of cover on soil loss. Reproduced with permission from Fryrear DW (1985) Soil cover and wind erosion. *Transactions of the American Society of Agricultural Engineers* 28(3): 781–784.

Since wind erosion tends to increase with field distance due to the avalanching effect, a reduction in field length is often accomplished using barriers. The barriers may consist of elaborate windbreaks called shelter belts that have several rows of vegetation, including trees and shrubs. The use of trees planted in shelter belts to control wind erosion is a technique that dates back to the US Dust Bowl days of the 1930s. Shelter belts were often planted along fence lines and around farmsteads to reduce the wind velocity and filter some of the suspended sediment from the air. In general, shelter belts reduce the wind velocity for a distance of 10 times their height in the downwind direction and three times their height in the upwind direction.

More recently, annual crops have been used as more closely spaced shelter belts within a field. Spacing of these 'wind strips' varies from two rows of wheat planted in the furrows between rows to 5–10-m-wide strips planted at 100–300-m intervals throughout the field (Figure 8). In emergency situations even physical barriers such as a snow fence may be used. The recommended porosity of barriers is about 40% for best results.

In some situations, it is not possible to maintain a vegetative cover on soil surface. In these circumstances, tillage may be used to maintain the soil in a cloddy condition (Figure 9). Care must be used to ensure tillage is done when the soil is moist enough to create clods. Tilling dry soils may exacerbate the problem by further pulverizing the soil and creating even more erodible particles.

Chemical treatments have also been used to control soil blowing. Chemical treatments tend to be expensive and are usually restricted to smaller or high-value areas. The chemical treatments bind the surface soil particles together. Any disturbance to the treated areas greatly reduces the effectiveness of the treatment.

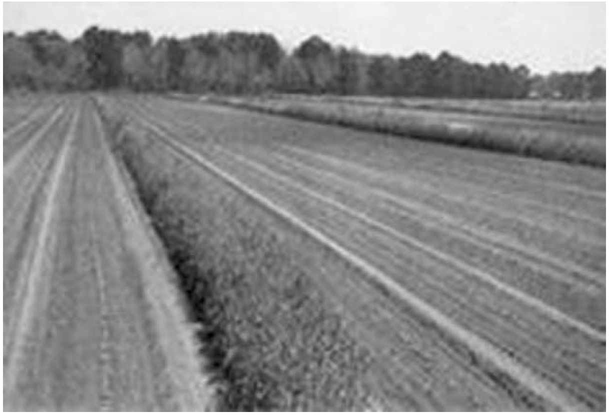


Figure 8 Perennial grass barriers protect cucumbers in South Carolina. Courtesy of USDA Natural Resources Conservation Service.

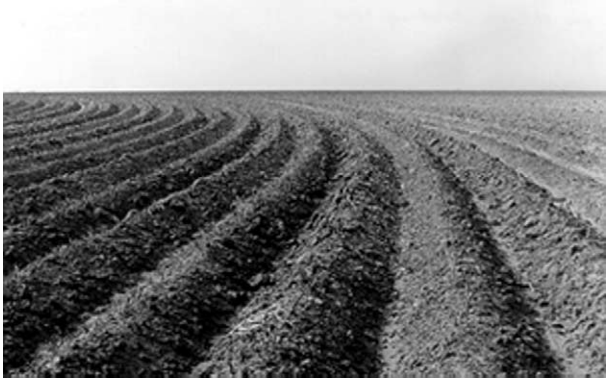


Figure 9 Tillage maintains this field in a rough cloddy condition. Courtesy of USDA Natural Resources Conservation Service.

List of Technical Nomenclature

u^*	Friction velocity (m s^{-1})
$u(z)$	Wind speed at height z (m s^{-1})
z	Height (m)
z_0	Aerodynamic roughness height (m)

Further Reading

Bagnold RA (1943) *The Physics of Blown Sand and Desert Dunes*. New York, NY: William Morrow.

Chepil WS and Woodruff NP (1963) The physics of wind erosion and its control. *Advances in Agronomy* 15: 211–301.

Fryrear DW (1990) Wind erosion: mechanics, prediction, and control. *Advances in Soil Science* 13: 187–199.

Fryrear DW and Bilbro JD (1998) Mechanics, modeling, and controlling soil erosion by wind. In: Pierce FJ and Frye WW (eds) *Advances in Soil and Water Conservation*, pp. 39–49. Chelsea, MI: Ann Arbor Press.

Hagen LJ (1991) A wind erosion prediction system to meet users' needs. *Journal of Soil and Water Conservation* 46: 106–111.

Marticorena B and Bergametti G (1995) Modeling the atmospheric dust cycle: I. Design of a soil-derived dust emission scheme. *Journal of Geophysical Research* 100: 16 415–16 430.

Shao Y (2000) Physics and modelling of wind erosion. In: *Atmospheric and Oceanographic Science Library,* vol. 23. London: Kluwer Academic.

Skidmore EL (1994) Wind erosion. In: Lal R (ed.) *Soil Erosion Research Methods*, 2nd edn, pp. 265–293. Delray Beach, FL: St. Lucie Press.

USDA, NRCS (1999) *National Agronomy Manual*. Title 190 Part 502. Wind Erosion, USDA Natural Resources

Conservation Service. Available online at:www.weru.ksu.edu/nrcs.

Wiggs GFS (1997) Sediment mobilisation by the wind. In: Thomas DSG (ed.) *Arid Zone Geomorphology: Process, Form and Change in Drylands*, 2nd edn. Chichester: John Wiley.

Woodruff NP and Siddoway FH (1965) A wind erosion equation. *Soil Science Society of America Proceedings* 29: 602–608.

Zobeck TM (1991) Soil properties affecting wind erosion. *Journal of Soil and Water Conservation* 46: 112–118.

ESSENTIAL ELEMENTS

E A Kirkby, University of Leeds, Leeds, UK

Introduction

Studies over many years of cultivating plants in water and sand cultures supplied or lacking in particular mineral nutrients have established a list of elements that are essential for the growth of higher plants. The essential elements that are required in large amounts were established in the nineteenth century and most of the trace elements, those elements usually taken up and required in only small amounts, mostly in the first half of the twentieth century. Techniques for avoiding contamination have improved and the list has been extended to include Ni for higher plants and, since 1970, As, Cr, Co, F, I, Pb, Li, and Se for animals.

The list of essential elements for higher plants is based on three criteria of essentiality. These are, that a lack of the element makes it impossible for the plant to complete its life cycle, a lack of the element gives rise to specific deficiency symptoms, and that the element plays a specific role in the nutrition and metabolism of the plant. From this approach the following are now defined as essential elements:

Carbon
Hydrogen
Oxygen
Nitrogen
Sulfur
Phosphorus
Potassium
Calcium
Magnesium
Iron
Manganese
Zinc
Copper
Boron
Molybdenum
Nickel

Sodium and silicon have also been shown to be essential for some species of higher plants.

When uptake is restricted either by a lack of the nutrient or because it is in an unavailable form in the soil, plant growth is depressed and nutrient deficiency symptoms can ensue which are typical for the lacking nutrient.

Classification of Essential Elements

Essential elements for both plants and animals are often considered in terms of macro- and micronutrients (or 'trace elements'), the macronutrients, C, H, O, N, P, S, K, Ca, Mg, being present in much higher concentrations in plant tissues than the micronutrients, Fe Mn, Cu, Zn, Mo, B, Cl, Ni. This approach provides no indication of biochemical behavior nor physiological function. One such attempt at physiological classification is shown in Table 1.

The elements in the first group include N and S, which in reduced form are covalently bonded constituents of organic matter. Phosphorus, B, and Si constitute another group and show similarity in biochemical behavior in that they are absorbed as inorganic anions or acids and in the plant are bound by hydroxyl groups of sugars forming phosphate, borate, and silicate esters. The plant nutrient cations and anions K, Na, Ca, Mg, Mn, and Cl make up a third group and are taken up in ionic form; and in the plant they remain as such or are adsorbed to indiffusible anions. These nutrients have osmotic- and ion-balance roles as well as more specific functions in enzyme catalysis. The fourth group of elements, including Fe, Cu, Mo, and Zn, are present in plants as structural chelates or metalloproteins; and the first three participate in redox reactions. Mn in its role in photosystem II can also be added to this list.

Nutrient Deficiency Symptoms

Observed visual symptoms of chlorosis and necrosis in nutrient-deficient plants reflect impairment of nutrient function and the degree of nutrient mobility within the plant (i.e., old leaves first showing symptoms indicate a mobile nutrient; young leaves presenting symptoms, indicate an immobile nutrient). A systematic approach to visual diagnosis can therefore be drawn up as shown in Table 2. It must be borne in mind, however, that symptoms only become clearly visible when deficiency is acute and the growth rate distinctly depressed. Additionally much natural vegetation adapted to nutrient-poor sites adjusts the growth rate to the most limiting nutrient so that visible nutrient deficiencies do not appear. Diagnoses can be especially difficult when more than one nutrient is involved either in deficient or toxic levels. Visual symptoms provide only a guide to nutrient status which must be supported by additional information, including soil and plant chemical analysis.

Nitrogen

Of all the plant nutrients, nitrogen is the most commonly deficient in soils. The two main forms of uptake are nitrate and ammonium ions, and maximum growth is favored by a combination of both. Ammonium N is assimilated in the root, whereas nitrate is mobile in the xylem and can be stored in the vacuoles of roots, shoots, and storage organs. Reduction of nitrate to ammonium via nitrate reductase occurs both in roots and shoots, and there are

Table 1 Classification of plant nutrients

Nutrient element	*Uptake*	*Biochemical functions*
1st group; C, H, O, N, S	In the form of $CO_2, HCO_3^-, H_2O, O_2, NO_3^-, NH_4^+, N_2, SO_4^{2-}, SO_2$. The ions from the soil solution, the gases from the atmosphere.	Major constituents of organic material. Essential elements of atomic groups which are involved in enzymic processes. Assimilation by oxidation-reduction reactions.
2nd group; P, B, Si	In the form of phosphates, boric acid or borate, silic acid from the soil solution.	Esterification with native alcohol groups in plants. The phosphate esters are involved in energy transfer reactions.
3rd group; K, Na, Ca, Mg, Mn, Cl	In the form of ions from the soil solution.	Non-specific functions establishing osmotic potentials. More specific reaction in which the ion brings about optimum conformation of an enzyme protein (enzyme activation). Bridging of the reaction partners. Balancing anions. Controlling membrane permeability and electrochemical potentials.
4th group; Fe, Cu, Zn, Mo	In the form of ions or chelates from the soil solution.	Present predominantly in a chelated form incorporated in prosthetic groups. Enable electron transport by valency change.

Reproduced from Mengel K and Kirkby EA (2001) *Principles of Plant Nutrition*, 5th edn. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Table 2 Some principles of visual diagnosis of nutritional disorders

Plant part	*Prevailing symptom*		*Disorder*
			Deficiency
Old and mature leaf blades	Chlorosis	Uniform	N (S)
		Interveinal or blotched	Mg (Mn)
	Necrosis	Tip and marginal scorch	N
		Interveinal	Mg (Mn)
Young leaf blades and apex	Chlorosis	Uniform	Fe (S)
		Interveinal or blotched	Zn (Mn)
	Necrosis (chlorosis)		Ca, B, Cu
	Deformations		Mo (Zn, B)
			Toxicity
Old and mature leaf blades	Necrosis	Spots	Mn (B)
		Tip and marginal scorch	B, salt (spray injury)
	Chlorosis, necrosis		Nonspecific toxicity

Reproduced from Marschner H (1995) *Mineral Nutrition of Higher Plants*, 2nd edn. San Diego, CA: Academic Press.

striking differences in distribution of activity between species. With increasing intensity of nitrate supply, the shoot generally plays a greater role in the reduction. The NH_3 formed cannot be reoxidized and provides the source for the synthesis of the low molecular weight N compounds, including amino acids, which are further metabolized to high molecular weight cell constituents containing N.

Nitrogen makes up about 20–40 $mg\,g^{-1}$ of the dry weight of the plant, most of which is present in reduced form as organic substances. These include low molecular weight compounds, including amino acids, amides, peptides, amines, and ureides, which are involved in intermediate metabolism and transport of N within the plant. Additionally N is a constituent of proteins, nucleic acids, and other compounds of high molecular weight such as phytohormones. As an enzyme constituent, N is involved in all reactions taking place in the cell, including energy metabolism, and, in the form of nucleic acids, is responsible for storage and transfer of genetic information.

Lack of nitrogen slows down the synthesis of protein and generally inhibits plant growth. Plants are small, with spindly stems, and senescence is enhanced, older leaves falling prematurely. Growth of roots is less depressed than that of shoots. Chloroplast development is disturbed and hence N-deficient leaves develop chlorosis, which is evenly distributed over the whole leaf. Since nitrogen is highly mobile within plants and is readily transported in reduced form from older to younger sites to meet demand for growth, older leaves first show the symptoms of chlorosis. The premature senescence associated with N deficiency particularly relates to a lack, in synthesis and translocation, of the N-containing phytohormones, cytokinins. When N is well supplied, not only is senescence delayed and growth stimulated, but shifts occur in the root-to-shoot ratio in favor of the shoot, all factors which appear to be related to induced changes in the phytohormone balance in favor of cytokinins.

Sulfur

The most important source of S to plants is sulfate taken up by the roots, although it can be absorbed through the leaves as sulfur dioxide. Uptake by the root appears to be sensitively controlled by shoot-to-root signals in the form of the tripeptide glutathione (GSH), the main long-distance transport form of reduced S. Sulphate is transported to the shoot and is reduced predominantly in the chloroplasts of mature leaves. Demand for S in the shoot is expressed by low concentrations of GSH in the phloem sap to the root, which favors the synthesis of SO_4^{2-} transporters and hence sulfate uptake.

Sulfur makes up between 1–5 $mg\,g^{-1}$ of the dry weight of plants. The first stable S-containing product in the reduction of sulfate is the amino acid cysteine, from which methionine can be formed, both amino acids being building blocks for protein. One of the main functions of S in proteins and polypeptides is in the formation of disulfide bonds between polypeptide chains. These so-called S-S bridges contribute to the conformation of enzyme proteins. Other essential S containing compounds include the vitamins thiamine and biotin, as well as coenzyme A, which is essential for respiration and the synthesis and breakdown of fatty acids in animals as well as plants.

Since S is an essential constituent of protein, S deficiency inhibits protein synthesis, and within the plant shoot there is an accumulation of non-S-containing amino acids as well as nitrate–N. Growth rate is reduced and frequently the plants are brittle and thin, with shoots more affected than roots. Chloroplasts are decomposed during S deficiency and leaves become uniformly chlorotic. Since S is not readily mobilized from older tissues, it is the younger leaves that usually first show the deficiency.

Sulfur deficiency in the field is becoming increasingly common. Sulfur-containing fertilizers are now less frequently applied to soils even though there is a greater demand of crops as a consequence of the higher application of other nutrients and the resulting greater annual offtakes of sulfur resulting from higher yields. Additionally since the early 1970s, the emissions of SO_2 have fallen dramatically so that many areas of Europe and North America are at risk from S deficiency.

Phosphorus

Phosphorus limitation of plant growth is widespread in soils. This is largely because of the insoluble forms in which phosphate is present in soil and the extremely low mobility of phosphate in soils in comparison with other plant nutrients. The uptake mechanism of phosphate is not usually a limiting step, since plant roots are capable of absorbing phosphate against a very steep concentration gradient from extremely low external P concentrations well below those of the soil solution. Supply is restricted because acquisition from the soil is dependent on the process of diffusion, which in many cases accounts for most phosphate supplied to plant roots. This contrasts to other nutrients such as nitrate, Ca, and Mg for which the supply to the root surface is largely by mass flow of nutrients in the soil solution driven by transpiration. For phosphate, this form of transport cannot meet plant demand, because the concentration in the soil solution is too low.

Under these limiting conditions of P supply, plants can adapt by physical and chemical modification to enhance phosphate acquisition. Root surface can be extended, as, for example, by root hair formation, to allow a more efficient exploitation of the soil. Mycorrhizal fungi associated with the root also fulfills a similar function, the mycelium of the fungus extending into the soil to enhance phosphate uptake, with the host root supplying the fungus with carbohydrates. Rhizosphere acidification either by proton or organic acid excretion is also a widespread response to P deficiency which can induce P release from soil minerals. One such example is citric acid, which is excreted by the proteoid roots of white lupin at high rates and amounts saturating confined root zones. The acid, which is a strong chelator of Ca, Fe, and Al, is thus able to mobilize phosphate from sparingly soluble compounds in both acid and calcareous soils.

Phosphorus generally accounts for about 1–5 $\mathrm{mg\,g^{-1}}$ of the dry matter of plants. Uptake by the roots appears to be regulated by a feedback signal of P-cycling in the phloem from shoot to root. Phosphate is at the center of metabolism. It is present in many sugar compounds involved in photosynthesis and respiration, and plays an essential role in energy metabolism as a component of ATP, ADP, AMP, and pyrophosphate (PPI). It is also a constituent of the phospholipids that occur in membranes and is an intergral component of the nucleotides RNA and DNA. In many seeds and fruits, phytate in the form of Ca and Mg inositol hexaphosphate acts as a phosphate reserve.

Plants showing P deficiency are typically stunted and often have a rigid, erect appearance. One of the earliest symptoms is a specific inhibition of leaf expansion and leaf surface area induced by restricted delivery of water. Older leaves often show a darkish green color by chlorophyll concentration and the stems may be a reddish color owing to enhanced formation of anthocyanins. Root growth is much less inhibited than shoot growth, and the roots act as a dominant sink for photosynthates and P from mature leaves. This response in favor of root growth at the expense of the shoot allows a greater exploitation of the soil for P as described above.

Potassium

Potassium is an essential element for all living organisms and is the most important nutrient cation not only in relation to its high concentration in plant tissues but also with respect to its physiological and biochemical functions. Unlike the elements so far discussed, K is taken up as a cation and remains in this form after uptake as the most abundant cation in the cytoplasm and vacuole, where it is required to neutralize organic acids and other anionic groups. Concentrations in plants range between 10–60 $\mathrm{mg\,g^{-1}}$ K dry weight, with younger tissues showing higher values. Potassium is highly mobile in plants, showing the highest cation concentration in both the xylem and phloem saps, and there is evidence that K transport from shoot to root in the phloem regulates K uptake by the root.

Potassium is closely related to meristematic growth associated with acidification and loosening of the cell wall. Cell extension results from the accumulation of potassium needed to stabilize the pH of the cytoplasm and increase the osmotic potential of the vacuoles. Potassium and some phytohormones react synergistically in this respect.

High concentrations of K are needed for the active conformation of many enzymes of intermediate metabolism and biosynthesis, including protein synthesis. Potassium also plays a most important role in regulating the water status of plants by lowering water potential. Similarly it also accumulates in guard cells in particularly high concentrations, thereby regulating the opening and closing of the stomata. Potassium also functions to promote photosynthesis and the translocation of photosynthates.

Plants can suffer from potassium deficiency without showing major symptoms, with only a reduction in growth, so called ‘hidden hunger.’ As the deficiency progresses, older leaves are first to show symptoms, as these supply the younger leaves with potassium. In many plant species, chlorosis and necrosis begin at the leaf margins and tips, where the localized lack of potassium has disturbed the water relations. In some species such as clover, however, irregularly distributed, white necrotic spots appear on the leaves. In field-grown crops, K deficiency can often be recognized by the decreased turgor of the leaves under water stress which appear flaccid. Lignification of vascular bundles can be impaired by K deficiency, thus weakening stems and making crops prone to lodging, i.e. flattening of the crop by wind and rain as a consequence of weakening of the culms.

Calcium

Higher plants contain appreciable amounts of Ca, usually in the range of 5–30 $\mathrm{mg\,g^{-1}}$ in the dry matter, but plant species and varieties differ greatly in their requirements. In order to obtain maximum growth, dicotyledons, particularly the legumes and herbaceous plants, have a much higher requirement than the cereals and other grasses. Calcium is taken up by plant roots in ionic form as the divalent ion and is depressed by competition from other mineral cations such as K, NH_4, and Mg. In comparison with K, the mechanism of uptake is not particularly efficient. The concentration of Ca in the soil solution is often relatively high, and the uptake of Ca and, to a large extent, also that of translocation within the plant as a whole is passive and reflects the flow of water from the bulk soil to the atmosphere driven by transpiration.

Uptake is limited to the apoplastic or free-space pathway mainly accessible in nonsuberized young roots, which is in accord with the restriction of Ca from the living parts of the cell. This relatively large, divalent cation is bound or is in an exchangeable form in the cell wall or at the exterior surface of the plasma membrane. Substantial quantities are also sequestered and stored in the vacuoles. The maintenance of an extremely low concentration of Ca in the cytoplasm to usually less than 1 μmol l^{-1} Ca is essential in order to prevent major disturbance to metabolism by precipitation of inorganic P, or competition with Mg at binding sites, thereby inhibiting the activities of essential enzymes. It is also a prerequisite for the role of Ca as a second messenger in regulating cellular functions in response to many stimuli in plants which are indicated by changes in cytosolic Ca concentration. This role of Ca in signaling in plants is currently the subject of much research.

Since Ca is greatly restricted from the living parts of cells, it is present in only very low concentrations in the phloem sap. Thus once Ca has entered a leaf it remains there and cannot be translocated to meet the requirements of younger leaves or fruits or storage organs. Transport of Ca from shoot to root is virtually absent and, in order to meet plant demand and supply via the xylem, there must be a continuous supply of Ca to plant roots from the soil. When uptake is restricted, the growth of meristematic tissue in young roots and leaves is rapidly disturbed. Membrane function is impaired and cells become leaky. Root growth stops and young leaves are twisted and necrotic. An undersupply of Ca can occur to the storage tissues of many fruits and vegetables. Ca-related disorders include ‘bitter pit,’ where the flesh and surface of the apple are covered in brown necrotic spots, and ‘blossom end rot’ in tomatoes, where there is a cellular breakdown at the distal end of the fruit.

Magnesium

Magnesium in some respects is similar to Ca in that it is taken up as a divalent ion and is present in cell walls in a bound or exchangeable form. On the other hand, it occurs in the cytoplasm in high concentrations, where it activates many enzyme systems. It also forms a complex with ATP which can then bridge with an enzyme through Mg and a N atom of the enzyme protein to activate the enzyme. Mg thus participates in all cellular actions involving ATP. Also, unlike Ca, it is highly mobile throughout the plant.

Magnesium is generally present in plants in concentrations from 1–5 mg g^{-1} of the dry weight and plays a most important role in photosynthesis. As well as being present at the center of the porphyrin structure of the chlorophyll molecule, it is additionally required in the synthesis of this molecule. However, in leaves only about 25% of Mg as a maximum is bound as chlorophyll. It is also needed in ionic form to activate ribulose bisphosphate carboxylase in the primary fixation reaction of CO_2 in the stroma of the chloroplasts of C_3 plants. Magnesium is also essential in the maintenance of structure and conformation of nucleic acids.

Since magnesium is highly mobile in plants, visible symptoms of deficiency occur first in the older leaves, usually as interveinal chlorosis. Deficiency decreases size, as well as disturbing structure and function of chloroplasts of the parenchymatous cells of the leaf. Magnesium deficiency in crops is often encountered on acid, well-leached soils, and there is increasing evidence of deficiency in forest ecosystems in Central Europe associated with air pollution and soil acidification. In animal nutrition, intensive grassland management can lead to low availability of Mg in the herbage and the occurrence of the acute disease hypomagnesemia or grass staggers in dairy cows.

Iron

In well aerated soil Fe is mostly oxidized and is relatively insoluble with the concentration of complexed ferric ion in soil solution an order of magnitude lower than that required for optimal plant growth. Two distinct strategies exist by which plants lacking in Fe are able to increase the solublity and absorption of Fe from the soil. The first, Strategy −1, is found in all plants except the grass family (*Poaceae*). Uptake is in the form of Fe^{2+} the production of which is dependent on the activity of the plasma membrane bound Fe (III) reductase of root cells to reduce Fe^{III} complexes from the soil. Plants utilizing this strategy respond to a lack of Fe by changes in the physiology and anatomy of the roots which increase the ability of the plant to acquire Fe. These responses include enhanced ferric reduction capacity at the root surface, stimulation of H^+ efflux from the roots, induction of transfer cells and increased formation of root hairs. By contrast, plants using Strategy −11, the *Poaceae*, respond to a lack of Fe by releasing Fe-chelating substances of the mugeneic acid family of phytosiderophores.

(MAs) from the roots. These phytosiderophores solubilize soil inorganic Fe^{III} compounds by chelation to form Fe^{III}–MAs complexes which supposedly traverse the plasma membrane of the root cells by a specific transport system. The transport of Fe from root to shoot in the xylem has been shown in a number of plant species to be mainly as Fe^{III} dicitrate.

Iron undergoes alternate oxidation and reduction between Fe^{2+} and Fe^{3+} and is present in numerous enzymes and proteins transferring electrons in the photosynthetic and respiratory chains. Chloroplasts provide the location for about 80% of Fe in leaves, and lack of Fe primarily affects photosynthetic activity. Protein and lipid synthesis are impaired, resulting in a disturbance in chloroplast structure and development, particularly of the thylakoid membranes. Moreover iron is also specifically required in at least two steps in the biosynthesis of chlorophyll.

In accordance with a major function of Fe in chloroplasts, iron-deficient plants are characterized by interveinal chlorosis of the leaves similar to that of Mg deficiency but occurring first in the younger leaves because of the restricted mobility of Fe within the plant. In some cases the leaves become completely white and devoid of chlorophyll and at a later stage may develop necrosis. Some plant species, including members of the rose family and various fruit trees, are particularly sensitive. The deficiency is especially common in high-pH calcareous soils and known as lime-induced chlorosis, where HCO_3^- ions may restrict Fe uptake.

Manganese

Manganese occurs in soil in three oxidation states in soils, as Mn^{2+}, as Mn^{3+}, and Mn^{4+} in the form of insoluble oxides and is mainly taken up as Mn^{2+} released from the higher-valency oxides under reducing conditions. In its biochemical function, Mn very much resembles Mg, and both of these ions activate a number of enzymes, e.g., in the TCA cycle. Both are also able to complex with ATP and bridge it with an enzyme complex (phosphokinases and phosphotransferases). The most well known role of Mn is in the photolysis of water mediated by a Mn-containing enzyme complex attached to photosystem II. The water-splitting reaction is cyclic, whereby O_2 is evolved with the concomitant reduction of Mn^{IV} to Mn^{III}, which in turn is reoxidized on the transfer of electrons to photosystem II (Hill reaction). As expected, a lack of Mn first depresses chloroplast function. Oxygen evolution is depressed, then, as deficiency becomes more severe, chlorophyll formation is decreased and the ultrastructure of the thylakoid membranes drastically impaired. Plant response to deficiency is species-dependent. Dicotyledons show interveinal chlorosis of the younger leaves, whereas, in cereals, greenish gray spots occur on the more basal leaves (gray speck in oats). Soils associated with Mn deficiency include acid soils derived from parent materials low in Mn, as well as high-pH soils containing free carbonates, especially when rich in organic matter.

Zinc

Zinc is taken up by plants as the divalent ion Zn^{2+}. Although only a few Zn-containing enzymes are known in plants, there are numerous enzymes activated by Zn which makes it difficult to interpret some of the complex changes that can occur in Zn-deficient plants. For example the disturbance of protein synthesis in Zn-deficient plants can be caused by a marked decrease in ribonucleic acid (RNA) as a consequence of a lower activity of the Zn-containing RNA polymerase, or by a decrease in structural integrity of the ribosomes or by enhanced RNA degradation. Membrane integrity can be impaired by Zn deficiency since, in healthy plants, Zn preferentially binds to SH groups of the membrane to stabilize structure; but additionally Zn is required in the synthesis of the fatty acids in membrane lipids. The Zn-containing isoenzyme of SOD (Cu-Zn superoxide dismutase) plays an important role in detoxification of the superoxide radical (O_2^-), protecting membrane lipids and proteins against oxidation. Under Zn deficiency, Cu-Zn-SOD is reduced and, particularly when light intensity is high, elevated levels of superoxide radicals are produced which destroy the membranes. This is the cause of 'sunscald,' a disorder associated with Zn deficiency which affects many fruit and vegetable crops, where enhanced lipid oxidation in the leaves leads to stunted growth, the destruction of chlorophyll, and the appearance of necrosis. The lower concentrations of the growth hormone indole acetic acid (IAA) in Zn-deficient plants is also probably related very closely to the well-known deficiency symptom of 'little leaf' and 'rosette' of apples, which describes the reduction in growth of young leaves and stem internodes. Deficiency occurs on high-pH soils and in crop plants symptoms of interveinal chlorosis of the leaf are common. In the monocotyledons, maize and sorghum chlorotic bands appear on either side of the midrib which later become necrotic.

Copper

Copper is similar to Fe in that it is present in redox enzymes and undergoes a valency change (Cu^+ and

Cu^{2+}) which allows electron transfer. It plays a role in photosynthesis as a constituent of plastocyanin, which accounts for about 50% of copper localized in the chloroplasts and enables photosynthetic electron transport. Also, in the form of the enzyme Cu-Zn-SOD, mainly located in the stroma of the choroplasts, Cu is directly involved in the detoxification of superoxide radicals which are generated during photosynthesis. Most of the other Cu-containing enzymes, for example, cytochrome oxidase, the terminal electron acceptor in the respiratory chain, react with oxygen to reduce it to H_2O_2. Other oxidases are phenolase and laccase, involved in the synthesis of quinones, melanins, and lignin.

Typical Cu-deficiency symptoms are especially expressed in the young shoot tissue because Cu is relatively immobile in plants. Chlorosis, white tip, necrosis, leaf distortion, and dieback are common. Enhanced tiller formation in cereals is a consequence of necrosis of apical meristems. Lignification is impaired, leading to lodging in cereals. The formation of vegetative tissue is less affected than the spectacular reduction in seed and fruit yields from Cu deficiency-induced male sterility.

Molybdenum

Molybdenum differs from the other micronutrients in that it is taken up from the soil as an anion (MoO_4^{2-}). Required in minute amounts, it is an essential component of two major enzymes in higher plants, nitrate reductase and nitrogenase, which is present in nodulated legumes and required in the N fixation process. The effective mechanism in both these enzymes depends on valency change of the Mo. Because of its role in NO_3 reductase, plants supplied with NH_4 have a lower requirement and deficiency symptoms are less severe or even absent than in plants supplied with NO_3-N.

Molybdenum deficiency, unlike that of all the other micronutrients, occurs on acid soils or soils high in iron oxides, which can fix Mo_4^{2-}. Also unlike the other micronutrients, deficiency symptoms are not confined to the younger leaves, since Mo is mobile within the plant. Deficiency can resemble that of N, with older leaves first showing chlorotic symptoms but with leaf margins becoming necrotic where NO_3^- accumulates. 'Whiptail' in cauliflower is one of the most well known symptoms, where the leaf lamina is not formed and only the leaf rib is present.

Boron

Boron is taken up as an undissociated boric acid. Uptake has long been held to be by passive following water flow, but recent evidence suggests an active component at low concentration of supply. Within the plant, B forms very stable complexes with organic compounds, including various sugars and their derivatives, uronic acid, and some *o*-diphenols all of which are abundant in cell walls.

Boron is the least well understood of all the plant nutrients and, in contrast to most other nutrients, is not the constituent of an enzyme. One of the most rapid responses to B deficiency occur in a matter of hours; when B is withheld there is a cessation in root growth. Cell wall development of apical meristems is disturbed, a response in accord with a function for B in the ultrastructural arrangement of the cell wall component in the apoplast (established in the late 1990s), the B-containing pectic polysaccharide B–rhamnogalacturonan II. This complex is composed of boric acid and two chains of pectic polysaccharides cross-linked through the borate diester bonding and forms a network of pectic polsaccharides, which is probably the ubiquitous form of B binding in the cell walls of higher plants.

The involvement of B in cell membrane function may also be inferred from the rapid recovery of metabolically linked ion transport following the addition of B to deficient roots. Additionally, membrane-bound ATPase activity, which is low in the B-deficient roots, is restored within 1 h of B application to the levels in the B-sufficient roots. Boron appears to stabilize the structure of the plasma membrane by complexing with membrane constituents such as glycoproteins and glycolipids, thereby keeping channels or enzymes at optimum conformation.

Because of the concentration of B at the cell wall–plasma membrane interface, primary effects of B deficiency on the morphology and physiology at this site are attracting increasing attention, including studies of accumulation of certain phenolics, inhibition of lignin synthesis, enhanced IAA activity, and decreased levels of diffusible IAA in response to B deficiency.

Boron deficiency is widespread. It is easily leached from soils and, as B availability decreases with increasing pH, it is common on alkaline and calcareous soils. In most plant species, B is immobile within the plant so that symptoms appear first at the apical growing points. However, in a number of species, including some fruit trees and vegetable crops, this is not the case, because B is phloem-mobile as it complexes with polyols, which are the primary photosynthetic product of these species. Boron is therefore more evenly distributed in these plants. An important role of B is in pollen germination and pollen tube growth. Both processes are highly sensitive to a lack of B as is also the viability of pollen grains. At low B supply therefore, seed, grain, and fruit production are very much more affected than is vegetative growth.

Chlorine

Chlorine is an unusual essential element as it is abundant in the environment and is taken up in high amounts by plants as Cl^-, whereas the concentrations required in its role as an essential element are minute. Indeed preventing contamination was particularly difficult in the first experiments to demonstrate the essentiality of Cl. In leaves Cl is accumulated in the chloroplasts and it is required as a cofactor to activate the Mn-containing water-splitting system in photosystem II. Chloride also functions in osmoregulation and together with K is often the main osmoticum in plant cells. At low concentrations, where Cl acts as a micronutrient, these osmoregulatory functions are restricted to specialized tissues and cells, including extension zones in roots and guard cells, where the Cl may be concentrated to much higher levels than in the bulk tissue.

Typical symptoms of deficiency include wilting of leaves, curling of leaflets, bronzing and chlorosis, and severe inhibition of root growth. Some plant species such as palm trees and kiwi have a high demand for Cl but it is questionable whether deficiency ever occurs in the field.

Nickel

Recently it has been recognized that Ni can be added to the list of essential elements. For about 25 years it has been known that Ni is an essential part of the enzyme urease which catalyzes the hydrolytic breakdown of urea to NH_4^+ and CO_2. Experiments with soybeans and cowpea have demonstrated that, regardless of the source of N-nutrition (urea, NH_4-N, NO_3-N, or N fixation), in the absence of Ni in the nutrient medium, urea accumulates in the leaves, producing severe symptoms of leaf-tip chlorosis. There is other supporting evidence that urea is a normal intermediate in nitrogen metabolism which, by the presence of Ni, is maintained at a low concentration, thereby preventing toxicity. Convincing evidence that Ni is essential for barley also comes from germination experiments using seeds obtained after growing plants for three generations deprived of Ni. The seeds were nonviable but viability could be restored by soaking in a Ni-containing solution.

See also: **Nutrient Availability**; **Nutrient Management**

Further Reading

Ae N, Arihara J, Okada K, and Srinnivasan A (eds) (2001) *Plant Nutrient Acquisition: New Perspectives*. Berlin, Germany: Springer-Verlag.

Bergmann W (1992) *Nutritional Disorders of Plants – Development, Visual and Analytical Diagnosis*. New York: Gustav Fischer-Verlag.

Brown PH, Welsh RM, and Carey EE (1987) Nickel: a micronutrient essential for higher plants. *Plant Physiology* 85: 801–803.

Clarkson DT and Hanson JB (1980) The mineral nutrition of higher plants. *Annual Review of Plant Physiology* 31: 239–298.

Epstein E (1999) Silicon. *Annual Review of Plant Physiology Plant Molecular Biology* 50: 641–664.

Fageria NK, Baligar VC, and Jones CA (1991) *Growth and Mineral Nutrition of Field Crops*. Marcel Dekker. Inc.: New York, Basel, Hong Kong.

Gerandas J and Sattelmacher B (1997) Significance of Ni supply for growth, urease activity and the contents of urea, amino acids and mineral nutrients of urea-grown plants. *Plant and Soil* 190: 153–162.

Marschner H (1995) *Mineral Nutrition of Higher Plants*, 2nd edn. San Diego, CA: Academic Press.

McDowell LR (1992) *Minerals in Human and Animal Nutrition*. San Diego, CA: Academic Press.

Mengel K and Kirkby EA (2001) *Principles of Plant Nutrition*, 5th edn. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Mordvedt JJ, Cox RF, Shuman LM, and Welch RM (eds) (1991) *Micronutrients in Agriculture*, 2nd edn. Madison, WI: Soil Science Society of America.

Xu G, Magen H, Tarchitzky J, and Kafkafi U (2000) Advances in chloride nutrition of plants. *Advances in Agronomy* 68: 97–150.

EUTROPHICATION

A J Gold, University of Rhode Island, Kingston, RI, USA
J T Sims, University of Delaware, Newark, DE, USA

Introduction

Eutrophication describes a cascade of processes that occur in aquatic and terrestrial ecosystems in response to an increase in nutrient inputs. Phosphorus (P) and nitrogen (N) are the nutrients that drive most eutrophication processes. In aquatic environments eutrophication can lead to excessive algal growth, low levels of dissolved oxygen, the death of fish, increased turbidity, and a loss of species diversity. These conditions threaten the long-term sustainability of fisheries and recreational uses of surface waters. In terrestrial environments, excess nutrient inputs to soils in managed ecosystems (e.g., agriculture, urban horticulture) can increase the risk of nutrient losses to groundwaters and surface waters. This can then lead to eutrophication of surface waters and also to human health problems if drinking waters are too enriched in nitrate-N or contaminated by toxic by-products produced through chlorination of eutrophic waters. Overloading natural ecosystems (e.g., forests) with nutrients can lead to soil acidification and changes in the composition and diversity of native plant species, both of which are usually ecologically undesirable.

Because N and P are essential nutrients for the growth and well-being of plants and animals – and because they are relatively inexpensive – many human activities result in the discharge of N and P into some sector of the environment. Both nutrients are discharged from point sources such as municipal wastewater treatment plants and concentrated animal-feeding operations (CAFOs). Years of effort and billons of dollars have been invested in reducing point-source nutrient pollution of surface waters by improving the nutrient-removal efficiency of wastewater treatment plants. N and P are also common nonpoint pollutants that are widely added to soils as soil amendments for crop production (e.g., fertilizers, animal manures, and other by-products such as composts, and municipal wastewaters and biosolids) or via soil-based wastewater treatment systems such as septic systems. Combustion of fossil fuels generates biologically available N and contributes to atmospheric contamination of many watersheds. Soil characteristics (physical, chemical, and biological) and soil-management practices profoundly influence the potential for nonpoint-source pollution of water bodies by nutrients and also determine the response of terrestrial biota to nutrient additions. Because of this, the implementation of 'best management practices' (BMPs) that reduce N and P losses from agricultural soils by processes such as erosion, runoff, and leaching, is a high priority worldwide today.

Aquatic Eutrophication

Causes

Additions of nutrients can generate remarkable changes in the primary productivity (the generation of biomass through photosynthesis) of aquatic systems. Because so many aspects of a freshwater ecosystem can be traced to nutrient supply, nutrient status is the basis for the widely used trophic (i.e., nutrition level) classification system applied to water bodies. Low-nutrient water bodies are classified as 'oligotrophic' (poorly nourished); high-nutrient water bodies are termed 'eutrophic' (highly nourished), and the intermediate state is referred to as 'mesotrophic' (Figure 1). Extremely nutrient-rich conditions do occur and these water bodies are classified as 'hypereutrophic.'

Limiting Nutrients

The nutrient status of a water body is not constant. Decreases can occur due to control of point-source pollution discharges and widespread implementation of BMPs to curb nutrient inputs from nonpoint pollution sources. More often, the nutrient supply within an aquatic ecosystem increases over time. This increase is known as eutrophication – it can occur naturally in response to slow increases in the stores of organic matter and sediment within the system. If the rate of nutrient increase is accelerated due to human activities, the process is known as 'anthropogenic eutrophication.'

In the nineteenth century, Justus von Liebig developed the 'law of the minimum' to indicate that growth of most organisms is controlled by the 'limiting nutrient,' the nutrient in least supply. At the beginning of the twentieth century, Brandt extended Liebig's insights to surface water, noting that plankton (microscopic, free-floating plant and animal organisms that form the base of the food chain in aquatic environments) abundance was correlated with nutrient concentrations in freshwater lakes in Germany. Later, A.C. Redfield pointed out that the Liebig law should be viewed in the context of the relative ratios of nutrients found within living algae. Redfield found

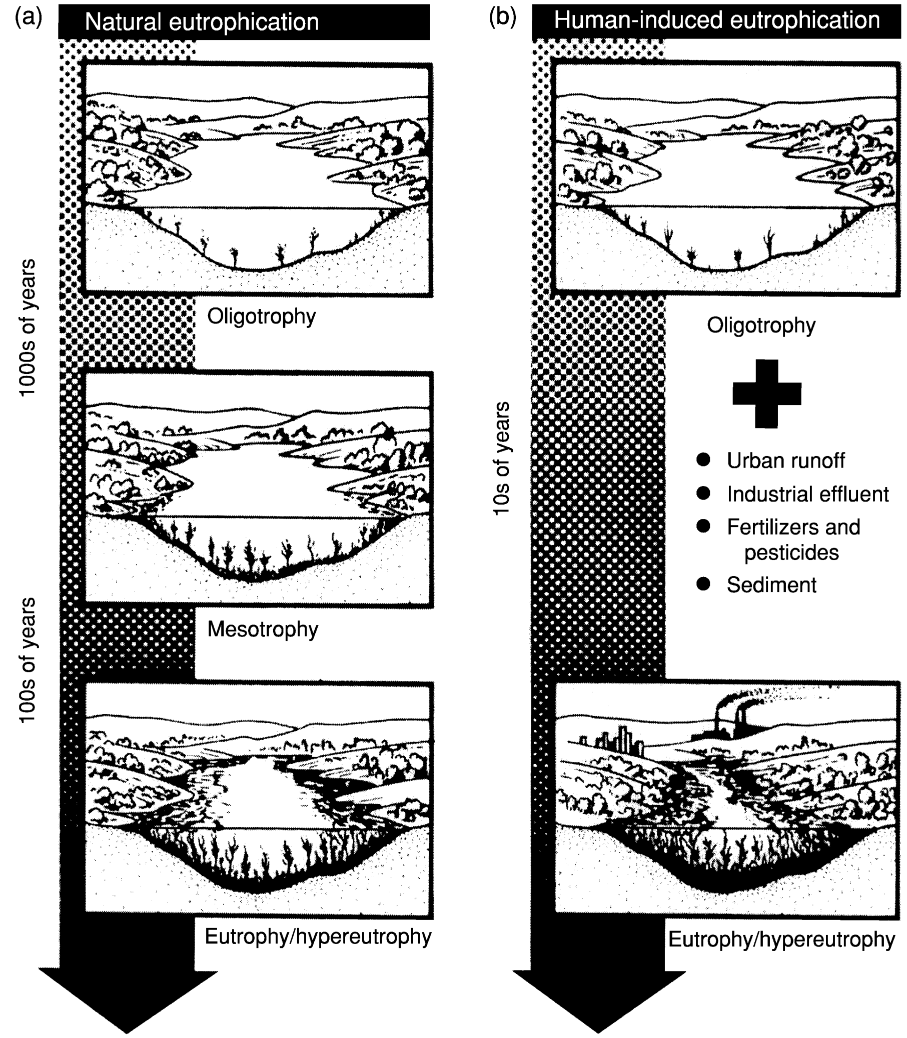


Figure 1 (a) Natural eutrophication describes the response of lakes to nutrient enrichment. Eutrophication increases primary productivity, leading to greater biomass and sedimentation. Oligotrophic lakes are nutrient-poor; eutrophic lakes are nutrient-enriched; (b) human activities induce anthropogenic eutrophication and generate rapid changes in nutrient enrichment and ecosystem characteristics. Reproduced with permission from NALMS (1990) *Lake and Reservoir Restoration Guidance Manual*. Madison, WI: North American Lake Management Society.

that aquatic algae and aquatic macrophytes are composed of fixed ratios of atoms:

106 C atoms:16 N atoms:1 P atom

By using the molecular weights of these atoms, we can translate these ratios into their mass ratios and express the composition in terms of the wet weight of living algae:

40 C:7 N:1 P per 500 wet weight of living algae

Thus, the law of the minimum suggests that additions of P will control primary production when the N:P mass ratio of available nutrients is greater than 7:1 and that the addition of 1 g of P can generate 500 g of algal biomass. If the N:P mass ratio is less than 7:1, N will be the limiting nutrient and additions of 1 g of N can be expected to stimulate 72 g of new algal biomass.

The limiting nutrient concept offers great practical value for the management of aquatic ecosystems. It suggests that decision-makers may be able to control ecosystem functions and values by managing inputs of a single nutrient. Today, most scientists agree that P is the limiting nutrient in freshwater ecosystems and that N limits primary production in coastal marine systems. C is rarely limiting, due to the presence of carbon dioxide in the atmosphere. These insights have emerged after heated debates within the scientific community, and exceptions to these ‘rules’ can be found in the literature.

Perhaps the most striking evidence for P limitation of lake eutrophication was demonstrated in a whole-lake study conducted in Canada by an international team in the 1970s (Figure 2). An oligotrophic lake with two similar basins was divided by a plastic curtain. One basin received additions of C and N, the other basin received C, N, and P. The basin receiving phosphorus developed intense blue-green algae blooms, while the other side remained clear and unchanged. The results of this experiment served to galvanize decision-makers around the globe to curtail P inputs into freshwater ecosystems.

A number of characteristics, such as rapid flushing rate, high depth-to-surface area ratio, or highly calcareous waters, can reduce the eutrophication response to nutrient inputs in freshwater lakes. Based on lake chemistry, morphology, and history of eutrophication, lakes can also develop internal stores of P that can continue to induce eutrophication even if external sources are controlled. P can accumulate on the bottom of lakes from the deposition of eroded soil particles transported to the lake by its tributaries. Accumulations of P in the form of organic sediments can also occur as eutrophication increases the quantity of organic matter (e.g., algae, macrophytes) in aquatic systems. Of equal importance, if aerobic conditions predominate in lake sediments and in the water overlying these sediments, dissolved P can chemically react with inorganic sediment constituents such as iron (Fe), manganese (Mn), and aluminum (Al) and form insoluble compounds that are much less bioavailable than soluble P. However, if anaerobic conditions develop in bottom waters and sediments, some of these compounds, especially iron phosphates, can dissolve and remobilize soluble P into the water column, where it will once again be available for uptake by algae and other aquatic species. In some cases, to mitigate the effects of these ‘internal loads,’ organic sediments are dredged to remove P physically from the lake, or chemically stabilized with additives such aluminum sulfate (alum) that precipitate soluble P into biologically unavailable forms.

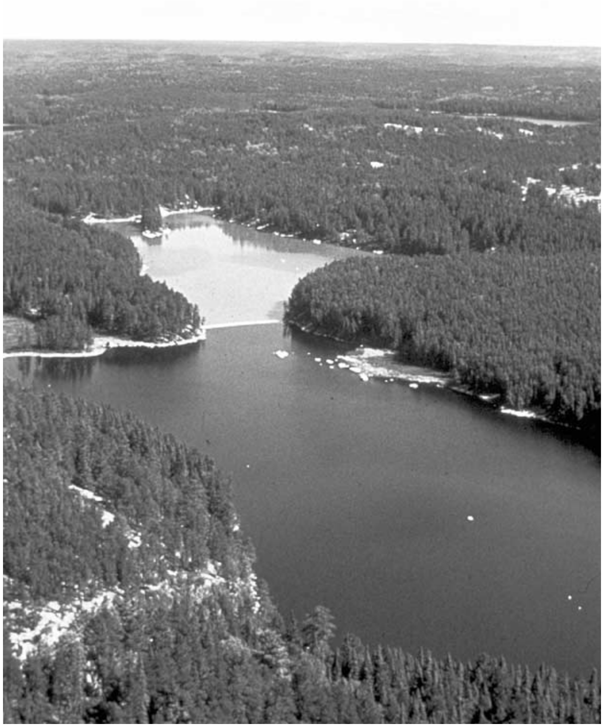


Figure 2 Additions of small amounts of phosphorus to one section of Lake 226 in the Experimental Lakes Area of Ontario, Canada, caused extensive surface blooms of blue-green algae (top) and vividly demonstrated the importance of phosphorus as a cause of excessive algal growth or eutrophication. Photo from Fisheries and Oceans Canada, Experimental Lakes Area.

Both ‘controlled addition’ experiments and whole-ecosystem studies support the premise that N is the limiting nutrient to eutrophication in most coastal marine systems. Nitrogen additions in mesocosms simulating the Narragansett Bay of Rhode Island have generated increases in algal biomass. In addition, long-term data from Swedish estuaries show that changes in algal abundance are closely related to changes in N inputs rather than P inputs (Figure 3).

The amount of P present in a water body is partially controlled by internal factors that affect solubility and by the import of soil (erosion of particulate P) and water (dissolved P) from the watershed and any ‘flushing’ processes that subsequently export P and carry it further downstream. Biological processes do not create or reduce the mass of P within an aquatic ecosystem. In contrast, N can be continuously added or removed from the land and water by biological processes. N-fixing plants such as legumes (terrestrial) or blue-green algae (aquatic) transform atmospheric N_2 into organic forms of N, thus adding N to soils and/or waters. Denitrifying bacteria transform dissolved nitrate-N into gaseous forms (N_2O, N_2) and thus remove N from aquatic and terrestrial systems. Although N can accumulate in bottom sediments, the dynamics of the N cycle provide opportunities for N to leave aquatic systems in gaseous form, reducing stores of N within the system. Because gaseous losses of P do not occur, P tends to accumulate in sediments to a greater extent than N.

Several factors combine to produce low N:P ratios (i.e., below the Redfield mass ratio of 7:1) and N limitation in coastal marine waters versus higher N:P ratios and P limitation in freshwater lakes. First, the sources of nutrients to lakes and coastal marine waters differ. Both coastal and lake ecosystems receive nutrients from terrestrial and atmospheric sources. However, coastal systems also receive inputs from

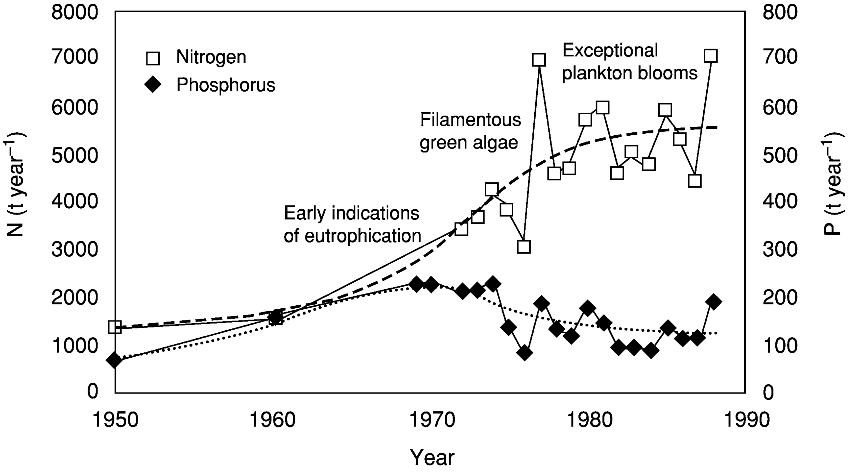


Figure 3 As evidenced by the biotic response of Laholm Bay in Sweden, nitrogen inputs generally control eutrophication in coastal marine waters. Reproduced with permission from Howarth RW, Anderson DM, Church TM *et al.* (2000) *Clean Coastal Waters: Understanding and Reducing the Effects of Nutrient Pollution.* Washington, DC: National Academy Press.

ocean waters with low N:P ratios due to denitrification on the continental shelf. Second, N fixation combined with slower flushing rates in lakes contributes to higher N:P ratios in lakes, whereas coastal marine waters experience little N fixation and any such additions are subject to rapid flushing from tides. Finally, sediment adsorption of P is often lower in coastal marine systems than in freshwater lakes, enhancing the relative supply of dissolved P in coastal waters.

Consequences of Aquatic Eutrophication

Stratification and Oxygen Depletion

The stimulation of primary production by nutrients can create algal blooms that reduce water clarity and lead to a number of other changes in aquatic ecosystems:

- Increased biomass of phytoplankton, suspended and attached algae;
- Decreased water column transparency;
- Shifts in phytoplankton composition to bloom-forming species, many of which may be toxic or inedible;
- Accumulation of carbon within the system;
- Changes in vascular plant production and species composition;
- Decrease in living aquatic habitats, including sea grasses and coral reefs;
- Depletion of deepwater oxygen concentration, resulting in hypoxia;
- Changes in fish species and fish production;
- Taste, odor, and water supply-filtration problems;
- Decrease in aesthetic values.

Oxygen is consumed as algae die and are decomposed by respiring microorganisms. The resulting oxygen depletion can be a major problem in stratified aquatic systems (Figure 4). Summer stratification occurs in freshwater lakes deeper than 3–5 m as the upper waters absorb solar energy and become warmer and less dense than the cool bottom waters. When stratified, a layer of rapid temperature change known as the ‘thermocline’ or ‘metalimnion’ effectively isolates the bottom waters and prevents oxygen diffusion from the atmosphere to waters below the thermocline. In estuarine systems, bottom waters can be isolated from the surface by both temperature gradients and a ‘halocline,’ a zone of rapidly changing salinity. Terrestrial water inputs create a low-salinity layer that floats above a heavier higher-salinity layer that is dominated by incoming waters from the open ocean. Regardless of the cause of stratification, dead and decaying algal biomass from surface waters will settle and, as decomposition proceeds, the oxygen supply of bottom waters becomes severely depleted. Oxygen depletion results in hypoxia (low oxygen) or anoxia (no oxygen). Both conditions directly kill many aquatic species. Anoxia can also alter sediment chemistry: reduced conditions can lead to sediment release of P, and production of hydrogen sulfide and methane, all contributing to water-quality deterioration. Hypoxia from eutrophication has damaged fisheries worldwide, from the Chesapeake Bay and Gulf of Mexico in North America, to the Baltic and Black Seas in Europe.

Changes in Species Composition

Eutrophication alters the composition and diversity of aquatic plants, affecting ecosystem structure and

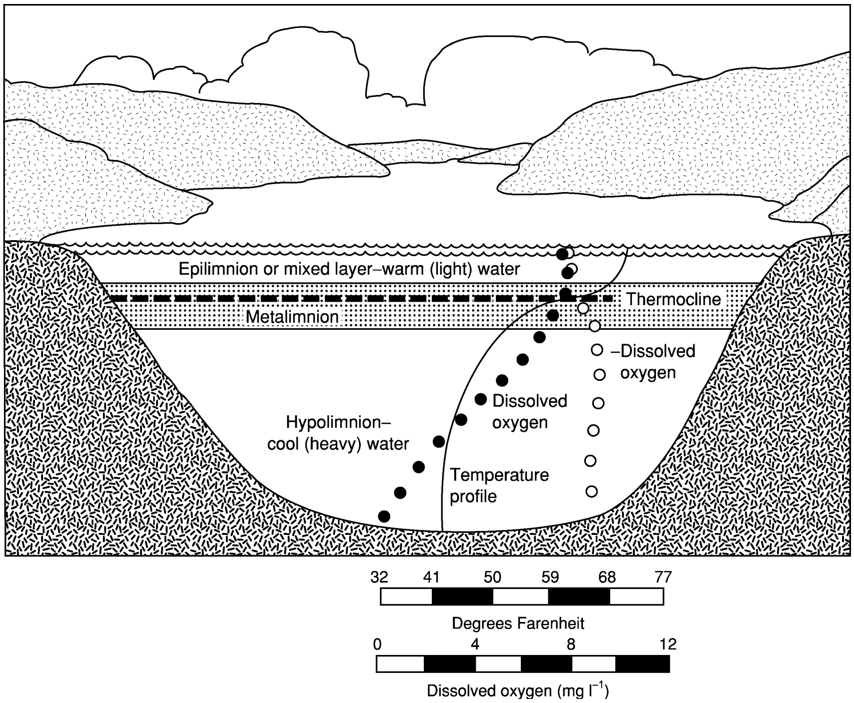


Figure 4 In temperate regions, most freshwater lakes deeper than 5 m undergo thermal stratification during the summer. Stratification separates a water body into distinct layers, isolating the cool bottom waters from atmospheric mixing and oxygen replenishment. Eutrophic lakes (solid circles) create substantial organic matter and its decomposition can deplete the stores of dissolved oxygen, leading to hypoxia or anoxia of the bottom waters. In contrast, oligotrophic waters generate little excess organic matter and minimal oxygen depletion occurs during stratification. Reproduced with permission from NALMS (1990) *Lake and Reservoir Restoration Guidance Manual*. Madison, WI: North American Lake Management Society.

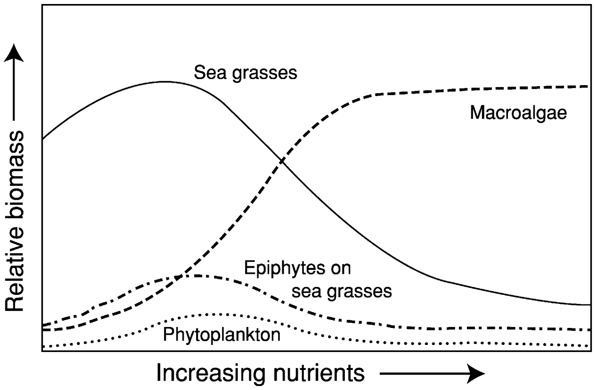


Figure 5 Increasing nutrients in shallow marine systems can shift aquatic plant communities from sea-grass beds that provide valuable habitats for marine organisms to nuisance macroalgae that cover the sediment with mats of rotting biomass. Nutrient enrichment stimulates the growth of phytoplankton in the water column and attached algae (epiphytes) on the sea grass, limiting light penetration below levels for sea-grass sustainability. Reproduced with permission from McComb AJ (ed.) (1995) *Eutrophic Shallow Estuaries and Lagoons*. Boca Raton, FL: CRC Press.

the food web (Figure 5). Increased inputs can shift algal composition in a freshwater lake from diatom-dominated systems, typical of oligotrophic lakes, to blue-green algae-dominated systems. Blue-green algae release toxins and are not readily ingested by secondary consumers. In addition many blue-green algae contain gas-filled vacuoles, causing the algae to float and accumulate on the water surface, effectively shading the lower waters and eliminating many important submerged plant species. Rotting masses of blue-green algae washed up on the shoreline of previously clear lakes is a discouraging sign that accelerated eutrophication has overtaken a lake's ecosystem. In coastal marine estuaries and bays, eutrophication has been linked to harmful algal blooms – often called 'red tides' – that cause widespread fatalities in fish and other marine organisms.

Sea-Grass Destruction

In shallow estuaries, increased nutrient inputs threaten the viability of sea grasses that serve as critical living spawning and nursery habitats. Sea-grass destruction rates rival the loss of tropical forests, creating serious problems in the Baltic Sea, Gulf of Mexico, and estuaries along the east coast of the USA. Sea-grass survival and growth are closely linked to light penetration. Elevated N inputs indirectly decrease light by stimulating both the growth of phytoplankton in the water column and attached algae (epiphytes) directly on sea-grass leaves. Because sea grasses stabilize bottom sediments, declines in sea-grass beds can increase suspended sediments, further decreasing water clarity and hastening the loss of submerged aquatic vegetation. Thus, restoration of sea grasses requires control of N inputs and stabilization of bottom sediments.

Eutrophication can stimulate nuisance blooms of benthic macroalgae (seaweeds) that have lower light requirements than sea grasses. The macroalgae create thick extensive mats over sea-grass beds and the sediment surface, effectively reducing habitats for fish species that require mineral substrates to spawn. During periodic die-off cycles, macroalgae decay results in oxygen depletion, fish kills, and accumulation of rotting algae along the shoreline.

Degradation of Corals

Nutrient enrichment is also a threat to coral reefs, with both N and P implicated in coral reef degradation. Although coral reefs are among the most productive ecosystems on the planet, they thrive in high-clarity, nutrient-poor environments. Only slight nutrient enrichment of coral reefs is needed to stimulate the growth of attached macroalgae, which inhibits coral propagation. In addition, macroalgae blooms generate periodic hypoxia on the reef. Nutrient degradation of coral reefs is a global problem, with notable examples reported from the Great Barrier Reef, the Caribbean, and Hawaii.

Modeling Eutrophication

Phosphorus-Limited Freshwater Systems

Reducing nutrient inputs to aquatic systems can reverse the effects of eutrophication. Great progress has been made in our ability to predict the extent of lake eutrophication through models that link P loading (a term used to describe nutrient input: mass per time) to hydrologic and morphologic lake characteristics, such as hydraulic residence time (years), mean depth, and lake volume. These models derive from the work of R.A. Vollenweider.

Two different equations can be used to predict lake response to P inputs. The first predicts mean lake P concentrations in a lake:

$$P = P_{\mathrm{i}}/(1 + T^{0.5})$$

This equation predicts that mean lake P concentration (P, in milligrams per cubic meter) will increase in proportion to the inflow concentration of phosphorus (P_{i}; milligrams per cubic meter), which is derived from P loading per outflow). P is expected to decrease with increasing hydraulic residence time (T, in years, obtained from lake volume per outflow).

The second predicts algal biomass (commonly quantified by the concentration of chlorophyll a (Chl a; in milligrams per cubic meter) – a photosynthetic pigment found in algae) – written as a function of mean P:

$$\mathrm{Chl\ a} = 0.68P^{1.46}$$

The use of these equations can provide valuable insight into the required reduction of P loading needed for lake and watershed management efforts to improve water quality.

Nitrogen-Limited Coastal Marine Systems

Because of the diversity and complexity of coastal marine systems, no single modeling approach can capture the range of responses of marine ecosystems to nutrient loading. A variety of modeling approaches are now underway and reflect basic differences in characteristics such as physical setting (i.e., fjords, shallow estuarine lagoons, or river estuaries), biological communities (i.e., mangrove swamps, sea grasses, or planktonic systems), and hydrodynamics (i.e., watershed inputs, tidal flushing, water retention, and stratification patterns). Some of these models have been tested for selected locations and have proven quite useful for the prediction of sea-grass response and phytoplankton production to nutrient enrichment.

Terrestrial Eutrophication

Nutrient enrichment can also affect both managed and natural terrestrial systems. In managed systems such as agriculture, the most common cause of nutrient enrichment is overapplication of fertilizers, manures, or other organic by-products. Overapplication frequently occurs where there is an inadequate land base to assimilate the amount of organic by-product produced (e.g., sewage sludge generation in urban areas or manure production from CAFOs that are geographically concentrated on a small land base). As a result nutrient sources are applied at rates and times of the year that overwhelm the capacity for plant uptake. P accumulates in soils to values well above those needed for crop production. For both N and P, overapplication increases the potential for losses to ground and surface waters by leaching and runoff. In natural ecosystems such as forests, regular additions of nutrients in fertilizers and organic by-products are rarely made. However, because the atmosphere has become enriched with plant-available forms of N as a by-product of fossil fuel combustion, terrestrial eutrophication of natural ecosystems has become a global problem. Atmospheric loading of plant-available N can exceed

$40\,\mathrm{kg\,ha^{-1}}$ per year within areas down-gradient of industrial zones, such as in southern California and northern Europe. These levels are more than 10 times greater than those found in much of the continental USA and represent a major increase over preindustrial times. High-density animal production systems also contribute to atmospheric loading of N. These locales can generate elevated emissions of gaseous ammonia-N that results in highly enriched atmospheric deposition of N to adjacent soils and surface waters. The long-term sustainability of natural ecosystems is threatened by these elevated inputs. In experiments conducted on grasslands in Europe and North America, increasing atmospheric deposition of plant-available N creates substantial declines in the species diversity of plants and insects.

Forested ecosystems can become 'N-saturated' as a result of eutrophication; atmospheric deposits of N can exceed the uptake capacity of the forest vegetation and soils. The excess N often converts to nitrate-N, a soluble anion that moves rapidly through soils and then travels to water bodies via groundwater discharge and surface runoff, thus contributing to eutrophication of aquatic systems. The transformation of nitrate from these external inputs also generates soil acidity that can contribute to depletion of essential cations in soils, particularly calcium (Ca) and magnesium (Mg), in turn diminishing the long-term productivity of the forest ecosystem. The increase in soil acidity can also mobilize Al ions and threaten aquatic life owing to the toxicity of Al to aquatic organisms.

Watershed Export and Models

The modern era is generating unprecedented levels of nutrient loading to surface waters. In the past 100 years, riverine discharge of N to large marine systems has increased four- to 10-fold, while P discharge has risen threefold. These increases in loading coincide with the expansion of human-generated nutrient inputs to the terrestrial environment. In particular, the global supply of available N has doubled since World War II, through production of synthetic N fertilizers, increased combustion of fossil fuels, and the expansion of land areas devoted to the cultivation of N-fixing leguminous crops.

Soils and soil management can play a profound role in the export of terrestrial nutrients to surface waters. Nutrient-balance studies conducted on watersheds ranging from local (i.e., less than 1000 ha) to regional scales (i.e., Baltic or Mississippi River Basin) demonstrate that rivers discharge only 10–30% of nutrient inputs to the watershed (Figure 6). Scientists now recognize the capacity of terrestrial environments to serve as nutrient 'sinks' that retain, remove, or transform nutrients and mitigate the effects of nutrient loading to surface waters. The soil N and P cycles contain a robust array of physical, chemical, and biological processes and transformations that control the fate of the terrestrial N and P inputs. Enormous quantities of N are immobilized by soil microbes and stored in soil organic matter. Agricultural soils retain 2000–5000 kg $\mathrm{N\,ha^{-1}}$, and even more N can be retained in the organic matter of

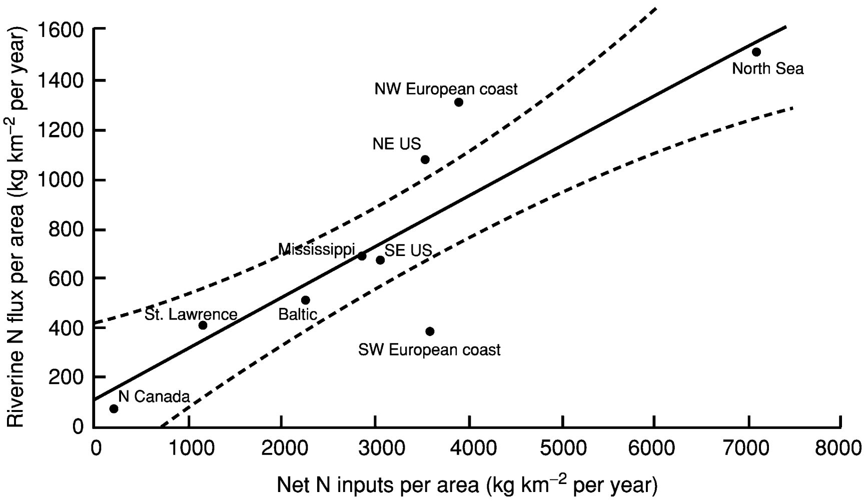


Figure 6 Nitrogen-balance studies conducted on large watersheds demonstrate that rivers discharge only 10–30% of nitrogen inputs. Transformations within the soil play an important role in reducing the effects of human-generated inputs to coastal marine waters. Reproduced with permission from Vitousek PM, Aber J, Howarth RW *et al.* (eds) (1997) Human alteration of the global nitrogen cycle: causes and consequences. *Issues in Ecology 1.*

wetland, forest, and grassland soils. Unfortunately, forest soils reaching N saturation will cease to function as N sinks, increasing the delivery of N to coastal waters.

Large quantities of P can be stored in soils by chemical fixation processes with soil constituents, such as Al, Ca, Fe, and organic matter. As with N, some soils in Europe and the USA are now becoming saturated with P due to long-term overapplication of fertilizers and animal manures, and may become sources of P in the future. Over time, human activity can alter the N stores in organic matter and the potential for losses of 'fixed' P in soils. Organic N can be mineralized and released through disturbance such as artificial drainage of wet soils, or conversion of forests and grasslands to routinely plowed croplands. P can be lost from soils as particulate P by erosion, a major problem worldwide, or as soluble P by leaching and surface runoff.

The control and reversal of aquatic eutrophication often rely on models to evaluate nutrient flux from forest, wetlands, riparian zones, and agricultural lands, in response to different management practices. Many modeling approaches are in use and reflect differences in the goals, setting, and scale of the simulation. Mathematical representations of soil hydrology, sediment transport, and biogeochemistry constitute the building blocks of these models – and the modeling approaches range from simple statistical relationships to highly parameterized mechanistic equations. Because of the inherent spatial variation in soil properties combined with the coarse scale of many available spatial databases, modelers recommend that indicators of uncertainty, such as confidence levels, be incorporated into estimates of nutrient export and that the complexity of the model match the available data and needs of decision-makers.

P flux models typically focus on soil processes and weather conditions that affect the movement of soluble and sediment-bound P forms in overland flow. In contrast, N export is usually dominated by groundwater flux, and models track nitrate losses by simulating N retention time, plant uptake, and microbial transformations within the root zone, and the timing and quantity of groundwater recharge out of the root zone.

Soil hydrologic models are central to nutrient export, because soil-moisture controls recharge to groundwater, rainfall/runoff relationships, and the extent of oxidized and reduced environments within the soil. Soil-moisture models typically generate a water balance based on infiltration, overland runoff, evapotranspiration, groundwater recharge, and changes in moisture storage within soil profiles in response to precipitation. Soil-moisture models range from mechanistic, highly parameterized schemes such as the Richards equation, to simple 'capacity' models that estimate soil-water flux in response to rules involving soil constants such as field capacity, permanent wilting point, and depth of the root zone. Intermediate, 'functional' models such as the SLIM model, developed by Addiscott, combine rate-based and capacity approaches and permit evaluation of nutrient dynamics in slowly draining soils. Once the soil-moisture status of the upper soil is established, overland runoff is often simulated through the curve number approach of the US Soil Conservation Service or through Green-Ampt techniques. Models such as AGNPS, developed by the USDA Agricultural Research Service, simulate runoff and sediment transport either in response to large 'design' storms or rely on daily soil-moisture balance approaches to estimate overland runoff.

Controlling Eutrophication: Future Trends

Natural and anthropogenic processes can mitigate nutrient losses and thus the potential for eutrophication. For example, advances in our understanding of soil denitrification can reduce the flux of nitrate from watersheds. This process generally occurs under anaerobic conditions where labile carbon is available to serve as the electron donor. Water table management through controlled drainage creates periodic drying and wetting cycles on cropland and shows promise for reducing nitrate export by stimulating nitrification (the microbial transformation of ammonium to nitrate that occurs in aerobic soils) and denitrification within the same location in soils. In addition, advances in tillage and crop residue management can take advantage of research that suggests that small fragments of easily decomposable organic matter can create localized 'hotspots' of respiration and denitrification, even in aerobic soils. In many watersheds denitrification in wetland soils receiving inputs of nitrate-laden groundwater is a major N removal process, particularly along corridors of undisturbed riparian (i.e., riverine) vegetation that intercept groundwater before it recharges streams. Thus, destruction of wetlands can result in disproportionate increases in nutrient loading due to the loss of the nutrient retention function of those locations. Conversely, reforestation of degraded lands can slowly sequester N. Efforts are under way to reduce P losses through the implementation of soil conservation BMPs (i.e., buffer strips, grassed waterways, terraces) that reduce soil erosion, the major source of P to surface waters. Clearly, comprehensive nutrient-management plans designed to prevent overapplication of manures and fertilizers is a critical aspect of nutrient control. Equally important to the control of anthropogenic eutrophication is the recognition of the limits of soil-based disposal of organic by-products from agricultural activities – and the development of value-added products for those wastes, such as composts and pelletized manure-based fertilizers, or fuels for 'bioenergy' plants.

List of Technical Nomenclature

Chl a	Chlorophyll a (mg m^{-3})
P	Mean lake phosphorus concentration (mg m^{-3})
P_i	Inflow concentration of phosphorus (mg m^{-3})
T	Hydraulic residence time (years)

See also: **Denitrification**; **Macronutrients**; **Nitrogen in Soils:** Cycle; **Nutrient Management**; **Phosphorus in Soils:** Overview; Biological Interactions; **Pollution:** Groundwater; **Septic Systems**; **Watershed Management**

Further Reading

Aber JD and Melillo JM (1991) *Terrestrial Ecosystems*. Philadelphia, PA: Saunders College Publishing.

Addiscott TM, Whitmore AP, and Powlson DS (1991) *Farming, Fertilizers and the Nitrate Problem*. Wallingford, UK: CAB International.

Doering PH, Oviatt CA, Nowicki BL, Klos EG, and Reed LW (1995) Phosphorus and nitrogen limitation of primary production in a simulated estuarine gradient. *Marine Ecology Progress Series* 124: 271–287.

Harlin MM (1995) Changes in major plant groups following nutrient enrichment. In: McComb AJ (ed.) *Eutrophic Shallow Estuaries and Lagoons*, pp. 173–187. Boca Raton, FL: CRC Press.

Howarth RW, Anderson DM, Church TM *et al.* (2000) *Clean Coastal Waters: Understanding and Reducing the Effects of Nutrient Pollution*. Washington, DC: National Academy Press.

NALMS (1990) *Lake and Reservoir Restoration Guidance Manual*. Madison, WI: North American Lake Management Society.

Nixon SW (1995) Coastal marine eutrophication: a definition, social causes and future concerns. *Ophelia, International Journal of Marine Biology* 41: 199–219.

Smith VH (1998) Cultural eutrophication of inland, estuarine and coastal waters. In: Pace ML and Groffman PM (eds) *Successes, Limitations, and Frontiers in Ecosystem Science*, pp. 7–49. New York: Springer-Verlag.

Vallentyne JR (1974) *The Algal Bowl – Lakes and Man*. Miscellaneous Special Publication 22. Ottawa: Department of the Environment.

Ver LMB, Mackenzie FT, and Lerman A (1999) Biogeochemical responses of the carbon cycle to natural and human perturbations: past, present, and future. *American Journal of Science* 299: 762–801.

Vitousek PM, Aber J, Howarth RW *et al.* (eds) (1997) Human alteration of the global nitrogen cycle: causes and consequences. *Issues in Ecology 1*.

Wetzel RG (2001) *Limnology: Lake and River Ecosystems*, 3rd edn. New York: Academic Press.

EVAPORATION OF WATER FROM BARE SOIL

C W Boast and F W Simmons, University of Illinois, Urbana-Champaign, IL, USA

Introduction

Evaporation of water from bare soil is often an important component of the soil water balance. If plants are living in the soil, evaporation of water from plant surfaces and evaporation from the soil are conceptually distinguished by calling the former transpiration and the latter evaporation, although both are evaporative processes – that is, both are characterized by a transformation of liquid water into water vapor. The combination, transpiration from plants and evaporation from soil, is called evapotranspiration. The bare soil surface of interest here can range from an entire plant-free (fallow) field to a small area between plants; for example, between trees in an orchard or between plant rows in a row-cropped field.

Evaporation Rate and Cumulative Evaporation

For an area of bare soil, it is often useful to know, at time t, the rate, $E(t)$ at which water is evaporating, per unit area of soil surface (quantity of water evaporating per square meter per second). The quantity of water is sometimes expressed by its mass, but more commonly by its volume, in which case the units of the evaporation flux density, $E(t)$, are (cubic meters per square meter per second = meters per second).

Alternatively, the evaporation process can be quantified by E_{cumul}, the cumulative amount of evaporation (per unit area of soil surface) which has occurred since some baseline time t_0. The relationships between $E(t)$ and E_{cumul} are:

$$E_{\mathrm{cumul}} = \int_{t_0}^{t} E(t)\mathrm{d}t \quad \text{and} \quad E(t) = \frac{\mathrm{d}E_{\mathrm{cumul}}}{\mathrm{d}t} \qquad [1]$$

If water volume is used to quantify the amount of water evaporating, that is, if $E(t)$ is expressed in meters per second, then cumulative evaporation is expressed in meters.

Processes Involved in Evaporation, and Latent Heat of Evaporation

In order for liquid water to change to water vapor, at a given location, three processes must occur: liquid water must be transported to, water vapor must be transported away from, and energy must be transported to the location. The relationship between the mass of water which is converted from liquid to vapor form and the amount of energy required is quantified by the latent heat of vaporization of water, denoted L_e. For pure, free water at 20°C, the value of L_e is $2.45 \times 10^6\,\mathrm{J\,kg^{-1}}$. Since L_e decreases only slightly with rising temperature, by approximately 0.1% per degree centigrade, L_e is often taken as constant over the range of temperatures encountered in natural systems.

The ultimate source of the energy which causes water to evaporate from a bare soil is electromagnetic radiation from the sun. The solar constant quantifies the average rate that solar (short-wave) radiation strikes the Earth's atmosphere: $1370\,\mathrm{W\,m^{-2}}$. The rate that short-wave radiation strikes the Earth's surface is denoted $J_s\downarrow$ (watts per square meter) and amounts to approximately one-half of the solar constant for a clear atmosphere, with the sun directly overhead. Integrated over a 24-h period, under more typical conditions, the daily amount of short-wave radiation reaching a bare soil surface can be as large as $20\text{–}30 \times 10^6\,\mathrm{J\,m^{-2}}$ or more. If this radiation load were to evaporate water at the rate $L_e = 2.45 \times 10^6\,\mathrm{J\,kg^{-1}}$, then, in order of magnitude the daily amount of water evaporated would be $10\,\mathrm{kg\,m^{-2}}$. Dividing by the density of water, $1000\,\mathrm{kg\,m^{-3}}$, gives this approximate upper limit of daily water evaporation as 0.01 m. This 1-cm daily evaporation serves as a rough benchmark for evaporation measurements. If the rate of evaporation E is expressed as kilograms per square meter per second, L_eE represents the rate that energy is consumed in the evaporation process (watts per square meter). On the other hand, if E is expressed as meters per second, and L_e is expressed as joules per cubic meter, then L_eE again represents watts per square meter.

Net Radiation

Seldom, even under ideal conditions, does daily evaporation amount to 0.01 m, because many other factors besides the rate, $J_s\downarrow$, of short-wave radiation striking the soil surface influence the actual amount of evaporation. First, a fraction, denoted α, of incoming short-wave radiation is reflected, and the rate that energy is reflected, $J_s\uparrow$ (watts per square meter) equals $\alpha J_s\downarrow$. The fraction α is called the soil surface's albedo, and, by definition, $0 \leq \alpha \leq 1$. Typical values of α range from 0.1, for wet, rough, dark-colored soil with the sun overhead, to 0.4, for dry, smooth, light-colored soil.

In addition to the downward-directed and upward-directed short-wave radiation, the net effect of electromagnetic radiation on the evaporation process includes two other terms: upward-directed long-wave radiation (emitted by the soil surface), whose rate is denoted by $J_l\uparrow$ (watts per square meter), and downward-directed long-wave radiation (mostly emitted by clouds), which is denoted $J_l\downarrow$ (watts per square meter). Combining all these terms gives the rate of net radiation, R_n (watts per square meter), as:

$$R_n = J_s\downarrow (1 - \alpha) - (J_l\uparrow - J_l\downarrow) \qquad [2]$$

During the daytime, the short-wave portion of this expression dominates, making the net radiation (directed toward the soil surface) positive and providing energy that can cause evaporation. At night the short-wave portion is essentially zero and the upward-directed long-wave radiation is usually greater than the downward-directed long-wave radiation (because the Earth's surface is usually warmer than the temperature of clouds or bare sky), so the net radiation (R_n, directed downward toward the soil surface) is usually negative. That is, the electromagnetic radiation energy balance at night is generally away from the soil surface, which provides no energy for evaporation, but provides an energy sink for vapor condensation at the soil surface and within the soil.

Energy Balance

Three possible fates of electromagnetic energy which arrives at the soil surface during the day are to heat the air (at a rate H, in watts per square meter), to heat the soil (G, in watts per square meter), and/or to evaporate water (L_eE, in watts per square meter). The idea that these are the only three possibilities is illustrated in Figure 1a and is expressed as an energy balance at the soil surface:

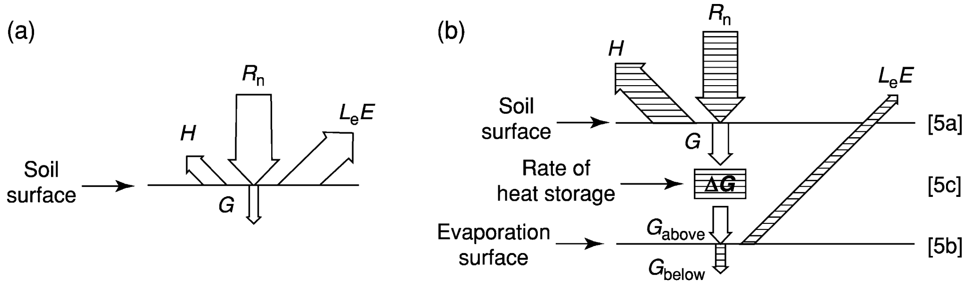


Figure 1 Energy balance (a) for the soil surface if all evaporation occurs there (Eqn [3]), and (b) (if all evaporation occurs at an 'evaporation surface' below the soil surface) energy balance for these two surfaces and the soil between them (Eqns [5] and (shaded) Eqn [6]).

$$R_n = H + G + L_eE \qquad [3]$$

The arrows in Figure 1 indicate the sign conventions, that is, an arrow points in the direction which is considered positive. During most of the daytime, each of the four terms in Eqn [3] is positive. At night, three of the terms, R_n, H, and G, are generally negative, reflecting the fact that heat is transferred toward the soil surface from air and soil, and that this heat is converted into outward-directed (longwave) electromagnetic radiation. At night, E can be either positive (continued evaporation from the soil) or negative (dew formation), depending on the relative magnitudes of R_n, H, and G. Over any given 24-h period, the cumulative amounts of R_n, H, and E are usually positive, with day-long or multiday cumulative values of G tending to be positive during spring warming and negative during autumn cooling.

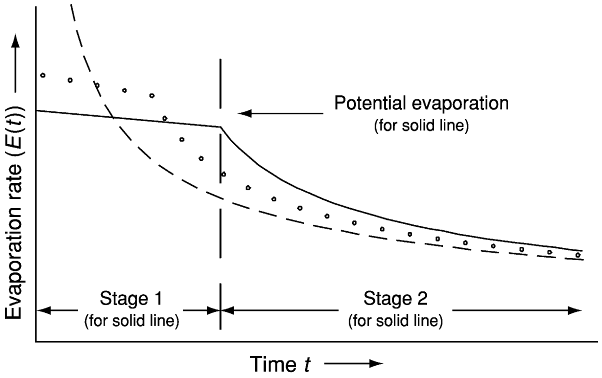


Figure 2 Schematics of (solid line) the two stages of evaporation from bare soil under conditions of constant potential evaporation, and (dashed line) soil evaporativity under conditions of unlimited potential evaporation. Dotted line is two-stage evaporation for slightly larger potential evaporation than solid line.

Stages of Evaporation

A striking feature of bare soil evaporation is illustrated by the solid line and curve in Figure 2. Under constant atmospheric and radiation conditions, evaporation from wet soil is observed to be nearly constant for a certain amount of time (stage 1, an almost horizontal line), then to decrease markedly (stage 2). The nearly constant rate of evaporation from bare soil during stage 1 is approximately equal to the rate of evaporation from an open body of water, or from a well-watered area with full plant cover, and this rate of evaporation is called the potential evaporation rate. On the other hand, atmospheric and radiation conditions vary over time, so the potential evaporation rate itself varies significantly, and evaporation from initially wet bare soil varies much more than is shown in Figure 2 during stage 1.

For both a constant and a varying potential evaporation rate, it has been observed that, after a period where evaporation of water from initially wet bare soil approximately equals the potential evaporation rate, the evaporation rate abruptly starts to fall below the potential evaporation rate. The time at which this occurs marks the boundary between stage 1 and stage 2 evaporation, and it is an important benchmark for describing the role of soil in the hydrologic cycle. One indirect way to detect when stage 2 begins is to measure the temperature of the soil surface. For a given set of atmospheric conditions, the daytime maximum surface temperature is generally higher, and the amplitude of the daily surface temperature fluctuation is larger during stage 2, than during stage 1. Also, for many soils there is a visible change in the color of the soil surface upon drying, and this change occurs at about the time when stage 1 evaporation ends and stage 2 begins. The change in color coincides with the disappearance of widespread water–air interfaces at the soil surface. (In other words, the lighter soil color indicates that the soil which is visible at the soil surface is 'dry.') Initially, the dry soil layer is very thin, and the dry layer becomes thicker over time.

Under certain circumstances a third stage of bare-soil evaporation occurs. For example, in the presence of a shallow water table whose depth is constant over time, it is possible for a steady upward flow from the water table to the soil surface to occur. In this case, the evaporation rate decreases until it approaches this upward flow rate. Even if the approach is asymptotic, there may come a time after which, for all practical purposes, the evaporation rate is constant. This situation is sometimes called stage 3 evaporation. However, because the approach to constant evaporation rate is asymptotic, there is usually no clear-cut definition of the boundary between stage 2 and stage 3 evaporation.

If potential evaporation is a larger (but constant) value than that shown for the solid line in Figure 2, then initially the evaporation rate is larger (nearly horizontal dotted line). After a period of stage 1 evaporation (of shorter duration than for the solid line), the evaporation rate decreases (dotted curve) more rapidly than for the solid curve. Thus, depending on the potential evaporation rate, an entire family of curves can occur. For the idealized condition of an unlimited potential evaporation rate, represented by the dashed line in Figure 2, evaporation from initially wet bare soil starts out very fast and rapidly decreases.

Evaporation under Conditions of Unlimited Potential Evaporation

Evaporation of water from a bare soil, under conditions of unlimited potential evaporation, can be analyzed by considering what happens at the transition from stage 1 to stage 2 evaporation. Two events occur simultaneously, or nearly simultaneously: the evaporation rate suddenly starts to drop below the potential evaporation rate, and the soil surface water content changes precipitously from a wet value to a dry value. The larger the potential evaporation rate the sooner these events occur after the onset of evaporation. In the extreme, for an effectively infinite potential evaporation rate, it makes sense to visualize that stage 1 evaporation does not occur at all and that the surface soil water content changes from a wet value to a dry value at the onset of evaporation.

For an idealized, infinitely deep, soil with uniform initial ('wet') water content, θ_w, whose surface water content instantaneously changes to a dry value, θ_d, at $t_0 = 0$, this scenario is expressed as the following initial condition and boundary conditions:
Initial condition:

$$\theta = \theta_w \quad \text{at} \quad t = t_0 \quad \text{for} \quad 0 < z < \infty \qquad [4a]$$

Boundary conditions:

$$\begin{aligned} &\theta = \theta_d \quad \text{at} \quad z = 0 \quad \text{for} \quad t_0 < t < \infty \\ &\theta \to \theta_w \quad \text{as} \quad z \to \infty \quad \text{for} \quad t_0 < t < \infty \end{aligned} \qquad [4b]$$

When the Richards equation, without its gravity term, is solved subject to these initial and boundary conditions, a remarkable result is obtained: regardless of the soil's hydraulic properties, the evaporation rate is inversely proportional to the square root of time, $E(t) = s(\theta_w, \theta_d)\, t^{-1/2}$. The proportionality constant, s, which is analogous to the sorptivity of infiltration theory, is a function (for a given soil) only of the initial water content, θ_w, and the surface water content, θ_d. This finding, although strictly applicable only (a) to a homogeneous soil, (b) of effectively infinite depth, (c) starting from uniform initial water content, and (d) uninfluenced by gravity, has been found to be applicable to situations which deviate greatly from these idealized conditions. And the seeming difficulty of an infinite initial evaporation rate ($t^{-1/2} \to \infty$ as $t \to 0$) is less troublesome when it is noted that (taking $t_0 = 0$, for simplicity of notation) the cumulative amount of evaporation after time t_0 is simply given by $2st^{1/2}$.

As a consequence of the simplicity of the square root of time behavior for evaporation under conditions of effectively unlimited potential evaporation, phase 2 evaporation has been modeled as a process whose rate equals $s'(t - t')^{-1/2}$, where s' and t' are parameters to account for the effect of stage 1 evaporation. One difficulty with this approach is that both s' and t' need to be chosen for each soil and each environmental situation, and it is very difficult to derive a theory which gives s' and t' as functions of measurable quantities.

Evaporation Below the Soil Surface

One of the salient features of evaporation from bare soil is the drying of the soil surface (first evident at about the time evaporation enters stage 2). As evaporation proceeds, a thicker and thicker layer of dry soil develops, and Eqn [3] ceases to be useful for determining the total evaporation rate. It can still represent the energy balance at the soil surface, but one of its terms, L_eE, represents only that part of the evaporation which occurs at the soil surface.

If the evaporation process is distributed over a range of depths, an exact statement of the energy balance must represent a range of depths. However, it has been observed that, at any given time, evaporation takes place over a quite narrow range of depths. Below this range of depths, water moves upward in liquid form, and above this depth range water moves upward as vapor. Because the range of depths is narrow, it is reasonable to approximate reality by supposing that all the evaporation occurs at a single depth (called 'the evaporation surface' or 'drying front' and shown in Figure 1b) and to write two energy balance equations:

$$\text{at the soil surface}: R_n = \mathrm{H} + \mathrm{G} \qquad [5a]$$

$$\text{at the evaporation surface}: G_{\text{above}} = L_e E + G_{\text{below}} \qquad [5b]$$

where G_{above} and G_{below} represent, respectively, the soil heat flux arriving at the evaporation surface, from above, and leaving the evaporation surface, to go below (both downward-directed). If ΔG represents the rate that heat is stored in the soil between the soil surface and the evaporation surface, then the equation:

$$G = \Delta G + G_{\text{above}} \qquad [5c]$$

describes the relationship between G and G_{above}. These three equations contain seven variables, four more than the number of equations. This can be compared with four variables in Eqn [3], three more variables than the one equation in that case. Thus, the situation is more complicated than for Eqn [3]: there is one more unrelated variable. The three Eqns [5] can be combined to give:

$$R_n = H + \Delta G + L_e E + G_{\text{below}} \qquad [6]$$

Two soil heat terms, ΔG and G_{below}, must be estimated in this equation, whereas only one such term, G, must be estimated in Eqn [3]. As is illustrated by the shaded arrows and the shaded rectangle in Figure 1b, Eqn [6] does not describe the energy balance at any particular location, because it contains terms at multiple locations; for example, R_n and H at the soil surface, and $L_e E$ and G_{below} at the evaporation surface. However, it does validly represent the relationship between energy terms at these two relevant locations. It is a tool for estimating the evaporation rate E (if effectively all evaporation occurs at an evaporation surface, below the soil surface), just as Eqn [3] is a tool for estimating E (if effectively all evaporation takes place at the soil surface).

Methods for Measuring Evaporation

Five strategies for measuring evaporation from bare soil are now described, three direct strategies and two approaches in which evaporation is indirectly measured by quantifying the amount of energy, $L_e E$, which goes into evaporating water (using either Eqn [3] or [6]). There are large equipment and labor costs for the direct methods, so a great deal of effort has gone into devising indirect methods to measure E.

Weigh the Loss of Water Mass in the Soil

One direct method is repeatedly to weigh a body of soil which is isolated from surrounding soil, but which is positioned such that the isolated soil is subjected to the same radiation and heat transport environment as the undisturbed soil. The soil plus its container is either called an evaporimeter or, to emphasize its separation from surrounding soil, a lysimeter. A microlysimeter consists of a cylinder (Figure 3a) that is small enough to be filled with undisturbed soil by pushing the cylinder into the soil (Figure 3b) and cutting off the soil at the bottom of the cylinder. To complete the installation, the bottom of a microlysimeter is covered (Figure 3c) such that the soil is hydrologically isolated from, but can be thermally connected to, the soil below, and the microlysimeter is weighed and then placed in a hole with the soil surface inside and outside the microlysimeter, as well as the cylinder rim, all at the same level (Figure 3d). Larger lysimeters, on the other hand, are often filled with disturbed soil that must be allowed to settle and reconsolidate, sometimes for years, before it can be considered representative of undisturbed soil.

The crucial property of an evaporimeter is that the rate of evaporation from the evaporimeter equal that from the nonisolated, undisturbed soil which the evaporimeter is supposed to mimic. The bottom of a microlysimeter blocks vertical flow of water, whether this is downward flow following rainfall or irrigation, or upward flow during a drying period. Thus the soil inside the microlysimeter can be either wetter or drier than nonisolated soil. Either way, the

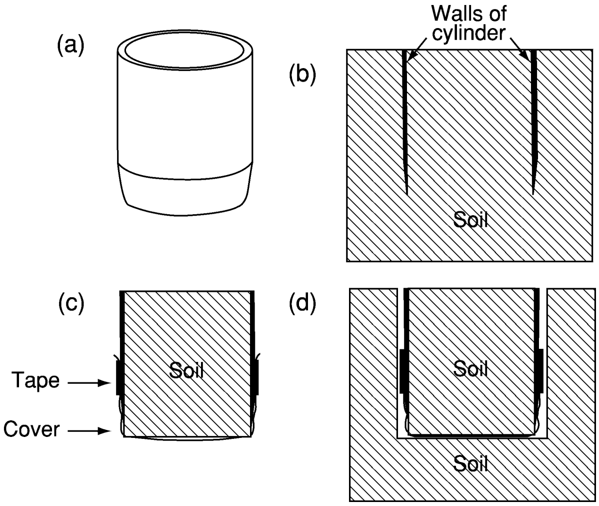


Figure 3 Procedure for microlysimeter determination of evaporation: (a) cylinder; (b) cylinder pushed into soil; (c) microlysimeter, partially taped, during preparation for first mass determination; (d) microlysimeter, in place, between mass determinations.

evaporation rate from the microlysimeter eventually deviates from what it is supposed to represent. And, in general, the shorter the microlysimeter the sooner this occurs. A rule of thumb is that the soil in a microlysimeter must be replaced (a) after any significant rainfall, and (b) almost every day during a drying period for a microlysimeter which is less than 0.1 m tall. The latter requirement can be relaxed to approximately weekly replacement for microlysimeters which are 0.3 m tall.

A variety of strategies have been employed to insure that a very large lysimeter, which is weighed *in situ*, does not deviate hydrologically from nonisolated soil. For example, if there is a shallow water table in the undisturbed soil, then the lysimeter can be equipped so that a water table is maintained at the same depth inside the lysimeter as the measured water table in intact soil. Once such a lysimeter is installed, it can be used indefinitely with a minimum of labor, whereas the labor cost of microlysimeters, which need to be periodically refilled with soil, can be very large. Microlysimeters have a low initial cost, and they can be used to measure evaporation for a small area of soil, for example to characterize bare soil evaporation in the space between plant rows for a row crop. A large lysimeter can characterize the combined effect of transpiration from plants and evaporation from soil, over an area which encompasses one or more row widths, vine spacings, or even a tree spacing, but large lysimeters generally cannot be used to determine separately water evaporation from soil and water transpiration from plants.

Detect the Arrival of Water Vapor Within the Air

Another direct method to quantify the rate of evaporation from bare soil is to measure the density of water vapor, ρ_v (kilograms per cubic meter) in a closed chamber containing air, above the soil surface, over a short interval of time. For example, if the evaporation rate is a rather low 0.06 mm h^{-1} (0.001 mm min^{-1}), then, for each square meter of soil surface area enclosed, 1 g of water vapor enters the air chamber per minute. The amount of water contained in air which is saturated with water vapor (at, for example, 15°C) is 12.7 g m^{-3}, so this rate of evaporation increases the relative humidity in, for example, a 0.8-m-tall chamber (which contains 10 g of water vapor per square meter at saturation) by 10% per minute. Since the evaporation rate can be 10 or more times as fast as in this example, the method can be quite sensitive for even a short closure time. Also, a reading must be completed fast enough so that the rising relative humidity of the air in the chamber does not reduce the evaporation rate.

Quantify the Upward Transport of Water Vapor in the Air

A somewhat less direct method to measure evaporation from a large field of bare soil is to use a model of vapor transport:

$$L_eE = L_e(\rho_{v,e} - \rho_{v,a})/r_v \qquad [7]$$

where E is expressed in kilograms per square meter per second, where $\rho_{v,e}$ and $\rho_{v,a}$ (kilograms per cubic meter) are the water vapor density at two positions, level ‘a’ in the air, and level ‘e’, either lower in the air than ‘a’ or at the evaporation surface, and where r_v (seconds per meter) is the resistance of the pathway from level ‘e’ to level ‘a’. In Eqn [7], the parameter which is most difficult to quantify is the resistance factor, r_v. If ‘e’ represents the evaporation surface for stage 2 evaporation, the soil portion of r_v can be modeled as a resistance to diffusion in a dry soil layer between an evaporation surface and the soil surface. And a variety of methods have been devised to estimate the air portion of r_v; that is, to characterize the wind-driven and/or buoyancy-driven mechanisms for transport in the air near the Earth. Two key methods are: (1) to directly measure vertical air movements; and (2) to use equations which express r_v as a function of measured horizontal wind speed at one or more heights, sometimes with corrections for atmospheric stability conditions.

One disadvantage of Eqn [7], as well as the indirect methods described below, is that these approaches can be inaccurate for anything except very large, uniform fields. If they are attempted near the edge of a bare soil area, then the measurements can be strongly influenced by neighboring land use. For example, the required fetch distance, upwind of the measurement location, can be on the order of 100 times the neighboring plant height.

Measure Enough Elements in the Energy Balance to Calculate L_eE

In Eqn [3] or Eqn [6], R_n can be measured with a net radiometer, and the soil heat terms can be either measured or estimated. This leaves the sum $L_eE + H$ known (by difference, in either Eqn [3] or Eqn [6]), but the individual values of L_eE and H unknown. The most common scheme for breaking L_eE+H into its two parts is practical only during stage 1 evaporation, but the first step in this scheme is more generally useful: the transport of heat into the air is modeled in a way which is very similar to how vapor transport is modeled in Eqn [7]:

$$H = \rho C_p(T_s - T_a)/r_h \qquad [8]$$

where T_s and T_a (degrees Kelvin) are the temperature at two levels with level 'a' in the air and with level 's' either lower in the air than level 'a' or at the soil surface (but not at the evaporation surface if this is below the soil surface), where ρ (kilograms per cubic meter) is the density of air, where C_p (joules per kilogram per degree Kelvin) is the specific heat of air at constant pressure, and where r_h (seconds per meter) is the resistance to heat transport from location 's' to location 'a'. As in Eqn [7], the entity which is most difficult to quantify is the resistance factor r_h.

One strategy for estimating evaporation is to define the Bowen ratio as $\beta = H/L_eE$, to replace H in Eqn [3] or Eqn [6] by βL_eE, and to solve for L_eE, giving, for Eqn [3], $L_eE = (R_n - G)/(1 + \beta)$, and giving a similar expression for Eqn [6]. If evaporation is occurring at the soil surface, during stage 1, then one strength of the Bowen ratio strategy is that (as portrayed in **Figure 1a**) H is usually smaller than L_eE, so β is small compared with 1, and even an approximate estimate of β can give a reasonably accurate estimate of the evaporation rate. A second strength of the Bowen ratio approach, for stage 1 evaporation, is that it is sometimes reasonable to assume that $r_h = r_v$. But only if level 's' in Eqn [8] is identical to level 'e' in Eqn [7]. Then β (the ratio of the right sides of Eqns [8] and [7]) can be determined from measurements or estimates of temperature and vapor density at the two levels.

These two strengths of the Bowen ratio approach are, for the most part, lost during stage 2 evaporation, in particular, when the factor r_v includes the resistance to vapor transport through a dry soil layer in addition to the resistance to transport through the air. Since heat transport, in the numerator of β, travels only through the air portion of this path, the factor r_h can be much smaller than r_v, and β can be large (or, as portrayed in **Figure 1b**, H can be much larger than L_eE, for dry soil). Thus, not only can one not assume that $r_h = r_v$ for stage 2 evaporation, the advantages which accrue when β is small, compared with 1, are lost.

Compare the Soil of Interest to a Completely Dry Soil

This method for measuring evaporation from drying bare soil requires establishment of a large body of dry soil, and taking enough measurements on the dry soil and on the drying soil to estimate the difference between their two energy balances. Combining Eqns [2] and [6] describes the energy balance of the drying soil:

$$H + \Delta G + L_eE + G_{\text{below}} = J_s{\downarrow}\,(1 - \alpha) - (J_l{\uparrow} - J_l{\downarrow}) \quad [9]$$

where it is assumed that all evaporation occurs at a single evaporation surface. (If the evaporation surface is at the soil surface, then ΔG equals zero and G_{below} could be replaced by G.) With the subscript 'zero' marking entities that are specific to the dry soil, the energy balance at the soil surface of the dry soil is:

$$H_0 + G_0 = J_s{\downarrow}\,(1 - \alpha_0) - (J_l{\uparrow}_0 - J_l{\downarrow}) \quad [10]$$

In this equation, there is no evaporation term, because there is no evaporation from the dry soil ($E_0 = 0$), and there are only two ways that the net radiation can be disposed of at the soil surface: heating air, at rate H_0, and heating soil, at rate G_0. Two terms in the right side of this equation are assumed to be the same for the dry soil as for the drying soil: the downward-directed short-wave and long-wave radiations, at rates $J_s{\downarrow}$, and $J_l{\downarrow}$.

Subtracting Eqn [10] from Eqn [9], and rearranging, gives:

$$L_eE = (H_0 - H) + (G_0 - \Delta G - G_{\text{below}}) + J_s{\downarrow}\,(\alpha_0 - \alpha) + \sigma(\epsilon_0 T_{s0}^4 - \epsilon T_s^4) \quad [11]$$

where the last term represents $J_l{\uparrow}_0 - J_l{\uparrow}$ (using the Stefan–Boltzmann law, with σ the Stefan–Boltzmann constant ($5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$), and with ϵ_0 and ϵ (dimensionless) the emissivities, and T_{s0} and T_s the soil surface temperatures of the dry soil and the drying soil, respectively). Because of evaporation, the surface temperature of the drying soil, T_s, is generally lower than that of the dry soil, T_{s0}, during the day. The dry soil's hotter surface causes more heating of the air there, so the first term in the right side of Eqn [11] is generally greater than zero during the day. Also, the solar reflection term is generally greater than or equal to zero, because the albedo of the dry soil, α_0, is greater than or equal to α for the drying soil. And, unless the emissivity of the dry soil is a lot smaller than that of the dry soil, the last term is greater than zero when $T_{s0} > T_s$. Generally, the only term in Eqn [11] whose sign is ambiguous during the daytime is the soil heat term. During the daytime the left side of Eqn [11] is positive, so, even if the soil heat term is negative, the other three terms in the right side of Eqn [11] more than compensate for this.

In general, two strategies have been employed to estimate the terms on the right side of Eqn [11]. In one strategy, the sum of the heat-flux term and the short-wave reflection term is assumed to be negligible compared with L_eE, and approximations are made in the other two terms so that their 24-h integrated value is proportional to the day's maximum $T_{s0} - T_s$ value. The advantage of this strategy lies in its simplicity, but these approximations are not always valid. In the other strategy, each of the terms in Eqn [11] are estimated, and the net integrated effect is calculated.

In either strategy, the $H_0 - H$ term in Eqn [11] is estimated by use of Eqn [8] twice, once for the dry soil, and once for the drying soil, and, generally, r_h is determined from wind speed measurements. Two key simplifying assumptions are that the resistance factor, r_h, be the same for the dry and drying soils, and that the same T_a can be used for both soils. The former assumption is analogous to the assumption, $r_v = r_h$ (in Eqns [7] and [8]), in the previous method. The method is relatively new, and many methodological questions have not been thoroughly explored. However, Eqn [11] quantifies a key difference between a dry soil and a drying soil, namely, that heat which is consumed in evaporating water in the drying soil (left side of Eqn [11]) must go elsewhere for a dry soil (right side of Eqn [11]). Soil surface temperature plays an important role in all (except for albedo) of the processes during stage 2 evaporation, so measurement of the surface temperature can provide much information in this scheme, and other schemes, for estimating the evaporation rate.

List of Technical Nomenclature

α	Albedo of drying soil (dimensionless)
α_0	Albedo of dry soil (dimensionless)
β	Bowen ratio, H/L_eE (dimensionless)
ΔG	Rate of heat storage in zone between soil surface and evaporation surface ($\mathrm{W\,m^{-2}}$)
ϵ	Emmissivity of drying soil (dimensionless)
ϵ_0	Emmissivity of dry soil (dimensionless)
θ	Volumetric soil water content (dimensionless)
θ_d	Volumetric soil water content at soil surface, during drying (dimensionless)
θ_w	Volumetric soil water content, before drying (dimensionless)
ρ	Density of air ($\mathrm{kg\,m^{-3}}$)
ρ_v	Water vapor density ($\mathrm{kg\,m^{-3}}$)
$\rho_{v,a}$	Water vapor density at level 'a' ($\mathrm{kg\,m^{-3}}$)
$\rho_{v,e}$	Water vapor density at level 'e' ($\mathrm{kg\,m^{-3}}$)
σ	Stefan–Boltzmann constant ($5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$)
C_p	Specific heat of air at constant pressure ($\mathrm{J\,kg^{-1}\,K^{-1}}$)
E	Rate of water evaporation at soil or evaporation surface ($\mathrm{m^3\,m^{-2}\,s^{-1} = m\,s^{-1}}$)
E_{cumul}	Cumulative water evaporation at soil or evaporation surface ($\mathrm{m^3\,m^{-2} = m}$)
G	Rate of heat transfer from soil surface downward into drying soil ($\mathrm{W\,m^{-2}}$)
G_0	Rate of heat transfer from soil surface downward into dry soil ($\mathrm{W\,m^{-2}}$)
G_{above}	Rate of heat transfer arriving at evaporation surface from above ($\mathrm{W\,m^{-2}}$)
G_{below}	Rate of heat transfer leaving evaporation surface downward ($\mathrm{W\,m^{-2}}$)
H	Rate of heat transfer from soil surface of drying soil into air ($\mathrm{W\,m^{-2}}$)
H_0	Rate of heat transfer from soil surface of dry soil into air ($\mathrm{W\,m^{-2}}$)
$J_s\uparrow$	Upward-directed short-wave radiation ($\mathrm{W\,m^{-2}}$)
$J_s\downarrow$	Downward-directed short-wave radiation ($\mathrm{W\,m^{-2}}$)
$J_l\uparrow$	Upward-directed long-wave radiation from drying soil ($\mathrm{W\,m^{-2}}$)
$J_l\downarrow$	Downward-directed long-wave radiation ($\mathrm{W\,m^{-2}}$)
$J_l\uparrow_0$	Upward-directed long-wave radiation from dry soil ($\mathrm{W\,m^{-2}}$)
L_e	Latent heat of vaporization of water ($2.45 \times 10^6\,\mathrm{J\,kg^{-1}}$)
R_n	Net radiation ($\mathrm{W\,m^{-2}}$)
r_h	Resistance to heat transport, from soil surface to position 'a' ($\mathrm{s\,m^{-1}}$)
r_v	Resistance to vapor transport, from evaporation surface to position 'a' ($\mathrm{s\,m^{-1}}$)
s	Proportionality constant for square root of time behavior during drying
s'	Parameter for stage-2-evaporation relationship
T_a	Temperature at level 'a' (K)
T_s	Temperature at surface of drying soil (K)
T_{s0}	Temperature at level 's' for dry soil (K)
t	Time (s)
t_0	Time at which evaporation commences (s)
t''	Parameter for stage-2-evaporation relationship
z	Depth (m)

See also: **Dryland Farming**; **Energy Balance**; **Evapotranspiration**; **Heat and Moisture Transport**; **Lysimetry**; **Radiation Balance**; **Thermal Properties and Processes**; **Thermodynamics of Soil Water**; **Water Cycle**

Further Reading

Ben-Asher J, Matthias AD, and Warrick AW (1983) Assessment of evaporation from bare soil by infrared thermometry. *Soil Science Society of America Journal* 47: 185–191.

Boast CW and Robertson TM (1982) A "micro-lysimeter" method for determining evaporation from bare soil: description and laboratory evaluation. *Soil Science Society of America Journal* 46: 689–696.

Brisson N and Perrier A (1991) A semiempirical model of bare soil evaporation for crop simulation models. *Water Resources Research* 27: 719–727.

Brutsaert W (1982) *Evaporation into the Atmosphere: Theory, History and Applications*. Norwell, MA: D. Reidel.

Evett SR, Matthias AD, and Warrick AW (1994) Energy balance model of spatially variable evaporation from bare soil. *Soil Science Society of America Journal* 58: 1604–1611.

Evett SR, Warrick AW, and Matthias AD (1995) Wall material and capping effects on microlysimeter temperatures and evaporation. *Soil Science Society of America Journal* 59: 329–336.

Hillel D (1998) *Environmental Soil Physics*. London, UK: Academic Press.

Lascano RJ and van Bavel CHM (1986) Simulation and measurement of evaporation from a bare soil. *Soil Science Society of America Journal* 50: 1127–1133.

Massman WJ (1992) A surface energy balance method for partitioning evapotranspiration data into plant and soil components for a surface with partial canopy cover. *Water Resources Research* 28: 1723–1732.

Monteith JL (1973) *Principles of Environmental Physics*. New York: American Elsevier.

Penman HL, Angus DE, and van Bavel CHM (1967) Microclimatic factors affecting evaporation and transpiration. In: Hagen RM, Haise HR, and Edminster TW (eds) *Irrigation of Agricultural Lands*, pp. 483–505. Monograph No. 11. Madison, WI: American Society of Agronomy.

Tanner CB (1967) Measurement of evapotranspiration. In: Hagen RM, Haise HR, and Edminster TW (eds) *Irrigation of Agricultural Lands*, pp. 534–574. Monograph No. 11. Madison, WI: American Society of Agronomy.

van Bavel CHM and Hillel DI (1976) Calculating potential and actual evaporation from a bare soil surface by simulation of concurrent flow of water and heat. *Agricultural Meteorology* 17: 453–476.

van de Griend AA and Owe M (1994) Bare soil surface resistance to evaporation by vapor diffusion under semiarid conditions. *Water Resources Research* 30: 181–188.

EVAPOTRANSPIRATION

G Stanhill, The Volcani Center, Bet Dagan, Israel

Introduction

Evapotranspiration is the total water loss to the atmosphere from a unit land surface area, usually expressed in units of depth; it includes the water vapor evaporating from the soil surface and from free water on plant surfaces as well as that transpired from within plant surfaces. The word 'actual' is sometimes added to distinguish it from 'potential' and 'reference evapotranspiration', terms used to describe the theoretical upper limit to total water loss from land surfaces under special experimental conditions.

Development of the Concept

'Evapotranspiration,' both actual and potential, was first defined by Thornthwaite in 1944, and the term became widely known and used following his 1948 publication in which potential evapotranspiration was calculated as a complex empirical function of air temperature and day length. However, the first published appearance of the word dates from 1937, albeit without explanation or definition and in a hyphenated form.

A major reason for the rapid and widespread adoption of the term was the success of the efforts made at this time, notably by Penman and later by Penman and Monteith, to estimate the total water loss to the atmosphere from standard climate measurements on a sound physical basis; another was the inadequacy of the methods then available to measure the different components of water loss under natural conditions.

However, Penman objected to the use of the word 'evapotranspiration' on the grounds that it was unnecessary, because 'evaporation' was an equally valid term for the separate components as well as for the total water loss to the atmosphere. There is historical support for this position in that when the use of the word ‘evaporation’ was first recorded in the middle of the sixteenth century it also covered transpiration, water loss from plants. The converse is also true, and the broad and imprecise use of both these words has persisted, even in the scientific literature, for more than 400 years.

Within two decades, emphasis on the need to lump together all the components of evapotranspiration diminished and recognition of the need to split the components returned with the appreciation of the importance of biological factors in controlling transpiration and of the strong coupling between plant growth and crop yield and transpiration, but not evaporation. This linkage is a basic feature of most of the dynamic, climate-driven models now widely used to simulate plant growth and yield.

As the transpiration stream can now be continuously and accurately measured under field conditions by heat pulse tracing systems, it is possible to study the relationship between the two fluxes accurately, previously referred to as the transpiration ratio and later as water-use efficiency, in natural surroundings.

Methods of Measurement

Only three of the almost innumerable methods of estimating evapotranspiration that have been described can be regarded as representing direct measurement; all other approaches require values of the constantly changing water conductance characteristics of the land surface under study, which are very difficult to measure and therefore have to be estimated from other measurements.

As the flux of water in evapotranspiration from a land surface is between two and three orders of magnitude greater than the flux of carbon to the surface in dry-matter assimilation and similarly greater than the changes in the plant’s water content, it follows that the rate of evapotranspiration can be directly measured with sufficient accuracy from changes in the mass of a sample land surface. Instruments for this purpose are known as weighing lysimeters, because measurement of drainage (the literal meaning of the word) is also required. Evapotranspiration measured with such an instrument is generally accepted as the standard against which other methods of measurement and estimation should be evaluated and calibrated. However, to serve as a standard, it is essential that the lysimeter contains a sample of the land surface whose water losses are fully representative of the surface of interest. To ensure that this is so requires that the depth and structure of the soil within the lysimeter, the height and density of the vegetation growing on it, and the microclimate above the lysimeter are indistinguishable from those of the surrounding surface it represents.

A range of weighing lysimeters meeting these demanding requirements has been described, including those with sufficient area and depth to support mature forest and orchard trees, those with undisturbed soil monoliths with water potential at their base matched to that of the surrounding soil, and instruments sufficiently sensitive to register the condensation of dew from the atmosphere. Unfortunately their cost, together with that of their installation and maintenance, has prevented the use of weighing lysimeters in sufficient numbers to measure the spatial variability of evapotranspiration.

Two meteorologic methods of directly measuring evapotranspiration are generally accepted as being of comparable accuracy with those obtained with weighing lysimeters; both of them have the advantages of portability.

The first to be used was the energy balance–Bowen ratio method, which requires measurements of the energy balance above the land surface of interest (i.e., the radiation balance above the vegetation surface and the heat flux below the soil surface) together with the ratio of the gradients of air temperature and vapor pressure above the land surface. The small size of these gradients, resulting from the need to confine them to within the shallow atmospheric layer whose microclimate is representative of the land surface of interest, is the major factor limiting the accuracy of this method of measuring evapotranspiration.

Recent technological advances have extended the use of the eddy–covariance flux method of measuring evapotranspiration which calculates the net eddy flux of water leaving the land surface from high-speed observations of the vertical component of wind speed and of humidity. However there are unresolved problems with this method, as shown by the frequent failure of energy-balance closure, i.e., the latent and sensible heat fluxes measured by this method are often less than the energy balance measured for the same land surface, suggesting that evapotranspiration is less accurately measured by the eddy–flux method than by the energy balance–Bowen ratio approach.

Magnitude of Evapotranspiration

Information on the magnitude of actual evapotranspiration is available from a number of water- and heat-balance studies of the Earth’s land surfaces. In the water-balance studies, evapotranspiration is estimated as the difference between regionally averaged measurements of precipitation and the runoff from the same area: the use of long-term, regional average annual values enabling the changing storage term to

Table 1 Evapotranspiration from the Earth's land surfaces

	Water balance estimate[a] *(mm year^{-1})*	*Heat balance estimate*[b] *(mm year^{-1})*
All land surfaces	480	442
Northern hemisphere	435	398
Southern hemisphere	572	535
Africa	582	502
North America	403	393
South America	946	882
Asia	420	417
Australasia	534	403
Europe	362	375

[a] Adapted from Baumgartner A and Reichel E (1975) *The World Water Balance. Mean Annual Global Continental and Maritime Precipitation, Evaporation and Runoff.* Amsterdam, the Netherlands: Elsevier.

[b] Adapted from Henning D (1989) *Atlas of the Surface Heat Balance of the Continents. Components and Parameters Estimated from Climatological Data.* Berlin, Germany: Gebruder Borntraeger.

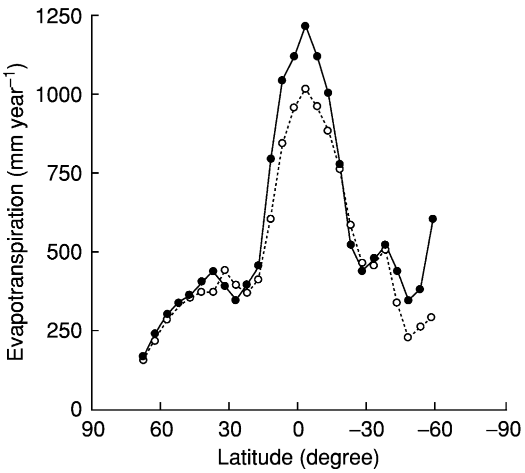


Figure 1 Evapotranspiration from the Earth's land surfaces, in millimeters per year. Empty circles, data source is Henning, 1989; full circles, data source is Baumgartner and Reichel, 1975. (Adapted from Henning D (1989) *Atlas of the Surface Heat Balance of the Continents. Components and Parameters Estimated from Climatological Data*. Berlin, Germany: Gebruder Borntraeger; Baumgartner A and Reichel E (1975) *The World Water Balance. Mean Annual Global Continental and Maritime Precipitation, Evaporation and Runoff.* Amsterdam, the Netherlands: Elsevier.)

be eliminated from the area's water balance. In the heat-balance approach, which can be used for periods less than a year, potential evapotranspiration is first calculated from climatologic measurements and then corrected to actual evapotranspiration using a function of measured precipitation.

The global, hemispheric, continental, and latitudinal values of evapotranspiration for the Earth's land surfaces presented in Table 1 and Figure 1 are taken from calculations using both the water-balance and heat-balance approaches. They agree to within 10% for the global, hemispheric, and most of the continental estimates; much of the larger differences found in the case of the latitudinal estimates can be attributed to the different values of precipitation used in the studies.

Maximum mean annual values of evapotranspiration reaching 1400 mm are found in the near-equatorial regions of South America and Southeast Asia, the specific areas coinciding with those with tropical forest cover and heavy rainfall. Maximum monthly values reaching 150 mm are found for these same areas of land cover in the mid-summer. On a daily basis, double this rate of evapotranspiration has been reported for tall vegetation subject to advective conditions, where latent energy is supplied from convective exchange with the passing air.

Control of Evapotranspiration

The major role of evapotranspiration in the Earth's water balance and humankind's growing shortage of water has focused attention on the possibility of controlling this so-called loss of water to the atmosphere: to do so in an environmentally responsible way requires an understanding of the processes controlling all the components of evapotranspiration, their interactions, and their biological significance.

In a purely physical sense, evapotranspiration can be understood as a transfer process in which water moves from the soil, a source of limited capacity and variable potential, and passes through parallel soil and plant pathways of variable conductance, into the atmosphere, a sink of variable potential but finite capacity. From this point of view evapotranspiration is controlled by either the source or sink strength, whichever is limiting. However, because the mechanisms controlling the rate of conductance of water through plants are biological and incompletely understood, this purely physical description represents a simplified view of the transpiration flux; and, when used to calculate this major component of evapotranspiration, empirical constants must be used.

The significance of the siting of the major biological control of conductance at the plant–atmosphere interface is that the plant is able to control its transpiration loss and so maintain its internal water balance at a favorable status and avoid dehydration. The major mechanism by which this is achieved is through the control of the size of the leaves' stomatal apertures: the plant's internal water status provides a feedback signal for this control. Feedforward systems of stomatal control based on the water status of the air outside the canopy and of the soil surrounding the roots have also been described.

A second component of evapotranspiration of hydrologic importance in a number of forest ecosystems is the fraction of precipitation that is intercepted by the vegetation canopy and then evaporated directly to the atmosphere without reaching the soil surface. This modifies the canopy microclimate and so reduces transpiration by an amount which depends on the ratio of the surface resistance of the wetted to the dry vegetation, and to a lesser extent on air temperature. Resistance ratios in temperate climates range from values of ≈0.2, typical of coniferous forest, to ≈0.8, typical of a number of field crops, so that the evaporation rate of intercepted water can be expected to be, respectively, 5 and 1.25 times that of transpiration from the same canopies when dry. For evergreen forests in high-rainfall climates, evaporation of intercepted precipitation can form a very substantial proportion of total evapotranspiration and is not insignificant for deciduous forest canopies even during their leafless phase.

The effect of evaporation of intercepted water on plant growth and yield is less clear than is its hydrologic significance. The use of misting and fogging irrigation systems in protected agricultural systems under conditions of high evaporative demand suggests a wholly favorable influence of the reduction in transpiration, even though the coupling of the transpiration and dry-matter production processes implies that this will have a negative effect on growth.

Precision drip-irrigation systems have also been used to reduce evaporation from the soil surface, the third component of evapotranspiration, by confining water application to below the surface and the crop row. The resulting dry surface imposes a high resistance to upward water flow, so reducing evaporation. It also modifies the crop microclimate, increasing air temperature and reducing humidity within the plant canopy, and so presumably leads to greater transpiration rates which perhaps in turn affect crop growth and yield. In many agricultural systems, cultivation of the soil surface is widely practiced with the aim of reducing evaporation from the soil as well as eliminating transpiration from weeds.

With natural land covers, the type and density of the vegetation can significantly affect the different components of evapotranspiration and these influences can be important for land-use planning and management in water catchment areas which have more than one objective. The ways in which land use influences evapotranspiration is central to the long-debated and complex question of the effect of aforestation and deforestation on water supplies and, on a wider scale, climate and rainfall.

A number of attempts have been made in both the laboratory and the field to control evapotranspiration directly. The two methods used to reduce transpiration have been stomatal closure and the reduction of plant radiation balance through increased leaf reflection. Most of these experiments were unsuccessful and in the few cases where water loss was significantly reduced this was accompanied by a corresponding reduction of plant growth.

Progress in this potentially important field requires a better understanding of the biological mechanisms controlling transpiration and of the coupling of this process to that of plant growth. With such an understanding it may be possible, and even environmentally advantageous, to use biotechnologic techniques to develop plants with lower rates of transpiration and/or higher levels of plant water use efficiency. A major benefit of such an achievement would be a reduction in irrigation requirements, which currently use two-thirds of the Earth's renewable water resources.

Climate Change and Evapotranspiration

The impact of climate change on evapotranspiration has been studied to evaluate its effect on human water supplies and indirectly on plant growth and crop yields. However, attempts to simulate future rates of evapotranspiration under future climate regimes are subject to such large uncertainties that at present they provide plausible, possible scenarios rather than estimates of known certainty which could be used for water planning.

Some indication of the magnitude of evapotranspiration changes during the last century can be obtained on the basis of the mean climate changes measured over this period. The direct effects of the 70 ppm (20%) increase in the atmospheric CO_2 concentration are twofold and are in contrast: reduced stomatal conductance and increased vegetative growth. Experiments comparing plant growth and evapotranspiration from plant stands growing under the current and a doubled CO_2 concentration indicate that these two effects cancel each other, leaving canopy water loss substantially unchanged.

Other climate changes measured during the twentieth century, attributed to the increase in CO_2 and other radiatively active gases, include an increase in mean land air temperature of $0.6 \pm 0.15°C$, with the increase in minimum temperatures twice that of the maximum. A mean overall increase of 9 mm (1%) in annual precipitation over the Earth's land surfaces was observed over the period and a decrease in global irradiance averaging 0.51 ± 0.05 W m^{-2} per year was recorded in the last 50 years. As the effect of the minor increase in air temperature would have been more than offset by the more substantial decrease in radiation, a decrease in potential evapotranspiration could be expected to have occurred. This has been verified from evaporation pan observations in the USA, the former Soviet Union, and India as well as from estimates in India and China based on the Penman equation.

There is no evidence that a corresponding change in actual evapotranspiration has occurred; this is shown by the absence of clear evidence for large, statistically significant, and widespread changes in the precipitation over the land surfaces of the Earth (the measured 1% increase is far below the error of its measurement) and in the annual streamflow and peak discharges of rivers. This lack of overall trend in the runoff component of the water balance has emerged from a major study of historical records of 142 major rivers throughout the world, possessing more than 50 years of data and representing drainage areas greater than 1000 km^2.

See also: **Energy Balance**; **Evaporation of Water from Bare Soil**; **Penman, Howard Latimer**; **Penman–Monteith Equation**

Further Reading

Baumgartner A and Reichel E (1975) *The World Water Balance. Mean Annual Global Continental and Maritime Precipitation, Evaporation and Runoff*. Amsterdam, The Netherlands: Elsevier.

Henning D (1989) *Atlas of the Surface Heat Balance of the Continents. Components and Parameters Estimated from Climatological Data*. Berlin, Germany: Gebruder Borntraeger.

Houghton JT, Meira Filho LG, Callander BA *et al.* (1996) *Climate Change 1995. The Science of Climate Change*. Cambridge, UK: Cambridge University Press.

Larcher W (1980) *Physiological Plant Ecology*, 2nd edn, pp. 258–267. Berlin, Germany: Springer-Verlag.

Molchanov AA (1963) *The Hydrological Role of Forests*. Jerusalem, Israel: Israel Program for Scientific Translations. (Translated from Russian *Gidrologicheskaya rol'lesa, Izdatel'stvo Akademii Nauk SSR Moskva 1960.)*

Monteith JL (1965) Evaporation and environment. In: Fogg GE (ed.) *The State and Movement of Water in Living Organisms,* vol. 19, pp. 205–235. Society of Experimental Biology Symposium. Cambridge, UK: Cambridge University Press.

Penman HL (1956) Evaporation: an introductory survey. *Netherlands Journal of Agricultural Science* 4: 9–29.

Penman HL (1963) *Vegetation and Hydrology.* Farnham Royal, UK: Commonwealth Agricultural Bureaux.

Rutter AJ (1975) The hydrological cycle in vegetation. In: Monteith JL (ed.) *Vegetation and the Atmosphere,* vol. 1, pp. 111–154. London, UK: Academic Press.

Stanhill G (1973) Evaporation, transpiration and evapotranspiration: a case for Ockham's razor. In: Hadas A, Swartzendruber D, Rijtema PE, Fuchs M, and Yaron B (eds) *Physical Aspects of Soil Water and Salts in Ecosystems. Ecological Studies – Analysis and Synthesis,* vol. 4, pp. 207–220. Heidelberg, Germany: Springer-Verlag.

Stanhill G (1986) Water use efficiency. In: Brady NC (ed.) *Advances in Agronomy,* vol. 39, pp. 53–85. Orlando, FL: Academic Press.

Tanner CB (1967) Measurement of evapotranspiration. In: Hagan RM, Haise HH, and Edminster TE (eds) *Irrigation of Agricultural Lands*, pp. 534–574. Agronomy Monograph No. 11. Madison, WI: American Society of Agronomy.

Waggoner PE (ed.) (1990) *Climate Change and US Water Resources*. Wiley Series in Climate and the Biosphere. New York: John Wiley.

F

FACTORS OF SOIL FORMATION

Contents

Biota
Climate
Human Impacts
Parent Material
Time

Biota

A H Jahren, Johns Hopkins University, Baltimore, MD, USA

Introduction

It is clear that organisms (such as plants, microorganisms, and soil invertebrates) that live in and on the soil both influence soil properties and are influenced by soil properties. A clear assessment of the isolated influence that a given type of biota has upon soil formation is best performed in the context of a biosequence, or series of sites in which all soil-forming factors are held constant, while the biota varies systematically. Biosequence studies are most often performed to evaluate the influence of differing plant communities, since vegetation at a site can be independent of climate, topography, parent material, and time. The most commonly performed biosequence studies assess the influence of grassland versus forest biota on soil-forming processes. Because grasses contribute a large portion of their tissue to the soil from within the soil (as below-ground tissue), grassland ecosystems result in significantly higher organic contributions to the soil, especially at depth, relative to forest ecosystems. Biosequence studies designed to evaluate the influence of deciduous versus conifer forests have also been performed, and highlight the unique role of the conifer needle litter layer as a discrete and chemically unique portion of the uppermost soil. The most recent studies have focused on the indirect effects of different vegetation on soil-forming processes, through the assessment of microbial populations or water infiltration characteristics along a biosequence.

The Biosequence

Efforts to explain soil characteristics in terms of the influence of biota are best facilitated by biosequence studies. These studies contain a series of soil profiles, across which the biotic soil-forming factor varies, while other soil-forming factors such as climate (cl), topography (r), parent material (p), and time of development (t) remain constant. Within the context of a biosequence, the effect of changing biotic factor (o) upon any soil property (s) can be assessed quantitatively, providing that all relevant factors and properties can be adequately described or assessed. The relationship between biota and soil properties can be formalized in order to highlight the functional influence that biota has on the soil, in the following definition of the biosequence:

$$s = \mathrm{f}(o)_{\mathrm{cl,r,p,t}} \qquad [1]$$

However, application of the biosequence study is problematic, for both conceptual and practical reasons. It is easy to imagine hundreds of ways in which organisms might affect soil characteristics: earthworms burrow through soil horizons, perhaps decreasing the bulk density; annual plants die and perhaps contribute organic matter to the soil profile; denitrifying bacteria might lower nitrate levels in soil solutions. However, acknowledging the influence of organisms on soil properties is fundamentally different from showing how organisms act as a soil-forming factor. The constellation of organisms present in any soil environment is dependent upon soil properties. For example, moisture levels in a given soil environment might partially control the amount of plant biomass that grows at the site. Thus the amount of vegetation, which influences soil characteristics, is actually dependent upon the soil properties. This interdependence, which can be seen in most examples of soil biota, significantly complicates the application of biosequence theory, in which all soil-forming factors are assumed to be independent. In application, it may be difficult to find sites where the biotic factor varies, while climate is constant, since most organisms are intimately sensitive to the climate in which they live. This may be overcome by focusing upon 'patchy' communities of organisms – constellations of organism type and species that exhibit regular heterogenity within a small geographical area. However, this raises another potential complication, namely how stable are the patches over the time of soil formation, since the time scales of community change and soil formation may be very different within the same ecosystem? Finally, there is the example that, at a given soil, only some plant species, of all the seeds that arrive at the site, eventually grow to maturity. Thus the actual vegetation at such a site is a subset of the potential vegetation, and may be dependent upon soil properties, or other soil-forming factors.

In light of these issues, biosequence studies are best performed where kind and composition of species change across sites (in preference to changes in quantity or yield of organisms, which may be governed by soil properties). It is also best that these studies are performed within a geographic area restricted to that which might be thought to be subject to the same overall potential biota, as well as the same climate regime. Given that these conditions can be met, biosequence studies yield valuable information about the specific effects of the biotic soil-forming factor.

The Biotic Factor

The biotic factor of soil formation (o) is multifaceted, given the myriad of organisms that conduct their life activities in the context of soil. A pedologist interested in evaluating the influence of the biotic factor might wish to specify a group of organisms to focus upon, and to acknowledge the distinct processes associated with vegetation (o_{v}), microorganisms (o_{m}), animals (o_{a}), and human activities (o_{h}). At present, there is a paucity of systematic studies evaluating the biotic factor associated with both microbes and animals on soil formation, although significant interest exists in these areas. The effect of human activities on soil properties and soil formation is treated extensively elsewhere in this volume and will therefore not be included in the discussion to follow. Given the wide environmental tolerances of many common plants, vegetation is considered to be the facet of the biotic factor most independent of climate, topography, parent material, and time in biosequence studies. Many stable and adjacent, yet highly distinct, vegetation communities exist in both the Old and New Worlds, and have formed the basis of most biosequence studies.

Vegetation as a Soil-Forming Factor

The elemental composition of soil differs from that of geologic materials in its striking enrichment of carbon and nitrogen compounds, relative to most rocks. The ultimate sources of this carbon and nitrogen are organic compounds contributed to the soil by life processes, and the death, of soil organisms such as plants, animals, and microorganisms. Of these contributions, those from vegetation dominate: plants contribute organic compounds to the soil in a variety of ways, including the senescence or necrosis of tissue, exudation or respiration from the roots, and the liberation of reproductive tissues such as pollen, seeds, and fruit. There are reasons to expect that different kinds of vegetation would contribute organic compounds to soils in different ways: deciduous trees contribute leaf material to the soil in one large annual pulse, while evergreen trees contribute leaf (or needle) tissue continuously throughout the year, as their foliage slowly turns over. In addition, the biomass produced annually as a potential contribution to the soil varies by two orders of magnitude across different vegetative biomes (Table 1).

The distribution of mass within the vegetative communities is variable as well: in most forests, the

Table 1 Leaf biomass production across vegetation and community type

Vegetation or ecosystem type	*Annual biomass production (air-dry) (kg m^{-2})*
Deserts[a]	0.04–0.12
Alpine meadow[b]	0.05–0.09
Tall- and short-grass prairie[b]	0.16–0.50
Grasslands (North American)[a]	0.09–0.90
Deciduous forest leaves[a]	0.40–0.85
Conifer forest needles[a]	0.80–3.00

[a]Source: Webb WL, Lauenroth WK, Szarek SR, and Kinerson RS (1983) Primary production and abiotic controls in forests, grasslands, and desert ecosystems in the United States. *Ecology* 64: 134–151.

[b]Source: Clements and Weaver (1924) and Swederski (1931), as reported in Jenny H (1941) *Factors of Soil Formation: A System of Quantitative Pedology*. New York: McGraw Hill. Republished in 1994 by Dover Publications, Mineola, New York.

belowground mass is about one-quarter of the total mass of the vegetation, while in some grasslands, more biomass exists belowground than above.

When plant material is contributed to the soil, it becomes soil organic matter only after the action of decomposition. The process of decomposition is mediated by microorganisms, and these bacteria, protozoa, and fungi preferentially decompose plant material of high quality. The quality of plant contribution can be quantified by its carbon-to-nitrogen ratio; high nitrogen concentrations in organic compounds are initially taken as an indicator of readily decomposable material. Therefore a low carbon-to-nitrogen ratio in plant material suggests abundant nitrogen in the substrate, that would readily decompose, and thus affect pedogenesis by forming soil organic matter. The carbon-to-nitrogen ratio of plant tissues is extremely variable from species to species, and between tissues within an individual plant (Table 2).

Conifer ecosystems contribute more tissue to the soil annually than do deciduous or grassland ecosystems, yet this tissue is of very low quality with respect to decomposition. In contrast, the A horizon of grassland soils exhibits organic matter with high quality, relative to deciduous or conifer systems. From this, we might hypothesize that a biosequence evaluating soils forming under conifer forests, deciduous forests, and grasslands would yield the following observations: (1) the conifer soil might have a thick undecomposed upper horizon, with relatively low levels of organic matter below; (2) the grassland might have well-decomposed and well-distributed organic matter throughout the soil profile; and (3) the deciduous system would exhibit properties intermediate to the conifer and grassland profiles.

In addition to the direct effect of organic contribution to the soil by vegetation, there are several

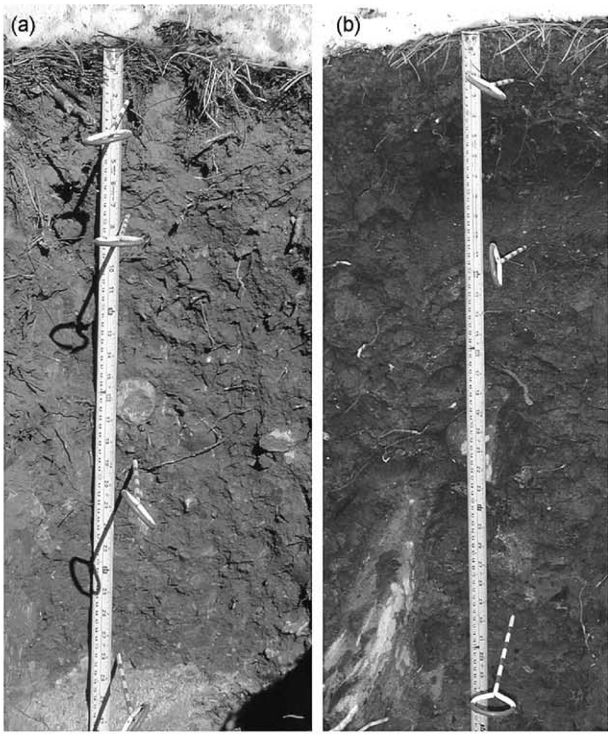
Table 2 Carbon-to-nitrogen ratio in plant tissue and resulting organic horizon

Vegetation or organic horizon type	*Carbon-to-nitrogen ratio (unitless)*
Shoot, root and leaf[a]	
Conifer, deciduous, grass and other	5–80
Pine needles (*Pinus coulteri*)[b]	57
Oak leaves	
Quercus pubescens[a]	45
Quercus dumosa[b]	39
Conifer forest A horizon	
Pine (*Pinus ponderosa*)[c]	33.5–54.9
Pine (*Pinus coulteri*)[b]	16.1
Fir (*Abies concolor*)[c]	26.0–42.3
Hardwood forest A-horizon	
Oak (*Quercus dumosa*)[b]	19.5
Mixed oak-hickory[c]	12.6
Bluestem grassland A-horizon[c]	11.7

[a] Source: Raschi A, Miglietta F, Tognetti R, and van Gardingen PR (1997) *Plant Responses to Elevated CO$_2$*. Cambridge: Cambridge University Press.

[b] Source: Quideau SA, Graham RC, Chadwick OA, and Wood HB (1998) Organic carbon sequestration under chaparral and pine after four decades of soil development. *Geoderma* 83: 227–242.

[c] Source: Bodman (1935) and Rost (1918), as reported in Jenny H (1941) *Factors of Soil Formation: A System of Quantitative Pedology*. New York: McGraw Hill. Republished in 1994 by Dover Publications, Mineola, New York.



(a) Fine-loamy smectitic mesic typic Haplustalf
Vegetation: pine, juniper, cactus, yucca, cedar
Parent material: volcanic (basalt)

Depth (cm)	*Horizon*	*Color*	*% Clay*	*Structure*	*pH*
0–6	A	7.5 YR 4/3	6	1fgr	6.5
6–19	Bt$_1$	7.5 YR 4/4	18	2fsbk	6.5
19–48	Bt$_2$	7.5 YR 3/3	25	3msbk	7.0
48–71	Bt$_3$	7.5 YR 3/4	35	3msbk	7.0
71–84+	C	7.5 YR 5/6		massive	7.0

(b) Very-fine smectitic mesic typic Haplustert
Vegetation: pine, juniper, yucca, various grasses
Parent material: volcanic (basalt)

Depth (cm)	*Horizon*	*Color*	*% Clay*	*Structure*	*pH*
0–5	A	7.5 YR 2.5/1	18	3mgr	7.0
5–26	ABtss	7.5 YR 3/1	50	3msbk	7.0
26–76	Btss$_1$	7.5 YR 3/2	80	3mabk	7.0
76–90+	Btss$_2$	7.5 YR 3/3	80	3mabk	7.0

Figure 1 Two soil profiles from the Kaibab national forest in Northern Arizona: (a) forest soil; (b) soil from a forest containing some grasses. The influence of grass vegetation can be seen as enhanced organic content at all depths in the profile.

indirect effects of vegetation that may influence soil properties. The decomposition of plant tissues produces humus, which contains many organic acid groups, including fulvic acids. The low pK of these acids may accelerate weathering processes and the production of secondary minerals from primary minerals. In this way the decomposition of plant tissues might facilitate pedogenic processes, such as the *in situ* production of secondary clay. The needles of conifer trees are well-known to result in low-pH soil conditions. The leachate of this evergreen foliage is usually ~2 units lower in pH value than leachate moving through deciduous litter, liberating organic acids that may participate in weathering processes. Vegetation type may indirectly influence the hydrologic properties of the soil: earthworm activity specific to oak vegetation results in water-stable aggregates that promote infiltration and aeration, and reduce erosion relative to nonaggregated soils. Also, the extensive fine-root systems present in grassland soils results in high percolation rates as water readily infiltrates subsurface horizons through a network of micropores.

Grassland–Forest Biosequence

The best-studied biosequence involves the 'Prairie-timber transition zone,' which is a coexisting grassland and forest ecosystem with abrupt transitions between the two vegetation types. A belt, or zone, of such transitions has been described in European Russia, Siberia, Canada, and the Midwestern USA. Within these zones, both grasslands and forests occur on the uplands as well as in the drainage channels, and rainfall is sufficient to support forest growth throughout the entire region. The sharp transition that forms the basis of the biosequence has puzzled ecologists and no satisfactory hypothesis has been put forward to explain the pedologically fortuitous configuration. Recent intruiging suggestions invoke long-term and long-range feedbacks between the grassland and the forests, perhaps via herbivores, acting to stabilize the discrete, but adjacent, ecosystems.

The influence of grasses on soil profile development can be partially isolated by examining soils under forests, and comparing them to those under forests containing some grasses near the grassland–forest transition (Figure 1).

Grass tissue is almost evenly distributed above- and belowground, therefore annual grasses readily contribute their tissue to the soil profile, not only at the surface, but from within the epipedon and subsurface horizons as well. Since increased organic matter results in dark soil colors, soils influenced by grasses in this way exhibit lower value and chroma than their forest counterparts, especially at the lower depths of the profile (Figure 1).

Various biosequence studies comparing grassland soils to forest soils reveal significantly higher levels of percentage carbon up to 1 m depth. The quality of this organic matter is higher under grasslands, with significantly lower carbon-to-nitrogen ratios than under forests, especially near the soil surface. Soil pH values are generally lower in forest soils, probably reflecting the influence of conifer litter, which results in acidic leachate. These differences in pH are often diminished with depth, however. Many studies have revealed more clay and enhanced structure under grassland vegetation, relative to forest soils, although

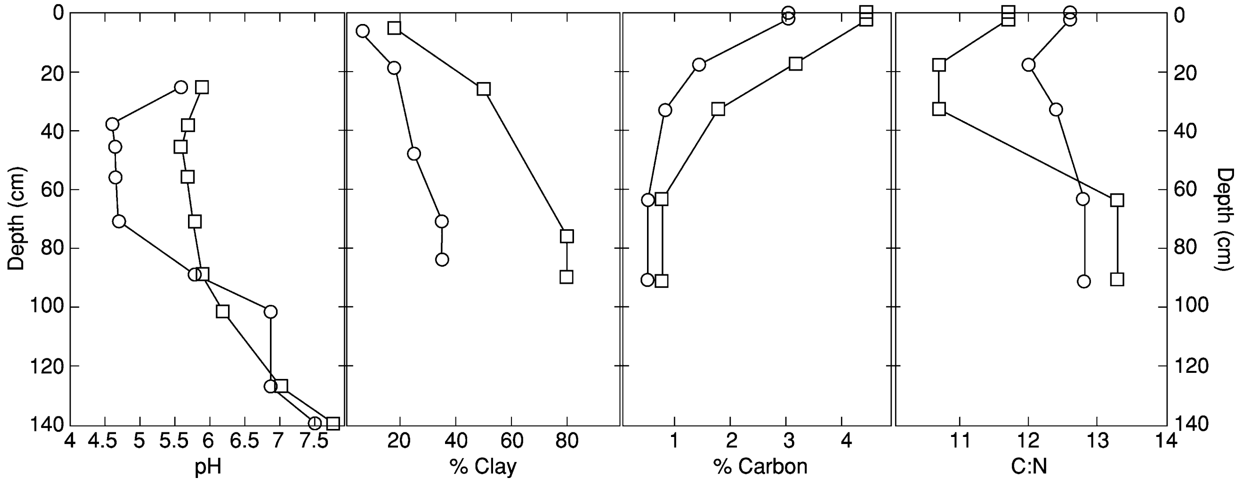


Figure 2 Comparison of soil characteristics below forest (circles) and grassland (squares) ecosystems, as reported in several biosequence field studies. pH after Jenny H (1941) percentage clay after Figure 1; percentage carbon and C:N after Rost (1918), as reported in Jenny H (1941) *Factors of Soil Formation: A System of Quantitative Pedology*. New York: McGraw Hill. Republished in 1994 by Dover Publications, Mineola, New York.

there is no agreement as to what mechanism acts to cause this relationship (Figure 2).

Conifer Forest–Deciduous Forest Biosequence

Recent studies focusing on soil development after the installation of lysimeters and using other long-term monitoring techniques have shed light on the effects of conifer versus deciduous forest ecosystems on soil development. An important example of such a biosequence resides within the San Dimas experimental forest in the San Gabriel mountains of California. Four decades of soil development have resulted in significantly higher soil carbon and nitrogen content under oak (deciduous) systems than under pine (conifer). However, such results may be highly species-specific: comparison of soils forming under beech (deciduous) versus spruce (conifer) vegetation along a biosequence in Denmark showed much higher levels of soil organic carbon under the conifer forest, at all soil depths (Figures 3 and 4).

The most notable result of deciduous versus conifer forest biosequence studies is that the conifer

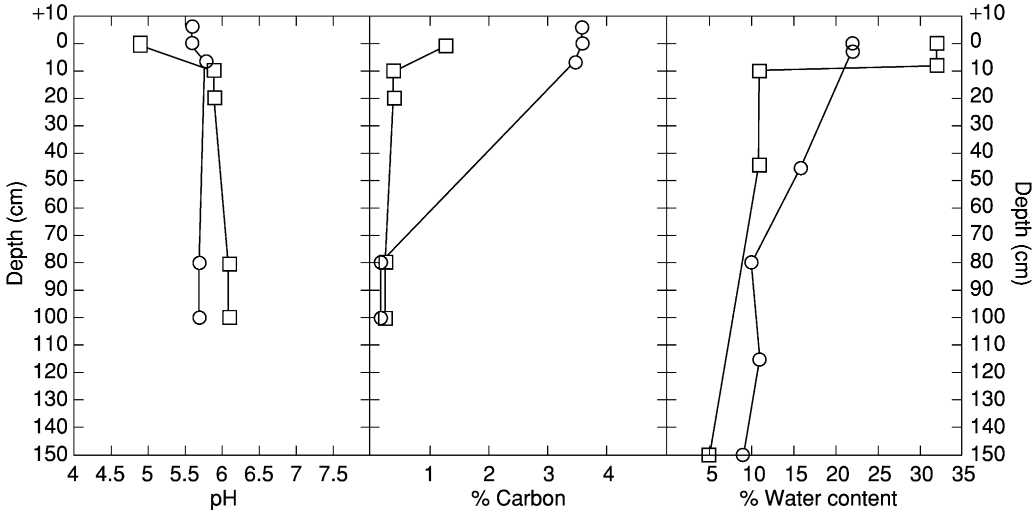


Figure 3 Comparison of soil characteristics below hardwood (circles) and conifer (squares) forests, as reported in several biosequence field studies. pH and percentage carbon after Graham RC, Ervin JO, and Wood HB (1995) Aggregate stability under oak and pine after four decades of soil development. *Soil Science Society of America Journal* 59: 1740–1744 (1995); percentage water content after Ekelund F, Rønn R, and Christensen S (2001) Distribution with depth of protozoa, bacteria and fungi in soil profiles from three Danish forest sites. *Soil Biology and Biochemistry* 33: 475–481.

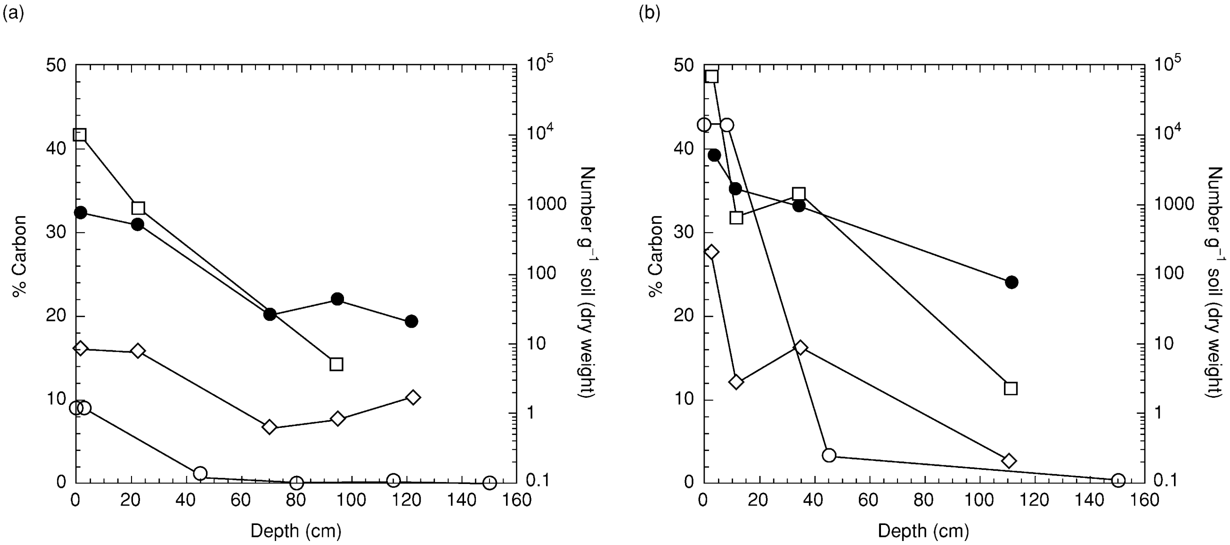


Figure 4 Comparison of microbial populations within (a) beech and (b) spruce forest soils. Open circles, percentage carbon; squares, protozoa; diamonds, bacteria ($\times 10^8$); filled circles, fungal hypha. Biosequence data after Ekelund F, Rønn R, and Christensen S (2001) Distribution with depth of protozoa, bacteria and fungi in soil profiles from three Danish forest sites. *Soil Biology and Biochemistry* 33: 475–481.

ecosystem dramatically influences the top ~10 cm of the soil, while the influence of the deciduous ecosystem may be more diffuse. A dramatic decrease in pH, increase in percentage carbon, and a large increase in water content, confined to the upper decimeter of soil, has been observed in soils forming under conifer forests by various biosequence studies (Figure 4). This may highlight the importance of the conifer needle litter layer as a stable and chemically unique portion of developing soils.

New areas of study have attempted to quantify microbial populations along biosequences and evaluate the effect of vegetation type on the microorganism community. Microorganisms control many soil processes, such as decomposition, and thus could be considered to affect soil processes indirectly as the result of changing vegetation along a biosequence. Deciduous versus conifer forest biosequences in Denmark show bacterial, fungal, and protozoan populations an entire order of magnitude higher under spruce (conifer) ecosystems, relative to beech (deciduous) forests. Although microbial populations decrease with profile depth under both types of forest, populations remain significantly higher throughout the profile forming beneath spruce (conifer) forests (Figure 4). Further study is needed to assess the different species composition and ecosystem significance of these differing microbial communities, but it is clear that vegetation biosequences show a wide range of both direct and indirect effects on soil formation that is highly variable across vegetation type.

List of Technical Nomenclature

% carbon	Carbon content (carbon g per 100 g of soil)
% clay	Clay content (total number per 100 particles that are clay-sized)
	'Clay-sized' (<2 μm)
% water content	Water content (water g per 100 g of soil)
B	Biomass production (kg m^{-2})
C:N	Carbon–nitrogen ratio (unitless)
pH	−Log [H^{+}] (unitless)

See also: **Biodiversity**; **Classification of Soils**; **Factors of Soil Formation:** Climate; Human Impacts; Parent Material; Time; **Jenny, Hans**

Further Reading

Bicki TJ, Fenton TE, Luce HD, and Dewitt TA (1988) Comparison of percolation test results and estimated hydraulic conductivities for Mollisols and Alfisols. *Soil Science Society of America Journal* 52: 1708–1714.

Ekelund F, Rønn R, and Christensen S (2001) Distribution with depth of protozoa, bacteria and fungi in soil profiles from three Danish forest sites. *Soil Biology and Biochemistry* 33: 475–481.

Graham RC, Ervin JO, and Wood HB (1995) Aggregate stability under oak and pine after four decades of soil development. *Soil Science Society of America Journal* 59: 1740–1744.

Jenny H (1941) *Factors of Soil Formation: A System of Quantitative Pedology*. New York: McGraw-Hill. Republished in 1994 by Dover Publications, Mineola, New York.

Quideau SA, Graham RC, Chadwick OA, and Wood HB (1998) Organic carbon sequestration under chaparral and pine after four decades of soil development. *Geoderma* 83: 227–242.

Raschi A, Miglietta F, Tognetti R, and van Gardingen PR (1997) *Plant Responses to Elevated CO$_2$*. Cambridge: Cambridge University Press.

Roy J, Saugier B, and Mooney HA (2001) *Terrestrial Global Productivity*. London: Academic Press.

Schlesinger WH (1997) *Biogeochemistry: An Analysis of Global Change*, 2nd edn. London: Academic Press.

Webb WL, Lauenroth WK, Szarek SR, and Kinerson RS (1983) Primary production and abiotic controls in forests, grasslands, and desert ecosystems in the United States. *Ecology* 64: 134–151.

Climate

O C Spaargaren, ISRIC/WDC for Soils, Wageningen, The Netherlands
J A Deckers, Catholic University of Leuven, Leuven, Belgium

Introduction

Climate is a key factor in the formation and potential use of soils. It influences vegetation and soil fauna, transforms rock into regolith, and regulates chemical reaction rates and transport flow of substances in soils. Climate plays a crucial role in human land-use options, determining the length of growing seasons and the crops that can be grown. Soils, on the other hand, act as a climatological archive, recording the conditions and changes in the recent past. These examples show the intricate interaction between soils and climate.

Key climatic components in relation to soil are the seasonal amount of precipitation, the amount of evaporation, the temperature and its fluctuations, wind, and solar radiation. The amount and seasonal distribution of precipitation, together with evaporation (or, combined with the influence of the vegetative cover, the evapotranspiration), determines the moisture content and the net flux of water in a soil. The temperature and the temperature variations (daily, seasonally, annually) largely regulate the rate of decomposition or alteration of organic and mineral substances in soils, and the faunal activity. Wind plays a key role in dry areas, and solar radiation is responsible for the necessary influx of energy.

Because climate has such an impact on soils, it is not surprising that the first realistic concepts of soil formation and soil distribution were linked to climatic conditions and climate belts (see example in Table 1). The concept of 'zonal soils,' i.e., soils that are in equilibrium with climate (and vegetation type), was first introduced by V.V. Dokuchaev during the 1880s in Russia. Rode describes Dokuchaev's concept of zonality of soils as:

> The zonal occurrence on the globe . . . consists in the fact that the areas occupied by soils of various types stretch in broad belts, on the whole latitudinally, like the climate zones. Such regularity is indeed the best proof of the very great role which climate plays in soil formation.

A very good correlation exists between the climatic, vegetation, and soil belts in Belarus, the Ukraine, and the European part of Russia. This is brought about by the fairly uniform parent material (somewhat sandier in the north, more loamy (loess) in the south), the regular increase in temperature from the north to the south, and a rainfall pattern showing an increase followed by a decrease also from north to south. It results in a latitudinal sequence of tundra soils (Gelisols), podzolic and forest soils (Spodosols, Alfisols), chernozemic and chestnut soils (Mollisols), desert soils (Aridisols), cinnamon soils (Ultisols), and Zheltozems (Oxisols), corresponding with arctic, boreal, steppic, desert, subtropical, and tropical climates.

Table 1 The relation between precipitation and the saturation deficit of the vapor pressure to correlate climate and soil type in Europe

Climate	*Soil type*	*Precipitation/ saturation deficit*[a] *(mm)*
Arid	Desert (Aridisols)	0–100
Semiarid	Chestnut brown (Mollisols)	100–275
Semihumid	Chernozem (Mollisols)	125–375
Humid	Brown forest (Alfisols, Ultisols)	275–500
Perhumid, cool	Podzol (Spodosols)	375–1200
Perhumid, cold	Tundra (Gelisols)	500–600

[a] The precipitation–saturation deficit quotient (in millimeters) is defined as the amount of precipitation divided by the saturation deficit, i.e., the saturated vapor pressure minus the actual vapor pressure.

Adapted from Meyer A (1926) Über einige Zusammenhänge zwischen Klima und Boden in Europa. *Chemie der Erde* 2.

The soil pattern in North America does not follow latitudinal belts. This is because, although the parent material is quite homogenous, at least in the interior part of the USA and Central Canada, the rainfall gradient is increasing from west to east, perpendicular to the temperature gradient, which is from north to south. Here Chernozems, Chestnut brown soils and Desert soils occur at the same latitude, unlike the pattern in Belarus, the Ukraine, and European Russia. Therefore, the concept of zonality has been difficult to apply in North America. This can be noticed from the soil maps published in the USA during the 1930s, and by the fact that, in developing the groupings of soils at that time, more consideration was given to soil characteristics and broad environmental conditions than climate alone. Only the highest category (Category VI, Solum Composition Groups), the distinction between the Pedocal and Pedalfer groups, being the presence or absence of accumulated secondary lime, respectively, reflects some climatic relationship (more evapotranspiration versus more leaching). This, however, only holds true if calcium carbonate is present in the soil system.

While in Russia the concept of zonality, with its close link to climate, has continued to govern the subsequent soil classification systems, the American Soil Taxonomy, introduced in 1975 and modified in 1999, focuses more on measurable soil properties and materials, and introduced soil climatic criteria (soil moisture regime, soil temperature regime) at the penultimate-highest (suborder) level. These soil climatic criteria still have some relationship to climatic conditions at the Earth's surface, concerning well-drained soils. For example, a xeric soil moisture regime (i.e., total soil moisture control section is, in normal years, dry for 45 or more consecutive days in the 4 months following the summer solstice and moist for 45 or more days in the 4 months following the winter solstice) is a typical moisture regime in areas of Mediterranean climates (moist and cool winters, dry and warm summers). Similarly, the (iso)hyperthermic soil temperature regime (i.e., mean annual soil temperature at 50 cm depth is 22°C or higher; 'iso-' indicating a temperature fluctuation of less than 6°C) is closely related to areas with a tropical climate.

As a basis for this discussion on the relation between soils and climate, the Köppen climate classification is used. This widely used, well-known, world-wide classification system combines temperature and precipitation into five main climatic zones, viz. tropical climates (A), dry climates (B), warm temperate climates (C), cool temperate climates (D),

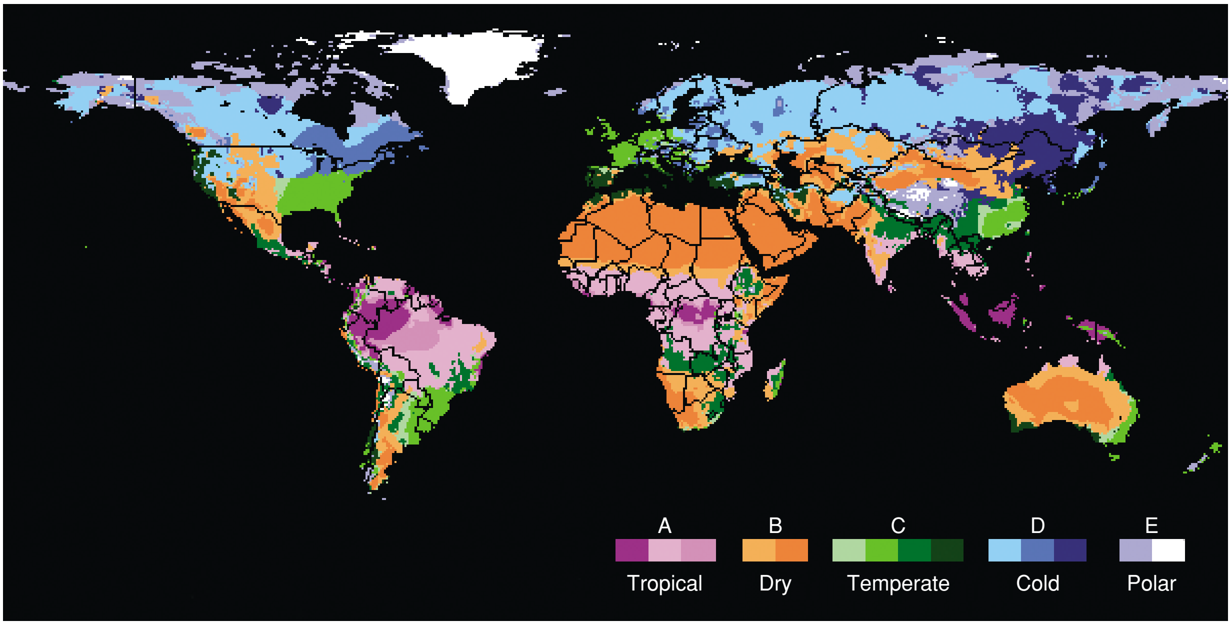


Figure 1 Koeppen's climate classification. Reproduced with permission from FAO Environment and Natural Resources Service, Sustainable Development Department (www.fao.org).

and polar climates (E) (Figure 1). Soils are classified according to both the second edition of *Soil Taxonomy* (ST) and the *World Reference Base for Soil Resources* (WRB). (*See* **Classification Systems:** FAO.)

Soils in Tropical Climates

The tropical climate in the Köppen system is defined as having an average temperature in every month of above 18°C. There is no winter season, and annual rainfall is large and exceeds the annual evaporation. Subdivision of this climate type is based on the absence or presence of a significant dry period, and the occurrence of monsoon rains (Af, tropical rainforest; Am, tropical rainforest with monsoon type of precipitation; Aw, tropical savannah).

The high temperature and precipitation of the almost-permanently moist tropical rainforest climates (Af and Am types of climate; (per)udic soil moisture and (iso)hyperthermic soil temperature regimes) yield an intense and deep chemical weathering and strong leaching of solutes. As a result primary minerals such as feldspars and micas are dissolved or transformed, leading to the loss of basic elements (Ca, Mg, K, Na) and silica, and the formation of such secondary minerals as kaolinite ($Al_2Si_2O_5 \cdot 2H_2O$), goethite (α-FeOOH), and gibbsite (α-Al(OH)$_3$). (*See* **Clay Minerals.**) On rare occasions also lepidocrocite (γ-FeOOH) may be formed. This process, known as ferralitization, is widespread in tropical regions, particularly on the older geomorphic surfaces. It gives rise to acid soils with low cation exchange capacity (CEC $<16\ \mathrm{cmol_c^{-1}}\ 100\ \mathrm{g^{-1}}$ clay) due to the prevalence of low-activity clay, and even soils with variable charge if iron hydroxides and/or aluminum hydroxides are dominating. Such soils are known as Oxisols (ST) or Ferralsols (WRB), and kandic Great Groups of the Ultisols (ST), comparable with Acrisols and Nitisols (*pro parte*; WRB). On younger deposits, not yet as strongly leached as the soils on the older geomorphic surfaces, chemical weathering of primary minerals such as smectites, vermiculites, and chlorites produces large amounts of exchangeable aluminum, resulting in Paleudults (ST) or Alisols (WRB). In sandy deposits, the strong leaching leads to the formation of giant Spodosols (ST) or Podzols (WRB) (Figure 2).

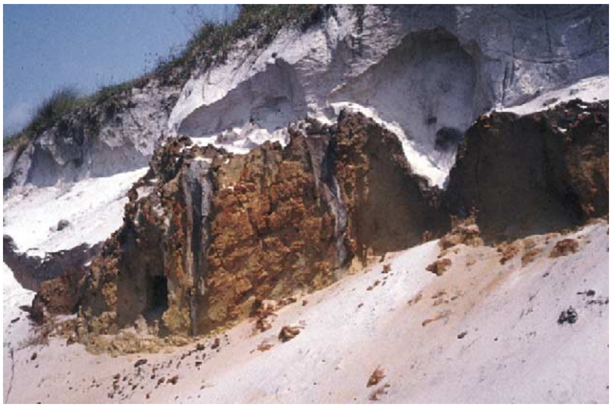


Figure 2 Giant Podzol in Malaysia. (Reproduced with permission from the International Soil Reference and Information Center, ISRIC.)

Under tropical savannah climates, which have a pronounced dry season (Aw type; ustic soil moisture and hyperthermic temperature regimes), weathering and leaching are less intense compared with the (almost) permanently moist tropical climates described above. This is due to less rainfall, the upward movement of solutes during the dry period preceding the evaporation of the soil moisture, and frequent additions of airborne dust. The Aw type of climates often border the climates characterized as dry (B). Dust derived from the deserts are can be transported over long distances. Well known are the Harmatan trade winds in West Africa, frequently obscuring the sunlight during daytime, and the Sahara dust traveling across the Atlantic Ocean to the Caribbean islands. This results in annual addition of dust, frequently calcareous, to the soils surrounding the dry regions of the Earth.

At the same time, the process of desiccation plays an important role. During the long, dry period, soils dry out completely (characteristic for the ustic soil moisture regime), and hydrated forms of iron and aluminium hydroxides may lose at least part of their crystal water. Goethite (α-FeOOH) is transformed into hematite (α-Fe_2O_3), following the reaction $2FeOOH - H_2O \rightarrow Fe_2O_3$. Similarly, one would expect dehydration of gibbsite (α-$Al(OH)_3$) into boehmite (α-AlOOH), but this has not yet been demonstrated with certainty. (*See* **Metal Oxides**.)

The dehydration of goethite and the formation of hematite are responsible for the striking difference in color between the soils of the humid tropical regions and the soils of the savannah-type climates. Soils under humid tropical climates are dominantly yellow or yellowish-brown (the color of goethite), whereas the soils under tropical savannah climates are dominantly red (the color of hematite). This reddening is known as the process of rubefaction and occurs in all climates with a pronounced dry season.

Because of the less-intense weathering and leaching, and the addition of airborne dust, the soils under the tropical savannah climate are less acid and have higher base saturation compared with their counterparts under the tropical rainforest climate. Here we find, next to Oxisols (ST) or Ferralsols (WRB), Kandiustalfs or Lixisols (Figure 3) and Nitisols (*pro parte*) (WRB), and Paleustalfs and Haplustalfs (ST), or Luvisols (WRB). Other soils frequently encountered are Natrustalfs (ST) or Solonetzes (WRB), where sodium plays an important role; and Albaqualfs (ST) or Planosols (WRB), which have pronounced water stagnation at the surface during the rainy season. In level areas rich in shrink-swell clays (montmorillonite), Vertisols (ST and WRB) occur.

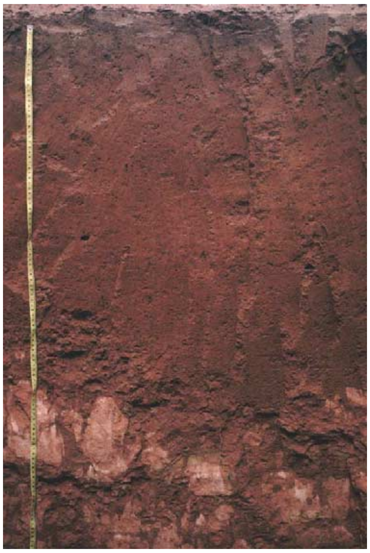


Figure 3 Deep red savannah soil over sandstone in northern Ghana. (Reproduced with permission from ISRIC.)

Soils in Dry Climates

Dry climates (B) in the Köppen climatic classification are distinguished from the more humid climates by the formula:

$$R = \frac{1}{2}T - \frac{1}{4}PW$$

in which R is rainfall (in inches), T is temperature (in degrees Fahrenheit) and PW is percentage annual rainfall in winter (half-year). In the dry climates potential evaporation exceeds precipitation on the average throughout the year. A distinction is made between a steppe climate (BS: ustic soil moisture regime, variable-temperature regimes) and a desert climate (BW; aridic or torric soil moisture regime, variable-temperature regimes), the first one being described by Köppen as semiarid, the latter one as arid. Half-R in the above formula designates the boundary between the two climates. In addition, both dry-hot (h) and dry-cold (k) climates are distinguished, with an additional option to indicate influence of fog (n) from cold ocean currents upon coastal deserts.

Evaporation exceeding precipitation leads to an upward movement of solutes in many soils of the dry climates. In steppe climates this leads to enrichment with secondary calcium carbonate ($CaCO_3$) and gypsum ($CaSO_4 \cdot 2H_2O$), whereas in desert climates even more soluble salts, such as mirabilite ($Na_2SO_4 \cdot 10H_2O$), thenardite (Na_2SO_4), hexahydrite ($MgSO_4 \cdot 6H_2O$), nahcolite ($NaHCO_3$), soda ($Na_2CO_3 \cdot 10H_2O$), trona ($Na_3CO_3HCO_3 \cdot 2H_2O$), and halite (NaCl), may precipitate. The accumulation of soluble salts is caused by the process known as 'salinization.'(*See* **Salination Processes**.)

The scarcity or even lack of vegetation means that wind action plays a significant role in these climates and has an important bearing on the soils. Wind speed not only regulates the amount of evaporation, also wind action is responsible for removal of fine particles (dust storms), and the accumulation of sands in dune complexes. Where fine particles are removed, often a gravely surface is left behind, giving rise to a so-called desert pavement (Figure 4). The gravels at the surface are shaped by scouring sand particles ('ventifacts') and have a shiny surface ('desert varnish').

Where cold ocean currents meet desert shores, such as the Benguela current along the coast of northwestern South Africa and Namibia, mist drifts inland, carrying moisture and solutes from the sea. This gives rise, over time, to deposition of, particularly, sulfates. The vast areas of accumulated gypsum, up to 150 km inland, in this region can largely be attributed to the influence of mist.

Soils in desert and steppe climates do not show much development apart from the accumulation of carbonates, gypsum, or salts. Under desert climates, various types of Aridisols (ST) occur, in particular Calcids (ST) or Calcisols (WRB), Gypsids (ST) or Gypsisols, and Durids (ST) or Durisols (WRB). Other soils are dominantly Entisols (ST), the sandy ones being Torripsamments (ST) or Arenosols (WRB), the others being Torriorthents (ST) or Regosols (WRB). Where salts have accumulated, Salids (ST) or Solonchaks (WRB) are found. Under steppe climates, where dominantly an ustic soil moisture regime prevails, Entisols (ST) or Regosols (WRB), Inceptisols (ST) or Cambisols (WRB), and Mollisols (ST) or Kastanozems and Chernozems (WRB) occur. Depending on the accumulated secondary material, Durustepts and Durustolls (ST) or Durisols (WRB), Calciustepts and Calciustolls (ST) or Calcisols and Kastanozems (WRB), and Argiustolls (ST) or Chernozems (WRB) occur. Other soils are Natrustolls and Natrustalfs (ST) or Solonetz (WRB), and Ustalfs (ST) or Luvisols (WRB).

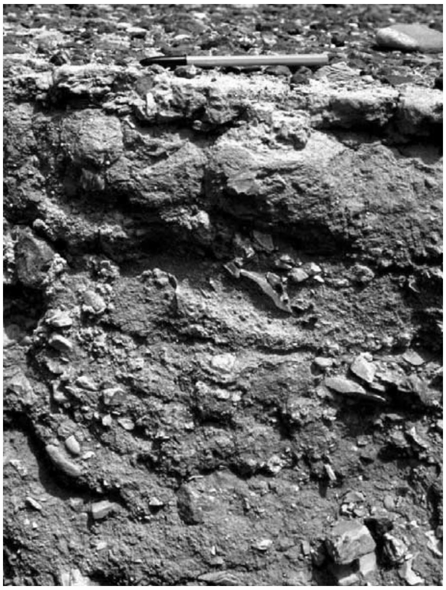


Figure 4 Typical desert soil with pavement in Israel. (Reproduced with permission from ISRIC.)

Soils in Warm Temperate Climates

Warm temperate climates (C) are defined in the Köppen climate classification as having a coldest month with an average temperature below 18°C but above −3°C. Thus there is a distinct summer and winter season. Subdivision within the warm temperate climates is based on rainfall pattern (permanently moist (Cf), dry winter (Cw), or dry summer (Cs), and maximum temperatures during the summer season (a, hot summer (warmest month, above 22°C); b, warm summer (warmest month, below 22°C); c, cool, short summer (less than 4 months above 10°C). The permanently moist climate type (Cf) coincides roughly with the (per)udic soil moisture regime, the one with a dry summer (Cs) with the xeric soil moisture regime. Soil temperature regimes are usually mesic and thermic, but areas of the warm temperate climate with a cool, short summer (Cfc) may experience a frigid soil temperature regime.

Warm temperate climates have, either seasonally or permanently, an excess of rainfall. This gives rise to chemical weathering and leaching, temporarily or all year around. Due to the large fluctuation in temperature and the varying rainfall patterns, chemical weathering rates will be variable as well. However, a net transport of substances, either in solution or as colloids, from the upper parts of the soils to the lower parts or to the groundwater is the common denominator in most of the soils in warm temperate climates.

The chemical weathering of noncalcareous parent materials leads to the release of free iron hydroxides (amorphous iron hydroxide, $Fe(OH)_3$; goethite, α-FeOOH) and the formation of clay, particularly illite. This clay formation is mainly the result of microdivision from muscovite, accompanied by partial or total loss of fixed K^+ ions. The release of free iron hydroxides gives the soils in warm temperate climates their characteristic yellowish-brown color, a process also known as 'brunification.' In warm temperate climates with a distinct dry summer (Cs), dehydration of the hydroxides leads to more red colors (rubefaction) due to the formation of hematite (α-Fe_2O_3). The chemical weathering of calcareous parent materials starts with the dissolution and removal of the carbonates present. It is only after the process of decalcification that other processes, such as brunification, will start.

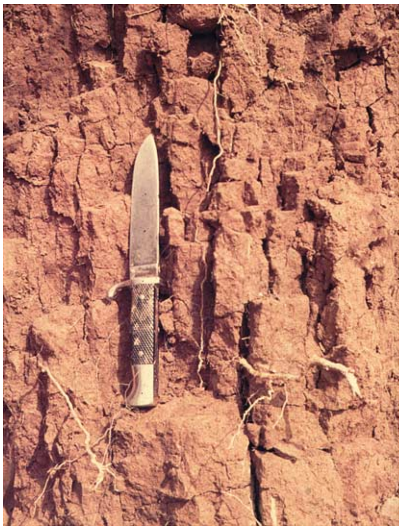


Figure 5 Red Mediterranean soil with well-developed argillic horizon in Greece. (Reproduced with permission from ISRIC.)

Of particular importance in these climates is physical translocation of clay particles by gravity water from the upper soil horizons to the lower soil horizons. This process, called clay illuviation, leads to clay-enriched, subsurface horizons, known as argillic (ST) or argic (WRB) horizons (Figure 5). The process is enhanced by a soil environment relatively poor in Ca^{2+} and Al^{3+} ions, both of which tend to flocculate the clay particles, thus inhibiting the dispersion necessary for the translocation of clay. It is further favored by desiccation of the soil, through which larger pores and cracks are created, and by the occurrence of heavy rain showers, creating the low-electrolyte environment needed for the translocation of clay. (*See* **Flocculation and Dispersion**.)

Soil types of the warm temperate climates are dominantly Alfisols, Inceptisols, and Spodosols (ST) or Luvisols, Cambisols, and Podzols (WRB). If dryness in the summer season prevails (climate types Csa and Csb), also soils with accumulation of calcium carbonate (Calcixerepts (ST) or Calcisols (WRB)), soils rich in sodium (Natrixeralfs (ST) or Solonetzes (WRB)), and shrink–swell clays (Vertisols (ST and WRB)) occur. In parts with high rainfall all year round, peat soils (Histosols (ST and WRB)) are found, together with soils having a thick, acid, and humus-rich surface horizon (Dystrudepts (ST) or Umbrisols (WRB)). Toward the boundary with cool temperate climates (D) Mollisols (ST) or Chernozems (WRB) are encountered.

Soils in Cool Temperate Climates

The cool temperate climate (D) in the Köppen climate classification is defined as having a coldest month with an average temperature below −3°C and an average temperature of the warmest month above 10°C. Distinction is made between a cool temperate climate, moist all-year round (Df) and climates with a dry winter (Dw). Further differentiation is based on maximum temperatures during the summer season: a, hot summer (warmest month, above 22°C); b, warm summer (warmest month, below 22°C); c, cool, short summer (less than 4 months above 10°C); d, very cold winter (coldest month, less than −38°C). The cool temperate climate approximately coincides with areas having frigid and (partly) cryic soil temperature regimes.

Large areas experiencing cool temperate climates have been formerly glaciated. This has resulted in deposition of glacial and periglacial deposits such as boulder clays, outwash material, fluvial deposits of braided river systems, lake sediments, loess, and cover sands. The heterogeneity of these sediments, in combination with the conditions of cool temperate climates, results in a variety of soil-forming processes.

Of importance in the cool temperate climates are the long, cold winters, during which the soils are frozen and covered by snow. In areas with slowly permeable parent materials (boulder clays, lake sediments, etc.) meltwater stagnates during springtime, causing waterlogging and temporary reducing conditions. This gives rise to dissolution and removal of iron, decline in soil structure in the surface horizons, and acidification. On better-drained parent materials (cover sands, braided river deposits, etc.), water stagnation may still be present due to the frozen subsoil, but not as long and as severe as on poorly drained parent materials.

The low temperature also hampers biological activity in the soils. Only during the warm summers will soil life be active and contribute to the process of homogenization. Consequently, humification and bioturbation are slow, particularly in the Dfc, Dfd, Dwc, and Dwd climates. On the other hand, where summers are sufficiently long and warm, and winters not very severe (Dfa, Dfb, Dwa, and Dwb climates), biological activity in the soils contributes greatly to the processes of homogenization and humification. This is especially so under grasslands with a high production of organic matter. (*See* **Grassland Soils**.) Burrowing soil animals homogenize the soils to a considerable depth, sometimes up to 2 m, and thoroughly mix the organic matter with the mineral soil constituents.

The processes described above lead, on the one hand, to well-differentiated soils with poor humification suffering strongly from water stagnation and, on the other hand, to well-homogenized and well-drained soils rich in organic matter. Dominant soils

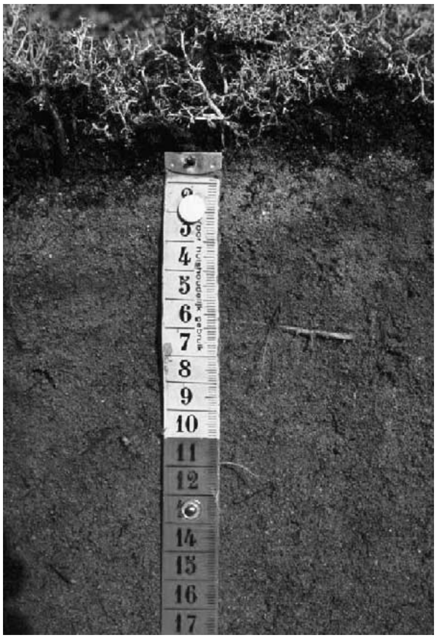


Figure 6 Mini-Podzol under tundra vegetation in northern Sweden. (Reproduced with permission from ISRIC.)

in the former group are Albaqualfs, Glossaqualfs, Epiaqualfs, Albaquults, and Epiaquults (ST) or Albeluvisols and Stagnic Luvisols (WRB). Mollisols (ST) or Chernozems (WRB) represent the group of well-homogenized and well-drained soils rich in organic matter.

The climate-induced coniferous forest in the colder part of the cool temperate climate contributes significantly to the development of acid soils, especially on sandy deposits. Here the process of cheluviation takes place, whereby organic acids, released from the vegetation debris, complex aluminium and iron that are subsequently leached down and deposited in lower parts of the soils. The result is the development of Spodosols (ST) or Podzols (WRB). Given the young age – geologically speaking – of the deposits and the slowness of the process because of the low temperatures, corresponding soils are often only weakly, though distinctly developed (Figure 6).

Soils in the Polar Climates

The polar climates (E) in the Köppen climate classification are defined as having an average temperature of the warmest month of less than 10°C. Two subclimates are recognized: the tundra climate (ET) and the climates of perpetual frost (EF). They coincide with the cryic soil temperature regime.

The low temperatures in regions with polar climates only permit physical weathering; chemical weathering is (almost) absent and, if any, is restricted to the short thawing periods. In addition, biological activity is low due to the low temperature.

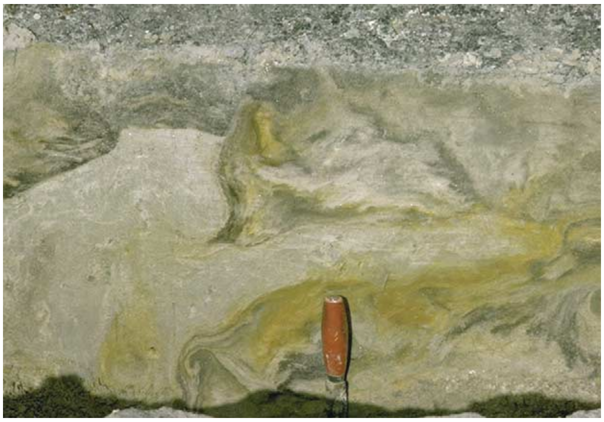


Figure 7 Horizon involutions in a Canadian soil. (Reproduced with permission from C. Tarnocai, Agriculture and Agri-Food, Canada.)

Characteristic for the soils in the polar climates are the effects of freezing and thawing and, in many cases, the presence of permafrost (permanently frozen subsoil with ice segregation). The annual freezing and thawing of the soils result in thermal contraction and expansion of minerals in rocks, thus slowly weakening the rock structure and permitting water to enter in small cracks. The freezing of water further widens the cracks, as ice has a lower bulk density than water (0.9 versus 1.0 kg dm^{-3}) and, over time, the rock will disintegrate into angular fragments. In soils, the annual freezing and thawing results in the process of 'cryoturbation,' giving such features as horizon involution (Figure 7), frost heave, cryogenic sorting, and patterned grounds (e.g., stone polygons).

The low rate of humification of organic matter and the wet conditions during the thawing periods result in the accumulation of peat at the surface. Peat soils are among the most widespread soils in the polar regions. Approximately half of the Earth's peat soils or Histosols (ST and WRB) occur in the circumpolar region of the northern hemisphere.

Other soils are dominantly Gelisols (ST) or Cryosols (WRB), soils that are characterized by the occurrence of permafrost within 100 cm depth, or within 200 cm depth in association with gelic materials above it, or having one or more cryic horizons within 100 cm of the soil's surface.

Soils as Climatological Archive

Some soil processes and resulting properties are clearly climate-related, such as the processes of ferralitization, rubefaction, and cryoturbation. Other processes such as brunification, decalcification, clay illuviation, and cheluviation occur under a variety of climate conditions. Again others such as salinization and formation of secondary calcium carbonate are rapidly counteracted when climates become moister, and are not preserved in the soil. Therefore, the number of soil properties usable as climatological indicators is limited.

The process of ferralitization, which occurs in the moist tropical climates, leads to accumulation of iron (hydr)oxides, aluminum (hydr)oxides, and kaolinitic clay, all of which are very stable in soil environments. The occurrence of soils with large amounts of the above components almost invariably points to former moist tropical climates. A typical example of the inheritance of kaolinite from earlier weathering stages is the so-called 'arène,' i.e., the deeply weathered, sandy regolith over granites in France which has often been attributed to deep weathering in the Tertiary. Other evidence in soils for humid tropical conditions in Europe during the Tertiary era is found in the occurrence of laterites in the Mainz region of Germany.

Rubefaction – the process of red coloring as a result of the formation of hematite – is associated with climates that have a long, hot dry season. Rubified soils, now occurring in climates without a distinct dry period, are evidence of earlier climates that had a distinct dry period. Examples of these can be found in the cool climatic belt in western and central Europe, particularly in association with limestone and marls.

Frost action results, amongst other things, in the formation of frost wedges and polygonal structures. It is particularly these two phenomena that are frequently encountered in former periglacial areas. They are evidence of the formerly cold climatic conditions that occurred during the ice ages of the Quaternary Era (Figure 8).

Despite clay illuviation occurring in many different climates, it requires a net flow of water under gravity conditions from the surface downward. It is therefore unlikely that such conditions occur in the dry climates of the Earth. Yet the presence of clay illuviation features in present-day dry soils, as recognized in *Soil Taxonomy* by the suborder of Argids (Aridisols with an argic or nitric horizon), points to formerly moister climates with conditions more favorable to clay translocation.

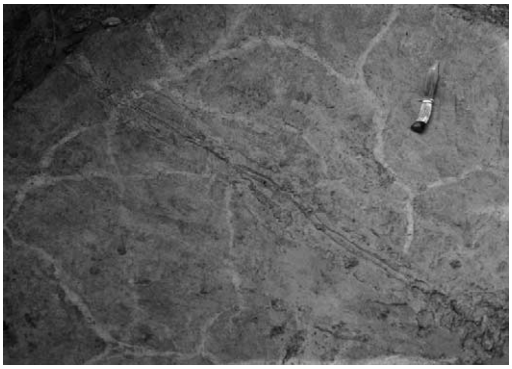


Figure 8 Polygonal subsurface structures in a horizontal section in a loess soil in Belgium. (Reproduced with permission from ISRIC.)

Summary

It is evident that soils bear a strong stamp of past and present climatic conditions. Essentially, the ratio between precipitation and evaporation determines the flux of water in the soil (downward or upward). Also temperature is a key climatic component, as it not only regulates the speed of weathering (the higher the temperature, the faster chemical processes take place), but also the biological activity in the soil, which plays a crucial role in the breakdown of organic matter. Wind speed is important in the rate of evaporation; and wind action, of particular importance in dry climates with very little vegetation, contributes both to the removal of mainly finer particles and to the accumulation of mainly coarser particles.

See also: **Classification Systems:** FAO; **Clay Minerals**; **Flocculation and Dispersion**; **Grassland Soils**; **Metal Oxides**; **Salination Processes**

Further Reading

Doebl F (1973) Ein "Aquitan"-Profil von Mainz-Weissenau (Tertiär, Mainzer Becken). Mikrofaunische, sedimentpetrographische und geochemische Untersuchungen zu seiner Gliederung. (An Aquitanian age cross section at Mainz-Weissenau (Tertiary, Mainz Basin). Microfaunal, sedimentpetrographic and geochemical study of its subdivision.) *Geologisches Jahrbuch Series A5*.

Duchaufour P (1998) *Handbook of Pedology. Soils–Vegetation–Environment*. Rotterdam, the Netherlands: Balkema.

FAO-ISRIC-ISSS (1998) *World Reference Base for Soil Resources*. World Soil Resources Report 84. Rome, Italy: FAO.

Hsu PH (1977) Aluminium hydroxides and oxyhydroxides. In: Dixon JB and Weed SB (eds) *Minerals in Soil Environments*, pp. 99–143. Madison, WI: Soil Science Society of America.

Köppen W (1936) *Handbuch der Klimatologie,* vol. 1, part C. *Das geographische System der Klimate*. Berlin, Germany: Borntraeger-Verlag.

Marbut CF (1935) *Atlas of American Agriculture,* part III, *Soils of the United States*. US Department of Agriculture, Bureau of Chemistry and Soils. Washington, DC: US Government Printing Office.

Meyer A (1926) Über einige Zusammenhänge zwischen Klima und Boden in Europa. *Chemie der Erde* 2.

Rode AA (1962) Soil Science (Pochvovedenie, translated from Russian). Moscow, Russia: Goslesbumizdat.

Soil Survey Staff (1999) *Soil Taxonomy. A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 2nd edn. Agriculture Handbook 436. US Department of Agriculture, Natural Resources Conservation Service. Washington, DC: US Government Printing Office.

Trewartha GT (1968) *An Introduction to Climate*, 4th edn. New York: McGraw Hill.

VASKhNIL (1987) *Classification and Diagnostics of Soils in the USSR*. (Russian Translation Series **42**). Rotterdam, the Netherlands: Balkema.

Human Impacts

J Sandor, C L Burras, and M Thompson, Iowa State University, Ames, IA, USA

Introduction

Humans, like many other organisms, are active agents of soil formation. Because soils comprise the dynamic, vibrant skin of the Earth's terrestrial surface, people have always interacted with, and therefore changed, soils and the course of their formation. While soils are subject to major change and even destruction by natural forces on the scale of geologic time, changes resulting from human activity usually occur on a much shorter time scale. People have impacted soil in a multitude of ways and extents, through farming, building, mining, and even war. In some cases, human activities enhance soils for particular uses. However, in a number of cases, the interplay between humans and soils has resulted in soil degradation, which is fundamentally a negative process of formation. Recognizing that soil use is literally and figuratively the base for most civilizations and that soil resources are essentially nonrenewable on the human time scale, understanding soil formation is imperative to developing agricultural and natural resource sustainability, and to protecting environmental quality.

Scales and Scope of Human Impacts on Soil Formation

To understand the effects of humans on soil, it is helpful first to consider soil-formation processes and their rates in natural environments to provide a frame of reference. From the perspective of pedology, the study of soil formation, soil is a complex assemblage of mineral and organic materials formed at the Earth's surface. In spite of the increasing impact of human activities on soil and the likelihood that all of Earth's ecosystems have been influenced to some degree by humans, many soils still retain their basic morphology imparted by natural processes and environmental factors. A hallmark of soil formation is the differentiation of horizons (layers). Soils are developed and organized into horizons by complex, interrelated physical, chemical, and biological processes that are determined by the factors of soil formation: climate, organisms, geology, topography, and time. The morphology of every soil is the expression of fundamental processes interacting with one another over multiple spatial scales (from micrometers to kilometers) and temporal scales (from days to millennia). Horizons at the surface (A horizons) are often enriched with organic matter, while deeper horizons (B horizons) may have accumulations of clay, calcium carbonate, metal–organic complexes, or other materials. Formation of A horizons is relatively rapid because organic matter accumulates in a few centuries to a millennium. Formation of B horizons commonly takes many thousands of years to become fully expressed. In this sense, soils can range from young to middle-aged to old. As soil formation progresses, soils generally tend to become increasingly anisotropic as they differentiate into greater kinds and numbers of horizons.

Landscape stability is a prerequisite for soil horizon development to proceed. Even so, disturbance and change are integral to the functioning of all natural ecosystems and their soils. Soils, landscapes, and associated biological communities are subject to disturbances ranging from minor perturbations such as low-intensity fires to major events such as volcanism, and to long-term climatic and environmental change. On the geologic time scale, soils are subject to major alteration, destruction, and renewal. In contrast to the relatively slow pace of natural soil development, soil changes resulting from human activity are often more rapid and far-reaching. Human-caused change may be so fast and irreversible that the impacted soil bears little resemblance to its original form. Human impacts usually reverse the anisotropic trend of soil formation, making soils simpler, less organized, and more homogeneous.

A wide array of land use and other human activities have altered paths of formation in many soils (Table 1). Corresponding soil changes also vary greatly in kind, intensity, time and spatial scale, and significance for soil and environmental quality (Tables 2 and 3). Human actions that change soil may act directly or indirectly by changing both soil morphology and the

Table 1 Human agents of change in soil properties and processes

	Examples
Agriculture and forestry	
• *Crop and animal agriculture*	Tillage, fertilization, cropping, flooding, irrigation, drainage, terracing, grazing
• *Forestry*	Altered vegetation type and cover, harvesting operations
• *Off-site effects*	Erosion and sedimentation, causing soil contamination and burial
Cities and industry	Excavation, urban cover, artificial fills, industrial pollution
Buildings, roads, and other structures	Excavation, artificial fills, pollution, paving, land leveling
	Dams and reservoirs, polders, dikes, mounds, other artificial landforms and fills
Mining and other earth and water resource extraction	Soil removal for pits and quarries, erosion and sedimentation associated with hydraulic mining, mixing and inversion of earth materials, reclamation, groundwater pumping
War and weapons testing	Bomb craters, military transport and engineering, contamination by radioactive elements
Human-linked environmental change	
• *Global climate change*	Atmospheric CO_2 increase, climate warming, increased climatic variability
• *Desertification*	Vegetation loss or change, accelerated wind and water erosion, salinization
• *Acid rain*	Soil acidification and other chemical changes

Table 2 Soil change resulting from human impact

Causes and characteristics of soil change[a]	*Examples*
Possible causes	
⌈Direct	Compaction
⌊Indirect	Slower or diverted water movement and reduced soil water storage after compaction
⌈Deliberate	Fertilization, liming, irrigation
⌊Unintended	Nutrient depletion or imbalance; salinization
⌈Constructive	Plaggen soil; agricultural terracing and drainage
⌊Destructive	Accelerated erosion, sulfide oxidation with drainage or exposure, salt/sodium buildup
Magnitude and extent	
⌈Low impact	Organic matter change from light grazing
⌊High impact	Urban expansion; land filling
⌈Part of soil	Thinning of A horizon from cultivation-induced erosion
⌊Whole soils	Removal of entire soil by intense erosion
Duration and rate	
⌈Short-term/ephemeral	Alleviation of plow pans by freezing and thawing; liming
⌊Long-term/permanent	Urban soil; mine soil
⌈Slow rate	Agric horizon development
⌊Fast rate	Oxidation following drainage
Response of soil to human impact	
⌈Susceptible	Base-poor soil susceptible to acid rain
⌊Resistant	Base-rich soil buffered against acid rain
⌈Reversible (resilience)	Alleviation of compaction is faster in organic matter-rich A horizons
⌊Irreversible	Laterite/plinthite hardening with exposure; construction of urban soils and soils on mined land
Outcomes of soil change for soil quality and use	
⌈Beneficial	Organic matter addition
Neutral/benign	Fertilizing that balances crop removal of nutrients
⌊Degradation (loss of productivity/ off-site environmental impacts)	A horizon erosion

[a]Bracketed terms illustrate range end members.

underlying soil-forming processes. They may be intentional changes based on land-management strategies to increase soil productivity, or rearrangements of landscapes and soil materials to construct buildings, roads, and other structures. Conversely, they may be unintended changes that can lead to degradation such as soil erosion and burial under sediment. Indirect soil change may result from off-site processes physically remote from the impacted soil, such as subsidence following groundwater pumping or downstream

Table 3 Spatial scales of human-induced soil change

Soil components	*Approximate spatial scale (m)*	*Examples of impacts*
Physical, chemical, and biological properties of soils (e.g., clay, microorganisms, organic matter)	10^{-10}–10^{-4}	Retention of pollutants such as heavy metals, pesticides, and industrial solvents
Morphological properties	10^{-3}–10^{-2}	Soil structure degradation; changes in texture, color, porosity, and pore distribution
Horizons	10^{-1}–10^{0}	A horizon erosion
Whole soils (pedon)	10^{0}–10^{1}	Plaggen soils; salinized soils; liming that changes Ultisols to Alfisols
Soil–watersheds–landscapes–ecosystems–biosphere	10^{2}–10^{7}	Broad changes in Mollisols, Histosols, Gelisols, and other soil orders

sedimentation. In land uses such as agriculture, surface horizons have usually been more altered than subsurface horizons, while engineering activities often change the entire sequence of soil horizons. Engineering activities lead to some of the most intensive impacts by removing soils entirely and replacing them with different earth materials, as well as by reshaping landforms or constructing new ones. Agriculture and engineering merge in management practices such as in paddy rice production and terracing of slopes, which also involve major geomorphic and soil change.

Some human activities such as agriculture have influenced soil processes and morphology for thousands of years, while others such as global climate change induced by human activity are becoming increasingly important. Some anthropogenic impacts on the environment, such as acid rain and other pollution, incur less visible chemical and biological changes, while processes such as desertification result in more sweeping changes through loss of vegetative cover, desiccation, wind and water erosion, and salt accumulation. While the atmospheric effects of global climate change have been detected relatively recently, their potential impacts on soil formation and distribution are large, because climate is a major determinant of soil formation over time.

Soils differ in their response to human actions in terms of their resistance to change, and their resilience, that is, their ability to rebound toward their original state. Processes of human-caused soil change may be reversible or irreversible, and the resulting changes may be ephemeral to permanent. Response depends on both external factors, such as the type of impact and environmental conditions, and internal soil properties. For example, soils with uniform textures or that are rich in organic matter tend to be resistant to compaction. After many thousands of years of natural formation, some subsurface soil horizons opened by deep plowing may return to their original condition in a few years (e.g., clay-rich argillic horizons), whereas more indurated horizons such as silica-cemented duripans remain fragmented longer.

Evaluating Human Impacts On Soil Formation

Human-caused soil change has been studied by pedologists in several ways. Hans Jenny and others have framed their views in the context of the factors of soil formation, placing humans within the biotic factor, while recognizing the unique role of human culture as distinct from other organisms. Jenny showed that human land use often alters soils through changes in the other soil-forming factors. For example, irrigation alters arid land soils by changing the climate factor, effectively increasing precipitation. Other researchers set humans apart from the natural factors of soil formation to emphasize the extraordinary scale and rate of anthropogenic effects on soils. In working with drastically altered soils such as mine soils and urban soils, pedologists have defined new soil taxonomic classes, because these soils bear little or no resemblance to their original state.

How is soil change detected and measured? While modern, large impacts on soil may be obvious, such as with wholesale change or destruction of soils and landscapes in urban development, longer-term soil changes are often more subtle and complex. Histories of soil change may be difficult to reconstruct, complicated by imprints of multiple land use activities and changing environmental conditions. An approach to evaluating long-term anthropogenic influence on soil has been to identify soils that are relatively undisturbed, or at least that have documented land-use histories, and to use these soils as reference points from which to measure soil change. Finding truly comparable sets of reference and altered soils is challenging and imprecise, and may not be possible in some situations. Although results of comparisons must be interpreted carefully, the paired-site approach represents one of the few methods available actually to measure long-term soil change. Soil changes from agriculture have been monitored using controlled experiments at a few locations such as Rothamsted Experimental Station (UK), Morrow Plots (Illinois, USA), and Sanborn Plots (Missouri, USA) for more than a century. Evaluating ancient agricultural soils up to about 1000–2000 years in age has extended the time perspective on anthropogenic soil change. Much can be learned about long-term human effects on soils, and about successes and failures in soil use and management, from past cultures at archeological sites, as well as from contemporary traditional cultures who have lived on the same land for many generations. Studies of past and present traditional land use contribute information for evaluating soils on longer time scales inherent in the concept of sustainable land use, as well as for modeling and predicting the future condition of land resources.

Changes in Soil Properties and Processes Resulting From Human Activities

To facilitate understanding of impacts and appreciation of the many ways people depend on soil for sustenance and support, soil changes are presented here in the context of the type of land use or other human activity. Examples from agriculture, engineering for urban development, mining, waste disposal, war, and global environmental change are selected to represent the myriad of ways people change soil. It is important to recognize that human effects on soil are complex, because soils are dynamic systems with interrelated components; they are themselves part of ecosystems within Earth's biosphere. Soils vary in their sensitivity to disturbance and threshold for change. Single land-use practices can affect many soil properties in a cascading process, and different land uses may initiate similar processes of change in soils. Soils may be impacted by more than one land-use practice simultaneously or diachronically. Land uses and their variations often differ in the way they impact soils in terms of magnitude, spatial extent, rate, and duration.

Agriculture

Agriculture has profoundly and extensively impacted soils since its inception about 10 000 years ago. In this discussion, agriculture is framed broadly to include all plant and animal production for food, feed, fiber, and fuel, including crop and livestock farming, as well as forestry. All of these land uses rely on many forms of soils worldwide.

Some soils are so impacted by agriculture that their original horizons are wholly transformed or buried. Examples of truly anthropogenic soils are plaggen soils, common in western Europe (Figure 1). These constructed surface horizons, which can be a meter thick, are the product of centuries of cultivation, with additions of organic materials such as manure and sod and inorganic amendments such as sand or marl. Soils that have been similarly transformed through management practices such as terracing and long-term applications of fertilizing materials are found in other regions with long histories of intensive agriculture such as Southeast Asia and the Andes.

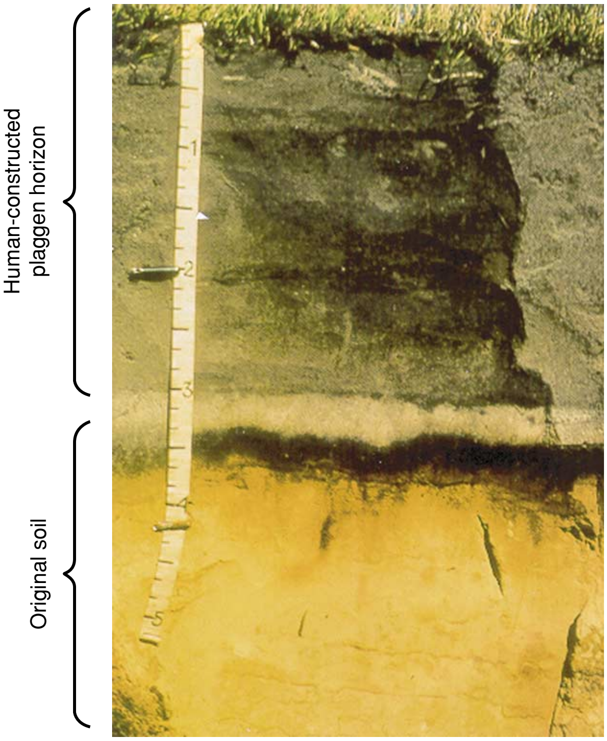


Figure 1 Plaggen horizon (approx. 1 m thick) from the Netherlands, created by long-term organic matter additions and cultivation, burying a natural Spodosol. (Reproduced from Soil Survey Staff (1999) *Soil Taxonomy,* 2nd edn. US Department of Agriculture. Washington, DC: US Government Printing Office.)

In contrast to both the intent and effect of constructed soils, many soils on slopes have been impacted by erosion accelerated by agriculture, some to the point of obliteration or deep burial under eroded sediments through mismanagement. Surface horizons, which generally contain the most plant nutrients and organic matter, are the most immediately vulnerable to erosion. Especially subject to major change are those soils with well-developed, organic matter-rich surface horizons such as Mollisols. The transformation of Mollisols by erosion to soils classified in other orders such as Alfisols, Inceptisols, and Entisols has been documented for past and present agriculture. A key reason for increased soil erosion is the loss of protective vegetation cover as natural ecosystems such as forests or prairies are converted to agricultural use. While accelerated erosion has been difficult to quantify on regional to global scales, estimated average erosion rates on cropland range from about 12 (USA) to 30–40 (Africa, Asia, South

America) metric tons per hectare annually, much greater than natural erosion rates. Human-induced water and wind erosion is estimated to have resulted in the degradation of approximately one-quarter of the world's cropland.

As surface horizons are eroded, subsurface horizons or bedrock effectively rise closer to the surface and may become exposed. Their incorporation into the topsoil under moderate to severe erosion alters soil properties such as color, structure, and texture, often lowering soil quality. If subsurface horizons are problematic for plant growth, serious degradation of productivity can ensue. In some tropical regions such as West Africa and Southeast Asia, exposure of horizons of cemented, iron-rich clay (petroplinthite) through erosion has resulted in abandonment of agricultural land. Erosion of Ultisols with highly weathered, acidic horizons of clay accumulation in warm, temperate to tropical environments has significantly reduced soil productive potential. Even in thick sediment parent materials that are considered 'forgiving' to erosion, productivity can diminish through exposure or shallow occurrence of dense or cemented horizons and layers that limit rooting, decrease available water capacity, or greatly slow drainage. For example, thick deposits of loess (wind-blown sediment rich in silt) such as in China and the Palouse and Midwest regions of the USA are favorable for crops but particularly susceptible to erosion. Erosion rates in loess can reach 100 metric tons per hectare per year, exceeding natural erosion rates by one to two orders of magnitude. This has resulted in the loss of A horizons, followed by the exhumation of buried soils (paleosols) or dense, brittle fragipans at shallower depths that reduce productivity.

Erosion also impacts soil formation off-site. As eroded sediments are transported downslope, soils in the lower terrain of watersheds may become buried by sediment that does not reach streams. Studies in several regions have measured burial of original soils by sediment that is centimeters to meters thick. Also, the freshly deposited sediment constitutes another parent material in which soil formation may begin anew. An opposite situation is where natural sedimentation processes and renewal of soil fertility are impaired by dam construction, as in the case of the Aswan dam on the Nile river. The extensive erosion and sedimentation that occur in many agricultural landscapes indicate that massive transformation of many soils is continuing. This situation underscores the need for documenting and responding to changes that undermine soil quality and productivity.

Because water deficiency or excess is of major concern in most agricultural systems, many soils have been altered by water-management practices. Some changes are direct, such as the effects of flooding soils for wet rice production, creating paddy soils with distinctive anthropogenic characteristics (Figure 2). 'Irragric' soils in arid to semiarid central Asia have been highly altered by long-term irrigation, including addition of suspended sediment in the irrigation water. Unintended chemical effects from irrigation in some regions have led to severe soil degradation through salt accumulation and transformation to saline or sodic soils. In contrast to paddy soil management, conversion of wetlands to row crop production by artificial drainage in management, regions such as the Midwest USA has changed soils from dominantly anaerobic to dominantly aerobic states, leading to changes such as organic matter oxidation. In some areas of the world, many meters of organic soil thickness have vanished through oxidation processes, resulting in significant land subsidence.

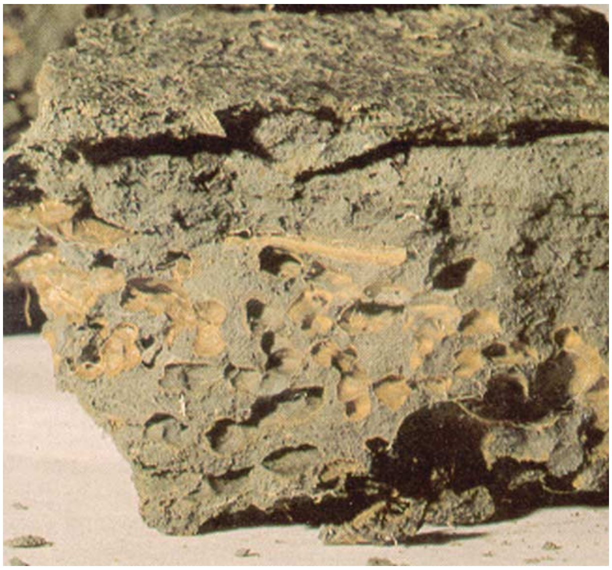


Figure 2 Sample (approx. 5 cm thick) of paddy soil managed for wet rice production in Southeast Asia. Anthropogenic features include variegated colors from flooding and alternating reduction/oxidation, and vesicles formed from gases trapped beneath the platy, puddled surface layer. (Reproduced with permission from Moorman F and van Breemen N (1978) *Rice: Soil, Water, Land*. Los Baños, Philippines: International Rice Research Institute.)

More subtle agricultural impacts may not completely alter the original soil, yet over time cause significant change. For example, since the mid nineteenth century, conventional cultivation of organic matter-rich, prairie-derived soils (Mollisols) throughout the Midwest USA has led to marked decreases in organic matter in upper soil horizons (Figure 3). Many Mollisols have lost approximately one-third to one-half of their original organic matter. In addition to erosion, a principal cause is the disruption of soil aggregates by cultivation, making previously protected humus accessible to microorganisms. Compared with native prairie ecosystems that are characterized by abundant organic matter and conservative nutrient cycling, agricultural soils have lower inherent fertility and are more 'leaky' with respect to nutrients such as nitrogen. Soil organic matter loss has a cascading effect on soil properties such as structure, reducing aggregate stability, and making soils more prone to compaction. Organic matter contents can be at least partially restored to soils through management, by using more diverse, conservation-oriented cropping systems, or by a return to more natural vegetation.

Cities and Industry

The growth of cities and industry has profound impacts on soils and their continued formation. Some of the impacts are direct; others are indirect. The creation of constructed urban soils is perhaps the most easily recognized direct impact of city growth on soils (Figure 4). Constructed soils have distinct profiles that are typically the product of the cheapest engineering fill available at the moment of creation. Soil scientist Phillip J. Craul wrote that the history of urban development has been based upon the premise 'dirt is dirt and it's cheap.' These human-created soils may be highly stratified and composed of very different types of fill in terms of their texture, mineralogy, and chemistry. Total fill – and consequently soil – thicknesses can reach multiple meters, with a soil's maximum age corresponding with the age of the city. That is, a constructed urban soil in Rome, Athens, Mexico City, and Beijing can be several thousands of years old, whereas a similar soil in New York or Buenos Aires must be less than 500 years. Of course, many new hectares of these soils are created each year in every city as new projects are completed. Many ancient cities occur on tells, human-created mounds or hills formed through time as cities were rebuilt on top of previous ones. Other construction materials used in urban soils include composted sewage sludge, municipal solid wastes, heavy metals, glass, plastic, and metal. Still others consist of clean sand mined from nearby rivers or topsoil transported from local farms. In other words, urban soils have an exceptionally high degree of disorganized variability.

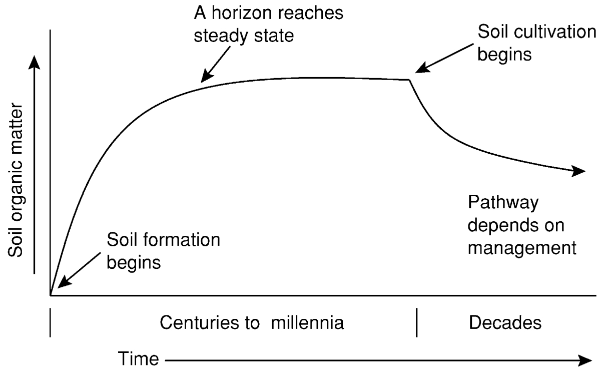


Figure 3 Accumulation of organic matter in topsoil (A horizon) during long-term natural soil formation, and subsequent rapid decrease following cultivation (based on the work of Hans Jenny).

A second direct impact of city development on soil formation is at the landscape scale as houses, gardens, parks, and streets replace the natural terrain. The result is fragmented landscapes, where soil-forming processes become controlled by constructed topographic, hydrologic, and ecologic factors. Thus, while the morphology of many of these soils does not show the dramatic changes of their downtown counterparts,

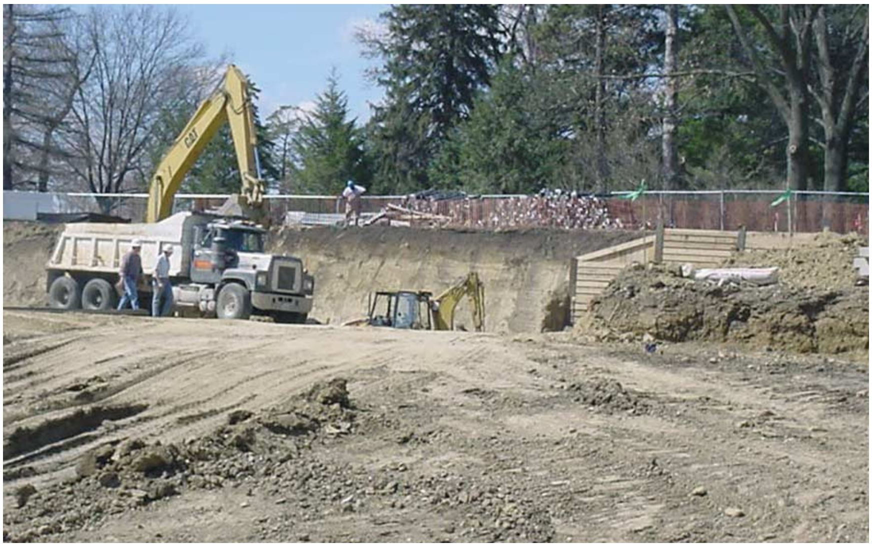
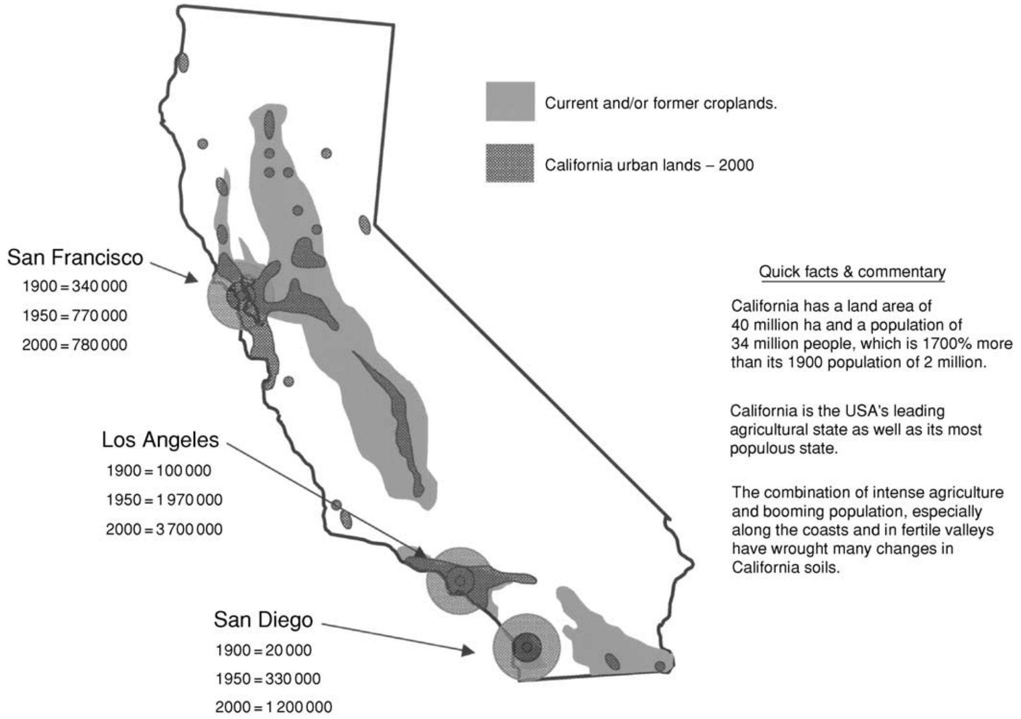


Figure 4 Urban soil being created after the natural soil has been removed. Note natural soil profile in background. Photo by Julie McLaughlin.



Quick facts & commentary

California has a land area of 40 million ha and a population of 34 million people, which is 1700% more than its 1900 population of 2 million.

California is the USA's leading agricultural state as well as its most populous state.

The combination of intense agriculture and booming population, especially along the coasts and in fertile valleys have wrought many changes in California soils.

Figure 5 California: an example wherein soils are increasingly a product of human activities such as agriculture and urbanization. Compiled from a variety of sources including US Census Bureau and US Census of Agriculture data and maps for c. 1900, 1950, and 2000, several atlases as well as the Great Valley Center and US Geological Survey's ''Preliminary Assessment of Urban Growth in California's Central Valley.''

their soil-forming processes are considerably altered. This can be illustrated by considering two hypothetical, adjacent gardens: one growing turf that is heavily fertilized and irrigated, while the other is planted to conifer shrubs and trees but receives no fertilizers and extra water. Over time, the properties of the soils under these two gardens will diverge in response to their different microclimates (wet versus dry), biota, and chemical inputs. Like their constructed counterparts, they will be characterized by disorganized variability, although in this case the disorganization occurs at the landscape (i.e., garden-to-garden) scale.

The impact of cities and their continued growth on soil formation is increasing as large areas of prime farmland as well as wild, often fragile, lands get converted to urban uses (Figure 5). In the USA the area of urban land increased from 6 million to 26 million hectares between 1945 and 1992, with some parts of California and New England developing into '100-mile cities' as urban areas became interwoven through express highways and suburbs. Urbanization is even more of a critical factor in developing nations, because their populations are growing at a rate fourfold greater than the world population. The World Resources Institute estimates that nearly 500 000 ha of rural soils are converted to urban land annually. For example, the urban core of São Paulo, Brazil, expanded from 18 000 to 90 000 ha between 1930 and 1988, with the overall metropolitan area encompassing more than 800 000 ha.

Industry, like cities, directly and indirectly affects soil formation. Land excavation and drainage are two important direct agents of change. Dredging of wetlands as well as sediments from beneath shallow waters to improve harbors, build canals, and meet other commercial needs is an example of industrial excavation. The US Army Corps of Engineers estimates that billions of cubic meters of sediment are dredged annually around the globe with over 300 million cubic meters being dredged annually just in the USA. Dredging has three direct impacts with respect to soils and their formation. First, the dredging itself destroys whatever natural soil existed. Second, the dredged material – or 'spoil' – is piled somewhere and is later used in soil construction. Soil created from dredge spoil may be used for buildings, roads, parks, farms, or even wild areas. Over time it will develop a sequence of horizons in response to its environment. The third pedological impact of dredging is oxidation of mineral and organic matter. This occurs because spoil that was excavated from below the water table is now directly exposed to the oxygen-rich atmosphere. In many cases, the oxidation is fairly benign in terms of soil properties and formation. A noteworthy problem occurs when the spoil contains sulfides. Exposure of sulfidic compounds to air results in the rapid formation of large quantities of sulfuric acid in these soils. Such acid sulfate soils are also created when naturally sulfidic soils are drained. It is estimated there are 24 million hectares of potential acid sulfate soils globally, with many located in prime settings for agricultural and urban development.

The indirect impacts of industry on soil formation are as manifold as the indirect impacts of urbanization. One of these is the addition to soil of contaminants such as lead, cadmium, and other heavy metals. In some locations these metals are deposited on soils as airborne contaminants that originated in factory or vehicle exhausts. In other cases they are added to soil directly and intentionally as constituents of fly ash and other wastes. 'Land farming' refers to disposing of waste by applying and incorporating it into soil a few tons per hectare at a time. These soils are subsequently cropped, although the crops are not normally used for human consumption. Compounds that contain sodium comprise another class of industrial contaminant that changes soil morphology – ultimately creating sodic soils. This process is increasingly noted worldwide because of the mind-boggling number of industrial uses for sodium and sodium compounds (e.g., coolant in nuclear processes; fumigants; solvents in manufacturing of brass, paper, ceramic glazes, textiles, and fertilizers; and in food processing) and the frequency at which sodium wastes are applied to land. Radionuclides are another class of industrial contaminants that alter soil processes. For example, 8.4 million hectares around Chernobyl, Ukraine, remain contaminated by ^{90}Sr and ^{137}Cs from the 1986 nuclear reactor disaster there.

The influence of most industrial processes on soil formation is proportional to the intensity of industrial activity and distance of the soil from the industry. This means that soils in areas having numerous factories, automobiles, and other industrial activities are more highly enriched in contaminants than soils where there is little industry. Likewise, the magnitude of impact radiates out with distance from the source of impact.

Mining

Humans have mined the Earth for metals, minerals, and fuel for thousands of years. Currently, 8 billion tons of minerals and oil are extracted from the Earth annually. Since the nineteenth century, mining activities have radically transformed landscapes and altered natural soil development processes. Mining requires that the overburden of soil and its parent materials be displaced so that the product (e.g., ore, rock, or coal) can be removed, concentrated, and transported off-site as economically as possible. Mining often results in large pits, high walls, and large accumulations of unconsolidated overburden, smelting residue, and finely ground residual ore. When abandoned at the Earth's surface, these materials often pose significant environmental risks due to subsidence, settlement, and slope instability. Spoil may also be contaminated with soluble metals. In addition, when some types of spoil are exposed to air and percolating water, oxidation reactions produce strongly acidic water with high metal concentrations, severely restricting subsequent biological activity. In the USA alone, it has been estimated that surface mining has disturbed more than 23 000 km^2.

Concern about mines and their spoils prompted many nations to enact environmental-quality laws beginning in the late twentieth century. For example, before the Resource Conservation and Recovery Act of 1976 and the Surface Mining Control and Reclamation Act of 1977, many mines in the USA were simply abandoned when the product had been exhausted or was too expensive to extract from the remaining ore. Since that time, however, planning for and completing reclamation of mined land has been part of every mining operation in the USA. Thus a number of practices are actively followed to allow the site to be used again for other purposes and to ensure that surface water and groundwater resources are protected. Construction of new soil at reclaimed mine sites is an example of the most radical of human impacts on soil formation.

Successful mine land reclamation must address a number of large-scale environmental issues while at the same time creating a near-surface environment that promotes vegetative growth and movement of water through the watershed. For example, a major problem in reclamation of coal and metallic ore mines is the oxidation of sulfide minerals and subsequent drainage and runoff of acidic water (Figure 6). Acidic spoil, surface water, and seeps must be treated with large amounts of calcite or hydrated lime to neutralize acidity before vegetation can be established. Organic matter and plant-available nutrients are commonly lacking in spoil, so application of sewage sludge or other organic amendments to the spoil surface is often helpful. Heavy applications of plant nutrients such as phosphorus and potassium may also be required.

Large-scale topographic reconstruction of mined land is expensive, but it is usually essential to ensure that reconstructed soils maintain ecologic and environmental quality. In some reconstructions, it is possible to approximate the general contours of the premining landscape (Figure 7). Throughout the process of consolidating, grading, and shaping the new landscape, control of settlement, slope instability, and surface erosion will minimize off-site damage of

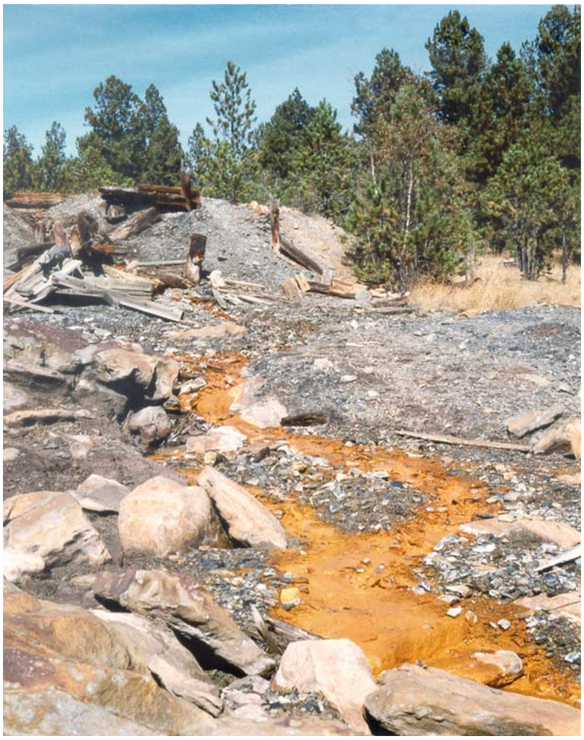


Figure 6 Water that drains through spoil from mines is often highly acidic because of the oxidation of sulfur and iron. This water is toxic to plants and aquatic organisms. Acid mine drainage at the Minnesota Ridge Mine, South Dakota. This site was remediated in 2001. (Photograph by Tom Durkin. Reproduced with permission from the South Dakota Department of Environment and Natural Resources, Minerals and Mining Program.)

sediments and drainage as well as maintain optimum conditions for plant growth and development of natural soil horizons. Water management in reclaimed mine lands includes designing watersheds for directing both run-on and runoff water. The new landscape may be designed to retain or slow the movement of water by passing it through constructed wetlands.

Where the original materials were not systematically stockpiled during mining, reclaimed soils often display extreme variability in particle size and composition. Problems of inhomogeneous mixtures of coarse and fine particles include both subsidence and inadequate water-holding capacity for plants to grow. For this reason, modern reclamation approaches usually call for restoration of the natural sequence of geologic materials (e.g., bedrock-derived spoil placed below till-derived spoil placed below organic matter-rich topsoil). Rates of soil development in reconstructed mine soils are slow. Reclaimed mine soils tend to have ‘AC’ horizon sequences, in which incipient topsoil is underlain by unconsolidated layers lacking soil structure or other pedogenic features. Still, where reclamation has been well managed, the thickness of surface horizons in the reclaimed soil may exceed those of nearby unamended soils that have been degraded by farming or logging operations.

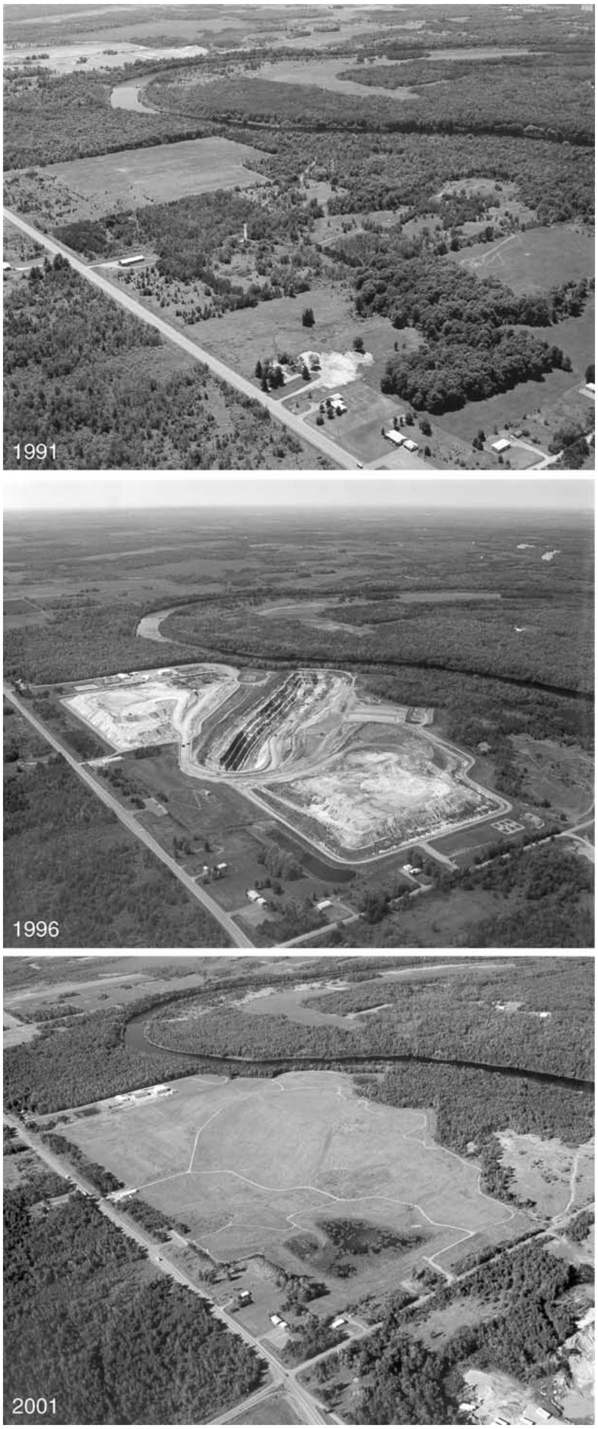


Figure 7 The Kennecott Flambeau Mine near Ladysmith, Wisconsin, was mined for copper, gold, and silver in the 1990s. At the end of mining, the open pit was completely backfilled and native vegitative communities were established. The site now includes high quality wetlands, native prairie, woodlands, and nature trails. Photographs before mining (1991), during mining (1996), and after restoration (2001) by T-B0 and courtesy of the Kennecott Flambeau Mining Company.

War

War, like mining, may have rapid and long-lasting impacts on soils. For example, nearly a century after an 11-month battle during World War I, extensive trench works and bomb craters remain near Verdun, France. Vietnam provides a documented example of the impacts of more recent wars. There, it has been estimated that bombs created some 21 million craters between 1965 and 1971, covering more than 50 000 ha (Figure 8). The explosion of a 2400-kg bomb typically created a hole approximately 10 m in diameter and 5 m deep; metal fragments from each bomb could be spread over an area of 0.5 ha. Some craters exposed soil horizons of iron oxide accumulation that subsequently irreversibly hardened to petroplinthite.

Similarly, land-clearing programs in Vietnam altered soil processes. Widespread and concentrated application of defoliating herbicides, as well as bulldozing of vegetation, resulted in extensive erosion, loss of organic matter-rich surface horizons, and increased flooding. Native mangrove forests have not returned to the drastically disturbed or herbicide-treated lands. More than 30 years after the application of herbicides ended, dioxin, a manufacturing contaminant in the herbicides, remains in the soil and sediments of the countryside. Thus soils are effectively beginning a new phase of soil genesis with topography, parent material, and vegetation altered from the previous state of dynamic equilibrium.

In addition to the persistence of war-related changes on soil properties and processes, there may be indirect effects that have a cascading environmental impact caused by displacement of civilian, agrarian populations. In rural Vietnam, bomb craters' soils contaminated with metal fragments, unexploded ordnance, and residual dioxin, and war-related destruction of water-diversion structures have made agricultural and forestry recovery very difficult for the inhabitants. In response, some farmers have moved to less-productive, marginal land that is more susceptible to erosion. Others have moved to urban slums, increasing the environmental pressures on urban areas. Thus the long-term impacts of war on soil properties and the processes of soil formation can extend beyond the zone of immediate impact.

Climate Change and Soil Formation

Soil formation is closely tied to climate, and, to the extent that human activities alter local or regional climatic variables such as temperature and effective precipitation, soil properties inevitably will be changed also. At the global scale, climatic change in the next century is likely to be driven by increasing atmospheric concentrations of 'greenhouse gases' such as carbon dioxide (CO_2) and methane (CH_4). As a result, it is expected that the mean temperature of the Earth's surface could rise as much as 1.5–4.5°C over the next 100 years. Although the local impacts of global warming are difficult to predict with certainty, significant shifts in the mean and seasonal variation of air temperature as well as in the total and seasonal distribution of rainfall are likely to perturb

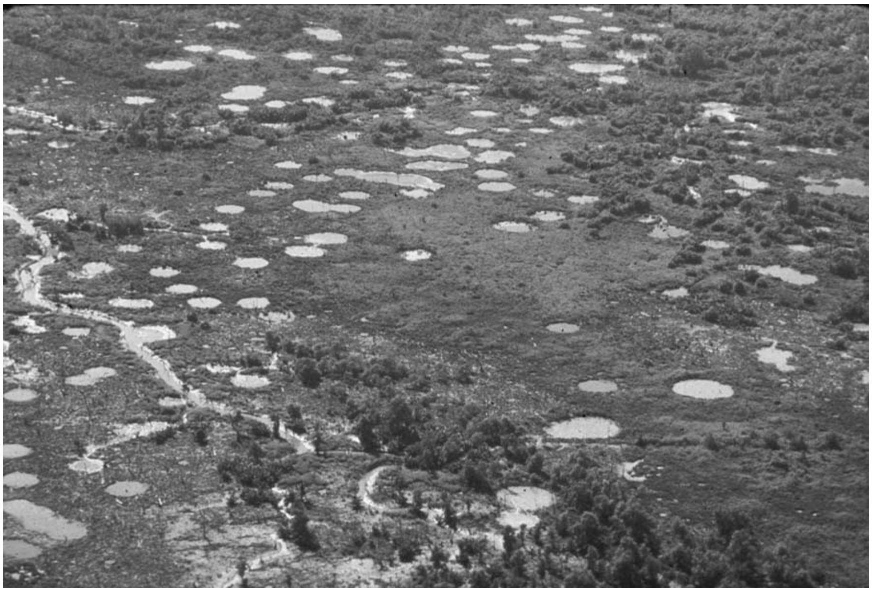


Figure 8 A cratered mangrove forest in Gia Dinh province in August 1971. Bomb craters created during the war in Vietnam disrupted agriculture and most other land uses. (Photograph courtesy of Arthur Westing. Reproduced from Westing AH and Pfieffer W (1972) The cratering of Indo-China. *Scientific American* 226: 20–29.)

both natural and agricultural ecosystems. The impact of a global increase in temperature will be unevenly spread in both space and time. Some of the impacts include: (1) new precipitation patterns in which today's wet regions get wetter and dry regions get drier; (2) migration northward of boreal and hardwood forests; (3) melting of ice sheets, sea ice, and glaciers, with concomitant rise in sea level and flooding of coastal soils; and (4) melting of permafrost in northern latitudes and oxidation of organic matter now stored in cold-region soils. For example, the temperature increase has the potential to markedly shrink the extent of approximately 11 million square kilometers of Gelisols, the cold-region soils with permafrost, where an estimated 20% of the world's soil organic carbon is stored.

Human activities, mainly burning of fossil fuels, will thus lead to some global and regional climate change over the next century. Both regionally and locally, climate change will directly and indirectly impact soil processes by altering vegetation patterns, encouraging increased water and wind erosion, favoring mass movement, increasing the leaching of nutrients and organic matter through soils, and favoring microbial oxidation of organic matter in surface horizons. Although we are not now able to predict how fast soil properties will be altered as a result of global and regional climate change, ultimately some types of soil horizons will be lost (for example, surface horizons composed entirely of organic matter) and some types will be newly developed where they did not previously exist (for example, spodic horizons under new boreal forests). As human populations adapt agricultural practices to new climate conditions, further changes in soils may be accelerated. For example, salinization may increase due to more extensive irrigation, and landscape instability may become widespread as more forest land is cleared for production of crops and livestock.

Significance and Future of Human Impacts on Soil Formation

Soil is a critical, dynamic natural resource and vital component of ecosystems, but one that is often neglected. This is in part because soil lies beneath the surface and so is not as familiar as other resources such as water, plants, or animals. As a natural resource, soil is crucial for food-, fiber-, and fuel-production systems, for construction materials and foundations, for replenishing and maintaining the quality of surface water and groundwater, and for waste processing and containment. Natural variation in pathways of soil formation and soil properties make for a diverse range of suitabilities for different land uses and sensitivities to anthropogenic change. Conservation of soil resources and soil quality is a critical priority globally because soil is fundamentally nonrenewable on a human time scale.

Human activities dictate that soil change is inevitable and necessary. Anthropogenic soil change has a long history, but it has exponentially increased in intensity, spatial extent, and rate since the nineteenth century. Some changes are advantageous and improve soil functionality. However, many soil changes induced by past and present land use have resulted in environmental degradation. In developed and developing nations alike, accelerated erosion, compaction and disruption of structural aggregates, lowered fertility, and contamination by pollutants continue to be sources of serious concern. Because soil serves as a filter, substrate, and reservoir linked to other land, water, and biological resources, degradation is not just detrimental to the soil internally and locally, but it extends to all parts of the larger hydrologic, geologic, and biological system. Some soils, like plants and animals and the habitats and ecosystems with which they are closely connected, have become endangered and in some cases extinct. Human-altered soil is a significant factor in global environmental change, as both a cause and a recipient of change. Oxidation of soil organic matter is a source of increased atmospheric carbon dioxide; soil properties stand to be further altered by global climate change; and yet soil also has the potential to help reverse the trend by sequestering more carbon. Soils also play a key role in other forms of global change such as desertification, acid precipitation, and loss of tropical forest ecosystems.

What can be done to counter the negative impacts of human activities on soils and environment and to support conservation and more sustainable land use? Improved knowledge of soil formation and soil change is an important starting point. By more clearly recognizing the extent of soil change and understanding its causes, mechanisms, and consequences, the better our chances become to develop and implement management practices that can sustain soil resources and restore damaged land. Traditionally, soil maps have portrayed soils in their relatively original, undisturbed state. However, because of the unprecedented scale of soil change and massive transformation, it is imperative that anthropogenic soil change be documented, monitored, and acted upon to a much greater degree. Efforts to deal seriously with the extent and significance of human impact on soil formation and distribution have led to a proposal to include Anthrosols in the US Soil Taxonomy, and to recognition of Anthrosols at the highest level in the current international soil classification system (World Reference

Base for Soil Resources). Improved knowledge of soil formation processes in relation to natural and human-altered pathways is essential to the restoration of ecosystems and the development of sustainable land use. The future of human society and the Earth we inhabit depend on soil formation processes, and increasingly, on how we respond to the changes human actions cause in soils and the biosphere.

List of Technical Nomenclature

A horizon Surface mineral horizon with some organic matter accumulation

argillic horizon Subsurface diagnostic horizon with significant clay accumulation

B horizon Subsurface mineral soil horizon (several kinds)

C horizon Relatively unweathered, unconsolidated mineral horizon or layer (or soft bedrock), usually beneath the active zone of soil formation

fragipan Dense, brittle subsurface horizon

irragric horizon or soil Anthropogenic horizon formed by long-term irrigation and accompanying additions of sediment

paleosol Soil formed in past geologic time

petroplinthite Hardened or cemented horizon of highly weathered, iron-rich soil

plaggen horizon or soil Anthropogenic horizon formed during long-term cultivation by additions of organic materials and/or inorganic amendments

spodic horizon Refer to Spodosol below

The following terms are eight out of the 12 soil orders of US soil classification:

Alfisol Mineral soil with subsurface clay accumulation, relatively low in organic matter, and relatively high in bases

Entisol Undeveloped mineral soil

Gelisol Mineral or organic soil of cold regions, with permafrost

Histosol Organic soil

Inceptisol Mineral soil with initial, but relatively weak development

Mollisol Mineral soil with thick, dark, organic matter-rich surface horizon, high in bases

Spodosol Mineral soil with subsurface accumulation of humus and amorphous aluminum, and commonly amorphous iron

Ultisol Highly weathered mineral soil with subsurface clay accumulation, relatively low in organic matter, and relatively low in bases

See also: **Acid Rain and Soil Acidification**; **Archeology in Relation to Soils**; **Carbon Emissions and Sequestration**; **Civilization, Role of Soils**; **Classification of Soils**; **Classification Systems:** USA; **Degradation**; **Desertification**; **Erosion:** Water-Induced; **Heavy Metals**; **Morphology**; **Organic Matter:** Genesis and Formation; **Pedology:** Basic Principles; **Quality of Soil**; **Salination Processes**; **Structure**

Further Reading

Amundson R and Jenny H (1991) The place of humans in the state factor theory of ecosystems and their soils. *Journal of Soil Science* 151: 99–109.

Amundson R, Guo Y, and Gong P (2003) Soil diversity and land use in the United States. *Ecosystems* 6: 470–482.

Bidwell OW and Hole FD (1965) Man as a factor of soil formation. *Soil Science* 99: 65–72.

Bryant RB and Galbraith JM (2003) Incorporating anthropogenic processes in soil classification. In: Eswaran H *et al.* (eds) *Soil Classification: A Global Desk Reference*, pp. 57–66. Boca Raton, FL: CRC Press.

Courty MA, Goldberg P, and Macphail RI (1989) *Soils, Micromorphology, and Archaeology.* Cambridge, UK: Cambridge University Press.

Craul PJ (1999) *Urban Soils: Applications and Practices.* New York: John Wiley.

Dudal R, Nachtergaele FO, and Purnell MF (2002) *The Human Factor of Soil Formation.* Transactions of the 17th World Congress of Soil Science, Bangkok, Thailand.

Effland WR and Pouyat RV (1997) The genesis, classification, and mapping of soils in urban areas. *Urban Ecosystems* 1: 217–228.

Gong ZT (1983) Pedogenesis of paddy soil and its significance in soil classification. *Soil Science* 135: 5–10.

Goudie A (1994) *The Human Impact on the Natural Environment*, 4th edn. Cambridge, MA: MIT Press.

Jenny H (1984) The making and unmaking of a fertile soil. In: Jackson W *et al.* (eds) *Meeting the Expectations of the Land*, pp. 42–55. San Francisco, CA: North Point Press.

Johnson DL and Lewis LA (1995) *Land Degradation: Creation and Destruction.* Oxford, UK: Blackwell.

Lal R (ed.) (1999) *Soil Quality and Soil Erosion.* Boca Raton, FL: CRC Press.

Lal R, Sobecki TM, Livari T, and Kimble JM (2003) *Soil Degradation in the United States.* Boca Raton, FL: CRC Press.

Ruddiman WF (2001) *Earth's Climate: Past and Future.* New York: WH Freeman.

Sandor JA and Eash NS (1991) Significance of ancient agricultural soils for long-term agronomic studies and sustainable agriculture research. *Agronomy Journal* 83: 29–37.

Sencindiver JC and Ammons JT (2000) Minesoil genesis and classification. In: Barnhisel RI *et al.* (eds) *Reclamation of Drastically Disturbed Lands*. Madison, WI: Soil Science Society of America.

Westing AH and Pfieffer W (1972) The cratering of Indochina. *Scientific American* 226: 20–29.

Yaalon DH and Yaron B (1966) Framework for man-made soil changes – an outline of metapedogenesis. *Soil Science* 102: 272–277.

Parent Material

K R Olson, University of Illinois, Urbana, IL, USA

Introduction

Parent material is the starting material from which soil develops and can include both consolidated rock and unconsolidated material. Mineral matter originating from rocks is referred to as 'soil parent material,' because it is the principal ingredient from which soils are formed. Unconsolidated material includes material deposited by gravity, wind, or water (Table 1) and consists of specific minerals of different sizes. However, the primary parent materials of organic soils are decomposing materials of various plant types.

Table 1 Bedrock type and parent materials by origin

Origin	*Parent material*	*Bedrock type*
Residual materials	Residium	
Sedimentary materials	Sedimentary	Sandstone
		Shale
		Limestone and dolomite
	Metamorphic	Marble
		Slate
		Quartzite
		Schist
		Gneiss
	Igneous	Basalt
		Granite
		Quartzite
Gravity deposits	Colluvium	
Ice deposits	Glacial till	
Water deposits		
Stream	Alluvium	
	Outwash	
Lake	Organic deposit	
	Beach deposit	
	Lacustrine	
Wind deposits	Loess	
	Eolian sand	
	Volcanic ash	

Defining Soil Parent Material and Role in Soil Genesis

Parent material is commonly defined as unconsolidated and chemically weathered mineral or organic matter from which the solum (set of horizons that are related through the same cycle of pedogenic or soil-forming processes) of soil is developed. 'Soil genesis' refers to the mode of origin of the soil, particularly the processes or soil-forming factors responsible for the development of the solum, or true soil, from unconsolidated parent material.

Soil-Forming Factors

The character, properties, and development of soils are controlled by external factors. Soil formation is difficult to view and often takes place over a long time. Parent material was initially identified as one of four significant soil-forming factors, as shown in the following equation:

$$S = f(\mathrm{cl}, \mathrm{o}, \mathrm{p})t^{\mathrm{o}} \qquad [1]$$

where S equals some attribute or measurable property of soil, $f()$ means 'as a function of,' cl is the climate of a region, o represents the organisms (both plants and animals), p is the geologic substratum, and t^{o} is the relative age (young, mature, or senile).

Later, a generalized equation, which related soil behavior to genetic factors and included relief, was developed:

$$S = f(\mathrm{cl}, \mathrm{o}, \mathrm{r}, \mathrm{p}, t, \ldots) \qquad [2]$$

where S equals some attribute or measurable property of soil, $f()$ means 'as a function of,' cl is the environmental climate, o is species of organisms, r is relief or topography (including hydrologic features such as the water table), p is parent material (the state of soil formation at time zero), and t is time (age of soil, absolute period of soil formation), with additional, unspecified factors.

The factors define the state of the soil system. The soil system has been defined as an arbitrary volume of the solum, a soil body, a pedon (three-dimensional body of soil with lateral dimensions large enough to permit the study of horizon shapes and relations in it), or an entire ecosystem of a component tessera (operation unit which is collected in the field, examined, and analyzed). The ecosystem approach avoids the impossible task of separating living material (nonsoil) from inanimate material (true soil) in a sample where they intermingle. The ecosystem is open to energy and matter influxes and outfluxes. Influx of energy includes solar radiation, heat transfer, and entropy transfer; outflux of energy includes heat radiation and reflection of light. Influx of matter involves gases entering the ecosystem by mass flow (wind), water in liquid and solid forms, solids dispersed in air or dissolved or dispersed in water, and organisms that migrate into the ecosystem.

A lithosequence (Table 2) has been defined using functional factorial analyses as a set of soils with property differences due solely to differences in parent material with all other soil-forming factors constant as expressed in the following equation:

$$S = f(\mathrm{pm}), \mathrm{cl}, \mathrm{o}, \mathrm{r}, t, \ldots \qquad [3]$$

where S equals some attribute or measurable property of soil, $f()$ means 'as a function of,' pm is parent material, cl is the environmental climate, o is species of organisms, r is relief or topography, t is time, and there are additional, unspecified factors.

Such an array or sequence of soils is difficult to recognize and establish because of the problem in establishing that all soils in the set have property differences due solely to parent material without some effects from environmental or local landscape position. Several sets of soils have been defined as approaching this condition, usually on young and relatively simple landscapes such as in glaciated regions. On these landscapes, attempts have been made to analyze the effects of differences in parent material composition.

Early approaches to soil survey and classification were based on the geology and composition of the soil-forming material. The geologic origin and composition of the initial material (Table 2) was identified by the terms 'granite soils,' 'glacial soils,' or 'till soils.' Soils that originate from a parent material inherit the mineral types found in them. Over time, these original minerals are weathered (dissolved) and new minerals form and accumulate in the soils.

Relief prior to and during soil formation is related to the nature of the starting soil material or parent material. In steeper topography, where the valleys below the mountain ranges are characterized by broad colluvial–alluvial fans, the starting or parent material near the mountain range contains more coarse and angular material than areas farther away from the mountain range. In broad river deltas, crests of natural levees near the stream channels contain coarser material than the areas beyond the levees, which are very nearly level and contain the finer-textured parent material.

Table 2 Lithosequence of soils with different composition and physical properties of till

Parent material	*Sequence of soil horizons*	*Soil order*
High-lime till	A, B, C	Mollisols Inceptisols
Medium-lime till	A, E, Bt, BC, C	Alfisols
Mixed sandstone and limestone till	O, E, Bs, BC, C	Spodosols
Sandstone and granite till	O, E, Bh, Bs, C	Spodosols

Weathering

'Weathering' describes the combined effects of all the physical and chemical processes that break down and transform preexisting parent materials at or near the surface of the Earth. These products are more stable under the physical and chemical conditions at the surface. Weathering of parent materials is the response to several forms of energy inputs as a function of time. The outcome of the weathering process is affected by physical and chemical factors. The products of weathering include liquids (including the solutions of salts present in rivers and the ocean) and solids (i.e., sediments and soils). The rock cycle, including erosion, transportation, and sedimentation, operates as a complex chemical sorting system which distributes weathering products of different composition to various sites.

Soil chemical composition is the product of parent material being altered over geologic time by the action of climate, topography, and biota. Physical disintegration and chemical weathering transform parent material into secondary minerals that often differ in type and composition with profile depth. Soils are transient bodies of the landscape. The initial soils will differ from the current and future soils. Parent material is often unrecognizable as a result of these chemical transformations and physical disintegration. The greatest change from the parent material occurs when soils in open systems gain new substances and lose their minerals through dissolution, leaching, or erosion. Minerals differ in their susceptibility to weathering. Physical disintegration and chemical weathering transform parent material into secondary minerals that often differ in kind and composition. Soils are the result of leaching, oxidation, and dissolution of surface materials by the percolation of groundwater and by humic acids derived from oxidized organic material. New minerals such as clays are formed by the chemical alteration of the parent material. The chemical weathering process develops a soil profile ranging from heavily weathered surface materials down to unaltered parent material.

Effects of Parent Material on Soil Properties

Soil formation starts after unconsolidated parent material is deposited on a stable landscape or after bedrock has been exposed at the Earth's surface and continues over time. Rate of soil formation depends on the climate, including temperature and rainfall. It also depends on vegetation and the activity of other organisms which live on or in the parent material. These organisms help convert parent materials to soil.

The properties of a modern soil are the result of the composition of the surficial layer which existed when the other factors began to impact and the alterations resulting from the effect of these factors over time. Properties of younger soils are significantly influenced by parent material. Weathering and pedogenic processes result in characteristics of the original parent material being eliminated. In old, weathered soils, extremely resistant starting material, such as quartz sand, may still exist. It can be difficult to separate the effects of the other soil-forming factors on the parent material of this soil, the nature (properties) of the initial material (Table 3) and its influence on soil, and the kind of 'preweathering' of the starting material before becoming parent material for the soil. Climatic and vegetational changes in the recent geologic past make it difficult to separate parent material influences on soil properties from the effects of other factors.

Modern soil properties are influenced by rock types (Table 1). The impacts of rock types on soil properties have been categorized into the following subdivisions: igneous, sedimentary, and combinations of mineralogically similar igneous and metamorphic rocks. Sedimentary parent materials (Table 1) include loessial deposits, unconsolidated glacial deposits, and coastal plain sediments, as well as consolidated rock such as sandstones, shales, limestone, and dolomites. Siliceous crystalline parent materials (Table 1) include both igneous and metamorphic rocks such as gneiss, granites, and schists, as well as more 'acidic' quartzose. Volcanic ash is a parent material composed of noncrystalline, glass fragments, bits of the easily weatherable ferromagnesian minerals, feldspars, and varying amounts of quartz. Dark-colored ferromagnesian rocks include andesites, diorites, basalt, and hornblende gneiss.

Table 3 Parent material and bedrock type effects on soils and soil classification

Parent material or bedrock type	*Diagnostic epipedon or subsurface layer*	*Soil order*
Alluvium	None	Entisols
Beach deposits	None	Entisols
	Spodic horizon	Spodosols
Outwash	Cambic horizon	Inceptisols
	Spodic horizon	Spodosols
Lacustrine	Cambic horizon	Inceptisols
Organic deposits	Histic epipedon	Histosols
Volcanic deposits		Andisols
Loess	Mollic epipedon	Mollisols
Glacial till	Argillic horizon	Alfisols
		Ultisols
Sandstone	Cambic horizon	Inceptisols
Shale	Argillic horizon	Alfisols
Limestone	Mollic epipedon	Mollisols

Mineral components of many soils are either developed *in situ* during the course of weathering and pedogenesis or primarily inherited from parent materials. Primary minerals such as igneous and metamorphic rocks are formed at high temperatures. Secondary minerals, including those in sedimentary rocks and soils, are those formed at lower temperatures. The elemental and mineralogical composition of the parent materials determine the soils' elemental and mineralogical composition. Approximately 96% by volume of the elemental composition of the Earth's crust is oxygen, silicon, aluminum, and iron.

Two of the most important properties of soil parent material are its mineral composition and its texture. Both of these characteristics are retained as properties of the soil formed from the parent material. They can be altered somewhat over time. Such alteration often reduces the particle size by weathering action. The fine particles are subject to removal from the surface layer by erosion, eluviation, and leaching.

Influence of Parent Materials in Soil Genesis

Lack of vertical uniformity in parent materials is a problem frequently faced by pedologists. Any of the parent materials (Table 1) could be located above or below another parent material and result in a lithologic discontinuity. Discontinuities can result from additions of loess, volcanic ash, sediments or colluvium over residual materials or older deposits. A lithosequence is defined as a set of soils with property differences due solely to differences in parent material with all other soil-forming factors constant. Parent materials affect soil property differences including diagnostic epipedons (surface layers) and subsurface layers which do affect the soils and soils classification (Table 3). In New York state, soil properties and soil horizons are attributed to differences in the composition and physical properties of the parent materials (Table 2). This set of four New York soils with different composition and physical properties of till (Table 2) is probably not a true lithosequence, since parent material differences may not account for all soil property and soil horizon differences, but approaches one sufficiently to illustrate the point.

The basic model of soil implies that soils are dynamic and geographic. Dynamics is more appropriate than statistics in soil systems because of implied processes or driving forces. Morphological properties of soil develop as a result of processes acting on parent materials.

Time has the effect of modifying the influence of the parent material so that only in young or relatively immature soils will parent material be the dominant factor in the soil-forming process. The influence of the parent material is an inverse function of time. To illustrate this, soil scientists have suggested that soils derived from dissimilar parent materials such as granite and basalt will become indistinguishable given sufficient time to attain equilibrium.

Parent Material and Soil Classification

The central concept of an Entisol is a slightly developed soil with properties determined primarily by parent material (Table 3). The weak soil development is usually due to youthfulness, wetness or dryness extremes, or resistant parent materials. Mineral components of many soils are inherited from parent materials, while others are developed *in situ* during the weathering and pedogenesis. Soils derived from different parent materials often classify as distinct soil series based primarily on properties inherited from the parent material (Table 3).

Summary

The five soil-forming factors – parent material, climate, vegetation, topography, and time – determine the distribution and nature of soils over the Earth's surface. Mineral components of many soils are inherited from parent materials, while others develop *in situ* during weathering and pedogenesis. Morphological properties of soil develop as a result of processes acting on parent materials. Soil scientists have attempted to show the controlling effects of parent material on soil properties to the extent that parent material is proposed as an independent soil-forming factor and defined as the state of the soil system at time zero of soil formation. The physical body of soil and its associated mineralogical and chemical properties are the basis of other soil-forming factors. Previously weathered rock or even a previous soil could be considered parent material, which can be modified over time. Parent material exerts its strongest influence on the soil-forming process of immature soils.

List of Technical Nomenclature

cl	climate
$f()$	as a function of
o	organisms
p or pm	parent material
r	relief
S	a quantifiable attribute of soil
t or t^o	time

See also: **Factors of Soil Formation:** Biota; Climate; Human Impacts; Time

Further Reading

Birot P (1960) *The Cycle of Erosion in Different Climates*. (Translated from French, 1968, by Jackson CI and Clayton KM.) Berkeley, CA: University of California Press.

Buol SW, Hole FD, and Mc Cracken RJ (1980) *Soil Genesis and Classification*, 2nd edn. Ames, IA: Iowa State University Press.

Chesworth W (1973) The parent rock effect in the genesis of soil. *Geoderma* 10: 215–225.

Chesworth W (1976) Conceptual models in pedogenesis. A rejoinder. *Geoderma* 16: 257–260.

Cline MG (1953) Major kinds of profiles and their relationships in New York. *Soil Science Society of America Proceedings* 17: 23–27.

Dokuchaev VV (1883) Russian chernozem (Russkii Chernozen). In: *Collected Writings (Sochineniya),* vol. 3. Moscow, USSR: Academy of Science.

Jackson C (1973) *Geology Today. Communications Research Machines*. Del Mar, CA: CRM Books.

Jackson ML (1964) Chemical composition of soils. In: Bear FE (ed.) *Chemistry of Soil*, 2nd edn, pp. 71–141. New York: Reinhold.

Jenny H (1941) *Factors of Soil Formation*. New York: McGraw-Hill.

Joffe JS (1949) *Pedology*, 2nd edn. New Brunswick, NJ: Pedology Publications, Rutgers University Press.

Russell RJ (1967) *River Plains and Sea Coasts*. Berkeley, CA: University of California Press.

Singer MJ and Munns DN (1996) *Soils: An Introduction*, 3rd edn. Upper Saddle River, NJ: Prentice-Hall.

Soil Survey Staff (1999) *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 2nd edn. Agriculture Handbook No. 436, USDA–Natural Resource Conservation Service. Washington, DC: US Government Printing Office.

Troeh FR and Thompson LM (1993) *Soils and Soil Fertility*, 5th edn. New York: Oxford University Press.

Troeh FR, Hobbs JA, and Donahue RL (2004) *Soil and Water Conservation for Productivity and Environmental Protection*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Time

E F Kelly and C M Yonker, Colorado State University, Fort Collins, CO, USA

Introduction

The many hundreds of thousands of soil types at the Earth's surface are each composed of unique biological, chemical, and physical properties. These properties are the direct result of a complex set of soil-forming processes that are conditioned by unique environmental factors. The environmental conditions under which soils and their properties develop are governed by unique combinations of the climate, biota, geology, and topography, which operate over time.

Climate affects soil development, as moisture and temperature control the rate of biological, chemical, and physical processes. The effect of biota is primarily related to controls on inputs and outputs of organic and inorganic materials; the effect of topography is the modification of temperature and precipitation relationships. The chemical and mineralogical composition of the parent material influences such key soil properties as nutrient status, alkalinity or acidity, and texture.

Whereas climate, biota, topography, and parent material have a direct effect on soil-forming processes, time has an indirect effect. Time affects soil-forming processes by controlling their duration and, therefore, the degree of soil development. Thus, we can differentiate between 'young' and 'old' soils in a given ecosystem on the basis of degrees of chemical and physical weathering and the resulting expression of soil properties.

Key Soil Properties that Vary with Time

The amount and distribution of soil organic carbon (SOC), carbonate, and various other elements, and the amount and distribution of clay and its mineralogy all change as a function of time in a generally consistent and predictable manner (Figure 1). The amounts and distribution of these soil constituents are often used to determine the relative age of a soil.

Once vegetation is established in an ecosystem, SOC is continually added through both above- and belowground inputs of living and dead plant tissues. The expectation is that older soils have a greater accumulation of SOC than younger soils. SOC accumulation eventually reaches a steady state wherein inputs equal outputs and there is no further net accumulation with increasing soil age. The length of time needed for SOC to reach steady state varies among environments, but is within a relatively short time frame of 10^2–10^3 years.

The accumulation of calcium carbonate ($CaCO_3$) in soils over time is generally limited to soils of semiarid and arid regions, where precipitation of carbonate occurs under conditions typical of dry soil environments, high pH, and high evapotranspiration. Under more humid conditions, the increased production of carbonic acid (H_2CO_3) results in the dissolution of $CaCO_3$ and a decrease in $CaCO_3$ content over time. Sources of calcium in arid and semiarid soils may be from *in situ* weathering of calcium-bearing minerals or from atmospheric deposition. Unlike SOC, $CaCO_3$ accumulation may not reach a steady state, but continues to increase through time.

The most abundant elements in soils, excluding oxygen and carbon, are silicon, aluminum, iron, magnesium, calcium, sodium, and potassium. Although these elements have varying solubilities, their relative abundance is a direct function of the degree of soil weathering, which increases through time. In general, the simple soluble cations, magnesium, calcium, sodium, and potassium, are readily released from the soil through weathering processes and are easily leached (Figure 2a). Barring an input of these elements to the soil system via the atmosphere or some other mechanism, their quantities steadily decrease as a function of time. Soil silicon, aluminum, and iron exist primarily as relatively insoluble hydroxides and are leached at a slower rate than the simple soluble cations (Figure 2b).

Phosphorus is an element which exists in soils in a variety of organic and inorganic forms, the distribution of which indicates the relative weathering of the soil profile and, therefore, soil age. As soil development begins and the soil matures, primary mineral sources are depleted, and organic and secondary, occluded mineral forms of phosphorus accumulate (Figure 2c).

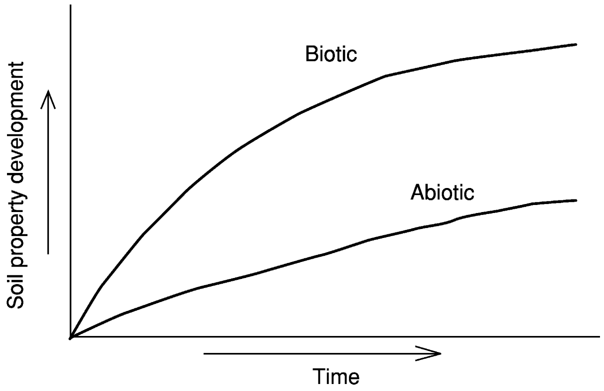


Figure 1 The effect of time on the development of biotic (e.g., soil organic carbon) and abiotic (e.g., calcium carbonate) soil properties.

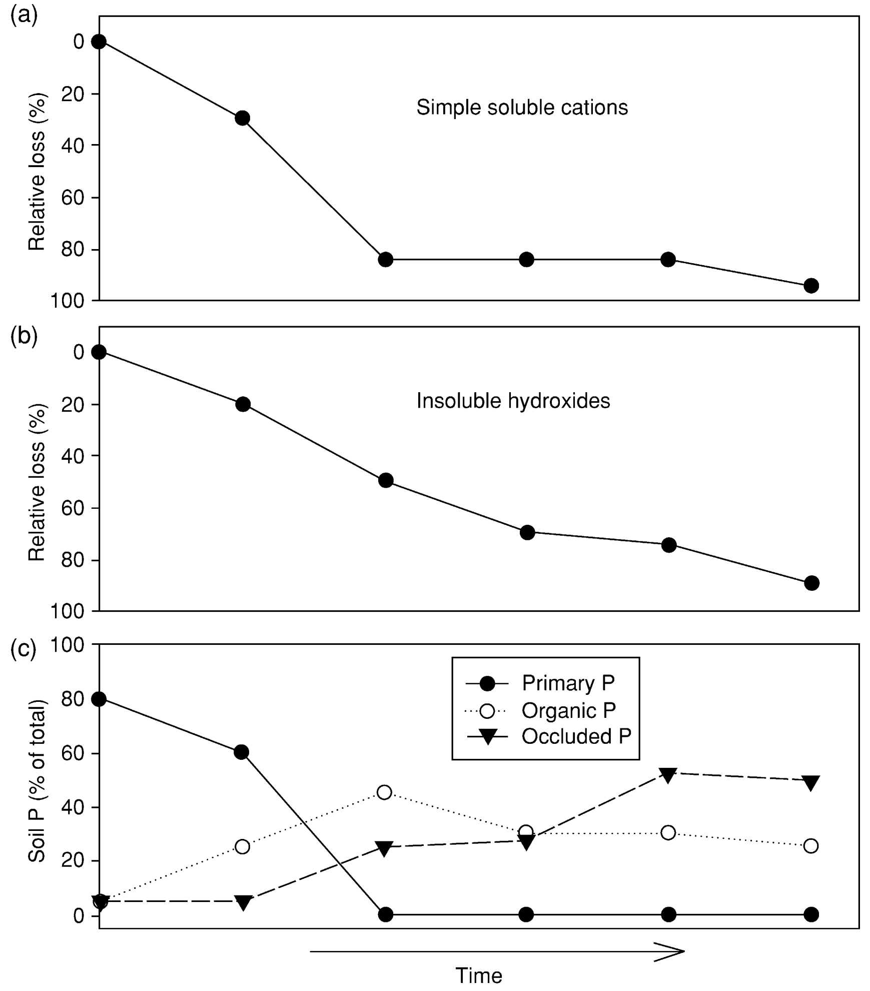


Figure 2 The effect of time on the loss of chemical soil constituents (a) and (b) and the transformation of phosphorus (c).

The accumulation and distribution of soil clay minerals are indicators of soil age. Clays accumulate through time as a function of eolian deposition of dust, translocation from superjacent soil horizons, and *in situ* formation of secondary clay minerals. Layers of accumulated clay can form in as little as a few hundred years if there are significant atmospheric inputs in the form of dust or mud rains. Where there is little atmospheric input, rates of accumulation are largely a function of translocation and weathering. Clay-enriched horizons can develop within 10^3–10^4 years, depending on whether the environment is humid or arid.

Clay mineralogy is partly a function of soil age because, over time, soils are increasingly more depleted in soluble framework cations and silicon, major constituents of clay minerals. Clay minerals, therefore, often form in a recognized sequence, with the earliest stages commonly consisting of illites and/or chlorites, intermediately weathering to smectites and/or vermiculites, typically resulting in kaolinite in the oldest soils that have undergone the lengthiest time of soil formation.

The preceding discussion theoretically describes the effect of time on soil development when climate is a constant variable. Given the same duration of soil formation, a soil forming in a warm, humid environment will have soil properties that are significantly more developed and appear to be older than a soil forming in a cool, dry environment. Increasing precipitation causes increased chemical and physical weathering, effectively mimicking the effect of a long period of soil formation in a relatively short time period. In humid areas, where precipitation exceeds potential evapotranspiration, biological and chemical weathering processes rapidly transform organic and mineral matter. Leaching transfers materials down, and eventually through, the soil profile. In drier areas, where precipitation is less than potential evapotranspiration, the same soil-forming processes operate; however, the availability of water limits the intensity and duration of these processes.

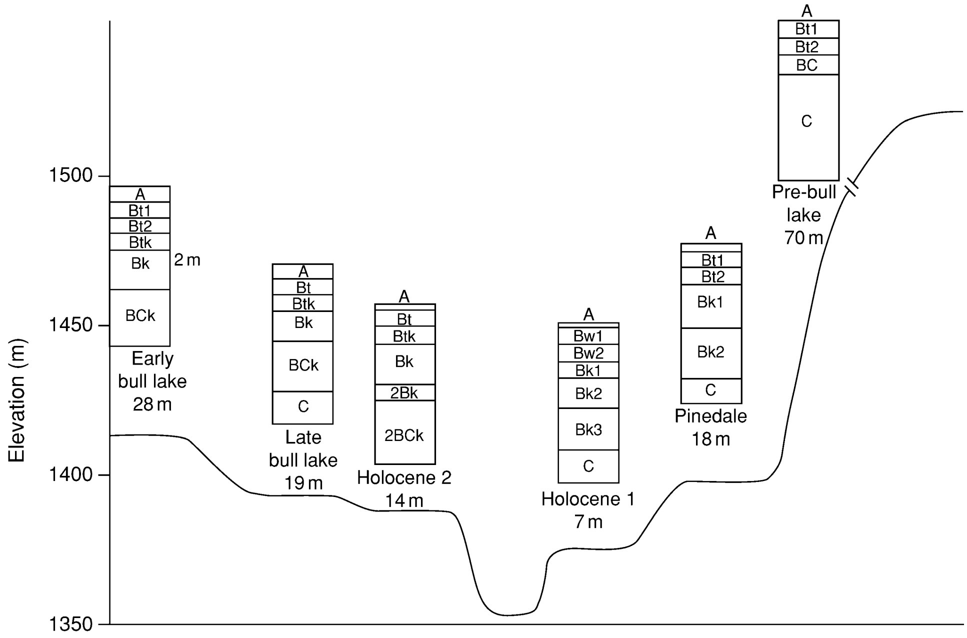


Figure 3 Soil chronosequence developing in a semiarid ecosystem. The youngest soil is located at the lowest elevation (7 m) above the floodplain; the oldest at the highest elevation (70 m). The soil ages are: Holocene (<10 000 years), Pinedale (13 000–20 000 years), Late Bull Lake (130 000–190 000 years), Early Bull Lake (190 000–300 000 years), Pre-Bull Lake (300 000–600 000 years).

Studies of Time and Soil Formation

The differences in properties between soils can be explained as a function of time if the soils being compared are forming in the same climate, maintain the same type of vegetation, have developed from the same or similar parent materials, and are located at the same topographic position. In other words, time is the only factor that varies between the soils being compared. Climate, biota, parent material, and relief can be held constant under a limited number of circumstances, including the comparison of soils developing on stream terraces sequentially incised by a single river system. In this case, the soils of the floodplain are developing in the most recently deposited material and are therefore the youngest; the soils on the highest stream terrace relative to the floodplain are therefore the oldest (Figure 3). Such a series of related terrace soils, which differ only with respect to their duration of formation, are called a chronosequence. Soils developing on chronosequences show broadly similar trends in morphology with time, including: (1) a thickening zone of organic matter accumulation; (2) increasing pedon thickness; (3) increasing horizon complexity; and (4) development of zones of secondary mineral accumulation. Such trends continue until such time that the soil becomes so weathered it begins to degrade.

Chronosequences are problematic in that, over millennial time scales, a constant climate cannot be assumed. Absolute soil age, or duration of soil development, can also be difficult to establish. In spite of these hindrances, chronosequences provide an excellent opportunity to study the influence of time on soil formation.

List of Technical Nomenclature

Calcium carbonate Carbonate mineral that forms coatings on soil that react with an acid to form carbon dioxide gas

Evaporation Water on the Earth's surface or in the soil absorbs heat from the sun to the point that it vaporizes or evaporates and becomes part of the atmosphere

Horizon An individual layer within the soil which has its own unique characteristics (such as color, structure, texture, or other properties) that make it different from the other layers in the soil profile

In situ Latin for 'the original position'

Soil horizons An identifiable soil unit due to color, structure, or texture

Soil profile The 'face' of a soil when it has been cut vertically that shows the individual horizons and soil properties with depth

Soil texture The way soil 'feels' when it is squeezed between the fingers or in the hand. The texture depends on the amount of sand, silt, and clay in the sample (particle-size distribution), as well as other factors

Translocation	The movement or transfer of material from one part of the soil profile to another
Transpiration	Water in plants escapes or transpires into the atmosphere as the leaf stomates open to exchange carbon for oxygen

See also: **Factors of Soil Formation:** Biota; Climate; Human Impacts; Parent Material; **Pedology:** Basic Principles

Further Reading

Birkeland PW (1999) *Soils and Geomorphology*, 3rd edn. New York: Oxford University Press.

Buol SW, Hole FD, McCracken RJ, and Southard RJ (1997) *Soil Genesis and Classification*, 4th edn. Ames, IA: Iowa State University Press.

Dixon JB and Weed SB (eds) (1989) *Minerals in the Soil Environment*, 2nd edn. Book Series No. 1. Madison, WI: Soil Science Society of America.

Gerrard AJ (1981) *Soils and Landforms*. London, UK: George Allen & Unwin.

Harland WB, Armstrong RL, Craig LE, Smith AG, and Smith DG (1990) *A Geologic Time Scale*. Cambridge, UK: Press Syndicate of University of Cambridge.

Jackson JA (ed.) (1997) *Glossary of Geology*, 4th edn. Alexandria, VA: American Geological Institute.

Jenny H (1941) *Factors of Soil Formation*. New York: McGraw-Hill.

Jenny H (1980) *The Soil Resource*. New York: Springer-Verlag.

Merrits DJ, Chadwick OA, and Hendricks DM (1991) Rates and processes of soil evolution on uplifted marine terraces, northern California. *Geoderma* 51: 241–275.

Schumm SA (1991) *To Interpret the Earth: 10 Ways to be Wrong*. Cambridge, UK: Cambridge University Press.

Soil Science Society of America (1997) *Glossary of Soil Science Terms*. Madison, WI: Soil Science Society of America.

FAUNA

T Winsome, University of California–Davis, Davis, CA, USA

Introduction

The soil fauna is tremendously diverse and range in size from the smallest protists to fossorial rodents. The focus here is upon the soil macrofauna, their natural history and role in ecosystem function.

Characterization of Soil Macrofauna

The soil macrofauna generally includes animals that are greater than 1 cm in length or 2 mm diameter. Many of the invertebrates share physiological adaptations to life in the soil, including reduced tolerance for desiccation, high temperatures, and light, but increased tolerance to carbon dioxide. Along with soil-dwelling vertebrates, many have evolved characteristic morphological adaptations to a fossorial lifestyle, including reduced or no eyes, and forelimbs modified for digging (Figure 1). Others have evolved streamlined body plans with limbs reduced or absent altogether to assist in tunneling.

Soil macrofauna may be differentiated between the euedaphic forms, such as earthworms, which spend their entire life cycles within the litter and soil horizons, and those that are transients, such as some beetles, which carry out their larval stage in the soil but live aboveground as adults. The macrofauna may also be differentiated on the basis of their preferred habitat within the soil profile. Epigeic macrofauna dwell within the litter layer and seldom burrow into the mineral soil. Endogenic macrofauna reside within the mineral soil horizons and may be further differentiated on the basis of whether they live in surface or subsurface horizons.

Finally, soil macrofauna may be classified in functional terms as: (1) saprovores and microbivores, which feed on dead and decaying matter and the associated microorganisms, (2) herbivores and granivores, which feed on live plants and seeds, (3) predators, and (4) omnivores, which feed on a variety of living and dead substrates (Figure 2). The remainder of this section provides a brief description of common, important members of each functional class. The families, genera, and species names included in the descriptions are intended as examples and are by no means exhaustive.

Saprovores and Microbivores

Millipedes (Diplopoda), isopods (Isopoda), roaches (Blattaria), termites (Isoptera), earthworms (Annelida) and some species of ants (Hymenoptera: Formicidae), beetles (Coleoptera), and flies (Diptera) are included in the saprovores and microbivores group.

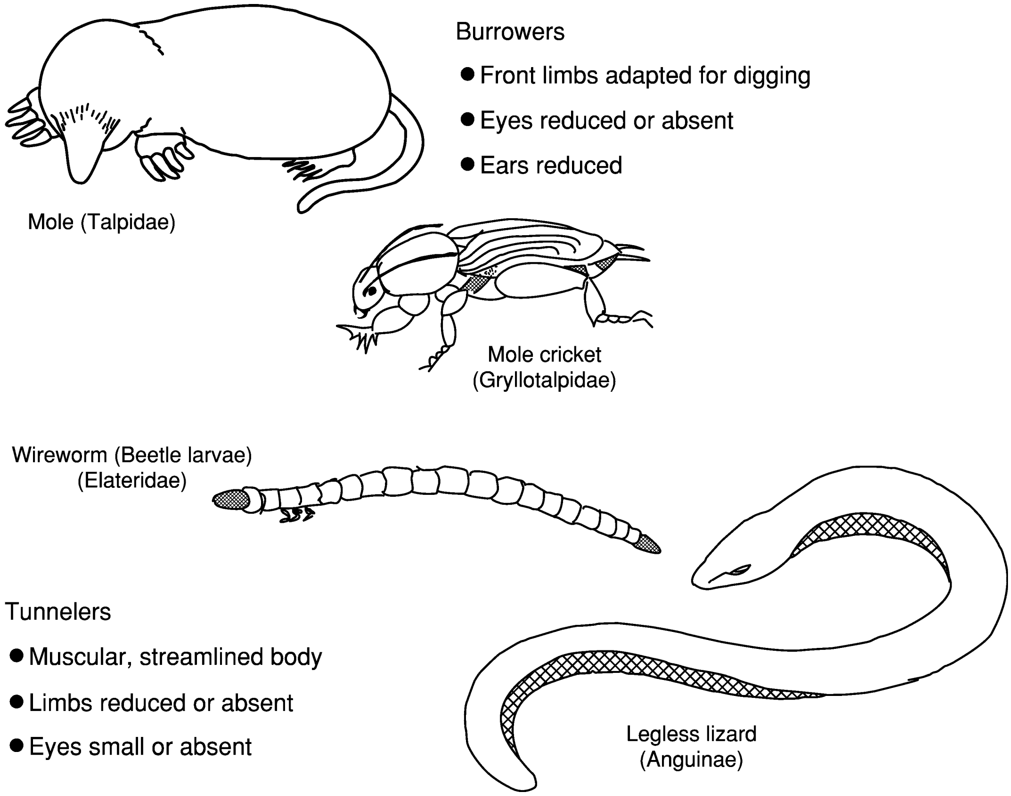


Figure 1 Morphological features for a life underground are shared by invertebrates and vertebrates alike. Animals that excavate soil, ''burrowers,'' have strong front limbs with large claws or claw-like appendages adapted for digging in soil or sand. Animals that push their way through the soil, ''tunnelers,'' have muscular, streamlined bodies with reduced or no appendages. Features common to both groups included reduced or no eyes, ears reduced in size, and sensory whiskers or appendages on or near the face.

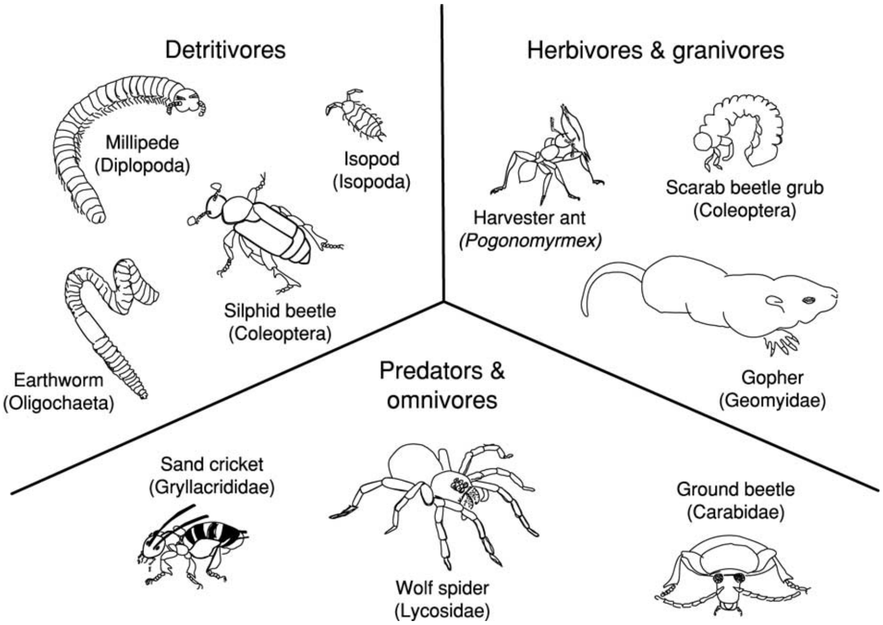


Figure 2 Commonly encountered soil macrofauna, grouped according to functional group. See text for details. (Not to scale.)

A few are generalist in their feeding habits and scavenge a wide variety of plant and animal residues, but most exhibit some degree of specialization with respect to diet. Many, perhaps most, of those that feed on plant residues derive their nutrition not from the residue itself but rather from the microorganisms that have colonized the residue, and a large number feed exclusively on fungi and other microorganisms.

The millipedes, isopods, and roaches are widely distributed in many habitats, including temperate and tropical forests, grasslands, and deserts. Densities vary from fewer than 10 to hundreds per square meter, depending on species, season, and location. Most isopod species are relatively small, 1–2 cm in length, and are generally epigeic. Millipedes can be either epigeic or endogeic, depending on species, and are amongst the largest soil macroinvertebrates, ranging from 5 to 6 cm in length for temperate species to greater than 20 cm for some tropical species. Most roaches are epigeic and can also be quite large, greater than 5 cm for subtropical and tropical species. Both groups feed primarily on plant residues and the associated microflora, although they will also feed on fallen fruit and animal dung, and isopods will feed on seedlings. In the process of feeding, millipedes, roaches, and isopods physically fragment organic residues, in a process known as comminution, and haphazardly distribute the fragments within the litter and surface soil horizons.

Beetle species of the family Scarabaeidae (scarab beetles) are important scavengers of dung and carrion. Both larvae and adults of other scarabaeid species and beetle families occur within dead and dying trees and are important in the early stages of wood decomposition. Some species of fly are specialists on the mushrooms of various species of fungi (Drosophilidae). The larvae of many flies (Calliphoridae and Sarcophagidae) and beetles (Scarabaeidae, Silphidae, Dermestidae) are necrophages and occur exclusively on carrion. Many of these species are useful in criminal forensics, as the timing of their arrival on a corpse is highly dependent on time of death, degree of decomposition, and location.

Termites and ants are social insects, with caste systems consisting of workers, soldiers, winged male and female reproductives, and a queen. They occur in almost all habitats and are particularly important in arid and semiarid ecosystems where they are the most abundant soil-dwelling animals. The size of termite and ant colonies varies widely, from a few dozen individuals to hundreds of thousands, and colony densities range from a few to more than 5000 ha^{-1}. Once established, colonies persist for the life of the queen, which may be up to 2 decades for some species. Termites are dependent upon microbial symbionts for nutrition. Wood-feeding termite species in some families (e.g., Kalotermitidae) digest cellulose with the help of protozoan symbionts that reside in special compartments within the gut. Some of these species are important pests and cause tremendous economic damage to houses and other structures. Many soil-feeding termite species (Termitidae) lack protozoan symbionts but rely instead upon a mixture of bacteria and fungi, with which they derive nutrition from humified soil organic matter. Other groups (e.g., *Macrotermes*) lack intestinal symbionts altogether and feed on decayed leaves and other substrates that have been colonized by fungi (*Termitomyces*) that the termites culture in special chambers within their nests. Detritivorous ants include the carpenter ants, which dismantle snags, stumps, and downed logs, and various litter-collecting species. Most of these species use these substrates as culture material for microorganisms within nest chambers in a manner analogous to the termites. Many species are also scavengers of dead animals and fallen fruit. Like millipedes, roaches, and isopods, termites and ants fragment and redistribute organic materials, but differ in that most collect organic materials in an organized fashion, transport them over relatively long distances, and concentrate them in storage chambers within their nest structures.

Herbivores and Granivores

The herbivorous members of the soil macrofauna include species of beetles, flies, ants, and fossorial vertebrates. The larvae of many species of beetles and flies feed on plant roots and seedlings, and some feed on seeds. Many soil-dwelling herbivores are important agricultural pests, but in natural communities may play an important role in plant community structure and species diversity by regulating plant populations. The leaf-cutter ants common in tropical rainforests can defoliate entire trees, cutting the leaves into small pieces and transporting them back to the nest for use in their fungus gardens. Granivorous ant species serve a critical ecosystem function by transporting and burying seeds. They collect the seeds of many different species and store them in nest chambers, where they are used as food for larvae. Those that are dropped along the way or escape predation go on to germinate or remain protected within the soil seed bank. Fossorial mammalian herbivores include the pocket gopher (*Thomomys*) and related genera within the Rodentia. Some, such as the bathyergid mole-rats of South Africa, are social and form huge, subterranean colonies. They feed on roots, bulbs, and plant material that they pull down from the surface, and often store food in cache chambers within the soil.

Predators and Omnivores

Predatory soil macroinvertebrates include spiders (Aranae), scorpions (Scorpiones), centipedes (Chilopoda), and some species of beetles, ants, and flies. Many are ambush predators; the trap-door spiders (Ctenizidae), for instance, construct a burrow with a 'trap door' of silk, soil, and litter residues behind which they hide in wait for prey to approach. Others, such as centipedes, ground beetles (Carabidae), and ants actively hunt for prey. These predators vary widely in size, from 1 to 2 cm to more than 15 cm in length for the largest centipedes and scorpions. Fire ants (*Solenopsis*) are capable of mobbing animals many times their own size, including small vertebrates. The victim is then dismembered and transported back to the nest. Predatory vertebrates include moles (Talpidae), shrews (Soricidae), and fossorial lizards and snakes (Reptilia). Many species of moles are specialized predators of earthworms, and often paralyze their prey with a bite to the head and then store them in cache chambers. Shrews are extremely small, epigeic mammals, most less than 10 cm in length. They have voracious appetites and feed on arthropods, earthworms, and small amphibians and reptiles, often attacking animals many times their own size. Some species have toxic compounds in their saliva that help to subdue prey. Fossorial reptiles feed on rodents and other soil fauna.

Common omnivores include mole crickets (Gryllotalpidae) and other members of the Orthoptera. These species feed on live plants, other soil fauna, and also scavenge dead plant residues. Omnivory is unusually common in soil food webs, and many of the species nominally assigned above as predators or detritivores in fact consume a wide variety of foods. Centipedes, for instance, long thought to be obligate predators, have recently been shown to include plant residues as part of their diet. The reason for the high incidence of omnivory in soil fauna remains poorly understood, but may include high mineral requirements or the need to supplement generally poor quality foods.

The Role of Soil Macrofauna in Ecosystem Function

Members of the soil macrofauna carry out three related functions that have ecosystem-level impacts: decomposition and nutrient cycling, modification of soil structure, and participation in both aboveground and belowground food webs.

Decomposition and Nutrient Cycling

Soil macrofauna have both direct and indirect effects on litter decomposition and nutrient cycling, but their net effects are best considered within the context of the entire soil fauna community (Figure 3). As a class, detritivores are characterized by relatively low assimilation efficiencies; estimates vary but range from 10 to 30%. (Some termite species are a notable

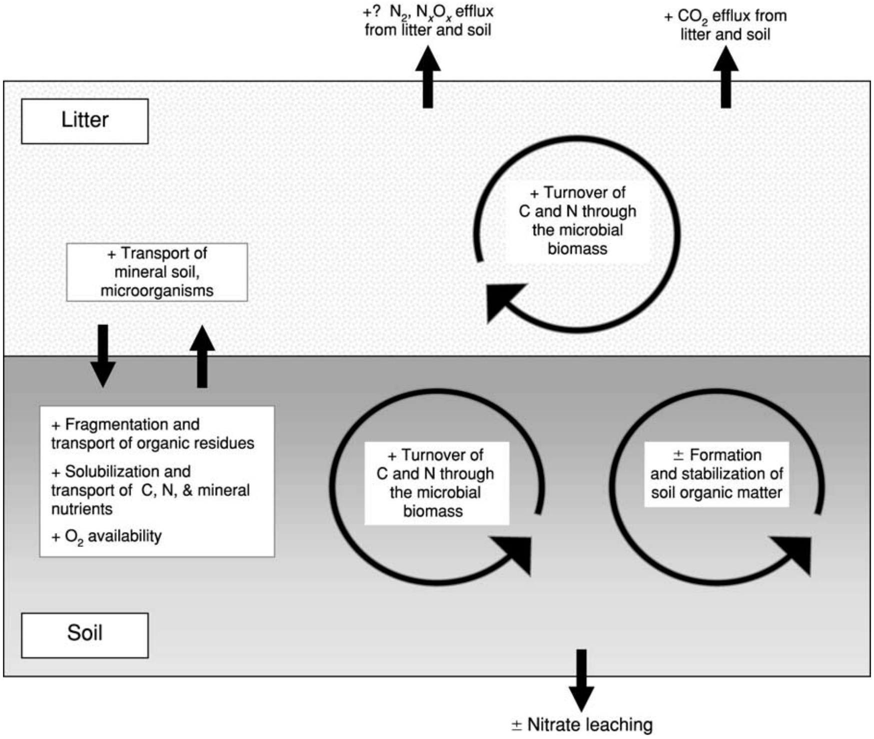

+? N_2, N_xO_x efflux from litter and soil
+ CO_2 efflux from litter and soil
Litter
+ Turnover of C and N through the microbial biomass
+ Transport of mineral soil, microorganisms
+ Fragmentation and transport of organic residues
+ Solubilization and transport of C, N, & mineral nutrients
+ O_2 availability
+ Turnover of C and N through the microbial biomass
± Formation and stabilization of soil organic matter
Soil
± Nitrate leaching

Figure 3 General effects of soil fauna on litter decomposition and nutrient cycles. A 'plus' symbol indicates enhancement; 'minus' indicates reduction. The 'plus-or-minus' symbol indicates that both effects can occur, depending on the organism(s) and ecosystem. The '+?' next to the efflux of gaseous forms of N (N_2 and N_xO_x) shows that evidence is inconclusive or contradictory.

exception and have assimilation efficiencies that exceed 50%). This requires a high consumption rate, and the soil fauna in total processes 20–40% of the annual litter input.

The process of litter decomposition is greatly facilitated by the physical comminution and redistribution of plant residues by soil macrofauna. These 'macroshredders' break down leaf and woody residues into smaller particles and, along with other soil animals, transport them from the surface deeper within the litter layer and soil mineral horizons. This physical fragmentation and transport of residues renders them more accessible as substrates for smaller fauna and soil microorganisms. Soluble nutrients are more readily leached into the soil from these small particles and enhance both microbial activity and plant growth. Decomposition rates are greatly reduced in the absence of soil fauna, especially in forests, where the volume of litter input is large and tends to be of relatively low quality, and in arid environments where climate limits microbial activity.

Soil macroinvertebrates also stimulate microbial activity through the production of large numbers of fecal pellets that are deposited throughout the soil profile and which serve as resource-rich microsites for microbial activity. Fecal pellets are generally associated with higher concentrations of soluble C, mineral forms of N and P, and available forms of mineral nutrients such as Ca, Mg, and K. The fecal pellets produced by earthworms and other soil-dwelling macroinvertebrates consist of a mixture of mineral soil particles and organic material, and contribute to the formation of stable soil aggregates. They are often characterized by numerous, small voids that provide habitable pore space for microorganisms (Figure 4).

The net influence of soil macrofauna on nutrient cycles at the ecosystem level is complex and remains poorly understood. Soil animal activity is generally associated with higher rates of C, N, and P mineralization. Termite mounds in arid regions of Australia, for instance, may account for up to 20% of the ecosystem flux of CO_2; in more mesic ecosystems, soil macrofauna contribute proportionally much less

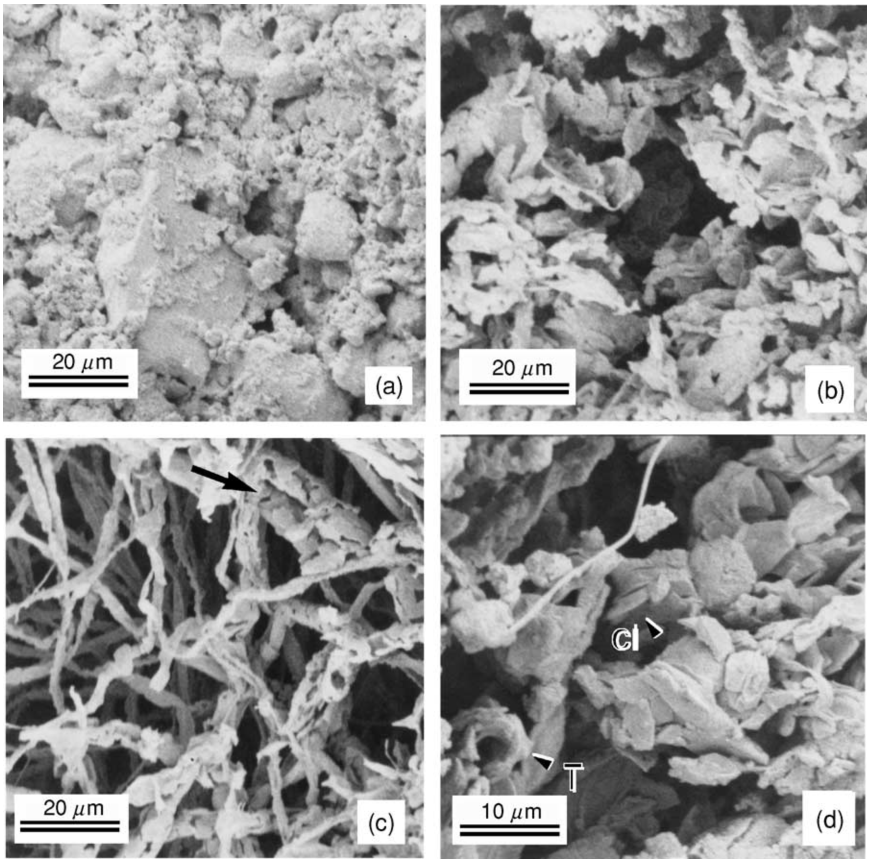


Figure 4 A comparison of micromorphological features of bulk soil and earthworm casts: (a) interior of bulk soil aggregate. Note dense, featureless structure; (b) interior of cast, with foliate structures surrounding voids; (c) interior of cast, showing fungal filaments bridging voids; (d) interior of cast, showing kaolinite domains (Cl) and clay particles forming a tube around fungal filaments (T). (Adapted from Winsome T and McColl JG (1998) Changes in chemistry and aggregation of a California forest soil worked by the earthworm *Argilophilus papillifer* Eisen (Megascdecidae). *Soil Biology & Biochemistry* 30: 1677–1687, with permission.)

to CO_2 flux, within the range of 2–10%. Experiments comparing N mineralization in soils with and without fauna (e.g., millipedes, earthworms) have shown that mineralization is enhanced by 10–30% in the presence of fauna. Through litter fragmentation and the production of fecal pellets, the fauna increase microbial production and activity, but through microbivory also regulate the degree to which nutrients are immobilized within the microbial biomass. While the fragmentation of detrital material enhances mineralization, the concomitant process of burial within the soil ensures that a proportion of it that would otherwise decompose on the surface is physically protected within soil aggregates and eventually contributes to the formation of stable soil organic matter. Feedback from all these processes may serve to maintain net ecosystem productivity.

Soil Modification

The physical disturbance and modification of soil horizons by animal activity is defined as 'biopedturbation.' This disturbance is extremely important in maintaining both spatial and temporal heterogeneity in ecosystems and has profound impacts on the structure and species diversity of both plant and animal communities. There are four general forms of biopedturbation, distinguished on the basis of spatial and temporal scale: (1) soil eject mounds, (2) burrow systems, (3) nest mounds, and (4) Mima and Mima-like terrain (Figure 5 and Table 1).

Soil eject mounds consist of the loose piles of soil thrown up on the surface as the animal burrows. These mounds are created by a wide variety of soil fauna, from ants and mole crickets to fossorial rodents. Accordingly, the size of these mounds varies from a few centimeters in diameter to greater than a meter. At any one point in time, the number of mounds formed by all species combined may total several hundred to thousands per hectare and occupy up to 10% of the surface. Over the course of time, the entire soil surface may thus be worked in this way. Burrow systems are created by various species of ants, termites, earthworms, and mammals, and consist of a network of burrows that run more or less parallel to

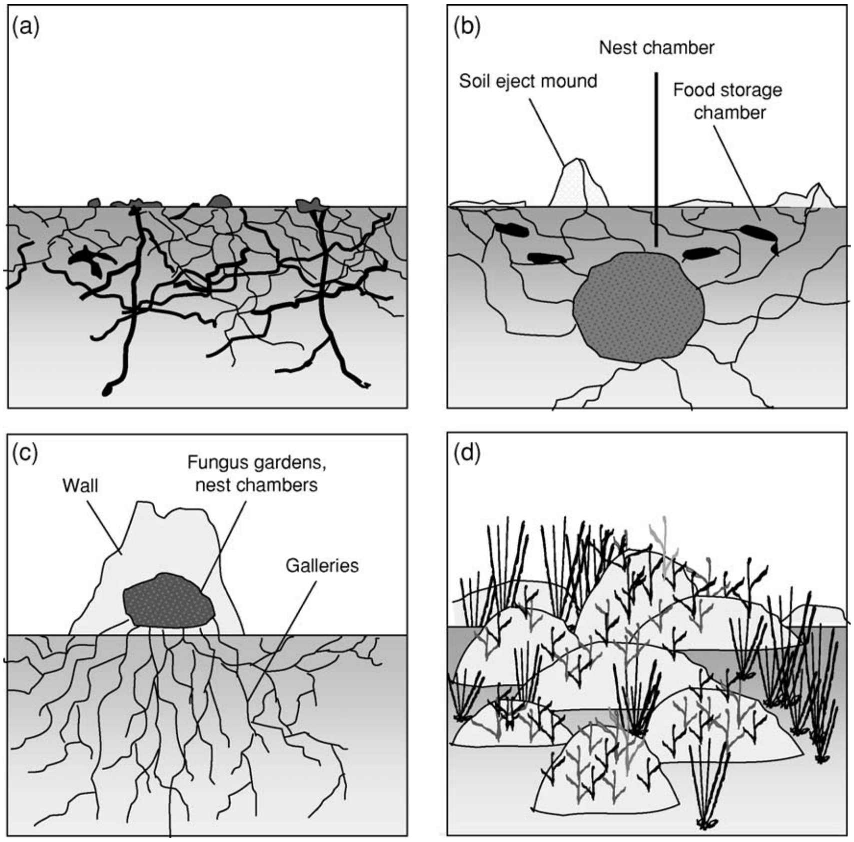


Figure 5 Forms of soil modification mediated by soil macrofauna: (a) burrow systems are dense, interconnected tunnels, created by a variety of soil animals, ranging from ants and termites to gophers; (b) subterranean nest structures are built by ants, termites, and rodents. The figure shows a central nest chamber, associated galleries and storage chambers, and soil eject mounds on the surface; (c) nest mounds are constructed by both ants and termites. Shown here is a mound constructed by fungus-cultivating termites; (d) Mima-type mound terrain is a landscape form, characterized by dense aggregations of low mound structures. Plant communities on the surface of the mounds are often very different from those that inhabit the intermound soils.

Table 1 Density, mass, or volume, and amount of soil displaced or transported to the surface by biopedturbation

Disturbance type, organism(s) and location	*Density (n ha^{-1})*	*Mass (Mg ha^{-1}) or volume (m^3 ha^{-1})*	*Soil displaced or transported to surface (Mg ha^{-1} y^{-1})*	*Source*[a]
Fossorial mound				
Tuco-tuco (*Ctenomys azarae*); Argentina	1050	14.7 Mg ha^{-1}	520	Roig *et al.*, 1988
Burrow system				
Earthworms, various species; USA, Europe, Africa, Asia	50 000–8 000 000[b]	—	2–91 (temperate) 3–507 (tropical)	Edwards and Bohlen, 1996
Nests and nest-mound systems				
Ants, various species; Australia	5100–15 400		0.1–0.31	Lobry de Bruyn and Conacher, 1990, 1994; Wiken *et al.*, 1976; Lee and Wood, 1971
Ants, (*Formica fusca*); British Columbia	1150		54 m^3 ha^{-1}	
Termite, mixed species; tropics	1–200	0.1–300 Mg ha^{-1}	1–4	
Mima-mound terrain				
Pocket gopher (*Thomomys bottae*); USA	20–50	15.0–74.8 m^3 ha^{-1}	101	Cox, 1984, 1990

[a] Sources:

Cox GW (1984) The distribution and origin of Mima mound grasslands in San Diego County, California. *Ecology* 65: 1397–1405.

Cox GW (1990) Soil mining by pocket gophers along topographic gradients in a Mima moundfield. *Ecology* 71: 837–843.

Edwards CA and Bohlen PJ (1996) *Biology and Ecology of Earthworms.* London, UK: Chapman and Hall.

Lee K and Wood TG (1971) *Termites and Soils.* New York: Academic Press.

Lobry de Bruyn LA and Conacher AJ (1990) The role of termites and ants in soil modification: a review. *Australian Journal of Soil Research* 28: 55–93.

Lobry de Bruyn LA and Conacher AJ (1994) The bioturbation activity of ants in agricultural and naturally vegetated habitats in semi-arid environments. *Australian Journal of Soil Research* 32: 555–570.

Roig VG, Loyarte Gonzalez MM, and Rosa ML (1988) Ecological analysis of mound formation of the Mima type in Rio Quinto, Province of Córdoba, Argentina. *Studies on Neotropical Fauna and Environment* 23: 103–116.

Wiken EB, Boersma LM, Lavkulich LM, and Farstead L (1976) Biosynthetic alteration in a British Columbia soil by ants (*Formica fusca* Linn). *Soil Science Society of America Proceedings* 40: 422–426.

[b] Measured on horizontal soil surfaces at various depths.

the soil surface. They can extend over many hectares and persist for decades, maintained and expanded by successive generations of animals.

Many soil macrofauna build large nests that are characteristic features of the landscape. Termites construct some of the most spectacular and complex nests. *Macrotermes* in central African savanna construct nests with 'chimneys' that can reach 9 m in height. In Australia, *Amitermes* constructs mounds several meters in height that have a flattened, fluted shape, while in tropical Africa *Cubitermes* constructs mushroom-shaped mounds. Although these shapes appear bizarre, they are in fact carefully engineered to regulate temperature and humidity within the nest or, in the case of the mushroom-shaped mounds, to prevent rainwater from collecting on the surface of the nest. In the process of building their nests, termites transport soil fines (clay-sized soil mineral particles) from the soil subsurface to the surface, altering the particle-size distribution of the soil in the nest area. Depending on species, they mix the soil with saliva, their feces, and/or organic matter and use this material to construct the galleries and chambers that form the interior of the nest. Termites that construct subterranean nests also transport soil fines and mix it with various materials to construct their nests. Their nests can be very large; for example, *Odontotermes* in Africa and India construct nests several meters in diameter, with foraging galleries that extend out 50 m or more from the central nest. For both mound-building and subterranean-nesting species, the area occupied by foraging galleries, storage chambers, and brood chambers can be quite extensive; for example, one nest system constructed by *Macrotermes michaelseni* in Kenya covered an area of 8000 m^2 and was associated with 6 km of galleries and 72 000 storage chambers. Some desert species construct vertical galleries that descend as much as 70 m to water sources.

Many species of ants also create large, elaborate nests from soil, with extensive subterranean galleries and chambers. Some species of carpenter ant build their nests within and beneath snags and stumps, cutting the wood away in small cubes. Over time, their nests come to resemble piles of sawdust. There are also a large number of vertebrates that forage aboveground for plants and seeds, but nest in the soil and have important impacts on soil properties. The black-tail prairie dog (*Cynomys ludovicianus*) of North America forages at the surface, but spends much of its time belowground in extensive burrow systems known as 'towns.' Other mammals with similar life habits include the kangaroo rat (*Dipodomys*) of the American southwest, and the gerbil (*Tatera* and related genera) of Africa and the Middle East.

Mima and Mima-like terrain is characterized by low mounds, ranging from less than 1 m to more than 3 m in height and approximately 1–20 m in diameter, that occur in a regular formation across the landscape. The name 'Mima' derives from their type locality, Mima Prairie, in the state of Washington. They occur worldwide, however, and their various origins have long been the subject of some controversy. Competing hypotheses focus either on strictly geomorphologic processes or on the long-term effect of animal activity. Whatever the cause of their origin, recent research has confirmed that the maintenance of these landscapes is often due to animal activity. For example, the heuweltjies landscape (pronounced hear-vill-keys) near Cape Town in South Africa consists of low, nearly circular mounds some meters in diameter that are up to 4000 years old and have been maintained by successive generations of termites. Similar landscapes in wetlands in Argentina are associated with fire ants (*Solenopsis*). Rodent activity is also frequently associated with mima mound terrain; some of the best-studied examples include pocket gophers (*Thomomys*) in southern California and tuco-tucos (*Ctenomys*) in Argentina.

An important feature of systems characterized by biopedturbation is a marked difference between the animal-worked soil and surrounding soil in soil physical, chemical, and biological properties (Table 2). Over time, the net effects of biopedturbation on soil physical properties are generally to reduce bulk density and improve water infiltration. These effects vary, however, according to disturbance type, species, location, and whether or not the structure is actively maintained or has been abandoned. Eject mounds serve to mix subsurface and surface soils, but are loosely structured and thus easily eroded by wind or water unless quickly colonized by plants.

The burrow systems of the smaller soil macrofauna, such as ants, termites, and earthworms, create numerous small channels that increase water infiltration if they are open to the soil surface. The burrow systems of termites and ants are often kept sealed from the surface while the colony is active and do little to improve infiltration, but greatly enhance water infiltration once they are abandoned and open to the surface. The larger burrow systems created by mammals may improve preferential flow, but may also increase surface erosion and runoff if the animals are active in removing the overlying vegetation and compacting the soil. Nest mounds and mima mounds tend to consist of fine-textured soil material, often underlain by a layer of gravels and stones. The fine material used in termite nest construction may become compacted over the mound and form a seal over the soil surface as it erodes away from the nest, effectively reducing water infiltration and negatively impacting plant growth. In areas with a lot of termite activity, much of the soil surface horizon may consist of material worked by termites that has eroded away from nests constructed over many thousands of years.

The burrow systems created by earthworms and millipedes are generally enriched in C, N, and other nutrients, due to the incorporation of organic materials within the mineral soil and the deposition of fecal pellets. The nest mound soils of ants, termites, and rodents are also often enriched in C, N, and mineral nutrients, due to the concentration of urine, feces, and stored food within the mound. Termite mounds, in particular, are enriched in Ca relative to surrounding soil, so much so that in parts of central Africa traditional farmers collect the soil from termite mounds and apply it to their fields for lime and fertilizer.

The soil biological properties of burrow systems and nest mounds depend on the species, ecosystem, and season. In the nest mounds of kangaroo rats, the concentration of organic matter and nutrients within the mound stimulates soil microbial activity relative to the surrounding soil; conversely, in some termite nests the high polyphenol content of the nest material suppresses microbial activity. Ant and termite nests often support large inquiline ('animals that share a common space') communities. The numbers of protozoa, nematodes, microarthropods, and beetles may be much greater in and around the vicinity of ant nests than the surrounding soil, due to the accumulation of organic debris and availability of moisture. Many of these species are dependent upon their host nests and occur nowhere else. Mima-like mounds provide space for burrowing mammals such as ardvaarks in the South African heuweltjies and armadillos in Argentina.

These differences in soil properties in turn affect the structure and diversity of associated plant communities. The porous, nutrient-enriched burrow systems created by earthworm and millipede activity invariably improve plant growth. Large eject mounds in

Table 2 The impacts of various types of biopedturbation on soil properties, plant communities, and other soil biota

Disturbance type organism and location	*Effect on soil properties*	*Effects on plant community*	*Effects on other biota*	*Source*[a]
Fossorial soil eject mound				
Pocket gopher (*Thomomys bottae*); California serpentine grasslands; USA	Reduced bulk density, increased Ni, Mn, and other minerals associated with serpentine	Fewer exotic species not adapted to serpentine soils		Koide *et al.*, 1987
Burrow system				
Earthworm (Lumbricidae); temperate grasslands; USA, Europe	Reduced bulk density, increased water infiltration, increased C, N, mineral N	Increased pasture biomass		Reviewed by Lee, 1985
Bathyergid mole rats		Reduced biomass; a change in community structure		Reichman and Jarvis, 1989
Nest mound				
Kangaroo rat (*Dipodomys*); Chihuahuan desert; USA	Reduced bulk density, decreased moisture, increased C, N	Fewer perennials, increased biomass annuals, increased species diversity, number of annuals	Increased microbial biomass and respiration increased microarthropods, nematodes, insects, lizards, other rodents	Ayarbe and Kieft, 2000; reviewed by Whitford and Kay, 1999
Harvester ant (*Pogonomyrmex barbatus*); Chihuahuan desert; USA	Litter accumulation and burial, increased mineral N, P, K	Seed burial and dispersal	Increased protozoa, microarthropods	Wagner *et al.*, 1997
Mima-type mound system				
Fire ant, (*Solenopsis richteri*); Argentina	Decreased salinity	Salt-tolerant community on mounds	Habitat for armadillo	Cox, 1992
Termites, (*Odontotermes*); heuweltjies, South Africa	Increased N, K, Ca, Mg, increased disturbance	Disturbance-adapted plant community on mounds, increased productivity	Habitat for ardvaark and other mammals	Esler and Cowling, 1985

[a] Sources:

Ayarbe JP and Kieft TL (2000) Mammal mounds stimulate microbial activity in a semiarid shrubland. *Ecology* 81: 1150–1154.

Cox GW (1992) Fire ants (Hymenoptera: Formicidae) as major agents of landscape development. *Environmental Entomology* 21: 281–286.

Esler KJ and Cowling RM (1995) The comparison of selected life-history characteristics of Mesembryanthema species occurring on and off Mima-like mounds (heuweltjies) in semiarid southern Africa. *Vegetatio* 116: 41–50.

Koide RT, Huenneke LF, and Mooney HA (1987) Gopher mound soil reduces growth and affects ion uptake of two annual grassland species. *Oecologia* 72: 284–290.

Lee KE (1985) *Earthworms: Their Ecology and Relationships with Soils and Land Use.* Sydney, Australia: Academic Press.

Reichman OJ and Jarvis JUM (1989) The influence of three sympatric species of fossorial mole-rats (Bathyergidae) on vegetation. *Journal of Mammalogy* 70: 763–771.

Wagner D, Brown MJF, and Gordon DM (1997) Harvester ant nests, soil biota and soil chemistry. *Oecologia* 112: 232–236.

Whitford WG and Kay FR (1999) Biopedturbation by mammals in deserts: a review. *Journal of Arid Environments* 41: 203–230.

grasslands may bury and kill the vegetation on which they fall, and thus create vegetation gaps that favor colonization by fast-growing, pioneer species. In the South African heuweltjies, the mounds support a plant community adapted to chronic disturbance, while the intermound space supports an entirely different community. In the fire ant-maintained mound terrain in Argentina, a high, brackish water table restricts the vegetation in intermound areas to salt-tolerant species, while the mounds are markedly lower in salinity and support a diverse community of salt intolerant species. Thus, the spatial heterogeneity created by the mounds results in an overall increase in plant species diversity in these ecosystems. However, with the recent spread of invasive plant species throughout many parts of the world, these mound systems often inadvertently promote the invasion of fast-growing, exotic plant species at the expense of native species. For example, gopher eject mounds in California grasslands are quickly colonized by fast-growing, exotic annual grasses at the expense of native perennials. Mima-like mounds created by rodents in Argentina are the preferred habitat for exotic trees and shrubs.

Participation in Belowground and Aboveground Food Webs

The soil macrofauna have been described as 'facilitators' or 'regulators' of interactions between other groups of organisms, and this may well be one of their most important roles in ecosystem function. Species that burrow extensively within the rhizosphere or transport organic material from the surface into soil horizons transport microorganisms to new habitats. For instance, a number of soil animals have been shown to disseminate the spores of mycorrhizal fungi throughout the rhizosphere, increasing the inoculation of plant roots. Soil animals have also been shown to serve as important vectors and reservoirs for microorganisms that cause both plant and animal diseases, including those that affect humans. Recent research has shown that the transfer of genetic material between different strains of bacteria is facilitated by the soil-mixing activity of earthworms, which may have important implications for the spread of disease-causing organisms.

Soil macrofauna serve as important food sources for a wide variety of animals, many of which are specialist feeders on certain species. Termites and ants are high in protein and fat and are consumed by humans in some cultures, as well as by a large number of other mammals, birds, and reptiles. Earthworms are also rich in protein and are consumed by many mammals and birds. In polluted environments, soil macrofauna form a critical link in the transfer of toxic compounds from the soil to the aboveground biota in a process known as bioaccumulation. In the course of feeding on soil and litter residues, detritivores accumulate in their bodies any heavy metal or chemical pollutants that may be present in low levels in the environment. The spiders, centipedes, and small mammals that feed on these animals further concentrate these compounds within their tissues. The birds and mammals that feed upon the soil animals ingest these compounds in concentrations that greatly exceed those in the soil and are frequently poisoned, or simply concentrate the compounds yet further and pass them along to their predators and scavengers.

Soil macrofauna play important and sometimes critical roles in ecosystem nutrient cycles, the shape and structure of landscapes, and the flow of energy and matter between belowground and aboveground components of ecosystems. They are also fascinating in their own right. Few groups of organisms are more diverse and complex with respect to their evolution, behavior, and ecology. From termites to trapdoor spiders, the capacity of the soil macrofauna to modify their physical and biological environment stands unparalleled in the animal kingdom.

See also: **Forensic Applications**; **Vadose Zone:** Microbial Ecology

Further Reading

Abe T, Bignell DE, and Higashi M (eds) (2000) *Termites: Evolution, Sociality, Symbioses, Ecology.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Benckiser G (ed.) (1997) *Fauna in Soil Ecosystems: Recycling Processes, Nutrient Fluxes, and Agricultural Production*. New York: Marcel Dekker.

Coleman DC and Crossley DA (eds) (1996) *Fundamentals of Soil Ecology.* London, UK: Academic Press.

Cox GW, Contreras LC, and Milewski AV (1995) Role of fossorial mammals in community structure and energetics of Pacific Mediterranean ecosystems. In: Kalin Arroyo MT, Zedler PH, and Fox MD (eds) *Ecology and Biogeography of Mediterranean Ecosystems in Chile, California, and Australia*, pp. 383–398. New York: Springer-Verlag.

Hopkins SP and Read HJ (1992) *The Biology of Millipedes.* Oxford, UK: Oxford University Press.

Jones CG, Lawton JH, and Shachak M (1997) Positive and negative effects of organisms as physical ecosystem engineers. *Ecology* 78: 1946–1957.

Lal R (1987) *Tropical Ecology and Physical Edaphology.* Chichester, UK: John Wiley & Sons, Ltd.

Lee KE (1985) *Earthworms: Their Ecology and Relationships with Soils and Land Use*. Sydney, Australia: Academic Press.

Lee KE and Wood TG (1971) *Termites and Soils*. New York: Academic Press.

Lobry de Bruyn LA (1999) Ants as bioindicators of soil function in rural environments. *Agriculture, Ecosystems and Environment* 74: 425–441.

Lobry de Bruyn LA and Conacher LA (1990) The role of termites and ants in soil modification: a review. *Australian Journal of Soil Research* 28: 55–93.

Meadows PS and Meadows A (eds) (1991) *The Environmental Impact of Burrowing Animals and Animal Burrows*. Oxford, UK: Clarendon Press.

Wallwork JA (1976) *The Distribution and Diversity of Soil Fauna*. London, UK: Academic Press.

Wardle DA (2002) *Communities and Ecosystems: Linking the Aboveground and Belowground Components.* Princeton, NJ: Princeton University Press.

Whitford WG and Kay FR (1999) Biopedturbation by mammals in deserts: a review. *Journal of Arid Environments* 41: 203–230.