



PRACTICAL APPLICATIONS OF GIS FOR ARCHAEOLOGISTS

A PREDICTIVE MODELING KIT

EDITED BY
KONNIE L. WESTCOTT AND R. JOE BRANDON

CD-ROM
INCLUDED

**Also available as a printed book
see title verso for ISBN details**

Practical Applications of GIS for Archaeologists

Practical Applications of GIS for Archaeologists

A Predictive Modeling Toolkit

EDITED BY

KONNIE L. WESCOTT
R. JOE BRANDON



First published 2000
by Taylor & Francis
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Taylor & Francis Inc
325 Chestnut Street, 8th Floor, Philadelphia PA 19106

Taylor & Francis is an imprint of the Taylor & Francis Group

This edition published in the Taylor & Francis e-Library, 2005.

“To purchase your own copy of this or any of Taylor & Francis or Routledge’s collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk.”

© 2000 Konnie L. Westcott and R. Joe Brandon

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Westcott, Konnie L.

Practical applications of GIS for archaeologists: a predictive modeling kit/Konnie L. Westcott and R. Joe Brandon.
p. cm.

Includes bibliographical references

1. Archaeology—Data processing. 2. Archaeology—Mathematical models. 3. Archaeology—Computer simulation. 4. Geographic information systems. 5. Excavations (Archaeology)—Data processing.

I. Brandon, R. Joe, 1967–. II. Title. III. Title: Practical applications of geographic information systems for archaeologists.

CC80.4.W47 2000

930.1 028–dc21 99–13217

CIP

ISBN 0-203-21213-4 Master e-book ISBN

ISBN 0-203-26953-5 (Adobe eReader Format)

ISBN 0-7484-0830-4 (Print Edition)

To our parents
To Jim and Jackie
To Theresa, Linda, Dan and Kristy

Contents

<i>Contributors</i>	viii
<i>Preface</i>	x
R.Joe Brandon	
<i>Acknowledgments</i>	xii
Konnie L.Wescott and R.Joe Brandon	
1 Introduction	1
Konnie L.Wescott	
2 A Predictive Model of Archaeological Site Location in the Eastern Prairie Peninsula	6
Robert E. Warren and David L.Asch	
2.1 Introduction	6
2.2 Predictive modeling	7
2.3 Materials and methods	10
2.4 Results	18
2.5 Discussion and conclusions	27
3 The Application of GIS Predictive Site Location Models within Pennsylvania and West Virginia	37
Richard B.Duncan and Kristen A.Beckman	
3.1 Introduction	37
3.2 Background	38
3.3 Predictive model development	39
3.4 Additional GIS Predictive Models	49
3.5 Discussion	53
3.6 Conclusions	56
4 Using a GIS to Model Prehistoric Site Distributions in the Upper Chesapeake Bay	62

Konnie L.Wescott and James A.Kuiper

4.1	Introduction	62
4.2	Aberdeen proving ground	63
4.3	The model	64
4.4	Results	71
5	Protecting Cultural Resources through Forest Management Planning in Ontario Using Archaeological Predictive Modeling	78
	Luke Dalla Bona	
5.1	Introduction	78
5.2	Background	79
5.3	Modeling methodology	80
5.4	Model research and development	83
5.5	Pilot projects	84
5.6	Oil and water can mix! Integrating archaeology into forest management planning	99
5.7	Summary	104
6	Considerations of Scale in Modeling Settlement Patterns Using GIS: An Iroquois Example	107
	Kathleen M. and Sydoriak Allen	
6.1	Introduction	107
6.2	Spatial scale and kinds of problems	108
6.3	Global, regional, and local views of the Iroquoian world	109
6.4	GIS and spatial scale: global, regional, and local views	111
6.5	Central New York region: regional and local	115
6.6	Conclusions	117
7	Construction of Digital Elevation Models for Archaeological Applications	121
	Jon B.Hageman and David A.Bennett	
7.1	Introduction	121
7.2	Why should the archaeologist care about interpolation?	122
7.3	What is interpolation?	123
7.4	Selecting an interpolation algorithm	126

7.5	A Belizean case study	127
7.6	Conclusion	132
8	The State of the Art in “Inductive” Predictive Modeling: Seven Big Mistakes (and Lots of Smaller Ones) James I.Ebert	137
8.1	Introduction	137
8.2	GIS is revolutionizing predictive modeling	138
8.3	Predictive modeling predicts and models the past	138
8.4	What we want to predict is site location	139
8.5	Proximity to environmental variables is important	140
8.6	Maps contain environmental variables	140
8.7	Map data is inaccurate	141
8.8	The accuracy of inductive predictive models can be determined	141
9	GIS Applications in Archaeology: Method in Search of Theory Tim Church, R.Joe Brandon and Galen R.Burgett	144
9.1	Current use of GIS in archaeology	144
9.2	Correlative predictive models	145
9.3	The resource landscape	152
9.4	An alternative: explanatory models from a landscape perspective	155
9.5	Tools to tackle landscapes	156
9.6	The temporal dimension	157
9.7	Scale and grain	158
9.8	The role of climate	158
9.9	Geomorphological processes and their impact	159
9.10	Summary	159
9.11	Concluding remarks	160
	<i>Index</i>	167

Contributors

Kathleen M.Sydoriak Allen
Department of Anthropology
3H01 Forbes Quad
University of Pittsburgh
Pittsburgh, PA 15260–7400 USA David L.Asch
Illinois State Museum
1011 East Ash Street
Springfield, IL 62703–3535 USA Kristen A.Beckman
Skelly and Loy, Inc.
520 Seco Road
Monroeville, PA 15146 USA David A.Bennett
Department of Geography
The University of Kansas
Lawrence, KS 66045–2121 USA R.Joe Brandon
Center for Advanced Spatial Technologies
Room 12 Ozark Hall
University of Arkansas
Fayetteville, AR 72701 USA Galen R.Burgett
U.S.Fish and Wildlife Service
423 Hatchery Circle
Spearfish, SD 57783 USA Tim Church
Lone Mountain Archaeological
Services, Inc.
#5 Butterfield Trail
Suite F
El Paso, TX 79906 USA Luke Dalla Bona
Woodland Heritage Services
548 Queen Street, 2nd Floor
Sault Ste. Marie, Ontario
Canada P6A 2A1 Richard B.Duncan
Skelly and Loy, Inc.
520 Seco Road

Monroeville, PA 15146 USA James I.Ebert
Ebert & Associates, Inc.
3700 Rio Grande Blvd. NW
Suite 3

Albuquerque, NM 87107–3042 USA Jon B.Hageman
Department of Anthropology
Southern Illinois University

Carbondale, IL 62901–4502 USA James A.Kuiper
Argonne National Laboratory
9700 South Cass Avenue
EAD, Building 900

Argonne, IL 60439–4832 USA Robert E.Warren
Illinois State Museum
1011 East Ash Street

Springfield, IL 62703–3535 USA Konnie L.Wescott
Argonne National Laboratory
9700 South Cass Avenue
EAD, Building 900
Argonne, IL 60439–4832 USA

Preface

During the 1995 Society for American Archaeology (SAA) Meetings in Minneapolis, Minnesota, I caught up with my old friend Konnie Wescott, who happened to be presenting a paper in a general session I was attending on Information Management and Remote Sensing. It appeared that our research interests had crossed paths again. While attending Northern Illinois University in the late 1980s, we had both worked at the site of Copan, Honduras, under Dr. William Fash, and now we found ourselves pursuing work tied to the many facets of geographic information systems (GIS).

Konnie had continued her affiliation with Argonne National Laboratory. As a principal investigator, she discovered the myriad of potential applications of GIS in archaeology, whether it be for predictive modeling, cultural resources management, or, even more generally, the environmental planning and management of both natural and cultural resources. I was coming in from a more technical direction, having worked at Fort Bliss, Texas, researching the relationship of archaeological sites with the dynamic desert landscape.

We outlined a plan of attack for the remainder of the meeting by which the two of us would be able to cover most of the GIS papers being presented. In the afternoons we would meet to discuss the pros and cons of the papers we had seen.

By Saturday afternoon we were both exhausted, and disappointed. With the exception of a few papers, it seemed the current use of GIS in archaeology was limited to including this catchy buzzword in the title of the presentations. We found that archaeologists, instead of using GIS as a tool to explore human interaction with the prehistoric and historic landscape, utilized GIS for nothing grander than “gee-whiz” visualization of data.

Konnie and I reflected on how little direction there was for this potentially useful tool. We were also concerned about a possible backlash to the poorly represented application of this technology. If it became the norm to use GIS as a catchphrase to draw people to a paper with little GIS substance, those archaeologists with potential interest in the technology might leave with a bitter taste in their mouths. The lessons of the 1970s when statistics came to center stage in archaeology were fresh in our minds. During its heyday, statistics had been waved above archaeologists’ heads as an “answer” to dealing with a

multitude of archaeological problems. In the end, after much yelling and arm-waving, most agreed that statistics were not an answer in themselves but, like GIS, an extremely important tool available for archaeological use.

That afternoon, slumped on a bench, as we watched the number of brown corduroy jackets with patches on their sleeves dwindle, we decided that for the 1996 SAAs we would organize a focused GIS symposium. Together we hammered out the idea to bring together more than just a catch-all symposium on GIS. Our symposium would be tightly focused on a theme with broad implications in archaeological research for both the academic and commercial worlds: predictive modeling. From the outset we orchestrated this symposium to be directed towards generating a publication on GIS uses for predictive modeling. We both felt that the archaeological community would benefit from a book with a narrower and more clearly defined focus. We also believed that by presenting GIS as what it is, a useful and robust tool, and not an answer in and of itself, we could help steer GIS applications in archaeology on a more solid course. Nor did it hurt our motivation when we realized the meetings would be in New Orleans at the end of Bourbon Street.

The symposium at the New Orleans SAAs in 1996 was well received. Our plan not to become mired in introductions as to what GIS is, or the intricacies of each particular GIS software and hardware product, was appealing to both the participants and the audience. This focused structure allowed the presentations to concentrate on the application and caveats when using GIS for predictive modeling. These concepts, such as significance of specialized layers, data development concerns, and theoretical considerations, provided the audience with a solid grasp of the important issues archaeologists need to address and work with in predictive modeling. Because of this tight focus on the real and long-term implications of GIS predictive modeling, we feel that the chapters in this book represent a robust body of research that will be as useful to archaeology in twenty years as it is today. This volume addresses the important issues of understanding and applying GIS and predictive modeling to the landscape.

R.JOE BRANDON
Fayetteville, Arkansas

Acknowledgments

We would like to thank all of the contributors to this volume (Kathleen Allen, David Asch, Kristen Beckman, David Bennett, Galen Burgett, Tim Church, Luke Dalla Bona, Richard Duncan, James Ebert, Jon Hageman, Jim Kuiper, and Robert Warren) for their participation and cooperation in this project. They have been a pleasure to work with and are all outstanding researchers. Although we were unable to present the work of Patricia Hansell and Anthony Ranere here, we would like to thank them for their participation in the symposium and to wish them well with their interesting research in Panama. We would like to thank Fred Limp for his thought-provoking discussion at the 1996 SAAs. We would also like to thank Taylor & Francis for this opportunity to share some of our thoughts and research with the academic and scientific community.

Konnie L. Wescott
R. Joe Brandon

CHAPTER ONE

Introduction

KONNIE L.WESCOTT

Geographic information systems (GIS) offer archaeologists an exciting and powerful research tool destined to have as profound an effect on the field of archaeology as did the introduction of carbon dating in the 1950s. Archaeological data is spatial and temporal in nature, and therefore especially suited to the basic principles driving the development and use of GIS. Until recently, archaeologists had to cope with hand-drawn maps and cumbersome paper databases which were difficult to integrate and manipulate. However, widespread commercial development of GIS software and easy access to powerful desktop (and even laptop) computers have enabled archaeologists to view and manipulate their data in a medium that reflects its complex origins without being prohibitively complex to use.

GIS is proving itself to be a powerful and efficient managerial tool for spatial data sets, allowing the land or resource manager the ability to access, analyze, and interpret large amounts of archaeological data in a fraction of the time previously required. When archaeological data sets are combined with ecological, hydrological, geological, and other data, an even more impressive land management planning tool is created. Less likely to be constrained by resource management and planning issues, academic archaeologists are using GIS to develop innovative approaches for analyzing data or, in most cases, to apply traditional methods to large data sets previously considered too complex and time-consuming to tackle. The ability to integrate multiple layers of information simultaneously is also providing research archaeologists with a new means for interpreting prehistoric and historic landscapes. GIS is emerging as a fundamental component of archaeological method, and is likely to have an increasing impact on archaeological theory.

Several books have now been published on GIS and archaeology (Allen *et al.* 1990; Gaffney and Stan i 1991; Lock and Stan i 1995; Maschner 1996) or related topics (Reilly and Rhatz 1992; Aldenderfer and Maschner 1996). These books provide a great deal of background information on GIS technology and concepts and demonstrate the wide range of possible archaeological or anthropological applications. As interest in GIS continues to expand and as members of the archaeological community become familiar with the basic

principles of GIS, the opportunities to publish more specialized material increase as well. With this book we chose to focus on a specific, yet widespread, practical application of GIS for archaeology, predictive modeling.

Predictive modeling is not a new endeavor in archaeology, and the topic in and of itself has generated a great deal of methodological and theoretical controversy over the years. However, predictive modeling continues to generate a great deal of interest, especially now with the computer tools that have been made available (Carr 1985; Kohler and Parker 1986; Judge and Sebastian 1988; Kvamme 1992). It is our intent with this book to provide the reader with a “toolkit” with which to approach the complexities of predictive modeling using a GIS. Each chapter in this book offers one or more concept(s) as part of the toolkit.

The first several chapters in the book focus on models that have been developed. These chapters provide the toolkit with methodological information and a number of lessons learned about what worked and what did not and what potential pitfalls to watch out for. Robert Warren and David Asch (Chapter 2) lead the series of modeling chapters with a brief introduction to predictive modeling and a comprehensive description of their model’s methods and results. The thoroughness with which the authors have written this chapter makes it a must-read for future modelers and a key component of the modeling toolkit. Cross-validation testing of their logistic regression model’s predicted site locations indicates that the predictions are approximately 73% accurate, an improvement of up to 51% over a random or chance classification.

Richard Duncan and Kristen Beckman (Chapter 3) address the formulation of a GIS modeling process by which the archaeological sensitivity of several areas can be determined despite differing geographic locations. Although the model is similar to, that employed by Warren and Asch, its use within a cultural resource management context adds another dimension to the model. Duncan and Beckman discuss the “disturbance factor” which takes site preservation into account beyond site presence or absence. The authors also introduce the interesting variable of solar insolation into their GIS model.

Konnie Wescott and James Kuiper (Chapter 4) also work within a cultural resource management framework that extends beyond locating sites to determining where sites are most likely to be adversely affected by a given action. The authors use a GIS predictive model which focuses on frequencies of unique combinations of variables in an area largely unsurveyed (and to some extent unsurveyable without incurring some additional risk and cost). This method relies more heavily on available data within a larger geographical area, but less on nonsite locations than the logistic regression models used by Warren and Asch and Duncan and Beckman. Wescott and Kuiper also address the need to adjust the model based on site type due to differences in environmental conditions favorable to particular site types. The authors describe the benefits of using GIS beyond just modeling site locations, including the benefits of combining GIS and modeling in the greater context of resource planning and

management. To this end, the potential for linking environmental impact models to the results is also mentioned.

The last of the modeling chapters is by Luke Dalla Bona ([Chapter 5](#)). Dalla Bona continues with the resource planning theme, as his chapter discusses the use of GIS predictive modeling for forest management plans in Ontario, Canada. The chapter provides a detailed look at the research and development of the archaeological models that will serve as a resource management tool in all new forest management plans for an area encompassing 45 million hectares. The need to communicate effectively using these models is stressed. Once a model is generated it must be interpreted and presented in a way that the land-use planners or others, most of whom are nonarchaeologists, can quickly understand. The planners want to know how the archaeological potential might affect proposed activities within the forest and what can be done about it. Clear explanation of model results and the explicit identification of management options is the means by which cultural resources can be successfully protected.

The chapters by Kathleen Allen ([Chapter 6](#)) and Jon Hageman and David Bennett ([Chapter 7](#)) focus on two serious considerations to be taken into account when developing a predictive model using GIS: issues of scale and choosing the proper digital elevation model, or DEM.

Allen's chapter ([Chapter 6](#)) uniquely identifies and elaborates on the issue of scale, whether it be global, regional, or local, with respect to modeling settlement patterns. In addition to reviewing a number of unpublished GIS studies on the Iroquois, Allen compares global patterning with regional and local patterning using an example from the Lake Cayuga watershed in central New York. She discusses an additional important consideration, the need to pay careful attention to the spatial resolution of the data so that it coordinates with the chosen scale of analysis.

Digital elevation models (DEMs) provide an important data set for use with GIS in creating an archaeological predictive model as they affect critical variables such as elevation, slope, and aspect. However, these models are not useful if the data inaccurately reflects the landscape. Hageman and Bennett ([Chapter 7](#)) discuss the implications of using the wrong DEM and how these pitfalls can be avoided. Four DEM interpolation methods are described and a case study is used to illustrate how one might go about choosing the proper DEM for a given terrain.

The last two chapters of the book switch gears yet again and discuss some of the theoretical implications of GIS and predictive modeling, or, more accurately, the implications of the current perceived lack of theory. Both chapters suggest quite strongly that there is a tremendous need to use a critical eye when evaluating other models and when developing one's own approach.

James Ebert ([Chapter 8](#)) critically discusses the application of GIS for archaeological predictive modeling, focusing primarily on what, in his opinion, are common mistakes modelers tend to make. His unique perspective is likely to elicit

lively discourse regarding both commercial and academic archaeological applications of predictive modeling and GIS in the future.

Tim Church, R.Joe Brandon, and Galen Burgett (Chapter 9) discuss how the way archaeologists look at and interpret the past is being revolutionized and freed from many of the constraints of analytical methods which trace their origins to half a century ago. They discuss the importance for archaeologists, while at this crossroads, to take a minute to critically examine their approach to using GIS for analysis. Many modern uses of GIS, while innovative, in these authors' opinion are merely allowing archaeologists to automate modeling activities that were once done by hand. In this chapter, the authors argue that archaeologists need to take a long, hard look at what goals they ultimately want to achieve. If the goal is simply to automate modeling and perform statistical analyses of probable site locales, then that goal is very close to being achieved. However, if archaeologists want to take this technology to the next level of analysis and, in the fashion of classic scientific methods, begin to generate testable hypotheses which will serve as the foundation for a robust body of theory, then they must delve into other disciplines, such as landscape ecology, to derive a more holistic view of the landscape. By using GIS as a means to generate and test hypotheses, managers and modelers alike will have a new and more powerful tool at their disposal to create even more accurate models of prehistoric land use.

In conclusion, the body of literature on GIS and predictive modeling is growing. Our intent with this book is to provide the reader with a basic toolkit of things to consider when embarking on a GIS predictive modeling project. This volume is certainly not, nor is it intended to be, all-encompassing. Hopefully, it is sufficient to start the many wheels out there turning. A substantial dialogue, innovative and critical thinking, and a great deal of experimentation are what it will take to realize the true potential of GIS.

References

- ALDENDERFER, M. and MASCHNER, H.D.G. (eds), 1996, *Anthropology, Space, and Geographic Information Systems*, Oxford: Oxford University Press.
- ALLEN, K.M.S., GREEN, S.W. and ZUBROW, E.B.W. (eds), 1990, *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis.
- CARR, C. (ed.), 1985, *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*, Kansas City: Westport Publishers.
- GAFFNEY, V., and STAN I, Z., 1991, *GIS Approaches to Regional Analysis: A Case Study of the Island of Hvar*, Ljubljana: Znanstveni Inštitut Filozofske Fakultete.
- JUDGE, W.J., and SEBASTIAN, L. (eds), 1988, *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, Denver, CO: US Department of the Interior, Bureau of Land Management Service Center.

- KOHLER, T.A. and PARKER, S.C., 1986, Predictive models for archaeological resource location. In M.B.SCHIFFER (ed.) *Advances in Archaeological Method and Theory*, Vol. 9, New York: Academic Press, pp. 397–452.
- KVAMME, K.L., 1992, A predictive site location model on the High Plains: an example with an independent test. *Plains Anthropologist* **37**:19–40.
- LOCK, G. and STAN I, Z. (eds), 1995, *Archaeology and Geographical Information Systems*, London: Taylor & Francis.
- MASCHNER, H.D.G. (ed.), 1996, *New Methods, Old Problems: Geographic Information Systems in Modern Archaeological Research*, Carbondale: Center for Archaeological Investigations, Southern Illinois University.
- REILLY, P. and RAHTZ, S. (ed.), 1992, *Archaeology and the Information Age: A Global Perspective*, London: Routledge.

CHAPTER TWO

A Predictive Model of Archaeological Site Location in the Eastern Prairie Peninsula

ROBERT E. WARREN AND DAVID L. ASCH

A geographic information system (GIS) was used to create a high-resolution predictive model of prehistoric archaeological site location in a poorly drained upland prairie region of central Illinois. The model is based on a logistic regression analysis of sample data using qualitative and quantitative measures of the natural environment as independent variables. Cross-validation testing indicates that the model's predictions are about 73% accurate and represent a gain of up to 51% over a random or chance classification. Sites are most probable along the margins of wooded stream valleys and on the crests of well-drained knolls in the upland prairies. In contrast, site probabilities are low across extensive tracts of flat to gently undulating prairies. The modeled distribution of settlement appears to reflect complex prehistoric strategies of resource use, but it also could have been affected by geomorphic processes of landscape evolution. The model's predictive capabilities may be a useful tool for modern land managers and development planners in the area.

2.1

INTRODUCTION

The Prairie Peninsula was a mosaic of tall-grass prairie and deciduous forest that existed historically in the midwestern United States (Transeau 1935). In central Illinois, the Prairie Peninsula environment consisted of extensive prairies on flat upland landforms and narrow strips of woodland along stream valleys (Anderson 1970). Aquatic and forest resources were clustered along the valleys, while grassland resources were more widely dispersed.

The prairie-forest mosaic presented interesting challenges and opportunities for human settlement. During the 1700s and early 1800s Native American and Euro-American farmers commonly settled along the ecotone between prairie and forest. Here they had ready access to water, timber, pasture, and tillable soil (Faragher 1986; Klippel 1976). The open prairies were less suitable for

settlement; they lacked wood for fuel and building material, they were susceptible to fire, and prairie sod was difficult to cultivate before John Deere's invention of the steel plow. Also, many prairies were poorly drained and became seasonal marshes in spring and early summer. Wet prairies were valued for their great flocks of migratory waterfowl, but they were not fit for cultivation until the advent of artificial drainage systems in the late 1800s (Winsor 1975).

Evidently, most prehistoric settlement in the eastern Prairie Peninsula focused on the forested river valleys. The largest and most complex settlements were located in and along major valleys, and site densities appear to be highest there as well (Bareis and Porter 1984; Brown 1981). We know far less about prehistoric settlement in the upland prairies, despite the fact that the prairie biome historically covered more than 60% of Illinois landforms (Iverson *et al.* 1989). If we are to understand patterns of human land use throughout the state, we must study the uplands as well as the valleys. Recent archaeological surveys in central Illinois have shown that upland resources did attract settlement throughout the Holocene (Ferguson and Warren 1991; Klippel and Maddox 1977; Warren 1995). However, patterns of upland land use appear to vary from region to region, and much remains to be learned.

In this chapter we examine prehistoric site distributions in an upland prairie area of central Illinois (Figure 2.1). We describe a formal predictive model of site location developed for the area using a geographic information system (GIS) and logistic regression analysis. The model is based on archaeological data from a systematic survey and environmental data obtained from maps.

2.2

PREDICTIVE MODELING

Predictive models are tools for projecting known patterns or relationships into unknown times or places. Such models are potentially useful in archaeology. Archaeologists have documented only a fraction of the millions of sites in the New World, while thousands of sites are destroyed each year to make way for ongoing land development. One way to help us understand and protect these sites is to create formal models capable of predicting where they are located.

Predictive modeling emerged only recently as an important component of archaeological research (Carr 1985; Kohler 1988; Kohler and Parker 1986). An underlying key to the success of these models is the fact that archaeological sites tend to recur in environmental settings favorable to human settlement. Predictive models take advantage of such redundancies; they exploit contrasts between the environmental characteristics of places where sites do and do not occur. With appropriate data it is possible to make predictions from a relatively small sample of known locations to a much broader area.

Most archaeological predictive models rest on two fundamental assumptions: first, that the settlement choices made by ancient peoples were strongly influenced or conditioned by characteristics of the natural environment; second,

that the environmental factors that directly influenced these choices are portrayed, at least indirectly, in modern maps of environmental variation across an area of interest. Given these assumptions, it is possible to develop an empirical predictive model for any particular area, as long as the area has been adequately sampled by archaeological surveys. Several criteria can be used to judge the adequacy of surveys, the most important of which is that they consistently distinguish between locations where sites are present and locations where sites are absent (sites versus nonsites, respectively).

The distinction between sites and nonsites is essential, as it provides a framework within which probabilities can be calculated (Kvamme 1983). This may be done by computing a statistical classification model that capitalizes on the measurable environmental differences between the two groups. Such models make it possible to predict the probability that a site occurs at a given location simply by measuring an appropriate set of environmental variables. A successful predictive model is one that minimizes classification errors (site versus nonsite) to such an extent that it offers a substantial gain in accuracy over null models arising from chance alone.

The practical benefits of predictive models stem from the fact that they can be applied to extensive unsurveyed tracts of land where the actual locations of sites and nonsites are not known. Predicted distributions are useful in a variety of ways. First, they provide archaeologists not only with images of the patterns of prehistoric settlement in an area, but also with evidence of the most important environmental determinants of site location. Second, they provide land managers with expected distributions of the resources they are charged with protecting. Conversely, they also provide development planners with preliminary guides to the places where cultural resources are least likely to be affected by future construction projects.

Inductive or empirical predictive models are formal devices of pattern recognition (Kvamme 1990; Warren 1990a). Most such models use statistical methods to extract from a sample of observations a formal decision rule, a rule that can be used to predict the composition or characteristics of future samples. One of the most powerful and widely used of the empirical methods is a set of procedures called probability models (Aldrich and Nelson 1984). These models are well suited for predicting the locations of archaeological sites, as they are designed to predict the responses of either-or situations (site presence versus site absence) to the interactions of independent variables (environmental measurements). The predictions themselves are expressed in terms of probabilities. Probabilities are readily interpretable and easily testable values that range between 0 (low probability) and 1 (high probability).

As noted by Carr (1985), archaeologists interested in predicting site location are abandoning the traditional methods of settlement-subsistence research in favor of procedures that are more appropriate for prediction (Kvamme 1983, 1985, 1988, 1989, 1990, 1992; Limp and Carr 1985; Parker 1985; Scholtz 1981; Warren 1990a, b). One consequence of this reorientation is a new focus on land

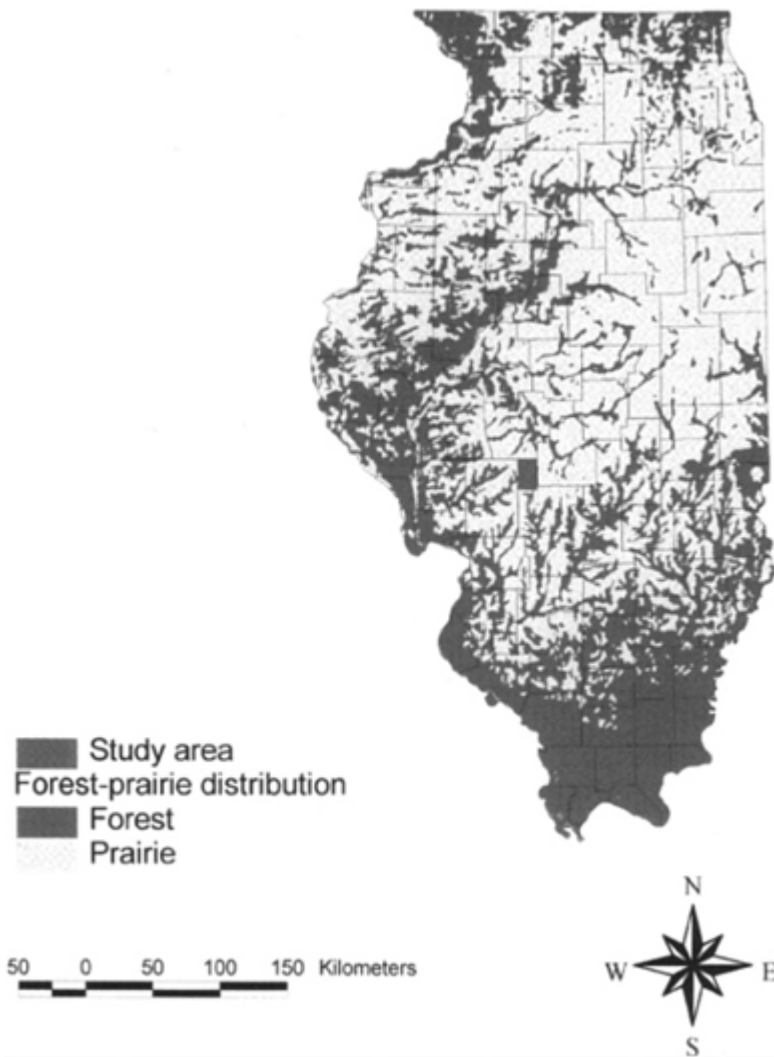


Figure 2.1 Location of the Montgomery study area In west-central Illinois.

parcels, rather than sites, as the basic unit of analysis. Other changes include a more widespread use of probability models, such as logistic regression analysis, and a healthy expansion of the environmental factors used as independent variables. Another unavoidable outgrowth of these developments is an increased reliance on computers—not just for analysis, but also for the collection of raw data and for the automated creation and measurement of variables. Computer-based geographic information systems (GIS) are needed to handle the vast amounts of data required for predictive models (Kvamme 1989; Kvamme and Kohler 1988).

One of the most powerful and flexible statistical techniques for predictive modeling is logistic regression analysis (Stopher and Meyburg 1979; Neter *et al.* 1983; Aldrich and Nelson 1984). In archaeological applications, logistic regression creates a prediction formula that uses independent environmental variables of virtually any scale to predict the probability that a site occurs on any given parcel of land. The formula defines an S-shaped probability curve of group membership that is oriented along an axis of intergroup discrimination (Figure 2.2). The axis comprises an interaction of environmental variables that best discriminates site locations from nonsite locations.

A logistic regression model can be tested for accuracy by predicting the group memberships of the locations used to develop the model (training sample). However, this approach yields overly optimistic results, as training-sample locations are not independent of the model (Kvamme 1988; Gong 1986). It is more realistic to run tests using locations that are truly independent of the training sample, locations that were either unknown at the time of model development or were randomly withheld from the modeling process (testing sample). In either case, accuracy is readily measured by calculating the percentages of correct and incorrect predictions along the probability scale of group membership (Kvamme 1988; Warren 1990a).

2.3

MATERIALS AND METHODS

This chapter describes a predictive model of archaeological site location in the eastern Prairie Peninsula. The project was supported in part by the Illinois Department of Energy and Natural Resources, which provided grant funds for developing predictive models of archaeological and paleontological resources in Illinois (Oliver *et al.* 1987; Warren *et al.* 1987).

2.3.1

Environmental setting

The study area is located in the northern panhandle of Montgomery County, Illinois (Figure 2.1). The panhandle occupies an interstream divide between the headwaters of Macoupin Creek, which drains to the west, and tributaries of the Sangamon and Kaskaskia rivers, which drain to the north and south, respectively.

During the Middle Pleistocene the study area was planed off by a series of Illinoian glaciers, the last of which retreated from the area by about 132,000 years ago (Johnson 1986). Surficial deposits consist of glacial drift overlain by about 1.5–2.5 m (5–7.5 ft) of loess. Since the Illinoian glaciation, the landscape has become submaturely dissected by its drainage networks. Upland landforms are relatively flat and stream valleys are incised less than 10m (30ft) below the elevation of the surrounding uplands. Elevations across the study area as a whole

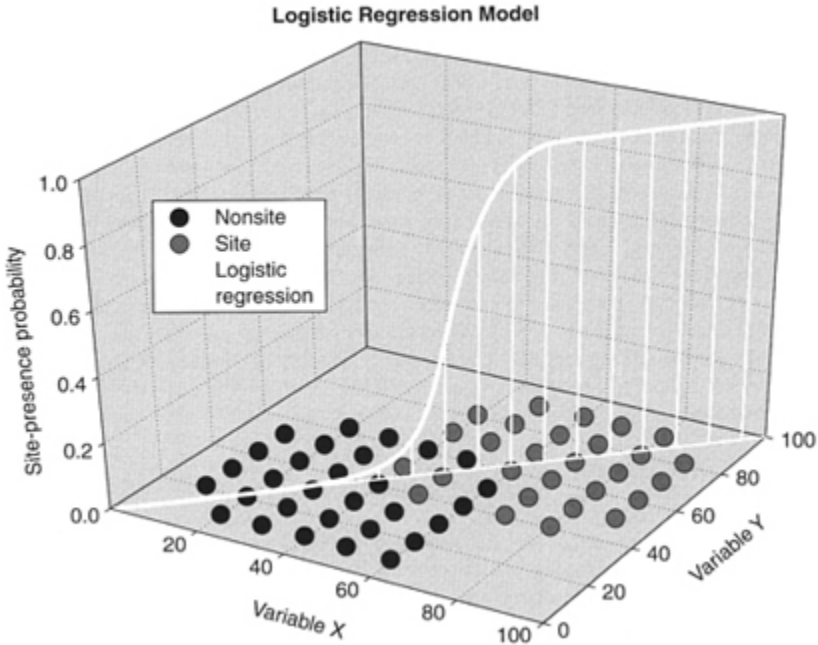


Figure 2.2 Idealized logistic regression of two groups of objects (sites and nonsites) across two independent variables (X and Y) (after Warren 1990a). The line running lengthwise through the horizontal scatter of points is the axis that best discriminates sites from nonsites. The vertical plane is defined by an S-shaped logistic regression line. This line shows an increase in site-presence probability from left to right along the axis of discrimination. [LogRegModel.jnb]

range from about 180 to 209 m (590 to 685 ft) above mean sea level. Although the upland surface is relatively featureless, it is crossed by a series of broad, gently sloping ridges and swales that tend to parallel one another and are oriented from northeast to southwest. Topographic relief of the ridges and swales is generally about 3 m (10 ft). The ridge-swale structure is evident on 5-ft interval contour maps (1:24,000 USGS quadrangles) and also on county soil-survey maps (Downey and Odell 1969). Ridge soils (e.g., Harrison silt loam, Herrick silt loam) are moderately well drained to somewhat poorly drained, whereas swale soils (e.g., Virden silty clay loam) are poorly drained and seasonally experience high-water-table elevations which are at or near the ground surface (0–0.3 m depth).

The native vegetation of the study area was almost entirely prairie; the only forests observed by General Land Office (GLO) surveyors in 1818 and 1819 were located in the western part of the area along Macoupin Creek and its tributaries. Paleoenvironmental records indicate that central Illinois was forested with cool temperate deciduous forests during the terminal Pleistocene and then warm temperate deciduous forests during the early Holocene (King 1981). Prairies evidently did not appear in the area until about 8000 years BP (King

1981; Webb *et al.* 1983) or perhaps later (Baker *et al.* 1992; Chumbley *et al.* 1990; Wright 1992).

2.3.2 Archaeological survey

The Montgomery County panhandle was selected for archaeological survey because of its suitability for developing predictive models of archaeological site location (Asch 1978). The area has a homogeneous environment with “very little relief, relatively little post-Pleistocene geomorphic change, no large deeply incised streams, virtually no forest, and a high percentage of cropland favorable to surface reconnaissance” (Asch 1978:6).

The study area comprises part or all of six congressional townships (T10–12N, R4–5W, 3rd PM). It extends about 23 km (14 mi) north-south and 14 km (9 mi) east-west, and has a surface area of about 322 km² (126 mi²). This area was systematically sample-surveyed from 1974 to 1977 in a project directed by David L. Asch, then of the Northwestern University Archeological Program (Asch 1975, 1978; Asch *et al.* 1981). The Asch survey was a stratified probabilistic sample that covered 13.7 km² of land, or about 4.3% of the study area.

Sampling units were widely dispersed (Figure 2.3). They include a series of square 40-acre (16-ha) quadrats selected at random from the entire study area (21 quadrats, 1.0% of study area), as well as a series of stratified random samples that often had irregular borders (3.3% of study area). The stratified samples focused survey activities on four specific environmental zones, including (1) square 40-acre quadrats containing areas of moderately well-drained prairie soil (14 quadrats); (2) irregular areas of moderately well-drained prairie soil (12 tracts); (3) linear segments of alluvium-floored valleys along streams (28 tracts); and (4) linear segments of headwater streams lacking well-defined valleys (8 tracts).

Most of the study area was farmed in the 1970s. The survey was restricted to cultivated fields with good ground-surface visibility, which, based on survey data, covered about 95.5% of the panhandle. Surveyors followed traverses spaced at 15-m (50-ft) intervals and inspected the ground for artifacts and other traces of prehistoric occupation. Traverse intervals were reduced to about 2.3 m (7.5 ft) upon the discovery of artifacts. Surface artifacts were flagged before they were collected to aid in the mapping of site boundaries on aerial photographs. Sites were defined as concentrations of three or more prehistoric surface artifacts. Surveyors also mapped find spots yielding isolated cultural debris (1–2 artifacts), but these locations were excluded from the present analysis.

The survey recorded 59 prehistoric archaeological sites (including 89 distinct artifact concentrations) ranging in age from terminal Pleistocene to late Holocene, or during the past 12 000 years (Asch 1978). Five sites covering a total area of about 1.7 ha (4.2 acres) were discovered in the 21 randomly selected survey quadrats. Surveyors covered a total land area of 324.8 ha (802.5 acres) in

the random quadrats, so prehistoric sites in the study area cover only about 0.53% of the ground surface. On the basis of these data, the *a priori* probability that a site occurs at any randomly selected location is $p=0.0053$.

2.3.3

GIS data and predictive modeling

Development of the Montgomery predictive model was basically a two-step process (Figure 2.4). The first step was to create digital map coverages of the area using a GIS. The second step was to create and test a formal predictive model of site location based on data contained in the digital coverages.

Geographic information systems are integrated systems of computer hardware, software, and peripheral equipment that can be used to create, process, and display spatial data (Burrough 1986; Kvamme 1989). We used a GIS running ARC/INFO software to generate the data needed to create the Montgomery predictive model (Environmental Systems Research Institute 1986).

The first step in this process was to transfer information from original maps to computer storage using a coordinate digitizer (Figure 2.4a). The digitized information included contour lines and stream courses from 7.5' USGS topographic quadrangles (1:24,000), soil types from county soil surveys (1:15,840; Downey and Odell 1969), and native vegetation from GLO plat maps (~1:45,000). We also digitized the archaeological survey data, including the locations of sites and surveyed areas. For convenience, we use the term *nonsite* to refer to surveyed areas lacking evidence of prehistoric settlement. We then edited and gridded the map images to create a series of *primary coverages* of elevation, streams, soils, vegetation, and archaeology. Each of the primary coverages was then transformed to create a suite of derivative *secondary coverages*, including topographic relief, distance to stream, soil drainage, distance to prairie-forest ecotone, and the locations of sites and nonsites in surveyed areas.

The basic unit of analysis in the Montgomery model was the grid cell. The grid consisted of a regular lattice of square cells, each measuring 50 m on a side and representing a land area of 0.25 ha. The survey region as a whole contained 1.3 million cells. The survey sample contained 5,473 cells, including 265 site cells and 5,208 nonsite cells. Each grid cell in the Montgomery data base was associated with dozens of numerical codes, including a sequential cell label, locational coordinates, a utility variable, and codes for each of the secondary environmental coverages.

The secondary environmental coverages included 24 independent variables derived from topographic maps, soil maps, and vegetation maps (Table 2.1). A majority of the variables were ratio scale, although several of the soil and vegetation coverages were nominal, ordinal, or interval scale (see Blalock 1979; Warren 1990a).

The topographic variables were derived from a digital elevation model (DEM) of surface landforms using a series of GIS transformations. The DEM model, in

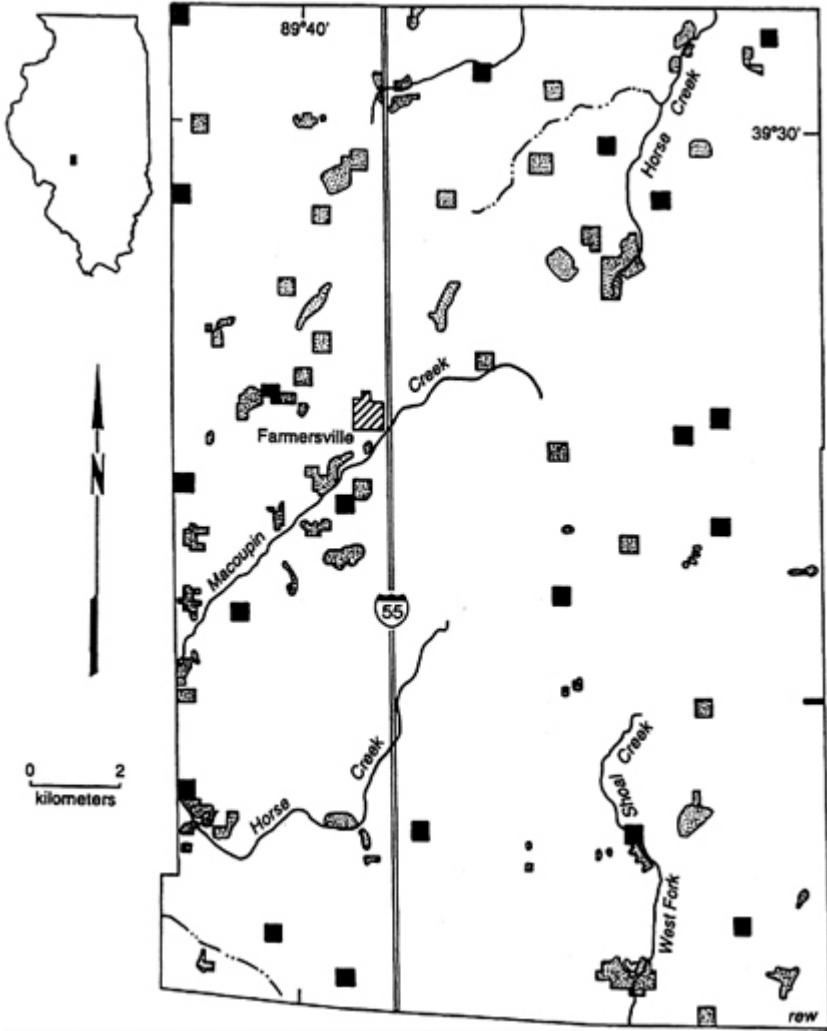
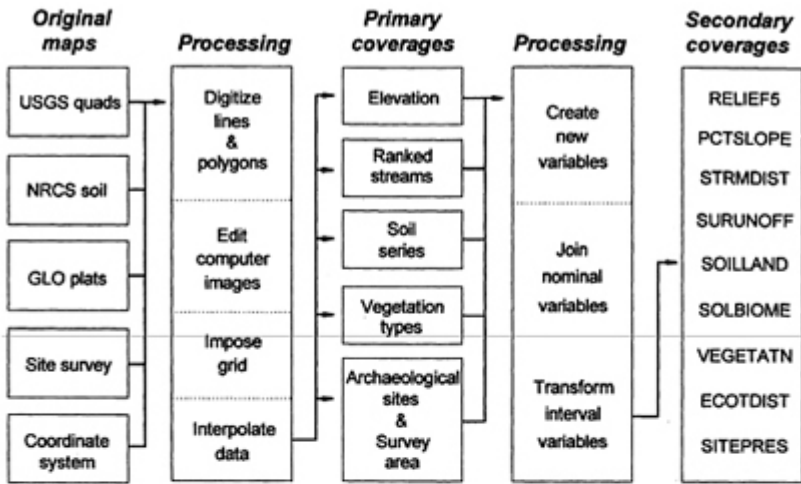


Figure 2.3 Map of the Montgomery study area, showing the distribution of archaeologically surveyed land areas.

turn, was derived from a triangulated irregular network (TIN) interpolation of elevation contours (Environmental Systems Research Institute 1986). The secondary topographic coverages included six relief variables, a measure of surface slope, and a measure of surface aspect (Table 2.1). The relief variables measured various elevation ranges—including total relief, above-site relief, and below-site relief—within 100-m and 500-m radius catchments of each grid cell. For example, a plot of total relief within 500-m catchments (Figure 2.5a) indicates that relief is highest in the darker areas along creeks and lowest in the lighter areas of glaciated uplands.

a. Creation of GIS Data Sets



b. Predictive Modeling

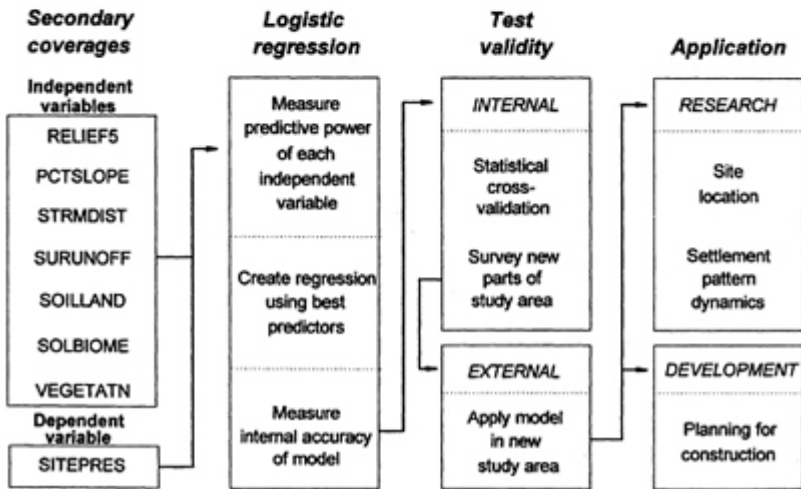


Figure 2.4 Generalized flow charts of the procedures used (a) to create computer files of geographical data using a geographic information system (GIS), and (b) to create, test, and apply probability-based predictive models of archaeological site location from GIS data sets (after Warren 1990a).

Stream courses were buffered to create two measures of stream distance: distance to nearest stream and distance to nearest permanent stream. For the

purpose of this analysis, permanent streams were defined as watercourses mapped on 7.5' United States Geological Survey quadrangles with rankings of 3 based on the Strahler (1957) method of stream rank-ordering.

Soil coverages in the study area plot the distributions of 43 distinct mapping units, including 17 soil series and their various surface-slope and erosional manifestations. Soils were classified and recombined across such properties as permeability, drainage, flood frequency, and landform type. For example, a plot of surface runoff (Figure 2.5b) indicates that runoff is most rapid in the darker areas along stream valleys and on sloping knolls scattered across the uplands. The light shading depicts areas of slow to ponded runoff, which are common in upland swales.

Univariate statistical tests were used to compare the environmental differences, if any, between site and nonsite locations for each of the variables listed in Table 2.1. Ratio- and interval-scale variables were tested with the Mann-Whitney rank-sum statistic; ordinal- and nominal-scale variables were tested with chi-square or Fisher's exact test (see Blalock 1979).

In developing the Montgomery predictive model, we used secondary environmental coverages as independent variables and site presence as the dependent variable (Figure 2.4b). The model itself was created by applying logistic regression analysis to a random subsample of site and nonsite locations. We used simple random sampling to select the nonsite cells and random cluster sampling to select the site cells.¹ The logistic regression program we used (BMD-PLR) is a stepwise procedure that measures the predictive power of each independent variable and calculates regressions using only

Table 2.1 Map sources and environmental variables used to describe sample locations (sites and nonsites) in the Montgomery study area.

Map source/ variable	Code	Scale	Measurement interval	Statistical test ^a
<i>Topographic maps</i>				
Total relief (dm) in 100-m catchment	RELIEF 1	Ratio	1 dm	Mann**
Total relief (dm) in 500-m catchment	RELIEF5	Ratio	1 dm	Mann**
Above-site relief (dm) in 100-m catchment	RELFABO1	Ratio	1 dm	Mann**
Above-site relief (dm) in 500-m catchment	RELFABO5	Ratio	1 dm	Mann**

Map source/ variable	Code	Scale	Measurement interval	Statistical test ⁴
Below-site relief (dm) in 100-m catchment	RELFBEL1	Ratio	1 dm	Mann**
Below-site relief (dm) in 500-m catchment	RELFBEL5	Ratio	1 dm	Mann
Surface slope (% grade)	PCTSLOPE	Ratio	1%	Mann**
Surface aspect (deviation from northerly aspect)	ASPECT	Ratio	1°	Mann
Distance to nearest stream (m)	STRMDIST	Ratio	50m	Mann**
Distance to nearest permanent stream (m; stream rank >3)	PERMDIST	Ratio	50m	Mann**
<i>County soil maps</i>				
Biome of soil formation	SOLBIOME	Nominal	—	Chi**
Soil landform	SOILLAND	Nominal	—	Chi**
Soil parent material	PARENTMA	Nominal	—	Chi**
Soil moisture regime	SOILMOIS	Nominal	—	Chi**
Soil drainage	DRAINAGE	Ordinal	—	Chi**
Soil permeability	PERMEABL	Ordinal	—	Chi**
Soil surface runoff	SURUNOFF	Ordinal	—	Chi**
Soil flood frequency	FLOODFRQ	Ordinal	—	Chi**
Soil erodibility (K factor)	ERODIBIL	Interval	100 K	Mann**
Soil productivity (basic management; adjusted for	SOILPROD	Interval	1 unit	Mann**

Map source/ variable	Code	Scale	Measurement interval	Statistical test ^a
slope and erosion)				
Minimum depth to seasonal high water table (cm)	WATERTAB	Ratio	1cm	Mann**
Distance to closed- depression soil (m)	DEPRESDI	Ratio	50m	Mann
<i>General Land Office plat maps</i>				
Native vegetation biome	VEGETATN	Nominal	—	Fisher*
Distance to prairie-timber ecotone (m)	ECOTDIST	Ratio	50m	Mann**

^a Statistical tests reflect differences between the environmental characteristics of sites (265 cells) and nonsites (5208 cells) in a survey sample of the Montgomery study area. Mann is the Mann-Whitney rank-sum test; Chi is the chi-square test; Fisher is Fisher's exact test (see Blalock 1979). Asterisks mark the 21 variables for which site and nonsite distributions are significantly different (* $p < 0.05$; ** $p < 0.001$).

the strongest combination of predictors (Engelman 1985). We tested the validity of the resulting model by measuring the accuracy of its predictions using both a training sample and a testing sample. The *training sample* consists of sample data that went into creating the model; the *testing sample* is an independent subset of sample data that was withheld from model development.

2.4 RESULTS

Univariate statistical tests indicate that site and nonsite locations in the Montgomery study area tend to occur in different environmental settings. The differences between sites and nonsites are statistically significant for all but three of the 24 independent variables created for the analysis (Table 2.1). For example, frequency distributions of site and nonsite grid cells differ significantly from one another on the soil-drainage variable (DRAINAGE); sites tend to occur on well-drained soils, whereas nonsites tend to be poorly drained (Figure 2.6). Given the many environmental contrasts between sites and nonsites, the Montgomery data set should be suitable for predictive analysis.

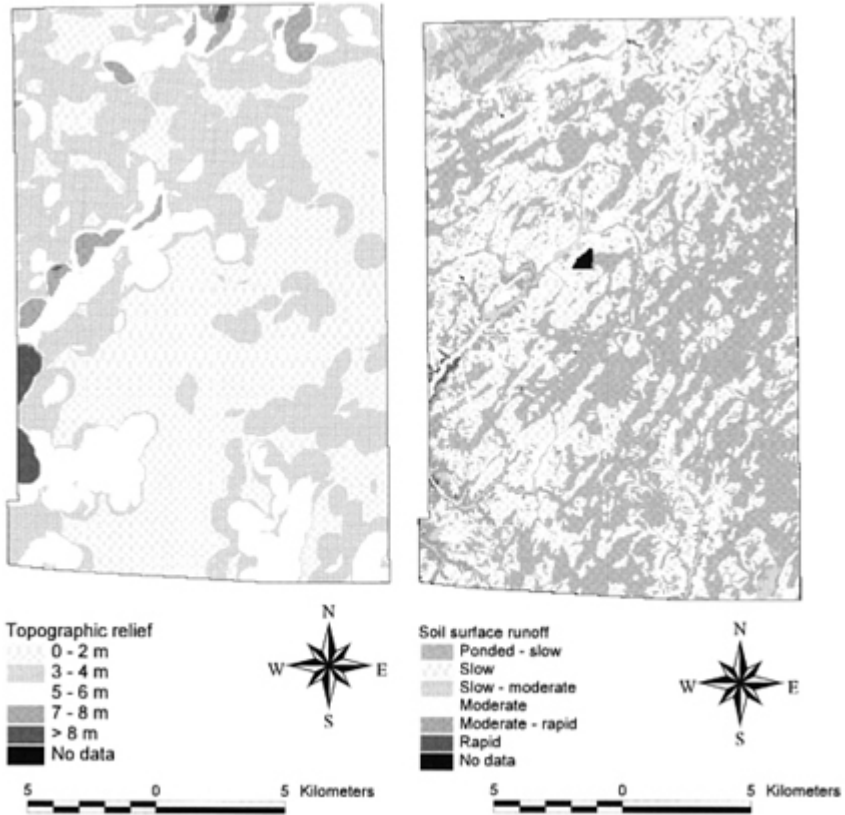
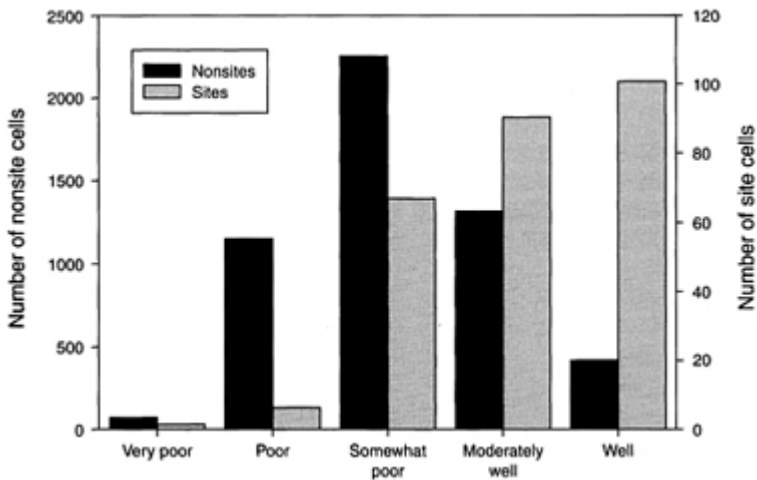


Figure 2.5 GIS coverages of the Montgomery study area, showing (a) topographic relief within 500-m radius catchments of each grid cell (RELIEF5), where relief ranges from 0 to 2 m (white shading) to >8 m (black shading); and (b) rate of soil surface runoff (SURUNOFF), where runoff ranges from ponded/slow (white shading) to rapid (black shading).

We used a training sample of 1,181 grid-cell locations to create the logistic regression model (Table 2.2). These cells included a random cluster sample of 162 site cells (representing 45 archaeological sites) and a random sample of 1,019 nonsite cells. The testing sample consisted of 4,292 cells, all of which were withheld from the process of model development.

The stepwise logistic regression program used F-to-enter scores to evaluate the predictive power of independent variables and select the most powerful combination of predictors (Figure 2.7). An assessment of F-to-enter scores at step 0 (i.e., prior to the inclusion of any variables in the model) showed that 21 of the 24 variables had significant predictive power ($p < 0.05$). The single most powerful variable at step 0 was topographic relief in 500-m catchments (RELIEF5), which provided greater



Natural soil drainage

Figure 2.6 Frequency distributions of site and nonsite grid cells on various categories of the soil drainage variable (DRAINAGE) in the Montgomery study area. The two distributions are significantly different ($p < 0.01$); sites tend to occur on well-drained soils, whereas nonsites tend to be poorly drained.

Table 2.2 Grid-cell composition of the training and testing samples used to develop and validate the Montgomery predictive model.

Sample	Site cells ^a	Nonsite cells	Total
Training sample	162	1019	1181
Testing sample	103	4189	4292
Total	265	5208	5473

^a Random cluster samples representing a total of 89 archaeological sites (45 sites comprise the 162 site cells in the training sample; 44 sites comprise the 103 site cells in the testing sample).

separation between sites and nonsites along the axis of discrimination than any other single variable (Figure 2.7). The program ultimately selected six independent variables for inclusion in the model, including two topographic variables (RELIEF5, PCTSLOPE), two soils variables (SURUNOFF, SOILLAND), and two hydrologic variables (STRMDIST, PERMDIST). Some significant variables were bypassed in the selection process because they correlated with more powerful predictors that had already entered the model.

The Montgomery predictive model consists of a logistic regression formula that may be separated into a probability component and a score component (Table 2.3). The score component consists of an intercept value and 15 regression coefficients. The formula can be used to calculate the probability that an archaeological site occurs at any given location in the Montgomery study area simply by measuring and classifying the environmental context of the location in

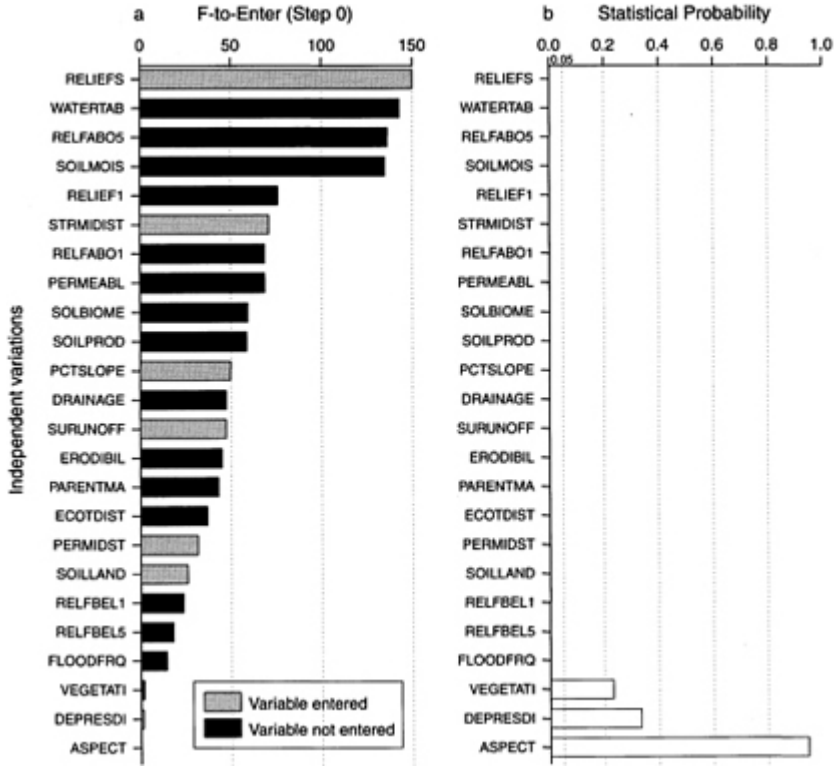


Figure 2.7 Comparison of the predictive power of all variables considered for inclusion in the Montgomery logistic regression analysis, (a) F-to-enter scores at step zero in the analysis, where scores increase with predictive power. Shaded bars denote variables that were entered and retained by the model, (b) Statistical significance of the F-to-enter scores in comparison with a significance level of $p=0.05$.

terms of the six variables comprising the model.² Regression coefficients for the four ratio-scale variables (RELIEF5, PCTSLOPE, STRMIDIST, PERMDIST) are simple multipliers that operate like standard regression coefficients (see Blalock 1979). However, regression coefficients for the categorical soil variables (SURUNOFF_{dl-d5}, SOILLAND_{dl-d6}) operate on design-variable codes (see Engelman 1985). Table 2.4 lists the design-variable coding sequence for each category. For example, the coefficients and codes for a location with rapid surface runoff would be as follows: SURUNOFF=-2.863(0)-1.96(0)-1.696(0)-0.67(0)+9.046(1)=9.046.

In the Montgomery predictive model, the axis of discrimination between sites and nonsites is a multivariate function of the six independent variables included in the logistic regression. Figure 2.8 compares the frequency distributions of sites and nonsites along the discriminant axis. The two distributions do overlap, indicating that the model does not provide a perfect separation between sites and

nonsites. However, there is a clear separation of modes; site cells tend to be most abundant at high site probabilities ($p[\text{site}] > 0.5$) and nonsite cells are more abundant at low site probababil

Table 2.3 Logistic regression formula predicting the site-presence probability of archaeological sites in the Montgomery study area, including the probability component (upper portion) and the score component (lower portion).

Probability component:

Score component:

SCORE=0.979	+0.020 (RELIEF5)	+0.176 (PCTSLOPE)	-0.196 (STRMDIST)
+0.012 (PERMDIST)	-2.863 (SURUNOFFd1)	-1.960 (SURUNOFFd2)	
-1.696 (SURUNOFFd3)	-0.670 (SURUNOFFd4)	+9.046 (SURUNOFFd5)	
+1.229 (SHILLANDd1)	-0.704 (SOILLANDd2)	-2.348 (SOILLANDd3)	
-0.147 (SOILLANDd4)	+1.380 (SOILLANDd5)	+0.719 (SOILLANDd6)	

Table 2.4 Design-variable codes for categorical soil variables in the Montgomery predictive model.

Variable	Category	Design variables					
		d3	d4	d5	d6		
SURUNOFF	Ponded to slow	-1	-1	-1	-1	-1	
	Slow	1	0	0	0	0	
	Slow to medium	0	1	0	0	0	
	Medium	0	0	1	0	0	
	Medium to rapid	0	0	0	1	0	
	Rapid	0	0	0	0	1	
LANDFORM	Upland flats	-1	-1	-1	-1	-1	-1
	Upland knolls	1	0	0	0	0	0
	Upland depressions	0	1	0	0	0	0
	Upland swales	0	0	1	0	0	0
	Valley slopes	0	0	0	1	0	0
	Small floodplains	0	0	0	0	1	0
	Large floodplains	0	0	0	0	0	1

ities ($p[\text{site}] < 0.5$). A log-likelihood chi-square value of $\chi^2 = 262.4$ ($df = 6; p < 0.001$) indicates there is less than one chance in a thousand that the observed separation of sites from nonsites could have arisen by chance (see Stopher and Meyburg 1979).

One can determine the accuracy of a predictive model by comparing the model's predictions with the actual archaeological characteristics of site and nonsite locations (Kvamme 1988). Accuracy can be measured at any given point along the gradient of predicted site probability ($0 \leq p[\text{site}] \leq 1$). Along this gradient one should expect to see a decline in the accuracy of site predictions and an increase in the accuracy of nonsite predictions (Figure 2.9). Measures of internal consistency indicate that the Montgomery model attains an optimum overall performance at a probability of 0.5 (Figure 2.9a). This is the point on the training-sample graph where the site and nonsite accuracy curves intersect; at this cut-point the model correctly classified 77.2% of sites and 77.4% of nonsites in the training sample. However, an independent cross-validation test based on testing-sample data indicates the model's true accuracy to be somewhat less (Figure 2.9b). Optimal results for the testing sample were obtained at a probability cut-point of about 0.5, where the model correctly classified 69.9% of sites and 73.3% of nonsites.

Another way to assess the performance of a predictive model is to measure its gain in accuracy over a random or null classification (Kvamme 1992). This can be done by assuming that the nonsites in a model are equivalent to the background environment or land area. This is a reasonable assumption in the Montgomery study area, where probabilistic sample data indicate that archaeological sites occupy only about 0.5% of the land surface. Performance curves plotted along the gradient of predicted site probability compare the percentages of sites and land area (=nonsites) incorporated in the model (Figure 2.10a, c) and the percentage gain in accuracy over a random or null classification (Figure 2.10b, d). With regard to the training sample (Figure 2.10a, b), the model incorporates 77% of the sites but only 30% of the land area at a site probability of 0.5 (Figure 2.10a). As shown in the adjacent graph (Figure 2.10b), this represents a rather substantial gain of 47% over a random classification. The testing sample also indicates good results (Figure 2.10c, d). There is a rather substantial separation between the site and land-area curves (Figure 2.10c) and a 51% gain over random at the 0.3 probability cut-point. The Montgomery predictive model clearly outperforms a chance classification.

One advantage of formal predictive modeling with a GIS is that models can be displayed in map form. Figure 2.11 plots the distribution of predicted site-presence probabilities across the entire Montgomery study area. It was created by applying the logistic regression formula (Table 2.3) to the independent environmental variables associated with each of the 1.3 million grid cells in the region. The map indicates that sites are most probable in two distinct and spatially limited environmental settings: (1) in and along the valleys of headwater streams, and (2) on isolated upland knolls scattered across the till plain. In contrast, site probabilities are low across extensive tracts of the upland surface.

Although maps of site probability are helpful for interpreting multivariate predictive models, a deeper understanding can be gained by assessing the

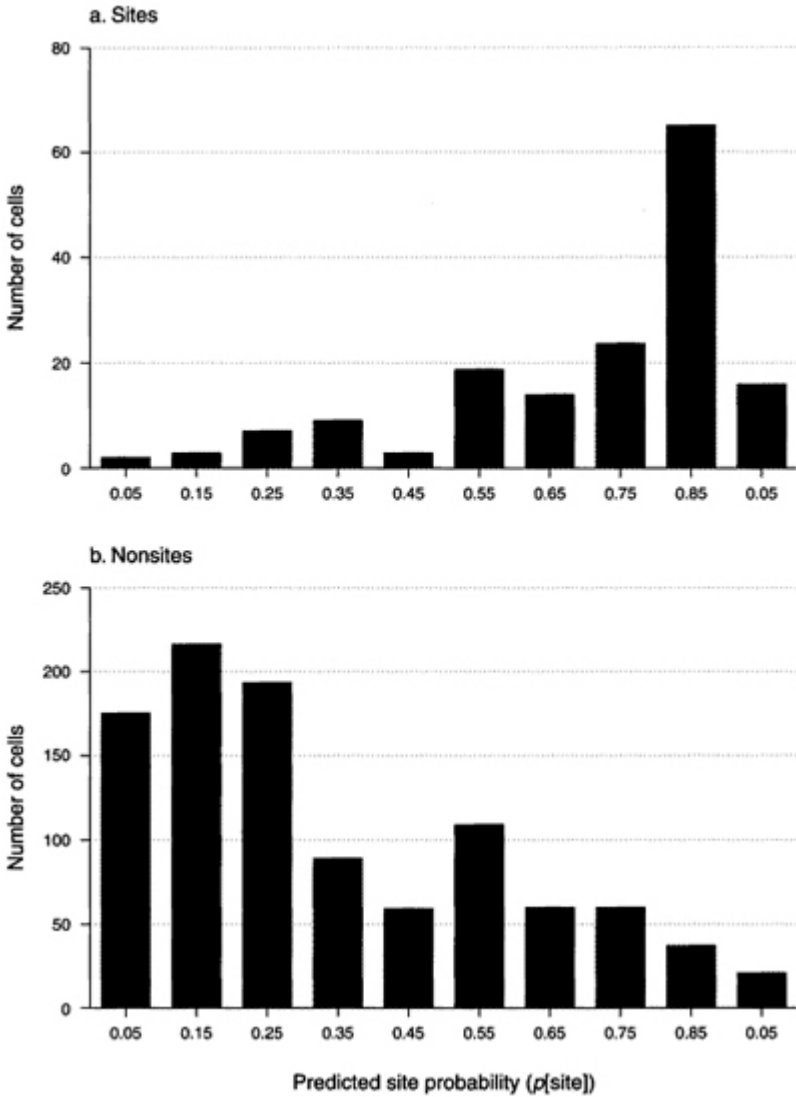


Figure 2.8 Frequency distributions of (a) site cells and (b) nonsite cells on the axis of discrimination ($p[\text{site}]$) of the Montgomery predictive model. The two distributions are significantly different ($p < 0.01$); site cells are relatively abundant at high site-presence probabilities, whereas nonsite cells are relatively abundant at low site-presence probabilities.

behavior of the particular independent variables incorporated in the model. To observe such trends, we calculated site probability curves for the observed ranges of specific variables while holding all other variables constant at their median or modal values. The four curves plotted in [Figure 2.12](#) illustrate the

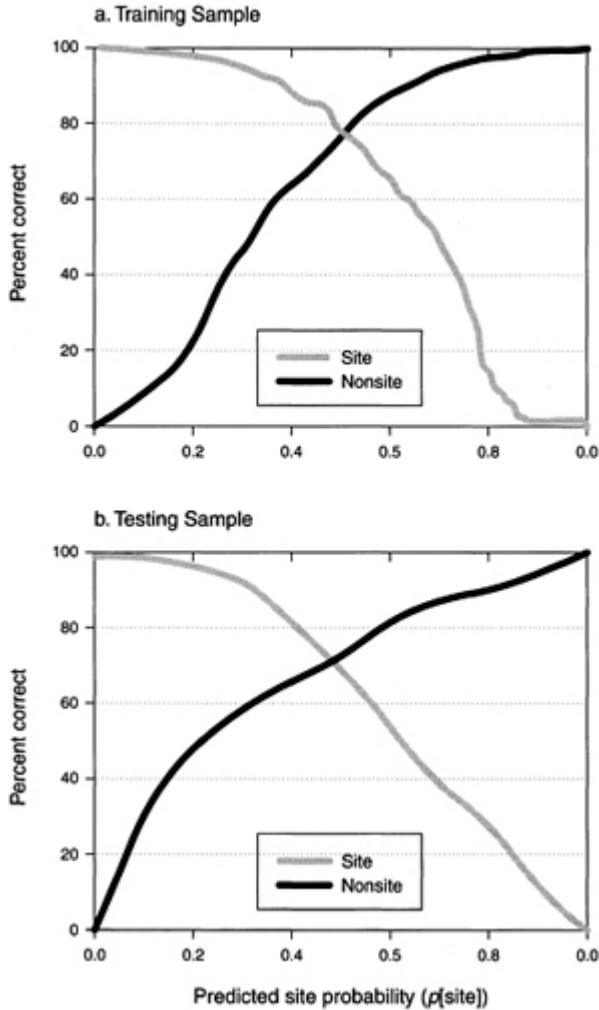


Figure 2.9 Accuracy of the Montgomery logistic regression model. The curves are percentages of correct predictions along a gradient of predicted site probability ($p[\text{site}]$) for (a) a training sample of 162 site cells and 1019 nonsite cells, and (b) a testing sample of 103 site cells and 4189 nonsite cells.

relationships between site probability and the four ratio-scale variables in the Montgomery model. First, site probability increases as one moves from flat landforms to more rugged landforms with greater topographic relief (RELIEF5; Figure 2.12a). Second, site probability increases as one moves closer to the nearest stream (STRMDIST; Figure 2.12b). Third, site probability increases as surface slope becomes progressively steeper (PCTSLOPE; Figure 2.12c). And fourth, site probability increases as one moves away from permanent streams (PERMDIST; Figure 2.12d).

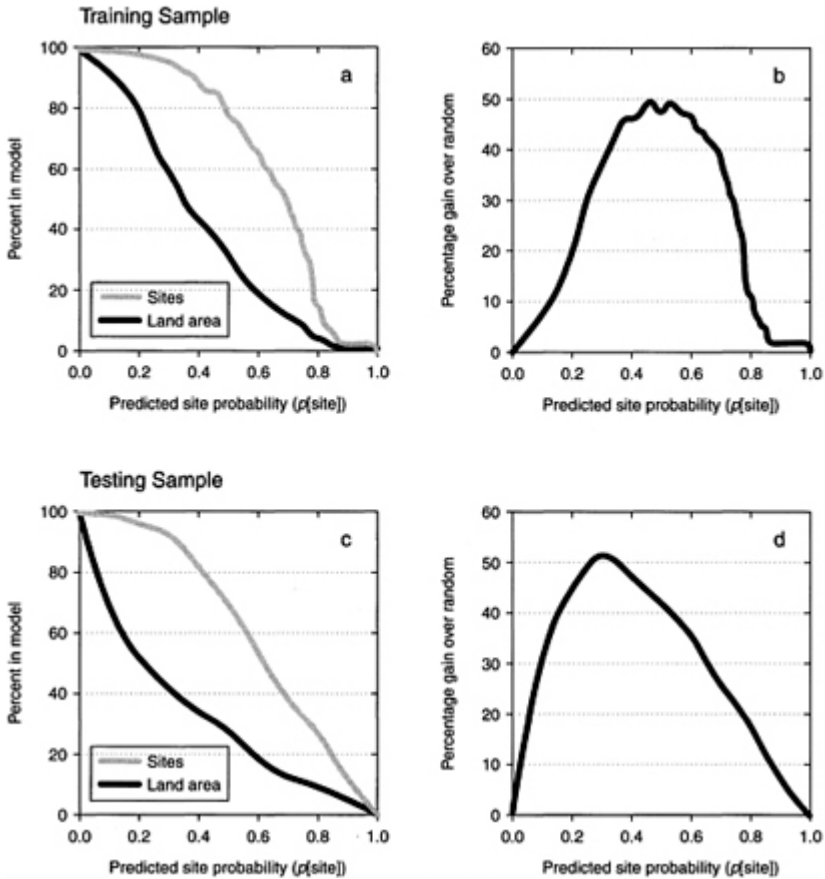


Figure 2.10 Performance curves (a, c) and percentage gain (b, d) of the Montgomery logistic regression model over a random or chance classification for (a, b) the training sample of 1181 grid cells and (c, d) a testing sample of 4292 grid cells.

Similarly, bar graphs can be used to compare site probabilities among the various categories of nominal- and ordinal-scale variables. [Figure 2.13](#) plots probabilities associated with the two soils variables in the Montgomery model. As indicated in the first graph, site probability tends to increase on soils with progressively more rapid surface runoff (SURUNOFF; [Figure 2.13a](#)).³ The only exception to this trend is the “Ponded to slow” category, which has a higher site probability than both the “Slow” and “Slow to medium” categories. The landform associations indicate that while site probabilities are lowest on upland swales, they are highest on upland knolls and on floodplains (SOILLAND; [Figure 2.13b](#)).⁴ Intermediate probabilities are evident for locations on upland flats, upland depressions, and valley slopes.

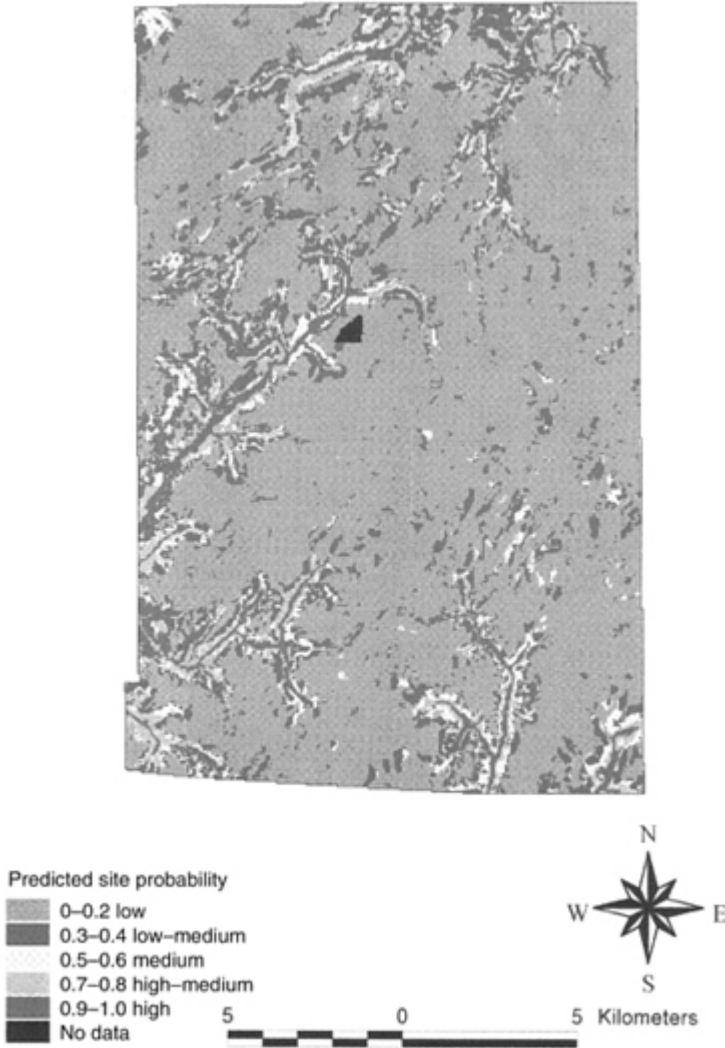


Figure 2.11 Map of predicted site-presence probability in the Montgomery study area.

2.5

DISCUSSION AND CONCLUSIONS

Predictive models developed with the aid of GIS can provide accurate probability estimates of prehistoric site location in sample-surveyed study areas. The Montgomery model joins a growing list of studies which demonstrate that fine-grained probability models can provide reliable predictions of where archaeological sites should—and should not—occur on a given landscape (Duncan and Beckman 1996; Kvamme 1983, 1985, 1988, 1992; Parker 1985;

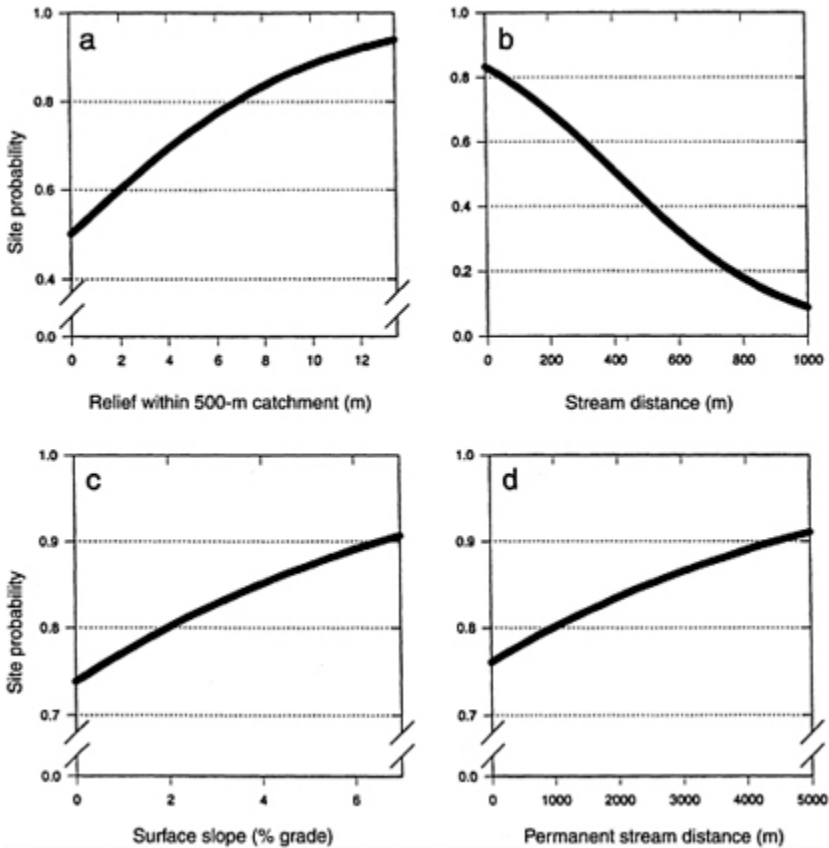


Figure 2.12 Variation in site probability for ratio-scale variables in the Montgomery predictive model, including (a) relief within 500-m catchments (RELIEF5; m); (b) distance to nearest stream (STRMDIST; m); (c) surface slope (PCTSLOPE; %grade); and (d) distance to nearest permanent stream (PERMDIST; m). The probabilities were calculated by holding constant at their median values all other variables in the model.

Warren 1990b). Cross-validation testing indicates that the model's predictions are correct about 70–73% of the time and represent a gain of up to 51% over a chance classification. The predictive power of the model should be of interest to any land managers or development planners in Montgomery County who wish to take into account the potential effects of their projects on cultural resources.

In the Montgomery study area, surficial traces of prehistoric human settlement are concentrated in two distinct environmental settings: (1) on sloping, well-drained valley landforms on or near the floodplains of headwater streams, and (2) on the crests of well-drained upland knolls scattered across the otherwise flat till plain. In general, sites are most probable in areas with relatively rugged relief, where surface slopes are steep and runoff is rapid. The valley locations are distinguished by their proximity to streams and their common occurrence on

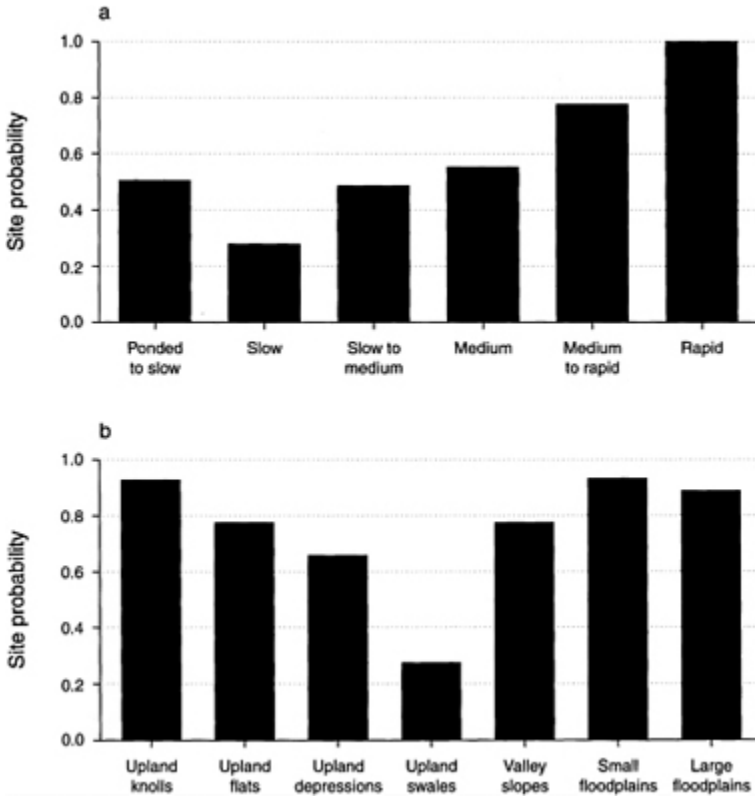


Figure 2.13 Variation in site probability for nominal and ordinal variables in the Montgomery predictive model, including (a) soil surface runoff (SURUNOFF); and (b) soil landform (SOILLAND). The probabilities were calculated by holding constant at their median values all other variables in the model.

floodplain soils. The knoll locations are defined by their occurrence on characteristic hilltop soils and their greater distance from permanent streams.

The complex distribution of prehistoric settlement in the Montgomery study area could have been shaped by cultural as well as natural processes operating over the past 12,000 years. Three hypotheses seem plausible.

1 *Bimodal settlement pattern.* The complex distribution of settlement may reflect a bimodal settlement pattern in which prehistoric hunter-gatherers focused their activities on both valleys and upland knolls. If so, it would appear that prehistoric land-use strategies were geared toward two sets of resources: (1) aquatic-riparian resources that were concentrated along upland stream courses, and (2) prairie or forest resources that were broadly dispersed across the glaciated uplands. The different modes of settlement could have had a seasonal dimension in which valley resources were exploited at one time of year and upland resources at another. Assuming that strategic differences are reflected in

the composition or utilization of toolkits, this hypothesis could be tested by comparing the artifact assemblages and occupational structures of valley sites with those of knoll sites. Consistent with the structural implication is Asch's (1978) observation that knoll sites are generally larger than valley sites and have more abundant artifacts. However, this difference could be related to the fact that knolls represent highly restricted zones of habitable living space in the poorly drained uplands of Montgomery County and were the focus of recurrent occupations.

2 Composite settlement patterns. Rather than a bimodal settlement pattern, the complex distribution of settlement in the Montgomery area could represent a composite of two different unimodal settlement patterns of different ages. Changes in upland settlement patterns have been documented in other parts of central Illinois, and these changes appear to have been related to the rather significant environmental changes that occurred in the eastern Prairie Peninsula during the past 12000 years (Ferguson and Warren 1991; Klippel and Maddox 1977; Warren 1995). Settlement in the Montgomery area could have shifted through time among valley and knoll locations in response to environmental change. However, a preliminary analysis of diagnostic artifacts from the different landforms suggests there were no major qualitative shifts in settlement location during the Holocene. For example, artifacts dating to the Early Archaic (~8,000–10,000 years BP) and Middle Archaic (~4500–8000 years BP) cultural periods were recovered from both valley and knoll locations in the Montgomery area (see Asch 1978). Hence, it does not appear that the complex Montgomery site distribution is a composite of unimodal settlement patterns.

3 Landscape evolution. Geoarchaeological research in the Midwest has shown that geomorphic processes of landscape evolution can affect the preservation and surface visibility of archaeological sites (Hajic 1990a, b; Van Nest 1993; Wiant *et al.* 1983). While some sites are destroyed or deflated by sediment erosion, others may be buried and obscured by sediment deposition. To the extent that erosion and deposition have destroyed or obscured archaeological sites in a given study area, site surveys designed to detect surface exposures of artifacts may lead to the development of incomplete and distorted models of settlement distribution. In the uplands of western Illinois between the Illinois and Mississippi rivers, Van Nest (1993) has shown that some prehistoric occupations are buried and preserved beneath the modern plow zone, which is generally about 20–30 cm thick. Several sites are buried in Holocene alluvium along ephemeral stream channels beneath thin veneers (0.5–1 m) of historic alluvium. Another, the Penstone site, is a discrete Late Archaic occupation dated to 3700 years BP that is partially buried in Holocene colluvium within a subtle upland swale. Van Nest (1993) suggests the Penstone occupation may have been buried by “developmental upbuilding” of the surface soil through the long-term assimilation of colluvium via the processes of soil creep or frost creep. If similar burial processes occurred in the Montgomery study area, some undetected sites may exist in the alluvium of stream valleys and in the colluvium of upland

swales. Although more than 95% of the area was plowed during the Asch survey, some occupations may have been located beneath the reach of the plow and were undetected by surveyors. However, modeled site probabilities are relatively high on floodplain soils despite the potential obfuscation effect of historic alluvium. It seems unlikely that buried prehistoric occupations would be common in upland swales, a poorly-drained landform where modern buildings are rarely constructed (Asch 1978). Nevertheless, it should be borne in mind that the Montgomery predictive model is based on surface-exposed sites and may not adequately portray the locations of any buried occupations that may exist in the area.

Although it is difficult to weigh the relative importance of the various factors that may have effected the complex distribution of Montgomery settlement, it appears that the bimodal-settlement-pattern hypothesis is most plausible. Prehistoric hunter-gatherers exploited the region throughout much of the Holocene and apparently located their occupations strategically to take advantage of various upland resources and to maximize the habitability of their living surfaces in a region with poor surface drainage. Geomorphic processes of landscape evolution may have affected the distribution of surface-exposed sites, but the importance of developmental upbuilding and other burial processes remains to be tested in this area.

In closing, it is worth noting that geographic information systems are indispensable for creating, testing, and interpreting complex predictive models of human settlement. The Montgomery data set consists of more than 36 million data values associated with 1.3 million grid-cell locations. Without the aid of a GIS, we would have required the services of a small army to help us create, test, and display the Montgomery predictive model.

Acknowledgments

The Montgomery County archaeological survey was directed by D.L.Asch and was supported by a grant from the Illinois Department of Conservation to the Northwestern University Archeological Program. The computer modeling project was directed by R.E.Warren and was supported by a grant from the Illinois Department of Energy and Natural Resources to the Illinois State Museum Society. We thank Sheryl G. Oliver and Ray E.Druhot for creating digital map coverages and data sets in the Illinois State Museum GIS lab. Drs Bonnie W.Styles and Michael D.Wiant provided support and encouragement. Dr Wayne M.Wendland arranged access to computing facilities at the University of Illinois, Urbana-Champaign. Jacqueline A.Ferguson assisted with the computer analysis. We also thank Kenneth L.Kvamme for helpful discussions of predictive modeling.

Notes

- 1 We selected site cells using random cluster sampling, where each cluster consisted of the one or more cells comprising an individual site. Cluster sampling was used to ensure that the sample of site cells used to create the predictive model (training sample) was truly independent of the sample of site cells used to test it (testing sample). We avoided simple random sampling of site cells because it would have introduced problems with spatial autocorrelation (Kvamme 1988). Spatial autocorrelation occurs when the values of adjacent cells are highly correlated with one another, which is clearly the case in the Montgomery study area in terms of the distributions of archaeological sites and their environmental characteristics. In our study, simple random sampling would have assigned some cells of multi-cell sites to the training sample and others to the testing sample, thereby violating the logical assumption that a model is independent of test data.
- 2 The y -intercept constant ($\beta_0 = 0.979$) in the score component of the logistic regression formula (Table 2.3) is an unbiased value that was adjusted to equalize the weight of site and nonsite cells in the model using the following equation: $\beta_0 = \beta_0 + \ln(n_2/n_1)$, where β_0 is the unbiased constant, β_1 is the biased constant ($\beta_1 = -0.860$), n_1 is number of site cells in the training sample ($n_1 = 162$), and n_2 is number of nonsite cells in the training sample ($n_2 = 1019$) (see Stopher and Meyburg 1979).
 For the purposes of field-testing the Montgomery predictive model, the unbiased y -intercept constant (β_0) should be readjusted to account for the *a priori* probability of encountering a site on a given parcel of land. In the 21 randomly selected quadrats in the Montgomery study area, 1.71 ha (4.23 acres) of site area were discovered in a total surveyed area of 324.8 ha (802.5 acres). The estimated prior probability of site occurrence is $p = (\text{site area}/\text{surveyed area}) = 0.0053$. The unbiased y -intercept constant (β_0) can be adjusted to account for the prior probability of site occurrence using the following equation: $\beta_0 = \beta_0 - \ln(N_2/N_1)$, where β_0 is the adjusted constant ($\beta_0 = -4.269$), β_1 is the unbiased constant ($\beta_1 = 0.979$), N_1 is site area in the 21 randomly selected quadrats ($N_1 = 1.71$ ha), and N_2 is nonsite area in the 21 randomly selected quadrats ($N_2 = 323.0$ ha). The adjusted constant (β_0) yields lower site probabilities than the unbiased constant (β_0); it reflects the rarity of prehistoric sites in the Montgomery study area and provides a more realistic estimate of site probabilities in the field.
- 3 Soil series and mapping units associated with various categories of the soil-surface runoff variable (SURUNOFF) are as follows: (1) *ponded to slow*: Cowden, Cowden-Piasa, Ebbert, Herrick-Piasa, Sable (=Harvel), Virden; (2) *slow*: Herrick, Lawson; (3) *slow to medium*: Clarksdale, Ipava, Oconee, Oconee-Tamalco, Radford; (4) *medium*: Harrison, Stoy, Tamalco; (5) *medium to rapid*: Blair, Downs (=Sicily), Velma; (6) *rapid*: Hickory.
- 4 Soil series and mapping units associated with various categories of the soil-landform variable (LANDFORM) are as follows: (1) *upland knolls*: Harrison, Oconee (0–2% slope), Oconee-Tamalco, Tamalco; (2) *upland flats*: Clarksdale, Cowden, Cowden-Piasa, Herrick, Herrick-Piasa, Ipava; (3) *upland depressions*: Ebbert, Sable (=Harvel); (4) *upland swales*: Virden; (5) *valley slopes*: Blair, Harrison, Hickory, Oconee (2–7% slope), Oconee-Tamalco, Downs (=Sicily), Stoy, Velma; (6) *small floodplains*: Radford; (7) *large floodplains*: Lawson. For

the purpose of this study, small floodplains are defined as the floodplain soils of low-order streams (Strahler ranks 1–3) and large floodplains are defined as the flood-plain soils of medium-order streams (Strahler rank 4).

References

- ALDRICH, J.H., and NELSON, F.D., 1984, *Linear Probability, Logit, and Probit Models*, Sage University Papers on Quantitative Applications in the Social Sciences No. 07–045, Beverly Hills: Sage.
- ANDERSON, R.C., 1970, Prairies in the prairie state, *Transactions of the Illinois State Academy of Science* **63**:214–221.
- ASCH, D.L., 1975, An archaeological survey of the upland prairie zone of westcentral Illinois. In *Preliminary Report of 1974 Historic Sites Survey Archaeological Reconnaissance of Selected Areas in the State of Illinois: Part I, Summary, Section A*, Urbana: Illinois Archaeological Survey, pp. 65–71.
- 1978, *Predictive Modeling of Archeological Site Location in an Illinois Upland Prairie Region: A Systematic Survey of the Montgomery County Panhandle*, Report submitted to Illinois Department of Conservation, Springfield, Illinois, Northwestern University Archeological Program, Kampsville, Illinois.
- ASCH, D.L., FARNSWORTH, K.B., UDESEN, H.C. and KOSKI, A.L., 1981, Upper Mississippi River and lower Illinois River units (III-South and VI). In M.K.Brown (ed.) *Predictive Models in Illinois Archaeology: Report Summaries*, Division of Historic Sites, Illinois Department of Conservation, Springfield, pp. 55–72.
- BAKER, R.G., MAKER, L.J., CHUMBLEY, C.A., and VAN ZANT, K.L., 1992, Patterns of Holocene environmental change in the midwestern United States, *Quaternary Research* **37**:379–389.
- BAREIS, C.J. and PORTER, J.W. (eds), 1984, *American Bottom Archaeology*, Urbana: University of Illinois Press.
- BLALOCK, H.M., Jr, 1979, *Social Statistics*, 2nd edition, New York: McGraw-Hill.
- BROWN, M.K. (ed.), 1981, *Predictive Models in Illinois Archaeology: Report Summaries*, Springfield: Division of Historic Sites, Illinois Department of Conservation.
- BURROUGH, P.A., 1986, *Principles of Geographical Information Systems for Land Resources Assessment*, Monographs on Soil and Resources Survey No. 12, New York: Oxford University Press.
- CARR, C., 1985, Introductory remarks on regional analysis, in C.Carr (ed.) *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*, Kansas City: Westport Publishers, pp. 114–127.
- CHUMBLEY, C.A., BAKER, R.G. and BETTIS, E.A. III, 1990, Midwestern Holocene paleoenvironments revealed by floodplain deposits in northeastern Iowa, *Science* **249**:272–274.
- DOWNEY, C.E., and ODELL, R.T., 1969, *Soil Survey of Montgomery County, Illinois*, Illinois Agricultural Experiment Station Soil Report No. 86, Washington, DC: US Government Printing Office.
- DUNCAN, R.B. and BECKMAN, K.A., 1996, Finding the needle: sitting down with a GIS based predictive site model, Paper presented at the 61st Annual Meeting of the Society for American Archaeology, New Orleans, Louisiana.

- ENGELMAN, L., 1985, Stepwise logistic regression. In W.J. Dixon (ed.) *BMDP Statistical Software Manual*, Berkeley: University of California Press, pp. 330–344.
- ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE, 1986, *ARC/INFO Users Manual Version 3.2*, Redlands, CA: ESRI.
- FARAGHER, J.M., 1986, *Sugar Creek: Life on the Illinois Prairie*, New Haven: Yale University Press.
- FERGUSON, J.A. and WARREN, R.E., 1991, *Prehistoric Archaeology of the Central Illinois Prairie: The Decatur-Pana Section of the FAP-412 South Highway Corridor*, Technical Report No. 91–590–9, Springfield: Quaternary Studies Program, Illinois State Museum.
- GONG, G., 1986, Cross-validation, the jackknife, and the bootstrap: excess error estimation in forward logistic regression, *Journal of the American Statistical Association* **81**: 108–113.
- HAJIC, E.R., 1990a, Late Pleistocene and Holocene landscape evolution, depositional subsystems, and stratigraphy in the lower Illinois River valley and adjacent central Mississippi River valley, unpublished PhD dissertation, Department of Geology, University of Illinois, Urbana.
- 1990b, *Koster Site Archeology I: Stratigraphy and Landscape Evolution*, Research Series Vol. 8, Kampsville, IL: Center for American Archeology.
- IVERSON, L.R., OLIVER, R.L., TUCKER, D.P., RISSER, P.G., BURNETT, C.D. and RAYBURN, R.G., 1989, *The Forest Resources of Illinois: An Atlas and Analysis of Spatial and Temporal Trends*, Special Publication No. 11, Champaign: Illinois Natural History Survey.
- JOHNSON, W.H., 1986, Stratigraphy and correlation of the glacial deposits of the Lake Michigan lobe prior to 14 ka BP, *Quaternary Science Reviews* **5**:17–22.
- KING, J.E., 1981, Late Quaternary vegetational history of Illinois, *Ecological Monographs* **5**: 43–62.
- KLIPPEL, W.E., 1976, Euro Americans and the Kickapoo in the Illinois prairie, *The Explorer* **18**:9–13.
- KLIPPEL, W.E. and MADDOX, J., 1977, The Early Archaic of Willow Branch, *Mid-Continental Journal of Archaeology* **2**:99–130.
- KOHLER, T.A., 1988, Predictive locational modeling: history and current practice. In W.J. Judge and L. Sebastian (eds) *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, US Department of the Interior, Bureau of Land Management Service Center, Denver, CO, pp. 19–59.
- KOHLER, T.A. and PARKER, S.C., 1986, Predictive models for archaeological resource location. In M.B. Schiffer (ed.) *Advances in Archaeological Method and Theory*, Vol. 9, New York: Academic Press, pp. 397–452.
- KVAMME, K.L., 1983, Computer processing techniques for regional modeling of archaeological site locations, *Advances in Computer Archaeology* **1**:26–52.
- 1985, Determining empirical relationships between the natural environment and prehistoric site locations: a hunter gatherer example. In C. Carr (ed.) *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*, Kansas City: Westport Publishers, pp. 208–238.
- 1988, Development and testing of quantitative models. In W.J. Judge and L. Sebastian (eds) *Quantifying the Present and Predicting the Past: Theory, Method,*

- and Application of Archaeological Predictive Modeling*, US Department of the Interior, Bureau of Land Management Service Center, Denver, CO , pp. 325–428.
- 1989, Geographic information systems in regional archaeological research and data management. In M.B.Schiffer (ed.) *Archaeological Method and Theory*, Vol. 1, Tucson: University of Arizona Press, pp. 139–203.
- 1990, The fundamental principles and practice of predictive archaeological modeling. In A.Voorrips (ed.) *Mathematics and Information Science in Archaeology: A Flexible Framework*, Studies in Modern Archaeology, Vol. 3, Bonn: Holos-Verlag, pp. 257–295.
- 1992, A predictive site location model on the high plains: an example with an independent test, *Plains Anthropologist* **37**:19–40.
- KVAMME, K.L. and KOHLER, T.A., 1988, Geographic information systems: technical aids for data collection, analysis, and display. In W.J.Judge and L.Sebastian (eds) *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, Denver, CO: US Department of the Interior, Bureau of Land Management Service Center, pp. 493–547.
- LIMP, W.F. and CARR, C., 1985, The analysis of decision making: alternative applications in archaeology. In C.Carr (ed.), *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*, Kansas City: Westport Publishers, pp. 128–172.
- NETER, J., WASSERMAN, W. and KUTNER, M.H., 1983, *Applied Linear Regression Models*, Homewood: R.D.Irwin.
- OLIVER, J.S., STYEES, T.R. and GRAHAM, R.W., 1987, *A Predictive Model for the Location of Paleobiological Sites in the Lower Illinois River Valley: A Tool for the Siting of Public Facilities and Construction Planning*, Springfield: Quaternary Studies Center, Illinois State Museum.
- PARKER, S., 1985, Predictive modeling of site settlement systems using multivariate logistics. In C.Carr (ed.) *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*, Kansas City: Westport Publishers, pp. 173–207.
- SCHOLTZ, S.C., 1981, Location choice models in Sparta. In R.Lafferty, J.Otinger, S.Scholtz, W.F.Limp, B.Watkins, and R.Jones (eds) *Settlement Predictions in Sparta*, Research Series No. 14, Arkansas Archeological Survey, University of Arkansas, Fayetteville, pp. 207–222.
- STOPHER, P.R. and MEYBURG, A.H., 1979, *Survey Sampling and Multivariate Analysis for Social Scientists and Engineers*, Lexington: Lexington Books.
- STRAHLER, A.N., 1957, Quantitative analysis of watershed geomorphology, *American Geophysical Union, Transactions* **38**:913–920.
- TRANSEAU, E.N., 1935, The Prairie Peninsula, *Ecology* **16**:423–437.
- VAN NEST, J., 1993, Geoarchaeology of dissected loess uplands in western Illinois, *Geoarchaeology* **8**:281–311.
- WARREN, R.E., 1990a, Predictive modeling in archaeology: a primer. In K.M.S.Allen, S.W. Green, and E.B.W.Zubrow (eds) *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis, pp. 90–111.
- 1990b, Predictive modeling of archaeological site location: a case study in the Midwest. In K.M.S.Allen, S.W.Green, and E.B.W.Zubrow (eds) *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis, pp. 201–215.

- 1995, Wetland settlement and environmental change in the eastern Prairie Peninsula, Illinois, Paper presented at a conference entitled “Hidden Dimensions: The Cultural Significance of Wetland Archaeology,” University of British Columbia, Vancouver, British Columbia, Canada.
- WARREN, R.E., OLIVER, S.G., FERGUSON, J.A. and DRUHOT, R.E., 1987, *Predictive Modeling of Archaeological Site Location: A Case Study in Southern Illinois*, Technical Report No. 87-210-9, Quaternary Studies Program, Illinois State Museum, Springfield.
- WEBB T., III., GUSHING, E.J. and WRIGHT, H.E., Jr, 1983, Holocene changes in the vegetation of the Midwest. In H.E. Wright, Jr (ed.) *Late-Quaternary Environments of the United States*, Vol. 2, *The Holocene*, Minneapolis: University of Minnesota Press, pp. 142-165.
- WIANT, M.D., HAJIC, E.R. and STYEES, T.R., 1983, Napoleon Hollow and Koster site stratigraphy: implications for Holocene landscape evolution and studies of Archaic period settlement patterns in the lower Illinois River valley. In J.L. Philips and J.A. Brown (eds) *Archaic Hunters and Gatherers in the American Midwest*, New York: Academic Press, pp. 147-164.
- WINSOR, R.A., 1975, Artificial drainage of east central Illinois, 1820-1920, unpublished PhD dissertation, Department of Geography, University of Illinois, Urbana-Champaign.
- WRIGHT, H.E., Jr, 1992, Patterns of Holocene climatic change in the midwestern United States, *Quaternary Research* **38**:129-134.

CHAPTER THREE

The Application of GIS Predictive Site Location Models within Pennsylvania and West Virginia

RICHARD B.DUNCAN AND KRISTEN A.BECKMAN

Skelly and Loy, Inc., of Monroeville, Pennsylvania has used a geographic information system (GIS) to formulate models for site potential or archaeological resource sensitivity within four areas of Pennsylvania and West Virginia. The models use easily available coded and digitized locational data for a variety of natural and cultural factors. The GIS models allow for the predictive evaluation of relative impacts of ground-disturbing activities in a flexible and cost-effective manner.

3.1 INTRODUCTION

The development and application of predictive models which assess the probability of prehistoric archaeological sites occurring across the landscape have greatly increased in recent years (Allen *et al.* 1990; Brandt *et al.* 1992; Carr 1985; Judge and Sebastian 1988; Kohler and Parker 1986; Kvamme 1983, 1986, 1990, 1992; Neumann 1992; Phillips and Duncan 1993). The driving force behind this growth in predictive model development has been the need for the identification, protection, and management of increasingly threatened cultural resources in a cost-effective and useful manner. The basis of such models is that the spatial distribution of cultural remains, which are often represented as archaeological sites, is the result of human decision-making activities within the possibilities and conditions presented by the environment. The development of most contemporary predictive models involves the consideration of multiple thematic layers of information relating to past environmental and/or cultural conditions. Interpreting the interplay between these multiple thematic layers and their various permutations may reveal identifiable patterns that reflect actual human behavioral patterns and choices (Kincaid 1988).

One major difficulty in effectively developing predictive site location models has been the synthesis of vast amounts of data pertaining to complex environmental and cultural factors at a sufficient level of geographic detail to

have cultural resource management and planning utility. Early efforts involved either intuitive and simplified methods or very labor-intensive manual computation systems. Within the past decade, innovations in computer hardware and software have led to the growing use of geographic information systems (GIS) within the realm of archaeological predictive modeling (Allen *et al.* 1990; Calamia 1986; Kvamme 1986, 1989, 1990, 1992; Savage 1989; Warren 1990). The application of GIS has brought a powerful geographic database tool to the modeling task, a tool that allows for greater quantity and complexity of data, more sophisticated quantitative methods, and an increased flexibility for site predictive modeling. The ability of GIS to rapidly manipulate vast amounts of disparate locational data from multiple map layers and to investigate potential relationships between these layers has made it possible to develop and implement detailed, complex, and yet effective predictive models. The purpose of this chapter is to describe the general process of developing and applying GIS predictive models within several studies located in Pennsylvania and West Virginia, and to briefly discuss some of the problems encountered within the process.

3.2 BACKGROUND

Since 1994, Skelly & Loy, Inc. has utilized a GIS—specifically ARC/INFO version 7.0 (ESRI 1992)—to formulate models for site potential or archaeological resource sensitivity within four study areas located in Pennsylvania and West Virginia. These four studies and the models developed for them are currently in various stages of completion. This chapter presents the overall results of these studies as examples of the application of GIS in archaeological predictive modeling. Although each project area was geographically different and the intended use of the final product varied from a basic-level sensitivity map to a more detailed quantitative model, our goal was to develop a process by which the GIS could be used with easily obtained and consistent data sets to produce a reliable but flexible model which could be applied to almost any geographic location. This goal has met with considerable success, although several areas of concern have been identified.

The locations of the four study areas are shown in [Figure 3.1](#): the Monongahela River Valley in southwestern Pennsylvania, the Central Susquehanna River Valley in north-central Pennsylvania, the Tygart Valley River in northern West Virginia, and the Kanawha River Valley near the Ohio border in western West Virginia. The discussion of the process that developed the GIS models will focus greater attention on the Monongahela River Valley study, as the three remaining studies and models are similar in structure. The current status and results of each study will be presented.

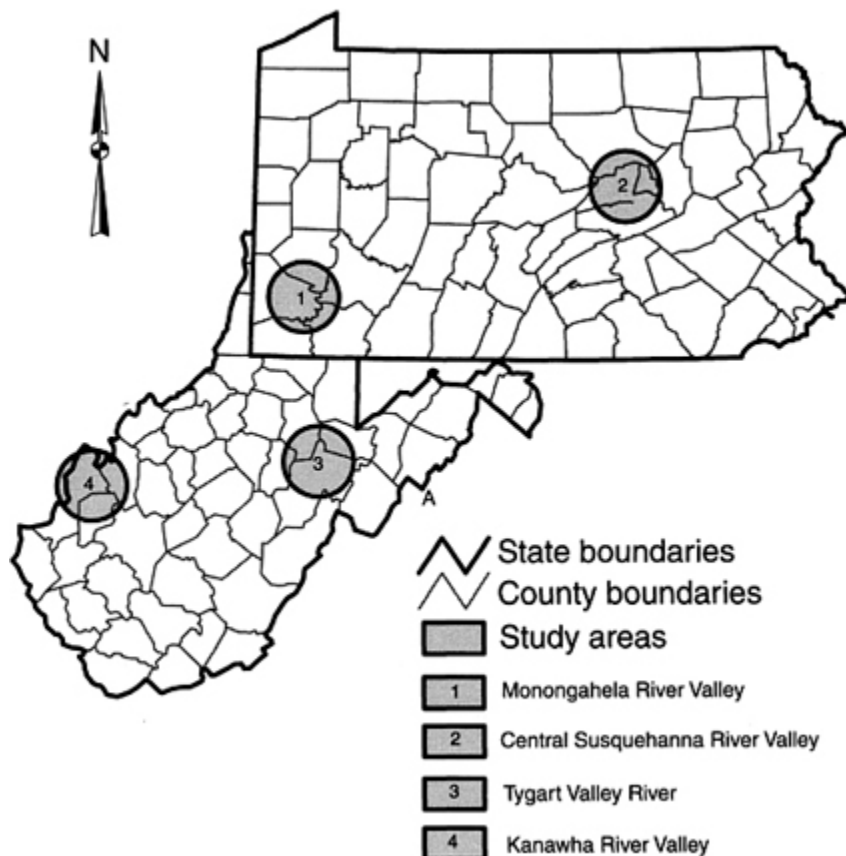


Figure 3.1 Study areas within Pennsylvania and West Virginia.

3.3

PREDICTIVE MODEL DEVELOPMENT

3.3.1

Monongahela River Valley, Pennsylvania

The first study area within which our GIS modeling process was employed was the Monongahela River Valley of southwestern Pennsylvania. The study area was defined as three US Geological Survey USGS. 7.5 quadrangles which contained a proposed highway corridor running across the uplands and tributary valleys west of the Monongahela River (Figure 3.2). The area is situated within the unglaciated Appalachian Plateau, a landscape of broad ridges, heavily dissected drainages, and narrow valleys and floodplains. The GIS-based archaeological predictive model was produced for this project in order to

evaluate a previous model that had been applied within the early stages of the highway corridor study. The GIS model was designed to be quantitative in nature, allowing the results to be statistically analyzed. The highway construction right of way had previously undergone full Phase I archaeological testing, the results of which were left out of the model construction process in order to create a valid external test of the model.

Methods

The following methodology is built upon the foundation of GIS modeling developed and reported by Kvamme (1983, 1986, 1989, 1990), Kohler and Parker (1986), Savage (1989) and Warren (1990). The basic assumption of the GIS model is that the prehistoric settlement and utilization within the study area was both dependent upon and restricted by local environmental conditions. Construction of the model relies upon a combination of deductive reasoning, statistical analysis, and the spatial analysis capabilities of the raster module within the GIS. Overall, the GIS model is an inductive model of site potential based on the correlation between known site locations and background environmental variables. The correlation of variables within the study area is mapped as a predictive surface consisting of a grid of 30-m×30-m cells or “land parcels,” each with a site potential score. This score represents the relative “attractiveness” of the cell for use by prehistoric populations. In addition, this score is augmented by a “disturbance factor” which relates to the current potential for preservation of archaeological remains within each cell.

The creation of the GIS model followed a stepwise process:

- 1 the collection of primary data sets;
- 2 the derivation of secondary data sets;
- 3 the sampling of the environmental variables with site locations and random background samples;
- 4 the exploration and statistical analysis of the two populations;
- 5 if appropriate, the implementation of logistic regression analysis;
- 6 the identification of significant variables to be used within the model;
- 7 the creation of a model formula, which is a weighted sum of the significant variable values;
- 8 the creation of the predictive surface from the formula;
- 9 the internal testing of the model against the model training sample;
- 10 the external testing of the model against an independent sample;
- 11 reiteration of the model formula and predictive surface given the testing results; and
- 12 continued updating of the model given future discoveries.

The GIS modeling process focused on several primary *data sets* which are relatively common to most areas and which can be easily obtained or digitized.

These primary data sets included recorded prehistoric site locations, historically documented Indian trails, roads and other disturbance factors, hydrologic features (e.g., rivers, streams, lakes, wetlands and springs), soils data based on the United States Department of Agriculture (USDA) soil surveys, and a digital elevation model (DEM).

From these primary data sets, secondary data sets such as slope, aspect, and distance to water were developed within the GIS. By focusing on a small portion of the study area (see [Figure 3.2](#), inset), various data sets and derived variables can be examined. The actual derivation of the secondary variables was a complex process, the description of which is beyond the intent or scope of this chapter (see Kvamme 1989, 1992; Kvamme and Kohler 1988). Of the 70 variables that were initially examined for potential significance, 26 were selected for use in the predictive model. Only a portion of these selected variables will be illustrated in this chapter.

Several primary data sets were purchased or digitized as vector data for use within the GIS. [Figure 3.3](#) shows the digitized data for streams and other hydrologic features, major drainage divides, roadways, and topographic contour lines within the inset area.

The DEM, obtained from USGS 7.5 data, is an extremely important data set within the model. A large number of variables are derived from or influenced by the elevation data set. In [Figure 3.4](#) elevation is shown by a grayscale range in which black represents the lowest elevations around the Monongahela River and white represents the highest elevations in the uplands.

There are several derivations of the elevation data set, including slope, aspect, terrain roughness, and relief. These derivatives are created by the GIS from the DEM. [Figure 3.5](#) presents a map view of the slope variable. Slope is used both as an independent variable and as a cost surface for many of the cost-distance functions. Using the grayscale range, black represents low slope and white represents high slope.

As an example of the complexity of some of the derived variables, [Figure 3.6](#) presents a map of the solar insolation gain of the landscape for the morning of 22 December, the shortest day of the year. Solar insolation gain relates aspect, slope, and elevation, to derive a data set that has more relevance to site location during colder seasons than the simple variable of aspect. In [Figure 3.6](#), the white areas represent locations with the highest solar gain or sun warmth on a cold winter morning.

In addition to DEM-derived variables, cost-distance analysis was performed using slope as the cost surface for water, Indian trails, fifth-order drainage divides, saddles, and vantage points. The results of the cost-distance analysis to permanent water resources are shown in [Figure 3.7](#), with black representing the lowest cost and white the highest cost. The cost-distance method provides a more relevant and accurate reflection of the effort which prehistoric populations had to expend to obtain a given resource.

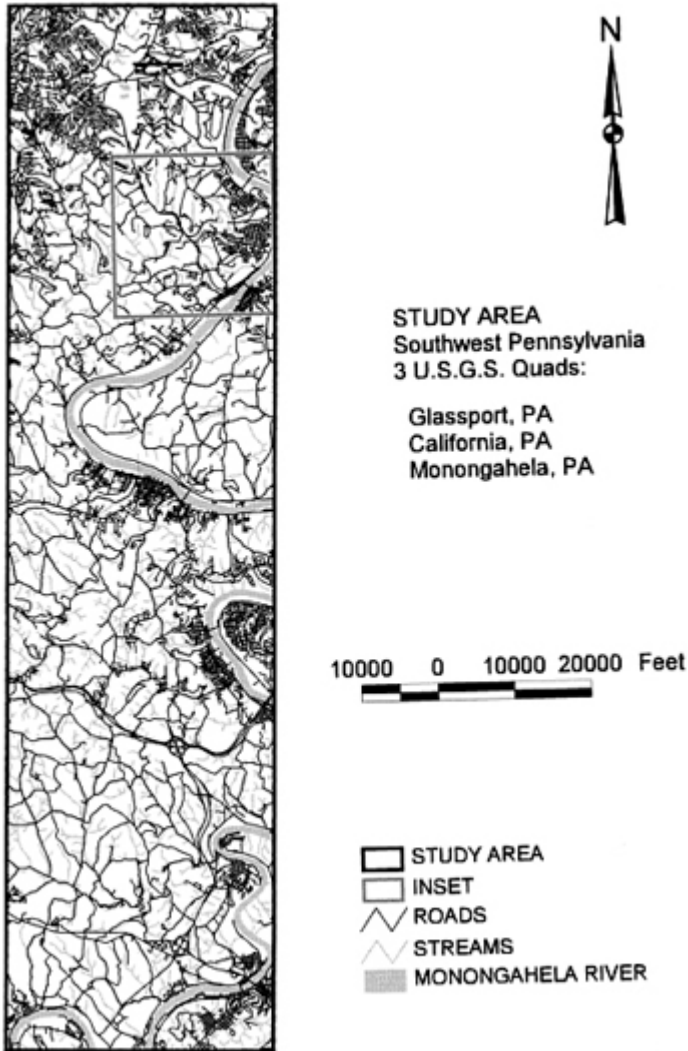


Figure 3.2 Monongahela River Valley study area.

The GIS was also used to identify certain landforms, such as saddles, peaks, vantage points and rims, so that subsequent correlations with site locations could be made. Figure 3.8 shows the saddles (black to white squares) and peaks (black triangles) identified by the GIS within a portion of the inset area. The squares denoting the saddles in Figure 3.8 are ranked from “major saddles” (black) to “minor saddles” (white) by the GIS. Saddles were particularly important landforms for site locations during certain cultural periods within this study area.

Selection of variables is based on the assumption that prehistoric site locations should occupy only a limited portion of the total variation present in the

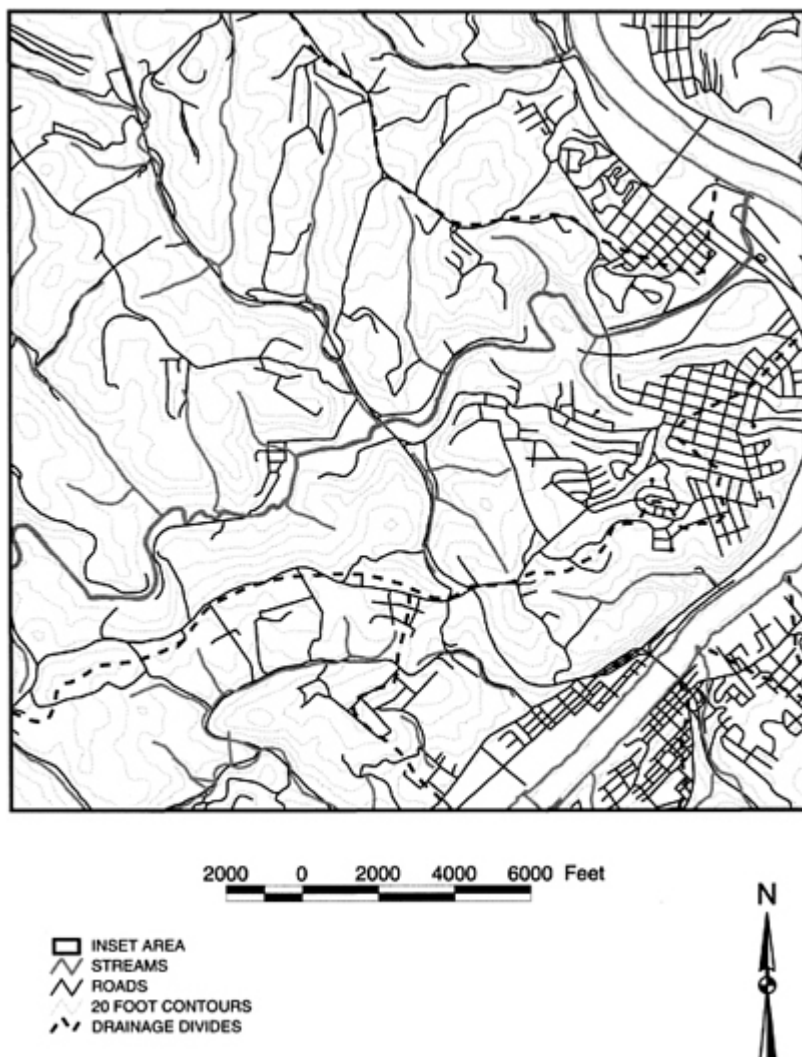


Figure 3.3 Inset location: streams, drainage divides, topography, and roads.

environment (Kvamme 1985). Exploratory and univariate statistics were used to compare the distributions of variable values for both site and background location samples. Univariate analysis consisted of the Kolmogorov-Smirnov two-sample test and/or the standard Chi Square test (Thomas 1986). [Figure 3.9](#) is an example of a bar graph comparing two populations—site cells and random background cells—for the slope variable. Once identified, the potentially significant variables were subjected to logistic regression (Kvamme 1988; Rose and Altschul 1988; Warren 1990), in which background cells were assumed to be

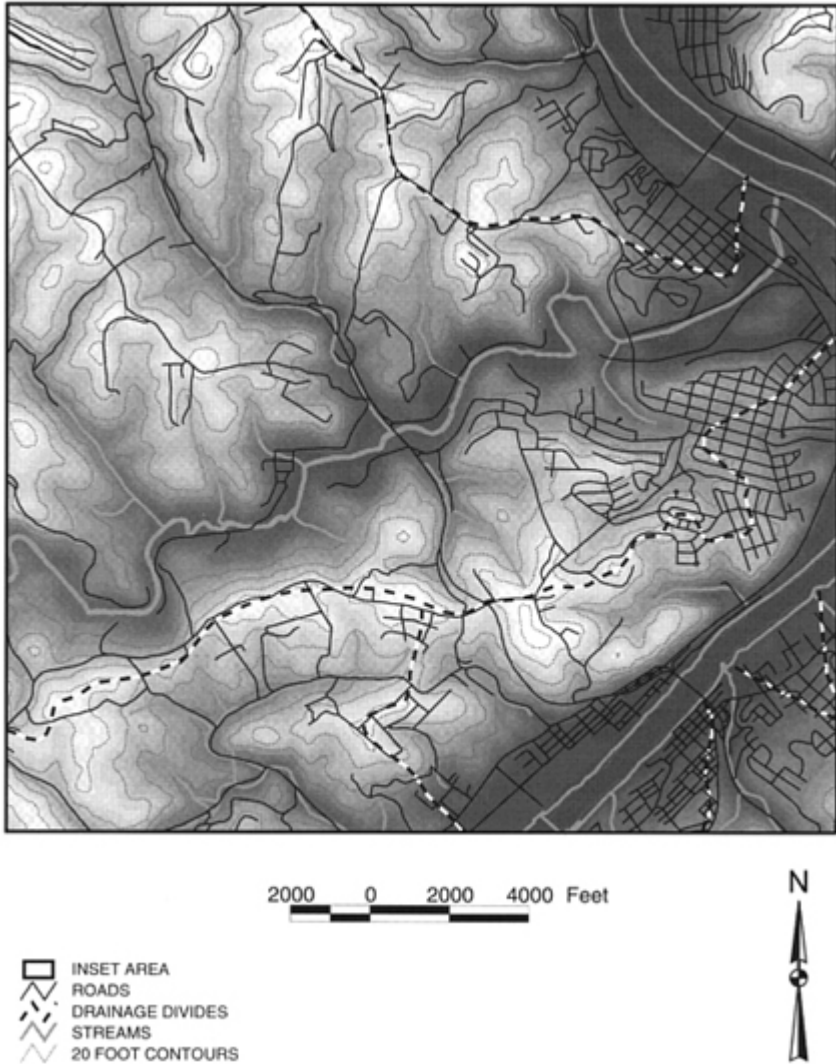


Figure 3.4 Digital elevation model within inset area.

nonsite locations. The results of the logistic regression were used to redefine and adjust the relative weights of the selected variables within the predictive model formula. This formula was applied to the study area using the raster map algebra function of the GIS. The predictive model formula was designed as a weighted sum of the selected variable values, resulting in a site potential score for each of the 500,000 or more 30 m cells contained in the study area. This score is essentially a combined measure of the “attractiveness” of the cell as a site location and the “potential for preservation” of any site remains within the cell. The score is not a

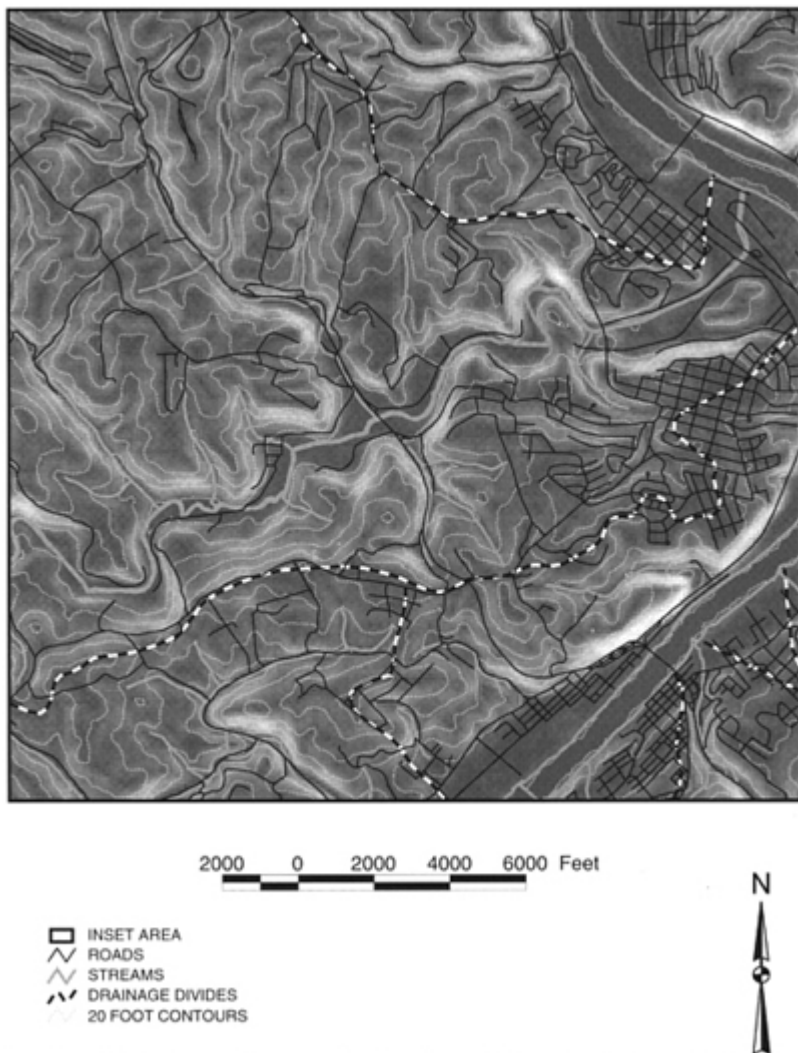


Figure 3.5 Slope within the inset area.

statistic of mathematical probability for the occurrence of a site in the cell. The GIS predictive model produced a range of site potential scores from 0 to 425 within the study area.

Internal model test

The model was tested internally against the training sample of 2082 site and random background cells. A graph of percentage of correct predictions in

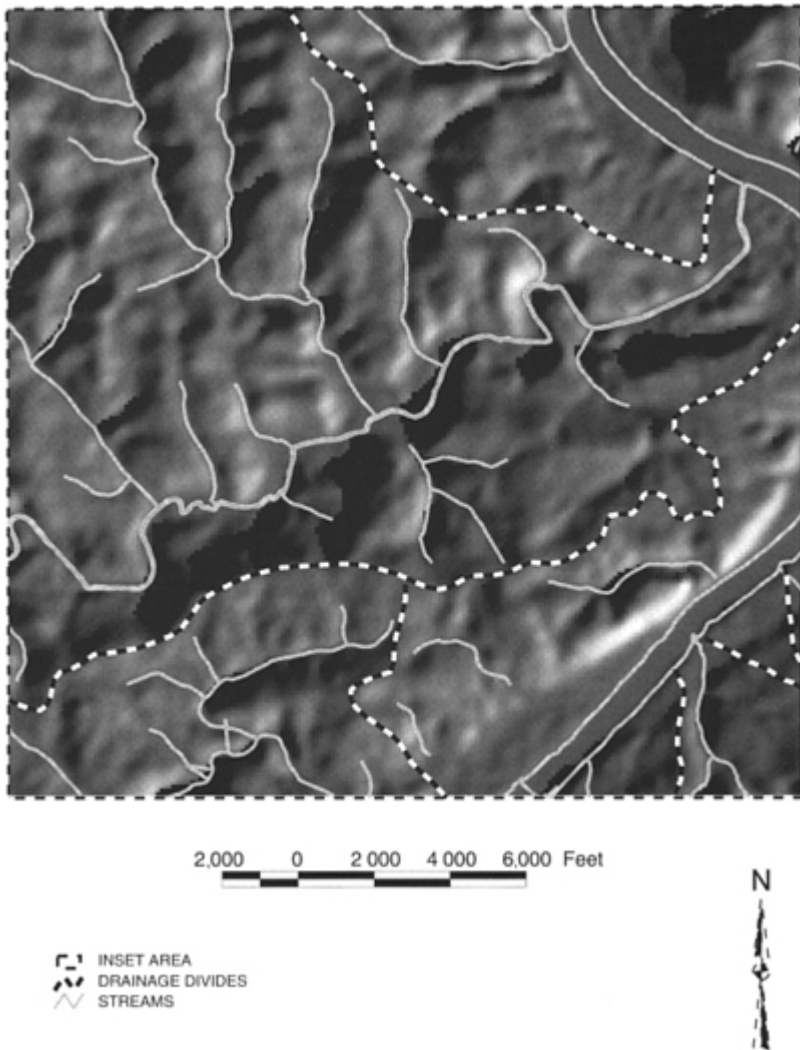


Figure 3.6 Solar gain on the morning of 22 December.

Figure 3.10 shows the results of this test, indicating that the highest level of correct predictions (approximately 75%) of both site and “nonsite” (or background) locations is at a site potential score of approximately 245. If this score is used as a cut-point or decision point, a two-by-two contingency table analysis indicates that the result is statistically significant. In spite of the statistically significant result, an internal test of a predictive model is by nature overly optimistic. Therefore, an external test was performed in order to validate the performance of the model.

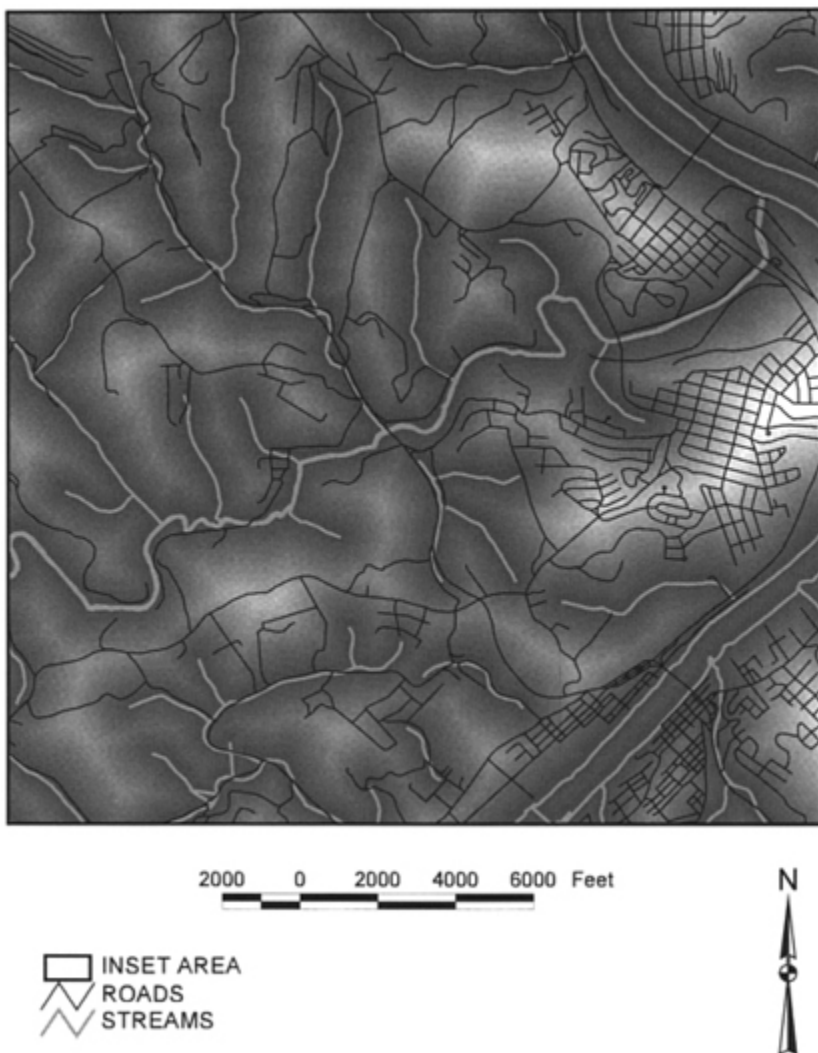


Figure 3.7 Cost distance to permanent water sources within the inset area.

External model test

The external test of the GIS model utilized the data set from the Phase I archaeological study of the proposed corridor. This data was omitted from the model construction process expressly for this purpose. The external data set consisted of 5262 cells that fell within the tested corridor. [Figure 3.11](#) indicates the results of the external test of the predictive model. The results were very similar to the internal test, with a peak accuracy of approximately 78% at a cut-

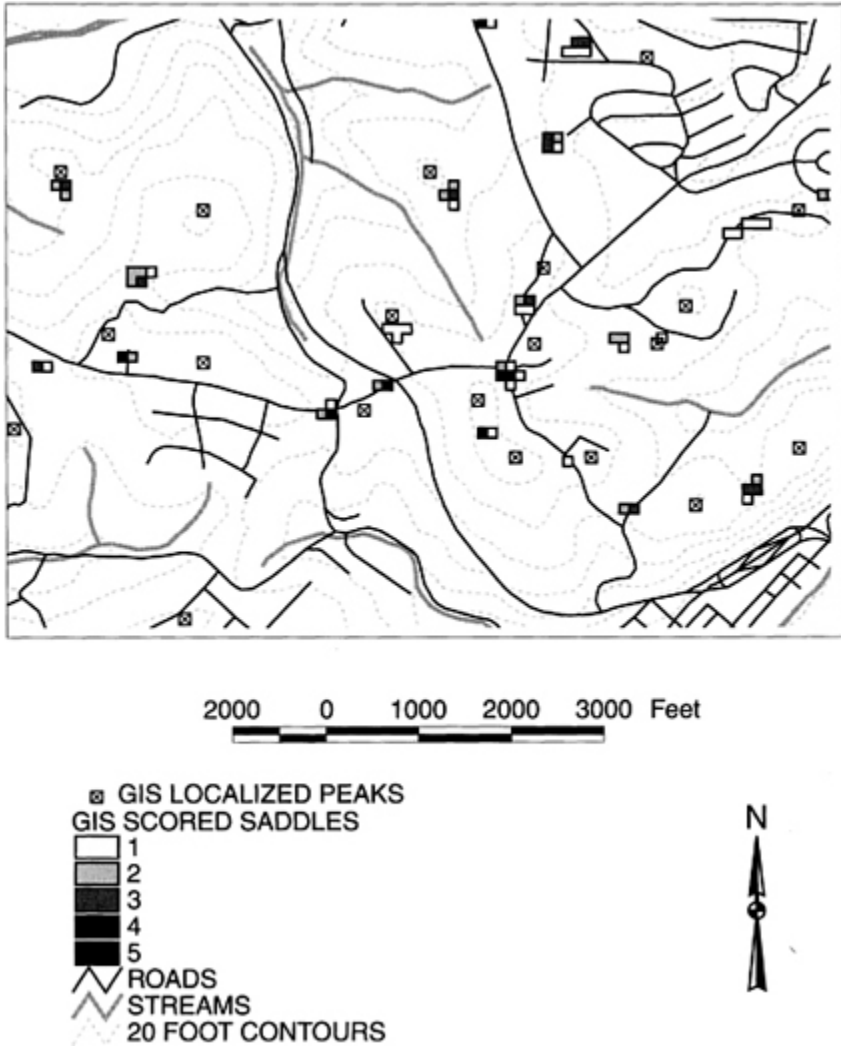


Figure 3.8 Landforms identified by the GIS: saddles and peaks. point score of 220. The contingency table analysis again indicated a significant result at this cut-point.

Results

The final result of the modeling process is the predictive surface applied across the entire study area. [Figure 3.12](#) shows the range of site potential created within the three USGS quadrangle study area. For display purposes, the 0 to 425 range

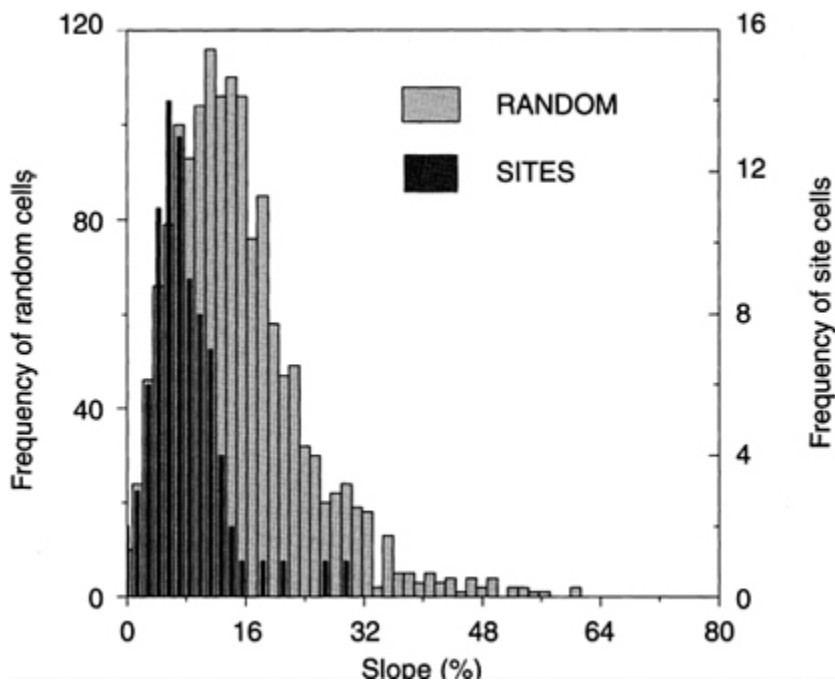


Figure 3.9 Population graph of the slope variable for random background and site cells. of site-potential scores has been reclassified into 20 intervals and assigned values along the greyscale, white for high potential, black for low potential.

In order to give some indication of the discriminating ability of the GIS predictive model, [Figure 3.13](#) is a close-up of an area tested by Phase I archaeological survey which was found to contain several artifact concentrations. As shown, the model was able to identify both high-potential locations (dark) and intrasite areas of less potential (light) which correlated well with the documented finds from the field-testing (dashed site boundaries).

3.4

ADDITIONAL GIS PREDICTIVE MODELS

3.4.1

Tygart Valley River, West Virginia

Since the development of the predictive model in the Monongahela River Valley, Skelly & Loy has utilized the GIS modeling process to predict archaeologically sensitive locations for three other projects whose settings and other characteristics are quite varied. The Tygart Valley River study area, shown in [Figure 3.14](#), is located in the Allegheny Mountain section of the Appalachian

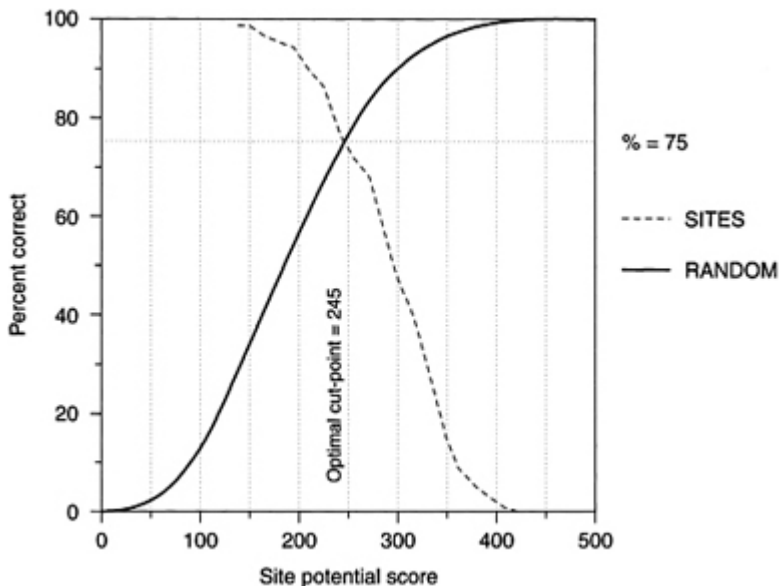


Figure 3.10 Results of the internal model test: percentage of correct predictions.

Plateau province in northern West Virginia. The study area is relatively small (two USGS 7.5 quadrangles) and consists primarily of floodplain and terrace landforms circumscribed by low mountains and steep, narrow ridgelines. The area is dominated by the broad valley and sweeping meander loops of the Tygart Valley River and its major stream confluences. As shown in [Figure 3.15](#), the predictive model estimates that the areas of higher site potential are generally confined to the well-drained lowlands in undisturbed locations. The site potential scores ranged from 0 to 1,100. An internal test of the model discovered an optimum cut-point score of 500 with an associated correct percentage for both site and nonsite cells of approximately 79%. An external test of the Tygart Valley River model is planned.

3.4.2

Central Susquehanna River Valley, Pennsylvania

The Central Susquehanna River Valley study area is located in the Ridge and Valley physiographic province of central Pennsylvania. The study area, shown in [Figure 3.16](#), consists of steep ridges, narrow valleys and numerous streams and river bodies. The site potential within the area is dominated by landforms along the rivers, major tributaries, and larger streams. A fully operational predictive model has not yet been prepared for the study area. However, a preliminary archaeological resource sensitivity map for the area has been produced and is presented in [Figure 3.17](#). The development of the sensitivity map used fewer data

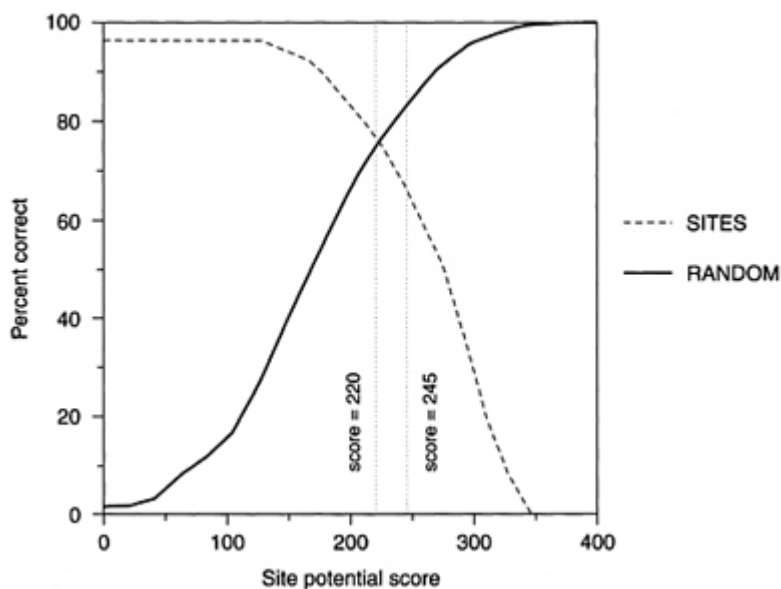


Figure 3.11 Results of the external model test: percentage of correct predictions.

sets and more cursory data analysis than the predictive model process described above. The sensitivity map was developed for use as an early-stage planning tool and may be upgraded to a full predictive model for later stages.

3.4.3 Kanawha River Valley, West Virginia

The Kanawha River Valley study area is located in the unglaciated Kanawha section of the Appalachian Plateau province in West Virginia, near the Ohio border (Figure 3.18). The study area is defined by nine USGS 7.5 quadrangles that contain the Kanawha River Valley and the uplands to the southwest. The area is characterized by the broad floodplains and terraces along the Kanawha River, the steep, narrow ridges of the uplands, and the winding “hollows” or valleys of the tributary streams. As in many areas, most recorded archaeological site data occurs in lowland, floodplain, and terrace locations. The documentation on upland sites in the area is extremely sparse. This disparity of site data necessitated the creation of independent upland and lowland models, which were later combined. In addition, it diminished the utility of an internal model test. The tentative model produced for the Kanawha River study area has site-potential scores ranging from 0 to 800 for 1.4 million cells. The highest site potentials occur within well-drained soils of the Kanawha River floodplain, as indicated in Figure 3.19. However, areas of high potential for archaeological sites do occur within the uplands, particularly in the upland “hollows” and “passes.” An

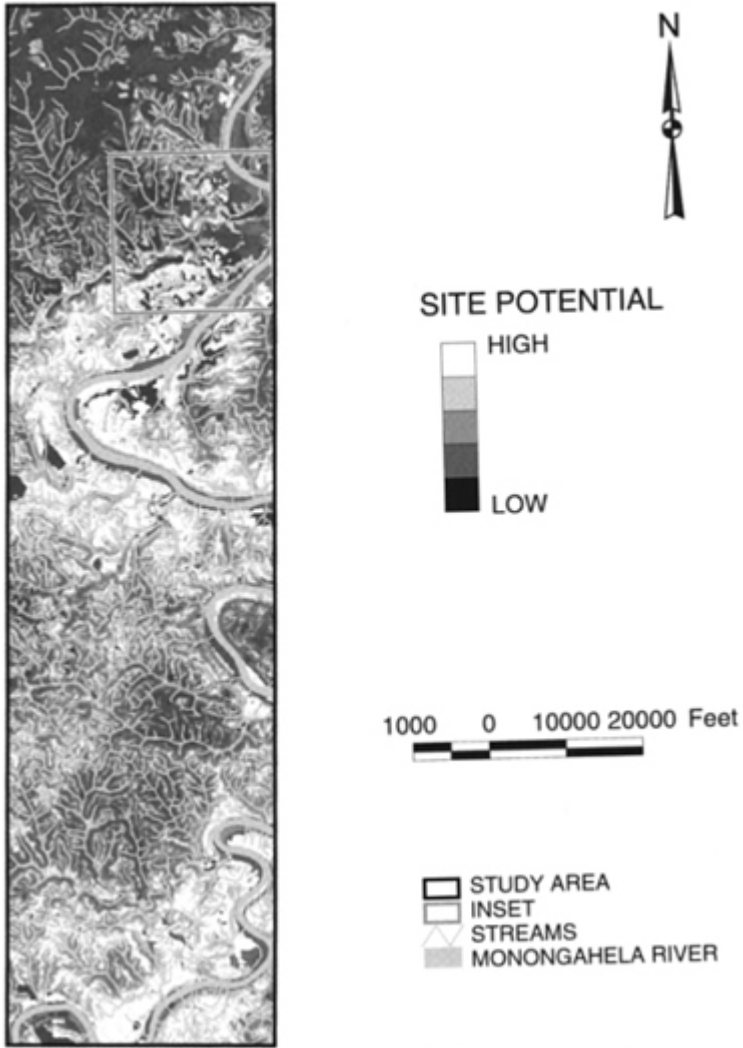


Figure 3.12 Predictive surface of site potential within the study area.

external test of the Kanawha River Valley predictive model is currently under way.

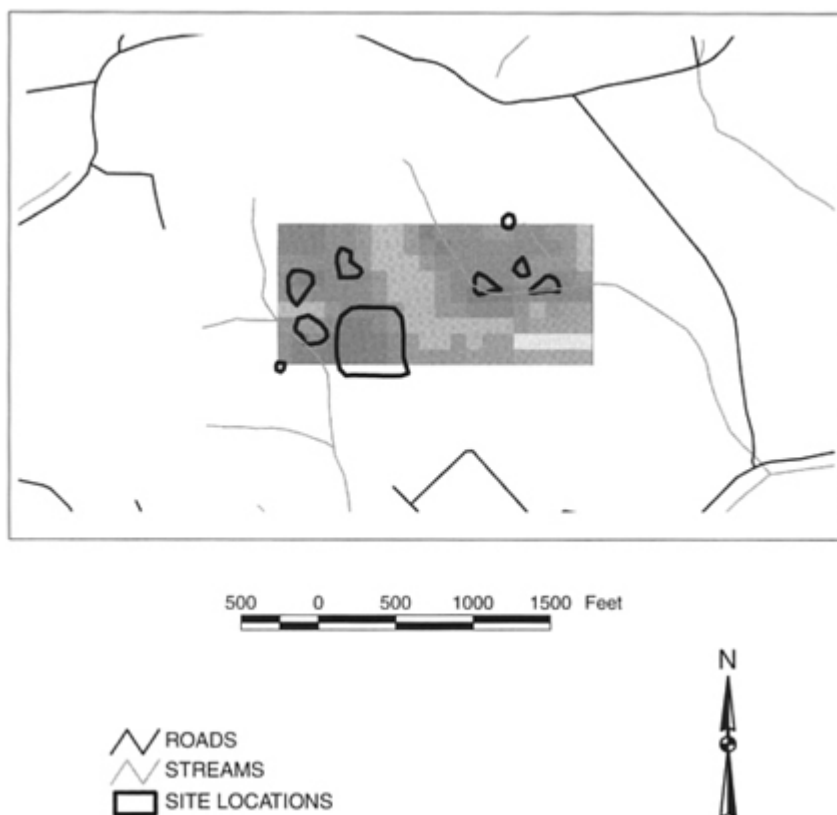


Figure 3.13 GIS model results compared to phase I archeological testing.

3.5 DISCUSSION

Although only one of the predictive models presented in this chapter has undergone external testing, the preliminary results indicate that the application of GIS to archaeological predictive modeling has been successful within a cultural resource management context. After we had applied the modeling process to the diverse regions outlined above, it was discovered that the process can be applied consistently from one study area to another, regardless of differences in the relative importance of selected variables as site predictors from region to region. Thus, the application of the GIS allows for the creation of a geographically flexible and adaptable modeling process.

The promotion, development, and application of GIS predictive modeling as a cultural resource management tool should be tempered by the consideration of a number of concerns, some of which will be mentioned here. As many researchers working with recorded site files can attest, there are often problems with the accuracy or completeness of data regarding site locations and cultural

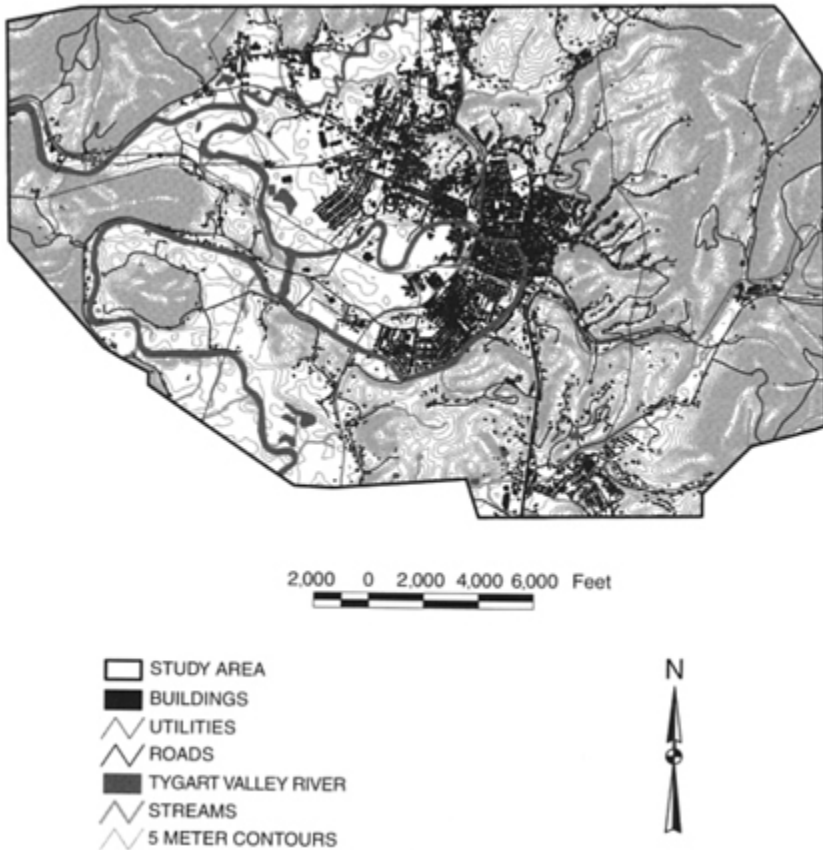


Figure 3.14 Tygart Valley River study area.

information. These problems may preclude any functional or temporal considerations within the site data set of the model. Inaccuracies in site location or extent may lead to erroneous environmental correlations. In addition, the site record is a highly biased sample of site distribution across the landscape, both in methodology and in documentation. When attempts are made to determine areas of relative probability for site locations, the lack of sufficient and reliable data ultimately controls the modeler's ability to create an effective inductive model. In all cases, the bias and insufficiency of the known site record is an inherent and pervasive problem that is difficult to overcome (Kvamme 1988).

Another consideration is the nature of the environmental data set. The modeling process assumes that the attractiveness of the land parcels in the past can be related, either directly or indirectly, to currently measurable modern characteristics across the landscape. This assumption may be faulty, and at times directly misleading. For instance, in the search for statistical relationships between sites and environmental variables within the Monongahela River Valley

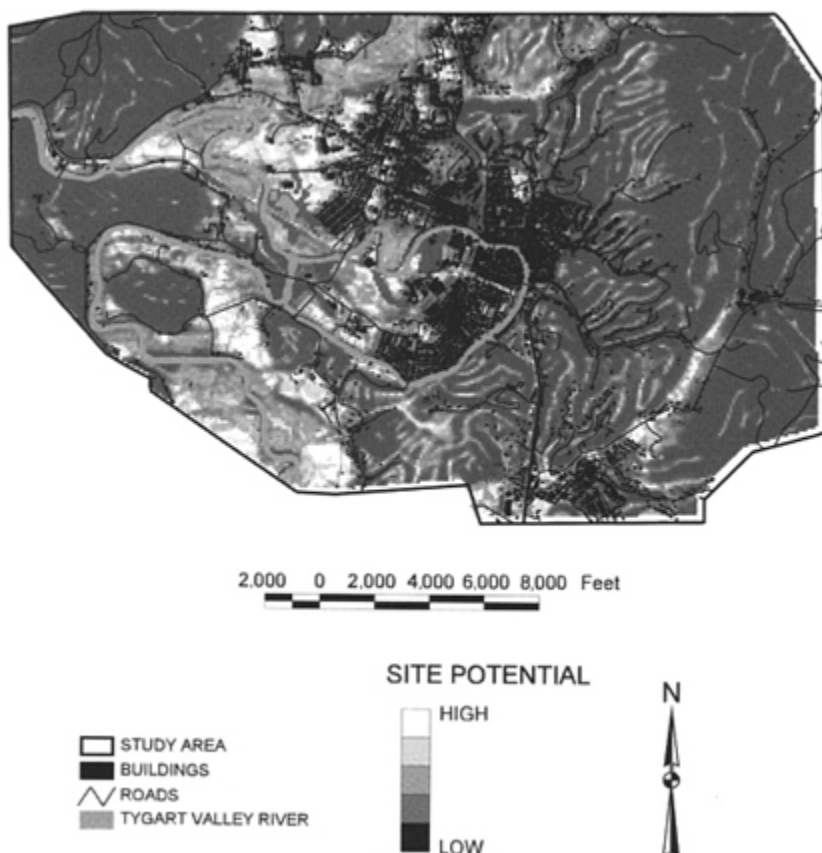


Figure 3.15 Predictive surface of site potential within the Tygart Valley River study area.

study area, no statistical correlation was found between site locations and the existence of springs as a nearby water source. This lack of correlation was counterintuitive, given the general archaeological knowledge in the area. Closer examination of the data sets determined that a number of factors contributed to this incongruity, including the fact that disturbance and/or development may have obliterated the location of springs; that springs are not consistently mapped features; that changes in the water table may have occurred; and that the areas which do have mapped springs tend to be those which have not been seriously investigated/documentated by archaeologists.

Numerous additional concerns have been raised, and many of them are well presented and discussed within the chapters of this book. Ultimately, the nature and accuracy of the data sets used for the development of the GIS model condition and limit the effectiveness and utility of the model, regardless of the sophistication of the process that follows. The data and their inherent

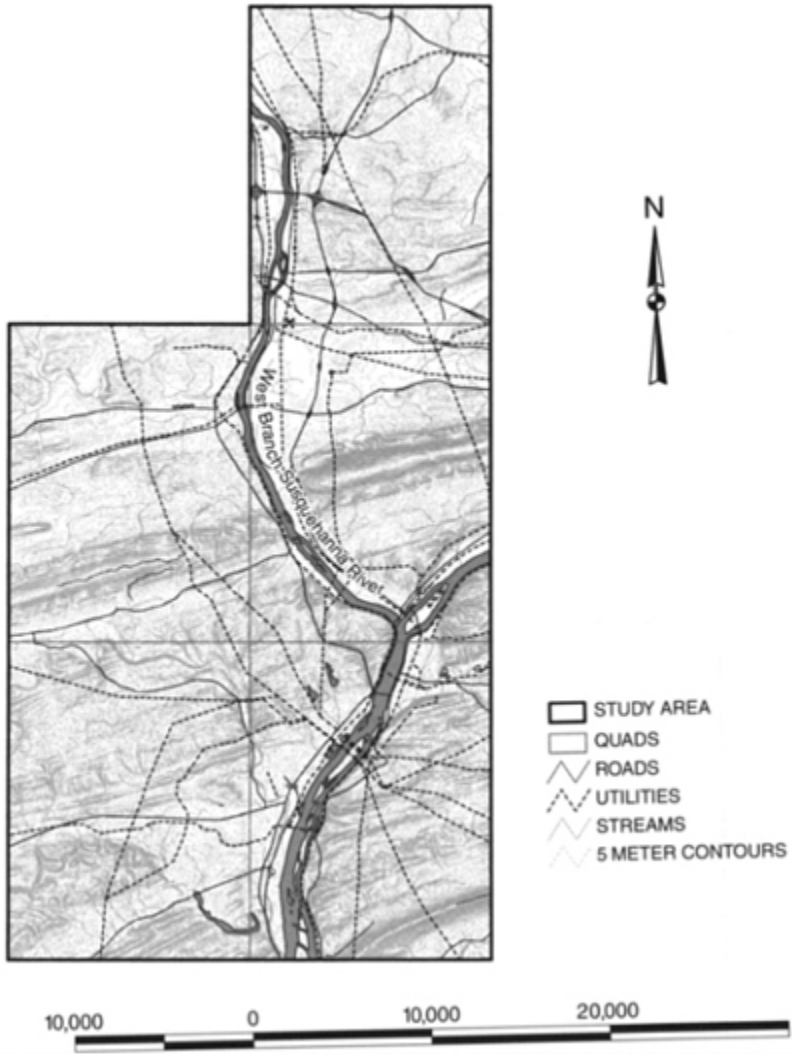


Figure 3.16 Central Susquehanna River Valley study area.

relationships and weaknesses must be carefully inspected and considered within the production of the predictive model.

3.6 CONCLUSIONS

An archaeological predictive model is always a work in progress. There is no absolute correlation between predictions and site locations, merely a level of

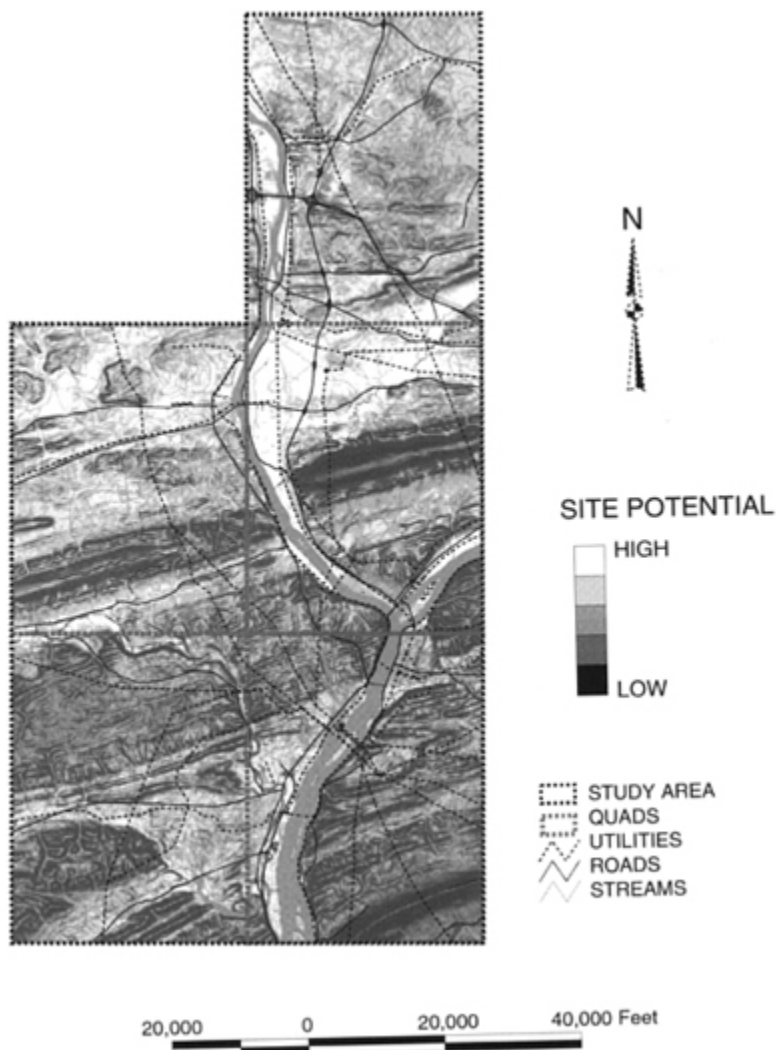


Figure 3.17 Sensitivity map for site potential within the Central Susquehanna River Valley study area.

confidence at which the model becomes a useful tool. The requirements of the particular study or project determine the level of confidence necessary for the model to obtain. There are many influences that a predictive model based on current levels of generally available data cannot account for, such as sociopolitical factors, or the occurrence of sites in less predictable locations, such as steep, generally uninhabitable terrain (rock shelters, quarries, etc.). In turn, the scoring of an area by the model as having a very high “attractiveness” for use or habitation does not mean that the probability of discovering a site in that area is

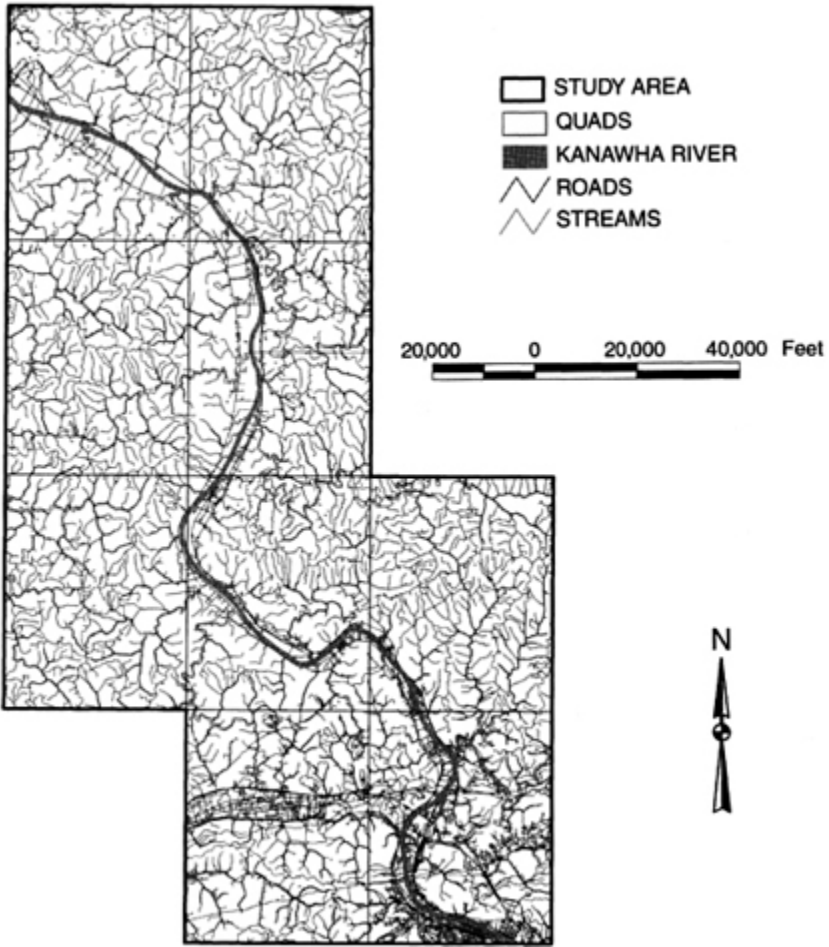


Figure 3.18 Kanawha River Valley study area.

100%. Sites are relatively rare concentrations of recoverable remains within an active landscape. Although the model may assign a high probability score to a parcel of land, it does not necessarily follow that a sufficient amount of human activity occurred at that location to have produced nonperishable and preserved remains. In sum, the GIS predictive model is a tool whose utility is relative to the nature of the data we employ, the expectations we make, and the limits of the creativity we bring to the task. Despite the technical challenges and the pitfalls of our assumptions, the application of GIS can be a valuable, flexible, and powerful tool for use in the development of archaeological predictive models.

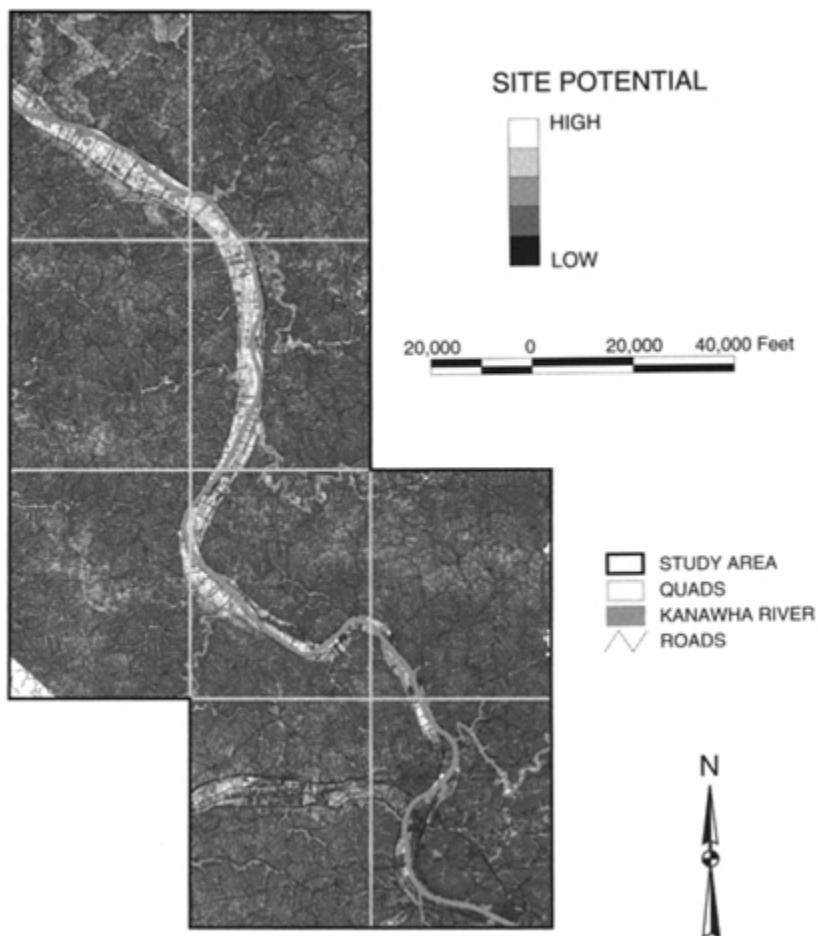


Figure 3.19 Predictive surface of site potential within the Kanawha River Valley study area.

Acknowledgments

The research discussed in this chapter was performed by Skelly & Loy, Inc., engineers-consultants, of Harrisburg and Monroeville, Pennsylvania. This work was supported by the Pennsylvania Turnpike Commission, the Pennsylvania Department of Transportation, and the West Virginia Department of Transportation. This chapter is based on a paper presented at the annual meeting of the Society for American Archaeology in 1996.

References

- ALLEN, K.M.S., GREEN, S.W. and ZUBROW, E.B.W. (eds), 1990, *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis.
- BRANDT, R., GROENWOUTD, B.J. and KVAMME, K.L., 1992, An experiment in archaeological site location: modeling in the Netherlands using GIS techniques, *World Archaeology* **24**(2): 268–282.
- CALAMIA, M.A., 1986, *Geographical Information System Applications for Cultural Resource Management*, Denver: Bureau of Land Management.
- CARR, C., 1985, Introductory remarks on regional analysis. In C.Carr (ed.) *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*, Kansas City: Westport Publishers.
- ESRI, 1992, *Understanding GIS, the Arc/Info Method*, Redlands, CA: Environmental Systems Research Institute.
- JUDGE, W.J. and SEBASTIAN, L. (eds), 1988, *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, Denver, CO: US Department of the Interior, Bureau of Land Management Service Center.
- KING AID, C., 1988, Predictive modeling and its relationship to cultural resource management applications. In W.J.Judge and L.Sebastian (eds) *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, Denver, CO: US Department of the Interior, Bureau of Land Management Service Center, pp. 549–569.
- KOHLER, T.A. and PARKER, S.C., 1986, Predictive models for archaeological resource location. In M.B.Schiffer (ed.) *Advances in Archaeological Method and Theory*, vol. 9, Tucson: University of Arizona Press, pp. 397–452.
- KVAMME, K.L., 1983, Computer processing techniques for regional modeling of archaeological site locations, *Advances in Computer Archaeology* **1**:26–52.
- 1985, Determining empirical relationships between the natural environment and prehistoric site location: a hunter-gatherer example. In C.Carr (ed.) *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*, Kansas City: Westport Publishers, pp. 208–238.
- 1986, The use of geographic information systems for modeling archaeological site distributions. In B.K.Opitz (ed.) *Geographic Information Systems in Government*, Vol. 1, Hampton: A.Deepak Publishing, pp. 345–362.
- 1988, Using existing archaeological survey data for model building. In W.J.Judge and L.Sebastian (eds) *Quantifying the Present and Predicting the Past: Theory, Method and Application of Archaeological Predictive Modeling*, Denver, CO: US Department of the Interior, Bureau of Land Management Service Center, pp. 301–323.
- 1989, Geographic information systems in regional archaeological research and data management. In M.B.Schiffer (ed.) *Archaeological Method and Theory*, Vol. 1, Tucson: University of Arizona Press, pp. 139–203.
- 1990, The fundamental principles and practices of predictive archaeological modeling. In A.Voorrips (ed.) *Studies in Modern Archaeology, Mathematics and Information Science in Archaeology: A Flexible Framework*, Vol. 3, Bonn: Holos Verlag.

- 1992, A predictive site location model on the high plains: an example with an independent test, *Plains Anthropologist*, **37**(138): 19–40.
- KVAMME, K.L. and KOHLER, T.A., 1988, Geographic information systems: technical aids for data collection, analysis and display. In W.J.Judge and L.Sebastian (eds) *Quantifying the Present and Predicting the Past: Theory, Method and Application of Archaeological Predictive Modeling*, Denver, CO: US Department of the Interior, Bureau of Land Management Service Center, pp. 493–547.
- NEUMANN, T.W., 1992, The physiographic variables associated with prehistoric site locations in the Upper Potomac River Basin, West Virginia, *Archaeology of Eastern North America* **20**:81–124.
- PHILLIPS, J.C. and DUNGAN, S., 1993, Archaeology and the geographic resources analysis support system: a preliminary model of archaeological site location in Santa Rosa County, Florida, *Florida Anthropologist* **46**(4):251–262.
- ROSE, M.R. and ALTSCHUL, J.H., 1988, An overview of statistical method and theory for quantitative model building. In W.J.Judge and L.Sebastian (eds) *Quantifying the Present and Predicting the Past: Theory, Method and Application of Archaeological Predictive Modeling*, Denver, CO: US Department of the Interior, Bureau of Land Management Service Center, pp. 173–255.
- SAVAGE, S.H., 1989, *Late Archaic Landscapes: A Geographic Information Systems Approach to the Late Archaic Landscape of the Savannah River Valley, Georgia and South Carolina*, Anthropological Studies No. 8, prepared by the University of South Carolina South Carolina Institute of Archaeology and Anthropology.
- THOMAS, D.H., 1986, *Refiguring Anthropology: The First Principles of Probability and Statistics*, Prospect Heights: Waveland Press.
- WARREN, R.E., 1990, Predictive modeling in archaeology: a primer. In K.M.S.Allen, S.W. Green and E.B.W.Zubrow (eds) *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis, pp. 90–111.

CHAPTER FOUR

Using a GIS to Model Prehistoric Site Distributions in the Upper Chesapeake Bay

KONNIE L. WESCOTT AND JAMES A. KUIPER

A GIS was used to predict the distribution of prehistoric sites in a largely unsurveyed coastal area of 39,000 acres (15,800 ha) in the Upper Chesapeake Bay. The potential occurrence of sites in this area was assessed on the basis of environmental variables recorded from over 500 known sites in the region. Shell midden sites and nonshell midden sites were treated as two separate data sets to avoid the possibility of competing environmental variables.

4.1

INTRODUCTION

At Argonne National Laboratory, natural and social scientists have been working together to develop a computer-based resource management system for large federal facilities. This system uses a geographic information system (GIS) as its primary means of organizing large amounts of spatial and tabular data for environmental assessment. We believe that GIS technology is having a profound effect on facility management and planning in regard to day-to-day facility operations, especially compliance activities.

Argonne has been working with the Directorate of Safety, Health, and Environment (DSHE) of Aberdeen Proving Ground (APG) in Maryland to develop such a system for environmental management and compliance. The primary goal of the system is to manage APG's environmental resources in an efficient and cost-effective manner. Areas of concern include air and water quality, wetlands, threatened or endangered species, and cultural resources. The first step was to gather all known information about each resource and store it in one place for quick and easy retrieval. The next step was to assess the potential for a series of proposed projects to impact those resources. GIS provides an effective mechanism for integrating management decisions for all of these areas of concern. As part of this effort, over 100 data layers have been compiled on a GIS for APG (DSHE 1996).

Very little of APG had been surveyed for cultural resources; therefore, little information on archaeological sites existed for use in assessing potential impacts. Historic map sources provided helpful information for identifying nearly 500 potential historic site locations; however, information regarding prehistoric sites was lacking. Development of a predictive model for prehistoric site locations using the GIS seemed a cost-effective solution for meeting cultural resource planning and management needs.

4.2

ABERDEEN PROVING GROUND

APG consists of 75 000 acres (30,400 ha) on the western shore of the Upper Chesapeake Bay in Maryland (Figure 4.1). Slightly more than half of the total area (39,000 acres, or c. 15,800 ha) is land, and the remaining area is water. APG has been owned and operated by the Army since 1917, and much of the facility has been spared the significant land disturbance that accompanies urbanization. Although military activity has affected the landscape, especially in core facility operation areas, over 25,000 acres (10,000 ha) of wetlands and woodlands still exist where natural and cultural resources are present and protected.

Only a small percentage of Aberdeen's land area (approximately 1%) has been intensively surveyed for cultural resources. This is largely a result of the unusual character and history of the facility. Access to many areas of the facility is restricted by mission activities (such as ordnance testing), and survey is constrained by the widespread occurrence of unexploded ordnance and, in some cases, chemical contamination. Over 45% of APG is under water, and many of the land areas are currently covered with marshes, which present additional difficulties for archaeological survey.

Forty-six prehistoric sites have been recorded at APG (Maryland Historical Trust (MHT) 1995). Most of these sites were recorded during archaeological work conducted in the late nineteenth and early twentieth centuries or during subsequent work by amateurs in the mid-twentieth century (Envirosphere Co. 1988). A coastal study of the Chesapeake Bay conducted in the 1970s did investigate a portion of APG that was not restricted; 22 sites were recorded (Wilke and Thompson 1977). The known sites are distributed predominantly along the coast (43 of 46 are within 50 m of a shoreline or stream) (Goodwin & Associates 1995). This distribution, however, is biased because of a lack of adequate inland survey. The sites range in age from PaleoIndian to the Late Woodland/Contact period. The predominant site types are shell middens and lithic scatters, although some sites contain ceramics.

4.3 THE MODEL

An extremely conservative predictive map for archaeological resources had been produced for the Aberdeen Proving Ground Cultural Resources Management Plan (Goodwin & Associates 1995). The map showed nearly two-thirds of the land area as having a high potential for containing prehistoric sites. Distance to water was determined to be the variable most likely to influence site location. Therefore, the map was made by creating a buffer area of a specified distance to water (in this case 100 m). However, specific water sources were not identified before applying the buffer, and the results consequently included areas near recent artificial drainage ditches, in addition to major stream courses, as having a high potential for containing sites.

For the purpose of effective impact assessment and for future planning needs, a more refined map was needed. This map could be derived from a predictive model for prehistoric sites developed in conjunction with the GIS data layers. This effort would not only result in an improved map of areas of high potential for prehistoric sites, but also provide a means to integrate the information with the overall management system being developed on the GIS.

The model was developed using “available data,” which is critical to understanding this model and the results. Ideally, a predictive model would be generated in a controlled environment. The unusual situation at APG did not easily allow for such an opportunity. As is often the case in cultural resource management, the time and money required to conduct an intensive controlled survey within APG from which to develop a rigorous model were not necessarily justified by the anticipated end use of the results. Additional clearance requirements, such as thorough metal detector sweeps for locating unexploded ordnance, would also have to be met. Ultimately, however, the safety issue over undetected unexploded ordnance would have seriously affected the amount of survey that it was reasonable to conduct at the facility; therefore, a predictive model based on “available data” was the most feasible approach.

Data was compiled on 572 prehistoric sites recorded throughout the Upper Chesapeake Bay region in areas most closely resembling the APG facility. Known survey areas and prehistoric site locations were transferred to US Geological Survey (USGS) topographic maps from the Maryland Historical Trust (MHT) Mylar over-lays. Data forms were completed for each site on the basis of the contents of the MHT site recording forms and included information on site type, distance to water, soil, topographic setting, slope, elevation, aspect, geomorphic setting, time period, dimensions, and contents (MHT 1995). On the basis of regional data, the assumption was made that APG is a representative subset of the region and any changes that occurred to the prehistoric environment had been uniform throughout the region. For example, paleoclimatic changes and associated fluctuations in sea level would have affected settlement patterns similarly throughout the region (Custer 1989).

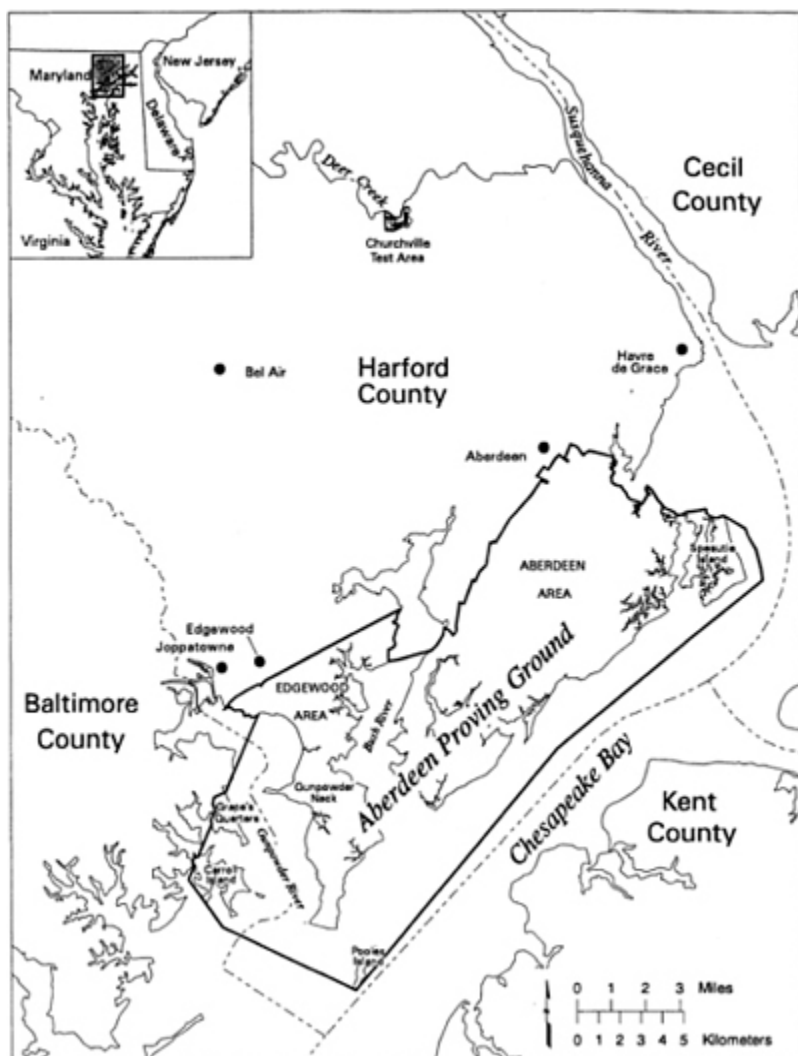


Figure 4.1 Aberdeen Proving Ground and surrounding area.

Two groups of data were separated on the basis of regional site types. Shell midden sites and “nonshell” sites, predominantly lithic scatters, were put into distinct data sets to avoid problems of competing variables. These two categories of sites were assumed to be located with respect to different (although overlapping) sets of variables, and thus have different distributions. Had the two data sets not been separated, the values of the variable(s) for which their distributions differed might have been averaged out within the model and might

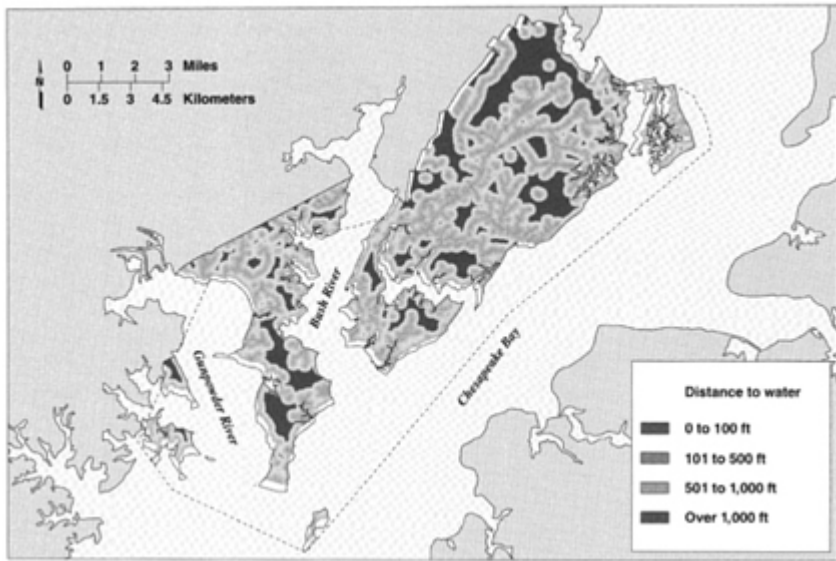


Figure 4.2 Input environmental layer: distance to water.

not have adequately reflected the distribution of either site type. This problem was previously identified by Warren (1990b).

Locations of known sites and survey areas were mapped in the GIS, and a copy of the site database was linked. GIS layers for each of the environmental variables in the database of known sites were also produced. The final layers for the model were made in raster format with a 100-ft (30.48 m) cell size. The final step was to mask the layers against the land area of the site so that edges of the layers would match exactly and could be modeled consistently. Most layers were derived from existing line or polygon layers, but some required a number of steps to produce the final result. On the basis of subsequent statistical analysis, the resulting GIS layers for aspect, slope, soil type, and soil drainage were not used in the final model.

The distance to water layer was produced from a generalized hydrology layer already existing in the GIS (Figure 4.2). Very small channels and known artificial channels or ditches were removed from this layer, and polygons were constructed for the remaining water bodies (i.e., lakes and ponds). Once the layer was converted to raster format, distance to shorelines was calculated with a Euclidean distance function; water bodies themselves were assigned a distance of zero. The water source type layer (described below) was used in a similar process to produce a distance to brackish water layer for use with the shell midden model (Figure 4.3).

Several analyses and processing steps were necessary to create the type of water source layer from available GIS information (Figure 4.4). Shorelines in the

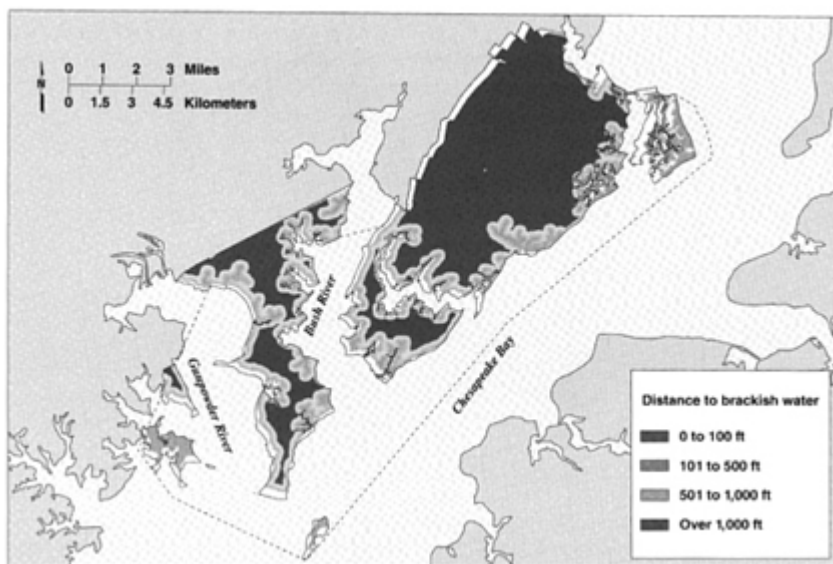


Figure 4.3 Input environmental layer: distance to brackish water.

hydrology layer used for distance to water were subdivided into bay, river, creek, and lake/pond categories. Water chemistry data in the US Fish and Wildlife Service National Wetlands Inventory (NWI) polygons were then used to separate the creeks into brackish and fresh categories. River mouths at APG widen as they reach Chesapeake Bay and are identified as brackish. After conversion to raster format, a Euclidean distance function was used with a distance of 1,000 ft to identify the water source type of all land areas within 1,000 ft (304.8 m) of a major water source. The remaining land area was then coded as having no major water source.

A digital elevation model (DEM) for the entire APG facility existed in the GIS database. The DEM was derived from 2-ft (60.9 cm) contour lines and point elevations for the Aberdeen peninsula and 5-ft (152.4 cm) contour lines for the Edgewood peninsula and the other APG areas. To match the more general elevation categories in the database of known sites, elevations were grouped into four ranges: 0 to 10 ft, 11 to 20 ft, 21 to 40 ft, and over 40 ft (Figure 4.5).

The topographic setting layer included floodplain/beach, terrace, bluff, hill top, hill slope, and interior flat categories (Figure 4.6). The DEM and hydrology data were used to locate the floodplains, which were defined as areas adjacent to shorelines with an elevation of 5 ft or less. Beach areas were designated on the basis of unconsolidated shoreline regions in NWI data and were combined with the floodplain category. Bluffs and hilltops were manually identified by examination of the contour and DEM data. Knowledge of the site, including locations of bluffs and human-caused topographic alterations, was taken into account for much of this work. Remaining land areas were categorized as hill

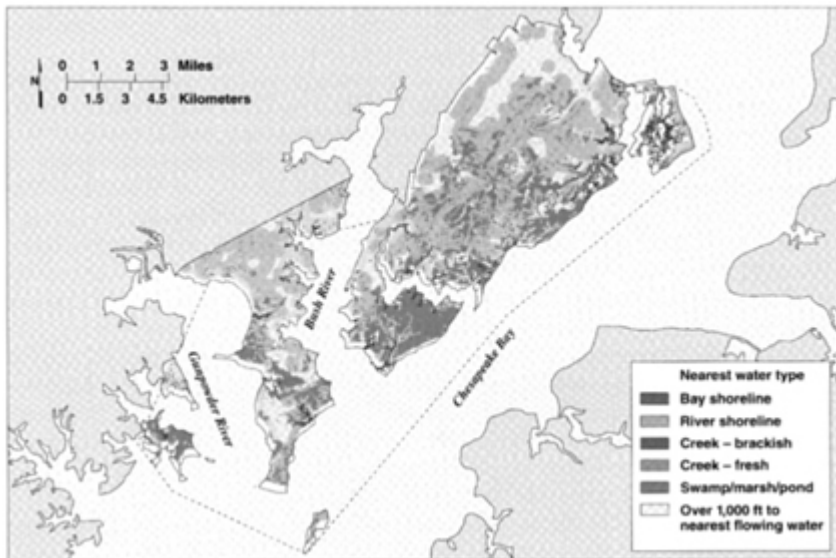


Figure 4.4 Input environmental layer: type of nearest water.

slope if they had a slope of at least 5% and were 150 ft (45.7 m) or greater in width; otherwise they were labeled as inland flat.

A set of 500 points within the APG boundary was randomly selected in the GIS, and the environmental variables associated with those points were derived from the raster data layers. This information was used as a background data set to which the regional site data could be compared.

Frequency tables were generated from the regional data and the APG background data for each of the environmental variables and for groupings of variables. These tables were used to eliminate a number of variables from further consideration. Aspect was eliminated because the predominant aspects for the eastern and western shores are different, and information on aspect is missing on many of the site forms. Topographic relief was too subdued for either slope or aspect to be a significant contributor to site distribution. Soil type was eliminated for a variety of reasons (e.g., the data source was not very current and was potentially unreliable). The considerable variability of the soils within the region also indicated that it might not be a strong or reliable predictor. Neither did soil drainage appear to be a strong predictor because the percentage of site locations occurring on well-drained soils was not significantly different than the percentage of well-drained soils occurring in the background data. (In other words, site locations appear to favor well-drained soils, but the background data indicates that well-drained soils dominate the data set.) Thus, the variables finally used to generate the predictive models were elevation, distance to water, water type, and topographic setting.

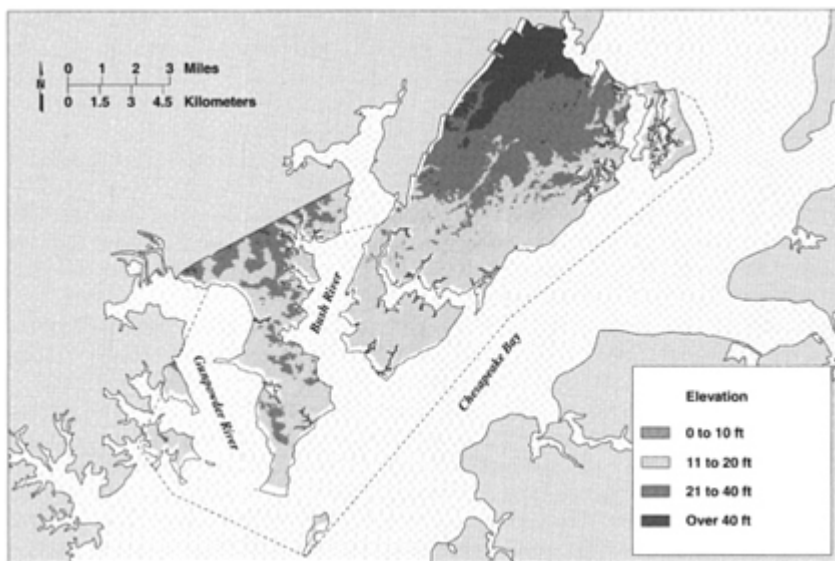


Figure 4.5 Input environmental layer: elevation.

Although logistic regression or log linear analyses typically have been used for this type of analysis (e.g., Kvamme 1985, 1992; Parker 1985; Warren 1990a; Carmichael 1990; Maschner and Stein 1995), our approach using available data does not meet all of the statistical assumptions of these models, such as the availability of nonsite data, or a sampling population of sufficient size.

Frequency tables of unique combinations for elevation (< 20 ft, >20 ft [20 ft = 6.1 m]), distance to water (0 – 500 ft, >500 ft [500 ft = 152.4 m]), water type (brackish, fresh), and topographic setting (terrace/bluff, floodplain/flat) were produced for shell and nonshell sites from the regional data (Tables 4.1 and 4.2). Different categorical breakdowns of variables were investigated, but the most useful results have occurred by using these simple categories that produce larger sample sizes. A high-potential designation was assigned to areas where unique combinations of the four variables occurred over 20% of the time in the regional database. Areas of medium potential for sites occur between 6.25% and 20%, and areas of low or no potential occur at less than 6.25% (an equal distribution of all variable categories is indicated at 6.25%). In the APG GIS, a raster image combining all four of the environmental variables was created. The unique combination of variables for each cell was determined, and then each cell was assigned the proper coding for site potential. Figures 4.7 and 4.8 show the site potential maps for shell and nonshell sites, respectively.

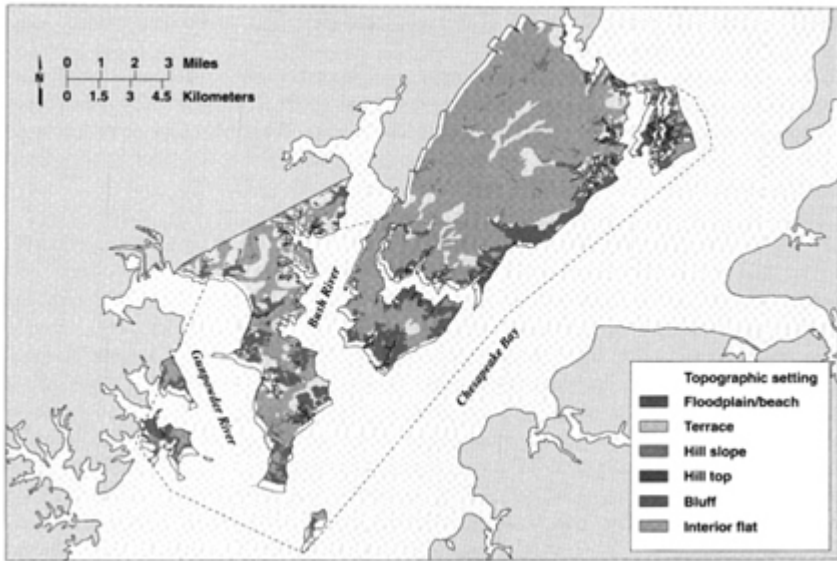


Figure 4.6 Input environmental layer: topographic setting.

Table 4.1 Frequencies of unique combinations for shell prehistoric sites

Distance to Water (ft)	water type	Elevation (ft)	Topography	Frequency	Percentage	Cumulative frequency	Cumulative percentage
0-500	Brackish	20	Terrace/Bluff	75	34.7	75	34.7
0-500	Brackish	20	Floodplain/Flat	81	37.5	156	72.2
0-500	Brackish	>20	Terrace/Bluff	14	6.5	170	78.7
0-500	Brackish	>20	Floodplain/Flat	2	0.9	172	79.6
0-500	Fresh	20	Terrace/Bluff	24	11.1	196	90.7
0-500	Fresh	20	Floodplain/Flat	10	4.6	206	95.4
0-500	Fresh	>20	Terrace/Bluff	4	1.9	210	97.2
>500	Brackish	20	Terrace/Bluff	4	1.9	214	99.1
>500	Fresh	20	Terrace/Bluff	2	0.9	216	100.0

Table 4.2 Frequencies of unique combinations for non-shell prehistoric sites

Distance to water (ft)	Water type	Elevation (ft)	Topography	Frequency	Percentage	Cumulative frequency	Cumulative percentage
0–500	Fresh	>20	Terrace/Bluff	89	27.3	89	27.3
0–500	Fresh	>20	Floodplain/Flat	14	4.3	103	31.6
0–500	Fresh	20	Terrace/Bluff	27	8.3	130	39.9
0–500	Fresh	20	Floodplain/Flat	23	7.1	153	47.0
0–500	Brackish	>20	Terrace/Bluff	27	8.3	180	55.3
0–500	Brackish	>20	Floodplain/Flat	8	2.5	188	57.8
0–500	Brackish	20	Terrace/Bluff	33	10.1	221	67.9
0–500	Brackish	20	Floodplain/Flat	35	10.7	256	78.6
>500	Fresh	>20	Terrace/Bluff	26	8.0	282	86.6
>500	Fresh	>20	Floodplain/Flat	34	10.4	316	97.0
>500	Fresh	20	Terrace/Bluff	2	0.6	316	97.6
>500	Fresh	20	Floodplain/Flat	2	0.6	316	98.2
>500	Brackish	>20	Terrace/Bluff	2	0.6	316	98.8
>500	Brackish	20	Terrace/Bluff	4	1.2	314	100.0

4.4

RESULTS

In the resulting raster image for shell sites at APG, 16.5% of the total land area was coded as high-potential, 2.5% as medium-potential, and 81.0% as low-potential. The corresponding raster image for nonshell sites included 2.7% of the area as high-potential, 44.3% as medium-potential, and 53.0% as low-potential.

Known shell and nonshell sites at APG (not included in the original data set that generated the models) were plotted as initial tests of the models. Of the 13 known shell sites on APG, 12 fell within the high-potential zone and 1 fell within the low-potential zone. Of the 33 known prehistoric sites at APG that did not

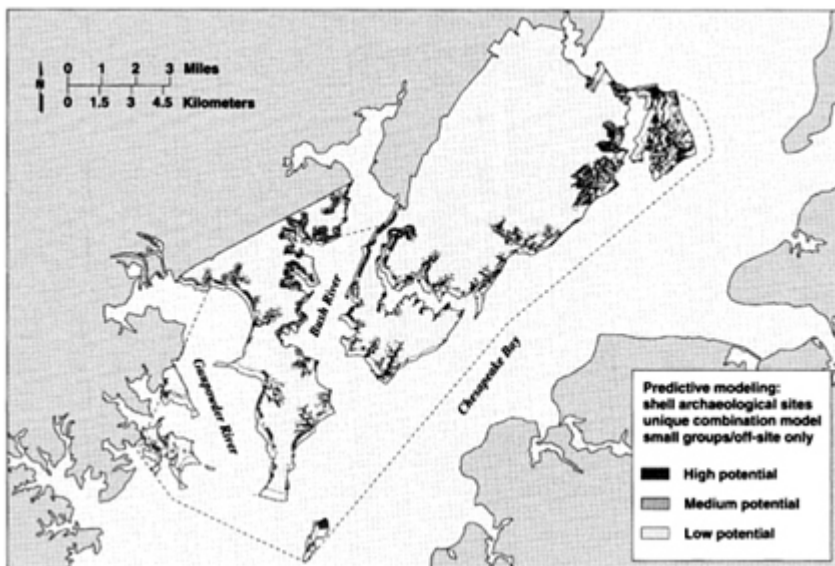


Figure 4.7 Shell predictive archaeological site model.

contain shells, 2 fell within the high-potential zone, 30 within the medium-potential zone, and 1 within the low-potential zone.

Model performance was evaluated using Kvamme's Gain Statistic ($1 - [\% \text{area} / \% \text{sites}]$) (Kvamme 1988). The shell model performed quite well, resulting in a 0.82 gain statistic for high-potential areas and 0.80 for medium-potential areas. These results are consistent with expectations based on the number and level of surveys conducted in coastal areas both regionally and at APG, and the large number of sites recorded.

The model performance results for nonshell sites were considerably lower. The gain statistic for high-potential areas was 0.55 and that for medium-potential areas was 0.52. These numbers are also consistent with our expectations. APG has no known/recorded noncoastal prehistoric sites, and very little of the interior has been intensively surveyed. Therefore, the percentage of known sites falling in high-potential areas was expected to be lower since the regional data that generated the model included results from interior as well as coastal surveys. However, the regional data also favor coastal surveys, otherwise the gain statistic might have been even lower. An additional factor that influenced the results is that less than 3% of APG's land area meets the regional criteria for high potential for nonshell sites, while nearly 45% of the land area is coded as medium-potential. Because less than 1% of APG has been surveyed, these numbers significantly affect the total number of sites falling within the high- and medium-potential areas. Hence, there is a higher number of sites falling in medium-potential areas (30) than in high-potential areas (2).

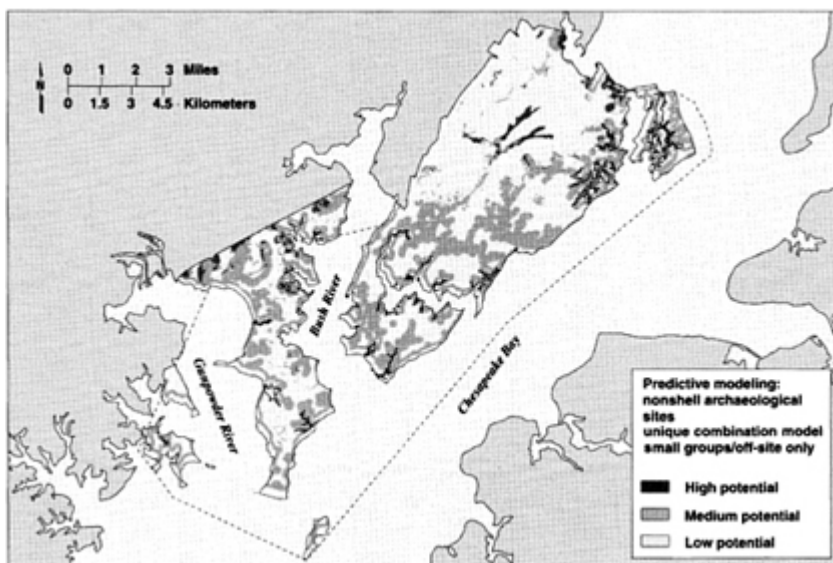


Figure 4.8 Nonshell predictive archaeological site model.

For overall compliance purposes, the two model outputs were combined in order to assess the potential for impact to any prehistoric archaeological sites, shell or nonshell (Figure 4.9). (Areas of potential historic site locations would also be overlain with the combined predictive model results when assessing impacts to cultural resources.) Within the resulting map of prehistoric site potential for APG, 19.2% of the area was coded as high-potential, 29.0% as medium-potential, and 51.8% as low-potential. Forty-two of the 46 known prehistoric sites fell within the high-potential zone, 4 within the medium-potential zone, and no sites fell within the low-potential zone (Figure 4.10). Applying Kvamme's Gain Statistic to these results, a 0.79 gain statistic is attained for high-potential areas and 0.52 for medium-potential areas.

Although the results obtained from the model turned out rather well, several problems or issues must be kept in mind when using these results. The model results are consistent with the bias toward shoreline surveys throughout the region that is apparent in the data set. This bias should be kept in mind as new survey and site data are collected, especially for inland areas. This new data will be invaluable in the future to refine and retest the results of this model.

The use of "available data" was also problematic. We were limited to the variables collected during previous surveys and by the quality of the data on the site forms, which varied considerably. Ideally, it would have been preferable to use a GIS to generate the data for different environmental variables on the basis of known site locations, but such data was not available. For example, more detailed information on soils would have been helpful. It was clear that in most

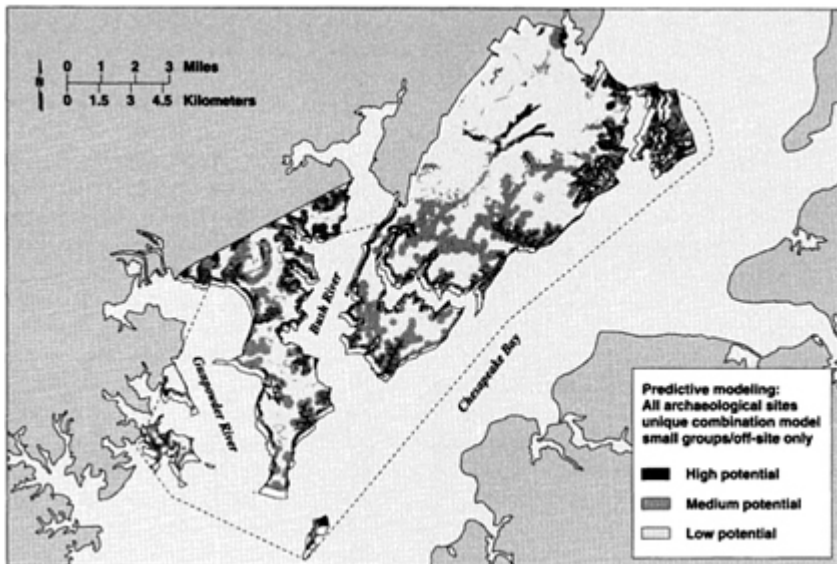


Figure 4.9 Combined predictive archaeological site model.

cases, the information on the site forms was simply taken from the county soil survey maps without field verification (along with slope information that was recorded from the county soil type descriptions). Likewise, soil data from APG had not been verified since the 1927 soils map was produced, again because of ordinance testing and restricted access to the facility.

In addition, a great deal of data regarding the site descriptions was missing, such as unrecorded or unknown site types and associated time periods. As a result, the analysis was limited to site content (shell versus nonshell), which offered reasonable sample sizes, in addition to logical distributional differences.

Despite these problems, the model still provides a useful predictive map that significantly refines and reduces areas of potential high probability for sites. Ground-truthing is a necessary follow-up procedure to determine the model's actual utility. In combination with ground disturbance data on the GIS, additional refinement to the areas of concern can be made by eliminating areas too disturbed to contain eligible sites regardless of their potential. Although a model can identify areas of high potential for sites, it in no way substitutes for or eliminates the need for intensive archaeological survey.

The uses of such a probability map are many. Staff at Argonne are developing ways to link impact models to the GIS data (Hoffecker 1997). For example, a watershed model, ANSWERS, developed by EPA to identify soil movement within watersheds (Beasley and Huggins 1981), has been modified and linked to the GIS. Potential impacts on cultural resources due to soil erosion and deposition can therefore be considered. By overlaying areas of known or

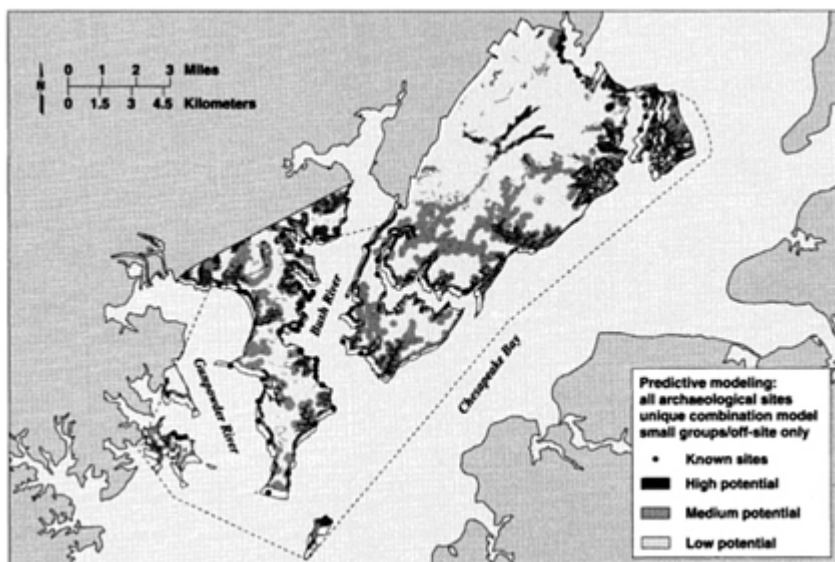


Figure 4.10 Combined predictive archaeological site model with known sites.

predicted archaeological sites with locational-based results of the watershed model, areas of potential impact can be predicted.

The model provides project planners with information about what areas would most likely require less time, effort, and money to develop from a cultural resource compliance point of view. For example, upfront avoidance of a high-potential area could result in less cost for survey and evaluation activities and avoid a potential project delay. The model is also helpful in determining priority areas (as well as evaluation, monitoring and mitigation efforts) for more efficient use of time, money, and human resources. These data can also be used for planning future projects at a level above cultural resource management (i.e., facility management) in which other environmental constraints (e.g., wetlands, contaminated areas) are also taken into consideration.

Acknowledgments

We would like to thank Reed MacMillan, David Blick, and other staff members of the Directorate of Safety, Health, and Environment at Aberdeen Proving Ground for their support and encouragement during the GIS project. We are also very grateful for the helpful comments on this manuscript from John Hoffecker, David Blick, and Georgia Anast. Special thanks go to John Hoffecker and Jennifer Blitz for their assistance in the data collection tasks, and to Richard Olsen for his management support. This project was sponsored by the US Army as part of a larger environmental analysis for Aberdeen Proving Ground. This

work was supported under a military interdepartmental purchase request from the US Department of Defense, US Army, through US Department of Energy Contract W-31-109-Eng-38.

Disclaimer

Neither the US Army, nor any of its employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Reference to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the US Army. The views and opinions of the authors do not necessarily state or reflect those of the US Army.

References

- BEASELY, D.B. and HUGGINS, L.F., 1981, *ANSWERS (Areal Nonpoint Source Watershed Environment Response Simulation) Users Manual*, EPA-905/9-82-001, US Environmental Protection Agency, Chicago.
- CARMICHAEL, D., 1990, GIS predictive modeling of prehistoric site distribution in central Montana. In K.M.S.Allen, S.W.Green and E.B.W.Zubrow (eds), *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis.
- CUSTER, J., 1989, *Prehistoric Cultures of the Delmarva Peninsula: An Archaeological Study*, Cranbury: Associated University Presses.
- DIRECTORATE OF SAFETY, HEALTH, AND ENVIRONMENT (DSHE), 1996, Geographic Information System, database, US Army, Aberdeen Proving Ground, MD.
- ENVIROSPHERE Co., 1988, *An Archaeological Overview and Management Plan, Aberdeen Proving Ground*, report to the National Park Service under contract CX4000-3-0018 for the US Army Materiel Development and Readiness Command, Philadelphia, PA.
- GOODWIN, R.C., & ASSOCIATES, INC., 1995, *Aberdeen Proving Ground Cultural Resource Management Plan*, prepared for US Army Corps of Engineers, Baltimore District, Baltimore, MD.
- HOFFECKER, J., 1997, Predicting impacts to prehistoric and historic sites with computer simulation models, Paper presented at the 62nd Annual Meeting of the Society for American Archaeology, Nashville, TN.
- KVAMME, K., 1985, Determining empirical relationships between the natural environment and prehistoric site locations: a hunter-gatherer example. In C.Carr (ed.) *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*, Kansas City: Westport Publishers, pp. 208-238.

- 1988, Development and testing of quantitative models. In W.J.Judge and L.Sebastian (eds) *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, US Department of the Interior, Bureau of Land Management Service Center, Denver, CO, pp. 325–428.
- 1992, A predictive site location model on the High Plains: an example with an independent test, *Plains Anthropologist*, **37**(138):19–40.
- MARYLAND HISTORICAL TRUST (MHT), 1995, Maryland archaeological site files and maps, Crownsville, MD.
- MASCHNER, H. and STEIN, J., 1995, Multivariate approaches to site location on the northwest coast, *Antiquity* **69**:61–73.
- PARKER, S., 1985, Predictive modeling of site settlement systems using multivariate logistics. In C.Carr (ed.) *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*, Kansas City: Westport Publishers, pp. 173–207.
- WARREN, R., 1990a, Predictive modeling in archaeology: a primer. In K.M.S.Allen, S.W.Green and E.B.W.Zubrow (eds) *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis, pp. 90–111.
- 1990b, Predictive modeling of archaeological site location: a case study in the Midwest. In K.M.S.Allen, S.W.Green and E.B.W.Zubrow (eds) *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis, pp. 201–215.
- WILKE, S. and THOMPSON, G., 1977, *Prehistoric Archaeological Resources in the Maryland Coastal Zone: A Management Overview*, Geo-Recon, Seattle, WA.

CHAPTER FIVE

Protecting Cultural Resources through Forest Management Planning in Ontario Using Archaeological Predictive Modeling

LUKE DALLA BONA

The Ontario Ministry of Natural Resources (OMNR) identifies and protects cultural resources through its forest management planning process. The OMNR is developing a predictive modeling methodology based upon weighted variables developed using a GIS in a deductive framework. This methodology, tested in stages from small areas to entire forest management units, has demonstrated a high level of predictive strength. Examples of models developed for use in Ontario's forests are described. The importance of presenting the model in a clear and understandable way to nonarchaeologists, such as land-use planners, so they know exactly how archaeological potential might affect proposed activities is also discussed.

5.1 INTRODUCTION

The Ontario Ministry of Natural Resources (OMNR) identifies and protects cultural resources through the forest management planning process. The problem that presents itself to cultural resource managers in northern Ontario, as well as much of the rest of the Canadian boreal forest, is one where the resources are known to exist, but their exact locations are unknown. So how do we manage a resource that we know to exist but we don't know where it is? In the late 1980s, OMNR identified archaeological predictive modeling as a means of addressing this situation and, given available knowledge, providing the best statement regarding the likelihood of archaeological resource existence. The OMNR sponsored three years of research and development that led to a first-generation predictive model. This was followed by three years of pilot projects which served to expand the applied base of the model from the original research and development area in northwestern Ontario and also to develop various means by which existing Ontario government digital databases can be incorporated into the archaeological predictive modeling process. The OMNR is now at a stage where it is ready to employ archaeological predictive models as a cultural resource

management tool in all new forest management plans—an area encompassing 45 million hectares of forested land.

5.2 BACKGROUND

In 1991, the Ontario Ministry of Natural Resources introduced the *Timber Management Guidelines for the Protection of Cultural Heritage Resources* (OMNR 1991). These guidelines outline the manner in which cultural heritage resources are protected through the forest management planning process. In addition to protecting known/verified archaeological sites, the guidelines explicitly state that areas determined to have a high potential for archaeological sites will also be protected. In 1995, the Ontario government legislated a new *Forest Management Planning Manual* (OMNR 1996) that changed the manner in which forests are managed in the province. Among the many changes was the introduction of guidelines for the protection of numerous *values* that previously were not formally considered in planning; values such as woodpecker habitat, impact on tourism, and protection of cultural heritage values, among others. While the protection of recorded archaeological sites has been a part of forest planning in Ontario for decades, it was never formalized and relied more upon the personal interest of the plan author or the ability of the regional provincial archaeologist to keep abreast of new forest management plans and schedules. It would be fair to say that, in spite of honest efforts prior to 1991, cultural heritage protection was a low priority in forest management planning.

The new *Forest Management Planning Manual* identifies cultural resources as one of many values that must be considered and protected in the planning process. There are seven steps outlined in the guidelines to be followed when identifying and protecting cultural resources (OMNR 1991:7–9; emphasis added):

- 1 Prepare a thematic overview of the heritage for the management unit...both the precontact and postcontact periods would be described.
- 2 Assemble known site databases for all four categories of heritage resources (i.e., cultural landscapes, structural remains, archaeological remains and traditional use sites).
- 3 *Apply and document appropriate site potential models for the management unit (or parts thereof). Assemble all relevant environmental and cultural data necessary to translate the models into maps showing areas of high potential for heritage resources.*
- 4 Rank the importance of the various types of known resources.
- 5 *Combine the maps of areas of high potential (Step 3) and of known sites (Step 2). The output of this step is the heritage component of the values map.*

- 6 Identify where the areas selected for operations during the 5 year term of the Plan coincide with heritage resource components of the values map. These coincident areas are the areas of concern for cultural heritage.
- 7 Identify a specific prescription for each cultural heritage area of concern.

In summary, not only is the Ontario government committed to using archaeological predictive modeling to protect cultural heritage resources, it is *required to do so*.

5.3 MODELING METHODOLOGY

The modeling methodology employed in OMNR's archaeological predictive models is a deductive model using the weighted value method. Kohler and Parker (1986: 432) see deductive models as encompassing three considerations:

A deductive model must:

- 1) consider how humans make choices concerning location.... This requires considering: (a) a *mechanism* for decision making; and (b) an *end* for decision making— what is the goal?;
- 2) specify the variables affecting location decisions for each significant chronological or functional subset of sites;
- 3) be capable of operationalization; it must propose a means for measuring each of the relevant variables and must allow for a set of predictions that can be compared with the archaeological data.

A number of interesting points can be raised when considering deductive models. For example, environmental variables are often considered by archaeologists to be important in conditioning the choice of activity location in the precontact past. Many predictive models, including this one, make the fundamental assumption that “settlement choices made by prehistoric peoples were strongly influenced by characteristics of the natural environment” (Warren 1990:202). This assumption figures prominently in determining which environmental characteristics or variables are used in the modeling process. An examination of the literature reveals some of the most basic environmental data used in predictive models: elevation, slope, aspect, and distance to water (Kvamme 1985; Parker 1985; Altschul 1990; Carmichael 1990; Warren 1990). However, most researchers recognize that a wide range of environmental considerations are important, including vegetation change over time as well as the use of various plants for medicinal purposes.

From the standpoint of human adaptation, patterns of local vegetation are of crucial concern. Many plants serve as primary food and technological

resources as well as secondary resources that attract economically important animals. The distribution of non-food resources, especially water and fuel, can be equally important to settlement decisions. Diversity is also beneficial when considering non-food resources. In addition to fuel, a variety of trees provide the raw materials for tools, utensils, shelter, and weapons, pitch for sealing seams, and fibers from the inner bark for cordage, bags and nets. A variety of plants can be used to make dyes, reeds can be woven into mats, and clay from local stream banks can be made into pottery. Evaluations of topography, water, soils, vegetation, precipitation, temperature, and availability of rock outcrops or glacial till exposures are all important in decisions about the adequacy of shelter and the availability of economic resources.

(Schermer and Tiffany 1985:220)

Dean (1983:11) has pointed out that people may search only for a few cues in their surroundings when identifying and selecting activity locations, rather than processing the entire range of environmental “cues” available. It may be only these basic variables that really have any association with archaeological sites. This raises interesting questions about the analysts’ choice of the proper environmental variables for inclusion in the modeling process:

Perhaps in building predictive models we are too ready to make the assumption that only a complex multivariate model can adequately account for human locational behavior, when in fact, a few (proxy?) variables, observed in the highly correlated database that is our environment, may be sufficient for forming locational decisions.

(Kohler and Parker 1986:433)

Support for this position lies in the fact that archaeologists have presented successful predictive models using very few variables. For example, Altschul (1990) developed a predictive model for the 9000-acre (3640-ha) Mount Trumbell area of Arizona. There were 228 known sites in the study area that had been sampled by various agencies in the past. Three environmental variables were identified which account for the majority of site locations: elevation, slope, and aspect (Altschul 1990:229–30). Altschul concluded that in this area “over 70 per cent of all component locations can be predicted with just three variables” (1990:234). However, he does not discuss what his three variables are measuring. What are they “proxy” variables for? Without this information, one is unable to discuss why sites are being found where they are, nor can explanations be offered for settlement systems in the area.

The weighted value method employed within a deductive framework begins with the assumption that each landscape variable contributed differently to ancient land-use decision-making. To account for this, each landscape variable is given a different numeric weight to reflect its assumed contribution to potential.

This is an arbitrary weighting scale and might offer a range for 0 to 5 where 0=no contribution to potential and 5=highest contribution to potential. The variables to be modeled are divided into categories and subcategories. Categories encompass broadly defined divisions such as *Proximity to Water, Soil, or Drainage*. Subcategories encompass detailed subdivisions of categories. For example, if the category was *Landform*, the subcategories might be *Moraine*, and the variables could be *End Moraine, Ground Moraine* and *Hummocky Moraine*.

A value V is applied by the researcher to the category to reflect its importance and contribution to the overall determination of potential. In addition, variables are assigned weights, W , to reflect differences within categories, again reflecting their contribution to potential. By multiplying the category value by the weight of the variable ($V \times W$), a weighted value is defined for each variable used in the modeling process (Table 5.1). The determination of the numerical weight and value scale is purely subjective, but there must be a basis upon which the researcher makes these numerical assignments. Reference may be made to previous archaeological work which has identified characteristics of the landscape presumed to be associated with archaeological sites. Ethnographic, ethnological, historical, or ethnoarchaeological studies may also be sources upon which the basis for weighting of variables is derived. The experience of the archaeologist, colleagues, and even the interested public, working or having experience in the area, may also contribute to determining a weighting scheme. Additionally, the nature of the project itself may have some bearing on the weighting applied to variables. For example, a researcher applying a predictive model within a given theoretical framework may give more importance to economic-related variables than to some geographic variables. In another instance, a researcher may combine their own experience with data obtained from the ethnographic literature and derive weights and values accordingly. In conclusion, the manner in which weights and values are applied is subjective, yet it is based upon data obtained and

Table 5.1 Example of assigned weights and values.

Category	Subcategory	Variable	Value	Weight	Weighted value
Drainage		Dry	4	5	20
Mixed wet/dry	4	3	12		
Wet	4	1	4		
Landform	Glaciofluvial	Delta	3	5	15
Esker	3	5	15		
Kame	3	3	9		
Outwash plain	3	4	12		
Raised beach	3	5	15		

Weights and variables are ranked on a scale of 0–5 with 5 having the highest contribution to potential.

evaluated by the researcher from a variety of sources and applied within project-specific frames of reference.

Because the weighted value method allows certain variables to have more “predictive strength” than other variables, it results in a model that better reflects the decisions made by people when they choose their activity locations. In addition, because it is imperative to clearly outline the manner in which the categories and variables are weighted, the contribution of each variable to the final model of archaeological potential is also clearly established. This final point is perhaps the greatest strength of this methodology. For any model to be valid, it must be reproducible and defensible. When the weighting factor of each variable is clearly defined, discussion can occur concerning the weights of individual variables, and the effects of changing weights and values can be tested. The results of these tests can then be evaluated. In the end, one has a model for precontact activity location that is clearly defined, testable, and reproducible.

5.4

MODEL RESEARCH AND DEVELOPMENT

Between July 1991 and March 1994, the OMNR and Lakehead University, through a memorandum of understanding, undertook research into the development of archaeological predictive models for use in forest management planning. This research took place through the Centre for Archaeological Resource Prediction (CARP) at Lakehead University in Thunder Bay, Ontario, and was based on a methodology developed by Dalla Bona (1993) for the northern Plains in Saskatchewan, Canada. The project was geared to answer the question, “Can archaeological predictive modeling be done in the boreal forest of northern Ontario?” Research was carried out in the areas of contemporary/historical boreal forest land use, predictive modeling history/methodology, and boreal forest archaeology.

A component of the research and development included archaeological field surveys that were carried out to collect baseline archaeological data and to provide initial indications of the predictive success of the archaeological predictive models. More than 50 km² of boreal forest were surveyed over two summers at as close to 100% coverage as possible. The purpose of the surveys was to return information about site density and distribution in those areas where forest harvesting activities were being carried out—precisely those areas where our archaeological knowledge was weakest. These initial, tentative surveys demonstrated that archaeological sites do exist in areas subjected to forest harvesting activities—perhaps not in the densities encountered elsewhere in the bush (i.e., along the shores of larger lakes and rivers) but more or less conforming to our understanding of boreal forest land use.

The results of this work were presented to the OMNR in March 1994 in six volumes that detailed the work completed and outlined a prototype predictive modeling methodology (Dalla Bona 1994a, b, c; Hamilton and Larcombe 1994;

Larcombe 1994; Hamilton *et al.* 1994). It is beyond the scope of this chapter to summarize these reports. However, it is important to note that the modeling results presented here are based upon models generated using the methodology described in Volume 4 of the CARP Reports (Dalla Bona 1994b).

5.5 PILOT PROJECTS

Upon completion of the research and development work, it was decided that additional work should be carried out to evaluate the reliability of this research product in an applied environment. A three-stage process was established to bring the predictive model from the realm of research and development into the realm of forest management planning: Stage 1 involved small-scale application of the model in an actual forest management plan; Stage 2 involved large-scale application of the model in an entire management unit, establishing the parameters it would be required to operate within the forest management planning process; and Stage 3 involved applying the model to a substantial area rich in cultural resource data to determine the correspondence of high-potential areas with verified sites.

5.5.1 Stage 1: small-scale applications

In January 1994, two management units were identified by the OMNR as being suitable for a pilot-project application of an archaeological predictive model as they were entering the first stages of the timber management planning cycle. The two management units identified were the Geraldton Management Unit and the Dog River/ Matawin Management Unit (Figure 5.1). The full pilot-project results are reported in Dalla Bona (1995). For reasons of brevity, only the Geraldton Management Unit pilot project will be discussed here.

Geraldton Management Unit

The Burrows Lake study area is located approximately 50 km north of Geraldton, Ontario. It is centered on Burrows Lake, a tributary of the Kenogami River, which meets the Albany River at the Albany Forks. Burrows Lake has three navigable waterways flowing into it, which exit from the lake eastward through Burrows River. The Burrows River in turn empties into the Kenogami River. Burrows River and False Creek flow into Burrows Lake from the south, and Murky Creek, which drains Poilu Lake and Arm Lake, from the west. Burrows Lake is therefore easily accessible by water.

All major rock types in the Burrows Lake area are early Precambrian in origin. The surficial geology of a major portion of the study area consists of lacustrine deposits of clay and silt with some sands. The remainder of the area is

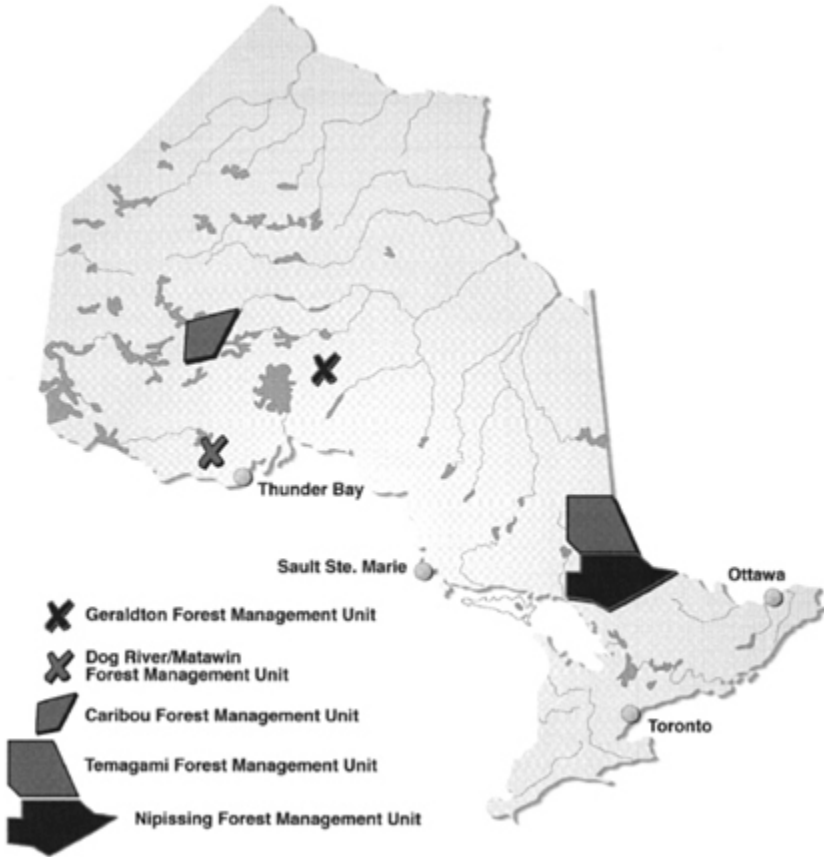


Figure 5.1 Location of predictive modeling applications discussed in this chapter.

ground moraine of silty to sandy till. An end moraine of sand, gravel, and boulders occurs east of Poilu Lake (Feldbruegge 1979:7). Burrows Lake has low, sloping, sandy to clayey banks on the eastern side of the lake while the northern and western sides have considerable bedrock exposures, boulders, and cobbles. Also of interest on Burrows Lake are the large quantities of chert pebbles that litter the beaches. While the majority of these pebbles may be too small for use in tool-making, it may be supposed that larger pebbles/cobbles existed in the past and were used for such purposes. The tree species in the Burrows Lake area consist mainly of white and black spruce, jackpine, poplar, white birch, and cedar (Feldbruegge 1979:7). Additionally, the different dates of harvesting through the past have resulted in a variety in the composition and age of forest stands.

The model was applied to a 148.2 km² area. This area is represented in the GIS database by a map 438×376 cells (164 688 cells total—134 001 when not counting cells identified as water), where each cell represents 28.57 m×28.57 m.

The information was derived from 1:50,000 topographic maps and 1:100,000 surficial geology maps. The following map layers were generated: digital elevation model, first-through fifth-order water, aspect, drainage, grade, landform, soils, topography, and archaeological sites. The predictive modeling methodology was applied to the digital map layers to produce a visual possibility statement—that is, a map indicating the potential for the existence of archaeological sites (Figure 5.2).

In the summer of 1979, an archaeological survey was conducted in the general Burrows Lake area that resulted in the discovery of 36 sites (Hems 1980), 22 of which fall within the pilot project study area. Hems identified 40 separate components at the 36 sites (Hems 1980: table 5.2). Most of the sites could not be assigned a cultural affiliation due to the lack of diagnostic material recovered. None of the sites recorded showed any evidence of having Paleo-Indian or Archaic components, while two Initial Woodland and six Terminal Woodland components were identified. One site was characterized by undifferentiated Woodland and four others had a Historic component. Two recent cabins were also recorded as components, as were 25 of unknown precontact affiliation. Very little material was found on the surface except on a large peninsula that divides the southern portion of Burrows Lake into two arms. This peninsula had recently been burned over and this had exposed large areas, which greatly facilitated surface collection. Nearly all the other sites were found in a relatively undisturbed context and were located and delineated through shovel testing (Hems 1980:174).

A second archaeological survey was conducted in 1994 and 14 new sites were identified (Dalla Bona 1995). Thus, the total archaeological site sample for the Burrows Lake study area is 37. During the 1994 surveys, approximately 60% of survey time was spent examining areas identified as being of high archaeological potential and the remaining 40% was spent examining areas that were identified as “not having high archaeological potential” (i.e., all other areas). While this is not a statistically valid sampling program, a deliberate effort was made not to focus solely on areas of high potential, including areas of “not high potential” that are removed from shorelines. Although the site database was not generated through the statistically-study principles, the fact that 37 sites exist cannot be ignored.

The 37 archaeological sites identified during the two surveys (1979 and 1994) are represented by 48 grid-cell locations in the digital database. Some sites fall across more than one grid cell, while the smallest sites are represented by only one cell, equivalent to an area of 816.24 m². The combined area covered by all the archaeological sites accounts for 0.03% of the total land base of the study area.

A one-sample Kolmogorov test (Kvamme 1990:373) was used to compare the observed pattern of weighted values at site locations against the pattern of weighted values in the background (expected). The background population is, in effect, the entire study area and represents the number of sites in each weighted

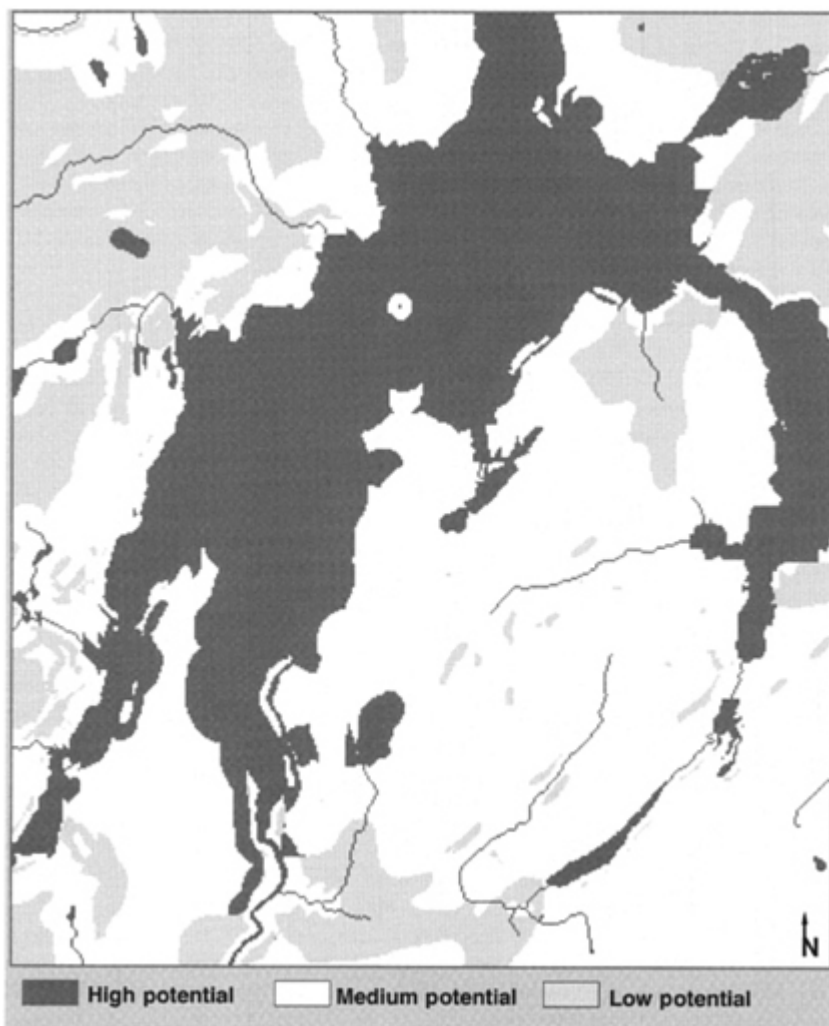


Figure 5.2 Visual possibility statement expressing archaeological potential for the Burrows Lake model test area, Geraldton Management Unit, Ontario, Canada.

value category. The null hypothesis states that there is no difference between the cumulative frequency distribution of the background and the site sample, thereby indicating that the occurrence of weighted values at sites locations is random. We can reject the null hypothesis if the maximum difference between the distributions exceeds a critical value. Accordingly, to reach significance at the 0.001 level, a value of D at least as large as $1.95/\sqrt{n}$ is required. A minimum value of $D=0.3205$ at the 0.001 level of significance results. Cumulative relative frequencies for both the background and the sample were computed and the

largest difference between the two is 0.7671, which exceeds the minimum expected value of 0.3205. We can therefore reject the null hypothesis of no difference between the background and the sample (site location) distributions and state with confidence that there is some difference between them.

An examination of the data provides illumination into patterns of association between site location and the distribution of cells across weighted values (Table 5.2). It is evident that if those areas where cells with weighted value equal to 89 occur, 91.67% of site cells (89.19% of sites) would have been identified. Those weighted values are represented at only 12.48% of the study area. Correspondingly, very few sites are found on cells with lower-weighted values. In fact, only 8.33% of all remaining site cells (10.81% of sites) are found on weighted values equal to 88. This relationship is demonstrated graphically in Figure 5.3. As the percentage of background cells increases, the percentage of site cells remains low, until weighted value 88. At this point, 87.52% of the land base has been counted, compared to only 10.81% of all sites.

With this information in hand, it is possible to refine categories of potential. It will be noted that an *a priori* assumption defined high-potential areas as those falling within the upper 10% of the scale of potential. From the cumulative frequency table (summarized in Table 5.2), three separate categories of potential become apparent. The zone of low potential is defined by weighted values 29 through 68. Together, they comprise 26.38% of the total land base but include no known archaeological sites. The zone of medium potential is defined by weighted values 69 through 88. These cells comprise 61.14% of the total land base but include only 10.81% of known archaeological sites. The zone of high potential is defined by weighted values 89 through 128 and comprises only 12.48% of the land base but includes 89.19% of known archaeological sites. There is a clear inverse relationship between the distribution of weighted value cells and archaeological sites/site cells. While there is a high percentage of the study area falling within zones of low and medium potential, a low percentage of known sites fall in those zones. Conversely, while a small percentage of the study area is represented as having high potential, a high percentage of known sites fall in that zone. Therefore it was concluded, at this early stage in the model evaluation process, that the application of the model in the Burrows Lake study area produced results which appeared to predict reasonably well where one could expect to encounter sites.

Table 5.2 Summary of cumulative frequency table—weighted values at site locations compared against weighted values in the entire background referent.

Zone of potential	Weighted values	No. of background cells	No. of Sites	% Background total	% Sites total
High	89–128	16722	33	12.48	89.19
Medium	69–88	81933	4	61.14	10.81

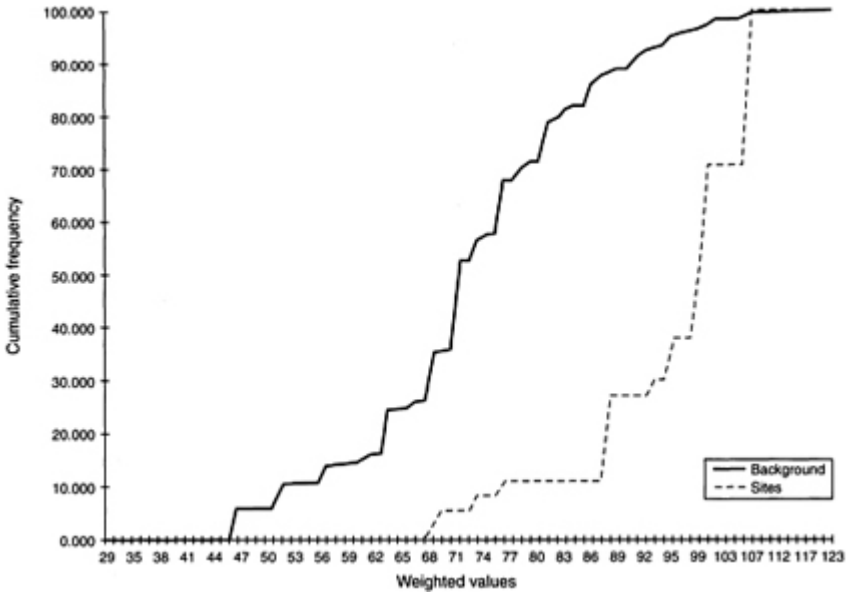


Figure 5.3 Comparison of percentage cumulative frequencies of the background and the site sample, Burrows Lake study area.

Zone of potential	Weighted values	No. of background cells	No. of Sites	% Background total	% Sites total
Low	29–68	35346	0	26.38	0

5.5.2

Stage 2: management unit-scale application

The Caribou Forest Management Unit (Figure 5.1) was selected for application of the archaeological predictive modeling methodology for two reasons. First, this unit was acting as a “test unit” for the application of the new forest management planning manual (OMNR 1996). A number of new guidelines and approaches that arose from the new manual were being applied for the first time—including the cultural heritage guidelines, in which archaeological predictive modeling was but one component. Second, this unit had seen minimal forest harvesting activities, when compared with units further to the south, and, coupled with the fact that it lies entirely within the boreal forest zone, made this unit well suited for archaeological study within the context of forest management planning in northern Ontario.

The physical geography of this 800,000-hectare forest management unit is characterized by Precambrian bedrock that exhibits numerous faults resulting in long, thin lakes extending for kilometers. There is considerable evidence for glacial activity throughout this management unit. Numerous sinuous eskers wind their way in lines, primarily from the northeast to the southwest. Ancient beach ridges dot the landscape in the southern reaches of the unit. Outwash plains, drumlins, deltas, and other features can be found throughout the study area. The unit is primarily vegetated by a conifer forest. This unit is at the northern limits of red pine, but jackpine and spruce are common, with aspen, poplar and birch occurring as well.

The archaeological predictive modeling methodology applied was essentially that developed by CARP (Dalla Bona 1994b) with some exceptions. First, the amount of existing digital information available for that management unit was extremely limited. While digital elevation models exist for most of the province, they did not exist for the Caribou Unit at the time (1995). As a result, I was faced with two choices: create the data myself or contract the work to an outside agency. The first option was possible but not within the time frame available. The second option was not possible due to the enormous cost involved (>\$50,000) and the unreasonable level of resolution (50–100 m per pixel). Therefore, no digital elevation model was used in the application of this predictive model, which effectively means that two variables could not be incorporated into the model: aspect, which reflects the compass direction parcels of land face; and slope, which reflects ground steepness. The latter variable is encompassed in part by the variable topography, which measures overall relief in a broad area but not at the level of detail reflected by that derived from elevation values.

An addition to the variables identified in the model developed by the CARP research and development project is the “portage” variable. Portages identified on OMNR values maps, National Topographic Series (NTS) maps, and NOGETS geology maps are digitized onto a map layer and weighted accordingly. This data layer complements rapids/waterfall data but strictly identifies one side of a water body or the other as having higher potential. In cases where portages go overland, this variable adds another dimension to identifying archaeological potential in areas away from lakes and rivers.

The GIS data for the Caribou Forest Management Unit was digitized by myself from 1:50,000 NTS topographic maps and 1:100,000 NOGETS surficial geology maps at an effective scale of 1:85,000 (resolution=29.4 m/cell) following the methodology detailed in Dalla Bona (1994b). The resulting digital database is 3,649 rows by 2993 columns, totaling 10,921,457 cells, of which 3,596,495 fall outside of the management unit boundaries. A total of 7,324,960 cells occur within the unit’s boundaries: 5,477,506 (74%) are classed as land; 1,597,493 (21.81%) are classed as water; and 249,961 (3.41%) are north of latitude 50° (where geological map data did not exist). Therefore, archaeological potential could not be generated for this area (north of 50° latitude) at that time (1995). The following digital map layers were generated: soils, landform,

topography, drainage, portages, rapids/waterfalls, first—through fifth-order water bodies.

The predictive modeling methodology was applied to the digital map layers to produce a visual possibility statement identifying archaeological potential (Figure 5.4). For the Caribou Management Unit, archaeological potential ranges from the lowest potential, where the numerical value equals 5, to the highest potential, where the numerical value equals 149. Following from the results of the pilot applications in the Stage 1 projects, archaeological potential is identified by the upper 15% of the landbase representing the highest weighted values.

Prior to 1995, there were no known archaeological sites for the 800,000 ha that comprise the Caribou Forest Management Unit. After two years of survey (Dalla Bona 1996, 1997), 23 archaeological sites were discovered in various parts of the unit, representing everything from the very earliest occupations (radiocarbon dated to 8100 BP; Pilon 1998) through to turn-of-the-century historic occupations. The surveys were not conducted according to the principles of random sampling theory, nor were prejudgments made about surveying in areas of high or not-high potential. Rather, these surveys were conducted to obtain a basic understanding of the density and distribution of archaeological sites in this part of the bush. Although not statistically valid, 22 of 23 sites were discovered to be in areas identified as having high archaeological potential.

The distribution of potential

The Caribou Forest Management Unit provided excellent information about the distribution of potential across the landscape. A common public perception of archaeological site location in northern Ontario is one in which sites are located adjacent to lakes and water bodies. This perception is transferred to an understanding of potential when it is surmised that protecting shorelines will also protect all the archaeological sites. An archaeological understanding of site distribution places a high density of sites along the shores of major lakes and rivers. We understand this to be the result of summer occupations, which are of longer duration and with a greater population concentration than occurs at other times of the year. However, that same understanding of land use also suggests that a lower-density but wider distribution of sites may be expected in areas high up in watersheds, well removed from the shores of major lakes and rivers (Larcombe 1994; Dalla Bona and Larcombe 1996).

The nonarchaeological-predictive-modeling-specialist's perception of archaeological potential as being highest along the shores of major lakes and rivers is not necessarily correct. The modeling methodology employed here clearly demonstrates that the distribution and "shape" of archaeological potential is as varied as the landscape itself (Figure 5.4). This is reflective of the power and utility of this methodology. When people consider where to set up a campsite, they do not paint the entire shore of a lake with the same stroke of the brush. There are selected shores of the lake that are preferred and there are other parts where it

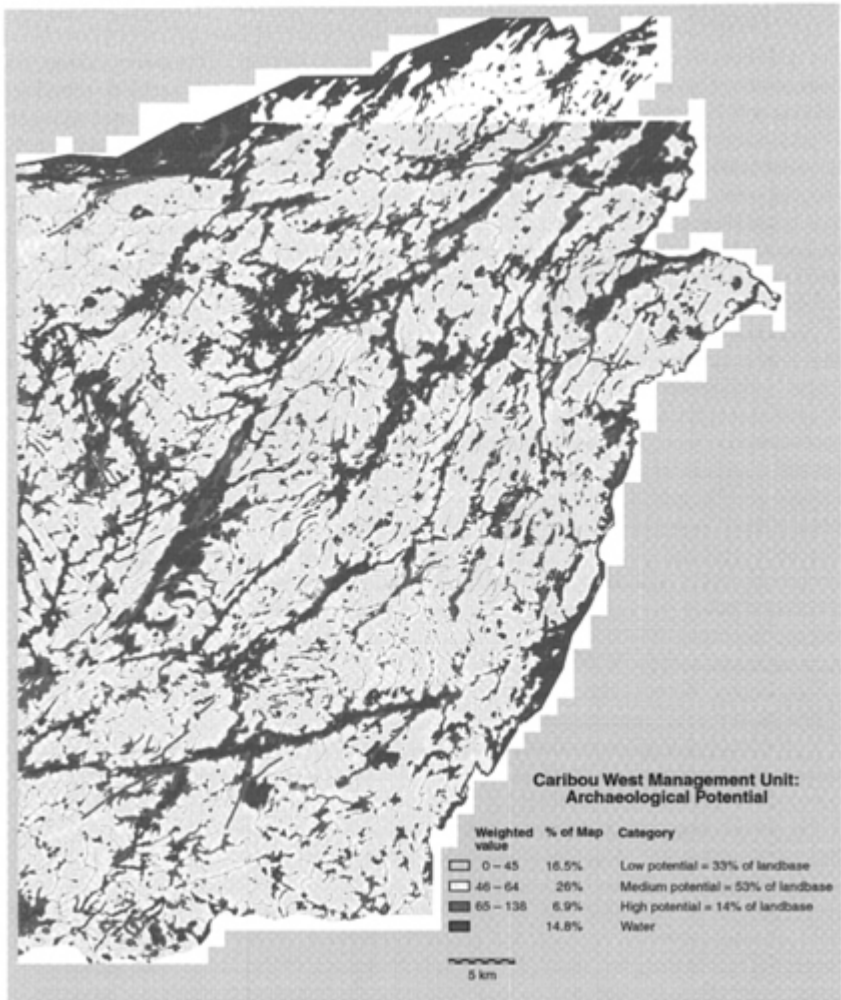


Figure 5.4 Archaeological predictive model generated for the Caribou Forest Management Unit, northwestern Ontario, Canada, classified into three zones—high, medium and low potential.

might not be physically possible to set up a camp (e.g. where there are cliffs or steep slopes). Three areas of archaeological potential, which correspond to aerial photographs of regions included in the archaeological survey of 1995, exemplify this point.

The Ragged Wood Lake area is illustrated in [Figure 5.5](#), and was the starting point of one of the 1995 surveys. Immediately apparent to the viewer is the large swath of high potential that loops through the lower center portion of the image area. This feature represents a raised beach identified from the geology maps and

is weighted high due to the expectation of identifying associated archaeological material. Examination of the image suggests that the highest potential for the existence of archaeological sites is on land south of the lake that intersects with the raised beach feature. Around the lake itself, archaeological potential is not uniformly distributed. Indeed, the northwest margins of the lake are of considerably lower potential than the northeast shore directly opposite. Clearly, the greatest influence on archaeological potential is the glacial beach that runs through the southern portion of the image. The only archaeological site (EcJt-1) identified in this image area is found on the point of land east of the small island near the center of the lake (Dalla Bona 1996).

A second area of aerial coverage is in an area south of Marchington River where harvesting has already taken place (Figure 5.6). This image is primarily characterized by numerous eskers running through the area, trending northeast to southwest. Archaeological potential in this image is distributed such that the areas with the highest potential occur considerable distances from the major water body, the Marchington River, which flows west across the top third of the image. In this area, the eskers clearly have the greatest influence on the distribution of potential, and the Marchington River, while a major contributor to potential itself, is not ringed in high potential. Rather, high potential crisscrosses through areas already harvested, well away from most of the water bodies.

A third area of aerial coverage is the central Fairchild Lake region (Figure 5.7). In this area, archaeological potential is distributed more in line with popular expectation. Highest potential is located along the margins of the lake and major creeks, and falls off as distance increases from the lake. There are no exceptional geological or glacial features in this area—a factor contributing to the fairly uniform distribution of potential. Two pictograph sites are located within the image area, on the south side of the small peninsula at the left center of the image. Two other sites are located outside the image: a third pictograph to the north and a small lithic scatter to the south.

The three examples above demonstrate that all potential is not distributed equally. Indeed, this is something that any archaeologist who has ever done any fieldwork already knows. The whole of the shoreline around a lake is not equally likely to contain a site. One side of a lake may be more favorable than the other and there are certainly going to be selected areas that are more likely to contain archaeological sites. This modeling methodology really does make explicit an attempt to quantify the subjective experience of archaeologists. The contributors to variation in potential are the innumerable combinations among the variables defining the model. Where there is little variation, potential is uniformly distributed; where there is great variation, potential appears distributed on the map like a pack of crayons melted in the sun. A model that reflects our understanding as professionals has some use. If it is accepted that we as professionals have a reasonable grasp on understanding the past, then we might find such a model useful.

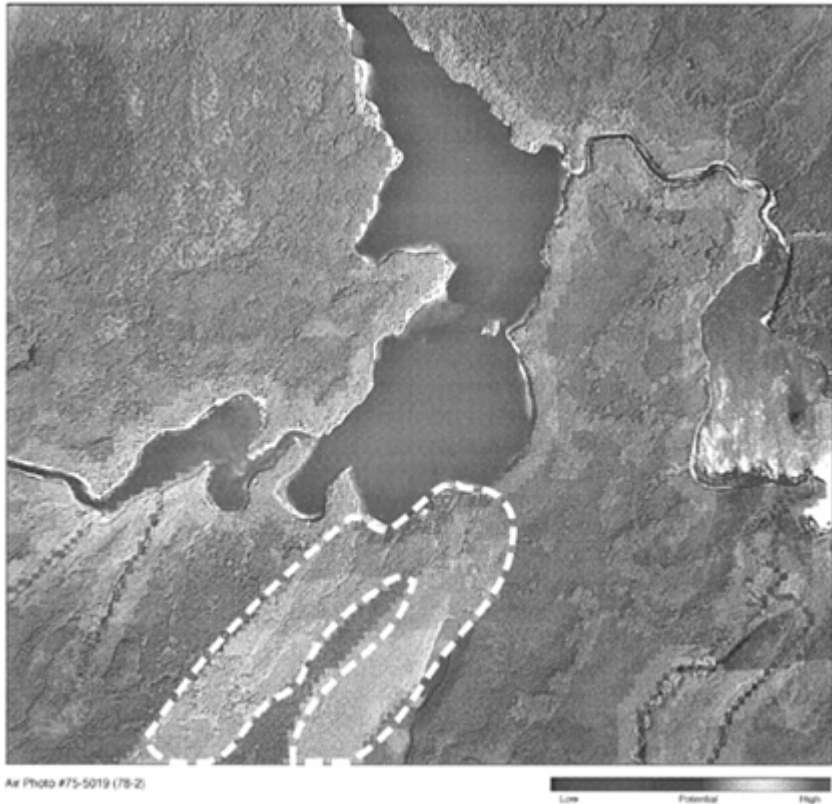


Figure 5.5 Southern Ragged Wood Lake, Caribou Forest Management Unit. Archaeological potential is superimposed onto an aerial photograph (Air Photo no. 75–5019/78–2). The highest archaeological potential occurs in the bottom center of the image, where the influence of a glacial beach ridge is clear.

5.5.3

Stage 3: application within forest management plans

The Temagami and Nipissing Forest Management Units in Ontario represent unique challenges to predictive modeling (Figure 5.1). First, these areas are well represented digitally with only detailed geology data lacking. More importantly, this area represents perhaps the best cultural heritage database in northern Ontario. Over a decade of study, resulting from intense scrutiny of land-use practices in these areas, has resulted in an excellent database that spans historical and archaeological categories. A full range of site types are documented, from simple one-flake wonders, to precontact native village sites, to spiritual sites with no physical evidence betraying their presence. Indeed, heritage sites, representing the rich mining and forest heritage of this area occupied early in the twentieth century, are also documented.

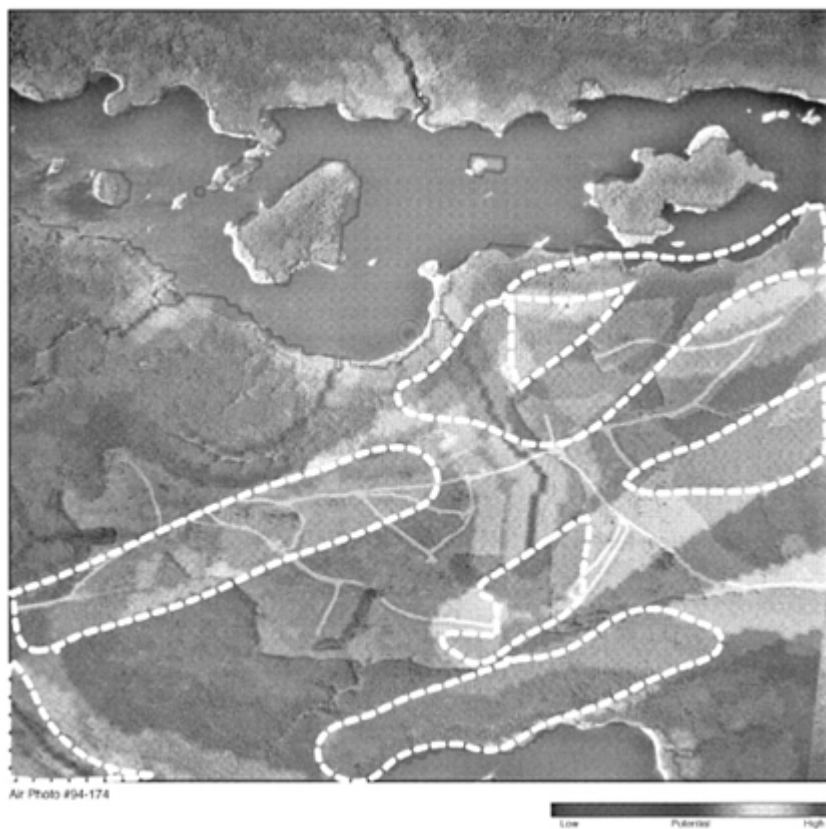


Figure 5.6 Marchington River Locality, Caribou Forest Management Unit. Archaeological potential is superimposed onto an aerial photograph (Air Photo no. 94–174). The highest archaeological potential occurs through the middle of the image where a series of eskers run northeast to southwest through an area already harvested.

There were two primary reasons for applying the model in this area. First, it represents an unparalleled opportunity to evaluate the correspondence of areas predicted to contain archaeological sites with a large database of verified archaeological sites. Second, planners in this area were in the initial stages of writing a new forest management plan and the opportunity existed to apply the methodology and incorporate the results into an actual forest management plan—and follow it through its five-year cycle.

Together, the Temagami and Nipissing Units cover an area of some 24,275 km² and stretch from just north of Algonquin Park, in the heart of the mixed-wood forest, north to the base of the Hudson Bay Lowland clay belt, through the boreal forest. While primarily a Precambrian bedrock environment, this study area is represented by a full range of postglacial features including dramatic glacial spillways, dunes, clay plains, and some of the highest elevations in Ontario.

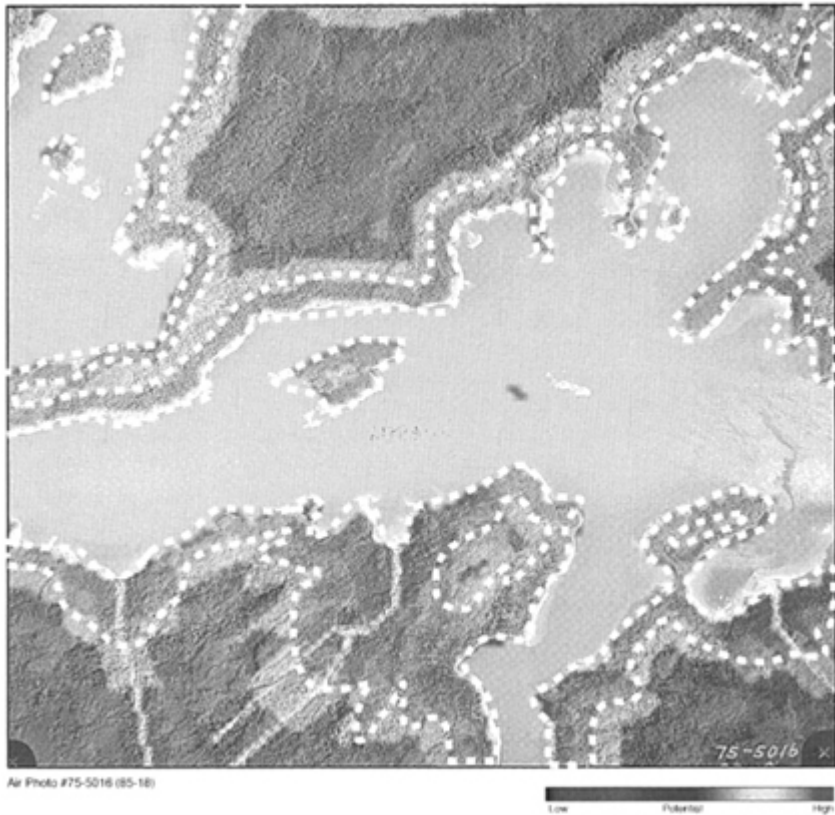


Figure 5.7 Fairchild Lake Locale, Caribou Forest Management Unit. Archaeological potential is superimposed onto an aerial photograph (Air Photo no. 75–5016/85–18). High potential is uniformly distributed around the lake except in the bottom center, where there is a large concentration away from the lake.

The predictive modeling methodology was applied to this area, resulting in a raster database 8000×6667 , for a total of 53,336,000 cells at 30 m resolution. Of that total, 26.3 million cells fall outside the management unit boundaries, 5.25 million cells are water and the remaining 21.7 million cells form the land base. The same methodology as described in Dalla Bona (1994b) was applied but again with the exclusion of elevation-related data. While there was digital terrain information for parts of the unit, it did not exist for the complete unit, and as a result it was not used. Consultation with John Pollock, a local archaeologist with considerable experience of working in the region, assisted in “tweaking” the weighting scheme applied to the variables. Perhaps the greatest difference between the original weighting (Dalla Bona 1994) and the weighting applied in the Temagami/Nipissing area was the increased importance of bedrock-related variables and the decrease of the importance of sandier soils. In the final

computations required for creating the model, ten different map layers (totaling one-half billion cells of information) were simultaneously juggled to produce the final visual possibility statement representing a map of archaeological potential (Figures 5.8 and 5.9). High potential was defined as 16.25% of the land base.

This map of potential was compared against the known site database. Table 5.3 summarizes the performance of the model. For the initial evaluation of model performance, and in the absence of full site records (still to be obtained from the provincial record-keeping agency), a total of 222 registered precontact archaeological sites were used as the comparative database. Of these, 83.8% (186/222) of the sites were found in areas identified as being of high archaeological potential, 15.8% (35/222) of the sites were found in areas identified as being of medium archaeological potential, and 1 site was located in an area of low archaeological potential. Two different statistical tests were conducted to evaluate the robustness of these predictions. First of all, it should be stated that the existing archaeological site database was not used to generate the model. Second, the archaeological data was not collected via statistically valid sampling methodology.

The first test conducted was Kvamme's (1990) one-sample Kolmogorov test. Once again, this test measures a sample of the background against the entire background. The sample in this case is the weighted values as they occur at known site locations. The background is every weighted value in every cell in the entire database. The null

Table 5.3 Summary of cumulative frequency table—weighted values at site locations compared against weighted values in the entire background referent—Temagami and Nipissing Forest Management Units

Zone of potential	Weighted values	No. of background cells ($n=2271078$ 5)	No. of sites ($n=222$)	% Background total	% Sites total
High	76–147	3 691 475	186	16.25	83.78
Medium	55–75	14696318	35	61.14	15.77
Low	29–68	15232258	1	26.38	0.5

hypothesis states that there is no difference between the cumulative frequency distribution of the background and that of the site sample, thereby indicating that the occurrence of weighted values at sites locations is random. We can reject the null hypothesis if the maximum difference between the distributions exceeds a critical value, which is computed to be $D=0.1308$ at the 0.001 level of significance. Cumulative relative frequencies for both the background and the sample were computed and the largest difference between the two is 0.6769, which exceeds the expected value of 0.1308. We can therefore reject the null hypothesis of no difference between the background and the sample (site

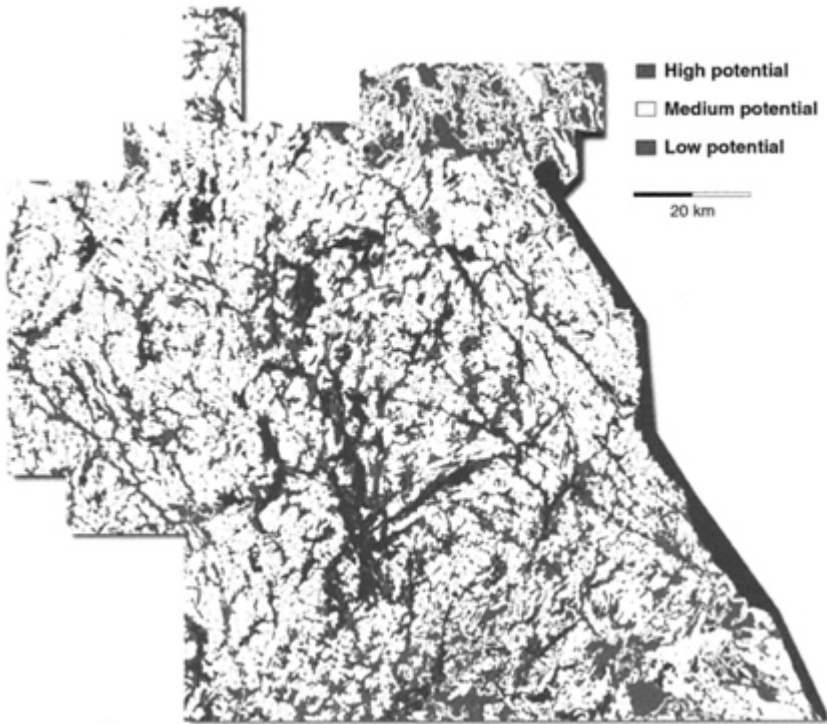


Figure 5.8 Predictive model of archaeological site location for the Temagami Forest Management Unit, Ontario, Canada.

location) distributions and state with confidence that there is some difference between them. We can examine this relationship graphically and visualize the dramatic separation between the two (Figure 5.10).

A second test conducted was a two-sample Kolmogorov-Smirnov test (Blalock 1979:266). In this case, rather than compare a sample of weighted values taken at site locations against all weighted values, the test compares two samples drawn from the population: sample 1 is weighted values taken at site locations, and sample 2 is a random sample taken from the background. The null hypothesis in this case states that to draw two independent random samples from identical populations results in two essentially similar cumulative frequency distributions. The K-S test produces a statistic that is the maximum difference between the two cumulative distributions. “If the maximum difference is larger than would be expected by chance under the null hypothesis, this means that the gap between the distributions has become so large that we decide to reject the hypothesis” (Blalock 1979:267).

As stated above, sample 1 is weighted values drawn from known site locations ($n=222$) and sample 2 is drawn randomly from the entire background of weighted values ($n=222$). The critical value, D , is computed to be $1.95/\sqrt{[(222$



Figure 5.9 Predictive model of archaeological site location for the Nipissing Forest Management Unit, Ontario, Canada.

$+222 / (222 \times 222) = 0.1851$ at the 0.001 significance level. Cumulative frequencies were computed for both samples and shown in graph form (Figure 5.11), and the largest difference between the two was 0.6669, significantly exceeding the expected value. Thus, the null hypothesis could be rejected, indicating that there is a significant difference between the distribution of weighted values in the background as against those at site locations.

In summary, a predictive model applied to two forest management units in northern Ontario appears to have a reasonable level of predictive reliability. When compared against a known-site database of 222 sites, almost 84% of the known sites fall within areas identified as having high archaeological potential (16.25% of the land base).

5.6 OIL AND WATER CAN MIX! INTEGRATING ARCHAEOLOGY INTO FOREST MANAGEMENT PLANNING

The preparation of policy statements and guidelines is a necessary step in the development of appropriate cultural heritage protection in any development environment. However, there is a big difference between presumed means of

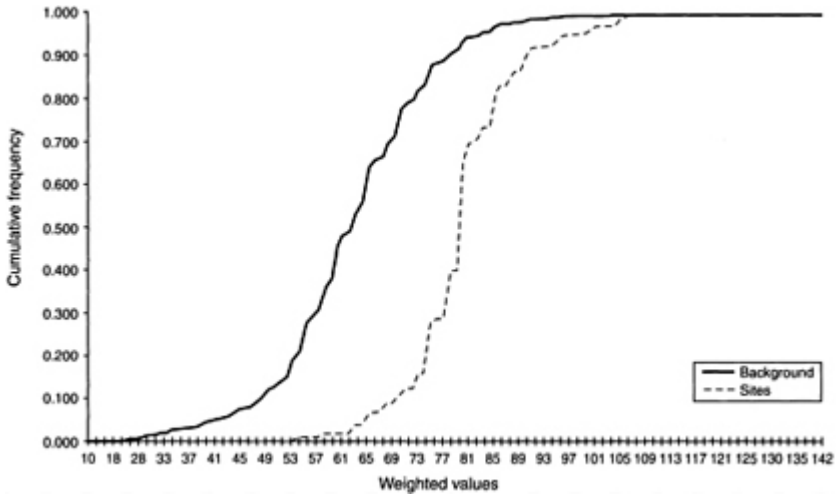


Figure 5.10 Cumulative frequencies of entire background referent versus site sample.

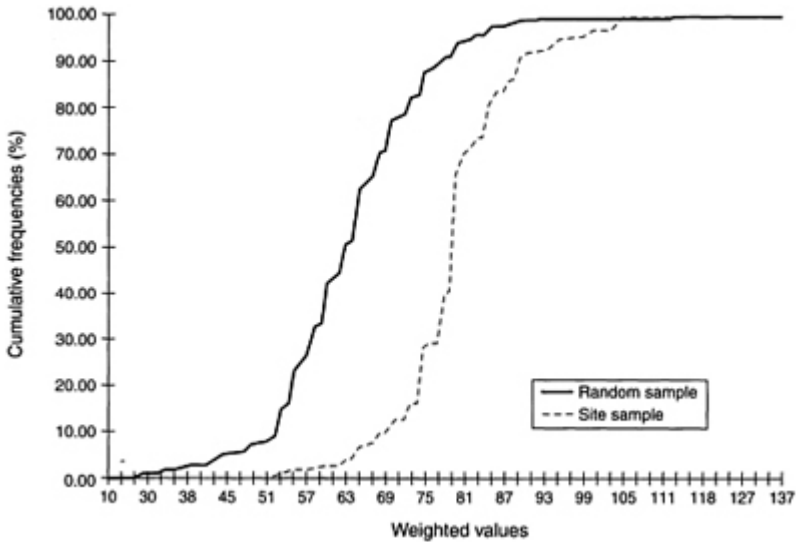


Figure 5.11 Cumulative frequencies of the random background sample versus site sample.

protection, arrived at through committee discussion, and actually implementing that protection in an operational setting. There are considerable hurdles to be cleared in getting an independent forest contractor to effect protection for a resource that we-think-might-be-there-but-aren't-really-sure-because-no-one-has-looked-for-it.

Perhaps the biggest hurdle facing archaeologists dealing with archaeological predictive models is the means by which the model is operationalized.

Generating a reasonable model result is a task in itself, but it must be remembered that a map of potential or a predictive model is not a final result; it is just a map, and all maps require interpretation and study. It is the product of a tool that allows us to make decisions regarding the appropriate types of activities that can be carried out in certain areas. While predictive models certainly have the capability to be used in theoretical research and purely academic exercises, we must recognize honestly that foresters and land-use planners are too busy to worry about the details of patch theory or biomass potential. They want to know how archaeological potential affects their proposed activities and we have to be able to interpret and present archaeological potential to people other than archaeologists in a manner that not only is understandable in an operational sense, but can be defended and justified when questioned.

For example, presenting a map of archaeological potential to a forest company will immediately elicit the question "What do I do with this now?" All users of the land base, be they developers or harvesters or cottage-lot owners, need to be instructed as to how they are being affected by the identification of a cultural heritage value. If high potential is identified as a value in a forest management unit, and that value has been identified as coming into conflict with a proposed activity, people want to know what they should do. Should a road be moved? Should a different site preparation or harvesting technique be employed? The map of archaeological potential essentially becomes another layer of information used by land-use planners in planning activities. This provides them with options and alternatives they can choose to exercise— as long as the implications of conducting activities in areas of high potential are clearly understood.

There is a common perception that forestry activities are destructive and damaging to the ground surface, which is where one tends to find the majority of precontact archaeological sites. Viewed from a distance, forestry-related activities appear to cause considerable disturbance to the ground. Road-building, mechanical site preparation, and borrow pits are all features which undoubtedly destroy cultural resources that might coincide with that activity location. However, forestry involves many more activities than road-building, and when these are examined in more detail many of them are not as destructive as might be perceived.

A detailed examination of forestry activities in the boreal forest of Saskatchewan was conducted by Western Heritage Services to evaluate the impact and effects different forest activities have on buried archaeological resources, with the intention of developing an impact classification system (Finnigan and Gibson 1993:92–3).

The kind of impact to a site, and its degree of intensity, will dictate the kinds of responses that can be taken to minimize site disturbance or mitigation cost. Direct impacts are said to occur as a consequence of an industry-related activity.... Indirect impacts occur not so much as a consequence of forestry activity, but as a consequence of forestry activity

having taken place, for example, when construction of a logging road into a formerly inaccessible area permits hunters or campers easy access....

The amount of damage that is caused by a given forestry activity depends on the activity being undertaken. Therefore, it is necessary to devise some kind of general classification scheme which can codify the severity of site disturbance. Disturbance would entail the alteration of a site in any manner from its natural state. Although the best information can be obtained from a site which has not been disturbed at all since it was created, in practice all sites become degraded to some extent by natural causes. Furthermore, some kinds of alterations, both natural and artificial, may ostensibly appear severe, but in fact may not constitute significant disturbance from the point of view of heritage data recovery.

Finnigan and Gibson went on to develop a six-stage Impact Classification System which basically categorizes forestry activities with respect to their impact on the ground surface—and, by extension, to archaeological sites. At one end of the scale, a Class “0” impact results in no ground damage and does not require pre-impact archaeological inspection. At the other end of the scale, a Class “5” impact completely destroys the ground surface and does require a pre-impact archaeological assessment (Finnigan and Gibson 1993:93–4).

The rationale behind this impact classification system is simple. The goal of cultural resource protection is to protect cultural resources—a goal that can be achieved using a variety of means, only one of which is complete avoidance of the locality. Protection can be achieved by understanding the effects of forest activities and the threshold of impact that specific “potential” localities can withstand. The challenge is to schedule activities which do not exceed the impact threshold of specific localities. Once this is understood, an entire range of options opens up to forest management professionals and land-use planners. In Ontario, the process of identifying a impact classification system has begun in earnest. Because Ontario’s forests span a biogeographical range, from primarily hardwood woodland stands to primarily conifer boreal stands, the types of forest activities conducted in those forests are broader than those described for the boreal forest of Saskatchewan (Finnigan and Gibson 1993). However, a preliminary classification scheme (Table 5.4) has now been completed, assessing forest activities and their impacts on archaeological sites.

The importance of Finnigan and Gibson’s work is that it provides a link between predictive models, as archaeological products that rigidly define areas of doubt and

Table 5.4 Interim impact classification system for high archaeological potential—boreal/mixed-wood forests of Ontario.

Category 1	Minimal disturbance	Maximum 1% random disturbance of mineral soil Disturbance of humic layer acceptable
------------	---------------------	--

		Soil compaction acceptable
		Chemical site preparation acceptable
		Does not require pre-impact assessment
Category 2	Localized disturbance	Between 1% and 50% random disturbance
		Tire rutting acceptable
		Disturbance of humic layer acceptable
		Compaction of soil acceptable
		Some mechanical site preparation acceptable
		Requires post-impact assessment
Category 3	Complete disturbance	More than 50% random disturbance
		Any regular area disturbance (landings, leveling)
		Any activity which results in complete alteration of the ground surface
		Requires pre-impact assessment

uncertainty, and the end user, who is the land-use planner. The true challenge that faces us is the process of getting archaeologists to communicate in a manner understandable to land-use planners. When cultural heritage requirements are presented in such a way that they can be operationalized by nonarchaeologists, protecting cultural resources will become yet another land-use planning exercise. Archaeological sites are fairly easy to protect: they do not move once discovered, and they have definable boundaries. Once the impact of various activities are made clear—from a heritage integrity point of view—then forest activities appropriate for specific locations, seasons, soil types, etc. can be scheduled. For example, a tree-cutting operation can be scheduled for winter, when the ground is frozen and after a sufficient snow pack has built up. In these conditions, one would be hard pressed to locate a tire track in that area the following spring. Even though certain land-use practices are being allowed, cultural resource protection is being achieved. If a company chooses to conduct an activity that will be detrimental to potential resources, it has arrived at that decision in light of other alternatives, and the cost of hiring archaeological professionals will have been weighed alongside other factors.

It is too difficult to tie prescriptions (instructions regarding protection of values) to specific forestry activities. While the activity may remain the same (e.g. harvesting with a feller-buncher), the impact will depend upon the season of the year, the soil type, the soil moisture content, etc. It makes much more operational sense to tie the prescription to the desired outcome of the activity. In this regard, forest operators can ask themselves, “Will Activity A result in a Category 3 impact?” If the answer is yes, they are required to hire an archaeologist before conducting the activity. Since all the variables (Activity A; Category 3) are defined, there can be no doubt or uncertainty as to the necessary actions that are required.

Once again, the goal is to work with foresters and other land users, not against them. If the only available alternative provided to land-use planners is avoidance

of the value, there will be hesitancy to work with archaeologists. However, if a range of alternatives is available, presented in a format compatible with existing activity scheduling, the working relationship between archaeologists and land users will be productive and can lead to forest companies considering cultural heritage first before all other forest values, as is happening in Saskatchewan (Gibson 1995, pers. comm.). There can be no compromising the requirement that potential cultural values be protected. There must be a change in opinion that complete avoidance of the value is the only means of protecting predicted archaeological site locations.

5.7 SUMMARY

The Ontario Ministry of Natural Resources has been actively protecting cultural resources through forest management planning since 1991. Concurrently, the OMNR has been developing a predictive modeling methodology based upon weighted variables developed in a deductive framework. This methodology, tested in stages from small areas to entire forest management units, has demonstrated a high level of predictive strength. The most recent applications of the model have resulted in almost 84% of known sites to be accounted for by high-potential areas, which comprise 16.25% of the land base.

The challenge faced by archaeologists is not with the development of archaeological predictive models. This is an area where archaeologists with differing theoretical or methodological backgrounds could spend entire careers debating the merits of different GIS systems, theoretical approaches, or validation scenarios. It is time to shift these debates to a forum which continues to contribute to the development of archaeological predictive modeling without impeding it. The real challenge lies in presenting the results of archaeological predictive modeling in a manner that is not decipherable only by the gods. Only through clear explanation of predictive modeling results and the provision of explicit instructions that detail management options will the protection of cultural resources be achieved. The forest industry, like all other land users, is under increasing pressure to demonstrate responsible and sustainable use of the forests and their resources. Cultural heritage is but one of hundreds of other concerns that must be addressed in land-use planning. Forest management professionals are not cultural heritage professionals: they look up at the trees and archaeologists look down at the ground. Success will be achieved when our eyes meet and we reach the common understanding that protecting cultural resources will not result in massive withdrawal of the available land base. It is not our responsibility as archaeologists to ensure that importance is given to the protection of cultural resources. That responsibility belongs to government legislators. It is our responsibility to identify the actual or potential location of the resources and provide management alternatives through which appropriate protection can be achieved. This process has already begun in Ontario's forests,

and the success that has already been realized serves as the foundation for continued protects of Ontario's cultural resources.

References

- ALTSCHUL, J.H., 1990, Red flag models: the use of modelling in management contexts. In K.M.Allen, S.W.Green and E.B.W Zubrow (eds) *Interpreting Space: GIS and archaeology*, London: Taylor & Francis, pp. 226–238.
- BLALOCK, H.M., Jr, 1979, *Social Statistics*, revised 2nd edition, Toronto: McGraw-Hill.
- CARMICHAEL, D.L., 1990, GIS predictive modelling of prehistoric site distributions in central Montana. In K.M.Allen, S.W.Green and E.B.W Zubrow (eds) *Interpreting Space: GIS and archaeology*, London: Taylor & Francis, pp. 216–225.
- DALLA BONA, L., 1993, A predictive model of prehistoric activity location for the Souris River, Saskatchewan, unpublished MA thesis, Department of Anthropology, University of Manitoba, Winnipeg.
- 1994a, *Methodological Considerations*, Cultural Heritage Predictive Modelling Project Report Series, Vol. 3, Report prepared for the Ontario Ministry of Natural Resources, Centre for Archaeological Resource Prediction, Lakehead University, Thunder Bay, Ontario.
- 1994b, *A Predictive Model of Prehistoric Activity Location for Thunder Bay District, Ontario*, Cultural Heritage Predictive Modelling Project Report Series, Vol. 4, Report prepared for the Ontario Ministry of Natural Resources, Centre for Archaeological Resource Prediction, Lakehead University, Thunder Bay, Ontario .
- 1994c, *Appendices—Reports of Archaeological Surveys*, Cultural Heritage Predictive Modelling Project Report Series, Vol. 6, Report prepared for the Ontario Ministry of Natural Resources, Centre for Archaeological Resource Prediction, Lakehead University, Thunder Bay, Ontario.
- 1995, *Cultural Heritage Resource Predictive Modelling: 1994/1995 Predictive Modelling Pilot Projects, Northern Ontario*, Report prepared for Lakehead University, the Ontario Ministry of Natural Resources, and the Ontario Ministry of Culture, Tourism and Recreation, 31 March.
- 1996, *Archaeological Surveys in the Caribou West Timber Management Unit, Sioux Lookout District*, Ontario License No. 95–113, Report prepared for the Ontario Ministry of Citizenship, Culture and Recreation, Toronto, January.
- 1997, *Archaeological Surveys in the Caribou Forest Management Unit, Sioux Lookout District*, Ontario License No. 96–09.
- DALLA BONA, L. and LARCOMBE, L., 1996, Modeling prehistoric land use in northern Ontario. In H.D.G.Maschner (ed.) *New Methods, Old Problems: Geographic Information Systems in Modern Archaeological Research*, Center for Archaeological Investigations, Southern Illinois University at Carbondale, Occasional Paper No. 23, pp. 252–271.
- DEAN, J.S., 1983, Environmental aspects of modeling. In L.S.Cordell and D.F.Green (eds) *Theory and Modeling: Refining Survey Strategies for Location of Prehistoric Heritage Resources*, Cultural Resources Document 3, Southwestern Region: Forest Service, USDI, pp. 11–27.
- FELDBRUEGGE, D., 1979, Burrows Lake area plan, Geraldton District, Ontario Ministry of Natural Resources.

- FINNIGAN, J. and GIBSON, T. (eds), 1993, *Cultural Resource Integrated Management Project*, Report prepared for Weyerhaeuser Canada Ltd, Mistik Management Ltd, Canada Saskatchewan Partnership Agreement in Forestry, Saskatchewan Heritage Foundation, by Western Heritage Services, Inc., Saskatoon.
- HAMILTON, S. and LARCOMBE, L., 1994, *Introduction to the Research*, Cultural Heritage Predictive Modelling Project Report Series, Vol. 1, Report prepared for the Ontario Ministry of Natural Resources, Centre for Archaeological Resource Prediction, Lakehead University, Thunder Bay, Ontario.
- HAMILTON, S., LARCOMBE, L. and DALLA BONA, L., 1994, *Summary and Recommendations*, Cultural Heritage Predictive Modelling Project Report Series, Vol. 5, Report prepared for the Ontario Ministry of Natural Resources, Centre for Archaeological Resource Prediction, Lakehead University, Thunder Bay, Ontario.
- HEMMS, D., 1980, Archaeological survey of the Burrow Lake area, Geraldton District. In C.S. Reid and W.R. Ross (eds) *Studies in West Patricia Archaeology No. 2:1979–80*, Toronto: Ministry of Culture and Communication, pp. 171–206.
- KOHLER, T.A. and PARKER, S.C., 1986, Predictive models for archaeological resource location. In M.B. Schiffer (ed.) *Advances in Archaeological Method and Theory*, Vol. 9, New York: Academic Press, pp. 397–52.
- KVAMME, K.L., 1985, Determining empirical relationships between the natural environment and prehistoric site locations: a hunter-gatherer example. In C. Carr (ed.) *For Concordance in Archaeological Analysis*, St. Louis: Westport Publishers, pp. 208–238.
- 1990, One-sample tests in regional archaeological analysis: new possibilities through computer technology, *American Antiquity* **55**(2):367–381.
- LARCOMBE, L., 1994, *Boreal Forest Aboriginal Land Use Patterns*, Cultural Heritage Predictive Modelling Project Report Series, Vol. 2, Report prepared for the Ontario Ministry of Natural Resources, Centre for Archaeological Resource Prediction, Lakehead University, Thunder Bay, Ontario.
- OMNR (Ontario Ministry of Natural Resources), 1991, *Timber Management Guidelines for the Protection of Cultural Heritage Resources*, Toronto: Queen's Printer for Ontario, September.
- OMNR, 1996, *Forest Management Planning Manual for Ontario's Crown Forests*, Toronto: Queen's Printer for Ontario, September.
- PARKER, S.C., 1985, Predictive modelling of site settlement systems using multivariate logistics. In C. Carr (ed.) *For Concordance in Archaeological Analysis*, St. Louis: Westport Publishers, pp. 173–207.
- PILON, J.-L., 1998, *1997 Field Activities in the Black Bay Region, Thunder Bay District and at the Allen Site (EcJs-1), Sioux Lookout District, Ontario*, Canadian Museum of Civilization, Hull, Québec.
- SCHERMER, S.J. and TIFFANY, J.A. 1985, Environmental variables as factors in site location: an example from the Upper Midwest, *Midcontinental Journal of Archaeology* **10**:215–240.
- WARREN, R., 1990, Predictive modelling of archaeological site location: a case study in the Midwest. In K.M. Allen, S.W. Green and E.B.W. Zubrow (eds) *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis, pp. 201–215.

CHAPTER SIX

Considerations of Scale in Modeling Settlement Patterns Using GIS:

An Iroquois Example

KATHLEEN M.SYDORIAK ALLEN

Issues of spatial scale and data resolution are examined using GIS in the analyses of Iroquoian settlement locations. Models for the evolution of Iroquoian settlement patterns conflict over the relative rate at which populations became sedentary and began to reside in semipermanent (but year-round) villages. Settlement locations for base camps and villages in the central New York State region were plotted and examined using GIS with reference to particular environmental features critical for subsistence needs. Spatial and temporal variability in settlement locations needed to be evaluated as to their implications for models of Iroquoian sedentarization.

6.1 INTRODUCTION

Archaeologists have long been aware of the necessity to consider scale as they model settlement systems. Although there is no consensus on the particular spatial scales at which settlement studies might be investigated, there are common themes. In early work, Clarke (1977) used the terms “micro”, “semi-micro”, and “macro” to refer to studies at the structure, within-site, and intersite levels. Butzer (1982) used “micro” (within-structure) and “semi-micro” (within-site in a limited or multiple-activity area) and added terms for larger spatial areas. “Mesoscale” was used for within-site structure aggregation areas and “macroscale” for intersite patterning related to environmental features in or around a node of administrative, economic, or ceremonial purpose (Butzer 1982: 232–3). Although there is some difference in the terminology employed, both Clarke and Butzer emphasized patterning at the local level. More recently, the increase in regional settlement pattern studies has provoked renewed emphasis on the broader spatial patterning of settlements at both the regional and larger levels.

In this chapter, I examine some issues of scale. In particular I consider some of the different kinds of questions that can be answered using studies at different

spatial scales of analysis and I compare several studies that have been done in the eastern Great Lakes region of North America.

My aim is not to suggest that we standardize the use of certain scales of analysis for different kinds of studies. Rather, I wish to bring a consideration of spatial scale to the forefront, especially with reference to the collection of data and the kinds of questions that are asked in settlement location studies. Scales of analysis, data scales, and data resolution will be considered. These issues are important in light of the increasing availability of data in a wide variety of forms and at widely varying levels of resolution that are incorporated into GIS studies.

6.2

SPATIAL SCALE AND KINDS OF PROBLEMS

There is some consistency in the terminology used in settlement studies. Regional or macroscale studies generally refer to areas that are larger than the site level and thus involve intersite analysis, although the area covered may vary from a single drainage to a broader physiographic region. The spatial scale used depends on the research problem and on the availability of data. Marquardt and Crumley (1987) suggest that switching the scale of analysis at various times during the course of a study may be necessary. A recent GIS study in the Netherlands moves through four spatial levels from the site through to the macro-region (c. 4500 km²) (Wansleben and Verhart 1995). Such work underscores the importance of suiting the scale of the study to the problem at hand and of planning to move among several scales as needed.

In this chapter, I use *global* to refer to studies at a broad regional level that may incorporate areas of roughly 10,000 km² or more, *regional* to refer to studies of roughly 5500 km², and *local* to refer to studies of approximately 100 km² or less. These areas are derived from the scales used in studies I review.

Global studies of settlement are most useful for studying the broad effects of environmental variables on settlement and adaptive patterns. In these studies, one can examine the particular broad environmental characteristics of a region and identify the relationship between these variables and settlement location. In addition, patterns such as widespread population movements that are most visible at a broad scale may be recognized. Regional-level studies focus on more specific variables of importance. While many of these variables will continue to be environmental in nature and may include soil types, temperature regimes, and vegetation, other factors such as proximity to different ethnic groups can also be considered. The smaller focus results in a more detailed view of settlement choices that can be compared from one area to another. At the local level of analysis, particular resources available in the immediate vicinity of the site can be identified and assessed for their importance to particular location choice. For example, at this level one might consider the importance of clay deposits, nearby wetland resources, and routes between the site and resources. While these can be

examined at the regional level, the spatial scale and consequent resolution of the base data are critical in making these comparisons meaningful.

The availability of data in a variety of digital and hard-copy forms that can be readily converted into GIS databases has made the careful consideration of map scale an important aspect of GIS design. Map scale refers to the relationship between a distance on a map and the distance in the real world. In the United States, digital data is widely available from the US Geological Survey (USGS) at scales of 1:250,000, 1:100,000, and 1:24,000, although not all areas of the country have complete coverages, especially at the largest scale (1:24,000). The resolution of data is the accuracy or level of uncertainty in point location or distance measurement. At a map scale of 1:250,000, a mapped feature may vary by 127 m; at a scale of 1:24,000, a feature may vary by 12 m (Marozas and Zack 1990:168–9). The accuracy of the mapped area required varies according to the kinds of questions being asked at the spatial level of interest. At the global and often at the regional level, very specific locations (within 15–30 m) may be less important. At local levels of analysis, a lower tolerance for error exists.

In addition to data resolution, consideration must be given to cell size within the GIS. Broadly speaking, the cell size should be as small as the smallest unit area in which one is interested. If one is studying the distribution of lithic scatters, the most appropriate cell size may be 10 m² or less. Concerns of data quality, resolution of the data layers, and size of the GIS also merit attention.

These issues of spatial scale and level of resolution of the data are framed below through an examination of several case studies of the use of GIS in the analysis of Iroquois settlement location. The studies incorporate global, regional, and local scales of analysis and so are particularly apt examples of the effect of spatial scale and data resolution on settlement analyses. Following this, a brief study of settlement in the Lake Cayuga region of central New York State is discussed and compared with the pattern presented in other studies.

In what follows, the term “global” is used to refer to all of those groups that share the basic Iroquoian cultural pattern. This incorporates most of New York State and southern Ontario. “Regional” refers to subareas that correspond to tribal groups, for example the Huron settlements in the Lake Simcoe region and those in western, central, and eastern New York State that roughly correspond to tribal groups. “Subregional” or “local” refers to particular drainages or clusters of settlements in close proximity to each other.

6.3

GLOBAL, REGIONAL, AND LOCAL VIEWS OF THE IROQUOIAN WORLD

Iroquoian peoples were slash-and-burn horticulturalists at the time of European contact and much of our modeling and understanding of them derives from the observations made by early missionaries and settlers. Subsistence models stem especially from Jesuit missionary observations in Ontario (as in the Jesuit

Relations, Thwaites 1959). The Iroquoians were dependent on maize, beans, and squash horticulture although there is substantial evidence for continued utilization of a variety of wild plant and animal resources (Fenton 1978).

The location of horticulturalists in the eastern Great Lakes was eased by the presence of the Great Lakes; Lakes Erie and Ontario have a moderating effect on climate such that the number of frost-free days is sufficient for agricultural production. South of the traditional Iroquois area in New York State on the Allegheny Plateau, the frost-free season is shorter and thus the area is less conducive for horticultural settlement. Iroquoians were surrounded by upland areas to the south and east (areas less conducive to horticulture). To the north was the Canadian biotic zone, where horticulture was not possible but where hunting, fishing, and gathering were important subsistence pursuits.

Villages were spaced over the landscape. The global Iroquoian pattern includes a number of different tribal groups in the eastern Great Lakes region. In Ontario, groups at contact were located in a relatively small area, especially in the vicinity of Lake Simcoe; in New York, Iroquoian groups were spread across the state in an east-to-west orientation characterized by settlement clusters. Because there was spatial separation between each of the groups in New York at the time of European contact, and because each of these groups had names for themselves, our understanding of prehistory has been underpinned by a concern for the regional and the local. The evidence for group coalescence within restricted areas seems to stem primarily from the late prehistoric period (late fifteenth to mid-sixteenth centuries) (Snow 1994; Tuck 1971). In earlier prehistoric times, concentration of populations into territories that correspond to tribal groups is not so readily apparent. However, we can identify some regions of population concentrations (i.e., western New York, the Genesee Valley, Central New York around Syracuse, the Mohawk Valley, and the Susquehanna Valley).

These tribal settlement clusters are on the regional level. Areas occupied by the different Iroquois tribes can be divided into regions (including the Mohawk Valley, the Lake Oneida area, the hills south of Syracuse, the central Finger Lakes region, the western Finger Lakes and Genesee River valley region, and the western New York State region). These areas roughly correspond to tribal locations as known during the early contact period. On this regional level, we can trace the origins of some of these groups to the early part of the Late Woodland period (*c.* AD 1000). In the Onondaga area, the formation of the Onondaga tribe has been identified on the basis of village movement and artifactual and settlement evidence (Tuck 1978).

It is also at the regional level that we can begin to identify the ways in which Iroquoian groups differed from each other. Aside from settling in areas of productive agricultural soils, what other factors come into play in the selection of village locations? What are the differences between tribal areas? Are certain resources more available in one area than another? How does this affect interaction between groups? An understanding of the evolution of tribal forms of

organization requires that we understand similarities and differences between regions. The global view tends to obscure differences in order to understand overall settlement patterning and subsistence.

With the Iroquoian subsistence pattern, soils became depleted over time and villages were moved every 15 to 20 years. This periodic movement of villages results in a pattern of village sequences. These sequences occur at the regional level and specific sequences identify local patterns of development. Several of these sequences have been identified. The longest sequence is known from the central New York Onondaga area (Tuck 1971). The most precise example is from the historic Seneca villages south of Rochester, New York, where a dual sequence of villages has been traced from AD 1550 to 1687 (Wray and Schoff 1953; Wray *et al.* 1987, 1991). These village sequences are part of the local pattern.

In particular village sequences, we are assuming that the same population is, in general, moving from one village location to the next and therefore that there is continuity in population (ancestor/descendant) over time. In terms of settlement location, the decisions that lead to the choosing of particular village sites are of interest. Factors of agricultural soil, defensibility, proximity to other resources, previous village placement and therefore ease of village relocation and the relationship of this population to others in neighboring valleys are all involved in such decisions.

6.4

GIS AND SPATIAL SCALE: GLOBAL, REGIONAL AND LOCAL VIEWS

Geographical information systems (GIS) can be used at a number of different scales depending on the problem being investigated. Many GIS studies have been done at the regional level and most have focused on the relationship between site locations and particular environmental variables.

The Iroquoian area has been a veritable hotbed of GIS activity from the late 1980s to the present. Several major studies have been conducted. These studies have concentrated on helping us understand the location of settlements primarily in relationship to environmental features. This, of course, is not surprising, given the relative ease with which environmental data can be incorporated into a GIS and the greater uncertainty that revolves around the measurement, quantification, and even identification of variables important to understanding the impact of social and political factors in cultural development.

Although there has been general agreement in the data categories used, there has been less uniformity in scale. Studies at the broad regional or global (eastern Great Lakes including southern Ontario and New York State) (Knoerl 1988; Hasenstab 1990), regional (New York State) (Hasenstab 1990), and local (western New York; Hunt 1990) levels have been conducted. Each of these studies will be discussed below.

Several studies of the Iroquois area have examined settlement location at close to the global scale. John Knoerl dealt specifically with issues of scale in his dissertation (1988). Knoerl identified three kinds of scale: spatial, temporal, and phenomenological (which includes populational, behavioral, and archaeological scale). He notes that some types of scale (particularly at the regional level) are more commonly used by archaeologists and it is important to identify the correct scale to use in researching problems. He suggests that patterns that appear random at one level may have meaning at another, a point also discussed in some depth by Ebert (1992). Knoerl notes that archaeological units often start at the lowest level with behavior (sets of activities that are performed over space and through time by groups of varying sizes) and are combined into larger units. He identifies the articulation of scale with the archaeological phenomenon under investigation as being of critical importance and suggests that one needs to isolate the relative spatial and temporal scales at which archaeological phenomenon form patterns so that one can look for causal phenomena at the same levels. Knoerl (1988:44) defines archaeological scale as a combination of space, time, and the behavioral grouping that is correlated to the archaeological unit of interest.

Knoerl's research, then, focuses primarily on spatial scale and establishes a database that encompasses the eastern Great Lakes region. He sets up a cell-based GIS with a grid cell size of 5 km, although his environmental data come from maps of a variety of different scales. He uses trend surface analysis as a means to identify regional, subregional, and local trends. He first demonstrates the utility of trend surface analysis to pick out these trends through the use of simulated data. Then he devises a maize suitability map, which includes information on frost dates, rainfall, soil associations, and physiographic diversity. His aim is to identify the spatial scale at which the data is patterned. By plotting the residuals from trend surface analyses, and comparing these with archaeological site locations from the eastern Great Lakes region (*c.* AD 1350–1650), he is able to demonstrate that the patterning in site location appears to be at the local level. He concludes that this is the case because areas of residuals (that is, where there are areas of higher or lower ratings for maize than appear in the trend) do not correspond with site locations. There is some overlap (*c.* 38%) but fully 62% do not overlap. The pattern for site occupation dates is more localized than it is for areas suitable for maize cultivation.

Knoerl's study is important in several ways. First, his detailed discussion of scale and resolution illuminates important issues in developing and analyzing large databases. Second, he does not stop at identifying broad patterns of settlement but also notes that simply looking at the broad patterns of settlement in relationship to environmental features important for maize cultivation does not suffice for an understanding of local patterns of settlement. His work encourages us to look for local as well as global and regional patterns in settlement location.

Hasenstab (1990) too conducted a major GIS study that looks at the patterning of settlements for the New York State Iroquois. His research area includes the central east-west spine of New York State. His study is directed towards identifying the extent to which movement of sites over time is associated with external forces—specifically with the impact of Mississippian cultures to the west who demanded furs—and with other groups to the west who moved into the western New York region.

The GIS Hasenstab develops is very detailed. His cell size is 0.5 km, while the scale of his environmental data source maps varies from 1:250,000 to 1:1,000,000 (Hasenstab 1990:29). He divides his research area into three regions based on physiography in association with drainages. These three regions are the Lake Plain, the Central Riverine Valley, and the Allegheny Plateau. He includes a very large area in the Central Riverine Valley region; it encompasses several large river valleys including the Genesee River and the Mohawk and Hudson River drainages. The largest central area is the Finger Lakes region, which includes all of the Finger Lakes that flow into the Seneca River, thence to the Oswego River and into Lake Ontario. This Central Valley area was home to all of the Five Nations Iroquois and is the area where Hasenstab identifies soil productivity as an important component of settlement location.

Hasenstab (1990) looks at changes in settlement location over time in each of the three regions through examining site location in relationship to a number of variables including wetlands, canoe-navigable waterways, forest diversity, soil associations, etc. He suggests that the broad pattern of village movement is indicative of depletion of animal resources that resulted from demands for furs, skins, and meat from peoples to the west and south. He makes other interesting observations as well, especially regarding broad patterns of settlement shifts over time in relation to environmental variables. Some of these are discussed more fully below in the section on patterning in the Lake Cayuga region.

Hasenstab clearly demonstrates the utility of broad-scale investigations. His large-scale study identifies causal factors for patterning of village locations that operate at the same spatial scale. Although he uses a relatively small grid size for site location, the level at which other data was mapped is coarser. For example, soils information was obtained from 1:750000-scale maps and is therefore relatively coarse.

Eleazer Hunt (1990) conducted a major GIS study of the western region of New York State that combined a regional and local approach to settlement location. He used ARC-INFO, a vector GIS, and a base map at a scale of 1:250,000 (Hunt 1990:103). Other environmental data maps appear to range from a scale approximating 1:20,000 (soils) to 1:500,000 (isotherm climatic maps). He considered a number of variables at the regional level including precipitation, snowfall, number of frost-free days, and soil associations. At the local level he examined soil associations in great detail for a small sample of sites. Hunt concluded that sites were located in horticulturally productive areas even prior to the time of dependence on cultivation. He also demonstrated that it was not

detailed determination of soil types and textures that was critical for settlement; rather, general classes of soil textures were selected for, in addition to favorable climatic regimes, such as appropriate length of the frost-free growing season and avoidance of exceptionally heavy-snowfall areas. While his study included an in-depth assessment of soils in the catchment area surrounding a village site, his work does not as carefully consider other factors that might influence site location at the local level.

Several additional local-level studies have been done in the Iroquois area that have also focused on characteristics of soils and village placement, although in neither case was GIS employed. Vandrei (1987) examined the location of the historic Seneca sequence of villages in relationship to categories of soil productivity, as well as in relationship to site size and defensibility. He used soil maps at the 1:20,000 scale. He found that, in general, sites in the western sequence of villages were located near more acres of highest-productivity soil. Sites in this sequence were also generally larger than those in the eastern sequence. In several cases it was clear that defensive considerations were more important in settlement location than was the amount of productive soil in the vicinity.

In the Mohawk area, Bond (1982) examined the relationship between soils (1:20,000-scale maps) and settlement in an attempt to identify whether groups were maximizing production through their choice of settlement locations. He concluded that while all sites were located in regions of highly productive soils, village location was also strongly influenced by defensive concerns. There was no attempt by these groups to maximize agricultural production.

Approaches to understanding village location in relationship to specific soils focus on the local level—either in plotting the placement of sites with regard to soil associations or in identifying proportions of particular soil types within 1 or 2 km of the village location. However, in all three cases above, it appears that the Iroquois performed quite a bit less of a detailed evaluation of soil types surrounding the village than did the researchers. Both Vandrei and Bond stress the importance of defensibility in at least some of the village locations selected. Note that this is a local cause for a local pattern and is most clearly visible with large-scale data (1:24,000).

Attention to both regional and local patterns is needed. At the local level, this could be approached from the perspective of those within the settlement, on the basis of what we know of Iroquois subsistence. The Iroquois were dependent on horticulture, so settlements had to be located near productive agricultural land. However, in the local studies that have been done, it does not appear that villages were located on or near to the *best* soils. There was no trend over time for village location selection to maximize the amount of prime agricultural land. Defensive considerations were also important. In addition, access to hunting and fishing resources was important for their contribution to subsistence. While these wild food resources may have been less crucial to specific village site location than was proximity to good agricultural soils, good soils in combination with access

to other food resources could have tipped the balance in favor of one site location over another.

6.5 CENTRAL NEW YORK REGION: REGIONAL AND LOCAL

Global studies have provided us with a good understanding of those factors important in conditioning Iroquois settlement. The general importance of good agricultural soils in village placement is clear. Regional and local studies have identified additional variables. In order to understand settlement in a smaller region within the Iroquois homeland, I have been compiling a database of sites in the Cayuga Lake watershed, the home of the Cayuga tribe of the Iroquois, preparatory to developing a GIS for the region.

As I have set about developing the GIS, however, I have faced the problems of identifying the kinds of data I need to acquire, the level of data resolution to use, and how to do this in the most efficient way possible yet still ensure that enough categories of data are included to model settlement in the region adequately. As a preliminary step, I decided to combine my data on specific site location with Hasenstab's environmental data to identify those categories of data that would be most useful to a subregional and local-level study and to help in identifying the level of resolution that is needed.

In order to assist with this process, several comparisons are made. First, the extent to which subregional and local patterns in the Cayuga area correspond with those identified for the Central Riverine Valley area used by Hasenstab is considered. Does the regional pattern he has identified correspond with subregional and local patterns? Is important local patterning not visible at higher levels? Second, the extent to which differences that are found are the result of problems of data resolution is addressed. What are the areas of incongruity between his results and what is known of the area at a larger scale? Third, to what extent can the patterns visible at the subregional or local level be ascribed to local causes? This, of course, takes us back to a concern voiced earlier: that is, the level of the patterning examined is the level at which one sees causes.

Hasenstab (1990) indicates that the most consistent pattern in this central valley region is the very strong association of village locations with soil attributes favorable to maize horticulture. The soil attributes most important for maize horticulture are overall ratings for maize suitability, lime status of soils, moisture availability, and physical condition of the soils. This is particularly the case in the Transitional and Iroquois periods (*c.* AD 1250–1550). Prior to this time, during the earlier Owasco period (*c.* AD 1250), settlements were in areas that exhibited high forest productivity. Due to settlement shifts away from the Seneca River lowlands with their higher concentration of wetland resources and forest diversity, and the greater focus on good agricultural soils, Hasenstab suggests that this provides support for the hypothesis that there was depletion of

game resources and subsequent movement away from the lowlands after the Owasco period (Hasenstab 1990:59–60). However, a movement into areas that are horticulturally productive but that contain fewer “wild” resources does not necessarily suggest depletion of game resources. “Wild” resources (including those resulting from hunting and fishing activities) can be obtained more easily at locations that are more distant from the settlement than can horticultural products.

In examining the data for the Lake Cayuga region, several patterns are evident. As Hasenstab has noted for the Central Valley region, in this area as well, virtually all of the village sites are located in areas with productive agricultural soils. The majority of sites (19 of 29) are located in places with the highest possible ratings for suitability for corn cultivation, lime status, potassium rating, physical condition, and moisture availability, and with no limitations aside from relatively slow permeability of soils. Given the level of resolution of the base map used for these determinations (1:750,000, a statewide map), this relative uniformity is not surprising. There are some exceptions, however. The sites in the immediate vicinity of the Seneca River, where there is easy access to wetland resources, have less than optimal ratings for lime status and potassium. On the west side of Cayuga Lake, five of the six village sites have less than optimal ratings again for the lime and potassium variables, and one has a low rating for all five variables, including the lowest rating for suitability for corn cultivation. Finally, on the east side of the lake, there is one village site with low ratings for all five of the soil variables.

In general, then, the view from the Lake Cayuga drainage supports Hasenstab’s conclusions about the region as a whole. That is, village locations are in areas with soils that are very suitable for corn cultivation, although there are some exceptions. As a second step in this comparison of global with regional and local patterning, the Lake Cayuga watershed was divided into five groups that broadly correspond to smaller drainages within the region. These include the Seneca River Lowlands, Paines Creek drainage, Salmon Creek drainage, Great Gully and several other small creeks that flow west directly into Lake Cayuga, and the set of two drainages on the west side of the lake in close proximity to each other, Trumansburg and Taughannock Creeks. As noted earlier, sites on the west side of the lake in general have slightly less favorable soils for corn cultivation. An examination of the mean frost-free growing season data suggests that all of these villages are below the norm as well, although they are all above the 120-day frost-free growing season limit (ranges for these sites are from 138 to 155 days). However, note that the mean number of frost-free days is an average, so there are undoubtedly years when this number falls well below 138 days. These sites, then, appear to be more marginal for corn horticulture than are other sites in the region.

Are there other compensating factors that would militate against the comparative marginality of this local area for corn cultivation? Forest productivity measures within 10 km are lower for this set of sites (44 to 46) (with

the exception of one site) than for the region as a whole (48 to 53). There is little difference, however, in forest productivity within 20 km of the village (considered the traditional tribal hunting territory). Forest diversity measures are slightly above those for villages east of the lake. On the other hand, large wetland resources are further away for these western villages. There is no clear evidence, then, for increasing importance of other environmental resources to balance the somewhat poorer soil characteristics.

An examination of the other drainages reveals some differences in environmental characteristics for several sites. In particular, two sites in the vicinity of the Salmon Creek drainage exhibit differences from the general pattern. One of these sites, Locke Fort, has low ratings for all five soil characteristics. This site and another one, Genoa Fort, have low values as well for mean number of frost-free days in the growing season and for forest productivity in both 10- and 20-km catchment areas.

In order to identify the importance of defensive considerations in village location, sites were coded on the basis of the natural defensibility of the landform (based on 1:24,000-scale maps) and/or the presence of palisades or other defensive structures. In general, sites that were located on naturally defensible landforms had evidence for palisades as well. In addition, several sites in every local drainage area were situated with regard to defense. Third, Genoa Fort and Locke Fort were among those villages sited on the most highly defensible landforms. Concerns for defense were important determinants for some of the villages in the region, and overrode requirements for good agricultural soils.

6.6 CONCLUSIONS

The patterning in site location that is apparent from an examination of villages at the subregional and local levels generally supports Hasenstab's findings, while providing additional amplification. Specific instances of what appear to be less than optimal placement of villages can be understood by looking at other factors such as landforms and detailed physiographic patterning. While explanations for village movements over time in the global view invoke global-level population influence and intrusions, at the regional and subregional level village locations can be seen as resulting from other needs, including defense.

In further investigation of these issues, more detailed mapping of soils especially is needed, as well as better estimation of frost-free seasons, taking extreme values into account rather than depending on means. This information can assist in the identification of the extent to which an apparently less optimal region is actually so. Using larger-scale soil maps will allow for the identification of possible areas of more productive soils in the immediate vicinity of sites. More detailed mapping of particular physiographic and geological features within the local environment is also important. In addition, characteristics of village size and material remains are required for a better understanding of local

and regional village patterning. For example, one of the village sites on the western side of the lake where I have been working has large quantities of deer bone. From the larger pattern of hypothesized depletion of animal resources and primary dependence of populations on corn cultivation, we might expect less evidence for hunting or perhaps, alternatively, greater dependence on a larger mammal with high return for the effort. Comparisons of faunal diversity and butchering practices between village sites will help in answering this question.

Finally, the usefulness of global-scale GIS for ideas about local processes is apparent. However, better understanding of the subregional and the local depends on the use of more detailed and more diverse data. Although the level of detail required must fit the particular questions under investigation, the fruitfulness of working with large-scale and small-scale data sources and identifying and playing off global, regional, and local views is clear. Larger-scale data (1:24,000) provides the kind of specific information on local environmental characteristics that is critical for understanding decision-making at the village or camp level. Smaller-scale data (1:250,000 or 1:100,000) permits the delineation of environmental factors of broad importance.

These observations mesh with some comments made in a recent article in a volume on GIS (Gaffney and van Leusen 1995). In a debate about the purposes for which GIS is employed and the extent to which GIS studies tend toward environmental determinism, Gaffney and coworkers note that small-scale data provides information on what is possible within a given environmental framework, while larger-scale data allows for the specific delineation of local environmental and social factors involved in selecting settlement location. Gaffney advocates concern with landscape history, with the local level at which people perceive the environment and in which they act (Gaffney and van Leusen 1995:377). The influence of cultural factors becomes more obvious at larger scales (i.e., 1:24,000) of research.

Using data at the largest possible scale provides the greatest flexibility in investigating problems at several levels of analysis. Household, village, tribal, and intertribal analyses are all possible with high-quality, large-scale data. In addition, large-scale data permits the greatest accuracy in locational information and increases the likelihood that landscape features perceived by past residents become visible in the present. In this way, cultural landscapes can be identified and followed through time and a better understanding of cultural change in response to environmental and cultural factors can be achieved.

Acknowledgments

Thanks are due to Konnie Wescott and Joe Brandon for inviting me to participate in the symposium. Appreciation is offered to the University of Pittsburgh for a grant to collect site data on the Cayuga Lake watershed and to the institutions that provided access to site location information, including the University at Buffalo Archaeological Survey, the Rochester Museum and Science Center, the

New York State Museum, the NYS Office of Parks, Recreation, and Historic Preservation, and the Public Archaeology Facility at the University at Binghamton. Grateful thanks to Konnie Wescott for her suggestions for revisions.

References

- BOND, S.C., Jr, 1982, The relationship between soils and settlement patterns in the Mohawk Valley. In D.R.Snow (ed.) *The Mohawk Valley Project: 1982 Field Season Report*, Institute for Northeast Anthropology, State University of New York at Albany, pp. 17–0.
- BUTZER, K.W., 1982, *Archaeology as Human Ecology*, New York: Cambridge University Press.
- CLARKE, D.L., 1977, Spatial information in archaeology In D.L.Clarke (ed.) *Spatial Archaeology*, New York: Academic Press, pp. 1–32.
- EBERT, J.I., 1992, *Distributional Archaeology*, Albuquerque: University of New Mexico Press.
- FENTON, W.N., 1978, Northern Iroquoian culture patterns In B.G.Trigger (ed.) *Northeast: Handbook of North American Indians*, Washington, DC: Smithsonian Institution, pp. 296–303.
- GAFFNEY, V. and VAN LEUSEN, M., 1995, Postscript—GIS, environmental determinism and archaeology: a parallel text . In G. Lock and Z. Stan i (eds) *Archaeology and Geographic Information Systems*, London: Taylor & Francis, pp. 367–382.
- HASENSTAB, R.J., 1990, Agriculture, warfare, and tribalization in the Iroquois homeland of New York: a G.I.S analysis of late woodland settlement, PhD dissertation, University of Massachusetts, Amherst, University Microfilms, Ann Arbor.
- HUNT, E.D., 1990, A consideration of environmental, physiographic and horticultural systems as they impact late woodland settlement patterns in western New York State: a geographic information systems (G.I.S.) analysis of site location, PhD dissertation, State University of New York at Buffalo, University Microfilms, Ann Arbor.
- KNOERL, J.J., 1988, The concept of scale and its application to prehistoric and historic Iroquois sites in the Lake Ontario drainage basin, PhD dissertation, State University of New York at Binghamton, University Microfilms, Ann Arbor.
- MARQUARDT, W.H. and CRUMLEY, C.L. (eds), 1987, *Regional Dynamics: Burgundian Landscapes in Historical Perspective*, San Diego: Academic Press.
- MAROZAS, B.A. and ZACK, J.A., 1990, GIS and archaeological site location. In K.M.S. Allen, S.W.Green, and E.B.W.Zubrow (eds) *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis, pp. 165–172.
- SNOW, D.R., 1994, *The Iroquois*, Oxford: Blackwell.
- THWAITES, R.G. (ed.), 1959, *The Jesuit Relations and Allied Documents 1610–1791*, 73 volumes, New York: Pageant.
- TUCK, J.A., 1971, *Onondaga Iroquois Prehistory: A Study in Settlement Archaeology*, Syracuse: Syracuse University Press.

- 1978, Northern Iroquois prehistory. In B.G.Trigger (ed.) *Northeast: Handbook of North American Indians*, Washington, DC: Smithsonian Institution, pp. 322–333.
- VANDREI, C., 1987, Observations on Seneca settlement in the early historic period, *The Bulletin, Journal of the New York State Archaeological Association* **95**:8–17.
- WANSLEEBEN, M. and VERHART, L.B.M., 1995, GIS on different spatial levels and the neolithization process in the south-eastern Netherlands. In G.Lock and Z.Stan i (eds) *Archaeology and Geographic Information Systems*, London: Taylor & Francis, pp. 153–169.
- WRAY, C.A. and SCHOFF, H.L., 1953, A preliminary report on the Seneca sequence in western New York (1550–1687), *Pennsylvania Archaeologist* **22**:53–63.
- WRAY, C.F., SEMPOWSKI, M.L. and SAUNDERS, L.P., 1991, *Tram and Cameron: Two Early Contact Period Seneca Sites*, Charles F.Wray Series in Seneca Archaeology, Vol. 2, Rochester Museum and Science Center, Research Records No. 21, Rochester, NY.
- WRAY, C.F., SEMPOWSKI, M.L., SAUNDERS, L.P. and CERVONE, G.C., 1987, *The Adams and Culbertson Sites*, Charles F.Wray Series in Seneca Archaeology, Vol. 1, Rochester Museum and Science Center, Research Records No. 19, Rochester, NY.

CHAPTER SEVEN

Construction of Digital Elevation Models for Archaeological Applications

JON B.HAGEMAN AND DAVID A.BENNETT

The use of interpolation in archaeology is becoming common. As archaeologists incorporate geographic information systems (GIS) and computer mapping programs into their research, questions of interpolation become fundamental considerations in the representation and manipulation of topographic data. To date, however, few archaeologists have dealt with these questions. Uncritical use of interpolation algorithms can result in unrealistic representations of the landscape in a mapping program or can result in an inaccurate digital elevation model (DEM) used in a GIS. This, in turn, can lead to an ineffective predictive model of site location. By carefully selecting an interpolation algorithm that is well suited to the data, statistical pitfalls and wasted effort can be avoided.

7.1 INTRODUCTION

The creation of digital elevation models (DEM) in archaeological applications of geographic information systems (GIS) has, with rare exceptions (e.g., Kvamme 1990; Warren 1990; Wiemer 1995; Madry and Rakos 1996), been largely ignored in print. As Kvamme (1990:123) has noted, archaeologists are usually concerned with the quality of archaeological data, not the quality of data obtained by computer means. Yet given the same data points, substantively different surfaces can be generated from alternative computer algorithms designed to accomplish the same task. These differences can have significant and unexpected impacts on archaeological investigations.

Consider, for example, the development of a site prediction model in which elevation, slope, and aspect are important independent variables. Available elevation data is likely to be incomplete and/or in a form that is not suitable for the calculation of slope and aspect (e.g., sampled data points or contour lines). To construct a usable DEM (e.g., a lattice of elevation points), an interpolation algorithm must be applied. Yet different algorithms can provide different

elevations for the same point in space. Landscapes constructed using alternative interpolation algorithms may superficially appear to be similar, but be both quantitatively and qualitatively different. The implications of this for the predictive modeling of archaeological site locations are critical, as landforms that had been used in the past may be under- or overrepresented in the derived data set. Thus, selecting an inappropriate interpolation algorithm could lead to a low degree of accuracy in the overall predictive model.

Unfortunately, guidelines for selecting a particular DEM or interpolation method do not exist in the archaeological literature. How, then, can the archaeologist determine which interpolation algorithm to use on a given set of points? Is there a single best algorithm? Are vector-based DEMs more accurate than raster-based DEMs?¹

The purpose of this chapter is to examine these questions. A data set from northwestern Belize will be used to describe methods by which archaeologists can evaluate the results of alternative interpolation algorithms that create DEMs for use in archaeological applications, including predictive models of site location. This comparison of algorithms will be made within the context of ARC/INFO Version 7.0, a commercially available GIS that is commonly used in academic, government, and private-sector applications.

7.2

WHY SHOULD THE ARCHAEOLOGIST CARE ABOUT INTERPOLATION?

Although digital elevation models are becoming increasingly available from government (e.g., USGS) and commercial (e.g., SPOT) sources, there are several reasons why archaeologists should develop a fundamental understanding of interpolation techniques. First, the spatial coverage of such databases is far from complete (Kvamme 1995:6; Madry and Rakos 1996:118). This is particularly true for Third World nations, but even in the United States many areas do not have the 7.5-minute quadrangle series in a digital format. For example, less than one-third of the state of Illinois was available in 7.5-minute digital format as of 1 August 1996 (USGS 1996).

As alluded to above, archaeologists work at a variety of spatial scales, from that of the region down to that of a single excavation unit. Available digital data sets tend to provide regional or continental coverage, and thus often lack the spatial resolution and accuracy needed for site-specific work (e.g., Biswell *et al.* 1995; Gaffney and Stan i 1991; Meffert 1995). Accordingly, archaeologists working in areas without existing digital topographic coverages at the appropriate resolution may have to create their own DEMs by digitizing topographic maps or by capturing elevation data using total stations and/or global positioning system (GPS) receivers (Forte 1995:232; Madry and Rakos 1996:118; Rick 1996). Points located with total stations or GPS are typically

recorded as three-dimensional (XYZ) coordinates, and can be entered into a GIS or mapping program to create a DEM and/or a contour map.

Finally, all DEMs are discrete approximations of a continuous phenomenon. How closely this approximation reflects reality depends on a variety of factors that include:

- 1 how many sample points were collected;
- 2 where they were collected;
- 3 the accuracy of the data collection device;
- 4 the skill and knowledge of the data collector; and
- 5 the explicit and implicit assumptions built into the interpolation algorithm (Robinson *et al.* 1995).

Just as the improper use of statistical tools can lead to a misinterpretation of data, the misuse of DEMs and interpolation algorithms can result in a misinterpretation of terrain. To maintain high levels of accuracy, therefore, the archaeologist using automated forms of data recording and manipulation should be aware of the assumptions that these tools bring with them.

7.3

WHAT IS INTERPOLATION?

A brief review of interpolation is appropriate at this point. Burrough (1986:147) defines interpolation as “the procedure of estimating the value of properties at unsampled sites within the area covered by existing point observations.” This is largely based on the rationale that two points that are near one another in space are more similar than two points farther apart (i.e., spatial autocorrelation exists in topo-graphical data²). The goal of interpolation is to model variation so that values at unknown locations may be estimated on the basis of known values in the vicinity. For the purposes of this chapter, interpolation algorithms take a set of data points in space and create a digital elevation model (DEM) from which a continuous surface may be inferred. Since the DEM is fundamental to locational modeling in archaeology, it is important to understand (1) the data structure of each method, (2) the assumptions of each method, and (3) how each method manipulates a data set to construct a DEM. Next, four types of interpolation methods will be reviewed. Each of these have either appeared in written reports on the use of GIS within archaeology or are commonly used in the geosciences. In this review we describe the characteristics of each algorithm. These descriptions are derived primarily from Burrough (1986).

7.3.1 Ordinary Kriging

Ordinary Kriging is an algorithm based on stochastic or random surfaces, rather than on mathematical smoothing functions (Ripley 1981:45). The product of this type of interpolation is a lattice. Kriging assumes that variation across a landscape can be expressed as a sum of: (1) a constant trend; (2) a random, spatially correlated component; and (3) random noise. The technique requires that the random, spatially correlated variation in a data set be relatively homogeneous, so that differences between known points are functions of the distance between those points. The semivariance is calculated from the sample data (often the variance is used). This semivariance is then used to determine weights for interpolation, since it is a function of the distance between sample points (Burrough 1986:155–6). With Ordinary Kriging it is usually assumed that there is no inherent trend in the data. By considering directional differences in the semivariance, such trends can be incorporated into the interpolation process. In sum, Ordinary Kriging looks to the data set to judge the area to examine for a specified number of known data points around the location to be interpolated. Data beyond this area is assumed to possess little predictive value.

Different semivariogram models can be fitted to the estimated semivariance, and some models fit the data better than others. ARC/INFO provides five different semivariogram models for use in Ordinary Kriging. Since a high degree of homogeneity is assumed between data points, Ordinary Kriging is usually not recommended for use in data sets that contain sharp breaks in the landscape, such as steep cliffs and ridges (Aronoff 1993:220). Ordinary Kriging can, however, handle even and uneven distributions of points. Ordinary Kriging is a frequently used interpolation method in the geosciences (Cressie 1993; Carr 1995).

7.3.2 Universal Kriging

Universal Kriging is another lattice-based interpolation method. Though similar to Ordinary Kriging in its general assumptions, Universal Kriging has the added assumption of well-defined, though not extreme, local variations or drift within the larger landscape. Accordingly, the random noise within the local variation is assumed to have a semivariogram within the locality (Lam 1983:133). As such, Universal Kriging is applicable to slightly more complex landforms than Ordinary Kriging. Burrough (1986:161) suggests that Universal Kriging can be used with smoothly varying landforms. If the local variation is too extreme, such as a cliff or ridge, it may be treated as random noise or residual error (the nugget in semivariograms). If this is the case, data sets with large residual error may stand to gain very little from using Universal Kriging instead of Ordinary Kriging (Webster and Burgess 1980 cited in Burrough 1986:161). In general, however, Ordinary Kriging has more restrictive assumptions but fewer

computational problems, while Universal Kriging has more generalized assumptions but places greater demands on processing time (Lam 1983:133). ARC/INFO provides two types of Universal Kriging: one with a linear local interpolator and the second with a quadratic local interpolator. As with Ordinary Kriging, Universal Kriging is widely used in geoscience applications of spatial statistics (Cressie 1993).

7.3.3

Inverse distance weighting (IDW)

A third algorithm is known as inverse distance weighting, or IDW. IDW is a lattice-based algorithm that calculates the unknown elevation at a point by computing an average value from a fixed distance, or window, from that point. The influence that a given sample point has on an interpolated value at a different point is weighted by the inverse of the distance between the two points (Burrough 1986:153). A certain minimum number of points (often $n=12$) is required to increase accuracy. Thus, as the window “moves” to a cell with an unknown Z-value, the nearest known n points are located and a weighted average is computed. This process is repeated until the elevation for each cell in the lattice has been calculated, resulting in a DEM. IDW assumes a more or less regular distribution of points, since clustering of data may create undesirable results (Ripley 1981:36–7). In contrast to kriging, which assumes a random component, IDW is more of a smoothing function. IDW was used by Robert Warren (1990) in his creation of a predictive model for archaeological site location within the Shawnee National Forest in southern Illinois.

7.3.4

Triangulated irregular network (TIN)

Triangulated irregular networks (TINs) are often used to construct DEMs for use in archaeological predictive modeling (e.g., Marozas and Zack 1990; Fedick 1994). In contrast to the lattice-based methods of DEM construction described above, the TIN is a vector-based structure. As such, it has a drastically different appearance, and often significantly smaller data storage requirements (Peucker *et al.* 1978). A TIN is composed of a set of triangular facets derived from irregularly spaced data points. TINs often are used to accurately represent stream channels and ridge lines. Accordingly, a major assumption of TIN utilized in this manner is that the digitizing process captures the overall landform as a set of topographically significant points rather than contour-line inflections (ESRI 1995). However, most TIN generation algorithms produce a Delaunay triangulation.

A triangulation is considered to be a Delaunay triangulation if the circle defined by the vertices of each triangle does not contain any other point in the data set. This circle rule generates triangles that are as equilateral as possible and

produces a triangulation that is the dual of the Thiessen diagram defined by the same data set (Worboys 1995). Edges within a Delaunay triangulation, however, will not necessarily follow such topographically important features as ridge and stream lines. In ARC/INFO, these features (referred to as break lines) must be imposed onto the Delaunay triangulation.

ARC/INFO provides two interpolators for TINs. The first is linear, and represents the landscape surface as the flat face of a triangle. The second is quintic, and can represent each facet with a curved surface, if appropriate. TIN has been used in predictive modeling efforts in Belize (Fedick 1994), in Montana (Marozas and Zack 1990), and in Hungary (Csáki *et al* 1995).

7.4

SELECTING AN INTERPOLATION ALGORITHM

The description of these four methods for constructing DEMs (Ordinary Kriging, Universal Kriging, IDW, and TIN) answers one of the questions asked above: Is there a single best interpolation algorithm? The answer is that no single algorithm is superior to all others across various applications. The consensus among geographers and others who deal with topographic modeling is that the selection of an appropriate interpolation algorithm depends largely on the type of data being used, whether the data fits the assumptions of the algorithm, the degree of accuracy desired, and the amount of time that can be spent on data processing (Aronoff 1993; Roman *et al.* 1995; Burrough 1986; Houk 1984:18; Lam 1983: 130). How, then, does one go about choosing between TIN or one of the lattice-based interpolation algorithms for the construction of a DEM?

Previous applications of GIS to archaeological predictive modeling do not provide much guidance in this endeavor. Though Kvamme (1990) has pointed out that different algorithms produce different results, most studies have not provided an explicit rationale behind the use of a particular algorithm in the creation of a DEM (e.g., Kvamme and Jochim 1985; Maschner 1996). Some appeal to “past experience” as the criteria used to select a particular method (Marozas and Zack 1990:167). Others allude to problems with the interpolation algorithm that was utilized in a particular study (Warren 1990:211). Otherwise, few guidelines exist in the archaeological literature regarding the selection of a particular type of DEM or interpolation algorithm for use in predictive modeling.

Researchers in other disciplines have conducted studies that compared various interpolation methods in an effort to select one that is best suited to their data set. A qualitative means of doing this is through the use of visualization, which consists of inspecting the DEM for any spurious data or undesirable effects produced by the interpolation algorithm. This allows the user to explore the pattern of error that might result from the creation of a DEM (Weibel and Heller 1991: 285; Wood and Fisher 1993:55).

Quantitative methods can also be used to compare the relative accuracy of DEMs. This research revolves around applying multiple algorithms to a single

data set, and comparing interpolated values with the actual elevations at known reference points (e.g., Monckton 1994; van Kuilenburg *et al.* 1982; Weibel and Heller 1991:285). Next, the root mean square error (RMSE) for each DEM is calculated; the individual RMSEs are then compared to one another. The RMSE provides an indication of how well interpolation algorithms represent the actual topography. The utility of this index depends in large measure on the number and location of real-world data points and the spatial variability of the terrain.

To ascertain which algorithm is best suited to a particular data set, it is necessary to consider a variety of factors, which include the type of data, algorithm assumptions, desired accuracy, and processing time. Archaeologists should perform the same qualitative and quantitative comparisons between DEMs generated by TIN or different lattice-based interpolation algorithms to identify the method that is most appropriate for their data set. To illustrate how this can be done, we will examine a real-world data set generated from paper maps.

7.5 A BELIZEAN CASE STUDY

7.5.1 The data

Data for this exercise was obtained from 1:50000-scale topographic maps of Belize, produced under the direction of the Director General of Military Survey of the UK Ministry of Defence and published in 1992. The study area consists of four maps, which represent a large portion of northwestern Belize (Figure 7.1). The resulting area measures 46 km×43 km (north-south by east-west), covering a total of 1978 km². Over 50000 points were manually digitized in the process of creating this topographic coverage.

This area of northwestern Belize is characterized by two rivers flowing southwest-northeast: the Rio Bravo to the west, and the Booth's River to the east. To the west of the Booth's River, a karstic topography dominates the landscape. To the east of the same river is a large coastal plain. The land surface is characterized by the Booth's River floodplain and an associated swamp. Elevations across the entirety of the study area range from 7 to 20 m above mean sea level (AMSL) on the coastal plain and 20 to >300 m AMSL in the karstic uplands. The contour interval for these maps is 20 m, making the choice of DEM extremely important so as not to compromise surface accuracy or detail. This highly variable data set is used in spite of, rather than in accordance with, the assumptions of the interpolators described above purely for the purpose of demonstrating qualitative and quantitative error assessment.

As mentioned above, this coverage was manually digitized from paper contour maps. The digitizing regime centered on representing contour inflections rather than capturing data about individual features or landforms. Prior to running the

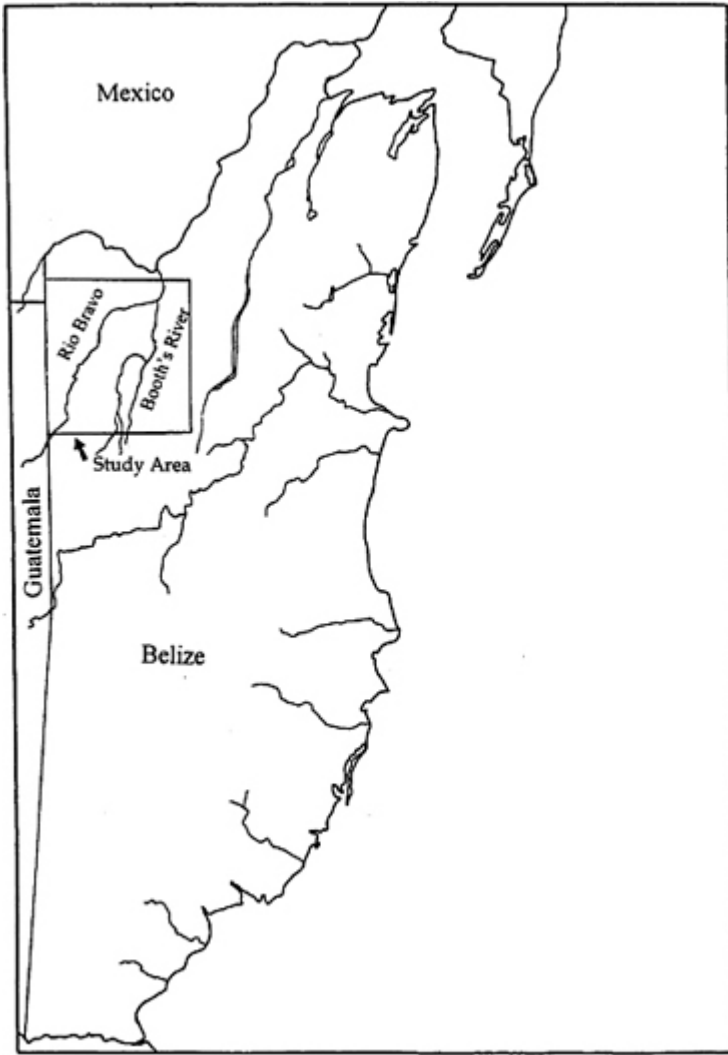


Figure 7.1 Map of Belize showing location of study area.

data set through various interpolation algorithms, a generalizing procedure was run. The effect of this procedure was to eliminate redundancy by removing points of equal elevation with virtually identical values located closely together that do not contribute to the form of the line.³ As a result, some 15,000 points were removed from the data set.

Prior to constructing a DEM from this database, we considered the assumptions of each of the algorithms described above. The data set consists of an irregular distribution of points that are not homogeneous in their elevations.

Since the digitizing method did not explicitly include break-line features, and contours rather than landforms were digitized, TINs were ruled out. At first glance, the uneven distribution of data points suggested that IDW would not be applicable here; however, the generalization procedure reduced the relatively uneven nature of this particular data set. Thus, IDW would be a consideration. Kriging, with its ability to handle unevenly distributed data, was also thought to be a favorable choice. Whether to prefer ordinary or universal kriging, however, was not obvious. Thus, both forms of kriging (with their accompanying variants) were deemed worthy of consideration.

In practice, this would suggest that a comparison of the results produced by IDW and various forms of kriging would be a productive means of identifying the best algorithm for this data set. For this study, however, all of the previously discussed algorithms were applied to the data set in an effort to illustrate the pitfalls associated with an inadequate understanding of interpolation techniques.

7.5.2

The analysis

All analyses were performed using ARC/INFO Version 7.0, operating on a Sun Sparcstation 5. Several common parameters for each algorithm were held as constant as possible in an effort to maintain comparability. For lattice-based interpolators, we used a maximum search radius of 250 m to search for the nearest 12 points. A 460×430 DEM was created to produce lattices with an individual cell size of 100 m×100 m.

The IDW procedure took about 10 minutes to complete. Five Ordinary Kriging procedures were run, based on circular, exponential, spherical, linear, and Gaussian models. Two Universal Kriging procedures were also performed, one with a linear local interpolator and one with a quadratic local interpolator. Each of the kriging operations required about 1.5 hours of processing time. Next the TIN was created; this took about 5 minutes. From this TIN, both a linear and a quintic interpolation were run. These procedures resulted in a total of ten DEMs.

One hundred reference points were then digitized from the original paper maps. Survey benchmarks and spot heights were used as reference points. None of these points was located on a contour line. The reference points and the predicted elevations for a given DEM were compared and the resulting error was calculated. Fifteen of the reference points were eventually disregarded as they were too near the edge of the coverage to provide accurate results.

In fact, a few interpolators returned negative elevation values near the margins of the DEM. This may have been due to edge effect. Edge effect is the artificial exaggeration of certain landscape trends (such as steep slopes) by an interpolation algorithm due to a shortage of information along the edge of the coverage, resulting in unrealistically high or low elevation values (Clarke 1995). Once these points with spurious elevations were identified and discarded, the root

mean square error (RMSE) for each DEM was calculated and recorded. RMSE is, essentially, standard deviation. The procedure of comparing reference and DEM values was repeated for each DEM.

7.5.3 The results

The results of this study are presented here in tabular form (Table 7.1). A low RMSE is desirable. As can be seen, no DEM perfectly matched the reference points. Some DEMs, however, are distinctly better than others. Kriging with a circular semivariogram model has the lowest RMS error at 7.99 m, and appears to best fit this data set. The TIN interpolations are extremely inaccurate; this could be due to the manner in which the contours were initially digitized (ESRI 1995). Essentially, TIN is more accurate when the data represents landforms rather than contour lines. In addition, the two methods with the most complex local interpolator, the Quintic TIN and Quadratic Universal Kriging, were also the least accurate. These algorithms seemed to be particularly inaccurate when estimating the elevation of points in close proximity to steep slopes, such as cliff edges.

Table 7.1 also verifies the rationale outlined above for selecting an algorithm. Given the nature of the Belizean data, IDW and one of the various forms of kriging were thought to be particularly suitable. TINs were considered to be inapplicable to this data set. As Table 7.1 shows, IDW and six forms of kriging have an RMSE within 3.3 m of one another.

Quantitative analysis provides only half of the picture, however.⁴ While Table 7.1 shows the overall error of each DEM, it does not indicate how this error is spatially distributed. A qualitative visual analysis can be used to identify the spatial distribution

Table 7.1 Comparison of digital elevation models (DEMs) by their root mean square error (RMSE).

DEM method	RMSE
Ordinary Kriging (circular)	7.990
Inverse distance weighting (IDW)	9.240
Ordinary Kriging (exponential)	9.994
Universal Kriging (linear)	10.180
Ordinary Kriging (spherical)	10.319
Ordinary Kriging (linear)	10.509
Ordinary Kriging (Gaussian)	11.311
TIN (linear interpolator)	17.129
TIN (quintic interpolator)	18.388
Universal Kriging (quadratic)	26.375



Figure 7.2 Map of study area (box delineates area represented in Figures 7.3 and 7.4).

of error (Wood and Fisher 1993). Many of the reference points used to assess RMSE came from peaks, which are notorious for being sources of error in DEMs. An examination of the reference points indicates that, for all methods described above, predicted elevations for hilltops and peaks were very low. This problem may be remedied by digitizing peaks in as points (rather than arcs) when such values are provided on the source map.

In addition, these features may be checked visually by generating a grayscale representation of the coverage and comparing it to the original source map. Potential sources of error, such as peaks, ridge lines, and streams, can be examined in this manner. Figure 7.2 shows a portion of the study area as represented by the source map, from which the digital data for this exercise was obtained. This area is characterized by uplands to the west, with a sharp drop to the Booth's River floodplain as one moves east. Figure 7.3 is a grayscale

representation of the same area created by ordinary kriging with a circular model. Note how the swampy area of the floodplain is represented as alternating dark and light shades of gray. The darker the color, the lower the elevation; thus, this area appears lower than the adjacent river. These alternating bands of color are an artifact of the interpolation algorithm, and could be remedied by including several points of the same elevation as the rest of the floodplain in the zone between the cliff line and the even-colored area to the east.

In contrast, [Figure 7.4](#) is the same area created by IDW. Note that the area appears to be characterized by periodic rises in the floodplain, without any trace of the low area seen in [Figure 7.3](#). Again, this is an unintended effect of the interpolation algorithm. A solution to this problem would be similar to that mentioned for the kriging example mentioned above: to include additional points on the floodplain east of the cliff edge in order to reduce the reliance of the algorithm on points located on the cliff edge for information. As can be seen, then, the two algorithms with the lowest RMSE are not perfect. Additional work beyond the digitizing of contours is required.

As noted above, the data set used in this study was generalized to remove redundant data. After the data set had been generalized, data points were more or less regularly distributed across the landscape. Thus, we were fortunate not to encounter a problem common to users of IDW. In their evaluation of different interpolation algorithms, two geographers, Wood and Fisher (1993), experienced a terracing, or “stair-step”, effect in their IDW-generated DEM. In his creation of a predictive model for southern Illinois, Robert Warren (1990:210–11) encountered an identical problem, in that the hill slopes of the resultant DEM appeared as “step-like tiers.” This is an artifact of the IDW algorithm applied to a data set composed of points that are clustered closely together. In these cases, both landscapes were represented inaccurately. In Warren’s (1990) case, this was identified as the primary factor in a predictive model of archaeological site location that was not very powerful. These particular situations are illustrative of the need for qualitative as well as quantitative examination of DEMs used in archaeological analysis.

7.6 CONCLUSION

As Kvamme (1990:123) has noted, archaeologists tend to be more concerned with archaeological data than data obtained from a computer. Warren (1990) found that an inaccurate DEM can be detrimental to the overall power of a predictive model, since coverages such as slope and aspect are ultimately derived from a DEM. This exercise, based on topographic data from northwestern Belize, has shown that the same critical eye an archaeologist casts toward a sample of artifacts can also be used to evaluate the overall quality of data manipulated within the context of a GIS. Given the fact that an archaeologist using a GIS or computer mapping program is likely at some point to encounter a situation in



Figure 7.3 Booth's River floodplain as created by Ordinary Kriging using a circular model.

which interpolation is a consideration, a basic understanding of the nature of interpolation is essential. Through an awareness of the assumptions of various algorithms, as well as the combined use of quantitative and qualitative methods, archaeologists can create more accurate representations of land surfaces and avoid the pitfalls inherent in the uncritical usage of spatial statistics. This, in turn, can lead to more powerful predictive models of archaeological site location.

Acknowledgments

The authors would like to thank the Office of Lands and Surveys, Ministry of Natural Resources, Belize, for kindly allowing them to reproduce portions of Topographic Sheets 9 and 14. We would also like to thank Charles Hargrove for providing us with digitized versions of Topographic Sheets 8, 9, 13, and 14.

Notes

- 1 Throughout this chapter, the terms "lattice" and "raster" are interchangeable.
- 2 For a recent review of this concept see Vasiliev (1996).
- 3 This procedure uses the Douglas-Peucker (1973) simplification algorithm.
- 4 For a discussion on the limitations of RMSE as an error estimation technique, see Morad *et al.* (1996).

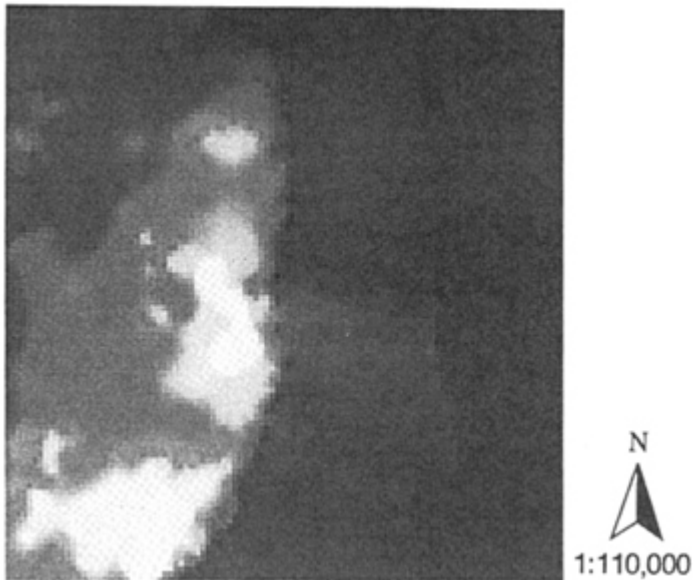


Figure 7.4 Booth's River floodplain as created by inverse distance weighting (IDW).

References

- ARONOFF, S., 1993, *Geographic Information Systems: A Management Perspective*, Ottawa: WDL Publications.
- BISWELL, S., CROPPER, L. EVANS, J., GAFFNEY, V. and LEACH, P., 1995, GIS and excavation: a cautionary tale from Shepton Mallet, Somerset, England. In G.Lock and Z.Stan i (eds) *Archaeology and Geographical Information Systems*, London: Taylor & Francis, pp. 269–286.
- BOMAN, G.K., MOLZ, F.J. and GUVEN, O., 1995, An evaluation of interpolation methodologies for generating three-dimensional hydraulic property distributions from measured data, *Ground Water* **33**:247–258.
- BURROUGH, P.A., 1986, *Principles of Geographical Information Systems for Land Resources Assessment*, Oxford: Clarendon Press.
- CARR, J.R., 1995, *Numerical Analysis for the Geological Sciences*, Englewood Cliffs, NJ: Prentice-Hall.
- CLARKE, K.C., 1995, *Analytical and Computer Cartography*, Englewood Cliffs, NJ: Prentice-Hall.
- CRESSIE, N.A., 1993, *Statistics for Spatial Data*, New York: John Wiley.
- CSÁKI, Gy., JEREM E. and REDŐ, F., 1995, Data recording and GIS applications in landscape and intra-site analysis: case studies in progress at the Archaeological Institute of the Hungarian Academy of Sciences. In G.Lock and Z.Stan i (eds) *Archaeology and Geographical Information Systems*, London: Taylor & Francis, pp. 85–100.

- DOUGLAS, D.H. and PEUCKER, T.K., 1973, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *Canadian Geographer* **10**: 112–122.
- ESRI (Environmental Systems Research Institute, Inc.), 1995, *ArcDoc: Documentation for ARC/INFO, Version 7.0*, ESRI, Redlands, CA.
- FEDICK, S.L., 1994, Ancient Maya agricultural terracing in the upper Belize River area: computer-aided modeling and the results of initial field investigations, *Ancient Mesoamerica* **5**:107–128.
- FORTE, M., 1995, Scientific visualization and archaeological landscape: the case study of a terramarra, Italy. In G.Lock and Z.Stan i (eds) *Archaeology and Geographical Information Systems*, London: Taylor & Francis, pp. 231–238.
- GAFFNEY, V. and STAN I, Z., 1991, *GIS Approaches to Regional Analysis: A Case Study of the Island of Hvar*, Ljubljana: Znanstveni Inštitut Filozofske Fakultete.
- HOUK, R.Z., 1984, A comparison of gridding algorithms. In *GeoTech '84: Personal Computers in Geology*, American Institute of Professional Geologists, Colorado Section, pp. 18–22.
- KVAMME, K.L., 1990, GIS algorithms and their effects on regional archaeological analysis. In K.M.S.Allen, S.W.Green, and E.B.W.Zubrow (eds) *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis, pp. 112–126.
- 1995, A view from across the water: the North American experience in archaeological GIS. In G.Lock and Z.Stan i (eds) *Archaeology and Geographical Information Systems*, London: Taylor & Francis, pp. 1–14.
- KVAMME, K.L. and JOCHIM, M.A., 1985, The environmental basis of Mesolithic settlement. In C.Bonsall (ed.) *The Mesolithic in Europe*, Edinburgh: John Donald Publishers, pp. 1–12.
- LAM, N. Siu-Ngan, 1983, Spatial interpolation methods: a review, *American Cartographer* **10**:129–149.
- MADRY, S.L.H. and RAKOS, L., 1996, Line-of-sight and cost-surface techniques for regional research in the Arroux River Valley . In H.D.Maschner (ed.) *New Methods, Old Problems: Geographic Information Systems in Modern Archaeological Research*, Occasional Paper No. 23, Center for Archaeological Investigations, Southern Illinois University at Carbondale, pp. 104–126.
- MAROZAS, B.A. and ZACK, J.A., 1990, GIS and archaeological site location. In K.M.S.Allen, S.W.Green, and E.B.W.Zubrow (eds) *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis, pp. 165–172.
- MASCHNER, H.D.G., 1996, The politics of settlement choice on the northwest coast: cognition, GIS, and coastal landscapes. In M.Aldenderfer and H.D.G.Maschner (eds) *Anthropology, Space, and Geographic Information Systems*, New York: Oxford University Press, pp. 175–189.
- MEFFERT, M., 1995, Spatial relations in Roman Iron Age settlements in the Assendelver polders, The Netherlands. In G.Lock and Z.Stan i (eds) *Archaeology and Geographical Information Systems*, London: Taylor & Francis, pp. 287–300.
- MONCKTON, C.G., 1994, An investigation into the spatial structure of error in digital elevation data. In M.F.Worboys (ed.) *Innovations in GIS: Selected Papers from the First National Conference on GIS Research UK*, London: Taylor & Francis, pp. 201–211.

- MORAD, M., CHALMERS, A.I. and O'REGAN, P.R., 1996, The role of root mean square error in the geo-transformation of images, *International Journal of Geographic Information Systems* **10**:347–354.
- PEUCKER, T.K., FOWLER, R.J., LITTLE, J.J. and MARK, D.M., 1978, The triangulated irregular network. In *Symposium on Digital Terrain Models*, St. Louis: American Society of Photogrammetry, pp. 516–540.
- RICK, J.W., 1996, Total stations in archaeology, *Society for American Archaeology Bulletin* **14**(4):24–27.
- RIPLEY, B.D., 1981, *Spatial Statistics*, New York: John Wiley.
- ROBINSON, A.H., Morrison, J.L., Muehrcke, P.C., Kimerling, A.J. and Guptil, S.C., 1995, *Elements of Cartography*, New York: John Wiley.
- USGS (UNITED STATES GEOLOGICAL SURVEY), 1996, *USGS 7.5-minute DEM Coverage* [online], URL <http://www.nmd.usgs.gov/metadata/7.5dem.html>.
- VAN KUILENBERG, J., DE GRUIJTER, J.J., MARSMAN, B. and BOUMA, J., 1982, Accuracy of spatial interpolation between point data on moisture supply capacity, compared with estimates from mapping units, *Geoderma* **27**:311–325.
- VASILIEV, I.R., 1996, Visualization of spatial dependence: an elementary view of spatial autocorrelation. In S.L.Arlinghaus (ed.) *Practical Handbook of Spatial Statistics*, Boca Raton: CRC Press, pp. 17–30.
- WARREN, R.E., 1990, Predictive modeling of archaeological site location: a case study in the Midwest. In K.M.S.Allen, S.W.Green, and E.B.W.Zubrow (eds) *Interpreting Space: GIS and Archaeology*, London: Taylor & Francis, pp. 201–215.
- WEIBEL, R. and HELLER, M., 1991, Digital terrain modeling. In D.J.Maguire, M.F.Goodchild, and D.W.Rhind (eds) *Geographical Information Systems: Principles and Applications*, Vol. 1, New York: Longman Scientific & Technical.
- WIEMER, R., 1995, Another way to deal with maps in archaeological GIS. In G.Lock and Z.Stan i (eds) *Archaeology and Geographical Information Systems*, London: Taylor & Francis, pp. 301–312.
- WOOD, J.D. and FISHER, P.P., 1993, Assessing interpolation accuracy in elevation models, *IEEE Computer Graphics and Applications* **13**(2):48–56.
- WORBOYS, M.F., 1995, *GIS: A Computing Perspective*, London: Taylor & Francis.

CHAPTER EIGHT

The State of the Art in ‘Inductive’ Predictive Modeling: Seven Big Mistakes (and Lots of Smaller Ones)

JAMES I.EBERT

Archaeologists have always been intrigued by new analytical directions coupled with technical means, be they stratigraphy, seriation, radiocarbon dating, remote sensing, and now GIS, as avenues to hopefully, rather automatically, understand the “past” using contemporary data. It can be argued that archaeologists have often used such data and analytical methods uncritically, without thinking very carefully about the real goals they must pursue, the meanings they must bring to the data, or the implications of their work in anthropological terms regarding prehistory. Predictive modeling cannot be an effective technique until we determine what factors drew prehistoric populations to certain places at certain times, until we begin to acknowledge the systemic activities of the people who made these sites. This requires some complex thinking and not just “camping thinking” among archaeologists.

8.1 INTRODUCTION

The topic of predictive modeling has commanded considerable interest within the profession of archaeology for at least a decade and perhaps somewhat more. A number of archaeologists have pursued predictive modeling relentlessly, and a few have even built most of their repertoires and reputations around it. A number of them have insisted that “inductive” predictive modeling, which consists basically of finding correlations between site locations and the proximity of “environmental variables” taken from maps, is a valid pursuit in and of itself, and that anthropological theory or explanation need not figure in predictive modeling efforts. “Inductive” predictive modeling needs at this time to be critically examined, and its basis and results analyzed. My discussion will be structured around seven major mistakes or misconceptions which have (mis)guided “inductive” predictive modeling.

8.2

GIS IS REVOLUTIONIZING PREDICTIVE MODELING

Most recent treatises on predictive modeling begin with the assertion that there is something new about predictive modeling, and what's new about it has more than a little to do with geographic information systems (GIS). While it is true that increasing numbers of archaeologists are becoming comfortable with the concept of using geographic information systems (one hears fewer and fewer complaints about how expensive or difficult to use geographic information systems software is), and while geographic information systems can streamline correlations of data extracted from maps, they are simply analytical tools, little different from word processing software or computers in general. Predictive modeling will be transformed into a worthwhile adjunct to archaeology and archaeological thinking only by the formulation of a body of explanatory propositions linking contemporary correlations with the past. In other words, it is productive, explanatory thought, and not computers, that can potentially raise predictive modeling above an anecdotal level.¹

8.3

PREDICTIVE MODELING PREDICTS AND MODELS THE PAST

Some archaeologists clearly seem to believe that the data they study is somehow directly and automatically linked with the past, and perhaps in no other area of archaeological "specialty" is this as pronounced as in predictive modeling.² Little real introspection is required to arrive at the realization that the variables and correlations which comprise the entire substance of "inductive" predictive modeling experiments are solely contemporary. Given this, how is the "past" involved in such predictive modeling? Almost invariably, it is only mentioned in the concluding paragraphs of reports or presentations, where the presumed preferences sought by past peoples in proximity to their sites are speculated about in at best an anecdotal manner. Predictive modeling cannot be a productive archaeological pursuit without the explicit realization that statistical tests and correlations can only inform us about coincidences in the present, which must then be linked with the past through the process of explanation. The insistence that "inductive" predictive modeling can be separated or distinguished from "deductive" or explanatory predictive modeling is a clear indication not only of the fact that predictive modeling has not yet reached the point where it can make a contribution to the science of archaeology, but (in my opinion) of the level of sheer indolence among those who think we can "stop" at "inductive predictive modeling." In any event, the concept that inductive predictive modeling somehow is a pursuit valid in itself is the source of many of the misconceptions of current predictive modeling.

8.4

WHAT WE WANT TO PREDICT IS SITE LOCATION

Whereas predictive modeling is so often extolled as a new archaeological direction, it is actually wholly grounded upon that most basic and unthinking assumption of traditional archaeology: that sites—i.e., discovered concentrations of artifacts or other archaeological materials—are where people lived or conducted activities and that they are therefore appropriate “analytical units,” and thus their locations are what we want to predict. Some of the corollary assumptions of a solely site-centered archaeology almost ubiquitous in predictive modeling are:

- sites are circumscribed and independent entities;
- sites occupy only a small percentage of the landscape, so there are many more places where sites are not³ than places where they are;
- predicting where sites are is the obverse of predicting where they are not;
- sites are where people do things and nonsites are not;
- some sites are “single-purpose” and others are “multi-purpose”;
- some sites are “single-component” and others are “multiple-component.”

Site-centered, inductive predictive modeling ignores a vast body of anthropological and archaeological evidence and thought that emphasizes that people, the things they do, the places they do those things, and all other aspects of human behavior are a systemically organized whole. “Sites,” in fact, are not independent entities at all, but components of systems—and their locations are dependent upon the locations of other components of that system, including other sites. This relationship is clearly not approachable through correlational analyses based on proximity of sites to one another. Instead, activities within human systems are demonstrably scheduled and planned within another systems component, time, and the location of a settlement, and what is done there should thus depend upon where the last settlement was, and where the next will be, and what was done or is planned at those places.

Another ethnographically demonstrated component of human systems that may be even more important than sites is travel between sites, where some studies suggest the largest part of human interaction with the environment occurs. Site locations should therefore be influenced by what lies between sites within a system, and not just what is near each of them. In the course of mobility, resources are extracted from the environment and transported—not just to the next site that is occupied, but through many occupations. Where a site is located and what is done there may therefore be quite independent of what resources are located nearby. What is located where sites are not is therefore probably just as important a factor in site placement as what is found where they are.

Other ethnographic and ethnoarchaeological studies have demonstrated that most settlement and activity locations are in fact reoccupied by human groups,

sometimes for the same activities but often for a variety of different purposes, for instance seasonally. Therefore, nearly all sites (particularly where cultural debris is concentrated) should be “multicomponent” sites, as well as “multiple-purpose” settlements. Properties of the environment at different times—seasonally, as well as properties that change more slowly through time—must therefore be taken into account in explaining their locations. Simple explanations about specific resources or situations required for “single-purpose” occupations should be expected to be “successful predictors” only very rarely, if ever.

8.5 PROXIMITY TO ENVIRONMENTAL VARIABLES IS IMPORTANT

Instead of considering the dynamics of human systems, however, site-centered, inductive predictive modeling assumes that the proximity of “environmental variables” is why people have placed their sites where they did.⁴ Properties of the environment that are closer to sites are more important factors in site placement than those that are farther away. A more sophisticated-sounding way of saying this is to cite principles of “cost-distance theory” or “least effort”—basically, that people consider the trade-off between what benefit they obtain and the amount of energy they will expend in making decisions about where to travel.

While this is something that can be easily modeled or simulated using GIS techniques, it clearly is not the way people or systems “behave.” Again, aspects of time must be considered. One such aspect is time utility gained versus energy expended; for instance, rather than walking down a valley and up the next, one might travel across the intervening mountain because it saves time. Another aspect of time that physical proximity correlations cannot address is the sequencing of activities across multiple locations in space.

8.6 MAPS CONTAIN ENVIRONMENTAL VARIABLES

The next step in correlating site locations with environmental “variables” in current inductive predictive modeling assumes that maps contain information that can be directly translated into variables. There are two kinds of map data that are correlated: site locations, and everything else. How much of everything else is used in various predictive modeling experiments seems to depend, more than anything else, on how much map data is available to the researcher in previously compiled and digitally automated form.⁵ Hence, “existing” map data is used, sometimes in surprisingly uncritical fashion. One environmental variable universally employed by inductive predictive modelers is distance to water, derived by the measurement of how far it is from the dot representing a site in the database to blue lines taken from a (US Geological Survey (USGS) topographic quad. As another chapter in this book emphasizes, those blue lines

(at least in many regions) can mean many things in terms of water quality, seasonality, and the like, some of which you wouldn't even want to be close to. Other “environmental variables” uncritically used include everything else one can get from topographic maps: that is, often dozens of variables derived simply from topography, such as elevation, slope, aspect, topographic “diversity,” and the like. Researchers fortunate enough to have automated sources of other environmental data, often at places like parks or government facilities, where many scientists have undertaken mapping projects, use habitat, vegetation, soils, and other even less transparent map data, and “variables” derived from them, as well. All of these derived data are then usually subjected to multiple regression analysis, a clear indication that the initial assumption is that many are unrelated and one needs many such “variables” to predict site locations.

8.7

MAP DATA IS INACCURATE

Another component of most reports or papers detailing predictive modeling experiments nearly invariably follows the explication of the environmental variables and the specific correlational analyses performed. In the very next section, the modeler now turns about-face and roundly criticizes the map data they have used. They now opine that map data is inaccurate and therefore inhibits the accurate carrying out of predictive modeling. Maps, after all, only show accurate locations of things to within National Map Data Accuracy standards, which are all but uninterpretable and untestable for anything but cultural features on maps. Often even more rabidly condemned is archaeological data; at least until the advent of Global Positioning System (GPS) methods for recording site locations, archaeologists must have recorded the locations of many sites incorrectly, and the recording of information about what sites are and what they contain is even more biased.

8.8

THE ACCURACY OF INDUCTIVE PREDICTIVE MODELS CAN BE DETERMINED

Why do inductive predictive modelers find it necessary to denounce archaeological and “environmental” map data? The answer is that their predictive models are not as accurate as they would like, so surely, they conclude, the data are to blame.

Virtually all inductive predictive modeling adherents advocate testing to determine the accuracy of one's model. A predictive model—that is, an array of correlations, and their strengths, among site locations and environmental variables in a study area—is derived on the basis of a “sample,” i.e., part of the known site locations in the area, and then tested with another “sample,” the rest of the known site locations in the area. This is referred to as “jackknife

sampling,” and what it amounts to is a grossly inefficient way to determine if there is inhomogeneity in one’s data. The goal of inductive predictive modeling is, of course, to find invariability in the data—i.e., to arrive at accurate predictions.⁶ The success of an inductive predictive model is based on its “gain,” the degree to which it predicts that sites will be found in a small percentage of the area considered by the model; predicting that there will be sites in places there aren’t is a “wasteful” error, while predicting that there will not be sites but then it turns out there are (however one would know) is a “gross” error.

For some reason, and I have yet to determine just why this may be, the reported accuracies of inductive predictive modeling seem to hover in the 60–70% range. Perhaps nobody wants to report “success” rates only minimally higher than 50%. Perhaps it has to do with the practice of some inductive predictive modelers of assessing the accuracy of their models by superimposing two cumulative and invariably somewhat logarithmic curves—predicted site probability versus percentage of correct predictions for sites as against nonsites—and noting their intersection as indicative of the accuracy of the predictive model. Sixty to seventy percent is not really bad but it is not very good either—certainly not good enough to justify spending a lot of money on doing this sort of predictive modeling as a substitute for doing “blanket” cultural resource surveys, the major expectation that initiated interest in inductive predictive modeling in the first place.

So what is inductive predictive modeling worth? And what will it become, when perfected? My assessment is, not very much. Inductive predictive modeling, which seems to be antithetical to using ethnographic observation and theory to approach an explanatory basis for its “modeling,” is not going to get any more accurate than it is right now, whether it is done with GIS or a stack of semitransparent map over-lays on a light table. It focuses on the wrong units of analysis, sites rather than systems, and attempts to relate their locations to “environmental variables” which not only are probably not variables at all, but cannot be warranted by any theoretical argument to be effective predictors of the locations of components of systems across landscapes.

Notes

- 1 It can in fact be supportably argued that using computers may have been, thus far, detrimental to analytical thinking and explanation. Since the mid-1980s, when personal computers became generally available, the focus of most archaeological discussion, literature, and, unfortunately, aspiration has shifted further and further from theory toward method.
- 2 This may also be due to a focus on computer methods to the exclusion of theoretical thinking on the part of many archaeologists who have embraced and based their professional renown on “inductive” predictive modeling.
- 3 The places where sites are not are quite tellingly referred to by site-centered predictive modeling specialists as “nonsites.”

- 4 In fact it might be more realistic to say "why a person placed his/her site where he/she did," for correlations which overlook entirely the operation of human systems can only offer individualistic, instantaneous, anecdotal "explanations" for such correlations.
- 5 Automating others' analog data is sometimes nearly as difficult as collecting one's own "independent" noncultural data, and few archaeologists have the time to do either.
- 6 This may be the direct inverse of what a noninductive predictive modeling approach, when developed, will seek: that is, variability among correlations which needs to be explained.

CHAPTER NINE

GIS Applications in Archaeology: Method in Search of Theory

TIM CHURCH, R. JOE BRANDON AND GALEN R. BURGETT

GIS applications in archaeology have been largely confined to data visualization, simple mapping, or predictive models of site locations. We argue that as such GIS has been used in an atmosphere bereft of theory and often using data of doubtful reliability or significance. This paper critically examines traditional predictive modeling as well as outlining what we consider to be the potential of a new archaeologist's toolbox, driven by models adapted from landscape ecology, and capable of producing theory. This toolbox is composed of techniques already in use in archaeology such as remote sensing, GIS and simulation studies. What we advocate is these techniques be used as an integrated whole to generate a landscape perspective of prehistoric dynamics rather than the traditional static models.

“Contrariwise,” continued Tweedledee, “if it was so, it might be; and if it were so, it would be: but as it isn't, it ain't. That's logic.”

9.1 CURRENT USE OF GIS IN ARCHAEOLOGY

The three typical applications of GIS in archaeology have been analysis or, more rightly, visualization; management; and the development of “predictive” models. A predictive model is defined as “hypotheses or sets of hypotheses which simplify complex observations whilst offering a largely accurate predictive framework structuring these observations” (Clarke 1968:32). Correlative predictive models are those that “identify and quantify relationships between archaeological site locations and environmental variables” (Sebastian and Judge 1988:4). Explanatory predictive models are “models that are deductively derived and attempt to predict how particular patterns of human land use will be reflected in the archaeological record” (Sebastian and Judge 1988:4). We would argue that except for management applications, the vast majority of GIS studies have confused either “pretty pictures” with innovation, as in the case of visualization, or the statement of simple correlations with theory, as in the case of predictive

models. This is not to say that previous predictive modeling efforts were wasted; indeed, without these early studies many of the tools discussed here would have remained undeveloped. However, predictive modeling in archaeology has entered a period of doldrums that will be overcome only when we shift from a methods orientation to a theoretical one. GIS grew out of efforts by land managers to construct a visually linked database that would allow them to track growth, would provide a basis for planning, and would allow them to isolate areas suitable for specific activities or facilities. Geographers and biologists then began to use GIS to conduct spatial analysis of various activities or populations. With the growth of cultural resource regulations in the 1970s coupled with large-scale impact projects (e.g., strip mines), land managers were immediately interested in the promise of predictive modeling. Predictive modeling efforts developed out of a desire on the part of land managers to categorize their lands as to the likelihood of site presence. There was an initial hope that lands placed in a low-likelihood category could be essentially exempted from further investigation. This attitude was quickly tempered into the present goal of flagging areas with a high likelihood for site presence as an aid in management decisions rather than as a way to exempt lands.

9.2

CORRELATIVE PREDICTIVE MODELS

The correlative model is the preferred model type in archaeology for a couple of reasons. The first is that it uses existing data. The second related factor is the modest time investment needed (still considerable) to compile and computerize the information. However, we believe that archaeologists need to take a cold, hard look at the precision, accuracy, and nature of the data we have collected (and continue to collect). Previous predictive modeling has, for the most part, relied upon indirect measures and/or implied cultural value of environmental variables. Thus, some models have established a correlation with certain soil types, suggested to result from a particular soil being a better resource (for a reason not always made clear) than the other soils present in the area. Adoption of these types of indirect measures may be expedient and was logically argued as a limitation of the resolution of the environmental data available for early predictive models. As Winterhalder has stated:

Typological thinking remains commonplace in Anthropology and permeates descriptions of the environments to which humans adapt. This paper has argued that normative description using spatial and temporal averages of environmental factors destroys the information necessary to analyze human adaptations...Environmental description must be suited to theory and adjusted to the spatial and temporal scales apposite to the organism (population) and function being studied.

(Winterhalder 1980:163)

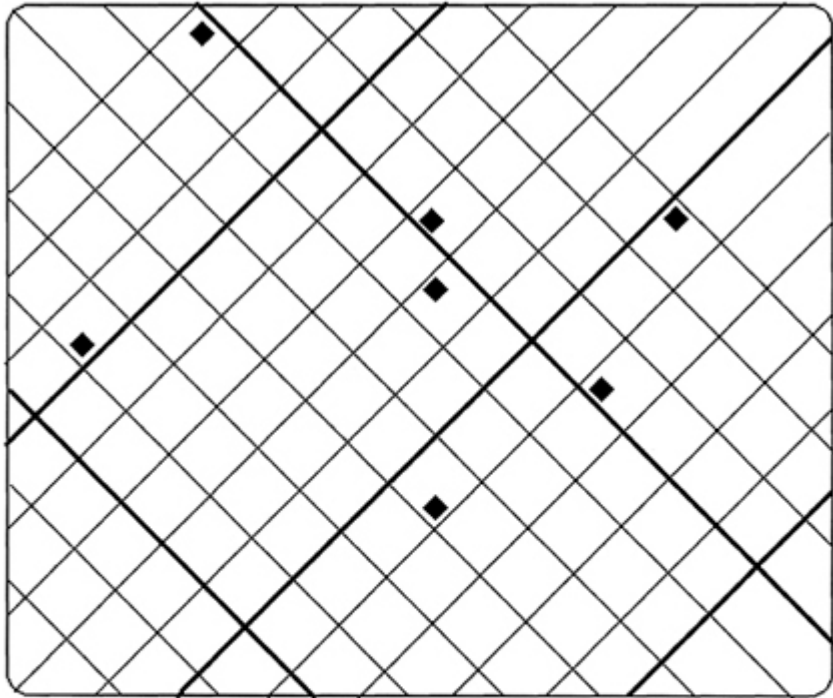
Expediency should not be an acceptable reason for anything other than initial, exploratory studies, and new technology has made environmental data available at a much higher resolution. In discussing the overlooked potential of GIS, Tosta states, “data are only the building blocks of information and ultimately of knowledge. Data alone do not provide understanding. It’s only when we put data in context that we create information to provide new understanding” (Tosta 1991: 46). The key variables selected for most correlative models are identified from the location of sites based on their statistical significance. The problem with many of these statistical studies is that they are only confirmatory in nature; that is, they seek to confirm preexisting statements. Butzer warns that “Proper statistical processing is an essential component of scientific research, but only as a means to an end. When the intellectual framework is too narrow, the results, no matter how elaborately programmed, cannot hope to allow high level interpretations” (Butzer 1978:191). Thus, a number of studies correlate such variables as distance to water, slope, aspect, relief, etc. with the location of sites. Use of these variables actually limits the model to addressing only issues of shelter and regards variables such as slope, aspect, etc. as environmental when in fact they are merely measures of terrain. There is also, however, no explicit explanatory component to these models. As Kvamme points out, “Without knowledge of the nature of the background environment, however, there is no way to ascertain whether the tendency [of certain environmental variables to correlate with site locations] is a result of real locational patterns exhibited by the sites or is merely a reflection of the nature of the background environment on that variable” (Kvamme 1990:367). This is key to understanding the limitations of correlative models, and one that has long been recognized in urban planning and suitability analysis. Such models are more appropriate for locating suitable campsites for contemporary hikers than for predicting prehistoric site locations, let alone understanding prehistoric behavior.

To give a contemporary example, suppose we wish to predict the best location for a new fast food business. In a correlative model we would look at the locations of existing fast food businesses. In doing so we would probably find high correlation with major streets. We would therefore conclude that the key factor for locating a fast food business would be distance to the nearest major street, with locations on major streets being most desirable. This can be simply demonstrated in [Figures 9.1–9.4](#). If we collect data about the locations of existing fast food businesses we will conclude, as stated previously, that there is a high correlation to streets ([Figure 9.1](#)). If, however, we change our perspective and look at the streets (landscape) we will conclude that there is a low correlation between streets and fast food businesses ([Figure 9.2](#)). This is a landscape perspective in a geographic sense but remains essentially descriptive in nature and continues to ignore the influence of a whole host of additional factors or resources such as residential income, zoning, parking, etc. More important are the interpretative limitations imposed by this perspective. Again, the interpretation based upon a correlative model is one that does little to tell why, in

explicit terms, locating a business along a major street is important. Figure 9.3 presents the same area but with additional resources and variables considered, and Figure 9.4 portrays the same model with time depth, providing a true dynamic landscape. With this information explanatory statements can be formed. The difference is in a contextual perspective where locations are isolated from their environment to one where locations are specifically put into a dynamic environment. Spurious correlation occurs when key resources are not recognized, and when they are recognized their parameters are not adequately established. That is, there is only an assumed relation to the truly critical resources of water, shelter, food, and lithics. A landscape location may be perfect in the sense that it is flat and south-facing, but if no exploitable resource is nearby there would be no reason to use the spot. Landform attributes may prove useful if we use them to further refine the basic model, but they do not in themselves provide a suitable foundation for modeling. For example, if we know, or can deduce, that the foothills of a mountain range contained a critical resource, say good populations of mule deer during the fall, then determining those specific areas suitable for campsites within this area could provide powerful predictive statements. This type of model has been termed a synoptic model which is based on “regional combination of variables relevant to archaeological site locations rather than specific site locations” (Custer *et al.* 1986). Synoptic models are rare and to date have also used only broad environmental classes.

Second, all the predictive models to date have relied on variables expressed in the contemporary environment. There is an acknowledgment that these may or may not reflect prehistoric conditions, but there is little else. To expect a model based on present-day environmental conditions to be of use to modeling the site locations of say, Paleo-Indian sites, is a tenuous assumption at best. The dynamics of the paleoclimate and its impact upon other environmental parameters need to be understood in order to provide a robust data foundation for any predictive model. Butzer observes that “archaeologists often take a static, classificatory approach to the environment, even when the human variables happened to be considered part of a dynamic system. It is my belief that ‘environment’ should not be synonymous with a body of static and descriptive background data” (Butzer 1980:418). The present-day environment is a good place to start, but a poor place to end.

Finally, the scale of the environmental data in almost all previous predictive models is broad—so general, in fact, that any recognized correlation may be spurious, more happenstance than fact. Much of this environmental information is at a scale of “zones.” Within these broad zones important resources are implicitly assumed to be uniformly distributed through time and space. This is the key weakness in the use of these broad categories, as it is demonstrable that important resources are almost never uniformly distributed spatially or temporally. There is often a passive acceptance of these correlations by archaeologists, particularly in a normative approach, and there is typically little discussion as to alternative reasons for the correlation. As Kvamme has stated:



— Minor street ◆ Fast Food business
— Major street

Problem: where should a new fast food business be located?

Data: drawn from neighborhood (environment).

Analysis: low correlation between streets and fast food locations, existing businesses concentrated in certain areas along or near major streets.

Interpretation: distance to major street is a contributing variable but other variables, as yet identified, probably are as, or more important.

Cause: data incomplete.

Action: further analysis of population structure, household income, city growth plans, etc.

Figure 9.1 Simplification of a landscape model.

A wide range of environmental phenomena, including hydrographie, landform, soil, and vegetation characteristics, has been examined for possible relationships with the immediate locations of prehistoric sites. These studies, however, have usually failed to offer objective evidence that the environmental phenomena examined are actually related to the presence or absence of sites.

(Kvamme 1985:208)

And as Keene has pointed out:

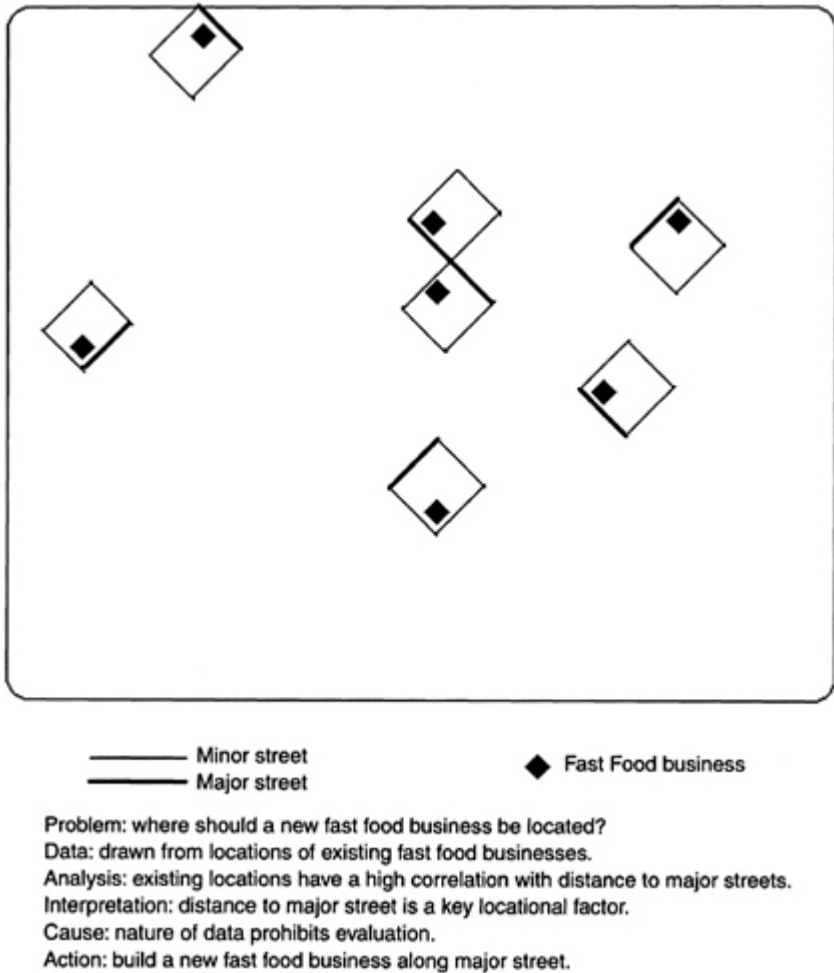


Figure 9.2 Simplification of a traditional correlative model.

Hence, we look not at adaptations to the highs and lows or the range of variability possible, but to averages or most common events. This practice tends to homogenize behaviors and environments and obscures any real variability we hope to understand.

(Keene 1983:147)

Traditional predictive modeling also only addresses residential site types to the almost total exclusion of other site types (such as lithic procurement, rock shelters, or rock art sites (see Bradley *et al.* 1994 for a study of rock art with a landscape perspective)). As these site types are often of regulatory or research

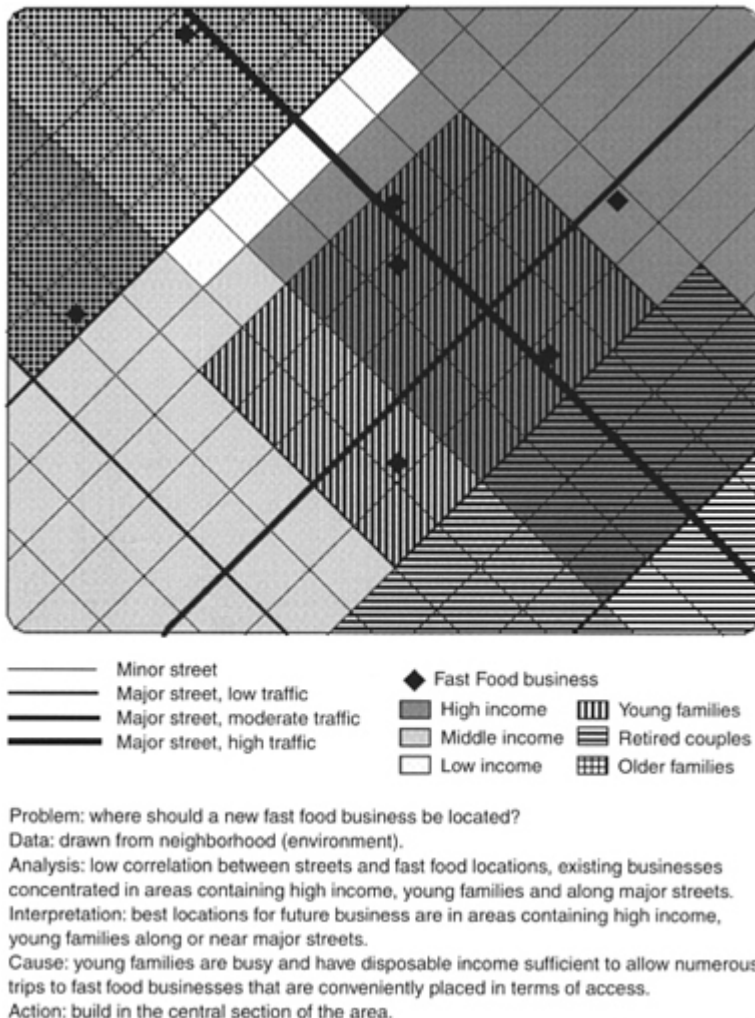


Figure 9.3 Landscape model with additional variables.

significance, overlooking them can limit the value of any modeling effort no matter the objective.

In summary, traditional correlative models can be questioned because of their reliance on inaccurate data, imprecise data, incomplete data, and static data. There is an axiom in data modeling, “the model is only as good as the data.” Prehistoric site data generated from inventory projects is inaccurate, spotty, and biased. Uncoordinated fact-gathering, as practiced under CRM, is not an adequate base for modeling prehistoric population behavior. Much of this data is reflective of the needs imposed by either management considerations or a culture

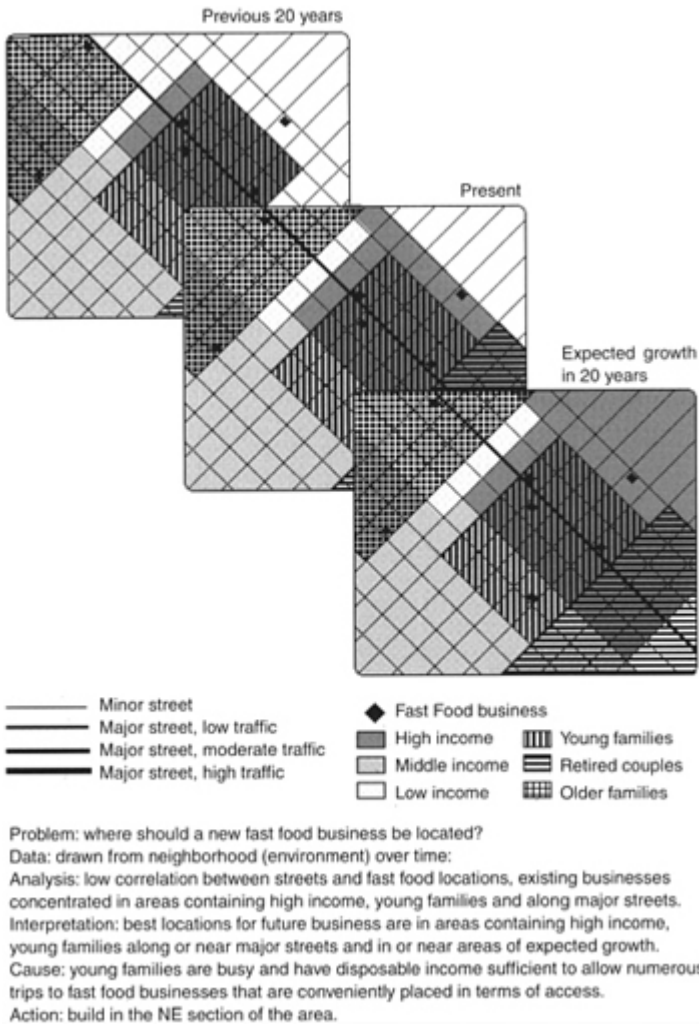


Figure 9.4 Landscape with temporal components.

history paradigm. Therefore, data currently gathered during cultural resource management projects is often unsuitable for landscape model building in a processual, explanatory paradigm. It is not that we do not have the tools to make major advances in the use of GIS, but, rather, we lack an integrated methodology incorporating these tools. We also lack a theoretical perspective within which to operate that enables us to recognize and gather valid and useful data.

9.3

THE RESOURCE LANDSCAPE

People move around a landscape in relation to a number of resources. The value of these resources in a cultural system is dynamic, as can be the resource itself. The term "resource" has been misused in a number of studies. In describing optimal location strategies Wood used Judge's early attempt at analyzing the locations of Paleo-Indian sites identified as overviews, trap areas, etc. as resources and assigned a value to these resources based on the distance to the "resource" (Wood 1978). In reality, what Wood was describing was travel cost to terrain features which were assumed to be valuable in obtaining the true resource. A predictive model using combinations of resources available across space and through time and related to adaptive value will be much more flexible in determining the influence of these resources upon populations and will provide a platform for explanatory interpretation rather than mere correlation.

The role of resource distributions and the impact of their dynamic nature upon populations has increasingly come under discussion in ecology and elsewhere (e.g., Cushman *et al.* 1988; Houston *et al.* 1995; O'Neill *et al.* 1988; Slobodchikoff 1984; Tilman 1980). Resources are four-dimensional in that their length, breadth, depth, and duration (time available) can be measured. Further, measures of these dimensions will shift through time due to disturbances. Disturbances can be caused by climatic, geomorphological, or cultural factors in a dynamic web of feedback loops. Key variables are suggested to include water, shelter, botanical, animal (food), and lithic resources.

By any model's measure water is a key resource. Despite this, water is often dealt with in a superficial, implicit manner in predictive models. Water is available naturally as rain or snowfall (short-duration area sources), in lakes (long-duration area sources), playas (short- to moderate-duration area sources), rivers and streams (moderate- to long-duration linear sources), springs/seeps (moderate- to long-duration point sources), and arroyos (short-term linear sources). The duration and spatial availability of water was enhanced by prehistoric populations through the construction of check dams, reservoirs, irrigation systems, and wells (e.g., Crown 1987; Evans 1951; Green 1962; Mobley-Tanaka *et al.* 1995; Scarborough 1988). Water availability is also directly influenced by climatic change (Hay *et al.* 1993).

Landscape features such as depressions and drainages are assumed to indicate the presence of water resources without regard to the duration (short and long term) of the water, its quality, or geomorphological and geological factors influencing its availability (an exception is Jackson 1988). Beyond this there is often only a descriptive, speculative statement made as to how these conditions may have limited use. As Jackson states:

However, the effectiveness of predictive settlement models is limited in this regard because although the close spatial association between

archaeological sites and water sources is recognized, the models cannot actually predict where adequate water sources exist apart from perennial streams and springs depicted on topographic maps.

(Jackson 1988:227)

Shelter resources are often implicitly dealt with in traditional predictive models with measures such as slope and aspect. These are in reality measures of terrain, not environmental variables. While certainly they are factors in the choice of shelter sites, other factors such as wind direction, exposure to sun and rain, temperature fluctuations, and defensive value can also be expected to condition choice and can be modeled with GIS (Lapen and Martz 1993). The location of caves and rock shelters as potential shelter sites is almost always ignored (an exception is Hall and Klippel 1988). Unfortunately this study falls into the same trap of trying to predict locations of specific features without the model containing any of the natural mechanisms that form these features (e.g., Thrailkill 1968).

A narrow view of food (botanical and animal) resources is often evident in those few previous predictive models that have addressed these resources. Besides the obvious use as a source of food, botanical resources can also provide fuel, tools, and construction materials. In certain cases, such as arrow shafts, very specific species of plants were preferred. Simulation models have often taken an optimizing economic viewpoint which focuses on food resources and their "cost" to procure versus their return in terms of calories (e.g., Lee 1969). However, others have placed this type of optimization analysis within a dynamic framework (Laferriere 1995) or an ecological one (e.g., Osborne 1993). Ecologists have used simulation to model animal behavior (e.g., Folse *et al.* 1989; Turner *et al.* 1993). Within archaeology there has been only limited investigation of the impact of dynamic food resources upon subsistence strategies (e.g., Walsh 1988), while GIS studies have started to appear that would provide the necessary data (Walker 1990). As Bamforth correctly notes, identifying the species of plants and animals exploited by the people in a region provides only part of the information needed to assess the relationship between environment and human adaptation. Thomas, Winterhalder, and McCrae have argued that human beings adapt to the overall spatial and temporal pattern of resource abundance and scarcity in a region and to the nature and degree of variation in this pattern rather than just to specific species of plants and animals found there. These authors argue that many anthropological explanations are irrelevant to societies other than the one for which they were initially formulated. This is because they rely on specific environmental characteristics, such as absolute temperatures or the species of plants and animals available, thereby obscuring structural similarities between superficially diverse regions (Bamforth 1988:16).

The exploitation of special use sites, such as lithic procurement sites, is conditioned by fewer environmental variables, fundamentally the presence of

suitable stone, as well as other factors such as accessibility, abundance, and quality. These factors have a high degree of stability; for example, except in the case of exhaustion of the source or burial by natural forces, most of these remain constant and may override other considerations such as the availability of water, etc. That said, it is important to acknowledge that the lithic environment is variable; that is, a model developed for one area may not be applicable in another. Custer *et al.* (1983) broached the subject of how differential distribution of lithic resources might result in different procurement systems. Gould and Sagger (1985) also used differential lithic resource distribution in an attempt to explain differences in material use at prehistoric Australian sites. Subsequent to that, other investigators attempted to hypothesize how this differential landscape would be reflected in archaeological assemblages.

Resource potential models have been extensively used in geology in helping to determine areas that have a high potential for the presence of minerals (e.g., Agterberg 1974; Griffiths 1978; Pan and De Harris 1992; Reddy and Koch 1988). Mineral resource potential has been defined in the literature of geology as “a measure of the likelihood of occurrence of valuable minerals or minerals that may become valuable within the foreseeable future” (Taylor and Steven 1983: 1268). These models are somewhat like the predictive models of site locations in archaeology. As with site predictive models, a resource potential model calculates the potential of any given area to contain the resource under investigation. These models range from strict statistical models to heuristic ones (Kliem and Petropulos 1990). The basic difference is that a resource model is environmentally based while a site predictive model is culturally based. Another difference is the scale: the typical geologic model has a scale from 4 to 10 km. A model that will be useful to the archaeologist must have a maximum scale of 1 km².

Lithic sources offer several distinct advantages for modeling. The first is that they are unmoving and relatively unchanging. They can only be physically depleted. This also means that they are present today and have been investigated and mapped to some extent by geological studies, thus providing a reliable base of information. Further, desired sources of chipped stone are a relatively rare occurrence in the landscape. Second, stone was a consistently sought resource until the introduction of metal. Prehistoric people could often decide to shift to other, comparable resources if needed. They could shift from big game to small game or from a game emphasis to gathering emphasis. No such option existed in terms of the stone needed to perform many day-to-day activities, although they might shift emphasis within the narrow set of types of stone.

Lithic procurement sites are unique in several ways:

- 1 Their value can override other environmental variables. A source of excellent material will be exploited despite being on a steep slope far from water, for example.

- 2 They are an essentially stable resource. They do not move and are subject to only exhaustion or burial.
- 3 Their occurrence, both vertically and horizontally, is fixed.
- 4 They are an extremely localized resource. Although some areas extend for miles, most areas are smaller than 1 acre (0.4 ha) in size.

Predictive modeling of lithic resources in archaeology was first proposed by Fanale in 1973, but very little additional work has been completed until recently. Predictive models of potential lithic resources have been done for the Appalachian Mountains (La Porta 1995) and on a smaller scale for the Bearlodge Mountains of Wyoming (Church 1996). In addition to these predictive models, a number of geological studies have made use of remote sensing and image analysis techniques to identify deposits or establish the spatial distribution of a number of rock types of interest to archaeology. These include silcrete (Densen and Peterson 1995), jasperoid (Murphy 1995), porcellanite (Clark 1981), silicates (Hunt and Salisbury 1970; Hunt *et al.* 1973), and sandstone (Vincent *et al.* 1972), as well as others.

Prehistoric people did not normally choose activity locations on the basis of the presence of a single resource, but rather on the presence of multiple resources and their perceived value at that time. Thus, an area may have the attributes of an excellent campsite in that the land is well drained, sheltered from wind, and warmed in the morning and cooled in the afternoon. But if other resources were not in proximity, then the most perfect campsite probably would have remained unused.

9.4

AN ALTERNATIVE: EXPLANATORY MODELS FROM A LANDSCAPE PERSPECTIVE

So, do we now have a suite of tools able to provide data of a suitable nature and a conceptual framework with which to intelligently structure this data and explore some of the hypotheses that have been proposed in the previous 30 years? We would answer yes, with the conceptual framework provided by a landscape perspective.

A truly useful predictive model, from both a management and a research perspective, is one that puts human use of the area into this environmental context. Not only will it define those environmental variables or combinations of variables that would attract human use and thus predict site location, but also it will address post-site formation processes that could obscure or destroy these sites. Butzer echoes this view by stating, “the primary goal of environmental archaeology should be the characteristics and processes of the biophysical environment that provide a matrix for, and that interact with, socio-economic systems as, for example, reflected in subsistence activities and settlement patterns” (Butzer 1980:419). Landscape is defined here as mosaics of temporally

and spatially dynamic resource patches in which ecological, geomorphological, and cultural systems operate at various scales. It is not an aggregate of types of sites, nor is it simply a large area. Much of what is discussed in this chapter is neither new nor unique. The chapter's perspective has drawn heavily upon the work of L.R.Binford, B.Winterhalder, L.Wandsnider, D.B.Bamforth, J.Ebert, K.W.Butzer, and A.Osborne in archaeology and R.T.T.Forman, B.Milne, R.H.Gardner, and R.V. O'Neill in ecology. What this chapter does offer, hopefully, is a holistic view in terms of both theoretical context and application methodology. That said, we do not advocate adoption of ecological concepts, but rather a critical exploration and adaptation.

An explanatory model seeks to establish dynamic relationships between variables: "they [explanatory models] are models that attempt to build the bridge between the dynamics of the living system and its observed outputs" (Kohler 1988:37). And as McGlade argues, the real questions that concern us, therefore, are those relating to change, to discontinuous transition, and ultimately to the structuring and restructuring of the socionatural environment. Such issues are fundamentally concerned with causality—and especially with the nonlinear dynamic processes and counterintuitive feedbacks that structure complex socionatural systems (McGlade 1995:114).

There might appear to be little real difference between the two approaches, but a couple of example statements point out the fundamental difference. Within a correlative model a statement such as "sites are located on level to moderately level slopes composed of soil type A and that are within X distance from water" is typical. In an explanatory framework a statement such as "residential hunting sites dating to the Archaic period are located in the foothills between elevations of X to Y on moderately level slopes and within X distance of water because the resources of food (mule deer) and water are present and these resources form the most reliable subsistence base available at that time" is possible. Further, prehistoric use of this area can be expected to fluctuate with the availability of resources, many of which will themselves fluctuate due to changing environmental conditions. Dincauze summarizes this point by stating, "The observation of a correlation is the beginning of the search for mechanism; the end is explanation of a relationship" (Dincauze 1987:319). The basic factors that have limited the number of explanatory models are the need for more detailed environmental data, a misunderstanding of the scope and nature of the needed initial theoretical foundation, and a basic passivity on the part of archaeologists in developing explanatory modeling techniques.

9.5

TOOLS TO TACKLE LANDSCAPES

The landscape tools include geographical information systems (GIS), remote sensing, and nonlinear simulation modeling. The exponential growth in the power of personal computers has helped to revolutionize the availability of GIS

applications for archaeologists. GIS programs that just a few years ago would have required a Unix workstation are now becoming increasingly available on the platforms of personal computers. Though possibly not as robust as a Unix-based GIS, all current PC GIS applications are capable of performing the basic functions of GIS well. The marriage of the landscape perspective with GIS (Johnston 1990) does not necessarily require the latest technology. Indeed, some GIS modules providing specialized functions of landscape analysis already exist (Baker and Cai 1992).

Likewise, there is an abundance of remotely sensed data available to the researcher in any part of the globe at almost any scale. As Roughgarden *et al.* point out, "Two of the most important tools for extrapolating understanding from local to regional scales are remote sensing and computer simulation modeling, two very synergistic technologies" (Roughgarden *et al.* 1991:1919–1920). The only real drawback to this flood of data is determining what is applicable to our specific research interests, including landscape analysis (Quattrochi and Pelletier 1991).

The vast majority of simulations in archaeology have been based on an economic optimization perspective and while some included ecological data the models were driven by economic concerns. Further, most, if not all, of these studies were linear, some limited to a Newtonian mechanical approach (Reidhead 1979). Since the early 1980s simulation studies in archaeology have almost ceased. This is in contrast to ecology, where simulation studies are now viewed as an important tool in understanding ecological dynamics. Indeed, a whole journal is devoted to the subject (*Ecological Modelling*). Modeling procedures are more thoroughly discussed by Aldenderfer (1991), Doran (1970), and Swartzman and Kaluzny (1987).

It is time to reassess the role of GIS, remote sensing, and simulation studies, not as techniques isolated from others but as an integral part of a triad of methods. By combining these three elements we can achieve dynamic, explanatory models of prehistoric behavior of value not only in a scientific context but also in a management context.

9.6

THE TEMPORAL DIMENSION

Time is the ultimate dynamic force. "One of the most challenging problems for both the experimental ecologist and the theorist is the interaction of biological populations with the time dimension" (Garsd 1984:199). In those predictive models that implicitly deal with time through the use of cultural sequences its impact is minimized to general observations. To more explicitly include time in modeling requires substantial methodological (Dunn *et al.* 1991; Johnson 1990; Langran 1992) and theoretical consideration. The traditional predictive models are based on large-scale environmental variables against which are superimposed a generalized cultural structure with only an implicit temporal structure. While

some have voiced the need to add in a temporal aspect to the mix, this too is at a gross level and has resulted in statements such as “during period X more sites are within landscape classification Y than during period Z.” Such statements are descriptively shallow and are contained in a vague conceptual twilight zone that denies the possibility of testing the statement. They are in essence a quantification of common sense rather than a true model.

9.7 SCALE AND GRAIN

An understanding of scale is essential to any attempt in model-building. “The problem of relating phenomena across scales is the central problem in biology and in all of science” (Levin 1995:311). Scale is not confined to the mere measure of geographic distances or in the number of “things” as typically recognized in archaeological studies but is also inherent in theory and method (Woodcock and Strahler 1987). It is also an inherent part of all natural systems, and all such systems operate on multiple scales (e.g., Clark 1985). Chronological resolution must be factored into models applied to archaeology (Jones and Beck 1992). Chronological grain is determined by the resolution of the data produced by the various dating methods. As radiocarbon dating is the most accepted and widespread, the chronological grain of most studies will be limited to 500 years (multigenerational).

9.8 THE ROLE OF CLIMATE

Climatic variation is the independent variable and one that has long been recognized as shaping the nature and compositions of ecosystems (Malanson 1993; Yeakley *et al.* 1994). It is the driving force behind the approach advocated here, and one that has not been incorporated into previous predictive modeling efforts, although it has been modeled in a GIS environment for ecological studies (Baker *et al.* 1991). This is not to say that numerous authors have not postulated a correlation between prehistoric adaptations and shifts in climate (e.g., Metcalfe 1987). However, most of these correlations are built upon generalized models of climate (Bryson 1994) rather than small-scale climate fluctuations (Matyasovszky *et al.* 1993; Strandman *et al.* 1993). “Traditional averaging procedures (annual average, climatic normals) have both theoretical and practical failings when used for the detection of small climatic changes” (Burt 1986: 279). This is not to say generalized models are not valuable; indeed, they provide the necessary framework which structures the development of smaller-scale models.

9.9 GEOMORPHOLOGICAL PROCESSES AND THEIR IMPACT

The geomorphological processes that affect the distribution of resources and, post-depositionally, the archaeological record, must also be taken into account (Brakenridge and Schuster 1986; Kirkby and Kirkby 1976). The initial measurement of the current geomorphological surface can be characterized using a number of remote sensing techniques (e.g., Connors *et al.* 1987; Pickup 1985; Pickup and Nelson 1984). Although a complex system, geomorphological modeling must be an integral part of the modeling process (e.g., Anderson 1988; Khanbilvardi and Rogowski 1986; Pickup and Chewings 1986), and we are heartened by the recent attempt to place these processes into a landscape perspective (Stafford 1995).

9.10 SUMMARY

- 1 We advocate that the body of theory and methods that have come to be termed “landscape ecology” has much to offer to the study of prehistoric populations. We are fully cognizant of the pitfalls in borrowing from other disciplines. In regards to this we agree with Keene, who stated, “The source of the problem is borrowing without modification and a tendency to adopt rather than adapt” (Keene 1983:142). However, we find that arguments against any use of ecological method and theory in archaeology are provincial and arrogant.
- 2 We accept the argument that, as societies have developed, the constraints imposed by the surrounding environment have been increasingly mitigated by cultural responses. However, we believe that during almost all of North American prehistory ecological forces have limited and shaped prehistoric population activities to a substantial degree. We therefore argue that an understanding of the ecological system and its interaction with the geomorphological and culture systems is essential to interpreting the archaeological record.
- 3 Arguments to the effect that human behavior is too complex to model and, therefore, any attempt to model cultural systems will be so generalized as to be useless are nihilistic and passive in viewpoint. No one is denying the complexity of the task, but that is the challenge, not an excuse.
- 4 We strongly believe that the current additive strategy in archaeology where information from a number of points or sites is used as the base to build a picture of regional prehistory is theoretically shallow, methodologically costly, and ultimately misleading. The archaeological record is more profitably used to validate hypotheses generated by models than as a basis for model-building itself.

- 5 Exploration of remote sensing, geographic information systems, and simulation modeling by archaeology during the past 20 years has been fruitful and necessary. However, much of the potential value of these methods remains limited, not by the techniques themselves, but by the theoretical perspective (or lack of one) that they have been operating in to date.
- 6 The role of scale, in terms of both phenomena and inquiry, in understanding these systems cannot be underestimated and must be explicitly addressed.
- 7 The challenges looming before us are in identifying those variables in the archaeological record that will provide the means to test these models and in the conceptualization of measurement units that relate those variables to the models.
- 8 The key factor that makes this proposed approach superior to traditional correlative models is the flexibility of data. Traditional predictive models are generalized models of static variables. With the predictive model structure proposed here, managers can generate data that incorporates temporally variable aspects, postdepositional processes that might obscure, alter, or destroy the archaeological record, as well as flag areas having a high probability of sites, including those more specialized sites that are often ignored in traditional predictive models. "Understanding patterns in terms of the processes that produce them is the essence of science and the key to the development of principles for management" (Levin 1995:278). We agree with Smith that "an adequate theory of adaptation must not only enable us to identify adaptations and adaptive processes, describe and measure them, and predict responses from specified environmental alterations; it must also provide solid deductive explanations for the existence of specific adaptive responses and general adaptive capabilities" (Smith 1979:56).

9.11

CONCLUDING REMARKS

Archaeologists' goal has been to assemble sufficient pieces of prehistoric life from archaeological sites to generate synthetic theories of behavior. After over 50 years of fact-gathering we have remarkably little synthetic, let alone explanatory, theory to show for our efforts. We liken this approach to trying to assemble a 10000-piece jigsaw puzzle without the box top. For archaeology to advance past mere description we need to look at the box top first, to put archaeological remains into context (as Butzer and others have argued). Costanza and coworkers put it quite well in stating:

Systems are groups of interacting, interdependent parts linked together by exchanges of energy, matter, and information. Complex systems are characterized by: (1) strong (usually nonlinear) interactions between parts; (2) complex feedback loops which make it difficult to distinguish cause

from effect; (3) significant time and space lags; discontinuities, thresholds and limits; all resulting in (4) the inability to simply “add up” or aggregate small-scale behavior to arrive at large-scale results.

(Costanza *et al.* 1993:545)

By combining these three technologies under a theoretical umbrella drawn from landscape ecology we can begin to examine issues of transportation costs (e.g., Brannan 1992; Jones and Madsen 1989; Rhode 1990; Soule and Goldman 1972) and travel corridors (Rice 1993), site exploitation territories (Bailey and Davidson 1983), the conditioning role of resource variation across space and through time upon human populations (e.g., Clapham 1976; Harpending and Davis 1977), the effects of ecological fluctuations on material culture (Fernstrom 1984), and adaptive responses to these changes (e.g., Pate 1986).

The approach we advocate is not new, being grounded in landscape ecology and archaeology and derived from the works of others. What we hope we have presented is a refinement of these pioneers’ approach and the advocacy of a suite of methods and tools that have the potential of providing an analytical basis for exploiting the landscape perspective. However, we must echo Thomas’s warning that “Blind acceptance of modeling results from the bowels of the computer can be as irrational as reliance on the honored and ancient skills used by the oracles in deciphering messages in the entrails of a sacrificial chicken” (Thomas 1986: xxi).

References

- AGTERBERG, P.P., 1974, Automatic contouring of geological maps to detect target areas for mineral exploration, *Mathematical Geology* **6**(4):373–395.
- ALDENDERFER, M., 1991, The analytical engine: computer simulation and archaeological research, *Archaeological Method and Theory* **3**:195–247.
- ANDERSON, M.G., 1988, *Modeling in Geomorphological Systems*, New York: John Wiley.
- BAILEY, G.N. and DAVIDSON, I., 1983, Site exploitation territories and topography: two case studies from Palaeolithic Spain, *Journal of Archaeological Science* **10**:87–115.
- BAKER, W.L. and CAI, Y., 1992, The r.le programs for multiscale analysis of landscape structure using the GRASS geographical information system, *Landscape Ecology* **7** (4): 291–302.
- BAKER, W.L., EGBERT, S.L. and FRAZIER, G.F., 1991, A spatial model for studying the effects of climatic change on the structure of landscapes subject to large disturbances, *Ecological Modelling* **56**:109–125.
- BAMFORTH, D.B., 1988, *Ecology and Human Organization on the Great Plains*, New York: Plenum Press.
- BRADLEY, R., BOADO, F.C. and VALCARCE, R.F., 1994, Rock art research as landscape archaeology: a pilot study in Galicia, north-west Spain, *World Archaeology* **25**(3): 374–390.

- BRAKENRIDGE, G.R. and SCHUSTER, J., 1986, Late Quaternary geology and geomorphology in relation to archaeological site locations, Southern Arizona, *Journal of Arid Environments* **10**:225–239.
- BRANNAN, J.A., 1992, On modeling resource transport costs: suggested refinements, *Current Anthropology* **33**(1):56–60.
- BRYSON, R.A., 1994, On integrating climatic change and culture change studies, *Human Ecology* **22**(1):115–128.
- BURT, J.E., 1986, Time averages, climatic changes, and predictability, *Geographical Analysis* **18**(4):279–294.
- BUTZER, K.W., 1978, Toward an integrated contextual approach in archaeology: a personal view, *Journal of Archaeological Science* **5**:191–193.
- 1980, Context in archaeology: an alternative perspective, *Journal of Field Archaeology* **7**:417–22.
- CHURCH, T., 1996, Lithic resources of the Bearlodge Mountains, Wyoming: description, distribution and implications, *Plains Anthropologist* **41**(156):135–164.
- CLAPHAM, W.B., Jr, 1976, An approach to quantifying the exploitability of human ecosystems, *Human Ecology* **4**(1):1–30.
- CLARK, G.R., 1981, The distribution of lithic material sources of coal burn origin in southeastern and east-central Montana: an application of remote sensing data to an archaeological problem, Paper presented at the 39th Plains Anthropological Conference, Bismarck, ND.
- CLARK, W.C., 1985, Scales of climate impacts, *Climatic Change* **7**:5–27.
- CLARKE, D.L., 1968, *Analytical Archaeology*, London: Methuen.
- CONNORS, K.F., GARDNER, T.W. and PETERSEN, G.W., 1987, Classification of geomorphic features and landscape stability in northwestern New Mexico using simulated SPOT Imagery, *Remote Sensing of Environment* **22**:187–207.
- COSTANZA, R., WAINGER, L., FOLKE, C. and MALER, K., 1993, Modeling complex ecological economic systems: toward an evolutionary, dynamic understanding of humans and nature, *BioScience* **43**(8):545–555.
- CROWN, P.L., 1987, Water storage in the prehistoric Southwest, *Kiva* **52**(3):209–228.
- CUSHMAN, R.M., FARRELL, M.P. and KOOMANOFF, F.A., 1988, Climate and regional resource analysis: the effect of scale on resource homogeneity, *Climatic Change* **13**:129–148.
- CUSTER, J.F., CAVALLO, J. and STEWART, R.M., 1983, Lithic procurement and Paleoindian settlement patterns on the middle Atlantic coastal plain, *North American Archaeologist* **4**(4):263–275.
- CUSTER, J.F., EVELEIGH, T., KLEMAS, V. and WELLS, L., 1986, Application of Landsat data and synoptic remote sensing to predictive models for prehistoric archaeological sites: an example from the Delaware coastal plain, *American Antiquity* **51**(3):572–588.
- DENSEN, T. and PETERSON, J., 1995, Mapping the Australian duricrusts: can distribution be derived from terrain maps?, *Australian Geographic Studies* **30**:87–93.
- DINCAUZE, D.F., 1987, Strategies for paleoenvironmental reconstruction in archaeology, *Advances in Archaeological Method and Theory* **11**:255–336.
- DORAN, J., 1970, Systems theory, computer simulations, and archaeology, *World Archaeology* **1**:289–298.
- DUNN, C.P., SHARPE, D.M., GUNTENSPERGEN, G.R., STEARNS, F. and YANG, Z., 1991, Methods for analyzing temporal changes in landscape pattern. In

- M.G.Turner and R.H. Gardner (eds) *Quantitative Methods in Landscape Ecology*, New York: Springer-Verlag, pp. 173–198.
- EVANS, G.L., 1951, Prehistoric wells in eastern New Mexico, *American Antiquity* **17**(1): 1–9.
- FANALE, R., 1973, Jasper formation and the prediction of quarry site locations. In W.M. Gardner (ed.) *The Flint Run Paleo-Indian Complex: A Preliminary Report 1971–1973 Seasons*, Occasional Paper 1, Department of Anthropology, Catholic University of America, Washington, DC, pp. 138–146.
- FERNSTROM, K.W., 1984, The effect of ecological fluctuations on exchange networks, Black Mesa, Arizona. In S.Plog and S.Powell (eds) *Papers on the Archaeology of Black Mesa, Arizona* (Volume 2), Carbondale: Southern Illinois University Press, pp. 189–208.
- FOESE, L.J., PACKARD, J.M. and GRANT, W.E., 1989, AI modelling of animal movements in a heterogeneous habitat, *Ecological Modelling* **46**:57–72.
- GARSD, A., 1984, Spurious correlation in ecological modelling, *Ecological Modelling* **23**:191–201.
- GOUED, R.A. and SAGGER, S., 1985, Lithic procurement in central Australia: a closer look at Binford's idea of embeddedness in archaeology, *American Antiquity* **50**(1): 117–136.
- GREEN, F.E., 1962, Additional notes on prehistoric wells at the Clovis site, *American Antiquity* **28**(2): 230–234.
- GRIFFITHS, J.C., 1978, Mineral resource assessment using the unit regional value concept, *Mathematical Geology* **10**(5): 441–72.
- HALL, C.L. and KLIPPEL, W.E., 1988, A polythetic-satisficer approach to prehistoric natural shelter selection in middle Tennessee, *Midcontinental Journal of Archaeology* **13**(2): 159–186.
- HARPENDING, H. and DAVIS, H., 1977, Some implications for hunter-gatherer ecology derived from the spatial structure of resources, *World Archaeology* **8**(3):275–286.
- HAY, L.E., BATTAGLIN, W.A., PARKER, R.S. and LEAVESLEY, G.H., 1993, Modeling the effects of climate change on water resources in the Gunnison River Basin, Colorado. In M.F.Goodchild, B.O.Parks and L.T.Steyaert (eds.) *Environmental Modeling with GIS*, New York: Oxford University Press, pp. 173–181.
- HOUSTON, A.I., MCNAMARA J.M. and WEBB, J.N., 1995, The evolutionary stable exploitation of a renewing resource, *Journal of Theoretical Biology* **177**:151–158.
- HUNT, G.R. and SALISBURY, J.W., 1970, Visible and near-infrared spectra of minerals and rocks: I.Silicate minerals, *Modern Geology* **1**:283–300.
- HUNT, G.R., SALISBURY, J.W. and LENHOFF, C.J., 1973, Visible and near-infrared spectra of minerals and rocks: additional silicates, *Modern Geology* **4**:85–106.
- JACKSON, T.L., 1988, Geologic setting and prehistoric settlement patterning in the central Sierra Nevada, California, *Journal of California and Great Basin Anthropology* **10**(2): 227–236.
- JOHNSON, L.B., 1990, Analyzing spatial and temporal phenomena using geographical information systems: a review of ecological applications, *Landscape Ecology* **4**(1): 31–43.
- JOHNSTON, C., 1990, GIS: more than just a pretty face, *Landscape Ecology* **4**(1):3–4.

- JONES, G.T. and BECK, C., 1992, Chronological resolution in distributional archeology. In J. Rossignol and L. Wandsnider (eds) *Space, Time, and Archaeological Landscapes*, New York: Plenum Press, pp. 167–192.
- JONES, K.T. and MADSEN, D.B., 1989, Calculating the cost of resource transportation: a Great Basin example, *Current Anthropology* **30**(4):529–534.
- KEENE, A.S., 1983, Biology, behavior, and borrowing: a critical examination of optimal foraging theory in archaeology. In J.A. Moore and A.S. Keene (eds) *Archaeological Hammers and Theories*, New York: Academic Press, pp. 137–155.
- KHANBILVARDI, R.M. and ROGOWSKI, A.S., 1986, Modeling soil erosion, transport and deposition, *Ecological Modelling* **33**:255–268.
- KIRKBY, A. and KIRKBY, M.J., 1976, Geomorphic processes and the surface survey of archaeological sites in semi-arid areas. In D.A. Davidson and M.L. Shackley (eds), *Geoarchaeology: Earth Science and the Past*, Denver: Westview Press, pp. 229–253.
- KLIEM, C. and PETROPULOS, T., 1990, Comparison of subjective and objective methods in quantitative exploration: case studies. In G. Gaal and D.F. Merriman (eds) *Computer Applications in Resource Estimation*, New York: Pergamon Press, pp. 83–96.
- KOHLER, T.A., 1988, Predictive locational modeling: history and current practice. In W.J. Judge and L. Sebastian (eds) *Quantifying the Present and Predicting the Past*, Denver: Bureau of Land Management, pp. 19–96.
- KVAMME, K.L., 1985, Determining empirical relationships between the natural environment and prehistoric site locations: a hunter-gatherer example. In C. Carr (ed.) *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*, Kansas City: Westport Press, pp. 208–238.
- 1990, Geographic information systems in regional archaeological research and data management, *Archaeological Method and Theory* **1**:139–203.
- LAFERRIERE, J.E., 1995, A dynamic nonlinear optimization study of mountain pima subsistence technology, *Human Ecology* **23**(1):1–28.
- LANGRAN, G., 1992, *Time in Geographic Information Systems*, New York: Taylor & Francis.
- LAPEN, D.R. and MARTZ, L.W., 1993, The measurement of two simple topographic indices of wind sheltering-exposure from raster digital elevation models, *Computers and Geosciences* **19**(6):769–779.
- LA PORTA, P.C., 1995, The spatial distribution of prehistoric lithic resources: geological models and the elucidation of a prehistoric mining district, Paper presented at the 60th Annual Meeting of the Society for American Archaeology, Minneapolis.
- LEE, R.B., 1969, !Kung bushman subsistence: an input-output analysis. In A.P. Vayda (ed.) *Environment and Cultural Behavior*, New York: Natural History Press, pp. 47–79.
- LEVIN, S.A., 1995, The problem of pattern and scale in ecology. In T.M. Powell and J.H. Steele (eds) *Ecological Time Series*, New York: Chapman & Hall, pp. 277–326.
- McGLADE, J., 1995, Archaeology and the ecodynamics of human-modified landscapes, *Antiquity* **69**:113–132.
- MALANSON, G.P., 1993, Comment on modeling ecological response to climatic change, *Climatic Change* **23**:95–109.
- MATYASOVSKY, I., BOGARDI, I., BARDOSSY, A. and DUCKSTEIN, L., 1993, Space-time precipitation reflecting climate change, *Hydrological Sciences* **38**(6): 539–558.

- METCALFE, S.E., 1987, Historical data and climate change in Mexico—a review, *Geographical Journal* **153**(2):211–222.
- MOBLEY-TANAKA, J.L., 1995, Check dams at Chimney Rock mesa: agricultural strategies in a marginal environment, *Kiva* **60**(3):429–442.
- MURPHY, R.J., 1995, Mapping of jasperoid in the Cedar Mountains, Utah, U.S.A., using imaging spectrometer data, *International Journal of Remote Sensing* **16**(6): 1021–1042.
- O'NEILL, R.V., MILNE, B.T., TURNER, M.G. and GARDNER, R.H., 1988, Resource utilization scales and landscape pattern, *Landscape Ecology* **2**(1):63–69.
- OSBORNE, A.J., 1993, Snowblind in the desert Southwest: moisture islands, ungulate ecology, and alternative prehistoric overwintering strategies, *Journal of Anthropological Research* **49**:135–163.
- PAN, G. and DE HARRIS, V.P., 1992, Decomposed and weighted characteristic analysis for the quantitative estimation of mineral resources, *Mathematical Geology* **24**(7): 807–823.
- PATE, D., 1986, The effects of drought on Ngatatjara plant use: an evaluation of optimal foraging theory, *Human Ecology* **14**(1):95–115.
- PICKUP, G., 1985, The erosion cell—a geomorphic approach to landscape classification range assessment, *Australian Rangeland Journal* **7**(2):114–121.
- PICKUP, G. and CHEWINGS, V.H., 1986, Random field modelling of spatial variations in erosion and deposition in flat alluvial landscapes in arid central Australia, *Ecological Modelling* **33**:269–296.
- PICKUP, G. and NELSON, D.J., 1984, Use of Landsat radiance parameters to distinguish soil erosion, stability, and deposition in arid central Australia, *Remote Sensing of Environment* **16**:195–209.
- QUATTROCHI, D.A. and PELLETIER, R.E., 1991, Remote sensing for analysis of landscape: an introduction. In M.G.Turner and R.H.Gardner (eds) *Quantitative Methods in Landscape Ecology*, New York: Springer-Verlag, pp. 51–76.
- REDDY, T. and KOCH, G.S., Jr, 1988, A generalized model for evaluating area-potential in a mineral exploration program, *Mathematical Geology* **20**(3):227–241.
- REIDHEAD, V.A., 1979, Linear programming models in archaeology, *Annual Review of Anthropology* **8**:543–578.
- RHODE, D., 1990, On transportation costs of Great Basin resources: an assessment of the Jones-Madsen model, *Current Anthropology* **31**(4):413–419.
- RICE, D.G., 1993, Modeling constraints arising from established communications pathways: the influence of trails, Paper presented at the 46th Annual Northwest Anthropological Conference, Bellingham.
- ROUGHGARDEN, J., RUNNING, S.W. and MATSON, P.A., 1991, What does remote sensing do for ecology? *Ecology* **72**:1918–1922.
- SCARBOROUGH, V.L., 1988, A water storage adaptation in the American Southwest, *Journal of Anthropological Research* **44**(1):21–0.
- SEBASTIAN, L. and JUDGE, W.J., 1988, Predicting the past: correlation, explanation, and the use of archaeological models. In W.J.Judge and L.Sebastian (eds) *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, Denver, CO: Bureau of Land Management, pp. 1–18.

- SLOBODCHIKOFF, C.N., 1984, Resources and the evolution of social behavior. In P.W.Price, C.N.Slobodchikoff, and W.S.Gaud (eds) *A New Ecology: Novel Approaches to Interactive Systems*, New York: John Wiley, pp. 227–251.
- SMITH, E.A., 1979, Human adaptation and energetic efficiency, *Human Ecology* **7**(1): 53–75.
- SOULE, R.G. and GOLDMAN, R.F., 1972, Terrain coefficients for energy cost prediction, *Journal of Applied Physiology* **32**(5):706–708.
- STAFFORD, C.R., 1995, Geoarchaeological perspectives on paleolandscapes and regional subsurface archaeology, *Journal of Archaeological Method and Theory* **2**(1):69–104.
- STRANDMAN, H., VAISANEN, H. and KELLOMAKI, S., 1993, A procedure for generating synthetic weather records in conjunction of climatic scenario for modelling of ecological impacts of changing climate in boreal conditions, *Ecological Modelling* **70**:195–220.
- SWARTSMAN, G.L. and KALUZNY, S.P., 1987, *Ecological Simulation Primer*, New York: Macmillan.
- TAYLOR, R.B. and STEVEN, T.A., 1983, Definition of mineral resource potential, *Economic Geology* **78**:1268–1270.
- THRAILKILL, J., 1968, Chemical and hydrologic factors in the excavation of limestone caves, *Geological Society of America Bulletin* **79**:19–46.
- THOMAS, J.W., 1986, Wildlife-habitat modeling—cheers, fears, and introspection. In J.Verner, M.L.Morrison, and C.J.Ralph (eds) *Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates*, Madison: University of Wisconsin Press, pp. xix–xxv.
- TILMAN, D., 1980, Resources: a graphical-mechanistic approach to competition and predation, *American Naturalist* **116**(3):362–393.
- TOSTA, N., 1991, More than maps: asking the right questions, *Geo Info Systems*, February: 46–48.
- TURNER, M.G., Wu, Y., ROMME, W.H. and WALLACE, L.L., 1993, A landscape simulation model of winter foraging by large ungulates, *Ecological Modelling* **69**: 63–184.
- VINCENT, R.K., THOMSON, F. and WATSON, K., 1972, Recognition of exposed quartz sand and sandstone by two-channel infrared imagery, *Journal of Geophysical Research* **77**(14): 2473–2477.
- WALKER, P.A., 1990, Modelling wildlife distributions using a geographic information system: kangaroos in relation to climate, *Journal of Biogeography* **17**:279–289.
- WALSH, F.J., 1988, The Influence of the spatial and temporal distribution of plant food resources on traditional Martujarra subsistence strategies, *Australian Archaeology* **26**: 88–101.
- WINTERHALDER, B., 1980, Environmental analysis in human evolution and adaptation research, *Human Ecology* **8**(2):135–179.
- WOOD, J.J., 1978, Optimal location in settlement space: a model for describing locational strategies, *American Antiquity* **43**:303–309.
- WOODCOCK, C.E. and STRAHLER, A.H., 1987, The factor of scale in remote sensing, *Remote Sensing of Environment* **21**:311–332.
- YEAKLEY, J.A., MOEN, R.A., BRESHEARS, D.D. and NUNGESSER, M.K., 1994, Response of North American ecosystem models to multi-annual periodicities in temperature and precipitation, *Landscape Ecology* **9**(4):249–260.

Index

- activity location 75, 77, 131, 145
aerial photographs 11, 86–7
a priori assumption 82
a priori probability 11, 29
Allen, K.M.S. xi, xv, 3, 101
Archaic 80, 146
 Early 27
 Middle 27
 Late 28
archaeological method 1
 see also method
archaeological predictive model(ing), 3, 6,
 34–5, 55–6, 73–5, 77–8, 83–1, 94, 97,
 117
 see also predictive model
archaeological scale 105
 see also scale
archaeological survey 6, 10, 45, 60, 77–8,
 80, 86
 see also survey
archaeological theory 1
 see also theory
ARC/INFO 12, 34, 107, 114, 116–17, 120
Asch, D.L. xi, xv, 2, 5, 11, 27–9
axis of discrimination 9, 18, 20–1
- Beatles
 see ecology
Beckman, K.A. xi, xv, 2, 33
Belize 114, 117–19, 121–5
Bennett, D.A. xi, xv, 3, 113
bimodal settlement pattern 27–8
Bond, S.C. 107
boreal forest *see* forest
botanical resources 143–4
- Brandon, R.J. xi, xiii–xv, 3, 111, 135
Burgett, G.R. xii, xv, 3, 135
Butzer, K.W. 101, 137, 139, 146, 150
- catchment 14, 18, 107, 110
categorical variables 19, 66
 see also variables
Cayuga, Lake 3, 103, 106, 108–9, 111
Cayuga tribe 108
cell size 62, 103, 105–6, 120
chance 8, 20, 92
chance classification 2, 5, 20, 23, 27
Chesapeake Bay 59–63
chi-square test 14, 41
chi-square value 20
Church, T. xii, xv, 3, 135
Clarke, D.L. 101
climate 103, 139, 148
climatic change 62, 143, 148
climatic maps 107
composite settlement patterns 27–8
 see also settlement pattern
contingency table 43, 45
contour(s) 2, 14, 38, 63, 66, 113, 117–18,
 120–1, 123
contour maps 10, 114, 120
correlation 36, 41, 55–6, 129–37, 142–3,
 146, 148
correlative model 135–7, 142, 146, 150
 limitations of 137
cost 2, 71, 84, 95–6, 143–4, 149–50
cost-distance 40–1, 132
cost effectiveness 33, 59–60, 71, 80, 134,
 136
cost surface 40–1

- coverage, or data set 11, 29, 103, 114, 118,
120–2, 124
primary 14, 36
secondary 14–15, 36
- CRM *see* cultural resource management
- cross-validation testing 2, 5, 20, 27
- cultural factors 33, 111, 143
- cultural heritage values 74, 94, 97
- cultural landscapes 74, 111
- cultural resource management (CRM) xiii,
2, 33, 55, 59–60, 62, 71, 73–4
- cultural resource protection 3, 8, 33, 60, 73–
4, 94–7
- cumulative frequency 81–2, 91–3
- cut-point score 42–3, 45, 49
- Dalla Bona, L. xii, xv, 2, 73, 77–8, 84, 90
- data accuracy 55, 69, 103, 111, 11–15,
133, 136
- data resolution 101–3, 108, 136, 148
- data sets 1, 3, 17, 28–9, 34, 36, 38–41, 44,
50, 55, 59, 62, 66–7, 69, 114–18, 120–4
- data visualization xiii, 118, 135
- deciduous forest *see* forest
- deductive model 73, 75–6, 97, 130
- defensibility 104, 107, 110, 143
- Delaunay triangulation 117
- DEM *see* digital elevation model
- development planners 5, 8, 27
- developmental upbuilding 28
- differential distribution 144
- digital data 73, 80, 102, 114, 122,
- digital elevation model (DEM) 3, 14, 36,
39–41, 63, 66, 80, 84, 113–24
- digitize 12–14, 33, 36, 38, 84, 118, 120–3
- disturbance 2, 36, 55, 60, 71, 95, 143
- Duncan, R.B. xii, xv, 2, 33
- Early Archaic *see* Archaic
- Ebert, J.I. xii, xv, 3, 105, 129, 146
- ecology 4, 135, 143, 146–9
see also landscape ecology
- ecosystem 148
- ecotone 5, 14
- environmental changes 27
- environmental impact 2, 139, 148
see also impact assessment
- environmental variables 75–6, 105, 129,
131–2, 134–5, 143–4, 146, 148
see also variables
- Euro-American 5
- explanatory model 130, 133, 135, 137, 146–
7
see also deductive model
- external test 35, 43–5, 49, 54–5
- Fisher's exact test 14–15
- floodplains 26–9, 35, 48, 53–4, 63, 66,
118, 123
- food resources 75, 107–8, 137, 143–4, 146
- forest 3, 6, 11, 27, 74, 80, 90, 95, 97, 106,
108–10
boreal 73, 77–8, 83, 90, 95
deciduous 5, 10
diversity 106, 108
harvesting 78
prairie-forest ecotone 14
prairie 5, 14
river valleys 6
see also woodlands
- forest management plan 2, 73–4, 77, 83,
89–90, 94, 97
- frequencies of unique combinations 2, 66
- frequency distributions 17, 20
see also cumulative frequency
- frost *see* variables
creep 28
dates 105
free days 103, 107, 109–110
- F-to-enter scores 18
- Gaffney 111
- gain 5, 8, 20, 27, 41, 69, 132–3
- Gaussian model 120
- General Land Office (GLO) 10, 14
- geoarchaeology 28
- geographers 117, 123, 136
- geographic information system (GIS)
analytical tool, 130, 147
cost-effectiveness 60
database 4, 34, 59, 62–3, 80, 102
data creation 11, 28
data manipulation 14
data interpretation 26

- data testing 28
- for modeling landforms 41
- history 34–5, 135
- management tool, 1–2, 48, 55–6, 59, 135–6
- map algebra 44
- methodology 142
- models 35–6, 56, 73
- model testing 45, 48
- pitfalls 34, 130, 135
- predictive modeling 2, 4, 11, 20, 28, 32, 34, 36, 44–5, 48, 59, 73, 97, 113, 117, 129–30, 135, 148
- raster/cell data 14, 36, 38, 62, 66, 80, 103, 105, 106
- scale 100–2, 105–6, 108, 110–11
- settlement patterns 101, 103, 106
- spatial and temporal variability 101
- watershed model 71
- geology (geological) 1, 79–80, 84, 86–7, 89, 110, 143–5
- geomorphology 143, 146, 149
 - process 28, 148–9
- geosciences 114–116
- GIS *see* geographic information system
- GLO *see* General Land Office
- global positioning system (GPS) 114, 133
- global scale 3, 102–6, 108–10
- GPS *see* global positioning system
- grain 148
- Great Lakes 101, 103–5
- grid cell 14, 80
 - see also* GIS-raster/cell data

- Hageman, J.B. xii, xv, 3, 113
- Hasenstab, R.J. 106, 108–9
- Holocene 6, 10–11, 27–8
- horticulture 103, 107–9
- human behavioral patterns 33, 76, 131, 149
- Hungary 117
- Hunt E.D. 106–7
- hunter-gatherers 27–8
- hunting 103, 107, 109–10, 146
- hydrology 62–3

- IDW *see* inverse distance weighting
- Illinois 5–6, 10, 27–9, 116, 124
- Illinois Department of Energy and Natural Resources 10, 29
- impact 136, 143
- impact assessment 33, 59, 61, 69, 74, 95–6
- impact classification 95
- impact models 2, 71
- independent variables 5, 8, 14–15, 17–18, 20, 23, 113
 - see also* variables
- inductive model 8, 36, 55, 129–34
- Initial Woodland *see* Woodland
- internal test 41, 43, 45, 49, 54
- interpolation 3, 14, 113–24
 - definition 115
 - importance of 114–15
 - selecting an algorithm 117–18
 - types of 115–117
- inverse distance weighting (IDW) 116–17, 120–1, 123–4
- Iroquoian settlement 101, 103, 108
- Iroquois 3, 101, 103–8

- Kanawha River Valley 34–5, 53–4
- knolls 5, 14, 22, 26–9
- Knoerl, J.J. 105–6
- Kohler, T.A. 35, 75
- Kolmogorov-Smirnov
 - one-sample test 80, 90
 - two-sample test 41, 92
- kriging 115–17, 120–1, 123
 - ordinary 115–17, 120, 123
 - universal 116–17, 120–1
- Kuiper, J.A. xii, xv, 2, 59
- Kvamme, K.L. 29, 35, 69, 90, 113, 117, 124, 137, 142

- land managers 1, 4, 5, 8, 27, 136
- land parcels 8, 29, 36, 55–6, 84
- landscape xii–iv, 1, 3–4, 10, 27, 33, 35, 55–6, 60, 74, 76, 83, 86, 104, 111, 113, 115–18, 121, 123–4, 131, 134, 135, 137, 142–50
 - holistic view 4, 146
- landscape ecology 4, 135, 149–50
 - see also* theory
- landscape evolution 5, 28
- land use 77–8, 86, 89, 96, 135

- land use patterns 6, 135
land use planners 3, 73, 76, 94–7
land-use strategies 27
large(r)-scale 78, 103, 106–8, 110–11, 136, 148, 150
Late Archaic *see* Archaic
Late Woodland *see* Woodland
lattice 14, 113, 115–18, 120, 125
see also raster
Limp, F. xv
linear interpolator 116–17, 120–1
see also triangulated irregular network (TIN)
lithic procurement 142, 144–5
lithic scatter 60, 62, 87, 103
local scale 3, 101–11, 116, 147
logistic regression 2, 5–6, 8–9, 15–20, 23, 29, 36, 41, 66
- macro-scale 101–2
management plans 2, 60
see also forest management plan
Mann-Whitney rank-sum 14–15
map data 84, 132–3
map scale 102–3
marsh 5, 60
Maryland 59–61
Maryland Historical Trust 60, 62
method
analytical 3
cost-distance 41
dating 148
empirical 8, 129
interpolation 3, 114, 116–8, 121
modeling 10, 35–6, 73, 75, 78, 80, 83–4, 86, 89–90, 97
quantitative 34, 118, 124
scientific 3
statistical 8, 90
traditional 1, 8
see also Strahler method;
weighted value method
method and theory 1, 136, 148
Middle Archaic *see* Archaic
model accuracy 8–9, 16, 20, 45, 114, 117, 133
models
see correlative model;
deductive model;
digital elevation model;
explanatory model;
Gaussian model;
impact models;
inductive model;
logistic regression;
multivariate model;
simulation modeling;
null models;
predictive model;
probability model;
simulation models;
statistical classification model;
synoptic model
Monongahela River Valley 34–5, 37, 39, 47–8
Montana 117
Montgomery County, Illinois 7, 10, 27, 29
multi-component sites 131
multiple regression analysis 132
multivariate logistic regression 20
multivariate model 23, 76
- New York 3, 101, 103–6, 108
nonlinear simulation modeling 147
nonsites 6, 9, 49, 66
as background locations 20, 41–2
definition of 6, 14, 41, 134
importance of 8
versus sites 6, 8, 16–8, 20, 22, 131, 133–4
Northwestern University Archaeological Program 11, 29
null classification 20
null hypothesis 81–2, 90–2, 94
null models 8
- one-sample Kolmogorov test 80, 90
Onondaga tribe 104
Ontario Ministry of Natural Resources 73–4, 97
Ontario 2, 73–4, 77–8, 83, 86, 89–90, 94–5, 97, 103–5
optimization analysis 144, 147
ordinary kriging *see* kriging

- paleontological resources 10
paleoclimate 62, 139
paleoenvironmental 10
Paleo-Indian 60, 80, 139, 143
Parker, S.C. 35, 75
pattern recognition 8
pattern variation 144
Pennsylvania 33–4, 49, 56
Phase I archaeological survey 35, 44–5
 see also archaeological survey and survey
Pleistocene 10–11
policy 94
prairie 5–6, 10–11, 27
prairie-forest *see* forest
Prairie Peninsula 5–6, 10, 27
predictive model
 as cultural resource management tool, 73–4
 assumptions 6, 36
 background 6
 empirical 8
 formula 41
 inductive 8, 55
 management perspective 146
 multivariate 23
 problems/pitfalls 55, 144
 see also GIS Predictive Model
performance
 accuracy of 20, 22
 assessment of 20
 research perspective 146
prehistoric landscape xiii, 1
prehistoric land use 4, 146
prehistoric occupation 10, 28
prehistoric settlement 6, 8, 14, 27, 36, 75
prehistoric site location (distribution) 5–6, 26, 33, 41, 59, 62, 137, 142
probability component 19
probability cut-point 20
probability map 71
probability model 8, 26
processual, explanatory paradigm 142
protection (of cultural resources) 33, 74, 94–7
quadratic local interpolator 116, 120–1
quantitative analysis 121, 124
quantitative method 34, 118, 125
quintic interpolator 117, 120–1
 see also triangulated irregular network (TIN)
radiocarbon dating 1, 84, 129, 148
random classification 2, 5, 20, 23
random cluster sample 17
random sample 11, 15, 17, 92
random sampling theory 85
raster 62–3, 66–7, 90, 125
 DEMs 113–14
 module 36
 map algebra 41
raster-based DEM 114
raster map algebra function 41
regional scale 3, 62, 66, 69, 101–10, 147
regression coefficients 19
remote sensing 129, 135, 145, 147, 149
resolution 5, 84, 90, 102, 106, 108–9, 114, 148
 see also data resolution or spatial resolution
RMSE *see* root mean square error
root mean square error (RMSE) 118, 121–3, 125
sample 6, 8, 11, 14, 82, 90, 92–3, 116, 133
sample data 5, 16, 20, 115
sample size 66, 71
sampling methodology 90
Savage, S.H. 35
scale 3, 8, 101–11, 114, 118, 136, 142, 145–50
score, of site potential 36, 41–2, 45, 49, 54, 56
season 41, 96
 frost-free 103, 107, 109–10
seasonality 5, 27, 131–2
semivariance 115–16
semivariogram 116, 121
sensitivity map 34, 50
 as planning tool 50
settlement 5–6, 36, 103, 108, 131
settlement choices 6, 75, 102
settlement models 143

- settlement patterns 3, 8, 27–8, 62, 101–7, 146
- settlement systems 76, 101
- shelter 75, 137, 143
- simulation models 144, 147, 149
 - see also* nonlinear simulation modeling
- simulation studies 135, 147
- site-centered archaeology 130–1, 134
- site potential 33–4, 36, 41–2, 45, 49–50, 54, 66, 69, 74
- site preservation 2, 28, 36, 41,
- site probabilities 5, 8, 11, 19–26, 28–9, 33, 41, 55–6, 71, 133, 150
- site probability curves 23
- site versus nonsite 6, 8, 17, 20, 22, 131, 133–4
- sites as activity locations 75, 77, 131, 145
- small(er)-scale 78, 110–11, 145, 148, 150
- soil
 - as a variable *see* variables
 - agricultural potential 104, 107–10
 - associations 105–107
 - APG classification 69
 - depletion 104
 - developmental upbuilding 28
 - drainage 14, 17, 62, 66
 - erosion 71
 - flood frequency 14
 - flood plains 27–29
 - hilltop 27
 - landform type 14, 29
 - moisture availability 108
 - movement 71
 - permeability 14, 109
 - physical condition of 108
 - poor drained 10
 - productivity 106–7
 - as a resource 136
 - runoff 18–19, 25, 29
 - scale of 106–7, 110
 - series 14, 29
 - soil-survey maps 10, 14, 69
 - surface 28
 - surface runoff 18–19, 25, 29
 - texture 107
 - tillable 5
 - type 66, 90, 96, 102, 107, 136, 146
 - USDA 36
 - variability 66
 - well drained 11, 17, 54, 66
- solar insolation 2, 41
- spatial
 - analysis capabilities 36, 136
 - autocorrelation 115
 - correlated 115
 - coverage 114
 - data 1, 12
 - distribution 33, 121, 145
 - dynamic resource 146
 - patterning 101
 - resolution 3, 114
 - scale 101–6, 114
 - statistics 116, 125
 - and temporal 1, 104–5, 136, 142–3, 150
 - types of 105
 - variability 118
 - see also* temporal
- standard deviation, 121
- statistical analyses 3, 35–6, 62
- statistical classification model 8
- statistical methods 8
- statistical significance 17, 43, 136
- stepwise procedure/process 15, 18, 36
- Strahler method of stream rank ordering 14, 29
- stratified random sample 11
- subsistence 107, 146
- strategies 144
 - activities 103, 146
- subsistence models 103
- subsistence pattern 104
- suitability 5, 10, 106, 108–9, 113, 137
- suitability map 105
- survey 11, 28, 60, 69, 78, 84, 133
 - adequacy 6
 - data 11, 59
 - sample survey 26
 - systematic survey 6, 11
- Susquehanna River Valley 34–5, 49–52, 104
- swales 10, 14, 26, 28–9
- synoptic model 138
- system 131, 134, 139, 142, 146, 150
 - (definition)

- temporal 55, 105, 147–50
 - patterns 144
 - resources 146
 - structure 148
- Terminal Woodland *see* Woodland
- terrain 3, 40, 56, 90, 115, 118, 137, 143
- testing sample 9, 16 (definition), 17, 22–3
- theoretical framework 76
- theoretical implications 3
- theoretical perspective 142, 149
- theory 135, 150
 - anthropological 129
 - archaeological 1
 - cost-distance 132
 - ecological 149
 - explanatory 150
 - needs 3, 134–5, 148–50
 - patch 94
 - random sampling 85
 - temporal considerations 147–8
- Thiessen diagram 117
- time 147–8
- TIN *see* triangulated irregular network
- topographic maps 12, 14, 62, 80, 84, 114, 118, 132, 143
- topographic relief 10, 14, 18, 23, 66
- traditional use sites 74
- training sample 9, 16 (definition), 17, 20, 22–3, 29, 36, 41
- trend surface analysis 105
- triangulated irregular network (TIN) 14, 117–18, 120–1
 - accuracy 121
 - DEMs 117–18
- two-sample Kolmogorov-Smirnov test 41, 92
- Tygart Valley River 34–5, 48–9
- upland resources 6, 27–8
- Undifferentiated Woodland *see* Woodland
- unimodal settlement patterns 27–8
- univariate statistical test 14, 16, 41
- universal kriging *see* kriging
- USGS map data 10, 12, 14, 62, 114, 132
- validation 97
 - see also* cross-validation testing
- valley resources 27
- Vandrei, C. 107
- Van Nest, J. 28
- variables 2, 8, 14, 18–19, 23, 36, 41, 55, 62, 66, 69, 75–7, 83–4, 88
 - aspect 3, 41, 137
 - categorical 19, 66
 - canoe navigable waterways 106
 - design codes 19
 - dependent 15
 - derived 41
 - DEM 41
 - distance to water 60, 137
 - distributions 41
 - economic-related 76
 - elevation 3, 39–0
 - environmental 8–9, 20, 36, 55, 59, 62, 66, 69, 75–6, 102, 105, 106, 129, 131–3, 135–7, 143–6, 148
 - forest diversity 106
 - frost-free days 107
 - geographic 76
 - geologic 90
 - hydrology (streams) 14, 18, 38
 - infrastructure (roads) 38
 - independent 5, 8, 14–16, 18, 20, 23, 40, 113, 148
 - interval-scale 14–15
 - landscape 76
 - lime 109
 - nominal-scale 14–15, 25–6
 - ordinal-scale 14–15, 25–6
 - portage 84
 - potassium 109
 - precipitation 107
 - ratio-scale 14–15, 19 23, 25
 - relief 14, 137
 - secondary 36
 - selection of 41
 - significant 41
 - slope 3, 40–1, 137
 - soil 14, 18, 19, 25,
 - drainage 17
 - landform 29, 107
 - surface-runoff 29
 - Solar insolation 2
 - snowfall 107
 - static 150

- topographic 14, 18, 36, 84
- utility 14
- vegetation 14
- weighted 73, 76–77, 97
- wetlands 106
- vector-based DEM 114, 117
- vector data 38, 107

- Warren, R.E. xii, xv, 2, 5, 29, 36, 62, 116, 124
- water 5, 36, 41, 55, 59–60, 62–3, 66, 75–6, 79, 84, 90, 132, 137, 143–6
 - spatial availability of 143
- watershed 71, 86
 - models 71
 - Lake Cayuga 3, 108–9, 111
- weighted value method 75–7, 80–2, 90–2
- Wescott, K.L. xii–xiii, xv, 1–2, 59, 111
- West Virginia 33–4, 48, 53, 56
- Woodland
 - Initial 80
 - Late 60, 104
 - Terminal 80
 - Undifferentiated 80
- woodlands 5, 60, 95
 - see also* forest