

ENGINEERING SCIENCES

Electrical Engineering



IMAGING WITH SYNTHETIC APERTURE RADAR

Didier Massonnet & Jean-Claude Souyris

EPFL Press
Distributed by CRC Press



IMAGING WITH SYNTHETIC APERTURE RADAR

IMAGING WITH SYNTHETIC APERTURE RADAR

Didier Massonnet & Jean-Claude Souyris

EPFL Press

A Swiss academic publisher distributed by CRC Press



Taylor and Francis Group, LLC
6000 Broken Sound Parkway, NW, Suite 300,
Boca Raton, FL 33487

Distribution and Customer Service
orders@crcpress.com

www.crcpress.com

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress.

This book is published under the editorial direction
of Professor Olivier Martin (EPFL).

Previously published in this Series:
Power Semiconductors
Stefan Linder



is an imprint owned by Presses polytechniques et universitaires romandes, a Swiss academic publishing company whose main purpose is to publish the teaching and research works of the Ecole polytechnique fédérale de Lausanne.

Presses polytechniques et universitaires romandes
EPFL – Centre Midi
Post office box 119
CH-1015 Lausanne, Switzerland
E-Mail: ppur@epfl.ch
Phone: 021/693 21 30
Fax: 021/693 40 27

www.epflpress.org

© 2008, First edition, EPFL Press
ISBN 978-2-940222-15-5 (EPFL Press)
ISBN 978-0-8493-8239-0 (CRC Press)

Printed in the USA

All right reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprint, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publisher.

Acknowledgements

This book is built on radar imaging activities conducted at the Centre National d'Etudes Spatiales (CNES), Centre de Toulouse, between 1986 and 2007. We gratefully acknowledge for their support two former directors of CNES, Michel Courtois for having encouraged us to undertake the writing of a comprehensive book on SAR imaging, and Pierre Moskwa for his support during the last steps of this project.

SAR images, whether airborne or spaceborne, are the core of this book. We thank the Agenzia Spaziale Italiana (ASI), the Canadian Space Agency (CSA), the Deutsche Zentrum für Luft- und Raumfahrt (DLR), the European Space Agency (ESA), the Japanese Space Agency (JAXA), the Jet propulsion Laboratory (JPL), the National Aeronautics and Space Administration (NASA), the NPO Mashinostroyeniye company the Office National des Etudes Aérospatiales (ONERA), for allowing us to use some of their results. We also wish to acknowledge the efforts of several colleagues and friends who supported this work in a variety of ways. While this is undoubtedly not a complete list, this book would be unfinished and inaccurate without the help Thierry Amiot, Pierre Borderies, Jérôme Bruniquel, Professor Paul-François Combes, Bruno Cugny, Pascale Dubois-Fernandez, Pierre-Louis Frison, Thuy Le Toan, Jon Sen Lee, Henri-José Mametsa, Philippe Marchal, Anjali Mathur, Professor Eric Pottier, Nadine Pourthié and her trainees, as well as Céline Tison, Ridha Touzi and Hélène Vadon.

The writing of this book progressed in a irregular fashion. We warmly thank Daniel Rosenfeld for having expressed keen interest in our project at a key moment of its conception, and having given the suggestion to publish with the EPFL Press. We also wish to thank our publisher Frederick Fenter from the EPFL Press, for his patience and support.

We thank our translators from the Coup de Puce company in Toulouse, especially Ian Margo and Gayle Trager Berthias, who spent a lot of time and extra-time in transforming our 'flowery' French into straightforward English. Last but not least, we have to thank an anonymous reviewer, who has greatly contributed to improve the quality of this work.

Jean-Claude Souyris and Didier Massonnet
March 2008

Preface

Why would anyone write yet another book about radar imaging? Although the field is relatively specialised and little known to the general public, there is already a lot of high quality literature available. We in fact have excellent reasons to do so.

The first is that the situation has changed due to new techniques and recent projects: the availability of data from new and efficient space and airborne systems since the 1990s has led to rapid technical progress. These new techniques have also been boosted by incredible progress in computing, which over the same period has allowed for both easier and cheaper calculations for radar studies. The kind of processing that used to require days of computation and an entire computer centre can now be done by anybody on a desktop PC. This has radically changed both data processing possibilities and the choice of algorithms. Our second reason was that we wanted to describe radar-imaging techniques from a different viewpoint. Because there are so many different technical fields in which radar techniques are used, there is a correspondingly wide range of approaches. Specialists who have come to radar imaging with a background in electromagnetic signal processing techniques are likely to prefer a purely formal approach. Specialists in the use of radar images will see it as being derived from real aperture radar. Specialists in other kinds of image processing such as optical imaging or seismic tomography might prefer different approaches yet again. Our intention in this book is to maintain as geometrical an approach as possible to radar imaging. There are several different reasons for this. We believe that geometry enables us to use a more universal language than any of the specialised approaches that we have mentioned above. Although the geometrical approach may often seem naive, it nonetheless remains very precise. We thus see it as the simplest possible way of approaching the subject without sidestepping any of the difficulties and complexities inherent in radar imaging. Lastly, the geometrical results produced by radar make for the most spectacular applications. The most practical way of approaching this technique is therefore directly via those aspects which enable the widest range of radar imagery applications.

We have always noticed that our fellow radar specialists are particularly enthusiastic about their work. No one who has been involved with radar processing can ever quite leave it behind, despite any change in career. Of course, attempts to explain this must necessarily be partly arbitrary and personal. We believe however that this invol-

vement depends on two factors. First of all, there is almost a rite of passage involved when confronting radar imaging, as it is necessary to make a leap of faith in order to understand the complexity of images which are purely computer-generated. While almost anybody can understand optical imagery, radar is harder to grasp; indeed, there is a significant barrier posed both by the abstract nature of the signal phase, as well as the elaborate reconstruction techniques employed. Those who then cross the barrier find themselves part of an elite club. Naturally, we do not claim that radar imaging is more complicated than other fields of observation techniques, at a specialist level, but the fact there is this barrier makes it appear far more complex. Furthermore, radar imaging has very fundamental links with the phenomena studied and the investigation methods employed. There is a striking similarity between the ambiguity in position and speed in radar imaging and Heisenberg's principle of quantum mechanics, which also implies complex-based formalisms. There are also several analogies between radar polarimetry and quantum mechanics. It is precisely the fundamental concepts used in radar imaging that lead to the variety of approaches we have mentioned above. We have divided this book into five parts. The first chapter (A Theoretical Emergency Kit for SAR Imagery) covers a few theoretical principles which by their very diversity place radar imaging techniques at the cross-roads between electromagnetism, signal processing and image processing. The propagation and polarisation of electromagnetic waves, the radiation of microwave antennas, the physics of radar measurements and the characteristics of the Fourier Transform are each dealt with in turn. The second chapter (SAR Processing: At the Heart of the SAR Technique) has nonetheless been written in such a way as to make it accessible to readers who are not familiar with remote sensing and who do not have previous knowledge of signal processing or radar physics. It describes radar processing in terms of geometry. The attentive reader should need no more than a basic scientific background. The third chapter (From SAR Design to Image Quality) deals essentially with radiometry aspects. Referring to these and to determination of the radar/target link budget it gives a detailed description of the decisive trade-off between the geometrical (resolution) and the radiometry (amplitude of the radar echo) features of a radar image. The fourth chapter (SAR Interferometry: Towards the Ultimate Ranging Accuracy) explains the principles and the main applications of radar interferometry which generally produces two types of information, most frequently combined in a single image: topographical information and information about ground movements. Finally, the fifth and final chapter (SAR Polarimetry: Towards the Ultimate Characterization of Targets) aims to explain the basics of polarimetry, which extends the possibilities of radar measurements by varying the polarisation.

Toulouse, France, February 2008
Didier Massonnet and Jean-Claude Souyris

Table of Contents

	Acknowledgements	v
	Preface	vii
CHAPTER 1	A THEORETICAL EMERGENCY KIT FOR SAR IMAGERY	1
1.1	The propagation and polarization of electromagnetic waves	1
1.1.1	Maxwell's Equations	1
1.1.2	The polarization of electromagnetic waves	4
1.1.3	Partially polarized waves	11
1.1.4	In passing: the elegant algebra of the $SU(2)-O^+(3)$ homomorphism	17
1.2	The electromagnetic radiation of microwave antennas . . .	19
1.2.1	Introduction	19
1.2.2	The electromagnetic radiation equation	19
1.2.3	Resolving the electromagnetic radiation equation	20
1.2.4	Antenna pattern, directivity and gain	20
1.2.5	The radiation of planar antennas	21
1.2.6	Array antennas	24
1.2.7	SAR antenna technology	26
1.3	Interaction between waves and natural surfaces	26
1.3.1	Introduction	26
1.3.2	Surface scattering	27
1.3.3	Volume scattering	35
1.3.4	Penetration properties of electromagnetic waves . . .	36
1.3.5	The effects of slope on radiometry	39
1.4	Elements of signal processing	40
1.4.1	Introduction	40
1.4.2	Fourier series of real periodic functions	40
1.4.3	Fourier transform	40

	1.4.4	Properties of the Fourier transform (FT)	41
	1.4.5	Fourier transforms of standard functions	42
	1.4.6	Sampling real signals	45
	1.4.7	Sampling theorem (Shannon's theorem)	46
	1.4.8	The Fast Fourier transform (FFT) algorithm	48
	1.4.9	The two-dimensional Fourier transform	50
CHAPTER 2		SAR PROCESSING: AT THE HEART OF THE SAR TECHNIQUE	53
	2.1	Introduction	53
	2.2	General principles of Synthetic Aperture Radar	54
	2.2.1	A different way of observing the Earth	54
	2.2.2	Range vision	55
	2.2.3	The three fundamental radar frequencies	56
	2.2.4	An intuitive geometrical approach to synthetic aperture	59
	2.2.5	Synthetic aperture, an analytic geometry formulation	64
	2.3	Frequency representation	69
	2.3.1	Phase distribution expressed in the frequency domain	70
	2.3.2	Non-Zero Mean Doppler	71
	2.3.3	Doppler locking	73
	2.3.4	Mean Doppler (or Doppler centroid) estimation .	74
	2.3.5	Mean reduced Doppler estimation (integer part) .	76
	2.3.6	Range migration	78
	2.3.7	Range processing	80
	2.3.8	Saturation effects	81
	2.3.9	Interference effects	82
	2.3.10	Motionless radar approximation	83
	2.4	SAR Synthesis algorithms	84
	2.4.1	A common preliminary step, range compression .	84
	2.4.2	Time or 'naive' processing	85
	2.4.3	Range-azimuth or 'classic' processing	86
	2.4.4	An alternative technique, the Ω_k processor	87
	2.4.5	Subtle distortion, chirp scaling	87
	2.4.6	PRISME architecture, a multistage processing technique	88
	2.4.7	Unfocussed processing, a special case	90

	2.4.8	A practical example of the unfocussed processing technique	90
	2.4.9	Another special case, deramping processing	91
	2.4.10	A radar processing technique without approximations	93
2.5		System constraints	94
	2.5.1	Radar system design	94
	2.5.2	Timing constraints	95
	2.5.3	The different types of radar data	96
	2.5.4	Tricks for cheating nature	97
2.6		Geometry characteristics	101
	2.6.1	A practical example of the effect of range on images	101
	2.6.2	Equations for geometric positioning	104
	2.6.3	Perpendicular radargrammetry	107
	2.6.4	Radargrammetry and interferometry	108
	2.6.5	Oblique radargrammetry	109
	2.6.6	Radar clinometry	109
2.7		An introduction to super-resolution	110
2.8		Radar processing and geometric specificity of bistatic data	112
	2.8.1	Direct echo	115
	2.8.2	Triangular resolution	115
CHAPTER 3		FROM SAR DESIGN TO IMAGE QUALITY	117
	3.1	Introduction	117
	3.2	Radar equation, Radar Cross Section (RCS) of a point target	118
		3.2.1 Loss terms	120
	3.3	Radar signature for extended targets - the backscatter	122
	3.4	Signal to noise ratio (SNR) of the radar-target link before SAR synthesis	123
	3.5	Modifying the SNR during SAR synthesis	123
		3.5.1 Point targets	124
		3.5.2 Extended targets	130
	3.6	Instrument Noise Equivalent $\sigma^0 (NE\sigma^{0inst})$	133
		3.6.1 Energy cost for improving resolution	133
	3.7	Impact of image ambiguities on the $NE\sigma^{0inst}$ - total	133
		3.7.1 Range ambiguities	134

3.7.2	Azimuth ambiguities	136
3.7.3	Combined processing of range and azimuth ambiguities	137
3.7.4	Total $NE\sigma^0$ ($NE\sigma^{0tot}$)	137
3.8	Volume of data generated onboard	139
3.9	Telemetry data rate	140
3.9.1	Source coding	141
3.10	Calibration and corresponding image quality requirements	143
3.10.1	Internal calibration	144
3.10.2	External calibration	144
3.10.3	Calibration requirements and expected scientific results	145
3.11	Speckle noise and image statistics	146
3.11.1	Physical origin	146
3.11.2	Statistics of fully developed speckle	148
3.11.3	Speckle noise: multiplicative nature and modeling	149
3.11.4	Texture effect	150
3.11.5	Speckle noise in multi-look images	151
3.11.6	Speckle reduction filters	154
3.12	The impulse response (IR)	158
3.12.1	Range impulse response (RIR)	159
3.12.2	Azimuth impulse response (AIR)	160
3.12.3	Complex image spectrum, ISLR, PSLR, weighting effect	161
3.13	Radiometric elements of Image Quality	162
3.13.1	Estimating and analysing $NE\sigma^{0tot}$	162
3.13.2	Estimating the ambiguity level	162
3.13.3	Radiometric resolution:	164
3.14	Geometric elements of image quality	165
3.14.1	Spatial resolution and pixel size	165
3.14.2	Geometric distortion	166
3.14.3	The image location	167
3.15	Radar image interpretation	169
3.15.1	Description of data	169
3.15.2	Assessment of the Doppler spectrum	172
3.15.3	Platform position	173

	3.15.4 Saturation effects	173
	3.15.5 Directional effects	174
	3.15.6 Is it possible to fully characterize a radar instru- ment only from an image?	175
CHAPTER 4	SAR INTERFEROMETRY: TOWARDS THE ULTIMATE RANGING ACCURACY	177
4.1	Principles and limitations of radar interferometry	177
4.1.1	A specific use for radar	177
4.1.2	A brief history of interferometry	179
4.1.3	Interferometry and the physical properties of radar images	179
4.1.4	Phase difference between radar images	180
4.1.5	Robustness of the phase signal in interferometry	181
4.1.6	Limitations due to geometric and surface changes	181
4.1.7	Eliminating systematic contributions	184
4.2	Implementing interferometry processing	184
4.2.1	Radar image co-registration	186
4.2.2	Calculating phase differences between images	186
4.2.3	Finishing tasks	189
4.3	Application for topography	190
4.3.1	Set of equations and error budget	190
4.3.2	Eliminating measurement ambiguity	192
4.4	Application for displacement measurement	193
4.4.1	Set of equations and error budget	193
4.4.2	Examples of use	194
4.5	How slope effects limit interferometry	198
4.5.1	Frequency interpretation of the slope effect	201
4.6	Interpreting the results	202
4.6.1	Typical signatures of different contributions	202
4.6.2	Methods for improving interpretation	202
4.6.3	Interferometry interpretation in practice	205
4.7	Availability and mission perspectives	211
4.7.1	Availability of archived radar data	211
4.7.2	Availability of processing resources	212
4.7.3	Principles of data selection	213
4.7.4	Possibilities for future dedicated space missions	213

4.8	Comparison of interferometry with other methods	221
4.8.1	Comparison with optical stereoscopy for topographic measurement	221
4.8.2	Comparison with GPS for measuring displacement	222
4.9	Robustness of coherent processing when faced with ambiguities	224
4.10	Permanent reflectors	226
CHAPTER 5	SAR POLARIMETRY: TOWARDS THE ULTIMATE CHARACTERIZATION OF TARGETS	229
5.1	Introduction	229
5.2	Radar polarimetry: operating principle	230
5.2.1	Timing analysis - impact on system design	231
5.3	The scattering matrix	232
5.3.1	The monostatic singularity - the backscattering matrix	233
5.3.2	Target vector	233
5.4	Standard forms of backscatter	234
5.4.1	Odd-bounce (single, triple) scattering	234
5.4.2	Even-bounce (double) scattering	236
5.4.3	Diffraction or dipole mechanisms	237
5.5	Polarization synthesis	238
5.5.1	Polarimetric signatures	239
5.6	Characteristic polarization and Euler parameters	240
5.6.1	The Huynen fork	241
5.7	Coherent decomposition of the polarimetric measurement	241
5.7.1	Decomposition into standard mechanisms - the group of Pauli matrices	241
5.7.2	Algebraic decomposition: the Cameron approach	243
5.8	Taking depolarization into account	247
5.8.1	Stokes formalism and Mueller matrix	247
5.9	Covariance matrix - coherence matrix	250
5.9.1	Covariance matrix	251
5.9.2	Coherence matrix	253
5.10	Incoherent decomposition of polarimetric measurements .	254
5.10.1	Decomposition into polarization states: the Huynen approach	254

5.10.2	Decomposition into standard mechanisms - the Freeman approach	255
5.10.3	Algebraic decomposition - The (H, α) approach	257
5.11	Practical cases of polarimetric analysis	262
5.11.1	Radiometric analysis	262
5.11.2	Entropy analysis	264
5.11.3	Average backscattering mechanism	265
5.12	Synoptic representation of polarimetric information	265
5.13	Future compact polarimetric systems	266
5.13.1	Another idea: compact polarimetry and the $\pi/4$ mode	268
5.14	Merging polarimetry and interferometry: PolInSAR	269
5.14.1	Interferometric coherence optimization	270
5.14.2	Application to the inversion of vegetation height	271
5.14.3	PolInSAR extensions	272
5.15	Conclusion	273

CHAPTER 1

A THEORETICAL EMERGENCY KIT FOR SAR IMAGERY

1.1 The propagation and polarization of electromagnetic waves

A friendly warning to the reader: this discussion on the propagation and polarization of waves may bring back memories of school physics from another age as it does for the authors. The wave phenomena were profusely described at the time with images of ripples in water and skipping ropes. One day electromagnetic waves appeared on the black board. Newtonian physics suddenly faded into insignificance in the light of this new mysterious and infinite perspective. Indeed, for the first time, the teacher had referred to Maxwell's equations with the fascination of a geographer for uncharted territories. According to him, they governed the behavior of this abstract universe, bedecked in symmetry.

Now, many years later, they remain as attractive as ever. Associated with constitutive equations and boundary conditions, they are the basis for this introduction. Their resolution is described for a standard case of propagation in a vacuum with no boundary conditions. An entirely deterministic *plane wave* structure is obtained, described by its polarization state. However the beautiful flow of this perfect wave is disrupted by the time or spatial fluctuations of the media it passes through. These perturbing encounters cause a depolarization effect and a dedicated formalism is needed to characterize it. The ultimate aim of Sect. 1.1 is to prepare the reader for [Chapter 5](#) of this book, dedicated to SAR polarimetry.

1.1.1 Maxwell's Equations

Maxwell's equations govern the propagation of electromagnetic waves in any medium. Defined by James Clerk Maxwell (1831–1879) in 1873, they confirmed the intuitions of the self-taught genius Faraday (1791–1867), and were then experimentally validated by Heinrich Hertz in 1888. The Berlin physicist characterized the propagation effects of the induction force of an electric current, which partly cleared up the mystery of the strange 'effects of sparks occurring at a distance' which had been observed by several

investigators. Albert Einstein¹ later built his special theory of relativity on Maxwell's equations. We shall limit ourselves here to a brief description:

$$\nabla \times \vec{e}(\vec{r}, t) + \frac{\partial}{\partial t} \vec{b}(\vec{r}, t) = 0 \quad (1.1)$$

$$\nabla \times \vec{h}(\vec{r}, t) - \frac{\partial}{\partial t} \vec{d}(\vec{r}, t) = \vec{j}(\vec{r}, t) \quad (1.2)$$

$$\nabla \cdot \vec{b}(\vec{r}, t) = 0 \quad (1.3)$$

$$\nabla \cdot \vec{d}(\vec{r}, t) = \rho(\vec{r}, t) \quad (1.4)$$

This set of four equations brings into play the vector quantities $\vec{e}(\vec{r}, t)$, $\vec{h}(\vec{r}, t)$, $\vec{d}(\vec{r}, t)$, $\vec{b}(\vec{r}, t)$ which are, respectively, the electric and magnetic fields, the electric displacement and the magnetic induction. They are the unknowns of any propagation problem. The terms appearing on the right-hand side of the equations, $\rho(\vec{r}, t)$ and $\vec{j}(\vec{r}, t)$, known as source terms, represent the electric charge and current densities which have to be generated locally in order to produce radiation. For example, a microwave transmitter placed at the focus of a parabolic antenna induces currents $\vec{j}(\vec{r}, t)$ on its surface, causing radiation.

The electric and magnetic vectors depend on four parameters (three space parameters x , y and z constituting the space vector \vec{r} , and a time parameter t), and are subject to a set of mathematical operators: the curl operator ($\nabla \times$), the divergence operator ($\nabla \cdot$) and the partial time derivative ($\partial/\partial t$). The $\nabla \times$ and $\nabla \cdot$ operators act on the space coordinates and the $\partial/\partial t$ on that of time. Equations (1.1) and (1.2), called Faraday's law and Ampere's generalized law, reveal the way in which the space variations of an electric field (respectively a magnetic field) generate time variations of magnetic (respectively electric) induction. Equation (1.2) is the generalization of the electrostatic case of an infinite straight wire, carrying a direct current, which creates circular magnetic field lines centered on the wire. Equations (1.3) and (1.4) are Gauss's laws for electric displacement and magnetic induction. They are deduced from Eqs. (1.1) and (1.2), and from the *equation of energy conservation* during the spatio-temporal ballet of electric and magnetic fields (by $\nabla \cdot \vec{j}(\vec{r}, t) + j \cdot \omega \cdot \rho(\vec{r}, t) = 0$). For this reason, Eqs. (1.3) and (1.4) do not provide any extra information about the unknown vectors $\vec{e}(\vec{r}, t)$, $\vec{h}(\vec{r}, t)$, $\vec{d}(\vec{r}, t)$, and $\vec{b}(\vec{r}, t)$. To conclude, Maxwell's equations limit the number of independent scalar equations to six.

1.1.1.1 The constitutive equations

Since each of the unknown vectors is formed of three space and time-dependant components, the problem expressed by Eqs. (1.1)–(1.4) includes 12 unknowns. Now, Maxwell's equations only provide six independent scalar equations. It is therefore necessary to provide this system of equations with more information about the propagation medium. For this purpose, we introduce *constitutive equations*, which are

¹Biographers reported that Einstein had decorated his office with three portraits of illustrious predecessors: Newton, Faraday and Maxwell, which probably indicated that he felt he had a scientific debt to the last two who were pioneers of electromagnetism. The aim of his general theory of relativity was to reconcile Newton's theory of gravitation with Maxwell's of electromagnetism altogether but this is another story.

vector relations allowing us to express $\vec{d}(\vec{r}, t)$ and $\vec{b}(\vec{r}, t)$ as a function of $\vec{e}(\vec{r}, t)$, $\vec{h}(\vec{r}, t)$, and the permittivity and permeability characteristics of the medium studied:

$$\vec{d}(\vec{r}, t) = \vec{\varepsilon} \cdot \vec{e}(\vec{r}, t) + \vec{\xi} \cdot \vec{h}(\vec{r}, t) \quad (1.5)$$

$$\vec{b}(\vec{r}, t) = \vec{\zeta} \cdot \vec{e}(\vec{r}, t) + \vec{\mu} \cdot \vec{h}(\vec{r}, t) \quad (1.6)$$

Equations (1.5) and (1.6) refer to the ‘reaction’ of the medium to electromagnetic illumination: the fields $\vec{e}(\vec{r}, t)$ and $\vec{h}(\vec{r}, t)$ of the incident wave create the displacement $\vec{d}(\vec{r}, t)$ and induction $\vec{b}(\vec{r}, t)$, which become the relevant quantities for describing propagation in the medium in question.

Readers wishing to delve further into the meaning of the tensors $\vec{\varepsilon}$, $\vec{\xi}$, $\vec{\zeta}$, and $\vec{\mu}$ can refer to [Hallikainen, et al., 1985]. Let us just note that the very singular diffraction phenomena observed in anisotropic, bi-anisotropic (such as plasma) or chiral-magnetic media are due to their very particular specific constitutive equations.

Let us now assume that the wave emitted by the radar propagates in a vacuum. Under these conditions the constitutive equations (1.5) and (1.6) take the simplified form:

$$\vec{d}(\vec{r}, t) = \varepsilon_0 \cdot \vec{e}(\vec{r}, t) \quad (1.7)$$

$$\vec{b}(\vec{r}, t) = \mu_0 \cdot \vec{h}(\vec{r}, t) \quad (1.8)$$

The vacuum is an isotropic medium (the vectors $\vec{d}(\vec{r}, t)$ and $\vec{e}(\vec{r}, t)$ are collinear, as are $\vec{b}(\vec{r}, t)$ and $\vec{h}(\vec{r}, t)$), characterized by the universal constants $\varepsilon_0 \approx 8.85 \times 10^{-12}$ Farad·m⁻¹ (called the dielectric constant) and $\mu_0 = 4\pi \times 10^{-7}$ Henry·m⁻¹ (called the permeability constant). Moreover in the vacuum:

$$\vec{j}(\vec{r}, t) = \vec{0}, \quad \rho(\vec{r}, t) = 0 \quad (1.9)$$

The hypothesis of a vacuum-like propagation of the electromagnetic wave assumes that we neglect possible interactions of the incident wave with some of the atmosphere’s component layers (troposphere and ionosphere) which lie between a space-based SAR and its targets.

1.1.1.2 Boundary conditions

Finally, Maxwell’s equations are completed by boundary conditions characterizing specific propagation cases. They express the need to impose a certain number of initial conditions in order to ensure a unique solution for a differential equation of a given order. An example of ‘academic’ boundary conditions is that of a perfectly conductive surface subjected to electromagnetic radiation. In terms of Maxwell’s equations, the electric field is locally orthogonal (perpendicular) to this surface, which is called an ‘electrical wall’. In the case of an electromagnetic interaction with a natural medium (Sect. 1.3), these boundary conditions express the dielectric discontinuities imposed on the incident wave when it ‘touches’ the natural medium.

1.1.2 The polarization of electromagnetic waves

1.1.2.1 Structure of electromagnetic waves in the vacuum

The propagation equation derived from Maxwell's equations relates the space and time variations of the vectors $\vec{e}(\vec{r}, t)$, $\vec{h}(\vec{r}, t)$, $\vec{d}(\vec{r}, t)$, and $\vec{b}(\vec{r}, t)$ to the constitutive equations for the vacuum:

$$\nabla^2 \vec{e}(\vec{r}, t) - \varepsilon_0 \mu_0 \cdot \frac{\partial^2}{\partial t^2} \vec{e}(\vec{r}, t) = 0 \quad (1.10)$$

$\nabla^2 \cdot$ is the *Laplace* operator (also noted $\Delta \cdot$). Once expressed in a reference basis $(\hat{x}, \hat{y}, \hat{z})$, it is written as:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (1.11)$$

In the absence of boundary conditions (an infinite vacuum), a solution for this differential equation is given by Kong [1990]:

$$\begin{aligned} \vec{e}(\vec{r}, t) &= \Re \left\{ \vec{E}_0 \cdot \exp \left[j \cdot (\omega \cdot t - k_x \cdot x - k_y \cdot y - k_z \cdot z) \right] \right\} \\ &= \Re \left\{ \vec{E}_0 \cdot \exp \left[j \cdot (\omega \cdot t - \vec{k} \cdot \vec{r}) \right] \right\} \end{aligned} \quad (1.12)$$

where j is the classical imaginary unit equal to the square root of -1 : $j = \sqrt{-1}$. \Re represents the real part; x, y, z and k_x, k_y, k_z are the respective coordinates of the space vector \vec{r} and the wave (or propagation) vector \vec{k} , whose components satisfy:

$$k_x^2 + k_y^2 + k_z^2 = \omega^2 \varepsilon_0 \mu_0 \quad (1.13)$$

The relationship (1.13) is better known as the *equation of dispersion*, with ω related to the frequency f_c of the wave by the expression $\omega = 2\pi f_c$. Finally, \vec{E}_0 is a constant vector with complex components contained in a plane perpendicular to \vec{k} , called the Jones vector (Sect. 1.1.2.2). The form of the electric field in Eq. (1.12) characterizes a wave mechanism (both in space and time) that propagates along the direction $\hat{k} = \vec{k}/\|\vec{k}\|$.

Equi-phase and equi-amplitude surfaces are planes that are orthogonal (perpendicular) to the propagation direction of the energy represented by the vector \hat{k} . For this reason, the wave is called a plane wave. Finally turning again to Maxwell's equations, it is easy to calculate the magnetic field $\vec{h}(\vec{r}, t)$. This gives:

$$\vec{e}(\vec{r}, t) = \xi \cdot \vec{h}(\vec{r}, t) \times \hat{k} \quad (1.14)$$

where $\xi = \sqrt{\mu_0/\varepsilon_0}$ is the impedance of the wave in the vacuum ($\xi \approx 120\pi$).

Equation (1.14) characterizes the structure of the plane wave propagating in the vacuum: the vector triplet $[\vec{e}(\vec{r}, t), \vec{h}(\vec{r}, t), \hat{k}]$ is right-handed. The wave has a transverse electromagnetic structure in which the fields $\vec{e}(\vec{r}, t)$ and $\vec{h}(\vec{r}, t)$ are in the plane

perpendicular to the propagation direction (Fig. 1.1). The transfer of electromagnetic energy thus occurs in a direction perpendicular to the electric or magnetic vibrations, similarly to waves rippling on a pond with local vertical displacements.²

Let us consider for example a plane wave propagating in the direction \hat{z} . Under these conditions, the transversal (or orthogonal) plane containing the fields $\vec{e}(\vec{r}, t)$ and $\vec{h}(\vec{r}, t)$ is the plane (\hat{x}, \hat{y}) , and the vector \hat{k} is parallel to \hat{z} (Fig. 1.1). During the propagation, the electromagnetic fields undergo a helical rotation centered around \hat{z} . Projected onto the transversal plane, this helix is reduced to an ellipse called the polarization ellipse.

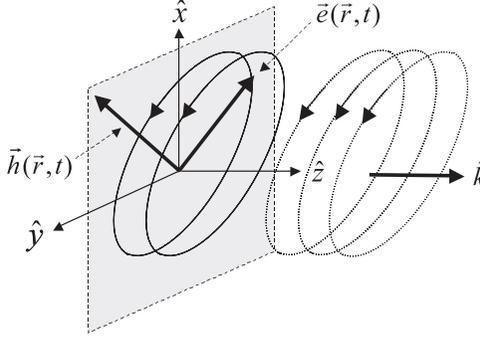


Fig. 1.1 Structure of plane wave in free space.

1.1.2.2 The polarization ellipse

The wave's polarization is defined by the projection of the curve traced by the leading edge of the electric field $\vec{e}(\vec{r}, t)$ onto the plane orthogonal to the propagation direction. To define this concept more clearly, let us look again at Eq. (1.12), expressed according to our simplifying hypotheses:

$$\vec{e}(x, y, z, t) = e_x \cdot \hat{x} + e_y \cdot \hat{y} = \Re \{ \vec{E}_0 \cdot \exp [j \cdot (\omega \cdot t - k \cdot z)] \} \quad (1.15)$$

in which the Jones vector \vec{E}_0 is:

$$\vec{E}_0 = E_{0x} \cdot \hat{x} + E_{0y} \cdot \hat{y} = a_x \cdot e^{j\delta_x} \cdot \hat{x} + a_y \cdot e^{j\delta_y} \cdot \hat{y} \quad (1.16)$$

where a_x (resp. a_y), δ_x (resp. δ_y) are the amplitude and the phase of the Jones vector coordinates along \hat{x} (resp. \hat{y}). The entity $k = 2\pi / \lambda$ is called the wave number.

This representation of the electric field taken independently of its spatio-temporal variations is also (incorrectly) called 'wave polarization'. From Eq. (1.15) and Eq. (1.16), this gives:

$$\begin{aligned} e_x &= a_x \cdot \cos(\omega \cdot t - k \cdot z + \delta_x) \\ e_y &= a_y \cdot \cos(\omega \cdot t - k \cdot z + \delta_y) \end{aligned} \quad (1.17)$$

²The propagation of acoustic waves is different with respect to the fact that the transfer of energy occurs in the same direction as that of the local variations of air pressure which cause the propagation.

The variations of the components e_x and e_y only depend on the space-time term $(\omega \cdot t - k \cdot z)$. From Eq. (1.17), we obtain:

$$\left(\frac{e_x}{a_x}\right)^2 + \left(\frac{e_y}{a_y}\right)^2 - 2 \cdot \frac{e_x}{a_x} \cdot \frac{e_y}{a_y} \cdot \cos(\delta_y - \delta_x) = \sin^2(\delta_y - \delta_x) \quad (1.18)$$

This results in an equation for a conic section whose determinant is ≥ 0 : this is an ellipse, called the ‘polarization ellipse’. Going back to the three-dimensional problem, the tip of the electric-field vector describes a helix inscribed in a cylinder with an elliptical cross-section (the cylindrical nature illustrates the principle of conservation of energy during propagation).

To avoid possible confusion between variations in space and in time, we arbitrarily choose a transversal reference plane from which we observe the movement of the electric field, for example, the plane $z = 0$. In this plane (as in any other plane), the polarization ellipse is characterized by two angular parameters: ψ and τ (Fig. 1.2). ψ characterizes the orientation of the major axis of the ellipse with respect to the \hat{x} axis, and τ is its ellipticity. The sign of τ (Sect. 1.1.2.3) indicates the direction in which $\vec{e}(x, y, 0, t)$ rotates around the polarization ellipse. Finally the value of the semi-major axis of the ellipse is linked to the amount of energy carried by the wave.

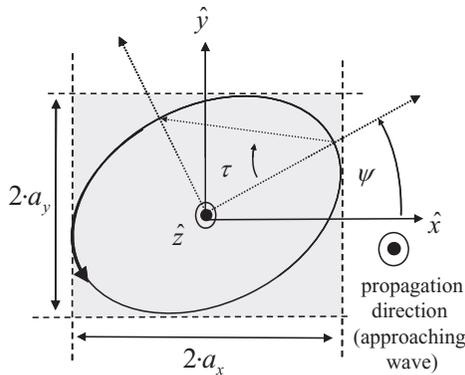


Fig. 1.2 The polarization ellipse in the polarization plane.

Using trigonometry, it is possible to derive the values (ψ, τ) from the values $(a_y/a_x, \delta_y - \delta_x)$ [Born, 1970]: the orientation of the major axis of the polarization ellipse, as well as its ellipticity, can be deliberately set given two orthogonal elementary excitations for which we are able to control the ratio of amplitudes and the relative phase offset. Given this, it follows that any wave with elliptical polarization may be considered as a linear combination of two elementary waves with perpendicular linear polarizations (along the directions \hat{x} and \hat{y}), to which are assigned respectively the complex coefficients $a_x \cdot e^{j\delta_x}$ and $a_y \cdot e^{j\delta_y}$. The complex nature of the weighting coefficients expresses the concept of coherent summing defined below.

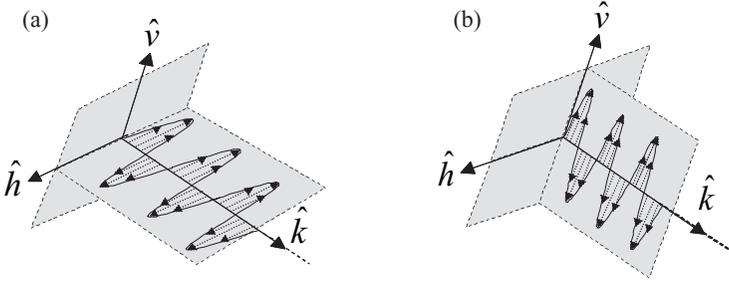


Fig. 1.3 (a) Spatial variations of the electric field in the case of a horizontal linear polarization (RADARSAT type); (b) Spatial variations of the electric field in the case of a vertical linear polarization (ERS type).

The Jones vector described in Eq. (1.16) by means of a_x , a_y , δ_x , and δ_y , can be expressed in a more geometrical way as a function of ψ and τ :

$$\begin{aligned} \vec{E}_0 &= A \cdot \begin{pmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{pmatrix} \cdot \begin{pmatrix} \cos \tau & j \cdot \sin \tau \\ j \cdot \sin \tau & \cos \tau \end{pmatrix} \cdot \begin{pmatrix} e^{-j\alpha} & 0 \\ 0 & e^{j\alpha} \end{pmatrix} \cdot \hat{x} \\ &= A \cdot [\bar{\bar{U}}(\psi)] \cdot [\bar{\bar{U}}(\tau)] \cdot [\bar{\bar{U}}(\alpha)] \cdot \hat{x} \end{aligned} \quad (1.19)$$

where A is the amplitude of \vec{E}_0 and α is an absolute phase term. To within the complex amplitude term $A \cdot e^{j\alpha}$, Eq. (1.19) reveals the coexistence in the wave structure of two vibrations in phase quadrature (term in j), respectively with amplitudes proportional to $\cos \tau$ and $\sin \tau$, in a basis which is rotated by an angle ψ with respect to the reference basis (\hat{x} , \hat{y}). Moreover, the matrices $[\bar{\bar{U}}(\psi)]$, $[\bar{\bar{U}}(\tau)]$ and $[\bar{\bar{U}}(\alpha)]$ have specific algebraic properties: all three belong to the group of special unitary 2×2 complex matrices (SU2). They satisfy the following expression:

$$\begin{aligned} \det [\bar{\bar{U}}] &= 1 \\ [\bar{\bar{U}}]^{-1} &= [\bar{\bar{U}}]^{T*} = [\bar{\bar{U}}]^+ \\ [\bar{\bar{U}}(x)]^{-1} &= [\bar{\bar{U}}(-x)] \end{aligned} \quad (1.20)$$

where the superscripts ‘ -1 ’, ‘ T ’ and ‘ $*$ ’ and ‘ $+$ ’ stand for ‘inverse’, ‘transpose’, ‘conjugate’ and ‘conjugate transpose’, respectively.

The linear polarizations

These are characterized by a constant orientation of the electric field during propagation. Following Eq. (1.17), this assumes a phase difference ($\delta_y - \delta_x$) equal to an integer multiple of π , or to the value $\tau = 0$ according to Eq. (1.19). In ordinary language, a linear polarization for which the electric field remains parallel to the illuminated surface (hence ‘horizontal’: Fig. 3(a)) is called **horizontal** and a linear polarization for

which the electric field is contained in the incident plane is called *vertical*³ (careful, this does not mean that the electric field is vertical with respect to the illuminated surface: cf. Fig. 3(b)).

The circular polarizations

Following Eq. (1.17), these require that two conditions be satisfied, namely that $a_y = a_x$ and that the phase difference $(\delta_y - \delta_x)$ be equal to an odd multiple of $\pi/2$ (the two elementary orthogonal excitations have the same amplitude and the same phase quadrature, hence $\tau = \pm\pi/4$ in Eq. (1.19)). The corresponding Jones vectors are written as $(1, j)^T$ and $(1, -j)^T$, respectively. We shall see (Sect. 1.1.2.3) that $(1, -j)^T$ is actually associated to with right-hand circular polarization.

The orthogonal polarizations

By definition, two polarizations are said to be orthogonal if their respective Jones vectors \vec{E}_0 and \vec{E}_0^\perp are orthogonal in terms of the Hermitian scalar product ($\vec{E}_0^T \cdot \vec{E}_0^{\perp*} = 0$). Returning to Eq. (1.19), the vector \vec{E}_0^\perp may be expressed in two ways:

$$\begin{aligned}\vec{E}_0^\perp &= A \cdot \left[\vec{U}(\psi) \right] \cdot \left[\vec{U}(\tau) \right] \cdot \left[\vec{U}(\alpha) \right] \cdot \hat{y} \\ &= A \cdot \left[\vec{U}(\psi + \pi/2) \right] \cdot \left[\vec{U}(-\tau) \right] \cdot \left[\vec{U}(\alpha) \right] \cdot \hat{x}\end{aligned}\quad (1.21)$$

Two orthogonal polarizations always form a basis in polarization space.

The pairs of orthogonal polarizations consist of linear polarizations oriented in orthogonal directions ($\psi = \psi_0$ in one case, $\psi = \psi_0 + \pi/2$ for the other), of circular polarizations with opposite rotation directions ($\tau = \pi/4$ in one case, $\tau = -\pi/4$ for the other), and, more generally, elliptical polarizations with the same ellipticity and with major axes orthogonal, but with opposite rotation directions.

Waves that propagate in opposite directions

What is the relation between the Jones vectors $\vec{E}_0(\hat{k})$ and $\vec{E}_0(-\hat{k})$ of two waves with the same polarization structure propagating in opposite directions? Propagation in an opposite direction means (by definition) a change in the propagation term $\exp[j \cdot (\omega \cdot t - k \cdot z)]$ of Eq. (1.15) to $\exp[j \cdot (\omega \cdot t + k \cdot z)]$. Consequently it is necessary to also conjugate the Jones vector $\vec{E}_0(\hat{k})$ to maintain the same spatio-temporal structure $\vec{e}(x, y, z, t)$. Thus, we can write:

$$\vec{E}_0(-\hat{k}) = \vec{E}_0^*(\hat{k}) \quad (1.22)$$

³The Canadian RADARSAT satellite operates with horizontal polarisation while the European ERS-1/2 satellites operate with vertical polarisation. For this kind of mono-polarized instrument, the choice of polarisation depends on the mission objectives. The RADARSAT mission which was designed to plot open sea routes in the Arctic area (by discriminating between water and ice) and to detect ships used horizontal polarisation (which minimises the sea clutter signal).

1.1.2.3 Direction of polarization

As has already been mentioned, the sign of τ determines the direction in which the edge of the electric field rotates on the polarization ellipse. Depending on the direction of this motion, it is referred to as right polarization or left polarization.

The terms ‘right’ and ‘left’ should be used with caution. Authors use different conventions to define polarization directions. Thus in [Born, 1970] (p.28), an elliptical polarization is said to be ‘right’ when the electric field $\vec{e}(x, y, 0, t)$ that describes the polarization ellipse turns in a clockwise direction for an observer looking towards the source of the wave. On the contrary, the standard IEEE definition refers to right polarization when this same field $\vec{e}(x, y, 0, t)$ turns in a clockwise direction for an observer looking in the direction of propagation. This is the case in Fig. 1.2, for a wave which propagates along \hat{z} . These two definitions are contradictory; we shall adopt the second which is in accordance with international standards.

This type of contradiction reveals a classic difficulty encountered when studying radars with polarization diversity; it is easy to get lost in a labyrinth of signs and conventions. There is an amusing anecdote, described in ref. [Pierce, 1974] referring to the mishaps of the TELSTAR-1 mission, which was used for the first transatlantic broadcast of television images (10/07/1962). During this historic rendezvous, the English station of Goonhilly and the French station at Pleumeur-Bodou, both watching out for signals emitted with circular polarization by the American satellite, did not have the same success. The French managed to receive the signals as their polarization direction was defined in the same way as that of the Americans, while the English, who had adopted an opposite convention to that of their partners, sadly regretted that their receivers remained mute.⁴

The diagram in Fig. 1.4 represents the polarization ellipse (as well as its track direction over time) in the transversal plane $z = 0$, for given values of a_x and a_y and variable phase difference $\delta = \delta_y - \delta_x$.

1.1.2.4 Polarization basis

We shall see Section 5.5 that an advantage of radar polarimetry is the possibility of synthesizing the response of a target to any given transmitted elliptical polarization. Successful use of such a principle requires correctly using the algebra for changes in polarization bases.

⁴The history of astrophysics tells us that two of the programme engineers from the Bell laboratories at Holmdel, in New Jersey, were none other than Arno Penzias and Robert Wilson. They became famous after having pointed a TELSTAR antenna towards the halo of our galaxy, which enabled them to detect radiation which was not known to exist at the time, with properties which strangely enough were identical in all directions. Was it an interference signal? No, without realizing it they had achieved the first radio observation of radiation from the cosmological background, which led to their Nobel Prize in Physics in 1978. This famous background radiation, with a temperature of 2.7 K, is the foundation stone of the Big-Bang theory and continues to excite researchers and engineers. Space agencies are currently considering developing probes which will be able to measure its polarization properties, which brings us back to the current subject.

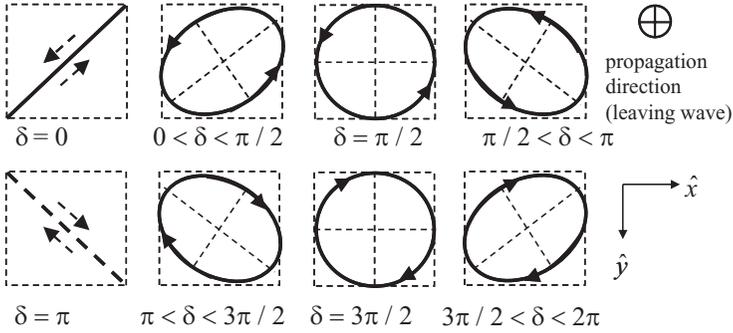


Fig. 1.4 The polarization ellipse in the plane which is transversal to the propagation plane, for given values of a_x and a_y (with the exception of the specific cases $\delta = \pi/2$, and $\delta = 3\pi/2$ where they are chosen equal), and a phase offset $\delta = \delta_x - \delta_y$ [Born and Wolf, 1970].

Let us consider any basis of orthogonal polarizations (\hat{a}, \hat{b}) . We want to determine E_{0a} and E_{0b} which satisfy:

$$\vec{E}_0(x, y) = E_{0x} \cdot \hat{x} + E_{0y} \cdot \hat{y} = \vec{E}_0(a, b) = E_{0a} \cdot \hat{a} + E_{0b} \cdot \hat{b} \quad (1.23)$$

as a function of E_{0x} and E_{0y} . \hat{a} and \hat{b} are defined via their characteristic parameters (ψ, τ, α) :

$$\begin{aligned} \hat{a} &= [\bar{U}(\psi)] \cdot [\bar{U}(\tau)] \cdot [\bar{U}(\alpha)] \cdot \hat{x} \quad \text{and} \\ \hat{b} &= [\bar{U}(\psi)] \cdot [\bar{U}(\tau)] \cdot [\bar{U}(\alpha)] \cdot \hat{y} \end{aligned} \quad (1.24)$$

which can be expressed via the transformation (transition) matrix:

$$[\bar{P}] = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = [\bar{U}(\psi)] \cdot [\bar{U}(\tau)] \cdot [\bar{U}(\alpha)] \quad (1.25)$$

The matrix $[\bar{P}]$ expresses in a concise way the transition from the vector pair (\hat{x}, \hat{y}) to the vector pair (\hat{a}, \hat{b}) :

$$\begin{bmatrix} \hat{a} & \hat{b} \\ p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{matrix} \hat{x} \\ \hat{y} \end{matrix} \quad (1.26)$$

Finally:

$$\begin{pmatrix} E_{0a} \\ E_{0b} \end{pmatrix} = [\bar{P}(\psi, \tau, \alpha)]^{-1} \cdot \begin{pmatrix} E_{0x} \\ E_{0y} \end{pmatrix} \quad (1.27)$$

It may be noted that

$$\left[\bar{\bar{P}}(\psi, \tau, \alpha) \right]^{-1} = \left[\bar{\bar{U}}(-\alpha) \right] \cdot \left[\bar{\bar{U}}(-\tau) \right] \cdot \left[\bar{\bar{U}}(-\psi) \right] \quad (1.28)$$

based on the properties stated in Eq. (1.20). $\left[\bar{\bar{P}}(\psi, \tau, \alpha) \right]$ and $\left[\bar{\bar{P}}(\psi, \tau, \alpha) \right]^{-1}$ also belong to the SU2 group.

1.1.3 Partially polarized waves

Up to now we have implicitly assumed that the parameters that define the polarization of the wave (ψ and τ , or in a dual way $a_x \cdot e^{j\delta_x}$ and $a_y \cdot e^{j\delta_y}$) are independent of time and space. We shall now consider the general case of partially polarized waves which will cause these parameters to fluctuate.

The structure of Eq. (1.15) reveals only sinusoidal temporal variations of the electric field by means of the $\exp [j \cdot \omega \cdot t]$ term. The wave is monochromatic. The terms $a_x \cdot e^{j\delta_x}$ and $a_y \cdot e^{j\delta_y}$ are intrinsic complex amplitudes characteristic of the wave's polarization state (likewise for ψ and τ). We may nevertheless imagine a more general case for which these terms are time dependent. Equation (1.15) may then be rewritten as:

$$\begin{aligned} \vec{e}(\vec{r}, t) &= \vec{e}(x, y, z, t) \\ &= \Re \left\{ \left(a_x(t) \cdot e^{j\delta_x(t)} \cdot \hat{x} + a_y(t) \cdot e^{j\delta_y(t)} \cdot \hat{y} \right) \cdot \exp [j \cdot (\omega \cdot t - k \cdot z)] \right\} \\ &= \Re \left\{ \vec{E}_0(t) \cdot \exp [j \cdot (\omega \cdot t - k \cdot z)] \right\} \end{aligned} \quad (1.29)$$

The term $\vec{e}(x, y, z, t)$ may be due to the scattering from a time-varying point target or, in a dual way, of an element of an extended static target which exhibits spatial fluctuations. In the second case, the time t in the vector $\vec{E}_0(t)$ will be replaced by an index identifying the element (i.e., the image pixel in radar imagery).

The temporal variations of $\vec{E}_0(t)$ are significant if the coherence time of the target (duration below which it may be considered to be static) is less than the duration of the measurement. Letting $\delta(t) = \delta_y(t) - \delta_x(t)$, the time variations of the quantities $a_x(t)$, $a_y(t)$, and $\delta(t)$ lead to a spectrum spread Δf of the initially monochromatic wave. Assuming $\Delta f \ll f_c$, and writing the measurement duration as Δt , the wave will be considered to be polarized as long as $\Delta t \ll 1/\Delta f$. If this condition is satisfied, the time dependency of the polarization parameters may be neglected. Otherwise, the fluctuation of the parameters $a_x(t) \cdot e^{j\delta_x(t)}$ and $a_y(t) \cdot e^{j\delta_y(t)}$ does not allow us to describe the polarization state in a relevant way, and the wave is said to be partially polarized. As the Jones vector is no longer useful, new parameters are needed to describe the polarization.

1.1.3.1 The coherence matrix

An electric field propagating according to Eq. (1.29) is detected by an antenna pointing in the direction of propagation (Fig. 1.5). The antenna is characterized by its receiving polarization \hat{E}_0^{rec} , which is the polarization of the field $\vec{e}^{rec}(\vec{r}, t)$ it would transmit if it were used as a transmitter (reciprocity theorem).

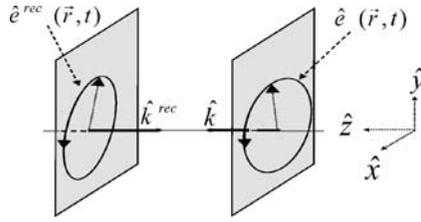


Fig. 1.5 Configuration of the wave incident to the receiving antenna and of its polarization.

The coupling of the receiving antenna and the incident field produces, at the antenna output, a detected voltage $V(t)$, with an average power proportional to:

$$P_{av} = \langle V(t) \cdot V^*(t) \rangle = \left\langle \left| \hat{E}_0^{recT} \cdot \vec{E}_0(t) \right|^2 \right\rangle \quad (1.30)$$

There are two points to be made about Eq.(1.30): (1) \hat{E}_0^{rec} is a dimensionless unit vector, independent of time. The projection of $\vec{E}_0(t)$ onto \hat{E}_0^{rec} expresses the antenna sensitivity to the polarization of the received field. (2) $\hat{E}_0^{recT} \cdot \vec{E}_0(t)$ is a Hermitian scalar product for which the second vector is conjugated twice. The first conjugation is inherent in the mathematical formulation of the Hermitian product, while the second is of physical origin: it expresses the coupling of two waves propagating in opposite directions (Sect. 1.1.2.2).

The average detected power may be written as:

$$P_{av} = \hat{E}_0^{recT} \cdot [\bar{J}] \cdot \hat{E}_0^{rec*} \quad (1.31)$$

where $[\bar{J}]$ is known as the wave coherence matrix. It is a 2×2 Hermitian matrix (since it is equal to its conjugated transposed matrix):

$$[\bar{J}] = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{bmatrix} = \begin{bmatrix} \langle E_{0x} \cdot E_{0x}^* \rangle & \langle E_{0x} \cdot E_{0y}^* \rangle \\ \langle E_{0y} \cdot E_{0x}^* \rangle & \langle E_{0y} \cdot E_{0y}^* \rangle \end{bmatrix} \quad (1.32)$$

The degree of coherence

The degree of coherence characterizes the correlation between the electrical excitations along directions \hat{x} and \hat{y} :

$$\mu_{xy} = \frac{J_{xy}}{\sqrt{J_{xx}} \cdot \sqrt{J_{yy}}} = \frac{\langle E_{0x} \cdot E_{0y}^* \rangle}{\sqrt{\langle E_{0x} \cdot E_{0x}^* \rangle} \cdot \sqrt{\langle E_{0y} \cdot E_{0y}^* \rangle}} \quad (1.33)$$

Due to the Schwarz inequality, the magnitude of μ_{xy} is between 0 and 1. There are two extreme cases:

- (1) *Polarized waves*: This gives the ‘conventional’ formulation for which the polarization parameters are independent of time. One then obtains $|\mu_{xy}| = 1$ and $\det([\bar{J}]) = 0$.

(2) *Unpolarized waves*: This is the case for natural light, for which there is no privileged polarization. The random relative phases $\delta(t)$ are uniformly distributed between 0 and 2π . This means that the off-diagonal terms J_{xy} and J_{yx} of the coherence matrix must be zero and that the diagonal terms J_{xx} and J_{yy} must be equal. One then obtains $|\mu_{xy}| = 0$.

In other words, the coherence matrix of an unpolarized wave is proportional to the identity matrix. In accordance with Eqs. (1.31) and (1.32), the average power generated on a receiving antenna is thus independent of its polarization.

Between these two extreme cases ($0 < |\mu_{xy}| < 1$), the waves carry part of the energy in polarized form (fixed polarization ellipse) with the remainder in an unpolarized form. When polarization diversity is available, a radar system can fully exploit the polarized component of the wave (maximum coupling), without however eliminating a ‘polarization noise’ generated by its unpolarized part. The subject of the next section will involve quantifying the proportion of each of these components in the composition of the wave.

1.1.3.2 The wave decomposition theorem

The previous section intuitively led to the representation of a partially polarized wave as the superposition of a polarized wave and an unpolarized wave. In [Born, 1970] (p. 551), the reader may find both a demonstration supporting this intuition and a proof of the uniqueness of this decomposition.

How do we proceed in practice for the ‘addition’ of two waves? Instinctively, we would simply add the respective Jones vectors (i.e. their polarization states). However this addition of electric fields (also known as the *coherent sum*) is only relevant if the phase offset between the two waves is constant, which requires that they be polarized. This is clearly not the case here because one of the two waves is unpolarized, which means only the addition of the energy terms is meaningful. In order to carry this out, we adopt the formalism of coherence matrices. According to Eq. (1.32), we see that the simultaneous effect of two waves received at an antenna is the same as that of a single wave whose coherence matrix is equal to the sum of the two coherence matrices. In this context, the energy terms are added (i.e. the J_{zt} terms) and the sum is referred to as *incoherent*.

Given the uniqueness of the decomposition, $[\bar{J}]$ must be written as:

$$[\bar{J}] = [\bar{J}^{upol}] + [\bar{J}^{pol}] = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} + \begin{bmatrix} B & D \\ D^* & C \end{bmatrix} \quad (1.34)$$

where A, B, C are real quantities and:

$$B \cdot C - |D|^2 = 0 \quad (1.35)$$

The structure of $[\bar{J}^{upol}]$ is characteristic of an unpolarized wave whereas condition (1.35) expresses the polarized nature of the wave. By substitution, we obtain:

$$A = \frac{1}{2} \cdot (J_{xx} + J_{yy}) - \frac{1}{2} \cdot \sqrt{(J_{xx} + J_{yy})^2 - 4 \cdot |\bar{J}|}$$

$$\begin{aligned}
B &= \frac{1}{2} \cdot (J_{xx} - J_{yy}) + \frac{1}{2} \cdot \sqrt{(J_{xx} + J_{yy})^2 - 4 \cdot |\bar{J}|} \\
C &= \frac{1}{2} \cdot (J_{yy} - J_{xx}) + \frac{1}{2} \cdot \sqrt{(J_{xx} + J_{yy})^2 - 4 \cdot |\bar{J}|} \\
D &= J_{xy}
\end{aligned} \tag{1.36}$$

with:

$$|\bar{J}| = |J_{xx} \cdot J_{yy} - J_{xy} \cdot J_{yx}| \tag{1.37}$$

Due to the Schwarz inequality, $|\bar{J}| \geq 0$.

The degree of polarization

After decomposition of the coherence matrix, we introduce a parameter to quantify the respective weights of the polarized and unpolarized components in the composition of the wave: the *degree of polarization* P_w is defined as the ratio of the power density contained in the polarized part of the wave to the total power density carried by the wave. By construction, P_w is between 0 and 1. Based on Eqs. (1.32) to (1.36), it may be shown that:

$$P_w = \sqrt{1 - \frac{4 \cdot |\bar{J}|}{(J_{xx} + J_{yy})^2}} \tag{1.38}$$

Only the determinant and the trace of the coherence matrix $[\bar{J}]$ enter into the degree of polarization P_w . These two quantities are independent of the choice of the basis $[\hat{x}, \hat{y}]$ of the polarization plane, as is P_w . For this reason, P_w is thus said to be ‘*polarimetric invariant*’. This is not the case for the degree of coherence μ_{xy} .

1.1.3.3 The group of Pauli matrices and the Stokes parameters

The polarization state of a wave is given by its coherence matrix (Sect. 1.1.3.1). We shall now express it in the Pauli matrices basis $([\bar{\sigma}_0], [\bar{\sigma}_1], [\bar{\sigma}_2], [\bar{\sigma}_3])$. The advantages of this approach will become obvious as we progress in our study of radar polarimetry:

$$\begin{aligned}
[\bar{\sigma}_0] &= \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & [\bar{\sigma}_1] &= \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\
[\bar{\sigma}_2] &= \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & [\bar{\sigma}_3] &= \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}
\end{aligned} \tag{1.39}$$

$[\bar{J}]$ is then written in the form:

$$\begin{aligned}
[\bar{J}] &= \frac{1}{\sqrt{2}} \cdot \{s_0 \cdot [\bar{\sigma}_0] + s_1 \cdot [\bar{\sigma}_1] + s_2 \cdot [\bar{\sigma}_2] + j \cdot s_3 \cdot [\bar{\sigma}_3]\} \\
&= \frac{1}{2} \cdot \begin{bmatrix} s_0 + s_1 & s_2 - j \cdot s_3 \\ s_2 + j \cdot s_3 & s_0 - s_1 \end{bmatrix}
\end{aligned} \tag{1.40}$$

The parameters s_i were introduced by Stokes at the end of the 19th century within the context of his study of polarized light. They provide the four components of the Stokes vector \tilde{S} :

$$\tilde{S} = \begin{pmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} J_{xx} + J_{yy} \\ J_{xx} - J_{yy} \\ J_{xy} + J_{yx} \\ j \cdot (J_{xy} - J_{yx}) \end{pmatrix} = \begin{pmatrix} \langle E_{0x} \cdot E_{0x}^* \rangle + \langle E_{0y} \cdot E_{0y}^* \rangle \\ \langle E_{0x} \cdot E_{0x}^* \rangle - \langle E_{0y} \cdot E_{0y}^* \rangle \\ 2 \Re \langle E_{0x} \cdot E_{0y}^* \rangle \\ -2 \Im \langle E_{0x} \cdot E_{0y}^* \rangle \end{pmatrix} \quad (1.41)$$

The respective weights of the polarized and unpolarized components of the wave may be evaluated by using the Stokes formalism. The positive nature of the determinant of $[\tilde{J}]$ is expressed by the inequality:

$$s_0^2 \geq s_1^2 + s_2^2 + s_3^2 \quad (1.42)$$

The degree of polarization P_W (Sect. 1.1.3.2) is then written as:

$$P_w = \frac{\sqrt{s_1^2 + s_2^2 + s_3^2}}{s_0} \quad (1.43)$$

The particular cases encountered are:

- (1) $P_W = 1$, the wave is polarized: $s_0^2 = s_1^2 + s_2^2 + s_3^2$;
- (2) $P_W = 0$, the wave is unpolarized: $s_1^2 + s_2^2 + s_3^2 = 0$.

Finally, the decomposition of waves into an incoherent sum of a polarized wave and an unpolarized wave (Sect. 1.1.3.2) is now written as:

$$\begin{aligned} \tilde{S} &= \begin{pmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{pmatrix} = \tilde{S}^{upol} + \tilde{S}^{pol} \\ &= \begin{pmatrix} s_0 - \sqrt{s_1^2 + s_2^2 + s_3^2} \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \sqrt{s_1^2 + s_2^2 + s_3^2} \\ s_1 \\ s_2 \\ s_3 \end{pmatrix} \end{aligned} \quad (1.44)$$

By construction, the degree of polarization associated with \tilde{S}^{upol} is zero (unpolarized wave), while that associated with \tilde{S}^{pol} satisfies the condition for the polarized wave ($P_W = 1$).

1.1.3.4 The Poincaré sphere

The Poincaré sphere is used to characterize the polarization state of a partially polarized wave described by a Stokes vector \tilde{S} and a normalized power density ($s_0 = 1$). The

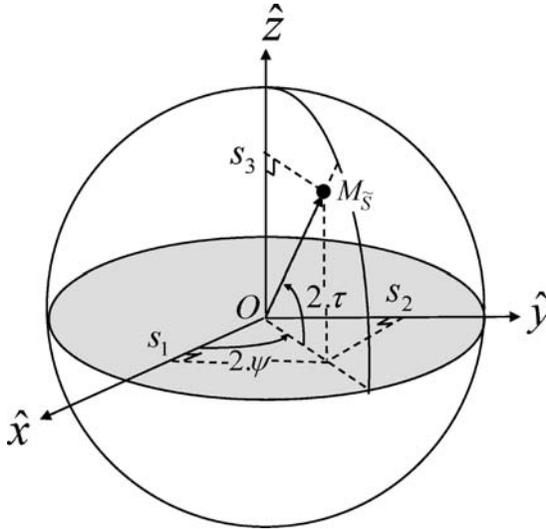


Fig. 1.6 The Poincaré sphere.

polarization state is represented by a point within a sphere of radius 1, with a reference coordinate system $[0, \hat{x}, \hat{y}, \hat{z}]$ (Fig. 1.6). For this, s_1 , s_2 , s_3 are expressed by means of two angles ψ and τ such that:

$$\begin{aligned}
 s_1 &= \sqrt{s_1^2 + s_2^2 + s_3^2} \cdot \cos(2\tau) \cdot \cos(2\psi) \\
 s_2 &= \sqrt{s_1^2 + s_2^2 + s_3^2} \cdot \cos(2\tau) \cdot \sin(2\psi) \\
 s_3 &= \sqrt{s_1^2 + s_2^2 + s_3^2} \cdot \sin(2\tau)
 \end{aligned} \tag{1.45}$$

In the context of the Poincaré sphere, the \vec{S} vector is associated with the internal point $M_{\vec{S}}$ that has the azimuth and elevation angles 2ψ and 2τ respectively and that is separated from the centre O of the sphere by a distance equal to its degree of polarization $\sqrt{s_1^2 + s_2^2 + s_3^2} \leq 1$.

In the specific case of a normalized polarized wave ($s_1^2 + s_2^2 + s_3^2 = s_0^2 = 1$), it can easily be demonstrated that the parameters ψ and τ of Eq. (1.45) are those introduced in the Sect. 1.1.2.2. to characterize the orientation of the major axis and the ellipticity (including sign) of the polarization ellipse. The representative point of a wave of this type is located on the surface of the Poincaré sphere. If the polarization of the wave is only partial, ψ and τ then characterize the polarization ellipse of its polarized part, and the wave's representative point 'sinks' into the volume of the sphere to the extent that its degree of polarization decreases. At the limiting value of 0, the point falls to the centre O of the sphere characterizing an unpolarized wave. Moreover, the points associated with linear polarizations ($\tau = 0$) are located on the equator, the right polarizations in

the Southern hemisphere ($\tau < 0$) and the left polarizations in the Northern hemisphere ($\tau > 0$). Finally the North and South poles represent, respectively, the left circular and right circular polarizations of perfectly polarized waves.

1.1.4 In passing: the elegant algebra of the $SU(2)\text{-}\mathbf{O}^+(3)$ homomorphism

It is not necessary to master the subject of this last section for the comprehension of the general physics of our subject. It deals with the relation between the Jones formalism and the Stokes formalism from an algebraic point of view which, will lead to an interesting generalization for the study of polarimetric targets (Sect. 5.8.1.1). The following development has been adapted from the excellent paper by Cloude [1986].

Equation (1.27) of Sect. 1.1.2.4 examines the transformation of the Jones vector undergoing a change of basis:

$$\vec{E}_{(\hat{a}, \hat{b})} = [\vec{U}_2] \cdot \vec{E}_{(\hat{x}, \hat{y})} \quad (1.46)$$

where $[\vec{U}_2]$ is an $SU(2)$ matrix, (Sect. 1.1.2.2).

There is a homomorphism⁵ of the $SU(2)$ group towards the group of real 3×3 orthogonal matrices, and this homomorphism associates the basis transition matrix $[\vec{O}_3]$ to any matrix $[\vec{U}_2]$ of $SU(2)$, with $[\vec{O}_3]$ defined by:

$$[\vec{O}_3]_{(p,q)} = \text{Tr} \left\{ [\vec{U}_2]^{*T} \cdot [\vec{\sigma}_p] \cdot [\vec{U}_2] \cdot [\vec{\sigma}_q] \right\} \quad (1.47)$$

where $\text{Tr} \{ \cdot \}$ is the trace operator.

The transformation equation uses Pauli matrices (Sect. 1.1.3.3), with the indices p and q varying from 1 to 3 (and not 0 to 3...). The equivalent of equation (1.46) in the real 3-dimensional space transform any real vector \vec{r} into a real vector \vec{r}' by:

$$\vec{r}' = [\vec{O}_3] \cdot \vec{r} \quad (1.48)$$

How can we now link the real vector \vec{r} , with dimension 3, to the initial complex vector $\vec{E}_{(\hat{x}, \hat{y})}$, with dimension 2? For this, we simply return to the coherence vector and the Stokes vector as defined by the Eqs. (1.32), (1.40) and (1.41). Even though the Stokes vector is real and has dimension 4, its basis transition matrix is of the same type as that operating on the \vec{r} vector. It can be shown that:

$$\vec{S}' = [\vec{M}] \cdot \vec{S} \quad (1.49)$$

⁵A homomorphism, ϕ , is an algebraic function relating two groups G (initial set) and G' (resultant set), with the operations “ \cdot ” and “ \circ ” respectively. For all elements x and $y \in G$, we obtain: $\phi(x, y) = \phi(x) \circ \phi(y)$.

with:

$$\left[\bar{\bar{M}} \right]_{(p,q)} = \text{Tr} \left\{ \left[\bar{\bar{U}}_2 \right]^{*T} \cdot \left[\bar{\bar{\sigma}}_p \right] \cdot \left[\bar{\bar{U}}_2 \right] \cdot \left[\bar{\bar{\sigma}}_q \right] \right\} \quad (1.50)$$

for indices p and q varying now from 0 to 3. $\left[\bar{\bar{M}} \right]$ is an extended form of the matrix $\left[\bar{\bar{O}}_3 \right]$:

$$\left[\bar{\bar{M}} \right] = \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & \left[\bar{\bar{O}}_3 \right] & \\ 0 & & & \end{array} \right] \quad (1.51)$$

These formulations have a precise mathematical meaning consistent with the results determined in the previous sections. They allow for the projection of a state of ‘pure’ polarization into the real 3-dimensional space, represented by the surface of the Poincaré sphere. Additionally, they include the case of partial polarization by assigning a 4-dimensional vector represented by a point within the Poincaré sphere to a Jones vector perturbed by space or time fluctuations.

1.1.4.1 On the Hermitian nature of the coherence matrix

The wave coherence matrix already described in Sect. 1.1.3.1, is Hermitian. Its eigenvalues λ_1 and λ_2 are thus positive, and its eigenvectors are orthogonal. It thus allows us to write a decomposition of the following form:

$$\left[\bar{\bar{J}} \right] = \lambda_1 \cdot \left[\bar{\bar{J}}_1 \right] + \lambda_2 \cdot \left[\bar{\bar{J}}_2 \right] \quad (1.52)$$

where $\det \left(\left[\bar{\bar{J}}_1 \right] \right) = \det \left(\left[\bar{\bar{J}}_2 \right] \right) = 0$. The two matrices are said to have “rank 1”. The decomposition (1.52) reveals an incoherent sum of two polarized orthogonal waves, which offers an alternative decomposition to that described in Sect. 1.1.3.2. If $\lambda_1 = \lambda_2$, the wave is unpolarized. If $\lambda_2 = 0$ (the smallest of the eigenvalues), the wave is polarized. The degree of polarization, expressed as a function of the eigenvalues λ_1 and λ_2 , is given by:

$$P_w = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \quad (1.53)$$

an equation which is in accordance with Eq. (1.38).

Please note that we reconstruct the decomposition described in Sect. 1.1.3.2 by rewriting Eq. (1.52) in the following form:

$$\begin{aligned} \left[\bar{\bar{J}} \right] &= (\lambda_1 - \lambda_2) \cdot \left[\bar{\bar{J}}_1 \right] + \lambda_2 \cdot \left(\left[\bar{\bar{J}}_1 \right] + \left[\bar{\bar{J}}_2 \right] \right) \\ &= (\lambda_1 - \lambda_2) \cdot \left[\bar{\bar{J}}_1 \right] + \lambda_2 \cdot \left[\bar{\bar{I}} \right]_2 \end{aligned} \quad (1.54)$$

where $\left[\bar{\bar{I}} \right]_2$ is the 2×2 identity matrix. $\left[\bar{\bar{J}}_1 \right]$ and $\left[\bar{\bar{I}} \right]_2$ then represent the polarized and unpolarized parts of the wave, respectively.

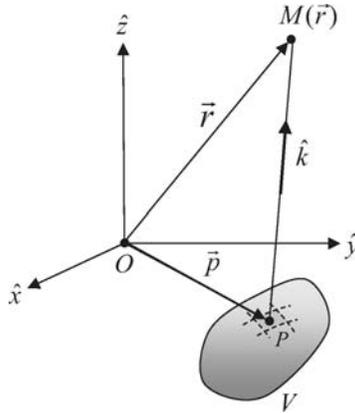


Fig. 1.7 Radiation into space from distributed electromagnetic sources confined within a volume V .

1.2 The electromagnetic radiation of microwave antennas

1.2.1 Introduction

The function of an antenna is to ensure the gradual transformation of electromagnetic energy transmitted along a feeder (coaxial cable, waveguide, etc.) into radiation in free space. In this section, we give the equations behind the principles of electromagnetic radiation; these will be detailed for the specific case of rectangular plane aperture. From these we can deduce the practical formulas that characterize its antenna pattern (3 dB beam width or Half Power Beam Width (HPBW), level of the side lobes). These are applied in the preliminary stages of SAR design. We then explain the principles of array antennas and conclude this rapid overview with a look at the technology of spaceborne SAR antennas.

1.2.2 The electromagnetic radiation equation

The Maxwell equations (Sect. 1.1.1), together with the *constitutive equations* in a vacuum $\vec{d}(\vec{r}, t) = \epsilon_0 \cdot \vec{\epsilon}(\vec{r}, t)$ and $\vec{b}(\vec{r}, t) = \mu_0 \cdot \vec{h}(\vec{r}, t)$, and the *conservation equation* that link the charge density $\rho(\vec{r}, t)$ and the current volume density $\vec{j}(\vec{r}, t)$ (Sect. 1.1.1) lead to the equation for electromagnetic radiation in an infinite medium including a volume V (Fig. 1.7), where the full range of sources are distributed ($\vec{j}(\vec{r}, t) = \vec{0}$ and $\rho(\vec{r}, t) = 0$ if $\vec{r} \notin V$):

$$\nabla^2 \vec{e}(\vec{r}, t) + k^2 \cdot \vec{e}(\vec{r}, t) = \frac{j}{\omega \cdot \epsilon} \cdot (\nabla(\nabla \cdot \vec{j}(\vec{r}, t)) + k^2 \cdot \vec{j}(\vec{r}, t)) \quad (1.55)$$

The radiated electric field $\vec{e}(\vec{r}, t)$ is formed both by the intrinsic nature of the electromagnetic propagation (given by the structure of the differential equation on the left

side), as well as by the distribution of the currents $\vec{j}(\vec{r}, t)$ inside V , (the right side of this differential equation). These source terms show how the current terms ($\vec{j}(\vec{r}, t)$) and the charge terms ($\nabla \cdot \vec{j}(\vec{r}, t)$, proportional to $\rho(\vec{r}, t)$) combine in the radiation process. For antennas it is usually assumed that $\rho(\vec{r}, t) = 0$, given the metallic nature of the surfaces.

1.2.3 Resolving the electromagnetic radiation equation

The resolution of Eq. (1.55) for time-harmonic fields gives an expression for the radiated field $\vec{e}(\vec{r}, t)$ throughout space, even in the immediate vicinity of the antenna (known as the near-field region) for which the expression $\vec{e}(\vec{r}, t)$ can be developed much more thoroughly [Kong, 1990]. The expression is less complex in the far field where the wave takes on a spherical appearance and is therefore locally plane (Sect. 1.2.5.2). Letting $\vec{k} = k \cdot \hat{k}$ (Fig. 1.7):

$$\vec{e}(\vec{r}, t) \approx -j \cdot \frac{\exp(-j \cdot k \cdot r)}{4 \cdot \pi \cdot r} \cdot k \cdot \xi \cdot \hat{k} \times \left[\hat{k} \times \int_s \vec{j}(\vec{p}, t) \cdot \exp(j \cdot \vec{k} \cdot \vec{p}) \cdot d\vec{p} \right] \quad (1.56)$$

with the wave impedance $\xi \approx 120 \cdot \pi$ in the vacuum. Point O can be any point in space, but is usually chosen as the center of a sphere centered on the antenna. Equation (1.56) expresses the coherent contribution of elements on the surface of V (with their respective currents $\vec{j}(\vec{p}, t)$) to the construction of $\vec{e}(\vec{r}, t)$. For a given direction, the complex amplitude of $\vec{e}(\vec{r}, t)$ varies as $\exp(-j \cdot k \cdot r)/r$, which is characteristic of the far-field radiation of a point source. It can also be shown that the electric field $\vec{e}(\vec{r}, t)$, and magnetic field $\vec{h}(\vec{r}, t)$ of the radiated wave in a far field satisfy the relationship: $\vec{e}(\vec{r}, t) = \xi \cdot \vec{h}(\vec{r}, t) \times \vec{k}$, which expresses the local plane structure of the radiated wave. The complex Poynting vector (whose amplitude is equal to the power density per unit of surface radiated by the wave) is given by:

$$\vec{P} = \frac{1}{2} \cdot \vec{e}(\vec{r}, t) \times \vec{h}^*(\vec{r}, t) = \frac{1}{2\xi} \cdot \|\vec{e}(\vec{r}, t)\|^2 \cdot \hat{k} \quad (1.57)$$

1.2.4 Antenna pattern, directivity and gain

The ability of an antenna to ‘focus’ radiated energy in a given direction is characterized with respect to the behavior of a theoretical isotropic antenna that transmits uniform power density in all directions. If we rewrite Eq. (1.56) as $\vec{e}(\vec{r}, t) = \exp(-j \cdot k \cdot r)/r \cdot \vec{F}(\vec{k})$, then the quantity $\|\vec{F}(\vec{k})\|^2$ represents the energy distribution law for energy radiated into space. Its directional variations define the pattern of the antenna, with its directivity given by:

$$D(\vec{k}) = \frac{4\pi \cdot \|\vec{F}(\vec{k})\|^2}{\int_0^{4\pi} \|\vec{F}(\hat{k})\|^2 \cdot d\Omega} \quad (1.58)$$

where $d\Omega$ represents a differential solid angle element. In the specific case of an isotropic antenna, $\|\vec{F}(\hat{k})\|^2$ is constant and $D(\hat{k}) = 1$. The antenna gain is given by:

$$G(\hat{k}) = \eta \cdot D(\hat{k}) \quad (1.59)$$

where η is its efficiency. The gain and the directivity are identical only for a lossless antenna.

1.2.5 The radiation of planar antennas

The exact resolution of Eq. (1.55) leads to Kottler's integral formulas for planar antennas. In the case of a SAR, the antenna is large compared to the wavelength ($> 10 \cdot \lambda$, which means that any perturbation created by the edge of the antenna can be ignored).⁶ The electric field radiated in the far field region is therefore given by Fresnel's formula:

$$\vec{e}(\vec{r}, t) \approx j \cdot \frac{1 + \cos \theta}{2 \cdot \lambda \cdot r} \cdot \iint_s \vec{e}(\vec{p}, t) \cdot \exp(-j \cdot k \cdot r') \cdot dS \quad (1.60)$$

We shall consider only rectangular planar apertures. If we rewrite (60) using the geometry shown in Fig. 1.8, we note that for the far field:

$$r' \approx r - p \cdot \cos(\vec{p}, \vec{r}) \quad (1.61)$$

and thus:

$$\begin{aligned} \vec{e}(\vec{r}, t) \approx j \cdot \frac{1 + \cos \theta}{2 \cdot \lambda \cdot r} \cdot \exp(-j \cdot k \cdot r) \cdot \iint_s \vec{e}(\vec{p}, t) \\ \cdot \exp[j \cdot k \cdot \sin \theta \cdot (x \cdot \cos \phi + y \cdot \sin \phi)] \cdot dx \cdot dy \end{aligned} \quad (1.62)$$

Along the axis ($\theta = 0$), the phase of the integral is zero at all points: all the elements of the antenna radiate in phase, producing a radiated field of maximum amplitude in this direction (as long as the phase shift due to the illumination does not change the direction of the beam, Sect. 1.2.6). For slightly off-axis (paraxial) directions, where $(1 + \cos \theta)/2 \approx 1$, Eq. (1.62) describes the *radiated field* $\vec{e}(\vec{r}, t)$, characterized by the direction variables $k \cdot \sin \theta \cdot \cos \phi$ and $k \cdot \sin \theta \cdot \sin \phi$, as *the two dimensional Fourier Transform (FT) of the illumination distribution* $\vec{e}(\vec{p}, t)$. This result is the basic law of antenna radiation which is used to select an illumination law to obtain the desired antenna pattern (the illumination law is the inverse Fourier transform of the pattern). The specific case of a uniform illumination law leads to a cardinal sine function.

⁶Note that in the case of low frequency radars (VHF or P band, i.e., for a carrier frequency ranging typically between 30 MHz and 600 MHz) the wavelength may not be much smaller than the antenna.

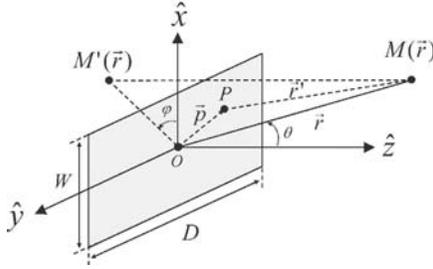


Fig. 1.8 Radiation into space from a rectangular planar antenna.

1.2.5.1 Antenna pattern of a rectangular planar antenna

The antenna pattern is generally characterized by its principal planes $\varphi = 0$ (elevation pattern) and $\varphi = \pi/2$ (azimuth pattern). Let us consider for example the azimuth pattern and a uniform illumination with linear polarization along \hat{x} : $\vec{e}(\vec{p}, t) = e_0(t) \cdot \hat{x}$ (vertical polarization). Under paraxial conditions, Eq. (1.62) gives:

$$\vec{e}(\vec{r}, t) \approx j \cdot \hat{x} \cdot e_0(t) \cdot W \cdot D \cdot \frac{\exp(-j \cdot k \cdot r)}{\lambda \cdot r} \cdot \text{sinc}(k \cdot \theta \cdot D/2) \quad (1.63)$$

where D and W are the antenna length and width, respectively, and $\text{sinc}(x) = \sin(x)/x$. The cardinal sine shape shows the effect of the FT (Sect. 1.4) for a uniform illumination law. This is the basic pattern of a SAR antenna in the azimuth direction. This is why it is necessary to account for the weight the echoes returned by a target T during radar illumination (Sect. 2.2.5.1). From Eq. (1.63), we derive the antenna radiation features needed for designing a SAR:

- *The HPBW (Half Power Beam Width - or 3 dB beam width):* The main lobe has a width given by $k \cdot \theta^w \cdot D/2 = 2 \cdot \pi$, which also gives $\theta^w = 2 \cdot \lambda/D$. The HPBW, defined by the angular sector (expressed in radians) in which the radiated power is larger than one half of that transmitted along the antenna axis, is approximated by:

$$\theta_{az}^{3dB} \approx \alpha \cdot \lambda/D \quad (1.64)$$

where α is a weighting coefficient (usually close to unity) dependent on the illumination law. Equation (1.64) is the equivalent of the resolving power, equal to the ratio λ/D , for an optical instrument with diameter D observing at a wavelength λ . The 3 dB lobe determines the size of an antenna's footprint.⁷ The larger the ratio of the antenna size to the wavelength, the smaller the footprint. The HPBW in elevation will thus be $\theta_{el}^{3dB} \approx \alpha' (\lambda/W)$. Space SAR antennas are much longer than they are wide, typically: $D \approx 10$ m, $W \approx 1$ m. In C-band ($\lambda \approx 5.6$ cm), $\theta_{az}^{3dB} \approx 5.7$ mrad, or approximately

⁷When considering backscattering (the signals are received after a round trip journey of the wave between the antenna and the target), the effect of the radiation pattern is applied twice: the HPBW corresponds, for this configuration, to the angular domain in which the power attenuation does not exceed 6 dB with respect to a round trip journey along the axis of the antenna.

0.3° (as a first approximation, we take $\alpha = 1$). At an altitude of 800 km, the size of the footprint in the azimuth direction will therefore be approximately 5 km, and in the transverse direction approximately 100 km (once the beam width has been ground projected as we are considering a side-looking radar here, cf. [Chapter 2](#)).

• *Gain*: From Eq. (1.58), the gain of a rectangular planar antenna along its axis is given by:

$$G^{max} = \eta \cdot 4\pi \cdot \frac{D \cdot W}{\lambda^2} = \eta \cdot 4\pi \cdot \frac{S_{ant}}{\lambda^2} = 4\pi \cdot \frac{S_{eff}}{\lambda^2} \quad (1.65)$$

The maximum gain of a planar antenna is approximately given by the ratio of its surface to the square of the wavelength multiplied by 4π (this approximation corresponds to the case for $\eta = 1$, which leads to directivity being equal to gain, and S_{ant} being equal to the effective surface S_{eff}). This formula will be used when selecting the power to be transmitted by radar imagers ([Chapter 3](#)).

• *Side lobes*: These are located in directions where the radiation from the various elements of the antenna are again in phase. The amplitude of the first side lobe with respect to that of the main lobe is given by $\sin c(k \cdot \theta \cdot D/2)$ for $k \cdot \theta \cdot D/2 = 3\pi/2$, which leads to a radiated power approximately -13.5 dB below that radiated along the main axis.⁸

If it becomes necessary to lower the level of the side lobes in order to improve the image quality of the SAR (Sect. 3.12.3), then we will use a non-uniform illumination whose amplitude decreases from the axis to the edges of the antenna. However, lowering the side lobes broadens the main lobe.

1.2.5.2 Defining the far-field zone

The far field of the antenna is the distance beyond which the radiated wave can be considered to be spherical (i.e., it is locally plane). This leads to the following conditions [[Combes, 1997](#)]:

$$r > 10 \cdot \Delta \quad (\text{amplitude condition}) \quad (1.66a)$$

and :

$$r > 2 \cdot \Delta^2/\lambda \quad (\text{phase condition}) \quad (1.66b)$$

where Δ is the largest dimension of the antenna being considered ($\Delta = \sqrt{D^2 + W^2}$ for the rectangular antenna). Condition (1.66(b)) becomes more constraining than (1.66(a)) when $\Delta > 5 \cdot \lambda$, which is the case for radar imaging antennas.

Example: In a spaceborne case, for a rectangular antenna with dimensions of $D \approx 10 \text{ m}$ and $W \approx 1 \text{ m}$, operating at C-band, ($\lambda \approx 5.6 \text{ cm}$) $\rightarrow 2 \cdot \Delta^2/\lambda \approx 3.5 \text{ km}$.

Note: Be careful with the synthetic aperture antennas that we discuss in [Chapter 2](#). Imagine for a moment that you are using an antenna 5 km in length. SAR processing is based on the principle of synthesizing the behavior of such an antenna, with its far field more than a million kilometers away from the radar instrument! A synthetic antenna always observes its target in its near field; we shall return to the implications of this comment in [Sect. 2.4.1](#).

⁸The value of the ratio in dB is 10 times the base 10 logarithm of this ratio.

1.2.6 Array antennas

Array antennas are composed of radiating elements powered by a common microwave source which distributes the available power in amplitude and in phase. Alternatively the power sources can be separated by using a Transmitter/Receiver (T/R) module for each element (active antenna) thus making the system more reliable.

By applying the law of energy distribution to the different radiating elements, we can model the antenna pattern of the array (this is the beam forming technique used in space telecommunications to cover a specific geographical area). When using SAR, the array effect allows the main lobe of the antenna to be oriented in specific directions, making such operating modes as SPOTLIGHT or SCANSAR possible.

We shall describe here the radiating principle of an array antenna for the specific case of a one-dimensional linear array made up of regularly spaced identical elements, as illustrated in Fig. 1.9. Each array element indexed i is powered uniformly with a complex distribution factor $a_i \exp(j \cdot \phi_i)$. Its contribution to the radiation is expressed as $\vec{e}_i(\vec{r}_i, t) \cdot a_i \exp(j \cdot \phi_i)$, where $\vec{e}_i(\vec{r}_i, t)$ is the radiated field and \vec{r}_i the position vector of the observation point with respect to the phase center of the radiating element.

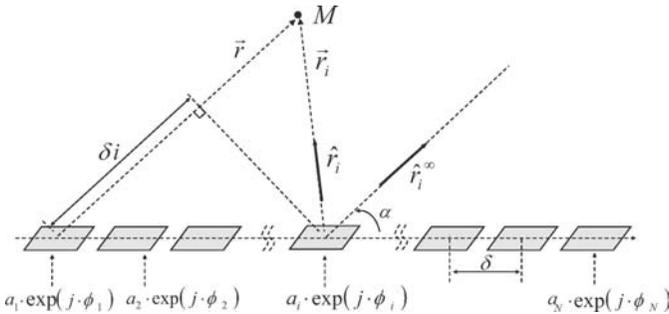


Fig. 1.9 Radiation from an array antenna in far field conditions.

In the far field, the observation directions \vec{r}_i^∞ are all parallel and the fields $\vec{e}_i(\vec{r}_i, t)$ are equal. If we take as a reference the distance r separating the observation point from the phase center of element No.1, the complex amplitude of the total radiated field can be written as:

$$\vec{e}^{tot}(\vec{r}, t) \approx \vec{e}^0(\vec{r}, t) \cdot \sum_{i=1}^{i=N} \exp(j \cdot \psi_i) \cdot a_i \exp(j \cdot \phi_i) \quad (1.67)$$

where $\vec{e}^0(\vec{r}, t)$ is the far field illuminated individually by any element of the array. If we now look at the geometry of Fig. 1.9, we can conclude:

$$\psi_i = (i - 1) \cdot k \cdot \delta \cdot \cos \alpha = (i - 1) \cdot \psi \quad (1.68)$$

1.2.6.1 Broadside arrays

When the energy is uniformly distributed among all the sources ($a_i = 1$, $\phi_i = 0$), and if we note that:

$$\begin{aligned} \sum_{i=1}^{i=N} \exp(j \cdot \psi_i) &= \frac{1 - \exp(j \cdot N \cdot \psi)}{1 - \exp(j \cdot \psi)} \\ &= \frac{\exp(j \cdot N \cdot \psi/2)}{\exp(j \cdot \psi/2)} \cdot \frac{\sin(N \cdot \psi/2)}{\sin(\psi/2)} \end{aligned} \quad (1.69)$$

the result is:

$$\begin{aligned} \vec{e}^{tot}(\vec{r}, t) &= N \cdot \vec{e}^0(\vec{r}, t) \cdot \frac{\exp(j \cdot N \cdot \psi/2)}{\exp(j \cdot \psi/2)} \cdot \frac{1}{N} \cdot \frac{\sin(N \cdot \psi/2)}{\sin(\psi/2)} \\ &= N \cdot \vec{e}^0(\vec{r}, t) \cdot F(\psi) \end{aligned} \quad (1.70)$$

Equation (1.70) reveals that a radiating array antenna has a behavior that can be described as the product of the number of radiating elements and the radiation characteristics of one element of the array, weighted by the array factor $F(\psi)$. Thus, working with an array causes an increase in gain by a factor of N^2 along the antenna axis (resulting in coherent summing of the contributions for $\psi = 0$, i.e. $\alpha = \pi/2$), but it also increases undesirable ‘secondary effects’ such as the creation of array lobes [Mailloux, 1994]. These appear when: $\psi/2 = m \cdot \pi$, where $m \in Z^*$, i.e.:

$$\cos \alpha = m \cdot \frac{\lambda}{\delta} \quad (1.71)$$

Array lobes no longer exist when:

$$\delta \leq \lambda \quad (1.72)$$

1.2.6.2 Offpointing the beam

The special SAR modes (SPOTLIGHT, Sect. 2.5.4.1 and SCANSAR, Sect. 2.5.4.2) require that the beam of the antenna be offpointed with respect to one of its principal planes (the azimuth or the elevation plane). The offpointing is achieved by introducing a phase gradient between elements. In this way, a weighting law such as ($a_n = 1$, $\phi_n = (i - 1) \cdot \phi_0$) leads to maximum radiation of the array in the direction α_0 such that

$$\cos \alpha_0 = -\frac{\lambda \cdot \phi_0}{2\pi \cdot \delta} \quad (1.73)$$

For this direction to exist:

$$|\phi_0| < \frac{2\pi \cdot \delta}{\lambda} \quad (1.74)$$

The minus sign in Eq. (1.73) indicates that the beam is shifted towards the elements with negative phase shifts. The array lobes are in the following directions:

$$\cos \alpha = m \cdot \frac{\lambda}{\delta} - \frac{\lambda \cdot \phi_0}{2\pi \cdot \delta} \quad (1.75)$$

They no longer exist when:

$$\delta < \frac{\lambda}{1 + |\cos \alpha_0|} \quad (1.76)$$

Apart from the off-pointing function, beam-forming techniques can be implemented using more sophisticated weighting laws (in amplitude and in phase) when associated with two-dimensional array geometries. The flexibility of this technique can be seen in Combes [1997].

1.2.7 SAR antenna technology

The design of current and future space SAR antennas depends on a variety of technologies. Passive antennas (such as those of ERS-1 and ERS-2), consisting of radiating elements whose role is limited to the energy transfer from a microwave source into free space, suffer from a lack of flexibility (neither the swath width nor the incidence angle can be changed).

If we seek better resolution (as in SPOTLIGHT mode), or a broader swath (as in SCANSAR mode), then the beam must be agile. This can be achieved using an array of phase-shifters at the interface between the power source and the radiating elements in the plane containing the phase gradient. The beam can be offpointed by 50° to 60° without mechanically manipulating the antenna. RADARSAT-1, launched in 1995, uses this technology in SCANSAR mode, for which the beam is offpointed in elevation.

Active antennas, which are made up of T/R modules that can be steered in both amplitude and phase, offer further advantages, such as reduced radiofrequency loss and increased reliability due to the distribution of active elements over the entire surface of the antenna. The ASAR/ENVISAT antenna uses this principle. It is composed of 320 T/R modules, allowing it to be offpointed in elevation (used for the SCANSAR mode).

The SPOTLIGHT mode, which aims to improve azimuth resolution, either by changing the yaw of the space platform or by offpointing the beam in azimuth, has been validated for use in the civil space field, specifically by TERRASAR-X and COSMO-SKYMED missions.

1.3 Interaction between waves and natural surfaces - the basics of radar measurement physics

1.3.1 Introduction

The purpose of this section is to make the reader familiar with the general principles of the interaction of electromagnetic waves with natural environments. Any interpretation of radar imagery, as well as the estimation of bio- or geophysical parameters on the basis of the imagery, requires a precise understanding of these interactions.

Electromagnetic radiation interacts with the natural environment causing surface and/or volume effects. Surface effects are conditioned by surface roughness, while volume effects are influenced by the density, the structure and the organization of the scatterers, for example within vegetation or snow. In both cases, dielectric properties, which reflect the water content of the observed media, play a key role. Finally, the radar echo is modulated by the local ground slope, which causes strong radiometric variations. In absolute terms, the use of interaction models is required to untangle the skein of all these contributions, but their development quickly runs into the difficulty – if not the impossibility – of simulating complex natural factors with mathematical models. Nevertheless, semi-empirical approaches inspired both by fundamental theoretical behavior (basically first-order surface scattering and volume radiative transfer) and by experimental observations, offer an alternative to the use of detailed models. We shall limit the theoretical discussion to a brief description of these fundamental mechanisms. We shall use a few examples to illustrate the main ‘physical’ effects observed in radar imagery: roughness, water content, penetration of waves and slope effect.

Finally, it should be noted that conventional, single-polarized radar does not give much information about the physical nature of artificial objects. They generally produce high-intensity signatures that are often saturated, thus spotting radar images with ‘bright points’. A more precise characterization of their structure requires polarization diversity in the radar waveforms ([Chap. 5](#)).

1.3.2 Surface scattering

Radar measurements can be used to evaluate ground-water content (by analyzing its dielectric variations), but in practice the impact of roughness is a severe obstacle to this estimation. This is one of the obstacles which has motivated the development of random surface electromagnetic scattering models. We shall discuss here the theoretical principles underlying these models.

1.3.2.1 Surface description

This refers to a dielectric surface separating two homogeneous semi-infinite media ([Fig. 1.10](#)). The relative dielectric constant⁹ ε_r of the lower layer characterizes the resistance of the medium to wave penetration. For a ground layer this depends on its free water content and, to a lesser extent, its texture. The roughness is characterized by a random process $z(x, y)$ reflecting local variations of the profile with respect to a mean reference plane. The random process $z(x, y)$ is usually assumed to be a stationary Gaussian variable with zero mean.

In the vertical direction, the r.m.s. (root mean square) height s characterizes the standard deviation of the rough surface with respect to the reference plane:¹⁰

$$s = \langle z^2(x, y) \rangle^{\frac{1}{2}} \quad (1.77)$$

⁹The relative dielectric constant ε_r is given by the ratio between the dielectric constant of the medium under study ε and that of the vacuum ε_0 : $\varepsilon_r = \varepsilon/\varepsilon_0$

¹⁰For cultivated surfaces, typical values of r.m.s. heights range between 1 and 2.5 cm for planted or harrowed plots and from 2 to 7 cm for plowed plots.

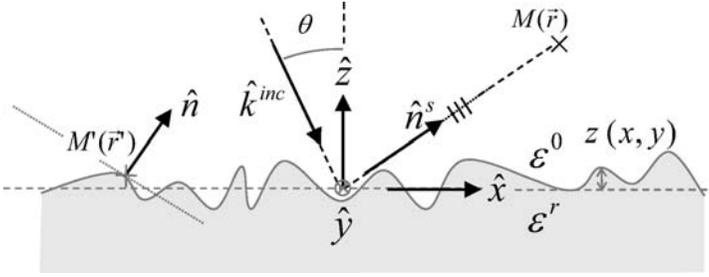


Fig. 1.10 Electromagnetic scattering from a rough surface.

In the horizontal plane, the surface auto-correlation function gives information on the rate of roughness variations in a given direction; for example, for the direction \hat{x}

$$K(\tau) = \frac{\langle z(x + \tau, y) \cdot z(x, y) \rangle}{\langle z^2(x, y) \rangle} \quad (1.78)$$

A similar formulation describes the rate of roughness variation for the direction \hat{y} by interchanging the variables x and y in Eq. (1.78). The correlation length l is the value for which:

$$K(l) = \frac{1}{e} \quad (1.79)$$

Sudden variations in roughness (broken profile) lead to a rapidly decreasing auto-correlation function, i.e. a short correlation length. The very nature of the auto-correlation function (generally assumed to be Gaussian or exponential) has a significant impact on the profile's morphology. Figure 1.11 shows a profile with an r.m.s. height $s = 2$ cm and a correlation length $l = 12$ cm (typical values), for the Gaussian case (Fig. 1.11(a)) and the exponential case (Fig. 1.11(b)).

Comments:

- (1) Mathematically, the auto-correlation function $K(\tau)$ is the spatial representation of the surface roughness spectrum (its Fourier transform). A rough surface is to some extent a superposition of randomly distributed periodic surfaces.
- (2) Using Eq. (1.78), we implicitly accept that the profiles are stationary, which means we can remove the space (x variable) dependence from the auto-correlation function. More realistic rugosity statistics (such as fractal statistics) do not allow for this type of simplification.
- (3) The hypotheses of stationarity and a single roughness scale do not generally match observations. An experimental soil profile of a cultivated surface (Fig. 1.12) shows that the soil/air interface consists of an aggregation of clumps of different scales and shapes, including numerous gaps and included volumes: excessive simplification often leads to poor modeling of the reality (compare Fig. 1.11 and Fig. 1.12!).

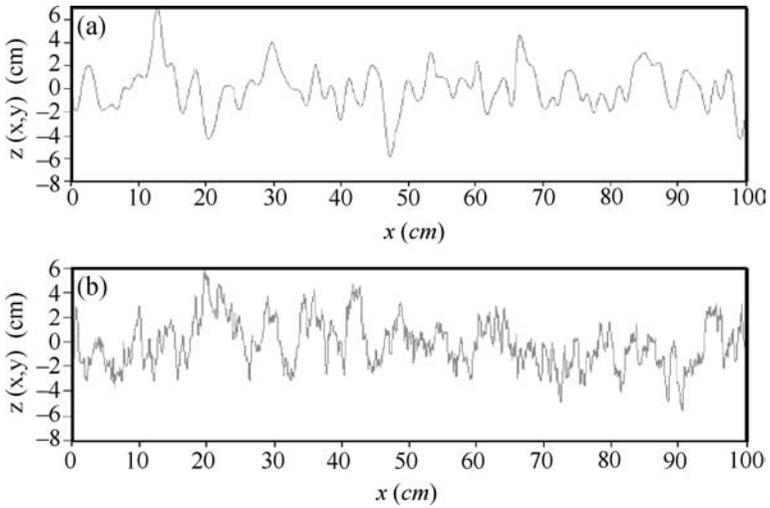


Fig. 1.11 Ground profile with r.m.s. height $s = 2$ cm and correlation length $l = 12$ cm; (a) Ground profile with Gaussian auto-correlation function; (b) Ground profile with exponential auto-correlation function.

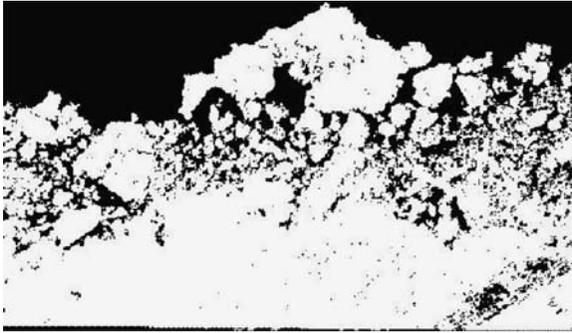


Fig. 1.12 Experimental ground profile acquired on cultivated land (Courtesy INRA).

The Rayleigh criterion

The effect of roughness on electromagnetic scattering depends on the transmitted wavelength, as the same surface may, for example, appear smooth in L band ($\lambda = 25$ cm) and rough in X band ($\lambda = 3$ cm). But is it possible, for a given wavelength, to distinguish between rough and smooth surfaces? The Rayleigh criterion stipulates that the roughness effect is negligible when the phase offset $\Delta\phi$ of the backscattered echoes generated by the altimetric variations of the surface (Fig. 1.13) does not exceed $\pi/2$. Taking into account the wave's round-trip path, if h is the maximum altimetric variation, one obtains:

$$h \leq \frac{\lambda}{8 \cdot \cos \theta} \quad (1.80)$$

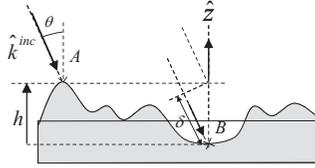


Fig. 1.13 Rayleigh criterion for a rough surface.

At an incidence angle of 45° , the $L(\lambda = 24 \text{ cm})$, $C(\lambda = 5.7 \text{ cm})$ and $X(\lambda = 3 \text{ cm})$ bands lead to maximum altimetric variations of 4.2 cm, 1 cm and 0.5 cm, respectively. Almost all natural surfaces thus behave as rough surfaces in X and C bands.

1.3.2.2 Calculation of electromagnetic scattering

Strictly speaking, this is governed by the Stratton-Chu equations, derived from Maxwell equations when they are coupled with the boundary conditions shown in Fig. 1.10. The field $\vec{E}^s(\vec{r})$ scattered at a point \vec{r} located far from the surface is given by the integral equation [Ulaby, 1986]:

$$\vec{E}^s(\vec{r}) = \alpha \cdot \hat{n}^s \times \int ((\hat{n} \times \vec{E}^s(\vec{r}')) - \xi \cdot \hat{r} \times (\hat{n} \times \vec{H}^s(\vec{r}'))) \cdot \exp(-j \cdot \vec{k} \cdot \vec{r}') \cdot dS' \quad (1.81)$$

where $\alpha = -\frac{j \cdot k}{4\pi R} \cdot \exp(-j \cdot k \cdot R)$ is a propagation factor. The parameter \hat{n}^s characterizes the propagation direction of the scattered wave, and \hat{n} is the normal to the exterior surface at a point M' (located at \vec{r}') on the surface. ξ is the wave impedance in the medium (Sect. 1.1.2.1). Finally, $\hat{n} \times \vec{E}^s(\vec{r}')$ and $\hat{n} \times \vec{H}^s(\vec{r}')$ represent the tangential components of the electric and magnetic fields at the interface and are the unknowns of the integral equation. Equation (1.81) should be compared to the radiation equation for antennas – Eq. (1.56) of Sect. 1.2. There is nevertheless one difference: the ground behaves like a dielectric antenna, whence the simultaneous contribution to its radiation of electrical and magnetic components.

The integral equation (1.81) takes into account all of the scattering mechanisms, whether single or multiple. This equation can only be resolved by using numerical methods (e.g. the method of moments). In the field of remote sensing we generally use asymptotic methods which are only valid for limited ranges of roughness.

Asymptotic method

The backscattering behavior¹¹ of rough surfaces is described by making approximations about the description of the interface and in the related electromagnetic calculations. The descriptive part is based on a ‘standard’ representation of the interface (stationary with no volume inclusions, most of the time with a single roughness scale), defined by its r.m.s. height s and the correlation length l):

¹¹Backscattering is the particular configuration for which the scattering is observed from the transmission direction. This is the case when the same antenna is both the transmitter and the receiver.

- The *Physical Optics* approximation (otherwise known as the *Kirchhoff* approximation) [Ulaby, 1986] breaks the surface down into a succession of facets (Fig. 1.14(a)). The validity condition¹² is expressed by $k \cdot l > 6$ and $l^2 > 2.76 \cdot s \cdot \lambda$.
- For very rough surfaces ($k \cdot s \geq 3$), the *Geometric Optics* approximation reduces the interaction to the facets orthogonal to the incident wave (Fig. 1.14(b)).
- The *Small Perturbations* model is based on the exact resolution of Eq. (1.81) to a given order. The first order is sufficient for characterizing backscattering for direct polarization (e.g. horizontal transmission, horizontal reception). An expansion up to the second order is necessary to evaluate cross polarization terms (e.g. horizontal transmission, vertical reception). The small perturbations method is only applied to surfaces that are slightly rough and that are not likely to significantly alter the wave structure (Fig. 1.14(c)). The validity condition depends on the r.m.s. of the slope (amounting to $\sqrt{2} \cdot s/l$ for a Gaussian auto-correlation function) and the product $k \cdot s$, with both being less than 0.3.
- The *Integral Equation Model (IEM)* [Fung, 1992] is more recent than the previous methods, with a validity domain that includes those of physical optics and small perturbations; ¹³ this extension was made possible by integrating geometric specificities of the backscattering configuration into the calculation.

1.3.2.3 Experimental simulations and observations

The case of bare soils

Despite their limitations, asymptotic models reveal the main tendencies of the electromagnetic behavior of bare soils. Figure 1.15(a) shows the backscattering coefficient σ_{vv}^0 , of the ERS configuration (C band, 23°, VV), simulated by the IEM model for different vertical roughness states (r.m.s. height) and water content.¹⁴ The auto-correlation function here has an exponential form while the correlation length is set to a typical value of $l = 8$ cm (generally difficult to estimate since l is often used as an fitting parameter).

For a given water content, we see an *increase in the radar response when the roughness increases*: in a radar image, the bright surfaces are rough surfaces and the dark surfaces are smooth surfaces as shown by the RADARSAT image (Fig. 1.15(b)), indicating a clear break between a smooth surface (loam fragments) and rough surface (limestone).

For a given roughness (fixed r.m.s. height), we see an *increase in the radar response as the water content increases* (of the order of 7 dBm²/m²): moist soils are ‘brighter’ than dry soils for comparable roughness values. In qualitative terms, a RADARSAT image of a farming region in the state of Washington (United States) (Fig. 1.15(c)) shows a collection of circular patches with variable radiometric content,

¹² $k = 2\pi/\lambda$ is the wave number.

¹³This may be applied for $k \cdot s \leq 3$.

¹⁴The definition of σ_0 and of its unity (dBm²/m²) is given in Sect. 3.3. At this stage of the discussion, let us assume that it is a quantity proportional to the amount of energy returned to the radar.

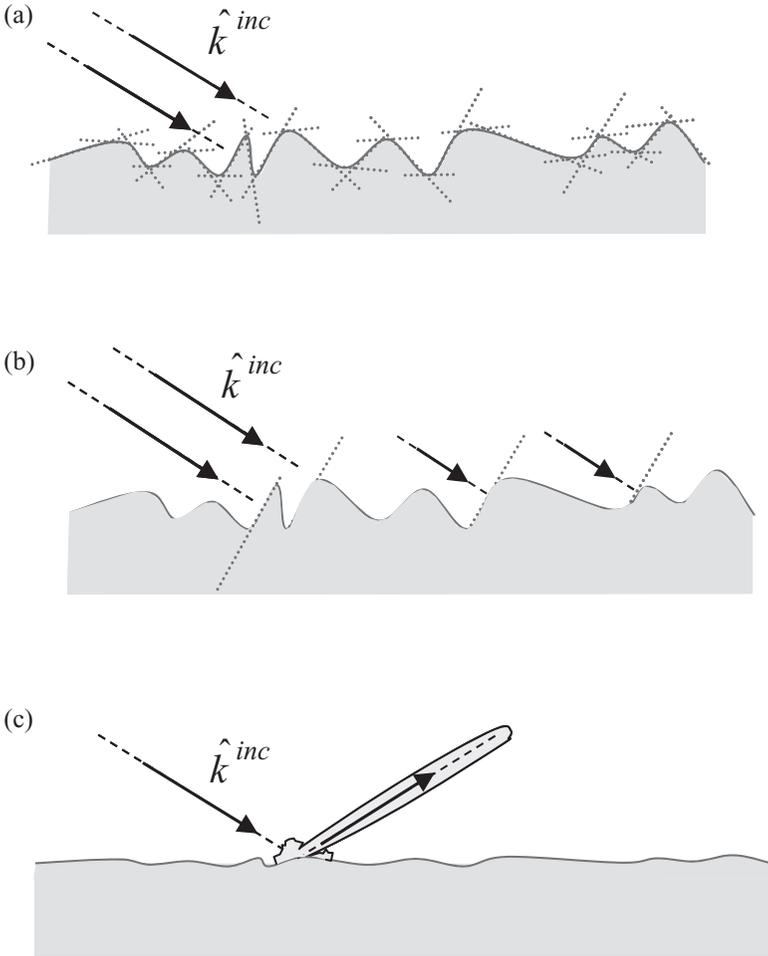


Fig. 1.14 (a) Physical Optics Approximation (or Kirchhoff approximation) for a rough surface; (b) Geometric Optics Approximation for a rough surface; (c) Small perturbation approximation for a rough surface.

each of which are an agricultural plot irrigated to a certain extent.¹⁵ Quantitatively, the experimental *ERS* observations (Fig. 1.15(d)) confirm measurement dynamics of the order of $6\text{--}7 \text{ dBm}^2/\text{m}^2$ for water contents ranging from 0 to 0.4 g/cm^3 .

In the light of these observations, attempts to estimate water content from radar measurements appear to be fraught with difficulties. The nested effects of roughness and water content lead to the *same level of backscattering for a rough and dry surface as for a smoother and damper surface*. The need to decorrelate water content and

¹⁵To reach this conclusion, it is necessary to assume the similarity of roughness states between two plots as well as the lack of vegetation.

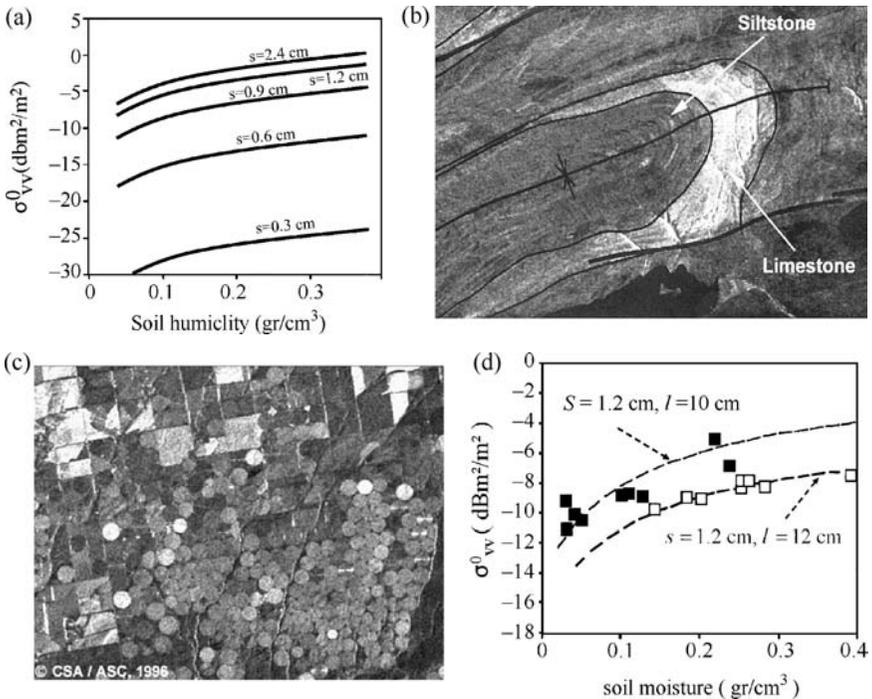


Fig. 1.15 (a) Theoretical backscattering coefficients for bare soils for the ERS configuration (C band, 23°, VV) as a function of vertical roughness (r.m.s. height) and water content. The auto-correlation function is exponential and the correlation length is set to 8 cm. (extract from [Le Toan, 1993]); (b) Effect of roughness on the SAR imaging radar. RADARSAT image (C band, 45°, resolution: 20 m) from: RADARSAT Geology Handbook (RADARSAT International), 1997; (c) Mode RSAT Stad 4, resolution 27 m, incidence angle 37°. Farming region in the state of Washington, USA. ©Canada Copyright Canada Centre for Remote Sensing - Applications in action!; (d) Experimental backscattering coefficient acquired by ERS (C band, 23°, VV) for two types of bare soil as a function of water content (agricultural area in Gharb, Morocco). The modeling (---) is conducted with $s = 1.2$ cm and $l = 10$ cm for the first zone (■), $s = 1.2$ cm and $l = 12$ cm for the second (□). Exponential auto-correlation function (from [Le Toan, 1993]).

roughness effects in radar measurements means that the radar information has to be diversified (for example, by using polarization diversity, Chap. 5).

Sea surfaces

At centimeter wavelengths, the sea is practically ‘impermeable’ to electromagnetic waves.¹⁶ Its surface reveals several scales of roughness (waves of capillarity on small scales, swell fields on larger scales, etc.). Relating wind speed and wave heights

¹⁶This is ignoring the effects of salinity likely to produce variations of the dielectric constant which are particularly sensitive to low frequencies.

reduces the surface description, in an overly simplified way, to a state information figure (expressed as a coefficient ss , for ‘sea state’, with values between 0 and 8 on the Sir Percy Henry Douglas scale (Head of the English Royal Navy Meteorological Survey in 1917): calm sea, weak to moderate, agitated, etc.

For backscattering, the incidence angle range is broken down into three regions: the specular zone (low incidence angles), the flat zone (average incidence angles) and grazing incidence angles (high incidence angles) (Fig. 1.16(a)). For an average sea state ($ss = 3$), the semi-empirical behavior [Morchin, 1990] of the backscattering coefficient σ^0 (averaged with respect to polarization effects and wind direction) indicates a preponderance of the specular component between 0° and 30° , with the width and amplitude of its lobe independent of the frequency. On the other hand, for the flat zone (where there is a predominance of incoherent scattering), σ^0 increases with the frequency with a $1/\lambda$ dependence. High radiometric variations (of the order of $20 \text{ dBm}^2/\text{m}^2$) are thus likely between 20° and 50° , corresponding to the junction between the specular and the flat zones, i.e., within the range of incidence angles typically covered by SAR. Radiometric variations which match this pattern have been observed in experiments with RADARSAT data (Fig. 1.16(b)), typically of the order of $10 \text{ dBm}^2/\text{m}^2$ for the swath of a Standard-S1 mode (20° to 27°).

The wind direction with respect to the radar viewing axis also affects the sea response. Compared to the ‘downwind’ configuration of the target (i.e., the wind direction and the radar beam are heading in opposite directions), the upwind σ^0 decreases on average by 7 to $8 \text{ dBm}^2/\text{m}^2$ (the wind direction and the observation direction are identical, and the sea has a smoother appearance) and by 3 to 4 dB with a crosswind.

For polarization, finally, we expect a higher backscattering for vertical polarization (VV) than for horizontal polarization (HH). When the incidence angle increases, the

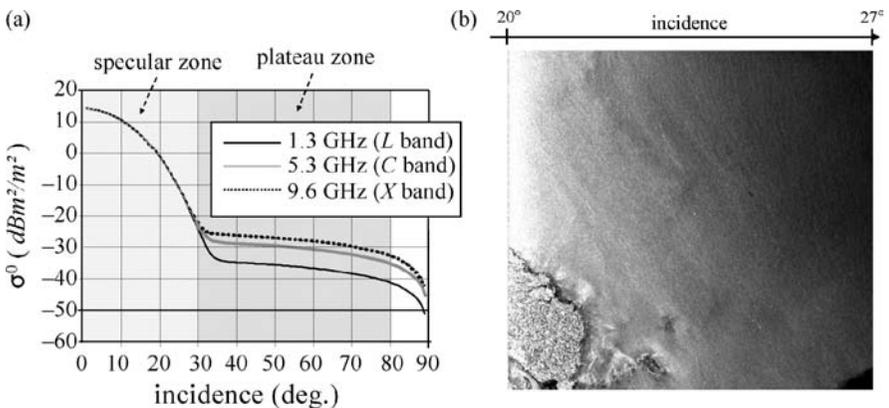


Fig. 1.16 (a) Semi-empirical behavior of the backscattering coefficient σ^0 as a function of incidence angle for an average sea state (Sea state $ss=3$ on the Douglas scale); (b) Incidence angle effect observed in a RADARSAT image, standard mode-1 ($20^\circ - 27^\circ$), acquired over sea (Ouessant rail, France, 09-03-1999), ©Radarsat international.

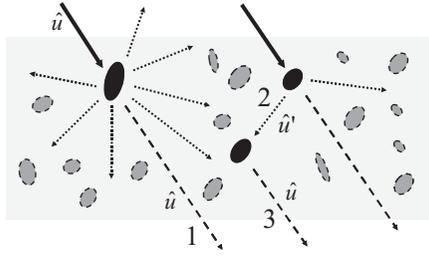


Fig. 1.17 Radiative transfer theory for a cloud of particles.

difference between HH and VV increases in favor of VV , with predictions locating σ_{HH}^0 at about $15 \text{ dBm}^2/\text{m}^2$ below the σ_{VV}^0 at a 60° incidence angle. The differences in polarization behavior have an impact on system engineering choices. The ERS-1/2 missions, which were originally designed for oceanographic studies, selected the VV polarization. The RADARSAT mission, on the other hand, chose the HH polarization (which minimized the sea clutter response and was thus more suitable for ship routing applications).

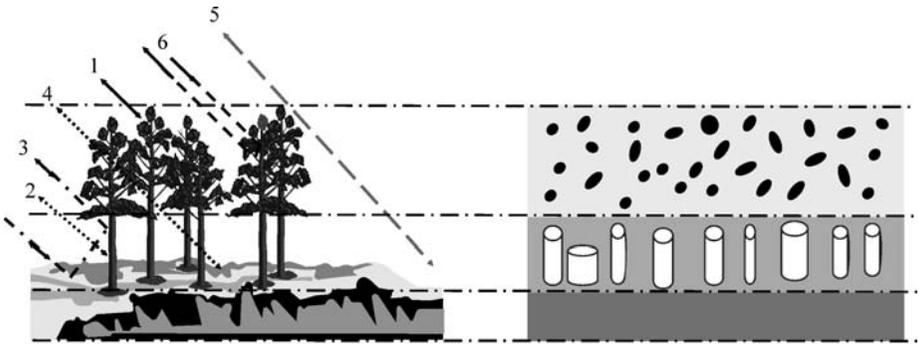
1.3.3 Volume scattering

The interaction of an electromagnetic wave with a cloud of scattering particles is described by the radiative transfer function [Chandrasekhar, 1960]:

$$\frac{d\tilde{S}(\vec{r}, \hat{u})}{du} = [\bar{\kappa}_e] \cdot \tilde{S}(\vec{r}, \hat{u}) + \int_{4\pi} [\bar{R}(\vec{r}, \hat{u}, \hat{u}')] \cdot \tilde{S}(\vec{r}, \hat{u}') \cdot d\Omega' \quad (1.82)$$

Here we see the Stokes vector \tilde{S} (Sect. 1.1.3.3) of a wave propagating in the direction \hat{u} simultaneously damped by the matrix $\bar{\kappa}_e$ - the medium extinction matrix - (path 1) and reinforced by the indirect radiation scattered by each particle in the propagation direction (path 2 + 3) (Fig. 1.17). At \vec{r} , the phase matrix $\bar{R}(\vec{r}, \hat{u}, \hat{u}')$ governs the deflection of energy beams from \hat{u}' to \hat{u} . The combined effect of the damping and reinforcing of \tilde{S} results in backscattering behavior which is both difficult to predict and dependent on several parameters: density, size, shape and scatterer arrangement, dielectric properties, surface mechanisms, etc.

Moreover, the radiative transfer does not allow for coherent summing of the contributions of different natures such as have been observed for a forest canopy (Fig. 1.18(a)). Radiative transfer modeling requires the assumption that the medium can be described as a stack of layers made of particles embedded in a host medium (Fig. 1.18(b)). We then limit ourselves to characterizing the average behavior of the wave within each layer, but the phase information is lost because of the averaging. This imperfection can be corrected by ‘coherent’ methods approximating an exact calculation of the interaction (for amplitude and phase). Currently, aside from a few promising



- | | |
|-----------------------------|---------------------------------|
| 1. Crown scattering | 4. Attenuated ground scattering |
| 2. Trunk scattering | 5. Direct ground scattering |
| 3. Trunk-ground interaction | 6. Trunk-branch interaction |

Fig. 1.18 (a) Interaction between radar incident beam and forest canopy: principal interaction mechanisms; (b) Modeling of forest canopy in accordance with the radiative transfer approximation.

tests for low frequencies ($f_c \approx 100$ MHz), for which the number of scatterers to be taken into account is limited, this approach requires too much computing time.

Stands of pinaster pines at different stages of development (i.e., with variable biomass rates) (Fig. 1.19), illustrate the complexity of backscattering including both surface and volume effects. In L band ($f_c = 1.25$ GHz), vertical polarization (Sect. 1.1.2) (Fig. 1.19(a)), the decreasing radar response of the sparsely populated forest (ranging from 0 to 33 t/h), reveals progressive damping of the ground's intrinsic response (-10 dBm²/m²); this is due to the increasing extinction of the vegetation. Beyond a critical volume of vegetation, multi-scattering mechanisms take over, augmenting the response from -12 dBm²/m² to a saturation on the order of -10 dBm²/m². The L band HV signal (emission in H , reception in V) (Fig. 1.19(c)) shows the very high sensitivity of the radar response to the biomass level. This sensitivity is enhanced in two respects: (1) HV is mainly generated by multi-scattering mechanisms, i.e. by vegetation. (2) The ground contribution, dominated by surface effects, produces practically no cross-polarization.

Finally, with C band VV polarization (Fig. 1.19(b)), the intrinsic contribution of the ground becomes highly preponderant (the roughness effects increase when the frequency increases with an intrinsic level for C band of about -3 dBm²/m²), until they wipe out the transition phase observed in L band: this is in stark contrast to the commonly accepted idea that radar response increases with increasing biomass!

1.3.4 Penetration properties of electromagnetic waves

To what extent are electromagnetic waves able to penetrate observed surfaces and glean information from the underlying structures? To reply to this question, let us return to

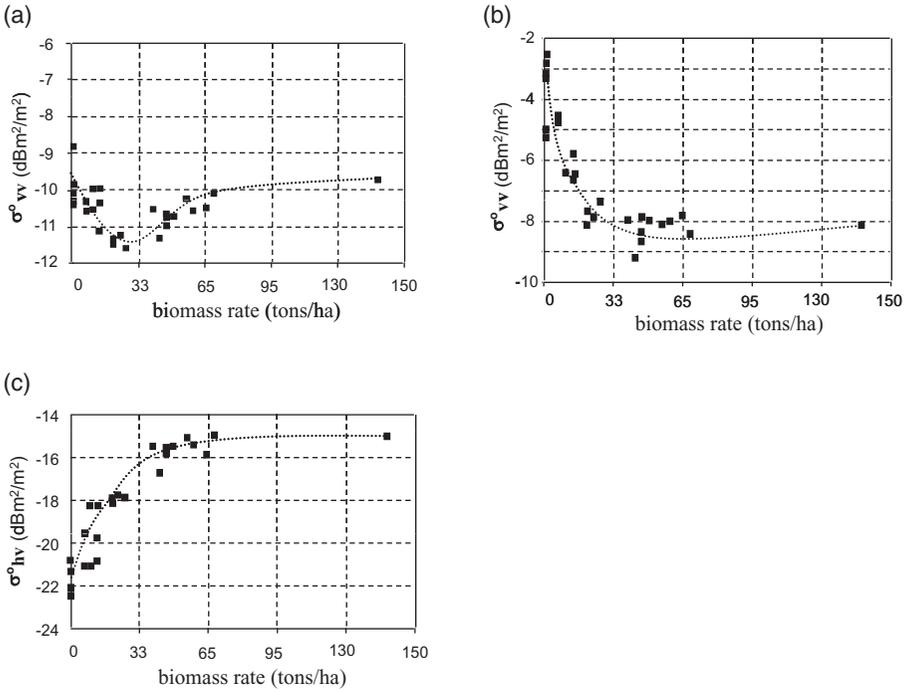


Fig. 1.19 Behavior of forest stands of various ages. SIR-C data, Landes forest, France (Source: CESBIO). Experimental data (■), interpolation (—); (a) *L* band, *VV* polarization, 26 °; (b) *C* band, *VV* polarization, 26 °; (c) *L* band, *HV* polarization, 26 °.

the equation for an electric field $\vec{e}(z, t)$ propagating in the \hat{z} direction of a homogeneous medium with a dielectric constant ϵ_r (Sect. 1.1.2.1):

$$\vec{e}(z, t) = \Re\{\vec{E}_0 \cdot \exp[j \cdot (\omega \cdot t - k \cdot z)]\} \tag{1.83}$$

Generally speaking, the dielectric constant ϵ_r is a complex quantity whose imaginary part expresses the attenuation of the wave during propagation. Extraction of its square root leads to a sign ambiguity, which is resolved by physical considerations. By letting the wave number $k = \alpha - j \cdot \beta$, we then get:

$$\vec{e}(z, t) = \exp[-\beta \cdot z] \cdot \Re\{\vec{E}_0 \cdot \exp[j \cdot (\omega \cdot t - \alpha \cdot z)]\} \tag{1.84}$$

Because of Eq. (1.84), β must be a real positive number (in the opposite case, the wave would amplify during propagation).

By definition, the penetration depth is the depth δp beyond which the amplitude of $\vec{e}(z, t)$ is attenuated by a factor larger than $1/\sqrt{e}$. From Eq. (1.84), it follows that:

$$\delta p = \frac{1}{2} \cdot \beta = -\frac{1}{2 \cdot k_0 \cdot \Im m(\sqrt{\epsilon_r})} \tag{1.85}$$

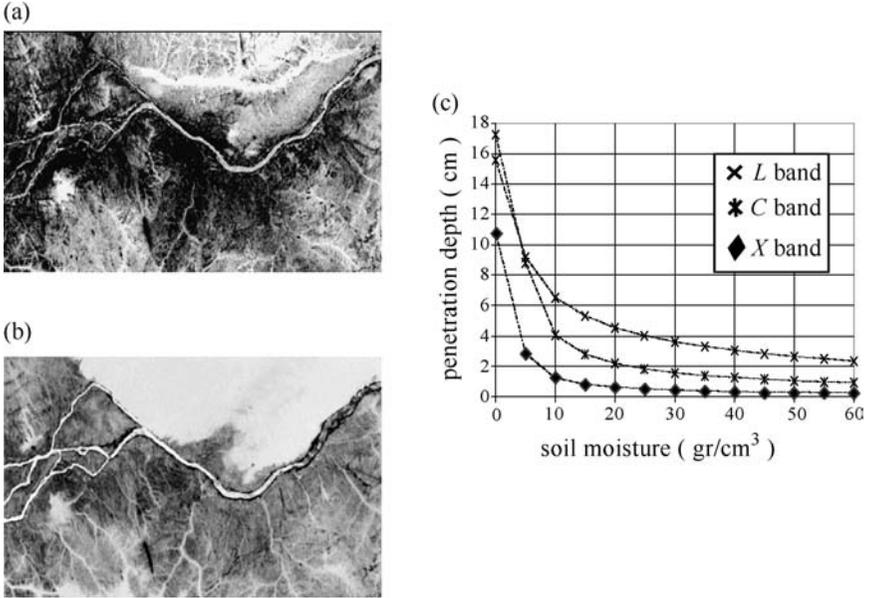


Fig. 1.20 (a) SIRC multi-frequency $\gamma(L + C)$ radar image, Nile Valley. (From: www.jpl.nasa.gov/radar/sircxsar.) ©Jet Propulsion Laboratory; (b) Infrared optical image of the same area. From : www.jpl.nasa.gov/radar/sircxsar. ©Jet Propulsion Laboratory; (c) Penetration depth for waves in bare soils as a function of their water content L band (×), C band (*), X band (◆).

In the case of slightly dispersive media¹⁷ ($|\Im m(\epsilon_r)| < 0.3 * \Re e(\epsilon_r)$), the penetration depth δp is approximated by:

$$\delta p \approx -\frac{\lambda \sqrt{\Re e(\epsilon_r)}}{2 \cdot \pi \cdot \Im m(\epsilon_r)} \quad (1.86)$$

Comments:

- 1) *The penetration depth increases with the wavelength:* low frequencies (L, P) penetrate better than higher frequencies (C, X), whence the interest of the scientific community for very low frequency SARs (from several tens to a few hundreds of MHz).
- 2) A total lack of dissipation $\Im m(\epsilon_r) = 0$ leads to an infinite penetration depth. Conversely, infinite dissipation ($|\Im m(\epsilon_r)| = \infty$) prohibits any penetration, as for a metal surface (or for a water surface with a high saline content). The effect of radar wave penetration has been used to observe the former bed of the Nile (Fig. 1.20(a)), buried under the sand, by means of *SIR-C* images acquired in C

¹⁷This is the case for most natural media except water.

and L bands, while the corresponding infrared optical image (Fig. 1.20(b)) does not reveal this underground trace.

Quantitatively, for bare soils, using the estimations of dielectric constants given by Hallikainen [1985], Eq. (1.86) gives the penetration depth δp : *at a given frequency, the penetration depth decreases with increasing water content.* X band hardly penetrates soils with more than 15% water content (Fig. 1.20(c)), whereas the L band can penetrate a few centimeters in ground with water content of up to 40%. In C band electromagnetic waves penetrate ground with average water content to a depth of about 5 cm.

For areas with dense vegetation cover and more specifically for forests, typical penetration values (*EOS SAR Instrument Panel Report Figures*, 1988) are 1 m for X band, 6 m for C band, and up to 20 m for L band.

1.3.5 The effects of slope on radiometry

Finally, the local slopes of the ground produce significant radiometric variations in radar imagery.¹⁸ Their impact may be observed on a ‘macroscopic’ scale in contrast to the scale of the effects previously mentioned (roughness, water content, etc.). We can consider the landscape as a succession of flat facets, each with a local orientation that affects the level of the signal response. In this view, a facet is a passive antenna which re-emits a signal whose power increases as the facet orientation approaches the perpendicular to the incident ray. The comparison of Fig. 21a (actual ERS image) and Fig. 21b (simulation of slope effects) shows how slope effects impact the global radiometry of an image.

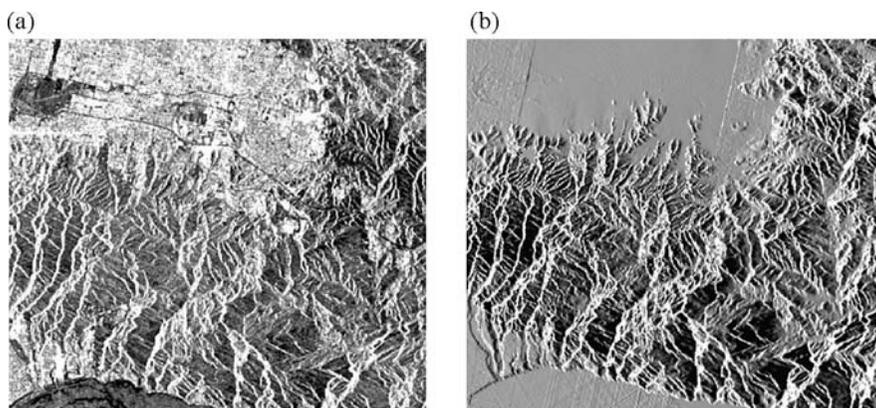


Fig. 1.21 (a) ERS image of a mountainous region; (b) Simulated ERS image.

¹⁸The slopes also affect the complex content of the radar signal in that they shift its frequency spectrum, thus deteriorating interferometric capability (see Sect. 4.5.1).

1.4 Elements of signal processing: the Fourier transform and the sampling theorem

1.4.1 Introduction

The Fourier transform is used very frequently in radar technology, from antenna radiation to SAR processing. The principle used is harmonic analysis in which any physical signal is broken down into either a discrete or a continuous set of sine waves with variable frequencies. We thus move the problem from a geometric domain into the more abstract frequency domain. In this section, we shall present the mathematical background of the Fourier transform (*FT*) and its expression for some standard functions and shall also review the principal features of real and complex signals sampling.

1.4.2 Fourier series of real periodic functions

Any real deterministic, square integrable signal $s(t)$, with period T , can be described by superposing a set of discrete sine wave functions. Indeed, all harmonic functions of the form $\exp(j \cdot 2 \cdot \pi \cdot k \cdot t/T)$, where k is a positive or negative integer, form an orthogonal basis of square integrable functions with period T . The resulting decomposition has the form:

$$s(t) = \sum_{k=-\infty}^{+\infty} X_k \cdot \exp(j \cdot 2 \cdot \pi \cdot k \cdot t/T) \quad (1.87)$$

with:

$$X_k = \frac{1}{T} \cdot \int_{\alpha}^{\alpha+T} s(t) \cdot \exp(-j \cdot 2 \cdot \pi \cdot k \cdot t/T) \cdot dt \quad (1.88)$$

The coefficient X_k represents the ‘strength’ of the harmonic component with frequency $f_k = k/T$ in the composition of $s(t)$, with α being any real number. As a result of Eq. (1.87), the frequency representation is a simple series of rays separated by $1/T$.

1.4.3 Fourier transform

The concept of Fourier series can be extended to non-periodic real functions. For this, we introduce a ‘gate’ function $s(t, T)$ such that:

$$s(t, T) = s(t) \cdot \pi(t/T) \quad (1.89)$$

where $\pi(t/T) = 1$ for $|t| \leq T/2$, and $\pi(t/T) = 0$ for $|t| > T/2$.

By taking the limit we obtain:

$$s(t) = \lim_{T \rightarrow \infty} s(t, T) \cdot \pi(t/T) \quad (1.90)$$

If we focus now on the $s(t, T)$ function in the interval $[-T/2; T/2]$, we may assume that it represents a pattern of a periodic function with period T , which can be decomposed into Fourier series. The spectrum of rays becomes denser as T increases, so that at the limit we end up with a continuous spectrum of rays. Thus the *non-periodic* signal $s(t)$ is described in the frequency domain by the superposition of an *infinity* of sine waves. At the limit, the Fourier transform $S(f)$ of $s(t)$ is given by:

$$S(f) = \int_{-\infty}^{+\infty} s(t) \cdot \exp(-j \cdot 2 \cdot \pi \cdot f \cdot t) \cdot dt \quad (1.91)$$

and in an equivalent way :

$$s(t) = \int_{-\infty}^{+\infty} S(f) \cdot \exp(j \cdot 2 \cdot \pi \cdot f \cdot t) \cdot df \quad (1.92)$$

Equations (1.91) and (1.92) are the continuous forms of Eqs. (1.87) and (1.88). The Fourier transform $S(f)$ of a real signal is a complex signal. Its modulus and its phase are referred to as the *modulus spectrum* and *phase spectrum* of the signal, which are, respectively, even and odd functions for real signals.

1.4.4 Properties of the Fourier transform (FT)

This is a review of the most widely used properties of the *FT* in the field of radar signal processing. Let us consider two functions $s(t)$ and $u(t)$, with the *FTs* $S(f)$ and $U(f)$:

- The *FT* is a linear operator:

$$\forall (\alpha, \beta) \in \mathbb{C}^2, \alpha \cdot s(t) + \beta \cdot u(t) \rightarrow \alpha \cdot S(f) + \beta \cdot U(f) \quad (1.93)$$

- A multiplication product in the time domain becomes a convolution product in the frequency domain and reciprocally:

$$s(t) \cdot u(t) \rightarrow S(f) * U(f) \quad (1.94)$$

$$s(t) * u(t) \rightarrow S(f) \cdot U(f) \quad (1.95)$$

$$\text{with: } s(t) * u(t) = \int_{-\infty}^{+\infty} s(t') \cdot u(t - t') \cdot dt'$$

- A translation in the time domain becomes a multiplication by a phase ramp in the frequency domain and reciprocally:

$$s(t - a) \rightarrow \exp(-j \cdot 2 \cdot \pi \cdot f \cdot a) \cdot S(f) \quad (1.96)$$

$$\exp(j \cdot 2 \cdot \pi \cdot f_d \cdot t) \cdot s(t) \rightarrow S(f - f_d) \quad (1.97)$$

Property (1.97) is of fundamental importance in SAR processing ([Chap.2](#)) because its physical principle requires that radar echoes be phase locked in the time domain (it is achieved by multiplication by phase ramps). In practice, implementation on computers is based on translations in the frequency domain, using the equivalence (1.97).

1.4.5 Fourier transforms of standard functions

1.4.5.1 The gate function

The envelope of a radar pulse, for example, can take the form of a gate function (Fig. 1.22(a)):

$$\begin{aligned} s(t) &= A \cdot \Pi(t/T) \rightarrow s(f) \\ &= A \cdot T \cdot \frac{\sin(\pi \cdot f \cdot T)}{\pi \cdot f \cdot T} = A \cdot T \cdot \text{sinc}(\pi \cdot f \cdot T) \end{aligned} \quad (1.98)$$

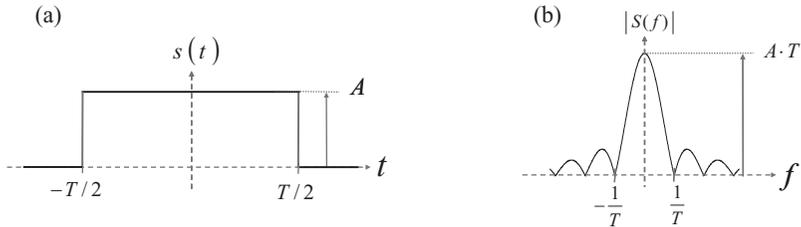


Fig. 1.22 (a) Gate function; (b) Modulus of the FT of the gate function.

The Fourier transform of the gate function is a cardinal sine (Fig. 1.22(b)), whose main lobe has a width of $2/T$. In practice it is assumed that the unbounded frequency representation barely exceeds the interval of the main lobe of the FT , thus limiting the frequency description to a ‘bandwidth’ of $2/T$ (the 3 dB bandwidth is of the order of $1/T$).

Please note: The FT shows that some power density is transmitted at negative frequencies. This mathematical curiosity is the result of decomposing the function in the basis of complex functions with the form $\exp(j \cdot 2 \cdot \pi \cdot k \cdot t/T)$. From the even and odd properties of the modulus and phase spectra, respectively, of $S(f)$ it is possible to reconstruct a real signal with the help of this negative frequency power density. It is therefore only a mathematical trick with no real significance in physical terms.

1.4.5.2 The Dirac delta function

This is an ideal impulse, whose value is zero everywhere except at 0 (Fig. 1.23(a)). The Dirac delta function $\delta(t)$ is only strictly defined in terms of distributions, i.e. using an integral operator:

$$s(t_0) = \int_{-\infty}^{+\infty} s(t) \cdot \delta(t - t_0) \cdot dt \quad \text{for any real } t_0. \quad (1.99)$$

$\delta(t)$ can also be considered as the limiting case of a gate function of unit surface and whose length tends towards 0:

$$\delta(t) = \lim_{T \rightarrow 0} \frac{1}{T} \cdot \Pi(t/T) \quad (1.100)$$

Consequently:

$$S(f) = \lim_{T \rightarrow 0} \frac{1}{T} \cdot T \cdot \frac{\sin(\pi \cdot f \cdot T)}{\pi \cdot f \cdot T} = 1 \quad (1.101)$$

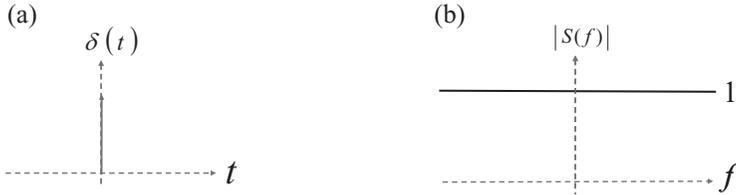


Fig. 1.23 (a) Dirac delta function; (b) Modulus of the *FT* of the Dirac delta function.

The frequency spectrum of the Dirac delta (Fig. 1.23(b)) is spread uniformly across the entire range of frequencies: an infinitely precise location of the signal in the time domain produces complete delocalisation in frequency domain.

Please note: Likewise the *FT* of a function which is constant and equal to 1 in the time domain is $\delta(f)$. Therefore, accounting for the effect of multiplication by a phase ramp on the *FT* operator, the *FT* of the harmonic function $\exp(j \cdot 2 \cdot \pi \cdot f_d \cdot t)$ is $\delta(f - f_d)$, where f_d is a constant frequency.

1.4.5.3 The Dirac comb

This is a periodic function $\delta_T(t)$ composed of an infinite number of Dirac functions separated by a sampling interval T . It is of fundamental importance in sampling theory (Fig. 1.24(a)):

$$\delta_T(t) = \sum_{-\infty}^{+\infty} \delta(t - k \cdot T) \quad (1.102)$$

Since it is a periodic function, the calculation of the *FT* of $\delta_T(t)$ is based on the Fourier series. As a result of Eq. (1.88):

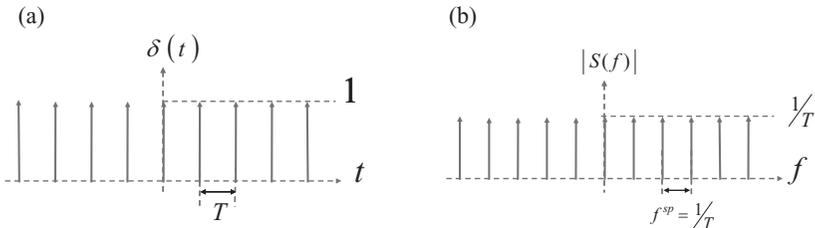


Fig. 1.24 (a) Dirac comb; (b) Modulus of the *FT* of the Dirac comb.

$$X_k = \frac{1}{T} \cdot \int_{-T/2}^{T/2} \delta t \cdot \exp(-j \cdot 2 \cdot \pi \cdot k \cdot t/T) \cdot dt = \frac{1}{T} \quad (1.103)$$

Substituting into Eq. (1.87) leads to:

$$\delta_T(t) = \frac{1}{T} \cdot \sum_{k=-\infty}^{+\infty} \cdot \exp(j \cdot 2 \cdot \pi \cdot k \cdot t/T) \quad (1.104)$$

Consequently, considering Eq. (1.97):

$$TF\{\delta_T(t)\} = \frac{1}{T} \cdot \sum_{k=-\infty}^{+\infty} \delta(f - k/T) = \frac{1}{T} \cdot \delta_{1/T}(f) \quad (1.105)$$

The Fourier transform of a Dirac comb with a sampling interval T and a modulus of 1 is a Dirac comb with an interval and a modulus of $1/T$ (Fig. 1.24(b)).

1.4.5.4 Monochromatic functions $\cos(2 \cdot \pi \cdot f_0 \cdot t)$ and $\sin(2 \cdot \pi \cdot f_0 \cdot t)$

Although these functions are not square integrable, a Fourier transform is nonetheless possible. With the use of ‘distribution theory’ it can be shown that (Fig. 1.25)

$$FT\{\cos(2 \cdot \pi \cdot f_0 \cdot t)\} = \pi \cdot \{\delta(f - f_0) + \delta(f + f_0)\} \quad (1.106)$$

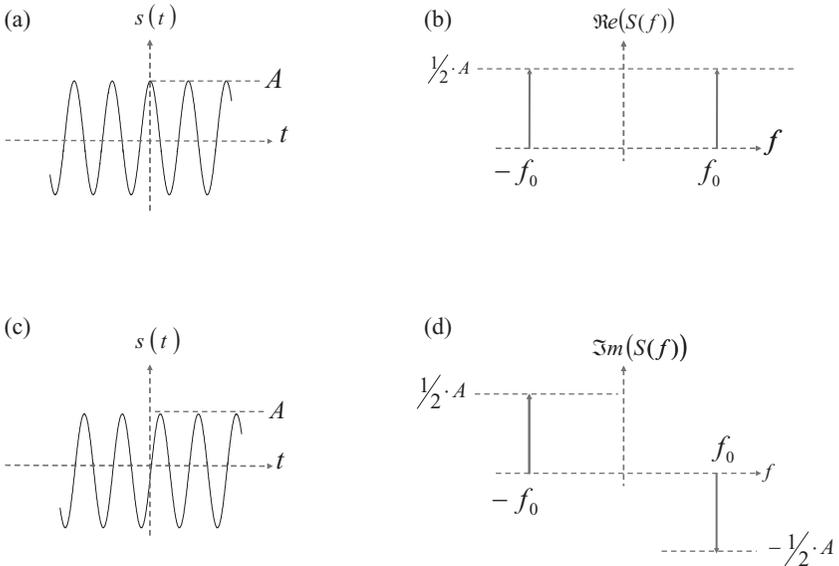


Fig. 1.25 (a) $f(t) = \cos(2 \cdot \pi \cdot f_0 \cdot t)$; (b) Real part of the FT of $(2 \cdot \pi \cdot f_0 \cdot t)$; (c) $f(t) = \sin(2 \cdot \pi \cdot f_0 \cdot t)$; (d) Imaginary part of the FT of $f(t) = \sin(2 \cdot \pi \cdot f_0 \cdot t)$.

and:

$$FT\{\sin(2 \cdot \pi \cdot f_0 \cdot t)\} = j \cdot \pi \cdot \{\delta(f + f_0) - \delta(f - f_0)\} \quad (1.107)$$

The frequency spectrum of a monochromatic wave is composed of two rays: the infinitely precise location of a signal in the frequency domain produces complete delocalisation in the time domain (a sine wave). The particular case of a function constantly equal to 1, with $FT \delta(f)$, expresses this property when f_0 equals zero (constant signal).

1.4.5.5 The auto-correlation function of a stationary signal

Consider a stationary physical signal $s(t)$. The fact that it is stationary means that its statistics are independent of time. Its auto-correlation function $K(\tau)$ is therefore defined by:

$$K(\tau) = \int_{-\infty}^{+\infty} S(t + \tau) \cdot s^*(t) \cdot dt \quad (1.108)$$

The Wiener-Khintchine theorem states that the spectral density $\Phi(f)$ of the signal $s(t)$, i.e. the power contained in the part of the spectrum between frequencies f and $f + df$, is the FT of $K(\tau)$:

$$\Phi(f) = \int_{-\infty}^{+\infty} K(\tau) \cdot \exp(-j \cdot 2 \cdot \pi \cdot f \cdot \tau) \cdot d\tau \quad (1.109)$$

It can also be shown that the modulus $|\Phi(f)|$ is equal to $|S(f)|^2$ for any frequency f . The Wiener-Khintchine theorem is used when estimating the mean Doppler frequency (also called Doppler centroid) of a radar signal from its auto-correlation function (Sect. 2.3.4).

1.4.6 Sampling real signals

The real physical analogue signal $s(t)$, is defined by a continuous set of values. We assume that the $FT S(f)$ of $s(t)$ (Fig. 26(a)) is band limited (Fig. 26(b)). To digitize $s(t)$, only discrete set of samples, separated by a constant interval T^{sp} is needed (Fig. 26(c)). For ideal sampling (i.e. the extraction of an instantaneous value of the signal during an infinitesimally short time), the resulting signal is:

$$s^{sp}(t) = \sum_{k=-\infty}^{+\infty} s(k \cdot T^{sp}) \cdot \delta(t - k \cdot T^{sp}) = s(t) \cdot \delta_{T^{sp}}(t) \quad (1.110)$$

If we consider Eqs. (1.94) and (1.105), we can see that:

$$FT\{S^{sp}(t)\} = S(f) * f^{sp} \cdot \delta_{f^{sp}}(f) = f^{sp} \cdot \sum_{k=-\infty}^{+\infty} S(f - f^{sp}) \quad (1.111)$$

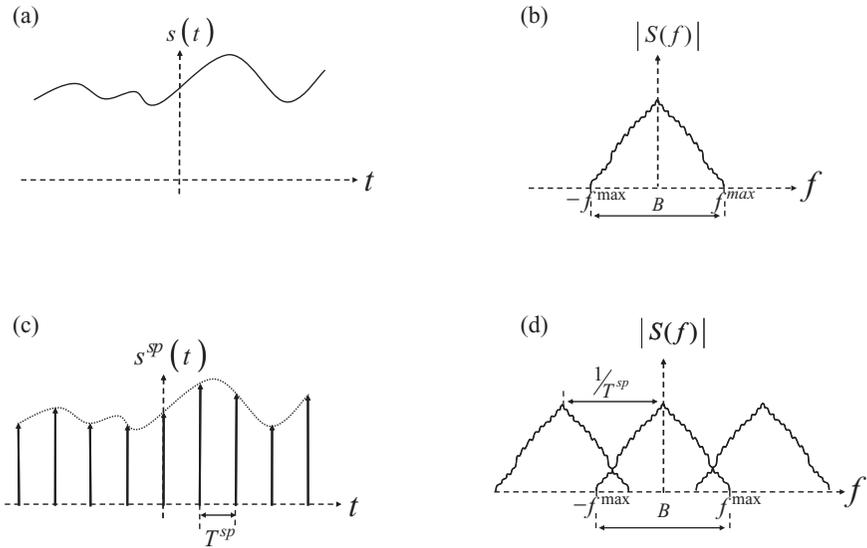


Fig. 1.26 (a) Continuous function $s(t)$; (b) Modulus of the FT of $s(t)$; (c) Sampling of $s(t)$; (d) Modulus of the FT of $s^{sp}(t)$.

with: $f^{sp} = 1/T^{sp}$. Using Eq.(1.104), we see that sampling in the time domain produces a periodic spectrum $S(f)$ (with period f^{sp}), and reciprocally (Fig. 1.26(d)). This is the underlying principle of the sampling theorem (also known as Shannon's theorem), which helps us avoiding aliasing, as described in the next section.

1.4.7 Sampling theorem (Shannon's theorem)

The sampling theorem defines the conditions under which it is possible to correctly reconstruct a continuous real signal from only a set of its sampled values. For a band limited signal (i.e. there is a frequency f^{\max} such that $S(f) = 0$ for $|f| > f^{\max}$), it is possible to isolate the frequency band by low-pass filtering of the signal as it existed before sampling, as long as its successive periodic patterns do not overlap. Under these conditions, low-pass filtering followed by an inverse FT allows the signal to be completely rebuilt, while avoiding 'frequency aliasing'. This is only possible (Fig. 1.26(d)) if the sampling frequency f^{sp} exceeds the signal bandwidth $B = 2 \cdot f^{\max}$ (applies only to real signals). This result is known as Shannon's theorem: a real signal can be reconstructed from its sampled values if f^{sp} is at least twice the maximum frequency f^{\max} of the frequency spectrum. In the specific case of a monochromatic signal of frequency f_0 , whose frequency spectrum is composed of two rays (one located at f_0 , the other at $-f_0$); f^{sp} must exceed $2 \cdot f_0$.

Please note: Unfortunately, the Paley-Wiener theorem states that the frequency spectrum of any physical signal is necessarily unbounded, with its envelope tending

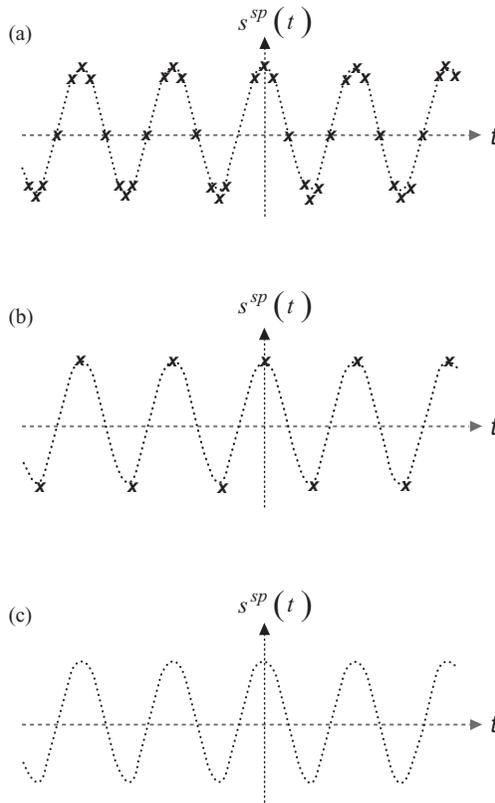


Fig. 1.27 (a) Sampled sine wave at $f^{sp} \gg 2 \cdot f_0$; (b) Sampled sine wave at $f^{sp} = 2 \cdot f_0$; (c) Sampled sine wave at $f^{sp} < 2 \cdot f_0$.

towards zero for high frequencies. However fine the sampling, it will therefore not be possible, in practice, to completely avoid frequency aliasing.

1.4.7.1 A ‘naïve’ interpretation of Shannon’s theorem

Let us observe the effect of the sampling frequency f^{sp} on the reconstruction of a pure sine wave of frequency f_0 . If sampling at f^{sp} provides knowledge of a sufficient number of points per period (Fig. 1.27(a)) then it is possible to reconstruct the signal intuitively: $f^{sp} \gg 2 \cdot f_0$. For $f^{sp} = 2 \cdot f_0$ (Fig. 1.27(b)), the reconstruction of the sine wave is still possible, though here we are stretching Shannon’s theorem to the limits. It is nonetheless necessary for the sine wave to be sampled at appropriate positions to allow correct restitution in amplitude, phase and frequency (any offset of the sampling relative to the maxima and minima will lead to a false estimation of amplitude and phase). Lastly, if $f^{sp} < 2 \cdot f_0$ (Fig. 1.27(c)), Shannon’s theorem no longer applies and the footprint of the initial sine wave is definitively lost.

1.4.7.2 Sampling of complex signals

As for real signals, avoiding frequency aliasing requires that $f^{sp} > B$. Nonetheless, Shannon's theorem must be interpreted appropriately for complex signals. Let us look at the behavior of a monochromatic signal $\tilde{s}(t) = \exp(j \cdot 2 \cdot \pi \cdot f_0 \cdot t)$. Unlike real monochromatic signals, with a frequency bandwidth $B = 2 \cdot f_0$ (and consisting of positive and negative frequencies) the frequency bandwidth of $\tilde{s}(t)$ is now in the positive frequency interval $[0; f_0]$, with width f_0 . Does this mean that the sampling theorem is only half as strict for the complex $\tilde{s}(t)$ as for the real signals?

The way to resolve this apparent paradox is to visualize the vector $\tilde{s}(t)$ by using a black spot located on the rim of a wheel with unit radius; the wheel is then lit with stroboscopic lighting (Fig. 1.28(a)). The fact that all the frequencies of $\tilde{s}(t)$ have the same sign indicates that the direction of wheel rotation is known. When f^{sp} varies (Fig. 1.28b), it is easy to determine the angular velocity of the wheel (since we know its direction of rotation) as long as $f^{sp} > f_0$. When $f^{sp} = 4/3 \cdot f_0$, if we do not know the direction of rotation, the impression given is that of a clockwise motion of frequency $f_0/4$. If the direction of rotation is known, then the observation reveals the exact angular velocity. This is rather like the case of the wagon wheels seen on film which seem to be turning backwards, but for which the context clearly indicates the actual rotational direction. For $f^{sp} = f_0$, phasing between the illumination and the wheel motion is perfect; this is the limit of Shannon's theorem. Finally, for $f^{sp} = 4/5 \cdot f_0$, the apparent direction of the wheel's rotation is the actual direction, but with an apparent frequency of $f_0/4$: the frequency information is completely lost and the spectrum of $\tilde{s}(t)$ is aliased.

1.4.8 The Fast Fourier transform (FFT) algorithm

The computation of the *FT*, using sampled signals, is called the Discrete Fourier Transform (DFT). It is usually implemented with the help of the Fast Fourier Transform (FFT) algorithm. Because of its computational efficiency, most of the synthetic aperture radar algorithms are applied in the frequency domain. Here, we describe the principle behind the FFT and determine the amount of calculation needed.

We first calculate the *DFT* of a function $s(t)$ determined by a regular sampling of N points, separated by the time sampling interval T^{sp} . In this way we describe a periodic pattern of width $T = N \cdot T^{sp}$, the $(N + 1)^{th}$ value being equal to the first. The function $s(t)$ is assumed to be periodic outside this interval:

$$\begin{aligned} s_k &= s(t_k), \quad \text{with: } t_k = k \cdot T^{sp} \\ \text{and: } k &= 0, 1, 2, \dots, N - 1 \quad \forall t, s(t + T) = s(t) \end{aligned} \quad (1.112)$$

This description in the time domain results in a frequency spectrum composed of a series of rays separated by a frequency interval $1/(N \cdot T^{sp})$:

$$S_n = S(f_n) \quad (1.113)$$

with: $f_n = n/(N \cdot T^{sp})$, and $n = -N/2, \dots, N/2$. Given Eqs. (1.112) and (1.113), the *DFT* acts as a linear operator, transforming a vector of N values (the s_k) into another

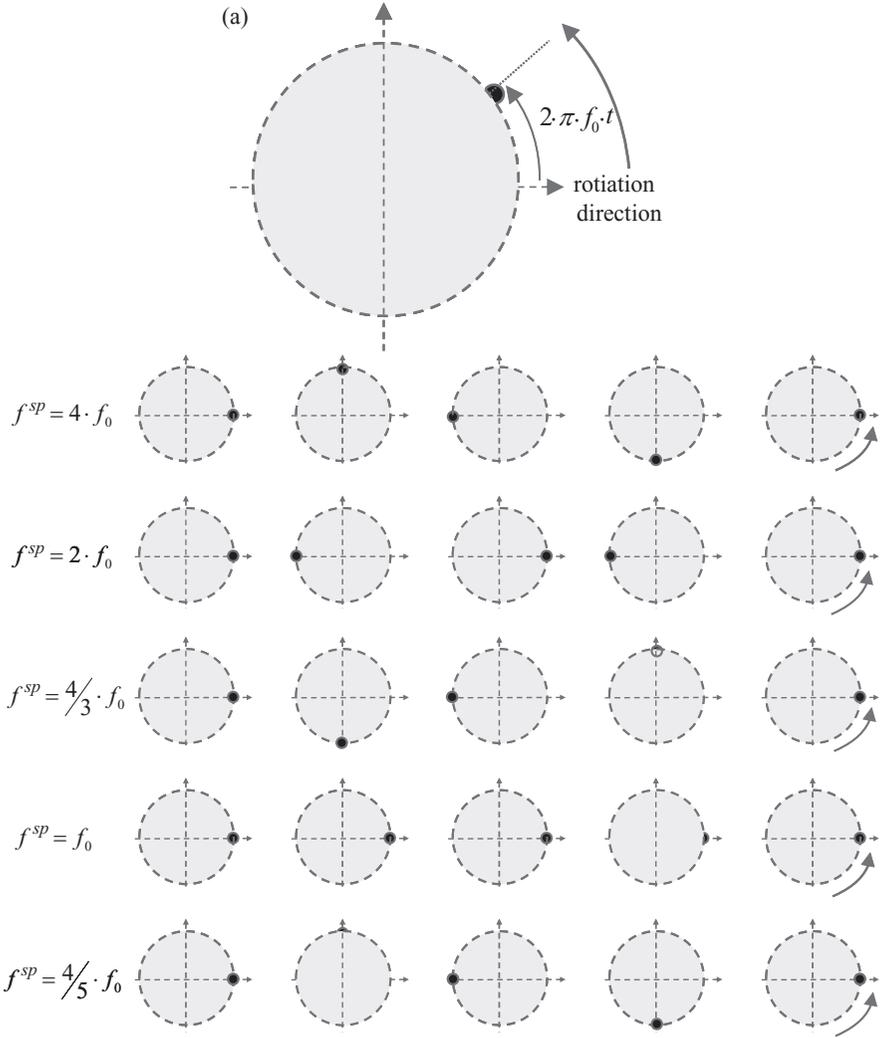


Fig. 1.28 (a) Location of a black spot on a wheel with a unit radius whose direction of rotation is known; (b) Effect of sampling on a wheel with a unit radius whose direction of rotation is known.

vector of N values (the S_n), where the terms $S_{-N/2}$ and $S_{N/2}$ are equal (period of the spectrum).

The *DFT* is expressed as a result of Eqs. (1.91), (1.112) and (1.113):

$$S_n = \int_{-\infty}^{+\infty} s(t) \cdot \exp(-j \cdot 2 \cdot \pi \cdot f_n \cdot t) \cdot dt \approx \sum_{k=0}^{n-1} s_k \cdot \exp(-j \cdot 2 \cdot \pi \cdot f_n \cdot t_k) \cdot T^{SP} \tag{1.114}$$

or:

$$s_n = T^{sp} \cdot \sum_{k=0}^{N-1} W^{nk} \cdot S_k \quad (1.115)$$

where

$$W = \exp(-j \cdot 2 \cdot \pi / N) \quad (1.116)$$

It can be seen from Eq.(1.115) that the matrix transformation from the (s_k) to the (S_n) requires approximately N^2 operations. The Fast Fourier transform (*FFT*) performs this transformation with fewer operations, based on the work of Danielson and Lanczos (1942) which was popularized by Cooley and Tukey (in the 1960s).¹⁹ Practically, if N is an even number, we can rewrite S_n as:

$$\begin{aligned} S_n = T^{sp} \cdot \sum_{k=0}^{N/2-1} s_{2k} \cdot \exp(-j \cdot 2 \cdot \pi \cdot n \cdot k / (N/2)) \\ + T^{sp} \cdot W^n \cdot \sum_{k=0}^{N/2-1} s_{2k+1} \cdot \exp(-j \cdot 2 \cdot \pi \cdot n \cdot k / (N/2)) = S_n^e + W^n \cdot S_n^o \end{aligned} \quad (1.117)$$

A *DFT* applied to N points is therefore the sum of two *DFTs*, each applied to $N/2$ points. This operation can be iterated on the even and odd parts S_n^e and S_n^o , for as long as N can be divided by 2. The extreme limit of decomposition is reached when N is a power of 2 (if this is not the case, it is a simple matter to correct). Determining S_n [Press, 1988] requires calculations on the order of $\log_2 N$ for a *DFT* with a length of 1 (i.e. using identity operations), leading to an overall number of calculations proportional to $N \cdot \log_2 N$ (with the operation being repeated for the entire set of S_n). In practical terms, passing from order N^2 to order $N \log_2 N$ saves considerable computing time. For an *FT* of a sample of 1000 points (a typical value in SAR processing), this results in a reduction of about a factor of 100.

1.4.9 The two-dimensional Fourier transform

An image is by nature a two-dimensional mathematical entity, even though it is often processed using one-dimensional algorithms. The two-dimensional extension of the *FT* is as follows:

$$S(f_1, f_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} s(t_1, t_2) \cdot \exp[-j \cdot 2 \cdot \pi \cdot (f_1 \cdot t_1 + f_2 \cdot t_2)] \cdot dt_1 \cdot dt_2 \quad (1.118)$$

¹⁹The FFT algorithm has been 'rediscovered' several times over the years. The work of Gauss included some very similar expressions as early as 1805.

or

$$S(f_1, f_2) = \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} s(t_1, t_2) \cdot \exp[-j \cdot 2 \cdot \pi \cdot f_1 \cdot t_1] \cdot dt_1 \right\} \cdot \exp[-j \cdot 2 \cdot \pi \cdot f_2 \cdot t_2] \cdot dt_2 \quad (1.119)$$

This reduces to the calculation of two one-dimensional *FTs*, the first of which is applied to the variable t_1 and the second to the variable t_2 .

References

- [1] M. Born, E. Wolf, *Principles of Optics*, Fourth edition, Pergamon Press, 1970.
- [2] S. Chandrasekhar, *Radiative Transfer*, New York. Dover, 1960.
- [3] S. Cloude, ‘Group theory and polarisation algebra’, *Optik*, 75, 1, 26–36, (1986).
- [4] P. F. Combes, *Micro-ondes: circuits passifs, propagation, antennes*, Dunod, Paris (1997).
- [5] A. K. Fung, Z. Li, and K. S. Chen, ‘Backscattering from randomly rough dielectric surfaces’, *IEEE Transactions on Geoscience and Remote Sensing*, 30(2) 256–369, (1992).
- [6] M. T. Hallikainen, F. T. Ulaby, M. C. Dobson, M. A. El-Rayes and L. K. Wu, ‘Microwave dielectric behaviour of wet soil. Part I: Empirical models and experimental observations’, *IEEE Transactions on Geoscience and Remote Sensing*, GE-23(1), (1985).
- [7] J. A. Kong, *Electromagnetic Wave Theory*, Second Edition, John Wiley & Sons, Inc. (1990).
- [8] T. Le Toan, P. Smacchia, J. C. Souyris, et al. ‘On the retrieval of soil moisture from ERS-1 SAR data’, in *Proceedings of the second ERS-1 symposium: space at the service of our environment*, ESA SP-361, 883–888, Hamburg, (1993).
- [9] R. J. Mailloux, *Phase array antenna handbook*, Artech House, Boston, London, (1994).
- [10] Morchin, W. C., *Airborne Early Warning Radar*, 75–76. Artech House, (1990).
- [11] W. H. Press, B. P. Flannery, S. A. Teuklosky, and W. T. Vetterling, *Numerical Recipes in C*, Chapter 12 (Fast Fourier Transform), Cambridge University Press (1988).
- [12] J. R. Pierce, *Almost everything about waves*, MIT Press, Cambridge, MA, 130–131. (1974).

- [13] F. T. Ulaby, R. K. Moore, A. K. Fung, *Microwave remote sensing active and passive. Volume II*, Chapter 12 (Introduction to random surface scattering and emission), Artech House, (1986).

CHAPTER 2

SAR PROCESSING: AT THE HEART OF THE SAR TECHNIQUE

2.1 Introduction

In this chapter we will explore the basic and most significant features of radar imagery. In a first part, we will introduce the general principles of synthetic aperture radar (SAR), highlighting some specific effects strongly related to the physics and geometry involved. In a second part, we will develop the key role of the frequency representation with its manifold advantages: strong physical signification, access to more efficient computing schemes, specific resampling procedures. In this part we will also show that in some cases, working in the time domain remains preferable. In the third part we will review the SAR synthesis algorithms in order to give their specific flavors rather than going in full details: what is important is the understanding of the main architectural choice attached to each design, together with its advantages and drawbacks. The fourth part is devoted to the system constraints, which is the way to accommodate the many geometric and time constraints in order to optimize such or such features of the radar images. The fifth part will focus on the geometric properties of radar imagery and the equations which govern the positioning of radar pixels on the ground. Some methods for topography reconstruction are also briefly evoked in this part. Finally, we address the specific processing and co-registration problems raised by bistatic systems, and we propose a geometric modeling for bistatic observations.

In this Chapter, we pay very little attention to the source of our data: the radar instrument. This is because, assuming its output is digital (now a very general situation), very few parameters allow a full characterization of the instrument, for example specific frequencies described in the first part of this Chapter or basic power budget figure. The only option left as far as the nature of the data is concerned (disregarding the number of bits per pixel or the possible use of compression/decompression schemes) is whether the data are sampled as complex numbers or real numbers at doubled sampling rate.

Nevertheless, [Figure 2.1](#) presents the typical block diagram of a radar instrument (here with a complex sampling). The “pure” carrier frequency f_c is provided by the

local oscillator and mixed with the “chirp” signal (Sect. 2.3.7), which gives a certain frequency span to the signal, inversely proportional to the desired range resolution. The signal is then amplified and transmitted. The returned signal follows another path. The main task of the circulator is to avoid leaks from the high energy output to the delicate receiving electronics, designed for very weak signals only. After proper amplification by the low noise amplifier, the signal is split into two parts mixed with replicas of the carrier frequency, but with one of them shifted by $\pi/2$. The two paths are then sampled into the real and imaginary parts (also called “in phase” and “in quadrature”) of the complex radar signal using Analog to Digital Converters (ADC).

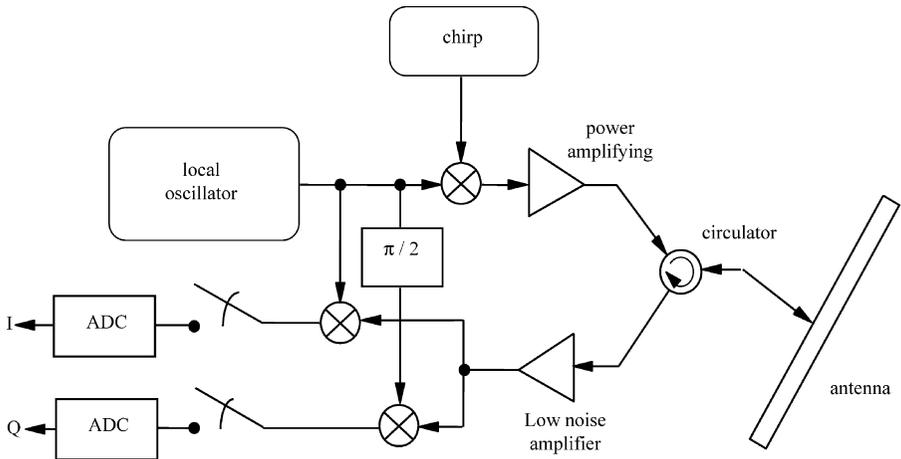


Fig. 2.1 Typical block diagram of a radar instrument (shown here with complex sampling).

2.2 General principles of Synthetic Aperture Radar

2.2.1 A different way of observing the Earth

When we try to observe the Earth using radio waves we immediately come up against an obstacle inherent in the laws of physics. The resolving power of an instrument with diameter D at a wavelength λ is given by the ratio λ/D (Sect. 1.2.5.1), i.e., two objects separated by a distance X and observed from a range R_0 will only be seen by the instrument as two distinct objects if the ratio X/R_0 exceeds λ/D .

The pupil of the human eye with its 3-mm aperture and collecting wavelengths near $0.5 \mu\text{m}$, has a theoretical resolving power of $1/6000$, with which it could distinguish two objects 1 m apart at a range of 6 km. The radar antennas that we can deploy in Space may be 10 m across, or even 15 m at their largest, but they use radiation in the centimeter range with wavelengths 100,000 times longer than that of light. A typical radar satellite, such as ERS-1 with its 10 m antenna and operating wavelength of approximately 5 cm, orbiting some 1000 km from the region that it is observing, is therefore not capable of resolving two points less than 5 km apart on the ground. The

natural resolution of such an image (Fig. 2.2(a)) would therefore be 5 km. To overcome this difficulty and create sufficiently detailed images (meaning with a resolution of a few meters, Fig. 2.2(c)), we apply techniques specific to each direction in the image which restore adequate resolution.

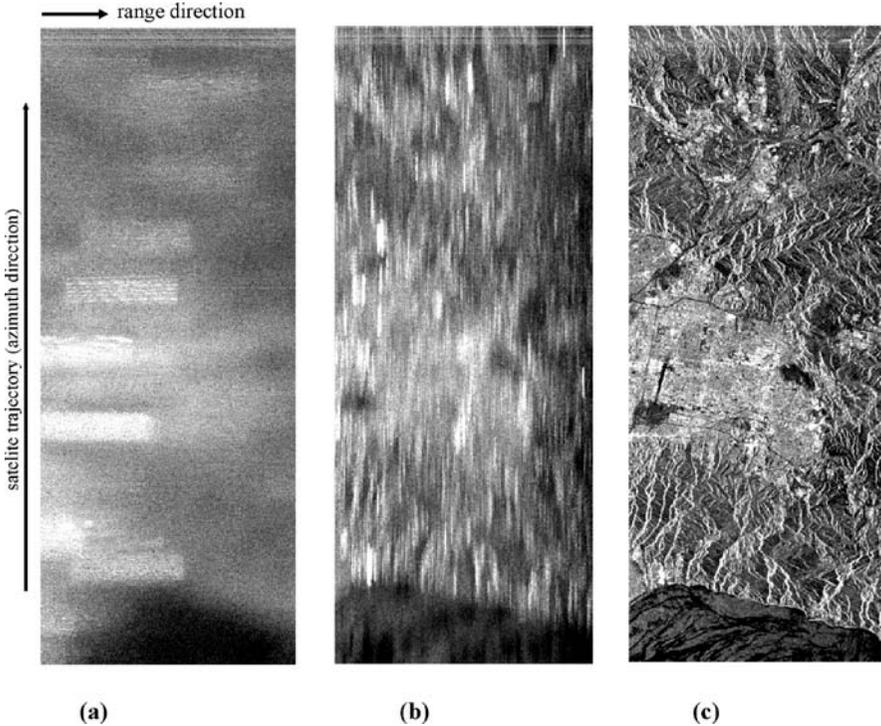


Fig. 2.2 (a) ERS image of raw echoes before SAR processing. Azimuth and range resolution: 5 km; (b) The same data after range processing. (Azimuth resolution: 5 km, range resolution: 20 m); (c) The same data after range processing and SAR synthesis. Azimuth resolution is: 5 m; range resolution is: 20 m. Northridge area, California, 1992.

2.2.2 Range vision

The first technique, which we will call ‘range vision’, takes advantage of the fact that the instrument is active—that is, it sends pulses—so that we can measure the round-trip time between the radar and the region observed. We can therefore distinguish between two points whose radar ranges differ by a few meters as long as the pulse duration is sufficiently short and the sampling rate of the return echo is sufficiently high.

Since short powerful pulses can damage the radar, a pulse compression technique is used to create short pulses from long coded pulses. Doing this also improves the resolution in the direction perpendicular to the trajectory of the radar, also known as ‘the range direction’ (Fig. 2.2(b)). We shall deal with the way in which the pulses are coded in Sect. 2.3.7 and Sect. 5.1.1. Using range to distinguish between objects

places geometric constraints on image acquisition, however, because two objects at an identical range are indistinguishable. This is why these radars cannot look directly down (at the nadir), as points located near the vertical are almost all at the same range from the radar. The instrument must therefore be side-looking so that the ground distance of a point from the nadir of the radar can be sorted as a function of its range from the radar. Ambiguous points located at the same ground distance from the vertical but in the opposite direction are not illuminated by the antenna and thus do not contribute to the return echo. With particularly steep terrain, a point with a greater ground distance from the vertical than another point lower down the slope may actually be closer to the radar itself, causing a layover effect (Sect. 6.1). The principal disadvantages of range vision, particularly acute in difficult terrain, are shown in Fig. 2.3 and are analyzed in detail in Sect. 6.1. Range vision is the basic principle behind all radar systems, even those that do not create images.

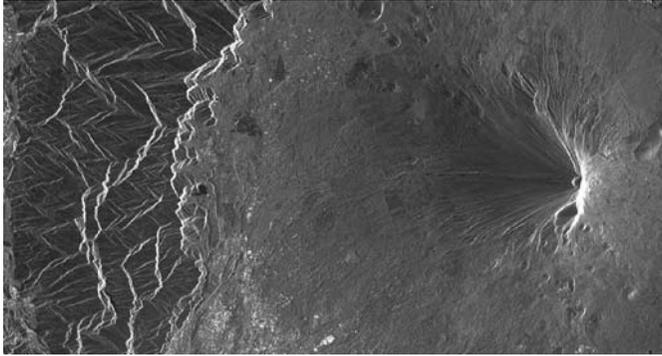


Fig. 2.3 Mount Fuji observed by COSMO-Skymed. COSMO-skymed image (Copyright 2007 ASI, Italian Space Agency, all rights reserved).

2.2.3 The three fundamental radar frequencies

A radar observation system is basically defined by the following three frequencies, described here from the highest to the lowest:

1. The radar's carrier frequency f_c is the frequency of the instrument's oscillator. This frequency defines the radar's wavelength λ by: $\lambda = c/f_c$, where c is the speed of light.
2. The sampling frequency in range f_d defines the size of the range pixel p_d by $p_d = c/(2 f_d)$ where the factor 2 takes account of the fact that the difference between two pixels includes the round trip. Naturally, to obtain the size of the pixel in ground range (projection, Fig. 2.4(a)) p_g , it must then also be divided by the local incidence angle of the wave θ :

$$p_g = \frac{c}{2 f_d \sin \theta} = \frac{p_d}{\sin \theta} \quad (2.1)$$

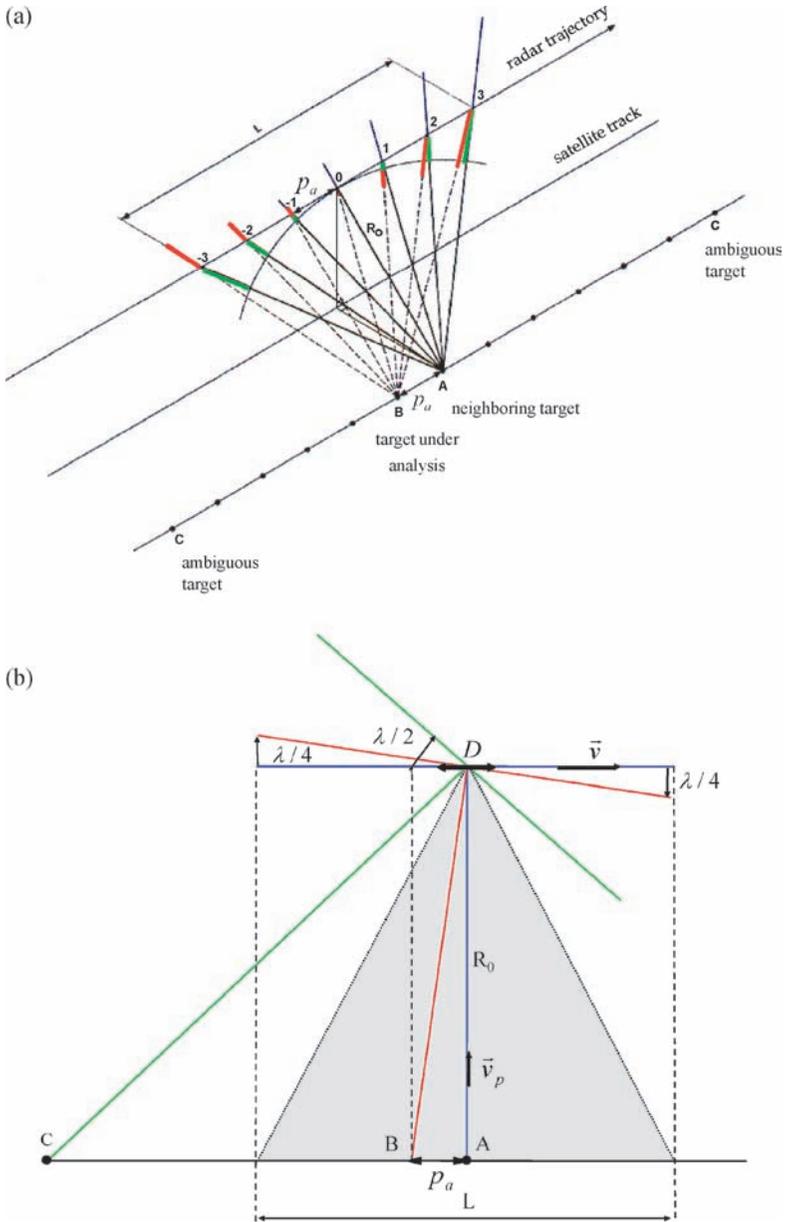


Fig. 2.5 Synthetic aperture (a) geometric approach; (b) geometric approach, 2D representation.

In the case of a satellite in space, whose trajectory is necessarily curved, a good approximation of p_a is obtained by:

$$p_a \cdot T_{orb} \cdot f_a = C_{earth} \quad (2.2)$$

where T_{orb} is the orbital period, approximately 6000 s for ERS, whose ground-track length is the circumference of the Earth C_{earth} , i.e. approximately 40000 km. This value is in fact slightly larger than the image ground-track because the greater the radar's incidence angle, the more the locus of the observed ground differs from a terrestrial 'great circle'. The low incidence angle of ERS makes it point two hundred kilometers from the satellite ground-track, or 2° of a terrestrial arc. The 'circumference' of the Earth as it is imaged is no longer 40000 km but $40000 \times \cos(2^\circ)$ km or 39976 km, a negligible difference in this case.

The sampling frequency and the pulse repetition frequencies may be adjusted in a given instrument, using simple arithmetic that relates them to the carrier frequency, e.g. by means of frequency divisions which can be programmed. As an example, on the ERS-1 satellite, the carrier frequency f_c is 5300 MHz (cf. Table 3.1). An intermediary frequency is first derived by $f_i = f_c/43$, and then the sampling frequency by $f_d = 2f_i/13 = 2f_c/559$. The pulse repetition frequency f_a , is programmable using:

$$f_a = \frac{f_d}{4 \cdot (N_{pri} + 2)} \quad (2.3)$$

where N_{pri} is an integer most often equal to 2820 or to 2822. On other satellites, such as RADARSAT, there are several possibilities for f_d ('fine' or 'wide' modes, cf. Table 3.1). For any satellite, all three frequencies form dimensionless fractions which we shall find useful in subsequent chapters:

$$Q = \frac{f_c}{f_d} \quad \text{and:} \quad \chi = \frac{f_d}{f_a} \quad (2.4\text{-ab})$$

For the ERS satellites, Q is exactly equal to 279.5 and χ is usually equal to 11200.

2.2.4 An intuitive geometrical approach to synthetic aperture

The technique used in the azimuth direction to restore resolution exploits a property of radio signals. We can record both their amplitude and their phase and then process them in a particular way to produce a purely geometric interpretation. Let us imagine an antenna of length D illuminating a region whose width is L and which contains a target A (Fig. 2.5(a)). We know that the wavelength λ , the observation range R_0 , and the illuminated ground width L , are related by:

$$\frac{\lambda}{D} \approx \frac{L}{R_0} \quad (2.5)$$

where L is the 'natural' resolving power of the radar instrument. Now let us also assume that the radar transmits a pulse each time it travels the distance p_a , with p_a being less than L . We shall see under which conditions the final resolution of the image can be p_a , rather than L . If the distance p_a is considerably less than L , a target A is illuminated by several pulses as the radar instrument passes over. The number of

these pulses will be equal to the ratio L/p_a . The target A will thus contribute to each of these pulses by returning an echo composed of an amplitude, i.e., a measure of the strength with which the target returns the wave towards the radar, and of a phase, which indicates how the dielectric properties (such as conductivity) or geometric properties of the target can modify the state of vibration of the wave (thus causing a phase shift). This phase will also be offset by the wave's round trip between the radar and the target. Each time the wave travels one wavelength, its phase completes a full cycle. Each pulse to which A contributes will have approximately the same amplitude, because the angle from which A is observed varies little due to the low ratio L/R_0 . Compared to the phase at the shortest distance, the phase of each pulse will be offset by the length of the green lines in Fig. 2.5(a) (in fact it will be double this length because of the round trip).

One simple solution is to modify the phase of the samples to which the target A has contributed, so that they all have the same phase, and then to add all the samples together. If they are all added together when in phase, the samples will reinforce each other. However, each of these samples also includes, for a width of ground L , the contribution of the other targets, for example target B. This is in the vicinity of A, separated by a distance p_a . In other words, B is the closest target to A that we want to distinguish from A. The ranges traveled by the pulses from the radar to B differ from those traveled to A by the lengths shown in red. Let us assume that the sum of the lengthening in red on the furthest left and the shortening in red on the furthest right is equal to one half of a radar wavelength. Taking into account the round trip, the phase compensation algorithm, which we use to favor target A, will give regularly distributed phase values to the contributions of target B. The sum of these contributions will therefore be zero as they will cancel each other out! The next target after B, located at a distance of $2p_a$ from A, will undergo the same treatment. The only difference is that the phases of the contributions of this third target will be spread over two cycles instead of one for B. We have here a little mathematical miracle which shows us that the processing applied to the contributions of A is 'lethal' for all the other targets contributing to the same signals as A.

It can be seen from Fig. 2.5(b) that target B is 'cancelled out' by the signal processing matched to A if the relative offset of the horizontal bar of the red 'T' shape reaches a quarter of a wavelength on either side (that is, one half of a wavelength in round trip). We therefore have a condition for canceling out contributions from neighboring targets (p_a must exceed a certain value):

$$p_a/R_0 \geq \frac{\lambda/4}{L/2} \quad (2.6-a)$$

and therefore:

$$p_a \geq \frac{\lambda R_0}{2L} \quad (2.6-b)$$

This relationship also ensures that the contributions of A's immediate neighbors (respectively 2, 3, etc.) are cancelled out because their phase shifts will also pass through all possible values one or more times, in the azimuth processing window.

What are the limitations of this technique? At some point, the red lines in Fig. 2.5(a) will add half a wavelength between each sample. For a round trip, the phases will therefore be the same as for target A, because a complete phase cycle does not change the signal. This is why target C, which is called an ambiguous target, will behave like A in the processing, as will any similarly located target on the other side of A. We must therefore ensure that the antenna cannot see A and C simultaneously, i.e. that the distance between A and C must be at least equal to L . This is the ‘ambiguous shift’:

$$\frac{\lambda/2}{p_a} \geq \frac{L}{R_0} \quad (2.7a)$$

or:

$$p_a \leq \frac{\lambda R_0}{2L} \quad (2.7b)$$

Lastly, if we consider the aperture ratio (2.5), it is easy to check that we must have:

$$\frac{\lambda}{2p_a} \approx \frac{L}{R_0} \quad (2.8a)$$

and therefore:

$$p_a \approx \frac{D}{2} \quad (2.8b)$$

A pulse must therefore be transmitted every time the radar travels a distance equal to one half of the length of its antenna. As a result, the resolution which can be obtained from an antenna whose length is D is equal to $D/2$. This is contrary to everything we know about antenna resolution because in this case the smaller the antenna, the better the resolution! Stranger still, this ground resolution does not depend on the observation range. A radar located at a distance ten times further away would have the same resolution. Since there is no such thing as a free lunch in this world, there is a price to pay for these strange phenomena, in this case the amount of processing needed (there is also a cost in terms of radiometry, shown in Chap. 3). The number of samples whose phase we need to change before adding them together is equal to L/p_a , or again to $\lambda R_0/Dp_a$, or expressed otherwise as $2\lambda R_0/D^2$, so the amount of processing necessary is proportional to the range. In Fig. 2.5(a), the amount of processing is for seven samples. In the real life situation of a satellite in orbit, with a 10-m antenna, a 5 cm wavelength and an observation range of 1000 km, the number of samples is 1000. There would therefore be 1000 phase-change operations and additions for each point on the image. Since a 100-km by 100-km image contains hundreds of millions of points, hundreds of billions of mathematical operations would be needed to restore proper resolution to a space radar image. Fortunately, with the power of today’s computers, these operations do not take too much time, considerably less than one hour, especially since signal processing techniques like the Fast Fourier Transform (FFT) (Sect. 1.4), make it possible to reduce the number of calculations. The memory of these computers is also large enough to hold all the points of an image simultaneously while constructing it.

Another factor that needs to be considered is the time needed for analyzing the target. A radar with a 5 km antenna footprint in azimuth will take approximately one second to observe the target at the typical velocity of satellites in orbit ($\approx 7 \text{ km} \cdot \text{s}^{-1}$). In some cases, the target does not ‘live’ as long as that. We might think, for example, of waves on the sea or leaves shaking in the wind. The analysis time may therefore limit the efficiency of synthetic aperture radar.

2.2.4.1 An intuitive but inaccurate image

The term ‘synthetic aperture’ seems to indicate that we create an antenna with length L by azimuth processing. Indeed, if this antenna existed, it would not give us the resolution that we just calculated ($D/2$); at the range R_0 such an antenna theoretically produces a resolution $R_0\lambda/L$ equal to $2p_a$ whereas we obtain p_a . The traditional image of the parabolic antenna drawn by the extremities of the green lines (Fig. 2.5(a)) is false because usually the target is not located at the focal point of the antenna and because the round-trip aspect of the observation is ignored. To use a vocabulary typical of the world of antennas (Sect. 1.2.5.2), the target seen by an antenna of such a size is not in the far field as defined by Fraunhofer’s condition:

$$R_0 \gg \frac{2L^2}{\lambda} \quad (2.9)$$

We are thus in near field conditions despite the enormous observation range R_0 . Indeed, because L is proportional to R_0 by definition of azimuth processing, we will never satisfy far field conditions! With a few simple substitutions, we can see that fulfilling Fraunhofer’s condition would require:

$$p_a \gg \sqrt{\frac{R_0\lambda}{2}} \quad (2.10)$$

i.e., an azimuth resolution of far less quality than even that of unfocused processing (also called ‘beam sharpening’) (Sect. 2.4.7).

2.2.4.2 Velocity aberrations

In the same way that range vision produces strange (and undesirable) effects on image geometry, synthetic aperture processing can create spectacular effects: the shifting of mobile targets. In the above explanation, the targets were assumed to be fixed. What happens when they are mobile? If target A is mobile and moves towards the radar, it will add a phase change which will not be taken into account in our model. In particular, the phase changes will not be correctly represented by the green lines in Fig. 2.5(a) because of an extra contribution. Will target A disappear, however? No, because there will be a point B on the same azimuth line as A (i.e. at the same range) whose phase distribution will be precisely that of ‘mobile A’. In this case mobile A will be correctly processed by the phase distribution planned for B. After processing, A will therefore be superimposed on B!

More precisely, let v be the velocity of the satellite and v_p the fraction of the velocity of the target directed towards the orbit of the satellite (Fig. 2.5(b)). During

analysis, the target's motion creates an additional change in range, Lv_p/v . The number of unexpected phase shift cycles (compared to the phase shift created by a motionless target) is, taking into account the round trip, $2Lv_p/\lambda v$. The mobile target will therefore be satisfactorily handled with a processing offset by the same number of cycles (appropriate for one of its neighbors) and will thus appear to be in the same position as that neighbor. The azimuth distance from this neighbor is obtained by multiplying the number of cycles by the size of the azimuth pixel p_a : $2Lv_p p_a/\lambda v$ (we have seen that a shift of p_a created a phase shift of λ over a distance L). Taking Eq. (2.8a) into account, the shift in azimuth of the mobile target can then be expressed simply as:

$$\Delta = \frac{R_0 v_p}{v} \quad (2.11)$$

Let us take a detailed example. Assume that target A is a ship sailing at 20 knots, i.e. approximately $10 \text{ m}\cdot\text{s}^{-1}$ towards the ground track of the radar. Depending on the incident angle, a fraction of this velocity (for example $4 \text{ m}\cdot\text{s}^{-1}$) is directed towards the radar. For a radar velocity of $\approx 7 \text{ km}\cdot\text{s}^{-1}$ (v), at a range of 1000 km (R_0), the phase distribution for 'mobile A' will coincide with that for a target shifted by a value Δ of approximately 600 m in our example!

This effect explains why, in radar images, ships are not necessarily found at the front of their wakes, (Fig. 2.6) and why trains do not run exactly on their rails. The wakes are almost immobile and are therefore in their proper place in the image, whereas mobile elements are shifted to the left or right of their real trajectory, parallel to the satellite ground track depending on whether they are receding from the instrument or



Fig. 2.6 Gibraltar Strait observed by TERRASAR-X (stripmap image). Copyright 2007 DLR – All rights reserved.

or, in terms of change in the round trip range $\Delta R(t)$:

$$\Delta R(t) \approx \frac{k^2 p_a^2}{R_0} \quad (2.14)$$

If we continue to expand the equation, the contribution to $\Delta R(t)$ of the next term is $-k^4 p_a^4 / 4R_0^3$. It can generally be considered negligible even for the highest values of k . Even with a deviation equal to the width of the antenna radiation pattern, or $k p_a = L$, this term is equal to $-L^4 / 4R_0^3$, typically 0.2 mm in the case of ERS, which is significantly less than the wavelength of 56 mm. This term may not be always be negligible particularly in the case of airborne imagery.

The azimuth axis can be described using various coordinates. We have used the number k in azimuth pixels (with size p_a) relative to the origin A. We could also use the time elapsed since the moment of closest point of approach to A, noted t . It then becomes clear that:

$$k p_a = vt = kv/f_a \quad (2.15)$$

where v is the scalar velocity of the radar on its trajectory and f_a is the pulse repetition frequency (PRF). This gives the equivalent equation:

$$\Delta R(t) \approx \frac{k^2 p_a^2}{R_0} = \frac{v^2 t^2}{R_0} = \frac{k^2 v^2}{f_a^2 R_0} \quad (2.16)$$

The round trip distance $\Delta R(t)$ is the path lengthening of the radar wave, relative to the reference round trip $2R_0$. This lengthening, after multiplying by $2\pi/\lambda$ results in a phase difference $\Delta\phi(t)$:

$$\Delta\phi(t) \approx 2\pi \cdot \frac{k^2 p_a^2}{\lambda R_0} = 2\pi \cdot \frac{v^2 t^2}{\lambda R_0} = 2\pi \cdot \frac{k^2 v^2}{\lambda f_a^2 R_0} \quad (2.17)$$

We will call the dimensionless factor the 'azimuth compression rate', N_a :

$$N_a = \frac{\lambda R_0}{2p_a^2} = \frac{\lambda R_0 f_a^2}{2v^2} \quad (2.18)$$

Consequently:

$$\Delta\phi(k) \approx \pi \frac{k^2}{N_a} \quad (2.19)$$

N_a represents the size of the azimuth processing window, or alternatively the number of points used to 'construct' the target's response. It is also the azimuth distance between a point and its ambiguous location (ghost target). In mathematical terms, in this context, N_a is the solution to the equation:

$$\Delta\phi(k + N_a + 1) - \Delta\phi(k + N_a) = \Delta\phi(k + 1) - \Delta\phi(k) \quad \text{for any } k \quad (2.20)$$

2.2.5.1 Coherent addition of radar echoes

A radar target C is generally characterized by the amplitude α_c and the ‘natural’ phase ϕ_c of its backscattered echo. Its complex expression is $\alpha_c \exp(j\phi_c)$, where j is defined by $j^2 = -1$. The contribution of the target C to a radar echo such that the latter is shifted by k azimuth pixels relative to the closest point of approach is therefore: $\alpha_c \exp(j\phi_c) \exp(j\Delta\phi(k))$.

In reality, the contribution of the target is also weighted by the azimuth antenna radiation pattern (Sect. 1.2.5.1) which depends on the position of the target, and thus on k (the index of azimuth pixels). The weighting coefficient will be set to 1 at the center of the antenna radiation pattern ($\beta(0) = 1$), equal to $\beta(k)$ elsewhere (usually in the shape of a cardinal sine curve). Although the antenna radiation pattern is not band limited, we assume that $\beta(k) = 0$ for $|k| > N/2$, where N is the number of samples contained in the width of the radiation pattern.

As a general rule, the signal $s(l)$, associated with the pulse number l , is made up of the full set of target contributions $\alpha_k \exp(j\phi_k)$ present in the antenna radiation pattern at position k , weighted by the latter and modified by the phase distribution $\Delta\phi(k)$:

$$s(l) = \sum_{k=l-\frac{N}{2}}^{k=l+\frac{N}{2}} \beta(k-l) \cdot \alpha_k \exp(j\phi_k) \exp(j\Delta\phi(k-l)) \quad (2.21)$$

The processing, when matched to the target whose closest point of approach to the radar is located at index m , involves adding the signals acquired when M is illuminated by the antenna. The signals are corrected for their respective phase distributions and radiation pattern weightings. The output $s_f(m)$ of this process is given by:

$$s_f(m) = \sum_{k'=m-\frac{N}{2}}^{k'=m+\frac{N}{2}} \beta(k'-m) \cdot s(k') \cdot \exp(-j\Delta\phi(k'-m)) \quad (2.22)$$

In the particular case where there is a single target with amplitude $\alpha_m \exp(j\phi_m)$, assumed to be located at m then $s(k') = \beta(k'-m) \cdot \alpha_m \exp(j\phi_m) \cdot \exp(j\Delta\phi(k'-m))$, and applying Eq. (2.22) leads to:

$$s_f(m) = \sum_{k'=m-\frac{N}{2}}^{k'=m+\frac{N}{2}} \beta^2(k'-m) \cdot \alpha_m \exp(j\phi_m) = \xi \cdot \alpha_m \exp(j\phi_m) \quad (2.23)$$

where ξ is a constant of the system.

In order to evaluate to what extent this processing favors the target m , let us consider the results of the processing for a target in the vicinity of m , located at $m+h$. It contributes the quantity $\beta(k'-m-h) \cdot \alpha_{m+h} \exp(j\phi_{m+h}) \cdot \exp(j\Delta\phi(k'-m-h))$ to the signals found between $k' = m+h-N/2$ and $k' = m+h+N/2$, which leaves a residual $r(h)$ such that:

$$\begin{aligned}
 r(h) &= s_f(m+h) = \alpha_{m+h} \exp(j\phi_{m+h}) \cdot \\
 &\sum_{k'=m+h-\frac{N}{2}}^{k'=m+\frac{N}{2}} \beta(k'-m) \cdot \beta(k'-m-h) \\
 &\times \exp(j\Delta\phi(k'-m-h)) \cdot \exp(-j\Delta\phi(k'-m))
 \end{aligned} \tag{2.24}$$

If we let $n = k' - m$, we finally get:

$$\begin{aligned}
 r(h) &= \alpha_{m+h} \exp(j\phi_{m+h}) \cdot \sum_{n=h-\frac{N}{2}}^{n=\frac{N}{2}} \beta(n) \cdot \beta(n-h) \cdot \exp(j\Delta\phi(n-h)) \cdot \\
 &\exp(-j\Delta\phi(n))
 \end{aligned} \tag{2.25}$$

Given Eq. (2.19), by expanding Eq. (2.25) and replacing the lower limit $n = h - N/2$ by $n = -N/2$ (since the additional values of $\beta(n-h)$ are null), the result is:

$$r(h) = \exp\left(j\pi \frac{h^2}{N_a}\right) \cdot \sum_{n=-\frac{N}{2}}^{n=\frac{N}{2}} \beta(n) \cdot \beta(n-h) \cdot \exp\left(-j\pi \frac{2nh}{N_a}\right) \tag{2.26}$$

By choosing f_a such that the compression rate N_a is equal to N (a correctly sampled system, cf. Sect. 3.7.2), for $h = 1$ (immediate vicinity), the phase term $-\pi 2nh/N_a$ varies between π and $-\pi$ on the interval $[-N/2; N/2]$. Here again the contributions cancel each other out.

2.2.5.2 Synthesis of mobile targets

Let us again consider the problem of mobile targets, treated here from an analytical point of view. Consider a moving target whose velocity includes a component \vec{v}_p directed towards the radar (Fig. 2.7). If we take the entire round trip into account, the actual phase distribution between the target and the radar will be:

$$\begin{aligned}
 \Delta\phi(t) &\approx 2\pi \frac{v^2 t^2}{\lambda R_0} + 4\pi \frac{v_p \cdot t}{\lambda} = \frac{2\pi v^2}{\lambda R_0} \left(t^2 + 2R_0 \frac{v_p}{v^2} t \right) \\
 &= \frac{2\pi v^2}{\lambda R_0} \left(\left(t + R_0 \frac{v_p}{v^2} \right)^2 - R_0^2 \cdot \frac{v_p^2}{v^4} \right)
 \end{aligned} \tag{2.27}$$

In addition to the constant phase $-2\pi R_0 \cdot v_p^2 / (\lambda v^2)$, everything happens as though the time t had undergone a shift $t_p = R_0 v_p / v^2$, i.e. a displacement along the azimuth axis. This displacement can be expressed as the length x_p or as a number of samples k_p by multiplying by the appropriate proportionality factor, recovering the results found in Sect. 2.2.4.2:

$$x_p = R_0 \cdot \frac{v_p}{v} \quad \text{and:} \quad k_p = f_a \cdot R_0 \cdot (v_p / v^2) \tag{2.28}$$

2.2.5.3 Effect of the radar's trajectory on the processing

The linear-trajectory hypothesis that we have been applying is not completely accurate. When a radar is installed on an aircraft, the trajectory can be perturbed by atmospheric turbulence or piloting. These can cause shifts in the trajectory which are greater than a fraction of the radar's wavelength. In this case, we can reconstruct the trajectory bit by bit if we have precise data on the trajectory, obtained from a navigation system.

When a radar is carried by a satellite, the trajectory is very regular, but it is curved as a result of acceleration forces. At its closest point of approach to a target A, the satellite's position is expressed by a vector \vec{s} , its velocity by a vector \vec{v} and its acceleration by a vector $\vec{\gamma}$; this last variable includes accelerations caused by gravity, the centrifugal force and the Coriolis force (we are assuming a non-inertial reference frame in which the Earth is fixed). If we assume that \vec{a} designates the vector giving the position of the target, the range R between the satellite and the target will vary according to:

$$R(t) = \left\| \vec{a} - \vec{s} - \vec{v} \cdot t - \frac{1}{2} \cdot \vec{\gamma} \cdot t^2 \right\| \quad (2.29)$$

If we expand this in the same way as for Sect. 2.2.5, we then get:

$$R(t) \approx \|\vec{a} - \vec{s}\| \cdot \left(1 + \frac{(\vec{s} - \vec{a}) \cdot \vec{v}}{\|\vec{a} - \vec{s}\|^2} \cdot t + \frac{(\vec{s} - \vec{a}) \cdot \vec{\gamma} + \|\vec{v}\|^2}{2 \cdot \|\vec{a} - \vec{s}\|^2} \cdot t^2 \right) \quad (2.30)$$

For the closest point of approach $(\vec{s} - \vec{a}) \cdot \vec{v} = 0$, we obtain:

$$R(t) \approx \|\vec{a} - \vec{s}\| \cdot \left(1 + \frac{(\vec{s} - \vec{a}) \cdot \vec{\gamma} + \|\vec{v}\|^2}{2 \cdot \|\vec{a} - \vec{s}\|^2} \cdot t^2 \right) \quad (2.31)$$

Let \hat{u} be the unit vector pointing from the target towards the satellite where $\hat{u} = (\vec{s} - \vec{a})/R_0$. The shortest range between the target and the satellite is $R_0 = \|\vec{a} - \vec{s}\|$. In terms of variation in the round-trip range $\Delta R(t)$, we have (Fig. 2.7):

$$\Delta R(t) \approx \left(\hat{u} \cdot \vec{\gamma} + \frac{\|\vec{v}\|^2}{R_0} \right) \cdot t^2 \quad (2.32)$$

If we apply Eq. (2.32) to Eqs. (2.16)–(2.18), we obtain a new expression for the compression rate N_a , which takes into account the satellite's acceleration:

$$N_a = \frac{\lambda R_0 f_a^2}{2(R_0 \hat{u} \cdot \vec{\gamma} + v^2)} \quad (2.33)$$

In the case of the ERS satellite, where $R_0 = 850 \text{ km}$ and at a latitude of 40° , $\hat{u} \cdot \vec{\gamma}$ and $\|\vec{v}\|^2/R_0$ have values of $-7.7 \text{ m} \cdot \text{s}^{-2}$ and $67 \text{ m} \cdot \text{s}^{-2}$ respectively. Lastly, in the event that the target should have its own velocity \vec{V}_p , the component \vec{v}_p directed towards the radar is $\vec{v}_p = \hat{u} \cdot \vec{V}_p$.

2.3 Frequency representation

It is also possible to represent the data in terms of frequency. The frequency of a periodic phenomenon is given by the number of phase cycles per second. If the phase is a linear function of time, and if its duration is infinite, the phenomenon is characterized by one single, perfectly determined frequency. Even if the phenomenon does not possess these properties, we can define an instantaneous frequency equal to the derivative of the phase of the phenomenon, normalized by 2π , which is the value of one phase cycle. We can therefore write:

$$f = -\frac{1}{2\pi} \frac{d\phi}{dt} \quad (2.34)$$

Frequency representation has numerous theoretical and practical advantages. Many natural phenomena are best expressed in terms of frequency, and many filters are best characterized by their frequency response (low-pass and high-pass filtering, etc.). We can all recall those simple electrical setups that demonstrate these functions, using coils and condensers. The main practical advantage arises from a fundamental property of frequency representation, namely, the possibility of applying a stationary (i.e. not changing in time) filter by simple multiplication in the frequency domain. Let $s(t)$ be the complex signal to be processed and $s_r(t)$ the reference signal. The convolution that leads to the processed complex signal $s_f(t) = s(t) * s_r(t) = \int_{-\infty}^{+\infty} s(\tau) \cdot s_r(t - \tau) \cdot d\tau$ is obtained by applying the frequency transformation (i.e. the direct Fourier Transform Sect. 1.4) to the signals $s(t)$ and $s_r(t)$, which produces the signals $S(f)$ and $S_r(f)$. Multiplying these two signals point by point, results in the transform $S_f(f)$ of the desired output. We then apply the inverse Fourier transform to obtain $s_f(t)$.

The practical advantage of this method lies in its computational efficiency. The Fourier Transform of a signal with N samples requires $N \cdot \log_2 N$ operations using the Fast Fourier Transform (*FFT*) (Sect. 1.4.8). If we assume that the reference signal is transformed once and for all, or that it is produced directly in the frequency domain, $N \cdot \log_2 N$ operations will be necessary to transform the signal, N operations for the point by point multiplication and lastly $N \cdot \log_2 N$ operations for the inverse transform. Neglecting edge effects, the number of operations per point processed is usually about $1 + 2\log_2 N$, whereas it is close to N for direct correlation.

On the other hand, a condition necessary for this efficiency is that the number of points N of the Fourier transform be significantly greater than the number of points N_a required by the processing. This minimizes the inconvenience of losing the incompletely correlated $N - N_a$ points. It is necessary that the correlation processing, i.e. the reference signal $s_r(t)$, remain stable (stationary) over this length N . This condition is not always fulfilled by airborne radar systems, because the phase distribution for the reconstruction is frequently perturbed by flight irregularities, which often exceed the scale of a wavelength.

2.3.1 Phase distribution expressed in the frequency domain

Equations (2.17) and (2.34) can be used to derive the instantaneous frequency of the signal at time t :

$$f(t) = -\frac{2v^2 t}{\lambda R_0} = -\frac{t f_a^2}{N_a} \quad (2.35)$$

For the radar signals that we are discussing, the instantaneous frequency is therefore proportional to time. (2.35) shows that the frequency span (the Doppler bandwidth B^{dop}) is proportional to the duration of the target's illumination. When the closest point of approach to the radar is located in the middle of the zone illuminated by the antenna (Fig. 2.8a), this Doppler bandwidth is centered on the zero Doppler frequency, giving what is called the 'zero mean Doppler'; we then have a 'zero Doppler centroid', acquisition.

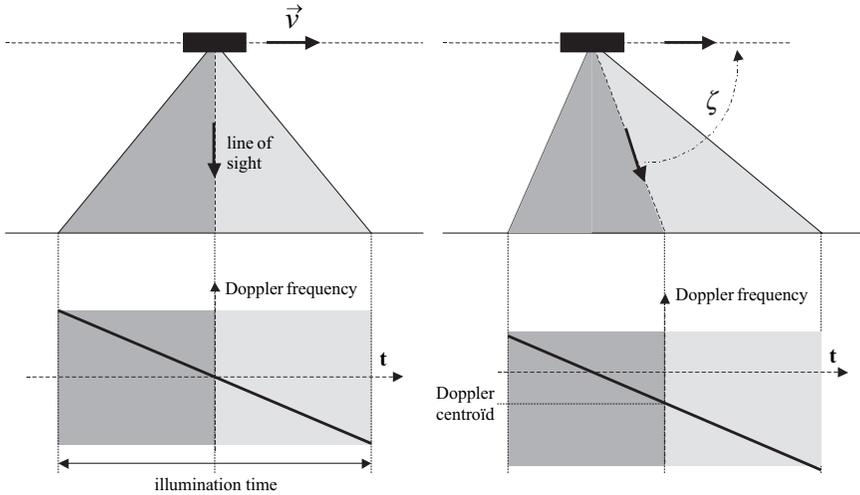


Fig. 2.8 (a) Deviation of the Doppler frequency in the case of an acquisition with zero mean Doppler (zero Doppler centroid); (b) Deviation of the Doppler frequency in the case of a Doppler acquisition with non-zero mean Doppler (non zero Doppler centroid).

We can correct the phase distribution $\Delta\phi(t)$ of the signal after it has been converted into frequency. Since:

$$t = -\frac{\lambda R_0 f}{2v^2} \quad (2.36)$$

we can express the phase distribution in the frequency domain:

$$\Delta\phi(f) = \pi \cdot N_a \cdot \frac{f^2}{f_a^2} \quad (2.37)$$

Using the reduced frequency f_r defined by $f_r = f/f_a$ we can obtain equations in an even more compact form. The phase distribution in reduced frequency then becomes:

$$\Delta\phi(f) = \pi \cdot N_a \cdot f_r^2 \quad (2.38)$$

Please note: a result of Eq. (2.36) is that: $B^{dop} = 2v^2 T^{ill}/(\lambda R_0)$, where T^{ill} is the duration of illumination of the target. Since we also have: $T^{ill} = \lambda R_0/(Dv)$ (where D is the width of the antenna), the final result is:

$$B^{dop} = 2v/D \quad (2.39)$$

The corresponding azimuth spatial resolution is given by $r_a = D/2$ as already established in Sect. 2.2.4.

2.3.2 Non-Zero Mean Doppler

Up until now, we have assumed that the closest point of approach of the radar was located in the middle of the zone illuminated by the antenna (Fig. 2.8(a)). In reality, this condition is never fully satisfied (Fig. 2.8(b)). Often, the closest point of approach is not illuminated at all by the antenna. Does this affect the complementarity between range and azimuth that we have mentioned previously? Not necessarily. For instance, an angle ζ of 80° , (i.e. 10° off-perpendicular), between the direction of the flight path and the radar to-target line of sight still allows us to process the raw data correctly in the two directions. We simply find that the two one-dimensional impulse responses (one in the range direction and the other in the azimuth direction, Sect. 3.12) can be non-perpendicular or ‘skewed’. The shift between the actual pointing and the pointing perpendicular to the flight path may exceed the width of a typical azimuth antenna radiation pattern. For ERS-1, this value is close to 0.3° , which is much smaller than the 10° skew.

For different radar satellite projects, the engineers may or may not seek to satisfy the condition of perpendicular pointing. In the case of the ERS-1 and ALMAZ satellites (Table 3.1), the attitude of the instruments is controlled in such a way as to minimize the deviation from perpendicular pointing. This minimization requires an attitude law that varies according to the satellite’s position on its orbit. The apparent velocity of the target observed by the radar is a combination of the velocity of the Earth’s rotation and the orbital velocity of the satellite. Observation satellites never follow an equatorial orbit; they cross the equator and ascend as far North as the inclination of their orbit will take them (the inclination of an orbit is roughly equal to the maximal latitude reached by ground projection of the satellite’s trajectory, whether to the North or to the South). They then fly back down to the equator, crossing it with an angle opposite to that of the northbound trajectory, before arriving at their extreme Southern point and starting the cycle again. To simplify, let us imagine a satellite whose inclination is 90° (a purely North-South trajectory). This satellite crosses the equator at a right angle when climbing towards the North. Its orbital velocity (\vec{v}), of approximately $7000 \text{ m}\cdot\text{s}^{-1}$ is combined with the rotational velocity of the Earth at the equator (\vec{v}_{rot}),

which is $463 \text{ m}\cdot\text{s}^{-1}$. To achieve pointing perpendicular to the resultant velocity, the antenna must be pointed slightly forward, with an angle α_p of 3.8° (whose tangent is equal to the ratio of the velocities) (Fig. 2.9). When crossing the equator southbound, the antenna will have to be pointed at the same angle but backwards. It is only when crossing the poles that the pointing will be perpendicular to the plane of the orbit. For non-equatorial orbits, maintaining perpendicular pointing requires the use of an attitude control law for the entire orbit. For most satellites, this law is based on an estimate of the satellite's attitude provided by dedicated sensors. This method is limited by the accuracy of the attitude sensors. Alternatively, the attitude can be controlled by the radar. In this case the attitude is maintained through a real-time estimation of the 'mean Doppler'¹ calculated from the radar data. This was the system used for the ALMAZ satellite. The mean Doppler effect perceived by the antenna should be zero when it is observing homogenous terrain perpendicularly to its flight path. Even if the radar data have not been azimuth processed at this stage, the positive Doppler effects from the leading edge of the zone illuminated by the antenna are compensated for by the negative Doppler effects created in the trailing part of this zone.

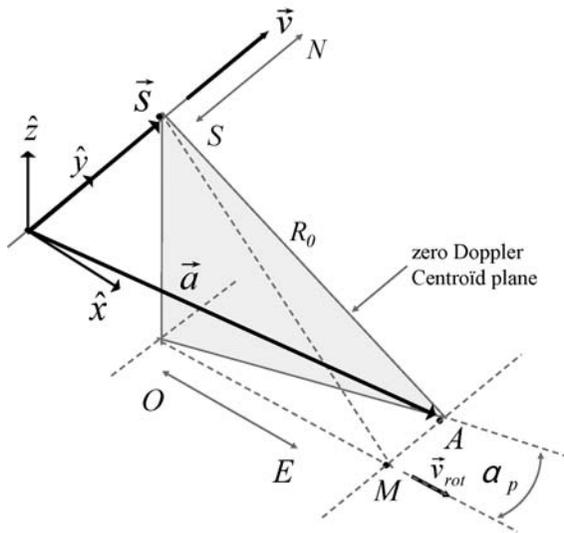


Fig. 2.9 Forward antenna off-pointing to compensate for the effect of the Earth's rotation to ensure zero mean Doppler acquisition (case of an ascending orbit).

The effect of the spacecraft's attitude on the perpendicularity of the observation depends principally on two of the three attitude angles (Fig. 2.9): the yaw (motion in the plane (\hat{x}, \hat{y}) , on either side of \hat{y}) and the pitch (motion in the plane (\hat{y}, \hat{z}) on either side of \hat{y}). The influence of roll (motion in the plane (\hat{x}, \hat{z}) on either side of \hat{z}) is usually negligible. The relative importance of yaw and pitch depends on the radar's incidence angle. The importance of pitch increases with the verticality of the line of sight. A

¹'mean Doppler' and 'Doppler centroid' are equivalent expressions.

horizontal line of sight is not possible in space borne radar; airborne radar, however, can come close to achieving a horizontal line of sight and, in this case, only yaw is significant. The value of mean Doppler f_m is given by:

$$f_m = \frac{2(\vec{s} - \vec{a}) \cdot \vec{v}}{\lambda R_0} \quad (2.40)$$

where \vec{a} is the vector representing the central position targeted by the antenna when the satellite is at position \vec{s} with velocity \vec{v} . We can obtain a value for a dimensionless mean ‘reduced’ Doppler f_{mr} , after dividing by the azimuth frequency f_a :

$$f_{mr} = \frac{2(\vec{s} - \vec{a}) \cdot \vec{v}}{\lambda R_0 f_a} \quad (2.41)$$

This is the value that we seek to extract from the radar data.

2.3.3 Doppler locking

Synthetic aperture radar processing requires the central azimuth position, which corresponds to the maximum of the antenna pattern. As we saw in Sect. 2.2.5, the azimuth phase distribution can be expressed in several ways – as a function of the time, of the position on the azimuth axis, of the azimuth sample number or of the reduced frequency, where the last two are dimensionless. To the small angle approximation (Fig. 2.7), the angle $\Delta\alpha \approx kp_a/R_0 = \lambda f_a f_r/2v$ is also a potential description variable along the track. We thus have:

$$\Delta\phi_a \approx 2\pi \cdot \frac{v^2 t^2}{\lambda R_0} = 2\pi \cdot \frac{x^2}{\lambda R_0} = \pi \cdot \frac{k^2}{N_a} = \pi \cdot N_a f_r^2 \quad (2.42)$$

In addition, we need to know in which interval the phase distribution will be applied, either $\{t_1; t_2\}$, $\{x_1; x_2\}$, $\{k_1; k_2\}$ or $\{f_{r1}; f_{r2}\}$. For example, for centered processing (zero mean Doppler), and using sample number representation, the interval is (cf. Sect. 2.2.5.1): $\{-N_a/2; N_a/2\}$. In the general case, the antenna beam is not strictly perpendicular to the relative motion of the targets on the ground. In order to determine the azimuth orientation of the antenna (i.e., the angle ζ of Fig. 2.8(b)), and consequently the position of the center of processing, we can use the radar data expressed in the frequency domain. The goal is to obtain the shape of the antenna radiation pattern, which gives a weighting to the targets as a function of their distance from the center of the radiation pattern (where gain is maximal). To obtain the shape of the antenna radiation pattern, we cannot use variables in the time domain, because for every target at the maximum point of the radiation pattern there are others which enter or leave and all these contributions are merged. In other words, each sample in the time domain contains contributions from targets in every possible position of the radiation pattern. In contrast, if we apply a Fourier transform to the azimuth data, the data will be ranked by frequency, meaning that they will be distributed according to the derivative of their instantaneous phase. Even if two targets have not crossed the antenna radiation pattern at the same time, their contributions to a given point on the antenna radiation pattern will be located at the same point in the frequency analysis. The amplitude of this point

must therefore reflect the shape of the antenna radiation pattern. This can be seen in Fig. 2.10, where several lines of data, after conversion into azimuth frequency, have been summed quadratically. We can also see the disadvantage of representation by frequency: its ambiguity. The position of the maximum of the response is only known to an integer multiple of the azimuth sampling frequency f_a (the pulse repetition frequency in the time domain leads to periodicity in the frequency spectrum, Sect. 1.4.6). At the point of minimum amplitude in Fig. 2.10 the aliased edges of the antenna radiation pattern, actually separated by f_a , overlap and combine with a ‘floor’ of white noise.

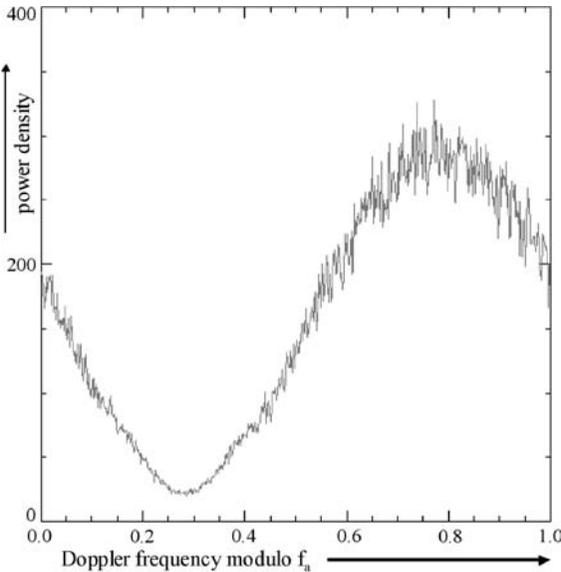


Fig. 2.10 Doppler azimuth spectrum of a SAR image (case of a non-zero mean Doppler).

2.3.4 Mean Doppler (or Doppler centroid) estimation

A classic procedure for establishing the mean Doppler consists in calculating the quadratic mean of Fourier transforms in azimuth, with lengths greater than the number of points necessary for synthesis. This precaution is taken to ensure that the contributions of the majority of targets involved are complete. Otherwise, there is the risk of a very powerful target biasing the measurements. (This could occur if it was only in the field of vision for a part of its passage through the antenna radiation pattern). We then determine the maximum of the resulting spectrum, which provides the mean fractional reduced frequency f_{mrf} , located between -0.5 and $+0.5$, whereas the real mean reduced frequency is equal to $f_{mr} = f_{mrf} + n$, where n is an integer.²

²The azimuth frequency bandwidth generally corresponds to a width of several kilometers on the ground (i.e. the beamwidth in the azimuth direction). It is possible to determine n from the satellite’s attitude, if its accuracy induces a pointing error that is less than this beamwidth. We will see, however (Sect. 3.3.5), that the integer n can also be determined from the radar data alone.

However there is a faster and simpler method for measuring the mean Doppler by calculating the correlation between neighboring points. Let $I(m, n)$ be the complex number representing sample n from the raw data resulting from pulse number m (m represents the line number, n the column number). The correlation τ is approximated by:

$$\tau = \sum_{i=1}^{i=long} \frac{I(m, i) \cdot I^*(m+1, i)}{\sqrt{I(m, i)^2 \cdot I(m+1, i)^2}} \quad (2.43)$$

The value of *long* may be lower than the number of samples of the pulse, depending on whether the correlation is estimated locally or globally in range, and on the accuracy required: the accuracy of the estimate will depend on the number of points used. Hence, the correlation can include the contribution of neighboring lines, i.e. by line $m+1$ combined with $m+2$, $m+2$ combined with $m+3$, etc. This way of calculating the correlation is much faster than calculating a Fourier transform because it is proportional to the number of points of raw data processed.

We shall now show that the phase ϕ_τ of τ is the estimation of the mean reduced Doppler frequency of the data multiplied by 2π : $f_{mr} = \theta_\tau/2\pi$. At the same time, its modulus $|\tau|$ is characteristic of the degree of similarity between neighboring lines. This modulus contains information about the noise affecting the signal.

To calculate the azimuth data spectrum, we perform the Fourier transform in azimuth of the raw data and then calculate the square of the modulus (possibly averaged over several lines). This modulus squared is none other than the Fourier transform of the auto-correlation of the raw signal in azimuth (cf. Sect. 1.4.5.5). Fig. 2.10 shows an example of this result. It can be almost perfectly modeled by adding a constant and a cosine function whose period equals the length of the spectrum. In terms of frequency, the constant represents the zero frequency and the cosine the first frequency (i.e. the lowest). The constant will therefore be the first term of the Fourier transform of the spectrum and the amplitude of the cosine, the second. But since the spectrum is itself the Fourier transform of the auto-correlation, the second term of the auto-correlation, which corresponds to a shift of 1, describes the cosine. More precisely, if w_n is the level of the noise and w_s that of the signal, the amplitude of the cosine is $\alpha \cdot w_s$, where α is a coefficient which depends only on the azimuth oversampling selected by the radar design. α is therefore constant if the antenna's geometry, the kinematics of the image acquisition and the frequency f_a do not change. Under these conditions:

$$|\tau| \approx \frac{\alpha \cdot w_s}{w_n + w_s} \quad (2.44)$$

A typical value of α is 0.3.

Finally, the phase of τ expresses the way in which the cosine function is phased in Fig. 2.10, and is therefore characteristic of the mean reduced Doppler frequency:

$$f_{mr} \approx \frac{\phi_\tau}{2\pi} \quad (2.45)$$

The accuracy of the estimation of this frequency is a function of the similarity of neighboring lines and therefore of the value of $|\tau|$. If this latter is close to its nominal value (≈ 0.3), the estimation is highly reliable. If not, then either the signal to noise ratio is poor (for example, $w_n \geq w_s$), or the lines used are not in fact neighbors. The algorithm can therefore also be used to detect missing lines in the raw data. When even a single line is missing, we actually compare a line to a neighbor offset by two pulses. The similarity measured by $|\tau|$ therefore becomes very low (the lines are almost independent). This makes it impossible to distinguish between one missing line and two or more missing lines.

2.3.5 Mean reduced Doppler estimation (integer part)

It is possible to calculate the integer n (Sect. 2.3.4) using only the radar data. We use a process involving two sub-images (also called sub-looks) created from the edges of the antenna radiation pattern (sub-look generation is explained in Sect. 3.2.5). To make this clear, let us assume that the fractional mean reduced frequency is f_{mrf} . In the frequency domain, the two sub-looks are centered on the reduced frequencies: $f_{mrf} - 0.25$ and $f_{mrf} + 0.25$. Each sub-look is processed on a quarter of the available azimuth frequency range, i.e. between $f_{mrf} - 0.375$ and $f_{mrf} - 0.125$ in the case of the first sub-look.

If the integer part of the mean reduced frequency is n_1 , the sub-looks will be corrected for range so that they appear to be at the range of ‘zero Doppler’. This correction, which is assumed to be constant within an entire sub-look, will be:

$$\mu_1 = \frac{N_a}{2Q} (f_{mrf} - 0.25 + n_1)^2 \quad (2.46-a)$$

and:

$$\mu_2 = \frac{N_a}{2Q} (f_{mrf} + 0.25 + n_1)^2 \quad (2.46-b)$$

After these corrections, the two sub-looks will appear to be at the same range, that of the closest point of approach. However, if the processing is carried out with an erroneous integer part n_2 of the mean reduced frequency, the corrections will also be erroneous and will be:

$$\mu'_1 = \frac{N_a}{2Q} (f_{mrf} - 0.25 + n_2)^2 \quad (2.47-a)$$

and :

$$\mu'_2 = \frac{N_a}{2Q} (f_{mrf} + 0.25 + n_2)^2 \quad (2.47-b)$$

The sub-looks will therefore be shifted in range by a number of pixels Δk equal to:

$$\Delta k = (\mu_1 - \mu'_1) - (\mu_2 - \mu'_2) \quad (2.48)$$

or:

$$\Delta k = \frac{N_a}{2Q}(n_2 - n_1) \quad (2.49)$$

In the case of ERS-1, the coefficient $N_a/2Q$ is usually 2.5, so that a shift of one in the estimation of n produces a shift of the sub-looks of 2.5 pixels in range. This is easily detected by correlating the sub-looks as it is only necessary to distinguish between cases where the shift is $-5, -2.5, 0, 2.5$ etc, (since $n_2 - n_1$ can only be an integer). It may seem that since the integer part of the reduced frequency is not known when the sub-looks are being processed, their image quality will be poor. In reality however, the sub-looks are perfectly acceptable because of the narrowed azimuth processing window. If we take the case of sub-look 1, the values of the range offset (called migration) at the edges of the processed bandwidth are, with the value n_1 :

$$\mu_{\text{inf}1} = \frac{N_a}{2Q}(f_{\text{mrf}} - 0.375 + n_1)^2 \quad (2.50\text{-a})$$

and:

$$\mu_{\text{sup}1} = \frac{N_a}{2Q}(f_{\text{mrf}} - 0.125 + n_1)^2 \quad (2.50\text{-b})$$

and with the value n_2 :

$$\mu'_{\text{inf}1} = \frac{N_a}{2Q}(f_{\text{mrf}} - 0.375 + n_2)^2 \quad (2.51\text{-a})$$

and:

$$\mu'_{\text{sup}1} = \frac{N_a}{2Q}(f_{\text{mrf}} - 0.125 + n_2)^2 \quad (2.51\text{-b})$$

If we process the n_1 case using n_2 , the sub-look 1 will be spread in range by a number of pixels Δk equal to:

$$\Delta k = \mu_{\text{inf}1} - \mu'_{\text{inf}1} - (\mu_{\text{sup}1} - \mu'_{\text{sup}1}) = \frac{N_a}{4Q}(n_2 - n_1)^2 \quad (2.52)$$

The sub-look will therefore be spread over a range corresponding to a degradation in range resolution less than the range offset between the two sub-looks. In our example with ERS-1, the spread in range is equal to 1.2 range pixels.

We have created two sub-looks separated by one half of the frequency bandwidth, each made up of one quarter of the frequency bandwidth available in the data. We could have chosen a larger offset between the central frequencies of the sub-images, in order to maximize the offset in range between the two sub-looks caused by an error of n . We could also decrease the frequency width used to process the sub-looks, in order to reduce their spread in range. Each of these possibilities may introduce more disadvantages than advantages. If we separate the sub-looks any further in frequency,

they will appear on parts of the antenna radiation pattern where the ambiguity is very high (cf. Sects. 3.7 and 3.13) and where the signal to noise ratio is low (Fig. 2.10). If we reduce the frequency width of the sub-looks, we will force the algorithm to function with a smaller fraction of the energy of the data, which may degrade the correlation between the sub-looks.

2.3.6 Range migration

Range and azimuth processing are not as independent of each other as we might hope. Variations in range between the target and the radar produce not only the phase distribution that we use for the processing, but also an offset in range that can reach several pixels. In other words, after range compression processing (Sect. 2.3.7), the different contributions in azimuth of a given target do not fall in exactly the same range cell (Fig. 2.11). The azimuth phase distribution can tell us the extent of this effect if we divide it by 2π and by Q , the ratio of the radar carrier frequency f_c to its sampling frequency f_d . More precisely from Eq. (2.42), which expresses the variation of the round trip phase as a function of the position of sample k relative to the closest point of approach, we can determine the corresponding variation in range by:

$$\Delta R(\text{mig}) = \frac{\lambda}{2} \cdot \frac{k^2}{N_a} \quad (2.53)$$

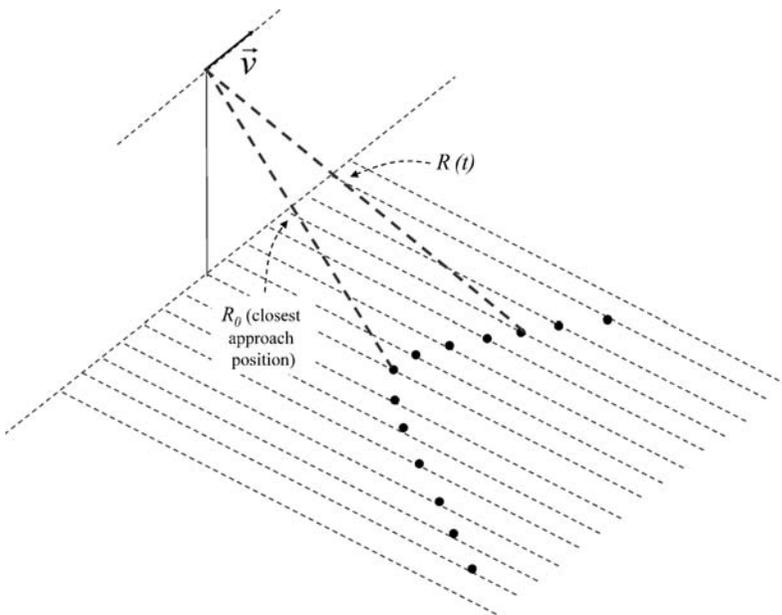


Fig. 2.11 Position of the different contributions of a given target after range compression (case of an acquisition with zero mean Doppler).

By definition, the migration $\mu(k)$, expressed as a number of pixels, is equal to the ratio of $\Delta R(mig)$ to twice³ the length of a range pixel p_d (sect. 2.2.3.):

$$\mu(k) = \frac{\Delta R(mig)}{2p_d} = \frac{\lambda}{2p_d} \cdot \frac{k^2}{2N_a} = \frac{f_d}{f_c} \cdot \frac{k^2}{2N_a} = \frac{k^2}{2QN_a} \quad (2.54)$$

or, as a function of reduced frequency:

$$\mu(f_r) = \frac{N_a f_r^2}{2Q} \quad (2.55)$$

These equations provide a very simple way of obtaining the order of magnitude of migrations. A typical example from the ERS-1 satellite, whose antenna is turned slightly forward and which observes between the reduced frequencies 0 and 1, leads to a migration range comprised between 0 and 2.5 range pixels (for this we take a realistic value of $N_a = 1400$). Considerable migration can occur in a satellite which is not maintained at zero Doppler. In the case of the J-ERS satellite (for which $Q = 74.67$), observation between the reduced frequencies 1 and 2, with $N_a = 3850$, gives migration between 25.8 and 103. Even if we only consider the variation of the migrations across the width of the azimuth processing we still obtain 77 range pixels.

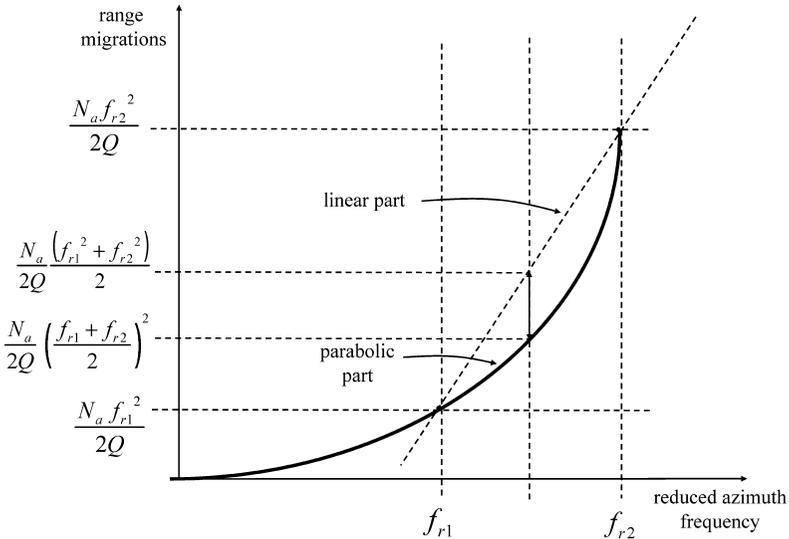


Fig. 2.12 Linear and parabolic parts of the range migrations.

A distinction is sometimes made between linear and parabolic migrations (Fig. 2.12). This distinction between two components of the same phenomenon comes from the processing algorithm. The linear part of the migrations can be dealt with during the

³The factor 2 takes account of the round trip, which is already incorporated in $\Delta R(mig)$.

range processing, whereas the parabolic part must wait until the azimuth processing begins. The amplitude of the parabolic part is usually much lower than that of the linear part. In the case of our example with the J-ERS satellite, the maximum parabolic amplitude is given by the value of the migration at the center of the azimuth analysis window ($f_r = 1.5$ for our example) less the average value of the extreme migrations, resulting in a value of -19 pixels. In the general case, the strictly parabolic part μ_p of the migrations can be calculated by:

$$\mu_p = \frac{N_a}{2Q} \cdot \left[\left(\frac{f_{r1} + f_{r2}}{2} \right)^2 - \frac{f_{r1}^2 + f_{r2}^2}{2} \right] \quad (2.56)$$

Qualitatively speaking, what is the impact of an error in migration processing? There is an effect in range and in azimuth. In range, there is a spreading effect: during azimuth processing, if an incorrect migration leaves the contribution of a point spread across five range pixels: the range resolution will clearly be degraded by a factor approximately equal to five. What is more surprising is the fact that the azimuth resolution will also be degraded by a factor of five. This is because of the actual bandwidth ‘seen by the target’ in azimuth: instead of remaining on the same line during azimuth processing, the contribution of the badly compensated target will drift across five azimuth lines. The frequency width of this contribution on each azimuth line will therefore only be one fifth of the azimuth processing frequency width, resulting in a degradation of the azimuth resolution by a factor of five.

2.3.7 Range processing

The range of targets is obtained by sorting the echo of each pulse as a function of its time of return. It is therefore linked to the sampling frequency of this echo f_d . The pulse duration also needs to be approximately the same as the sampling period⁴ $1/f_d$. In the world of signal processing, we know that a short pulse with a typical duration of $1/f_d$ has a typical frequency bandwidth f_d (Sect. 1.4.5.1). It leads to a spatial resolution in range of $r_d = c/2/f_d$. For technical reasons however, it is difficult to create a very short pulse (whose duration should be approximately 50 ns for a radar like ERS-1) that has the quantity of energy necessary to distinguish the targets from thermal noise, a quantity defined by the radar equation (Sect. 3.2). It is therefore preferable to create a much longer pulse, which contains a frequency bandwidth at least equal to f_d (the sampling frequency must remain slightly greater than this bandwidth, in order to satisfy the conditions of Shannon sampling, Sect. 1.4.7). This long pulse includes a phase modulation (Sect. 3.5.1.1), known as a ‘chirp’ which is very similar to that of the radar signal in azimuth, with the difference that it has a constant amplitude and that it is ‘artificial’, whereas the phase modulation is ‘natural’ since it is caused by the kinematics of the image acquisition.

⁴Fig. 2.4(a) illustrates this condition: without further processing, the pulse duration T^{pulse} creates uncertainty about the measurement of the duration of the round trip and consequently makes it pointless to have any sampling period finer than T^{pulse} .

More precisely, this gives a pulse phase distribution:

$$\Delta\phi_d(t) = \pi\rho t^2 \quad (2.57)$$

where ρ is a frequency rate expressed in $\text{Hz} \cdot \text{s}^{-1}$ (or in a similar manner in Hz^2). In other words, if B_d is the modulation bandwidth and T^{pulse} the duration of the long pulse, ρ is given by $\rho = B_d/T^{\text{pulse}}$.

As in azimuth processing, it is preferable to use dimensionless coefficients based on the number k of range samples, which are the values which will actually be manipulated. We shall call this dimensionless factor, which is characteristic of the quadratic phase distribution, the 'range compression rate' and note it N_d . Unlike N_a , which is always positive, N_d can be either positive or negative depending on the choice of the radar's designer. Starting from $N_d = f_d^2/\rho$, it follows that:

$$\Delta\phi_d(k) = \pi \cdot \frac{k^2}{N_d} \quad \text{and thus :} \quad \Delta\phi_d(f) = \pi \cdot N_d \cdot f_{rd}^2 \quad (2.58)$$

The quantity f is now the frequency variable of the FT applied to the radar signal in the range direction, and f_{rd} is its reduced frequency ($f_{rd} = f/f_d$). As in azimuth processing, the length of the pulse, expressed as a number of samples, is usually N_d . In the case of the ERS-1 satellite, N_d equals 704.

Range compression introduces an extra calculation stage in the processing of raw data but this disadvantage is now negligible as computers progress. The use of a long frequency-coded pulse has other benefits because it increases the convolution of the image. The filtering core, which is restricted to a few thousand points in azimuth processing (1300 for ERS-1), is more than a million (1500 times 704) with range compression. It also limits the risk of a powerful target saturating the raw data. A source would need to have more energy than the approximately one million neighboring targets, rather than just a few thousand (Sect. 2.3.9). Since it is more convoluted, the raw data is less sensitive to transmission losses or to interference from ground sources.

The ALMAZ satellite is a notable exception to the principle of range compression. The engineers succeeded in making it transmit 60-ns pulses with peak power of 270 kW. This technical *tour de force* meant that range compression was no longer necessary. Despite their quality, images from ALMAZ are much more difficult to re-sample because of the on-board pre-processing, which makes it difficult to exploit them through further processing, in particular for coherent applications such as interferometry.

2.3.8 Saturation effects

Radar data can only be recorded over a certain dynamic range. If a target is exceptionally powerful, it can saturate the receiver, meaning that it pushes the radar amplitude to the maximum of the ADC receiver. Beyond this maximum amplitude, the receiver will remain blocked at plus or minus that value. The powerful target will therefore be recorded as a 'bang-bang' phenomenon, with the detected signal swinging from the maximum positive value to the maximum negative value on the real and imaginary

parts according to the value of the phase. This recording is the equivalent of a recording on one bit. It reduces the quality of this powerful signal after processing, because the phase is only known to modulo $\pi/4$. For targets sharing the data with the saturating target however, this saturation is more dramatic. These targets have no influence on the raw data and therefore disappear from the final image. After processing, the saturating target seems to lie in the center of a black line in azimuth, if the pulses are not chirp-coded (an unusual case arising more frequently in airborne than in space borne radar). This phenomenon can be seen in the series of images shown in Fig. 3.28 of Chap. 3 (the dark slightly parabolic arc), and is discussed in Sect. 3.15.4. If the pulses are chirp-coded, the saturating target seems to lie at the center of a dark halo, because it 'extinguishes' all the targets that share its raw data in azimuth or range.

2.3.9 Interference effects

As we shall see in Sect. 3.2, a radar's link power budget is proportional to the fourth power of the observation range. If an active instrument on the ground transmits energy in the frequency band observed by the radar, this energy will be attenuated by only the square of the observation range. The energy situation therefore favors the interfering agent. This property is exploited by transponders, which are active instruments used in radar calibration. In energy terms, it is easy to jam imaging radar however. Space borne systems can be easily jammed if the spacecraft orbit is known because the jamming device can be pointed straight at it if the radar's frequency band is known. Then all the jamming energy can be transmitted into the frequency window used by the radar. The fact that jamming is relatively simple is a major problem for military radar observation systems. We shall not pursue this issue here, as is it outside the scope of this book, other than to say that partial technical solutions exist which are both complex and costly. We shall, however, look at a few cases where the interference is not malicious but the result of unintentional radio pollution from ground instruments.

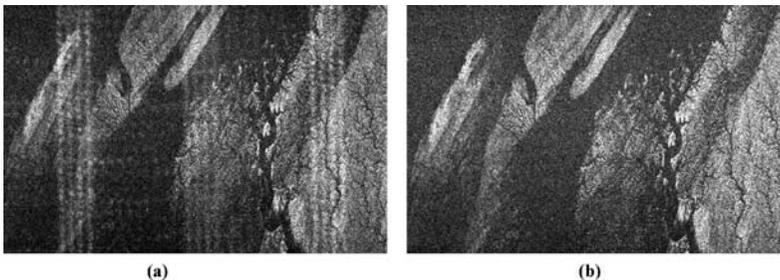


Fig. 2.13 (a) SIR-B image of Saudi Arabia corrupted by an interfering signal from the ground; (b) Restored SIR-B image.

A typical example of interference in raw data is illustrated in Fig. 2.13(a). The influence of the 'alien' signal can clearly be seen imprinted on this extract of an echo. The sample comes from the SIR-B system which samples in the real domain. As the polluting signal has a completely different shape from the signal expected by the radar

processing algorithm, the interfering energy has spread through the image in the form of characteristic bands or lines as a result of the processing. If we do not have enough knowledge of the interfering signal to be able to subtract it, we can simply eliminate those parts of the signal that contain the interference, which can be easily identified using the results of local statistical analysis (such as the value of the signal's standard deviation over a few dozen points). The parts that are eliminated can be replaced either by noise-like random values or by zeros (Fig. 2.13(b)). Since the number of samples influenced by the interference is usually very small in the raw data, their elimination has no influence on the quality of the final image, which is completely restored. As an example, some of the heavily corrupted images in the SIR-B data are fully restored after destruction of only about 0.2% of the raw data chosen on simple statistical criteria. Visual inspection showed that only 0.1% of the data were actually affected. It is preferable to destroy a few 'not guilty' data than to risk leaving corrupted data.

2.3.10 Motionless radar approximation

Throughout this book, we shall deal with radar as though it were stationary. This 'stop and go' hypothesis supposes that the radar instrument does not move between the moment when the pulses are transmitted, and when echoes are received by the instrument after being backscattered from the ground. In an airborne system traveling at $100 \text{ m}\cdot\text{s}^{-1}$ observing targets at a range of 20 km, the aircraft can only advance by 13 mm during the time necessary for the round trip by the wave traveling at the speed of light. This approximation is therefore fully justified because the movement of the aircraft is considerably less than the size of an azimuth pixel. In the case of a space borne system, the observation range can be as long as a thousand kilometers and the velocity of the satellite about $7 \text{ km}\cdot\text{s}^{-1}$. The return time is so great that the echo received just after the pulse transmission does not come from that pulse but from one that was transmitted several cycles earlier (Sect. 2.5.2). For the ERS satellite, observation from 850 km implies a round trip lasting 5.7 ms, whereas the time elapsed between each pulse transmission is about 0.6 ms. The echo of the first pulse transmitted is therefore received after the transmission of the tenth! Right after the radar is switched on, this characteristic can be used to measure noise before the return of the first echo.

Let us look at some of the consequences of this stationary radar hypothesis. Compared to the position of the satellite at the moment the pulse is transmitted, the zero Doppler position of the target (meaning its line of sight perpendicular to the velocity of the satellite or its closest point of approach) will occur further along the satellite's path. If we draw a very long triangle formed by the target C seen at the closest point of approach and the positions A and B of the satellite at the moments of transmission and reception for a given pulse, we can see that the position of the target corresponds to $AC = BC$ and that it can therefore be calculated in a plane that includes the middle I of segment AB and perpendicular to the velocity of the satellite at I.

To put it another way, the delay τ of a target located at a range R relative to the position A of the satellite at the moment of transmission is equal to $2R/c$. The number of pixels in azimuth crossed during this period is $2f_a R/c$. Unfortunately, the number is

not constant across the width of the image because it depends on R_1 . It is therefore not just a question of bias or shift of the origin of time but of image distortion. If we consider this effect as a bias at the first observation range R_1 , we will obtain a slanted image whose deviation at the last observation range R_2 will have become $2f_a(R_2 - R_1)/c$, measured in number of pixels in azimuth.

Now, the range ambiguity condition (Sect. 3.7.1) gives us a mathematical inequality: $f_a \leq c/2(R_2 - R_1)$ or, equivalently, $\chi/f_d \geq 2(R_2 - R_1)/c$. Distortion arising from the stationary radar hypothesis will therefore always be less than the value of one azimuth pixel. In the case of ERS, the useful listening time corresponds to 5000 range pixels, so the distortion is $5000/\chi$, or 0.44 pixels. This distortion is similar to that of an additional Doppler effect and can be corrected in the same way.

2.4 SAR Synthesis algorithms

In this section, we discuss the commonest processing techniques, whether high-quality or simplified, with an explanation of their architectural choices. Each processing technique accepts a different trade-off between quality of result and investment in calculation. The processes which aim for the highest quality make advanced applications such as interferometry possible because they preserve the signal's phase.

2.4.1 A common preliminary step, range compression

The first step in radar processing is usually range compression. The goal is to obtain an adequate range resolution using the frequency content of long coded pulses. The processing is straightforward as it is (1) one-dimensional, with the lines processed independently from each other, (2) sequential, because the lines are processed in their order of arrival and the computer only needs enough memory to store one line at a time (and the various products of intermediate calculations) and (3) stationary, because the range compression filter, also known as the 'range replica' has the shape of the transmitted pulse, and remains absolutely identical for all points on a line and for all the lines.

Given these conditions, the best processing technique uses the Fast Fourier Transform (Sect. 1.4.8). The signal of a line $s(t)$ is transformed, resulting in its Fourier transform $S(f)$, which is then multiplied either by the conjugate of the Fourier transform of the chirp, (the chirp is characterized by its phase distribution and a generally constant amplitude), or directly by the frequency expression of the chirp (Sect. 3.5.1).

It is worth noting that chirp responses are available in the telemetry data of certain radar satellites. It is possible to use them for range processing but as far as we know there is no advantage in doing so. Some satellites (such as ALMAZ) and airborne radar systems do not use range compression or perform it on board using analog systems (surface acoustic wave frequency mixers) or digital systems, while others (such as ERS) give a range of choices. In Sect. 2.3.8 we have seen the advantages of processing from range uncompressed raw data.

It is possible to determine the mean Doppler of the data before performing the range compression. It is then possible to predict the amplitude of the range migrations that the processing will need to take into account. The migrations depend on the range and thus vary across the image. They cannot be fully corrected during the range processing but their amplitude can be reduced. Indeed, the average value of the linear shift from one line to the next can be compensated for by a shift applied during the range processing. Let us look again at the J-ERS example in Sect. 2.3.6 and assume that the azimuth rate of the image N_a changes from 3,700 to 4,000 from the near range to the far range. Pursuing our previous example, when we process between the reduced frequencies 1 and 2, the value of the migration (range offset during the azimuth processing) at near range will increase from 24.77 to 99.1 cells, whereas this increase will be from 26.78 to 107 at far range. Scaled to the respective length of the azimuth responses, the linear part of the migrations will be 0.0200 of a range cell per azimuth cell crossed at near range and 0.0206 at far range. We can compensate for part of this shift during the range processing: in this example using the mean shift corresponding to the median range of the image, or 0.0203 range cells per azimuth cell crossed. In order to implement this more easily we choose a shift value that will produce an integer rapidly (in this example, $1/50$), so that the number of different shift values will be limited. The integer shifts are easy to perform during range compression as the fractional shifts are obtained by applying an extra phase ramp in the frequency domain. To achieve a shift of one range pixel after every m azimuth pixels, we will apply a phase distribution gradient:

$$2\pi\phi(k) = 2\pi \frac{k}{Nm} \quad (2.59)$$

where N is the length of the Fourier transform used, and k is the range sample number in the frequency domain (cf. Sect. 1.4.4). This gradient or its multiples can be applied repeatedly to obtain any given shift.

2.4.2 Time or ‘naive’ processing

The most natural kind of radar processing consists in adding the appropriate raw data on each final point of the desired image, applying phase correction as necessary. For each position of the satellite or aircraft at the moment of pulse transmission, we calculate the range and angle relative to the plane illuminated by the antenna. If the range is between the minimum and maximum range of the radar, and if the point is in the zone illuminated by the antenna, then the pulse under consideration participates in constructing the point. We can then resample the raw data locally to create a sample located exactly at the range of the point (after correcting for the migration). Then we compensate for the phase of the samples relative to that of the closest point of approach of the radar (by using the azimuth phase distribution). Aside from the tests that we have mentioned earlier (which are negligible in terms of computing time), each point on the ground usually requires that the group of operations (resampling and phase compensation) be performed N_a^2 times (N_a is the azimuth compression rate).

Although azimuth time processing is not very efficient, it does have certain advantages which have led us here at CNES to develop a ‘modern’ version of this technique, whose specific characteristic is that the points are created directly on a geographic grid with appropriate phase compensation. The final product is therefore a complex image corrected geometrically in both position and phase. The raw data can be entered into the application, one file giving the positions of the radar instrument for each line, a second file giving the attitude of the antenna for each line (the pointing file) and the geographic grid on which the image will be reconstructed. The advantages are:

- Easy interferometric combination (see [Chap. 4](#)) of the results produced by such processing is possible. Since the phase distributions specific to each orbital track have been compensated for in their corresponding images, the interferogram is simply produced by the difference of the phases between the two images. If the geographic grid includes elevation information, most topographical effects are also removed from the interferogram. It then contains only the possible effect of meteorological propagation, effects of displacement, and/or residual topographical information resulting from imperfect knowledge of the elevation in the geographic grid used. The number of two by two combinations of radar scenes increases rapidly with the number of scenes (190 combinations for 20 scenes for instance). It may therefore be more efficient, on sites with a large number of scenes, to devote more time to processing each scene in order to easily combine them later.
- All types of super-resolution, whether in azimuth or in range (Sect. 2.7), are also easily obtained by simply adding two or more images. It is only necessary to correct the phase of the added images by adding point-by-point the phase difference observed between each of them and the reference image.
- The special modes are processed with the same software as the standard mode, in particular for ‘Spotlight’ processing (Sect. 2.5.4.1) which consists in extending the phase distribution correction to the appropriate number of azimuth samples.
- Data from airborne instruments are processed in exactly the same way as satellite data because time processing does not require the stationary hypothesis assumed by the Fourier transform. The processing is capable of dealing with lines of data stored in random order provided we have correct information on the trajectory and pointing.

The disadvantages of the time domain architecture are its large requirement for processing power and memory. But considering the constant progress in the computing industry, we are confident that this architecture has a promising future.

2.4.3 Range-azimuth or ‘classic’ processing

Range-azimuth processing is the standard technique for computationally efficient radar processing. The first stage is range compression, which may involve the total or partial

elimination of the linear part of the range migration. Azimuth processing starts by transforming the data into frequencies in azimuth. Azimuth phase distribution is then applied in the frequency domain before performing range resampling in order to compensate for any remaining migrations. The resampling can be performed using spline functions or any other classic interpolation tool working on three or four successive points. These interpolators resample the complex data by working separately on the real and imaginary parts of the data.

Alternatively, an additional conversion to the range frequency domain and back, would allow resampling by ‘phase ramping’.

2.4.4 An alternative technique, the Ω_k processor

Observation of the frequency ‘map’ of the raw radar data suggests a different way of rearranging the data to facilitate processing [Prati, 1992]. We do not intend to go into the physical paradigm developed by the authors, which is rather complicated, and comes from processing seismic data. Briefly, this type of processing rescales the data in the azimuth frequency domain so as to obtain the same parabolic shape regardless of the range coordinates. Migration correction can then be performed in a single shifting operation because the parabolic shift value is the same for each azimuth frequency value.

2.4.5 Subtle distortion, chirp scaling

Another way of managing migration variation as a function of range consists in using a ‘twisted’ deviation principle by applying a preliminary phase distribution before transforming to range frequency coordinates [Raney, 1994]. Migration compensation that varies linearly from the minimum to the maximum range provides a perfect first order correction. To obtain this, the range data are first multiplied by a phase distribution having the form πak^2 in the time domain (where k is the sample number). As explained in Sections 1.4.4 and 2.3.1, this quadratic phase distribution has the effect of altering the frequency distribution of the data. When the data are translated into frequency in the range direction, this alteration causes a change of scale that makes it possible to correct the linear part of the variation of the parabolic migration as a function of range. Range compression is carried out in the same operation, but with a phase distribution whose coefficient must differ slightly from the nominal coefficient of the range phase distribution. Once back in the time domain, the data are corrected for a residual phase distribution resulting from the preceding operations.

Despite these extra steps, the scaling process has significant advantages as a variant of the range-Doppler algorithm, because it corrects first-order migrations perfectly, while it only involves multiplications by phase ramps and Fourier transforms, both of which are simple operations that are optimized for rapid processing in computer software libraries.

2.4.6 PRISME architecture, a multistage processing technique

The principle behind PRISME architecture [Massonnet, 1994], developed by CNES, is based on two observations:

- (1) In the azimuth frequency domain, the quality of the contributions to the final image depends heavily on their offset from the central frequency. Fig. 2.10 shows clearly that the data located at the center of the antenna radiation pattern has a significant signal to noise ratio (the noise, which we can assume to be white noise, makes an equal contribution throughout the spectrum). The data close to the center frequency are not very corrupted by mixing with ambiguous contributions (Sect. 3.7.4). However, the data located at the frequency transition, which corresponds to the minimum on Fig. 2.10, shows a signal to noise ratio that is much lower and which, most significantly, is made up of equal parts of useful and ambiguous signals. The compensation of parabolic migrations is specific to each of these contributions. This gives a slight advantage to the useful part by not fully focusing the ambiguous signal. However, the energy contribution of the latter is not reduced (Sect. 4.9).
- (2) Migration compensation does not require the highest level of azimuth sampling because, unlike the case with phase, azimuth migrations vary only slowly. The Q factor (cf. Sects. 2.2.3, and 2.3.6) indicates that we shift much more quickly from one phase cycle to the next than from one range pixel to the next. Since the migration and the phase compensation follow the same geometric law, an azimuth sampling much lower than Shannon's requirement is sufficient for migration correction;

These two observations encouraged us to adopt the following architecture (Fig. 2.14):

- The first step is to transform the raw data into the azimuth frequency domain using a much smaller number of azimuth points than the length of the azimuth radar response (for example, we would perform a Fourier transform on 128 points). We then choose the half of the azimuth spectrum containing the most interesting contribution, i.e. the half centered on the maximum of the spectrum (64 points in our example).
- We then transform the selected data into the range frequency domain, using range windows of a size compatible with maximum computational efficiency (for example, windows of 2048 points which each produce one quarter of the swath in the case of ERS-1.).
- We then perform range compression and migration compensation simultaneously. This involves applying the opposite of the range phase distribution expressed in the frequency domain (Eq. (2.58)), i.e., multiplying the range frequency response by the term $\exp(-j \pi N_d f_{rd}^2)$, modified by the appropriate phase ramp which will shift the response to the migration value (Eqs. (2.55) and (2.59)). This operation is possible because we are still in the azimuth frequency domain.

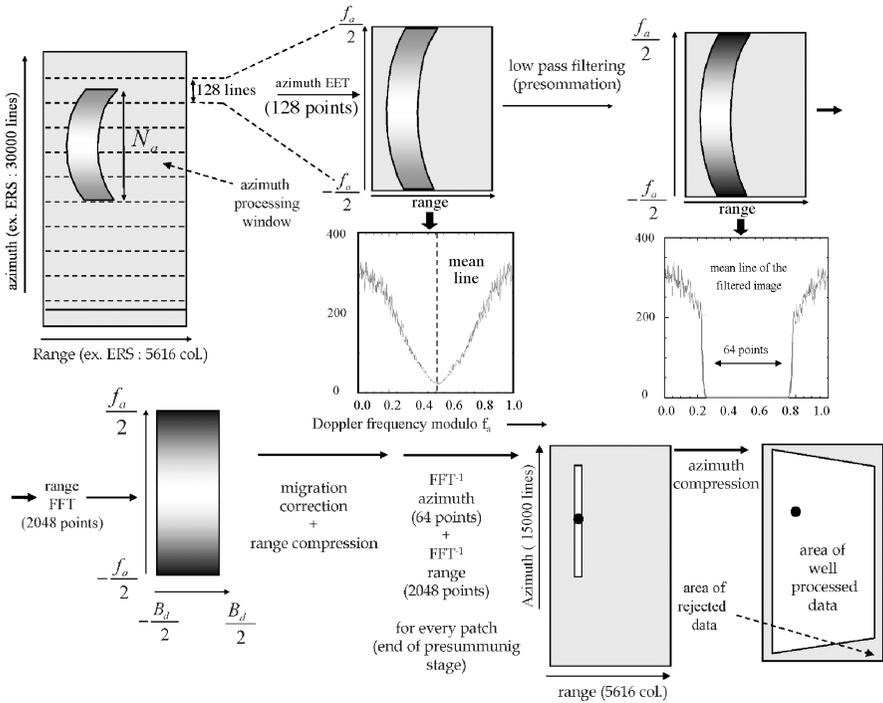


Fig. 2.14 General diagram of PRISME architecture.

The coarse azimuth sampling is not a problem because working on 128 azimuth samples rather than 2048 is the equivalent of giving the same range shift value to 16 neighboring samples in azimuth. Since this correction is stationary, the same migration correction is applied to the whole range being processed (a quarter of the swath, for instance, in our example).

- We then bring the data back into the time domain by inverse Fourier transform in range and in azimuth. We now only have half as many azimuth points as in the raw data. All these operations are called the pre-summing stage.
- We then still have to perform the azimuth compression. This operation is made easier because we have half the azimuth density which in turn, as a result of Eq. (2.18), divides the length of the azimuth processing window by four (340 instead of 1360 for ERS data).

This architecture can be adapted by modifying the parameters, such as retaining only a quarter of the azimuth spectrum and thus dividing by sixteen the azimuth compression rate to be applied. We can also apply a pre-summing stage for range. When PRISME was being designed, only Fourier transforms with lengths of an exact power of 2 were used. The use of transforms on different decompositions gives greater

flexibility for choosing the factor to be used in pre-summing stages, which can be for instance adjusted to one third or two thirds. PRISME architecture is used in the Diapason interferometry processing software developed at CNES.

This architecture is very economical in terms of processing power despite an additional conversion into the azimuth frequency domain and back. It is also economical in memory because of the reduced size of the memory block used for the pre-summing stage and the small size of the final azimuth response.

The disadvantages of this architecture are:

- The use of the same migration compensation value on an azimuth sector equal to $N_a/128$ (if we assume an azimuth buffer of 128 points).
- The use of the same migration compensation value for a range sector equal to a quarter of the swath (if we continue the case of our example), which amounts to accepting a maximum error corresponding to the range migration change on of the swath.
- The discarding of part of the useful data (Fig. 2.10).

2.4.7 Unfocused processing, a special case

As we have seen, full radar processing includes range compression, azimuth phase distribution compensation and migration compensation. In contrast, unfocused processing is a technique for very rapidly obtaining a radar image with lower quality. It consists in summing a batch of successive echoes for which the phase distribution of the echo being added is not too far from that of the central echo of the batch whatever the range. This addition may not include a phase correction since if the phase distribution differs from an ideal distribution by less than $\pi/2$, we can consider that an addition is still 'constructive' for the processing; the scalar product of two vectors separated by an angle belonging to $[-\pi/2, +\pi/2]$ is still positive. The result of this condition is:

$$\Delta\phi(t) \leq \frac{\pi}{2} \quad (2.60)$$

Or, using Eq. (2.42):

$$|k| \leq \sqrt{\frac{N_a}{2}} \quad (2.61)$$

2.4.8 A practical example of the unfocused processing technique

The unfocused processing technique described in the previous paragraph is not useful for processing real data as these are rarely centered on 'zero Doppler'. We will now describe an unfocused processing technique that can be used in all circumstances. We know that rapid algorithms allow calculation of mean Doppler (or Doppler centroid) by analyzing only a few echoes (Sect. 2.3.4). These algorithms can be used to establish which 'Doppler quarter' is closest to the mean Doppler,

i.e. whether the mean Doppler is closest to 0, to 0.25, to 0.5 or to -0.25 in terms of reduced frequency. Once this is done we process the data, summing the echoes four by four before any range processing. Let A, B, C, D be four points at the same range in four successive echoes. If the mean Doppler is close to 0, we simply form $A + B + C + D$. If we wish to filter in the neighborhood of 0.5, we form $A - B + C - D$. For a Doppler value close to 0.25 we form $A + jB - C - jD$ and finally, for -0.25 , we form $A - jB - C + jD$. These linear combinations apply phase ramps to the data in the time domain, so that their azimuth spectrum is translated while the summation downfilters them. In the preceding operations, there is no multiplication. Indeed, multiplying a complex point by j is achieved by simply interchanging the real part and the imaginary part, and changing the sign of the latter.

We can then apply the simple azimuth processing described in the previous paragraph to the filtered and zero-Doppler centered data, and the number of points to be summed, originally equal to $\sqrt{2N_a}$, is actually divided by four because of the preliminary grouping described above.

We only perform the range compression at the very end, so that it only needs to be applied to a reduced number of echoes. Performing the processing in this order is acceptable here because the simplified azimuth processing changes very little with range. It is therefore legitimate to apply it even though the data are still spread over the width of the range pulse. We will therefore only compress one range line every fifty lines (a typical value for ERS-1), with which we can create an image with reduced resolution: approximately 200 m in azimuth, and 20 m or more in range (depending on the amplitude of the migrations, which are not compensated for in this type of processing). In terms of computing time, it costs only one or two operations per point of raw data on average.

2.4.9 Another special case, deramping processing

Radar processing by deramping (also called *Specan* processing) is based on differential processing with respect to a reference point A . The raw data are then corrected according to the phase values of this reference point.

The time phase distribution (Sect. 2.2.5) corresponding to the target crossing the antenna beam is given by the Eq. (2.17), expressed in time, in along-track distance or in terms of the number of echoes counted from the point of closest approach. These phase distributions can also be expressed in the frequency domain (Eq. (2.38)).

As we have seen, Eq. (2.38) governs the synthesis when conducted in the frequency domain, which consists in (1) transforming the raw data into azimuth frequency using the Fast Fourier Transform, (2) applying the opposite of the initial phase distribution law, i.e. multiplying the frequency response by the term $\exp(-j\pi \cdot N_a f_r^2)$, which eliminates the data's phase coding and lastly (3) returning to the time domain using the inverse Fourier Transform.

What would happen if we processed the raw data by first correcting the phase in the time domain, using a particular target as a reference? Let us take a standard target A with phase distribution $\Delta\phi(f)$, whose point of closest approach gives the time

origin. Now let us take a second target B whose range at its point of closest approach is identical to that of the first target, and whose time at its point of closest approach we will call t_c . This time represents a shift along the track with respect to the original target that we can express as a number of azimuth pixels k_c : $k_c = t_c \cdot f_a$. Equation (2.17) shows that the phase distribution of target B is therefore:

$$\Delta\phi = 2\pi \frac{v^2(t + t_c)^2}{\lambda R_0} \quad (2.62)$$

If we assume that the antenna radiation pattern is infinite (meaning that we ignore any edge effect), we can easily predict the residual phase distribution of target B, if the phase of the raw data is corrected in the time domain according to the phase values of the reference target A:

$$\Delta\phi_r = 2\pi \frac{v^2(t + t_c)^2}{\lambda R_0} - 2\pi \frac{v^2 t^2}{\lambda R_0} = 2\pi \frac{v^2(2tt_c + t_c^2)}{\lambda R_0} \quad (2.63)$$

If we ignore the fixed phase term ($2\pi v^2 t_c^2 / (\lambda R_0)$), which behaves like a constant offset applied to target B's own phase (which we may nevertheless have to take it into account, especially when performing interferometry or polarimetry), we obtain the distribution:

$$\Delta\phi_r = 2\pi \frac{2v^2 t t_c}{\lambda R_0} \quad (2.64)$$

If we then apply a Fourier transform to the raw data corrected in this way, we will obtain for the second target a reduced frequency:

$$f_r = -\frac{1}{2\pi f_a} d(\Delta\phi_r)dt = -\frac{1}{2\pi f_a} \frac{d(2\pi 2v^2 t t_c / (\lambda R_0))}{dt} = -\frac{2v^2 t_c}{\lambda R_0 f_a} \quad (2.65)$$

This reduced frequency is actually constant, and can be expressed as:

$$f_r = -\frac{2v^2 t_c}{\lambda R_0 f_a} = -\frac{f_a t_c}{N_a} = -\frac{k_c}{N_a} \quad (2.66)$$

Since target B's contribution 'sits' on a clearly defined frequency sample, as would any other target, the processing is completed!

The reduced frequency f_r is fractional. The Fourier transform is generally performed on N samples, with $N \approx N_a$. The samples from the Fourier transform have the same order and spacing as in the time domain. What are the advantages and disadvantages of deramping?

The main advantage is the shorter calculating time, which is only about half as long because there is only one Fourier transform. However, there are several disadvantages. To start with, the domains of targets A and B are not exactly the same, because some samples of raw data contain contributions from target A and not from target B and vice-versa. Since the shift between points A and B causes B's azimuth response to be

recorded incompletely, the shift must be small with respect to N_a . We end up with a much smaller number of points after this operation than after standard filtering by Fourier transform, where the only points rejected correspond to the length of the filter. This disadvantage essentially cancels out the gain in computing time that the method provides. In addition, the method is non-stationary by nature. Image quality therefore varies as a function of the azimuth distance between a given point in the image and its ‘deramping reference point’. The corresponding variation in image quality is called scalloping. Lastly, since the deramping process does not require expressing the data in the frequency domain, the operations usually carried out in this domain (calculating mean Doppler, correcting migrations, etc.) must be either abandoned or performed with other methods.

Why should deramping therefore still be considered when computation costs are no longer an issue and we have alternative methods? Because in the Scansar mode (Sect. 2.5.4.2), which is based on collecting discontinuous batches of echoes, the disadvantages of deramping are inherent: non-stationary image quality (scalloping) and partial impulse responses.

2.4.10 A radar processing technique without approximations

Just as nothing prevents us from using the exact form of the range distribution between the radar and the target, rather than stopping at second order, we can produce a frequency distribution without approximations. The exact form of the range variation is (Fig. 2.7):

$$2 \cdot (R(t) - R_0) = 2 \cdot (R_0^2 + k^2 p_a^2 - R_0) \quad (2.67)$$

which gives the exact phase distribution:

$$\Delta\phi(k) = \frac{4\pi}{\lambda} \cdot (\sqrt{R_0^2 + k^2 p_a^2} - R_0) \quad (2.68)$$

that, when expressed as frequency, gives:

$$f(k) = -\frac{1}{2\pi} \frac{d}{dt} \{\Delta\phi(k)\} = -\frac{1}{2\pi} \frac{d}{dk} \{\Delta\phi(k)\} \frac{dk}{dt} \quad (2.69)$$

We find that the quantity dk/dt is f_a , the pulse repetition frequency. Hence:

$$f(k) = -f_a \frac{d}{dk} \left\{ \frac{2}{\lambda} \cdot (\sqrt{R_0^2 + k^2 p_a^2} - R_0) \right\} = -\frac{2f_a}{\lambda} \cdot \frac{k p_a^2}{\sqrt{R_0^2 + k^2 p_a^2}} \quad (2.70)$$

With this formula we can find k corresponding to the frequency $f(k)$ by:

$$\lambda^2 R_0^2 f^2(k) + \lambda^2 k^2 p_a^2 f^2(k) = 4 f_a^2 k^2 p_a^4 \quad (2.71)$$

Or:

$$k^2 = \frac{\lambda^2 R_0^2 f^2(k)}{4 f_a^2 p_a^4 - \lambda^2 p_a^2 f^2(k)} \quad (2.72)$$

We can therefore compensate for the phase distribution $\Delta\phi_a(k) = \pi \cdot k^2/N_a$ in the frequency domain for any frequency $f(k)$ by substituting the corresponding frequency for k^2 . Finally, in terms of reduced frequency we have to compensate for the distribution:

$$\Delta\phi_a(f_r) = \frac{\pi\lambda^2 R_0^2 f_r^2(k)}{[4p_a^4 - \lambda^2 p_a^2 f_r^2(k)] \cdot N_a} \quad (2.73)$$

Note: The quantity $\lambda^2 p_a^2 f_r^2(k)$ appears as the correcting factor compared $4p_a^4$ to in the equations laid out in Sect. 2.2.5, where the approximation $2 \cdot (R(t) - R_0) \approx k^2 p_a^2 / R_0$ is made. This is fully justified for ERS-1, but may not be valid in other contexts (for example in the case of very high resolution imaging or for acquisition modes using a large off-pointing angle).

2.5 System constraints

2.5.1 Radar system design

Radar system design must respect strict limitations which taken together leave very little freedom to the designer. These constraints include legal issues concerning the choice of authorized frequencies and the corresponding bandwidths, geometry issues linked to the physical dimensions of the antennas, dynamic issues linked to the laws of mechanics for orbiting bodies (in the case of satellites) and power issues which govern the strength of the signal relative to the different sources of noise.

The legal issues are decided at international conferences which fix the wavelength authorized for radar and their bandwidth as well as the frequencies authorized for transmitting the data to the ground. These authorizations also cover power flux limits. These legal decisions tend to change as new techniques are developed, or as different groups of users gain or lose influence. At the time of writing, synthetic aperture radars can be designed in the *Ku* band on 500 MHz around 13.5 GHz, in the *X* band on 300 MHz around 9.65 GHz (a possibility of extension to over 500 MHz is currently under discussion), in the *C* band on 320 MHz around 5.41 GHz, in the *S* band on 200 MHz around 3.2 GHz (this frequency has the limitation that a narrow bandwidth must be left untouched in the middle of the permitted band because it is very useful in radio astronomy) and lastly in the *L* band on 85 MHz around 1.26 GHz. There is also the possibility of using in the *P* band a very small 6 MHz bandwidth around 435 MHz. The bandwidths allocated to each of the frequencies govern the maximum range resolution that they can attain in range (cf. Sects. 2.3.7, and 3.12.1).

2.5.2 Timing constraints

Optical systems usually possess an array or matrix of detectors for simultaneous acquisition of data from several points on the ground. Radar systems however, with some exceptions, are intrinsically one-dimensional. They can therefore only listen to

the specific stream of echoes received by their antenna. This property is the cause of considerable geometric constraints. The resulting chronology of events linked to the pulses is called the chronogram.

In the first place, a radar cannot listen while it is transmitting, in order to preserve its highly sensitive amplifier, designed to detect very faint signals which cannot be exposed to the kilowatts needed for the transmission. Furthermore, the radar cannot listen to the echoes of two pulses simultaneously. When it is listening to a given echo, it is therefore necessary that the echoes from preceding or following pulses concern points which are outside of the antenna's ground footprint (the swath width), in the range direction. In some cases the following pulses may all be in free space where they produce no echoes (Sect. 3.7.1).

We have here a big difference between airborne and space-borne radar. Airborne radar, even long range (for example 50 km), receives the most distant response of a pulse only 0.33 ms after transmitting it. Even if the velocity of the aircraft is 300 ms^{-1} , it only travels 10 cm during the return trip. There is then no rush about transmitting the following pulse after completing reception of the first. The only requirement is that the antenna must be at least 20-cm long.

A satellite 750 km from its target has to wait 5 ms for the return echo, during which it travels 30 m. As a radar satellite can certainly not carry a 60-m antenna it transmits several pulses before receiving the first. On the ERS satellite, just after the radar is switched on, the spare time before receiving the return of the first pulse is exploited for measuring the noise, because the lines are recorded in each listening period but cannot yet contain echoes.

For a satellite the antenna range radiation pattern enables us to distinguish between the contributions of successive pulses, as is done for targets with azimuth ambiguity. We then speak of targets with range ambiguity.

The range difference of targets observed at the same moment by two successive pulses is, for the round trip:

$$\Delta R = \frac{c}{2f_a} \quad (2.74)$$

Hence, the number of range pixels (of size $c/2f_a$) separating the real image from the ambiguous image is equal to the ratio between the range sampling frequency and the azimuth sampling frequency f_a . This is the χ factor defined in Sect. 2.2.3.

For any target observed at range R , the radar listens to the echoes of the other pulses located at ranges $R + n\Delta R$, where n is an integer. So that these ambiguous targets make only a negligible contribution to the signal, the antenna radiation pattern must illuminate only the ground footprint corresponding to a range shift no greater than ΔR . This is the range ambiguity condition (Sect. 3.7.1). Expressed as ground range and supposing that the mean incidence angle is θ , this condition becomes:

$$\Delta R_s = \frac{c}{2f_a \sin \theta} \quad (2.75)$$

This condition about the width of the zone illuminated on the ground (the swath) goes hand in hand with a phasing condition. Care must be taken that the echo does not

arrive back at the radar during a transmission period. We could for example require that the central observation range R corresponds to a time equal to a integer number of pulse cycles plus a half-cycle, or:

$$R = c \cdot \frac{n + 1/2}{2f_a} \quad (2.76)$$

where n is an integer number and c the speed of light. This condition may require f_a to be changed slightly.

The chronogram is usually designed to exclude the echo from the nadir point (vertical echo) from the listening window. Although the nadir is generally located outside the main lobe of the antenna, its echo is very powerful because the radar wave reaches the ground perpendicularly and is mirrored back to the radar (Sect. 3.7.1). This can more than compensate for the weak antenna gain in the vertical direction. If H is the radar's altitude (Fig. 2.7), the vertical echo appears in the chronogram at time $2H/c - k/f_a$, where k is the largest integer such that the difference is positive. We might, for example, decide to force the vertical echo to return at the moment when a pulse is being transmitted: $f_a = kc/2H$. Alternatively, if this solution is not compatible with the other constraints, we could design the antenna gain so that it is very small in the vertical direction (creating a 'hole' in the antenna pattern) while also adjusting f_a so that the nadir echo does not merge with the useful echo.

2.5.3 The different types of radar data

There are two main types of numerical representation of radar data, complex or real, the latter being oversampled by a factor of 2. They correspond to two slightly different ways of handling the backscattered signal. This signal has a frequency bandwidth B_d that can be no larger than the range sampling frequency f_d and is centered on the carrier frequency f_c . In the first case, the signal is mixed with the carrier, the component at twice the carrier frequency is discarded, and the remaining baseband signal is sampled in phase and in quadrature. This gives a series of complex samples whose real part corresponds to the in-phase sampling while the imaginary part corresponds to the in-quadrature sampling. Most radar systems use this scheme. For the case of real types, however, the backscattered signal is mixed with the carrier after its frequency is offset by one quarter of the frequency bandwidth. This single channel is then sampled, but at twice the rate (i.e., $2f_d$ whereas complex sampling is performed at f_d , Sect. 1.4.7). For the processing, a real-to-complex Fourier Transform is used to return to the complex domain. There is therefore very little difference for the user. Concerning system design, the choice is between one sampler running at twice the speed or two samplers. Different calibration precautions may also be required.

2.5.4 Tricks for cheating nature

Radar system designers are never short of a trick or two and have conceived several solutions for getting around the harsh limitations dictated by geometry, physics and

orbital dynamics. These solutions are very tempting as long as they represent a good trade-off between the added capacities and the added complexities.

2.5.4.1 ‘Spotlight’ mode

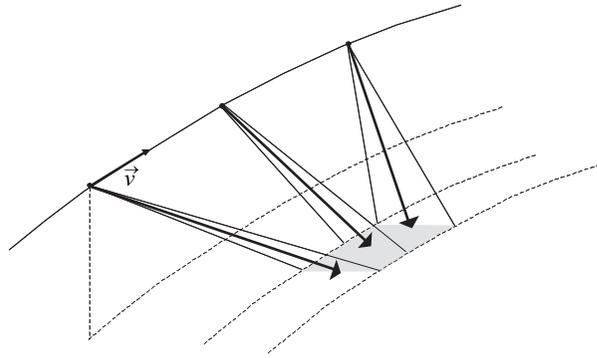
To improve azimuth resolution it is necessary to reduce the length of the antenna (resulting in a corresponding increase in the azimuth antenna crossing time and therefore Doppler bandwidth) and to increase the frequency f_a proportionately (in order to satisfy the Shannon conditions for proper sampling). But this would force us to reduce the swath to avoid range ambiguities, and therefore to increase the height of the antenna. In order to conserve the entire swath while improving azimuth resolution, we could follow the target with the antenna for a longer time than that needed for crossing the azimuth lobe of a fixed antenna. For this we can either move the antenna so that we aim at the target well before its point of closest approach and long after we have crossed it, or get the same result with electronic pointing using programmed phase shifters on an active antenna (Sect. 1.2.6.1). This procedure is commonly known as ‘Spotlight’ mode (Fig. 2.15(a)), recalling the way an actor can be followed across the stage by a theatre spotlight.

The signal budget for a given target in Spotlight mode has a significantly larger observed Doppler bandwidth than that resulting from simply crossing the antenna’s lobe. The overall Doppler bandwidth is approximately proportional to the time during which the target is observed, because of the time/frequency proportionality (Sect. 2.3.1). However, this overall Doppler bandwidth remains compatible with the sampling frequency f_a because the instantaneous observed Doppler frequency bandwidth is still tied to the antenna’s lobe which leads to a nearly constant instantaneous frequency bandwidth. Processing such data requires either the use of ‘time-domain’ architecture which will naturally take into account the way the signal changes, or a preliminary stage of ‘spectral rearrangement’ (Fig. 2.15b). After this stage, the data are equivalent to the data which would have been obtained with a frequency higher than f_a (in this case a magnitude close to $3f_a$). During spectral rearrangement, each sub-spectrum is nested inside a ‘larger’ one, around its own mean Doppler, f_{m1} when approaching the target, 0 at the closest point of approach and f_{md} when leaving the target (left to right in the diagram). A slight attenuation of the signal may result from this off-pointing, due to the less than optimal antenna gain in these observation conditions.

The disadvantage of ‘Spotlight’ mode is that it does not allow continuous and/or stationary data acquisition because the antenna footprint is delayed on the ground relative to the satellite. It is necessary to stop the observation in order to point the antenna ahead of the satellite. Thus, long stretches of continuous data can no longer be acquired.

Far from ‘breaking the mold’ of satellite radar design, Spotlight mode involves trading one resource, the orbital arc that is used, against another, the available azimuth resolution for a given antenna geometry and swath. It therefore grants increased flexibility. This mode also has consequences concerning image quality, since ambiguous azimuth contributions are strongly affected and dispersed by the spectral rearrangement (Fig. 2.15(b)). They are less focused but lose none of their overall energy. The same

(a)



(b)

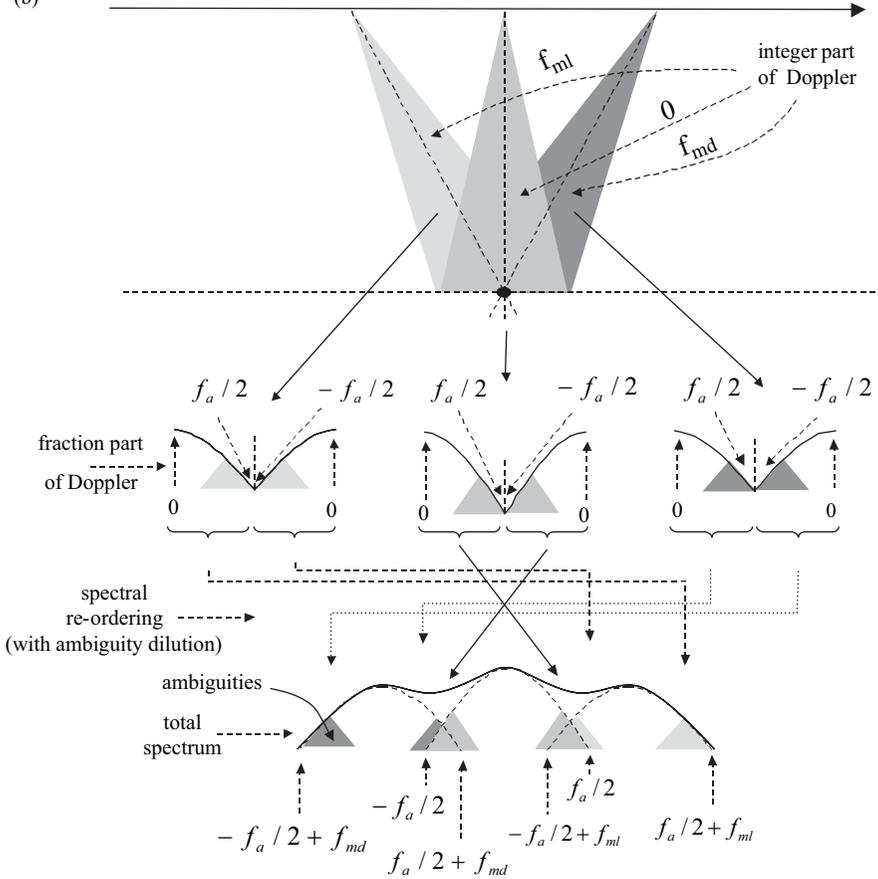


Fig. 2.15 (a) Principles of Spotlight mode; (b) Spotlight mode spectral recording.

is true of ambiguous range contributions, because of the increased azimuth compression rate N_a , which more efficiently defocuses points that are not at the appropriate range.

2.5.4.2 ‘Scansar’ mode

In order to increase the swath width, it would be necessary to increase the antenna length so as to decrease the frequency f_a proportionally. This illustrates a trade-off between azimuth resolution and swath width. However, most designers want to keep the possibility of high resolution which requires a relatively small antenna length. The solution is to use ‘Scansar’ mode, which was put into operation for the first time on the Canadian satellite RADARSAT, launched in 1995.

This mode consists in processing several sub-swaths simultaneously (Fig. 2.16(a)), which is achieved by sending successive ‘bursts’ of pulses to each of them, making sure that a complete sequence of sub-swath illumination does not exceed the length N_a of the azimuth processing window. For example, for a four fold ‘Scansar’ (i.e. treating four sub-swaths) with a mean value of 2000 for the azimuth compression rate, a burst of 500 pulses will be sent to each of these sub-swaths before restarting the cycle. Since each sub-swath is revisited after an interval which is less than the time taken to cross the azimuth antenna lobe, each target contributes to the signal and can therefore be reconstructed through processing. On the other hand, each target in a sub-swath will only be seen under a Doppler bandwidth equal to one quarter of the nominal Doppler bandwidth. This means that the azimuth resolution will be reduced proportionally. In addition, depending on its phasing with respect to the bursts, the Doppler frequencies seen by the target will comprise only a portion of the Doppler bandwidth. For example, a target whose point of closest approach occurs when a burst of pulses is directed at its sub-swath will see the quarter of the Doppler spectrum around zero Doppler. A target located between two bursts will see a frequency bandwidth split between 0.5 and -0.5 reduced Doppler. This difference in configuration causes ‘scalping’ which can produce bands of variable radiometry on the image.

There are several basic differences between ‘Spotlight’ and ‘Scansar’ modes even though both share the principle of a trade-off between resolution and coverage. ‘Scansar’ mode has already shown its usefulness in operation. Fig. 2.16(b) shows a SCANSAR image acquired by RADARSAT over the Mediterranean. The swath obtained is close to 500 km (compared to 100 km for ERS). Extending the use of SCANSAR to the field of interferometry is possible, but it requires special precautions, such as making sure that the batches of pulses are phased the same way in the two images. ‘Spotlight’ mode has only been tested on a small scale with a radar carried on board the Shuttle (during the SIR-C mission in 1994). More recently, ‘Spotlight’ images have been delivered by TERRASAR-X and Cosmo - Skymed missions. As far as the technology is concerned, ‘Scansar’ mode requires very rapid commutation of the antenna in elevation which can only be obtained electronically. ‘Spotlight’ mode only requires an attitude sweep that can be performed by a traditional antenna. This difference may become less significant if the use of electronic antennas expands. Finally, for image quality, ‘Scansar’ mode, like the deramping method, is subject to scalping.

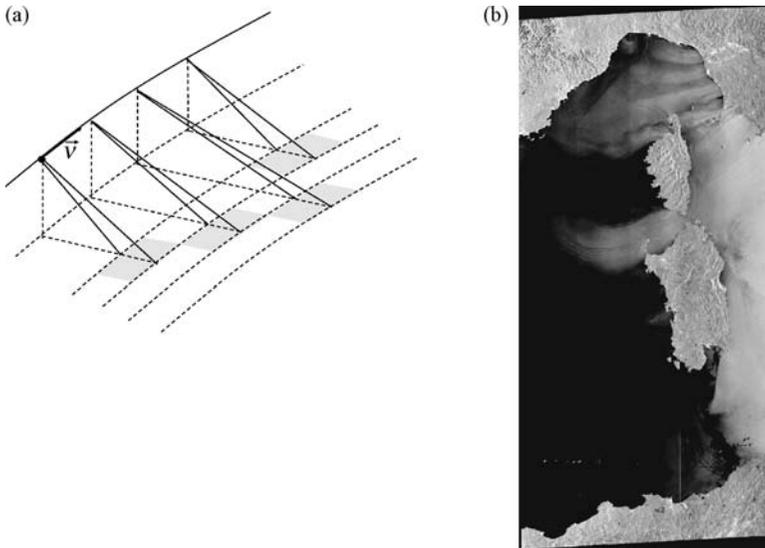


Fig. 2.16 (a) Principles of Scansar mode (with 3 subswaths); (b) Scansar/Radarsat image of Corsica-Sardinia (500 km wide).

2.5.4.3 Other ideas

It is not our intention to discuss in detail all of the ideas that have been suggested for facilitating the design of radar systems. However, one promising idea is to use an antenna with rapid-pointing capability to follow the backscattered signal in the swath. More precisely, a sufficiently wide antenna illuminates the swath without working at maximum resolution during the transmission of each pulse. The angular position of the echo as a function of the return time can be known approximately, ignoring topography-induced variations. This angular position can be followed with full resolution using permanent and rapid repointing of the antenna. Such a system would certainly be considerably more complex but would have the potential of higher gain and much better protection against range ambiguities.

We could also alternate the sign of the range compression rate N_d for each pulse. Doing this would cause the pulse preceding or immediately following the processed pulse to be 'spread'. The first order range ambiguities would therefore be totally unfocussed (but their energy would remain in the data in the form of 'noise'). The second order ambiguities would be processed normally but would be much weaker. In certain favorable cases, the nadir echo would also be spread, depending on its ambiguity rank.

Lastly, if we had heavy digital computation on the ground, and large telemetry capabilities, we can imagine recording the contributions of different parts of the antenna separately, bypassing the 'natural' analog combination performed by the antenna. During ground processing, these different contributions can be reorganized into beams

pointing at different places, which would break up the ‘one-dimensional’ aspect of the radar signal that we underlined in Sect. 2.5.2. This ‘computed synthetic beam antenna’ would significantly increase the quantity of recorded data, however, and complicate on-ground processing. TERRASAR-X, launched in June 2007, offers some of these capacities as it is able to record independently the data on two sub-apertures.

2.6 Geometry characteristics

2.6.1 A practical example of the effect of range on images

One of the most obvious characteristics of radar imaging is the distortion caused by range sampling which we have already mentioned in Sect. 2.2.2. In order to explain this more clearly, we shall evaluate an example in quantitative terms in Fig. 2.17(a) and the blown-up extract in Fig. 2.17(b).

This image shows the amplitude of a radar image acquired over Paris by the Russian ALMAZ system, built by the NPO Mashinostroyeniye Company. A study of this system carried out by CNES in 1991 [3] was fascinating for several reasons. ALMAZ worked at an unusual wavelength (*S* band, $\lambda = 9.6$ cm) and the image of Paris was taken from an unusually high incidence angle of 50° on 6 June, 1991. A satellite like Radarsat is now capable of doing this, but at that time data available from SEASAT as well as data expected from ERS-1 featured a low incidence angle of 23° .

The radar observation system was also very original, because it used no range compression, which required it to transmit an ultra-short pulse ($T^{pulse} \approx 40$ ns) with an extremely high peak power of 270 kW (as compared to 4.8 kW for ERS-1)! Another technical innovation was that the instrument was carried on board a pressurized orbital module with a mass of some twenty tonnes, designed to be used as an element of a space station. Since the instrument was designed for operating in pressurized atmospheric conditions, it was very different from its Western equivalents. Lastly, the platform was directly servo-controlled to ‘zero Doppler’ using the data acquired by the radar, an original attitude control system. This radar system, which fell back into the atmosphere in 1992, has unfortunately had no successor.

We can take advantage of this nice image of Paris to analyze the radar signal of the Eiffel Tower. The Tower is made of metal positioned at various angles. It is clearly visible as a small bright triangle, sharply contrasted, in the middle of the Champ de Mars, which is darker (Fig. 2.17(b)). It is worth recalling that in an optical image the tower’s brown color gives much lower contrast. The tower is almost vertical and is subject to the layover effect described in diagrams given in Figs. 2.18(b) and 2.19. Layover is the extreme case of ‘foreshortening’, which can be observed on slopes facing towards the radar (Fig. 2.18(a)). The radar sees the summit of the tower before its base and the tower seems to lean towards the radar. This effect can reveal the direction from which the radar is illuminating the target.

The size of a range pixel is equal to $p_d = c/2/f_d$ (Sect. 2.2.3) where f_d is the range sampling frequency, which is 28.8 MHz for ALMAZ. The range pixel length is therefore 5.2 m. Since we know the scale of the reproduction of Fig. 2.17(b), we

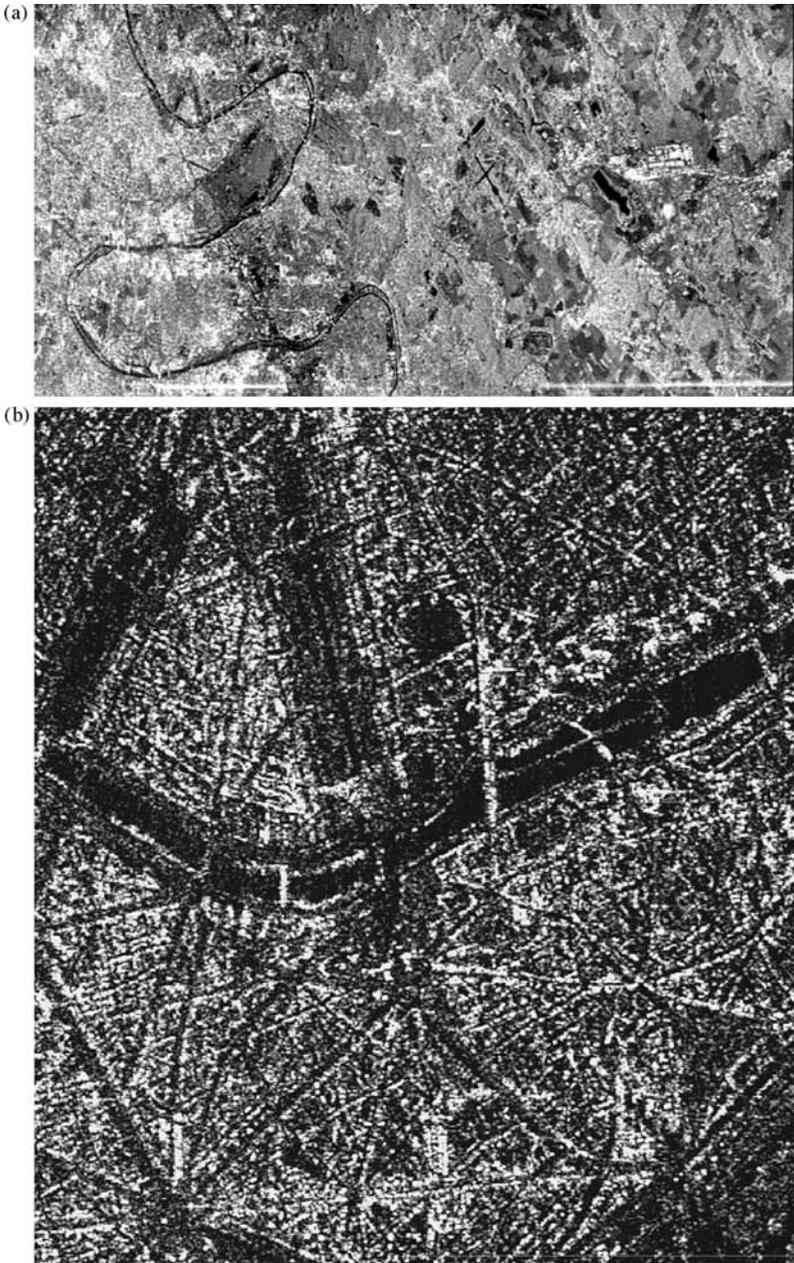


Fig. 2.17 (a) ALMAZ image of Paris (overall view) (b) Image of Paris acquired by ALMAZ (detail, Eiffel Tower area).

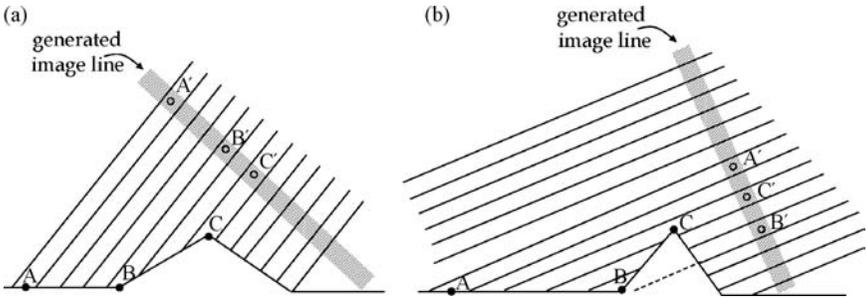


Fig. 2.18 (a) Foreshortening effect; (b) Layover effect.

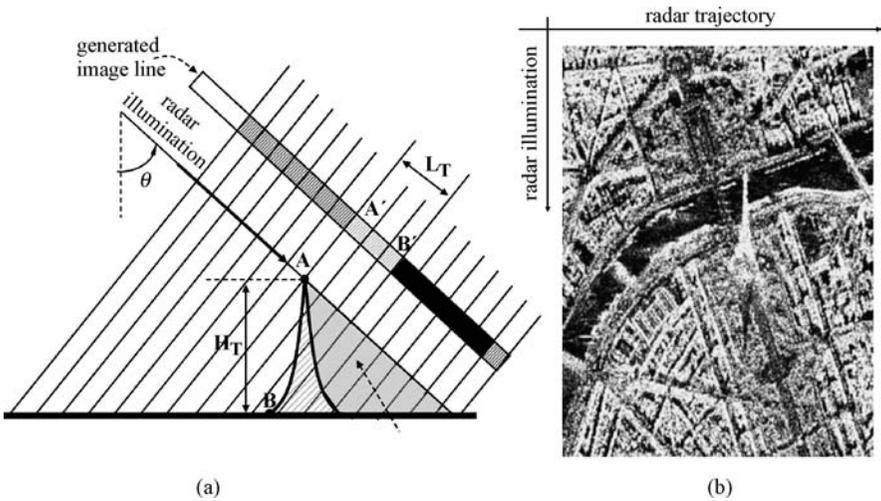


Fig. 2.19 Layover effect. (a) Modeling the phenomenon on a ‘theoretical tower’; (b) Layover effect on the Eiffel Tower. Airborne SAR ‘Sethi’ image, C band (resolution: 3 m).

can deduce that the length of the tower is 45 pixels or $L_T = 234$ m. Considering the diagram on Fig. 2.19(a) and taking into account the incidence angle of $\theta = 48^\circ$, we find a height of $H_T = L_T / \cos \theta$ or 350 m, fairly close to the actual value of 312 m.

Finally, the image of the same area observed by the airborne ONERA/CNES/Sethi radar, with a resolution of 3 m (Figs. 2.19(b) and 2.20), helps show Paris in a more familiar way, and allows us to conduct similar calculations. Apart from the resolution difference we notice differences in image quality (cf. Sects. 3.13, and 3.14.) to the advantage of the airborne image.

2.6.2 Equations for geometric positioning

In order to place a point \vec{P} of a radar image correctly on the surface of the Earth, three equations are used:

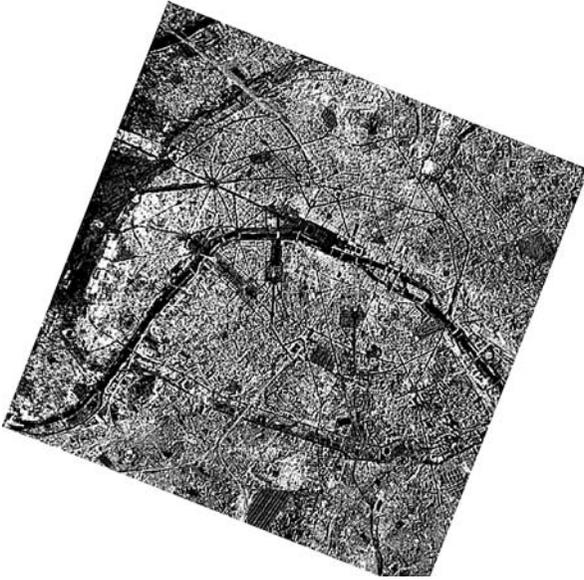


Fig. 2.20 ONERA/CNES Sethi image, Paris (resolution: 3 m).

- The range equation which indicates that the range between point \vec{P} and position \vec{S} of the satellite is known and equal to R :

$$|\vec{P} - \vec{S}| = R \quad (2.77)$$

- The Doppler equation which indicates that point \vec{P} is seen with the Doppler centroid f_m . If \vec{v} is the velocity of the satellite and λ is the wavelength used, then this can be expressed as

$$f_m = 2 \frac{(\vec{S} - \vec{P}) \cdot \vec{v}}{\lambda R} \quad (2.78)$$

- The equation for the shape of the Earth giving the norm of \vec{P} as the local radius R_E , modified if necessary by the local elevation:

$$|\vec{P}| = R_E \quad (2.79)$$

The geometric interpretation of the combination of these equations is the intersection of a sphere of radius R centered on the satellite, a hyperboloid and another sphere centered on the Earth (or the center of the osculating sphere used to describe the surface of the Earth). The intersection of the two spheres, when it exists, usually gives a circle. The intersection of this circle with the hyperboloid generally gives two points of intersection located on either side of the satellite

ground track, i.e. the projection (over time) onto the surface of the Earth of the vector \vec{S} (see Fig. 3.26 for clarification).

The choice between these two solution points therefore depends on whether the satellite is right-looking or left-looking. Certain satellites such as SEASAT or ERS-1 and ERS-2 were always side-looking (right) in the same direction. The need for maximum flexibility will lead increasingly to radar satellites offering the possibility of looking on either one side or the other (as is the case with RADARSAT-2). To indicate the pointing direction, we introduce the scalar value ϵ which is 1 if the projection of vector \vec{P} onto the vector $\vec{S} \times \vec{v}$ is positive, meaning it is left-looking. The scalar ϵ will equal -1 if it is right-looking.

Once this has been established, we can express the vector \vec{P} as a combination of vectors \vec{S} , \vec{v} and $\vec{S} \times \vec{v}$. These vectors are always linearly independent (a satellite is never traveling straight down!), even if \vec{S} and \vec{v} are not usually strictly orthogonal. We therefore have:

$$\vec{P} = \alpha\vec{S} + \beta\vec{v} + \gamma\vec{S} \times \vec{v} \quad (2.80)$$

In order to find the vector \vec{P} we only need to calculate the coefficients (α , β and γ). By substitution into the range equation (2.77), we have:

$$|(\alpha - 1)\vec{S} + \beta\vec{v} + \gamma\vec{S} \times \vec{v}| = R \quad (2.81)$$

or:

$$(\alpha - 1)^2 S^2 + \beta^2 v^2 + \gamma^2 (S^2 v^2 - (\vec{S} \cdot \vec{v})^2) + 2(\alpha - 1)\beta\vec{S} \cdot \vec{v} = R^2 \quad (2.82)$$

By substitution into the range equation (2.79), we have:

$$|\alpha\vec{S} + \beta\vec{v} + \gamma\vec{S} \times \vec{v}| = R_E \quad (2.83)$$

or:

$$\alpha^2 S^2 + \beta^2 v^2 + \gamma^2 (S^2 v^2 - (\vec{S} \cdot \vec{v})^2) + 2\alpha\beta\vec{S} \cdot \vec{v} = R_E^2 \quad (2.84)$$

We can subtract Eq. (2.82) from Eq. (2.84) to obtain a simple result:

$$\alpha S^2 + \beta\vec{S} \cdot \vec{v} = \frac{R_E^2 - R^2 + S^2}{2} \quad (2.85)$$

The Doppler equation (2.78) gives the following result:

$$-\lambda R f_m = (\vec{P} - \vec{S}) \cdot \vec{v} \quad (2.86)$$

or:

$$-\lambda R f_m = ((\alpha - 1)\vec{S} + \beta\vec{v} + \gamma\vec{S} \times \vec{v}) \cdot \vec{v},$$

which is equivalent to:

$$\alpha \vec{S} \cdot \vec{v} + \beta v^2 = \vec{S} \cdot \vec{v} = -\lambda R f_m \quad (2.87)$$

Equations (2.85) and (2.87) can be used to resolve the system into α and β :

$$\alpha(S^2 v^2 - (\vec{S} \cdot \vec{v})^2) = \frac{R_E^2 - R^2 + S^2}{2} v^2 - \vec{S} \cdot \vec{v} (\vec{S} \cdot \vec{v} - \lambda R f_m) \quad (2.88)$$

$$-\beta(S^2 v^2 - (\vec{S} \cdot \vec{v})^2) = \frac{R_E^2 - R^2 + S^2}{2} \vec{S} \cdot \vec{v} - S^2 (\vec{S} \cdot \vec{v} - \lambda R f_m) \quad (2.89)$$

We can therefore find the value of γ by substitution, using Eq. (2.84), which can be rewritten:

$$\alpha(\alpha S^2 + \beta \vec{S} \cdot \vec{v}) + \beta(\alpha \vec{S} \cdot \vec{v} + \beta v^2) + \gamma^2(S^2 v^2 - (\vec{S} \cdot \vec{v})^2) = R_E^2 \quad (2.90)$$

Given equations (2.85) and (2.87), we get:

$$\alpha \left(\frac{R_E^2 - R^2 + S^2}{2} \right) + \beta (\vec{S} \cdot \vec{v} - \lambda R f_m) + \gamma^2(S^2 v^2 - (\vec{S} \cdot \vec{v})^2) = R_E^2 \quad (2.91)$$

Since $A = S^2 v^2 - (\vec{S} \cdot \vec{v})^2$, $B = (R_E^2 - R^2 + S^2)/2$ and $C = \vec{S} \cdot \vec{v} - \lambda R f_m$, Eq. (2.91) becomes after multiplying both sides by A :

$$\alpha AB + \beta AC + \gamma^2 A^2 = AR_E^2 \quad (2.92)$$

Incorporating conditions (2.88) and (2.89), the result is:

$$(Bv^2 - \vec{S} \cdot \vec{v} C)B - (B\vec{S} \cdot \vec{v} - S^2 C)C + \gamma^2 A^2 = AR_E^2 \quad (2.93)$$

$$B^2 v^2 - 2\vec{S} \cdot \vec{v} BC + S^2 C^2 + \gamma^2 A^2 = AR_E^2 \quad (2.94)$$

From this we can calculate γ with the help of scalar ϵ , as a function of the pointing direction:

$$\gamma = \frac{\epsilon}{A} \sqrt{AR_E^2 - B^2 v^2 + 2\vec{S} \cdot \vec{v} BC - S^2 C^2} \quad (2.95)$$

We now know α , β , γ . We have therefore fully calculated the value of \vec{P} with the help of the preceding results, which can also be expressed in condensed form:

$$\alpha = \frac{Bv^2 - C\vec{S} \cdot \vec{v}}{A} \quad (2.96(a))$$

and:

$$\beta = \frac{CS^2 - B\vec{S} \cdot \vec{v}}{A} \quad (2.96(b))$$

This process may require iterations when the Earth's radius R_E needs correcting for a local elevation which itself could be strongly dependent on the position.

Zero Doppler geometry is intellectually simpler. We only need to imagine the intersection of two spheres and the plane (the plane perpendicular to the flight path and containing the position of the satellite, Fig. 3.26). This plane is the same as the degenerate version of the preceding hyperboloid. In practice however we do no more than simplify the expression for C which does not affect the calculations much.

Instead of finding the direct geometric solution, it is much simpler to solve the inverse geometric problem, that is to say: "Given a point P , what are the range/time coordinates in the image?" Using the 'zero Doppler' formulation, we only have to find the position of the satellite at its point of closest approach to P . This is a 'classic' orbital problem for which solutions have existed for a long time outside of the field of radar imaging.

The equations in the previous section are considerably simpler when Doppler is zero because the surface of the iso-Doppler is a plane. We can easily reproduce this situation without error even if the Doppler is not zero. This is because, in radar coordinates (azimuth and range), the trajectory of any given point forms a parabola. From our departure point (the mean Doppler used for processing) and the focus (the processing rate), we can deduce the position of the summit of this parabola and perform the azimuth and range shifts which will place the processed point at this position. This is the same position it would have had if it had been observed at zero Doppler by the radar.

Poor knowledge of the Doppler never leads to azimuth and range errors, because the Doppler used by the processing and not the 'natural' Doppler, determines the position of the target. We can exclude the possibility of an integer rank error on the Doppler which would have produced a poor quality image and would have been easily detected due to the resulting major geometric error (equal to the width of the azimuth processing window). The error therefore comes solely from uncertainty about the compression rate.

2.6.3 Perpendicular radargrammetry

Two radar images taken from different orbits will generally have different incidence angles relative to a given point of the terrain. Their respective range pixels, sharing the same size, will be projected onto the ground with different sizes (Fig. 2.21). Even on flat terrain, the images will therefore have a different scale. In the event of a change in elevation, the slope of the terrain will modulate this difference in scale and lead to a stereoscopic or 'radargrammetric' effect, allowing us to reconstruct the topography from a detailed analysis of the images' geometric distortions.

2.6.4 Radargrammetry and interferometry

In geometrical terms, radargrammetry is strictly identical to interferometry. Both techniques detect and measure the same geometrical deviations. Interferometry however (cf. Chap. 4) is less robust and more difficult to implement particularly because the conditions on incidence angle and Doppler similarity are stringent. It also leads to an ambiguous result because of fringe wrapping. It is, however, more precise than

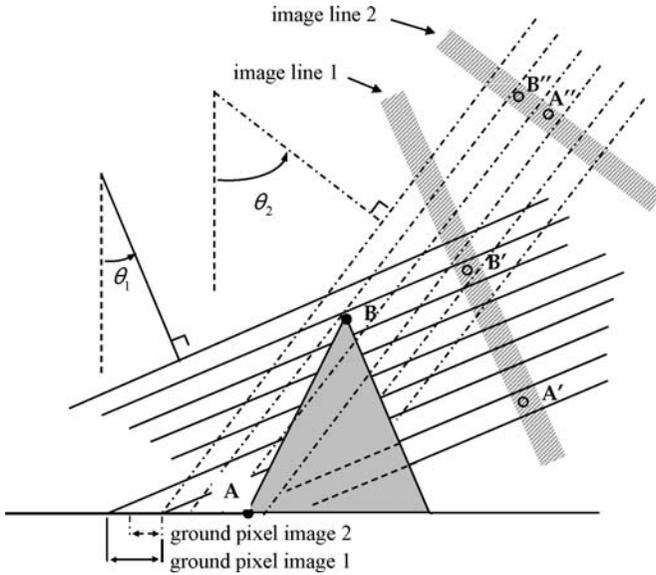


Fig. 2.21 Principles of radargrammetry.

radargrammetry because the phase measurement, which can be significant down to a millimeter, is more precise than the correlation measurement since the precision of the latter is typically limited to one tenth of a pixel. This difference in geometric efficiency generally compensates for the very small stereoscopic baseline in interferometry, compared to that permitted by radargrammetry. One argument in favor of radargrammetry is that it is unaffected by atmospheric influence. Atmospheric activity (Sect. 4.6.1) on radar imagery is obviously the same whether we use the image in radargrammetry or in interferometry. However, its order of magnitude (a few centimeters) is similar to the path differences measured in interferometry and considerably less than those measured in radargrammetry. Atmospheric activity is therefore very harmful for the former but not for the latter.

Let us attempt a more quantitative comparison between the two techniques. Consider a radar observation with incidence angle θ , wavelength λ and a ratio $Q = f_c/f_d$ (Sect. 2.2.3). In interferometry, the maximum tolerable deviation between viewpoints $\Delta\theta$ is such that (Sect. 4.5):

$$\tan(\Delta\theta) = \frac{\tan(\theta)}{Q} \quad (2.97)$$

whereas in radargrammetry the difference can be on the order of θ . To first order, the two techniques should have similar results, if the range pixel offset can be sliced as precisely as the phase offset.

2.6.5 Oblique radargrammetry

Generally speaking, two radar images taken from different viewpoints will display not only local range difference but also local azimuth differences. This second effect, which can be seen even if the two radar scenes are referenced at zero Doppler, may result from the angle of orbital paths. One might think intuitively that this is an effect very similar to the backward-forward looking stereoscopic effect created by some optical instruments, but this would be incorrect. In zero Doppler geometry, any point on the ground is located in a plane perpendicular to the velocity at a given range, and therefore on a circle. Each plane corresponds to the time t when the point is at its closest to the radar. With a single view, we cannot tell where the point is on the circle. With two views, however, we can if the trajectories, and therefore the planes, are parallel and if the circles are not concentric (meaning that there must be a stereoscopic baseline, either for radargrammetry or interferometry). In the current example, the paths are not parallel. A variation in elevation e of a point on the radar image taken as the reference, will shift it on the circle of this image (from A to B on the time t plane) but on the second image, the point will not remain on the same time plane. This means that the point will change its corresponding time on the second image as a function of its elevation. In other words, the change in elevation will be detected as an azimuth shift as (Fig. 2.22).

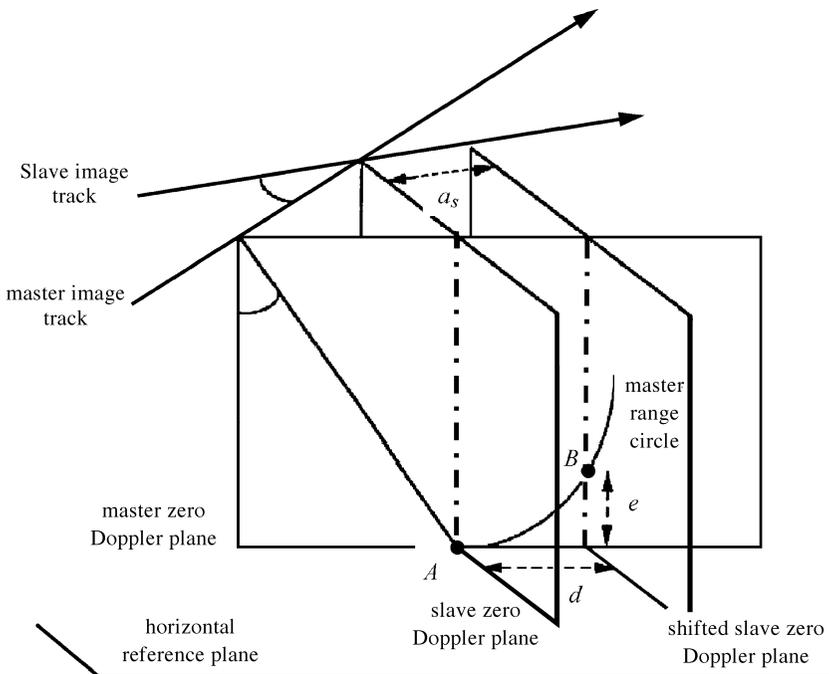


Fig. 2.22 Geometry of oblique radargrammetry.

In the case of interferometric processing, the shift in azimuth is not exploited, but we must compensate for it anyway in order to obtain two perfectly co-registered images before interferometric combination.

2.6.6 Radar clinometry

An uncommon technique for calculating relief is clinometry, which consists in interpreting variations in backscatter on a single radar image as resulting solely from local terrain slopes. The aim is to derive a topographic map from estimates of the slopes. This technique has the advantage of only needing a single image. Nonetheless there are several disadvantages, the first of which is oversimplification in its hypotheses:

- The orientation of the terrain relative to the incident wave depends not only on the its rotation component in the azimuth direction, the one used by this technique as the slope in the cross-track direction, but also on the rotation component in the line of sight direction (where the direction of the rotation is the direction around which the terrain rotates).
- The reconstruction of the topography from its derivative (the slope) must be treated cautiously as it is subject to noise.
- Variations in backscatter can arise from factors other than slope (soil moisture, vegetation cover, etc.).

In short, clinometry gives something like an ‘artist’s impression’ of elevation as a result of examining the image. Considering that the overall topography of the Earth is already available, with greater precision than that which clinometry could provide, this technique is now something of an historic curiosity or, at best, an empirical technique to be used when only one set of data is available (such as in planetary studies).

2.7 An introduction to super-resolution

Two radar images taken under slightly different incidence angles can be combined into a single image with better resolution. In order to understand how this is possible, we shall use a simple diagram in which we see the cross-section of a radar pixel perpendicular to the flight path of the radar instrument. Let us imagine that this pixel contains two ‘elementary targets’ A and B in any configuration, but not in the same half of the pixel (Fig. 2.23). These targets, illuminated by two passes of the radar instrument, return their signals during the first pass along the line of the full arrows, where we assume that they are summed ‘in phase’ (i.e. their difference in round-trip range $2\delta R$ is an integer multiple of the radar’s wavelength). For the pixel in question, the radar will therefore receive the signal ‘A + B’. If on its second pass, the radar listens in the direction of the dotted lines, the distance between A and B will differ from the previous pass by one quarter of a wavelength, and therefore half a wavelength for the round trip. Targets A and B will be received ‘in phase opposition’ and, in this case, the value of the pixel

will be 'A - B'. If the corresponding pixels for the two passes are added together, we obtain twice the contribution of A. If they are subtracted, twice the contribution of 'B' is obtained. We have therefore succeeded in distinguishing targets A and B by creating these two new pixels. This means that we have improved the resolution by up to a factor of two. We can calculate the typical angular offset $\Delta\theta$ corresponding to the angle between the full and dotted arrows. In the case of the ERS-1 radar satellite, flying at an altitude of 785 km, twice the size of the range pixel is 279.5 times the wavelength (5.6 cm in this case), with an incidence angle of 23° . A and B may be separated by 140 times the wavelength at 23° . At an angle of 23.043° the range would be 140.25 times the wavelength while at an angle of 22.957° it would be 139.75 times. The difference is 0.087 degrees. At the range of the satellite, this angle corresponds to a separation known as the critical perpendicular baseline. In the case that we have just calculated, the observation range is approximately 850 km and the critical perpendicular baseline is 1300 m (this limit applies to the interferometric technique, Sect. 4.1.6).

The reality is more complicated because most pixels are covered with many randomly distributed elementary targets. Nonetheless, the value of the critical baseline that we have defined remains valid and marks the limit beyond which the two images received are completely independent. It then becomes difficult, or even impossible, to combine them. We will therefore try to stay below the critical baseline, while at the same time trying to maximize the baseline if we seek extreme super-resolution or high sensitivity to topography.

There is another notion of critical baseline which sets the limits of the image independence in the direction parallel to the flight path of the radar instrument. In this direction, images may be subject to various Doppler effects, depending on the way the orbiting velocity is projected along the pointing directions of the passes (for example, the pointing could shift slightly backwards and forwards).

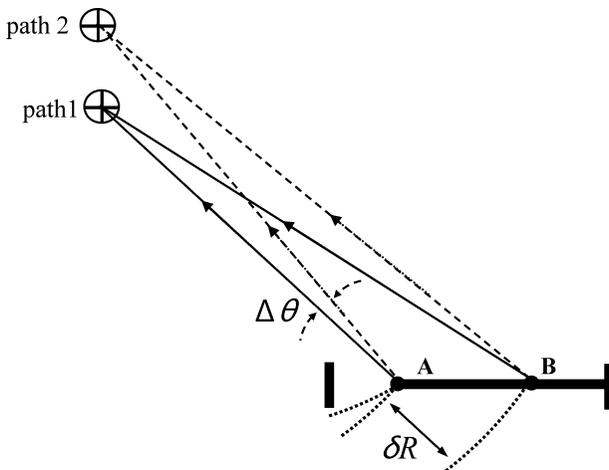


Fig. 2.23 Principles of super-resolution.

The combination of these two critical baselines makes it possible to define an ‘interferometric surface’ for each radar pass. The size of the intersection of these surfaces reflects what the data have in common, and to what extent they can be used for interferometric comparison (Sect. 4.7.4.1). The union of these surfaces reflects the total amount of information that can be used to improve the resolution.

2.8 Radar processing and geometric specificity of bistatic data

A system such as the interferometric cartwheel (Sect. 4.7.4) is a case of bistatic (even multistatic) system. Bistatic systems, for which the transmitting and receiving antennas are not colocated, raise specific processing and co-registration problems, due to the high squint angle typical of bistatic operations. The a priori knowledge of the topography is an asset for efficient processing in terms of autonomy and speed. It can be appropriate to use this knowledge very early, for example during the processing from raw to resolved data.

Fig. 2.24 gives a general geometric layout of the bistatic acquisition: we observe a point P with cartesian coordinates \vec{P} . The origin of time t is chosen when the transmitter is at its closest distance (R_{0T}) to P . At $t = 0$, the transmitter T is characterized by its position \vec{T}_0 , its velocity vector \vec{v}_T and its acceleration $\vec{\gamma}_T$. The situation summarizes as follows:

$$\vec{T}(t) = \vec{T}_0 + t \cdot \vec{v}_T + \frac{t^2}{2} \vec{\gamma}_T \quad (2.98a)$$

$$(\vec{T}_0 - \vec{P}) \cdot \vec{v}_T = 0 \quad (2.98b)$$

$$(\vec{T}_0 - \vec{P})^2 = R_{0T}^2 \quad (2.98c)$$

For convenience, we also introduce the following scalars: $F_T = (\vec{T}_0 - \vec{P}) \cdot \vec{\gamma}_T + \vec{v}_T^2$, $v_{\gamma T} = \vec{v}_T \cdot \vec{\gamma}_T$, and $\gamma_{sT} = \vec{\gamma}_T^2/4$. The transmitter T generally illuminates close to, but not necessarily at zero Doppler. The scalar F_T is directly proportional to the azimuth compression factor N_a (Sect. 2.2.5).

Let us assume that a bistatic receiver R follows at a distance. With similar conventions and obvious notations, we have :

$$\vec{R}(t) = \vec{R}_0 + t_2 \cdot \vec{v}_R + \frac{t_2^2}{2} \vec{\gamma}_R \quad (2.99a)$$

$$(\vec{R}_0 - \vec{P}) \cdot \vec{v}_R = 0 \quad (2.99b)$$

$$(\vec{T}_0 - \vec{P})^2 = R_{0R}^2 \quad (2.99c)$$

As previously, we introduce the scalars $F_R = (\vec{R}_0 - \vec{P}) \cdot \vec{\gamma}_R + \vec{v}_R^2$, $v_{\gamma R} = \vec{v}_R \cdot \vec{\gamma}_R$, and $\gamma_{sR} = \vec{\gamma}_R^2/4$. The origin of time t_2 corresponds to the crossing of P by the receiver R . We have : $t_2 = t - t_0$ where t_0 is the delay between the transmitter and the receivers at P crossing.

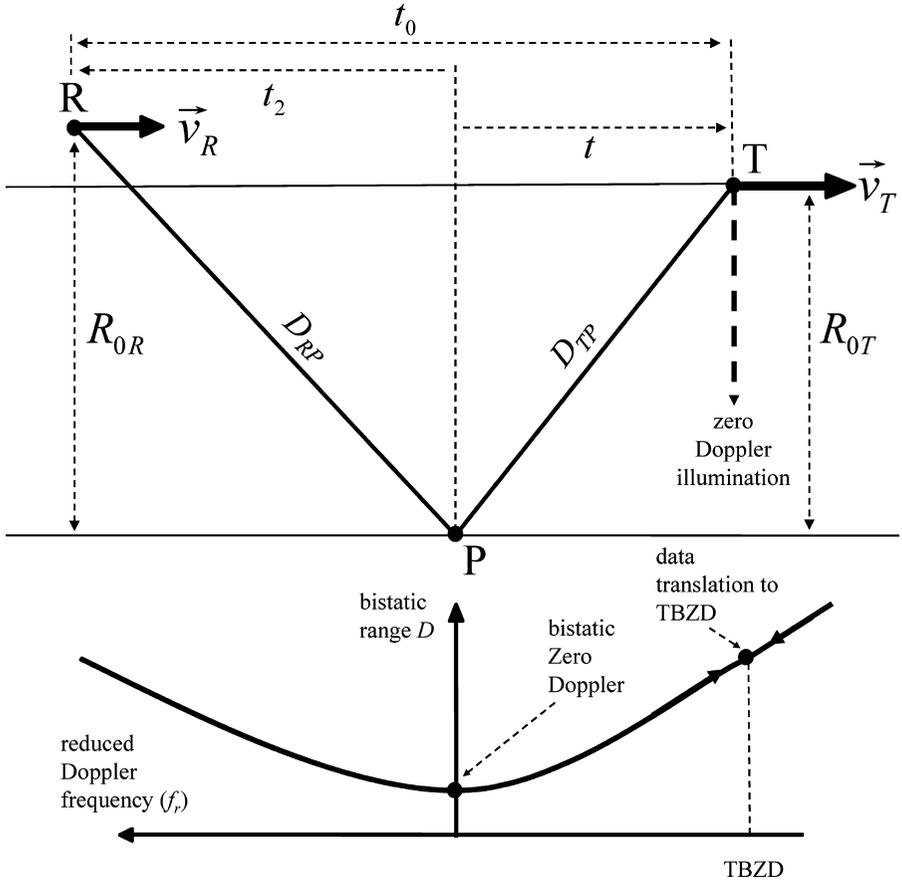


Fig. 2.24 Geometry of the bistatic acquisition. The sign of the frequency depends whether we consider the reconstruction frequency or the signal frequency.

The distance between T and P as a function of time is given by $D_{TP} = \sqrt{(\vec{T}(t) - \vec{P})^2}$, which develops into:

$$D_{TP} = \sqrt{R_{0T}^2 + F_T t^2 + v_{yT} t^3 + \gamma_{sT} t^4} \quad (2.100)$$

Similarly, the distance between R and P is:

$$D_{RP} = \sqrt{R_{0R}^2 + F_R t_2^2 + v_{yR} t_2^3 + \gamma_{sR} t_2^4} \quad (2.101)$$

The bistatic range D is given by $D = D_{TP} + D_{RP}$. The bistatic zero Doppler is obtained when the bistatic range derivative is cancelled, i.e.

$$\frac{dD}{dt} = \frac{dD_{TP}}{dt} + \frac{dD_{RP}}{dt} = 0 \quad (2.102)$$

Given Eq. (2.101), we have:

$$\frac{dD_{RP}}{dt} = \frac{2F_R t_2 + 3v_{\gamma R} t_2^2 + 4\gamma_{sR} t_2^3}{2D_{RP}} \quad (2.103)$$

A very interesting condition arises when dD_{TP}/dt is null, indicating the illumination is made at zero Doppler. By definition, the corresponding time is t_0 , and the corresponding instantaneous Doppler is derived by substituting t_0 to t_2 in Eq. (2.103).

Once it has been divided by the azimuth sampling frequency f_a , the so-called **Transmitter-Based Zero Doppler** (TBZD) is given by:

$$f_{orT} = \frac{1}{\lambda f_a} \frac{dD_{RP}}{dt} = \frac{2F_R t_0 + 3v_{\gamma R} t_0^2 + 4\gamma_{sR} t_0^3}{2(D - R_{0R})\lambda f_a} \quad (2.104)$$

As in the monostatic case (Sect. 2.4.3), the processing of the data will provide an observed reduced azimuth Doppler frequency f_r . Instead of translating the data to zero Doppler, as we would do for a monostatic observation, we will move them to the TBZD. The observed Doppler sets the time of emission by the transmitter by solving:

$$f_r = \frac{1}{\lambda f_a} \left(\frac{dD_{RP}}{dt} + \frac{dD_{TP}}{dt} \right) \quad (2.105)$$

leading to

$$f_r = \frac{2F_R(t - t_0) + 3v_{\gamma R}(t - t_0)^2 + 4\gamma_{sR}(t - t_0)^3}{2D_{RP}\lambda f_a} + \frac{2F_T t + 3v_{\gamma T} t^2 + 4\gamma_{sT} t^3}{2D_{TP}\lambda f_a} \quad (2.106)$$

Once t has been determined, the rest follows.

2.8.1 Direct echo

Here we assume that the conditions allowing the detection of the direct echo are met, and that this one is recorded in the data. The advantages of detecting the direct echo from the transmitter are threefold: (1) The direct echo is received at a very low Doppler, thanks to the very small difference of modulus between \vec{v}_T and \vec{v}_R . Summing a few ten of pulses gives a very accurate measurement of this difference. (2) The direct echo provides a timing reference for the sampling by the receivers. Disregarding some minor corrections, the delay is very close to t_0 and the sampling of the receivers can be based on this starting time, which can also be used for selecting the SWST (Sect. 2.2.3). (3) Finally, a minor advantage is in having a copy of the transmitted pulse for the optimization of the pulse compression.

2.8.2 Triangular resolution

A given point in the receiver's data stream is characterized by its instantaneous reduced Doppler frequency f_r (observed on the raw data) and its timing compared to the arrival

of the direct echo. A processing choice can be to use the zero Doppler of the transmitter as a geometric reference. It is particularly interesting when there are several receivers, such as for the Cartwheel (Sect. 4.7.4). Under this assumption, the correction in azimuth and range will move the target from the position where the bistatic range is minimal to the plane perpendicular to the transmitter's orbit, at an appropriate time and range (for instance the time and range of closest approach of the target with respect to the transmitter). However, this operation depends on the knowledge of the topography.

We can quantify this dependency by assessing the increase of altitude which offsets the real position in azimuth by one pixel, with reference to the transmitter's zero Doppler plane. The inaccuracy of the preliminary topography must be such as to create an offset in azimuth less than the accuracy required for further pixel superposition in azimuth.

In case of high squint angle, which is likely to be the rule for cartwheel operations, we must use a distance law without approximation (Sect. 2.4.10) instead of a square law approximation, without modifying the general principle.

We have described here some of the geometric features of bistatic radar observations. These configurations cause azimuth offsets that are very similar to the offsets found with non-parallel orbits (Sect. 2.6.5, Fig. 2.22). These effects could be used *per se* as a mean of computing topography. In this case, the distance between the passive system and the transmitter must be sufficient for proper accuracy. Otherwise it must remain below the threshold set by the accuracy of the a priori knowledge of the topography.

References

- [1] J. C. Curlander, R.N. McDonough, *Synthetic Aperture Radar*, Wiley Series in Remote Sensing, (1991).
- [2] D. Massonnet, F. Adragna, M. Rossi, 'CNES general-purpose SAR correlator', *IEEE Transactions on Geoscience and Remote Sensing*, 32(33), (1994).
- [3] D. Massonnet, F. Adragna, E. Xenofontov, I. Elissaevetina, 'Advanced use of Almaz radar data', in *Acta Astronautica*, 35(7), 461–465, (1995).
- [4] C. Prati, F. Rocca, 'Focusing SAR data with time-varying Doppler centroid', *IEEE Transactions on Geoscience and Remote Sensing*, 30(3), (1992).
- [5] R. K. Raney, R. K. H. Runge, R. Bamler, I. G. Cumming, and F. H. Wong, 'Precision SAR processing using Chirp Scaling', *IEEE Trans. Geosci. Remote Sens.*, 32, 786–799, (1994).

CHAPTER 3

FROM SAR DESIGN TO IMAGE QUALITY

3.1 Introduction

In this chapter, we shall first look at SAR design from a radiometric perspective. The preceding chapter dealt with a geometrical approach to radar processing. It seemed to leave us with a paradox (Sect. 2.2.4): if D is the length of the radar antenna in the azimuth direction, the azimuth resolution obtained after SAR synthesis is $D/2$, which seems to contradict the common understanding of the separating power of a microwave antenna (Sect. 1.2.5.1). This is not really so surprising: if D is reduced, the azimuth radiation pattern is broadened, which increases the illumination time for a target, and thus the width of the Doppler spectrum. Hence the azimuth resolution is naturally improved. So, why should we try to give imaging radars long antennas? There are several reasons. A ‘geometrical’ reason, discussed in Sects. 3.7.1 and 3.7.2, concerns the conditions under which ambiguities appear in the image. But there is also a ‘radiometric’ reason: the shorter the antenna, the lower its directivity and thus the weaker its gain. The amplitude of the radar echo is attenuated proportionally to the square of the antenna’s gain (because the wave crosses the antenna’s main lobe twice). It is thus necessary to be careful when designing the antenna to ensure an adequate level of collected power, which requires a minimal value for D . In these few words we have just described the trade-off between the image’s geometric qualities (i.e., its resolution) and its radiometric qualities (which depend on the amplitude of the radar echo). This is the central issue in system design. We shall keep this in mind when considering essentially radiometric issues in the following sections.

We shall first write the radar equation in its simplest theoretical form. This will allow us to define the concept of Radar Cross Section (RCS), both for point targets (Sect. 3.2) and extended targets (Sect. 3.3). From this, we infer the signal-to-noise ratio (SNR) of the radar antenna (Sect. 3.4). We shall then discuss how the SNR of the radar antenna affects the image produced and how this SNR is transformed during SAR synthesis (Sect. 3.5). We shall see that we need to distinguish between the behavior of point targets and extended targets. Finally, the ‘noise floor’ of the resolved image ($NE\sigma^{0tot}$) includes a contribution from the radar antenna ($NE\sigma^{0inst}$, Sect. 3.6),

increased by the azimuth and range ambiguity contributions (Sect. 3.7). To complete the characterization of the link budget, we must calculate the volume of data generated (Sect. 3.8) and the corresponding telemetry dataflow (Sect. 3.9).

Now that the SAR sensor has been designed, and we are acquainted with the link budget, we can ask if we can determine the nature of a radar sensor and the processing applied to its data by carefully examining the resulting image. Does the product reveal the secrets of the instrument? This is one aspect of Image Quality (IQ) assessment.

Determining the quality of a SAR image involves evaluating radiometric and geometric aspects which affect each other. We shall first refer to the calibration of radar images in Sect. 3.10. This is an important part of SAR analysis. We have chosen to deal with it before other aspects of image quality. Speckle noise is covered in Sect. 3.11. After first investigating its physical origin, it is described statistically. We refer briefly to the traditional speckle filtering algorithms. We shall not go into detail as this has been extensively covered in the literature. Readers may refer to [Touzi, 2002] who gives a thorough review of filter basics and enhancements. Section 3.12 considers IQ analysis in detail. The radiometric part is addressed first, even though it is impossible to fully separate radiometric and geometric aspects of image quality. Image quality parameters are generally described as being the effect of a double weighting. The first one comes from the antenna pattern (Sect. 3.7.2) and leads to $NE\sigma^{0tot}$, the other is applied to the digitized radar data. We shall be discussing here the ‘downstream’ weighting applied to the radar data in the frequency domain. This weighting affects the resolution of the image via the ‘impulse response’ but it also affects the level of radiometric pollution from the nearby environment. The geometric part (Sect. 3.13) briefly reviews concepts of resolution and pixel size and then addresses the evaluation of distortion and localization accuracy characterizing the image products. Finally, in Section 3.14, we present a practical example based on a series of images which illustrates the amount of information accessible through in-depth radar photo-interpretation.

3.2 Radar equation, Radar Cross Section (RCS) of a point target

The Radar Cross Section (RCS) of a point target characterizes the ‘strength’ of its echo in the radar signal. The interaction between the incident wave and the target is expressed by the radar equation, which summarizes a sequence of three events:

- (1) The radar transmits a pulse with peak power P_e , amplified by the antenna gain G (Sect. 1.2.5.1). When located at a range R from the radar, the target is illuminated by a power density dP^{inc} (expressed in W/m^2) equal to:

$$dP^{inc} = \frac{P_e \cdot G}{4 \cdot \pi \cdot R^2} \quad (3.1)$$

This equation treats the antenna's radiation as isotropic, that is, in all directions the radiation is the same as that along the radar line of sight.¹ The quantity $P_e \cdot G$ is known as the *effective isotropic radiated power* (EIRP).

- (2) Illuminated by the radar, the target now behaves as an isotropic radiating element returning to the radar a power density dP^s of

$$dP^s = dP^{inc} \cdot \frac{\sigma}{4 \cdot \pi \cdot R^2} \quad (3.2)$$

which suggests that dP^{inc} has been intercepted and reflected by a surface σ (the RCS, expressed in m^2) which behaves isotropically and *independently of the observation range*. If the target actually radiates isotropically (as would be the case for a conducting sphere whose radius R_s is much larger than the wavelength), the RCS is exactly the surface seen by the radar (here its actual geometric cross-section $\pi \cdot R_s^2$) (Fig. 3.1(b)). In all the other cases, a target radiating preferentially in the backscatter direction has an RCS greater than its geometric cross-section (Fig. 3.1(a)), whereas a target 'evacuating' energy in a direction other than the backscatter direction has a lower RCS (Fig. 3.1(c)).

Fig. 3.2 gives some typical values for the RCS of point targets, as well as their scattering beam width along the main axes. For example, the corner reflector made up of three orthogonal triangular plates, gives a maximum RCS of

$$\sigma^{trihedral} = \frac{4\pi \cdot a^4}{3 \cdot \lambda^2} \quad (3.3)$$

where a is the vertex length and λ the transmitted wavelength. For a 1-m corner reflector, the RCS is 1289 m^2 at C band, whereas the equivalent sphere in RCS terms would have a radius of 20 m! This powerful signature combined with its very open radiation pattern makes it a widely used radar calibrator².

- (3) Finally, the power P^{rec} collected by the receiving antenna with an effective area S_{eff} (Sect. 1.5.1) is given by:

$$P^{rec} = dP^s \cdot S_{eff} \quad (3.4)$$

Equations (3.1), (3.2) and (3.4), combined with the relationship linking effective area and antenna gain (Sect. 1.2.5.1), lead to the radar equation (assuming that antenna gain is identical for transmission and reception):

$$P^{rec} = P_e \cdot \frac{\lambda^2}{(4\pi)^3} \cdot \frac{G^2}{R^4} \cdot \sigma \quad (3.5)$$

¹In reality, the antenna's radiation is not isotropic at all! However, erroneous hypotheses formulated for other directions than that along the radar's line of sight have no effect on the final result.

²In the military field on the other hand, electromagnetic discretion requires that structures be designed with the lowest RCS possible: when seen from the front, the *F-104 Starfighter* has an RCS equivalent to 5 m^2 between 1 and 10 GHz and in the same conditions, the *MIG-21* has an RCS of about 3 m^2 [Morchin, 1990]. The figures given here show how different the perception can be between the eye and radar when target radiation is no longer isotropic!

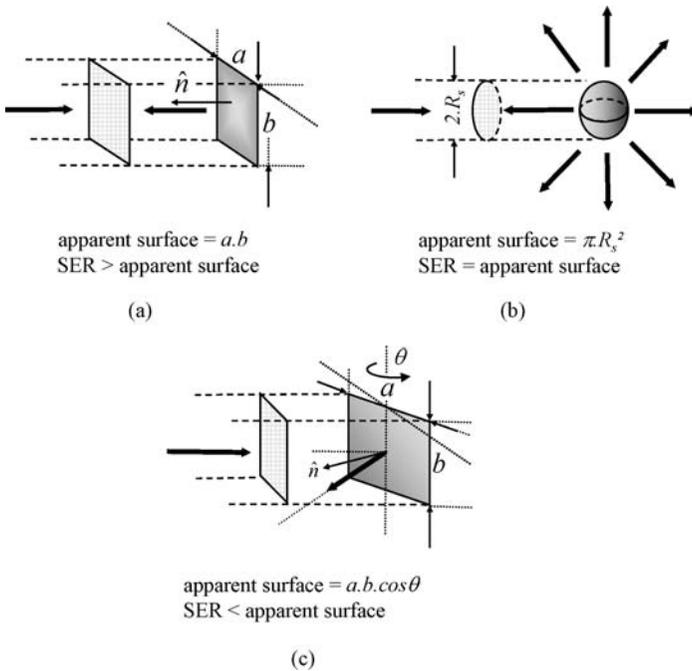


Fig. 3.1 Comparing RCS and actual surfaces in standard situations; (a) actual surface = $a \cdot b$; $RCS >$ actual surface; (b) actual surface = $\pi \cdot R_s^2$; $RCS =$ actual surface; (c) actual surface = $a \cdot b \cdot \cos \theta$; $RCS <$ actual surface.

The power received increases in direct proportion to the square of the antenna gain. Another element in the link budget is the term λ^2/R^4 , which expresses the fact that attenuation increases as frequency increases. This attenuation is also in direct proportion to R^4 .

Equation (3.5) predicts the expected echo signal levels reaching the radar. For example, a 1-m^2 parabolic antenna transmitting 1-kW peak power at C band ($\lambda \approx 5.7\text{ cm}$) towards a 1 m radius conducting sphere at a range of 1 km will receive a peak power of the order of 10^{-8} W !

3.2.1 Loss terms

Up until now, we have assumed that the radar/target link has no losses. In reality, the power received is less than what might be expected. Losses have several causes, which we list here in order of decreasing magnitude:

- Losses in the transmitting and receiving devices (antenna, preamplifier, Radio-Frequency losses and encoding losses): because of these losses, it is necessary

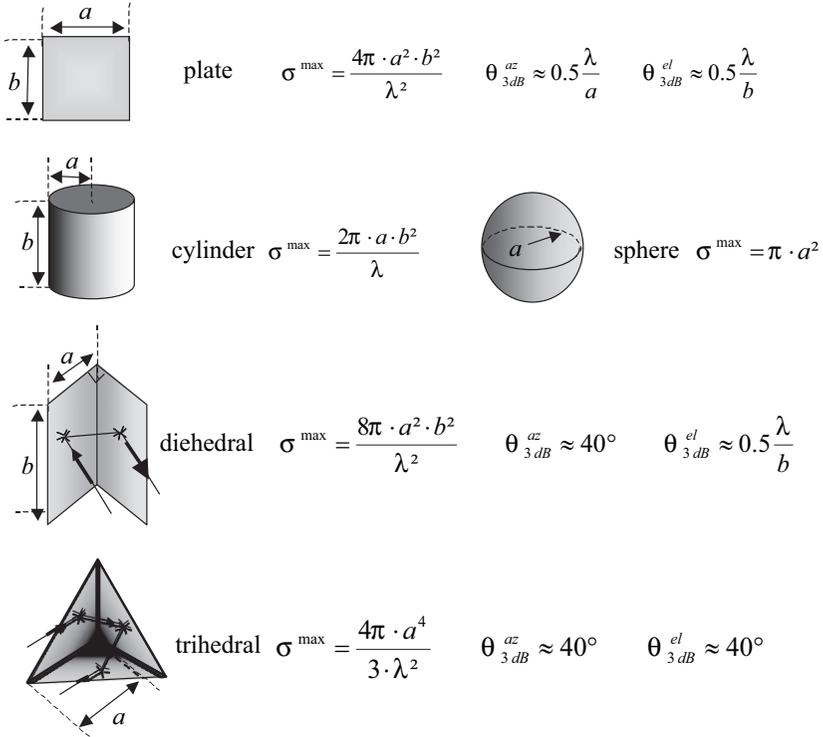


Fig. 3.2 Examples of the RCSs of standard targets (after [Ruck, 1970]). The indicated values correspond to configurations generating maximum signatures. θ_{3dB}^{az} and θ_{3dB}^{el} represent 3-dB apertures for azimuth and elevation radiation patterns respectively.

to distinguish between the radiated power P_e (which is an element in the radar equation) and the power actually produced by the high power amplifiers with an efficiency of the order of 50%.³

- Losses during atmospheric propagation: Propagation models predict low attenuations in the atmosphere, though certain situations may lead to higher atmospheric losses, for example when short wavelengths (X band; K_u band) encounter rain. This effect could become more critical for the K_a band, around 35 GHz. In addition, the troposphere creates losses that affect frequencies in X band and beyond. All these atmospheric losses (from atmospheric gases, clouds, rain and

³The microwave source used on the ERS-1 and 2 satellites (C band, VV polarization, 23°) was a Travelling Wave Tube (T.W.T.) with an output power of 4800 W for a radiated power of 2524 W. A more complete energy budget for the platform would require taking into account the efficiency of the T.W.T. and of its power source (about 40 % overall). However, this efficiency does not affect the link budget, which begins with the amplifier output, but only the design of the power sub-system. For an active antenna the efficiency of *Solid State Power Amplifiers (SSPA)* is less than that of a T.W.T. (~ 25 to 30%), but propagation losses within the receiver are reduced, as the SSPAs are located in the immediate vicinity of the radiating elements.

ionospheric scintillation) are also influenced by the geographic location. Variations of 0.1 dB to 0.3 dB can be found in C band and 0.3 dB to more than 1 dB in X band.

- Losses linked to polarization effects: There are two types, firstly the possible rotation of the wave's polarization during propagation (the Faraday effect when crossing the ionosphere). This effect can be critical for frequencies lower than 1 GHz. secondly, Depolarization effects (Sect. 1.1.3).

All these losses together are combined in the *Loss factor* which is included in the denominator of the radar equation. These losses can have a magnitude of 2 to 3 dB for each *one way* trip of the wave.

3.2.1.1 Noise factor

The noise factor F is the parameter characterizing noise of electronic origin affecting receiving devices. A receiver whose ambient temperature is T_0 behaves as though the temperature of its electronics was $T_n = F \cdot T_0$. The temperature $T_n = F \cdot T_0$ is called the 'noise equivalent temperature'. The noise factor F is approximately 0.5 to 1.5 dB depending on the frequency and the bandwidth.

Finally, the term expressing overall *Loss* (losses + noise factor) is of the order of 6 to 9 dB. This is the value usually used for a first approximation when designing SAR systems.

3.3 Radar signature for extended targets - the backscatter coefficient σ^0

How can we adapt the concept of RCS to extended targets? If these are stationary, they backscatter towards the radar an amount of energy proportional to their physical surface S . In order to get around the backscatter's dependency on S , a normalized signature is introduced :

$$\sigma^0(dB) = 10 \cdot \log_{10}(d\sigma/dS) \approx 10 \cdot \log_{10}(\sigma/S) \quad (3.6)$$

σ^0 is the radar backscatter coefficient expressed in m^2/m^2 (the exponent 0, indicates that the RCS is normalized).⁴ Most natural surfaces have negative σ^0 (when expressed on a logarithmic scale, i.e. in dB m^2/m^2): therefore $1 m^2$ of their physical surface has an RCS of less than $1 m^2$.

⁴Measurements are usually expressed on a logarithmic scale because of the large range of dynamics involved.

3.4 Signal to noise ratio (SNR) of the radar-target link before SAR synthesis

This SNR, hereafter denoted SNR^{raw} , is calculated from the radar equation and from the thermal noise equation P_n :

$$P_n = k \cdot T_n \cdot B_n \quad (3.7)$$

where k is Boltzmann's constant ($= 1.38 \times 10^{-23}$ J/K) and B_n is the receiver's bandwidth (in Hz). From Eqs. (3.5) and (3.7), we get:

$$SNR^{raw} = \frac{P^{rec}}{P_n} = P_e \cdot \frac{\lambda^2}{(4\pi)^3} \cdot \frac{G^2}{R^4} \cdot \frac{1}{Loss} \cdot \frac{1}{k \cdot T_n \cdot B_n} \cdot \sigma \quad (3.8)$$

σ is either the RCS of a point target or the product $\sigma^0 \cdot S$ for the RCS of a stationary extended target with physical surface S . In the latter case it is necessary to determine the surface area S to be included in the link budget:

- In the azimuth direction, the surface's length is tied to the azimuth HPBW, which gives a length of approximately $R \cdot \lambda / D$ (D being the length of the antenna), to within a coefficient which is close to 1 and depends on the antenna's illumination law (Section 1.2.5).
- In the range direction, T^{pulse} being the duration of the pulse transmitted by the radar, any signal received at time t_0 was generated by the portion of the pulse transmitted between $t = \tau_t$ and $t = \tau_t + \delta\tau_t$, then reflected by a strip of terrain located between range r and range $r + \delta r$, such that $\tau_t + 2r/c = t_0$. If we now consider the entire pulse duration, we collect at the same time the contributions from a strip of terrain (located at a mean range R), with width $c \cdot T^{pulse} / (2 \cdot \sin \theta)$, where θ is the local incidence angle.

Thus, at a given time, the surface which has contributed to the response is $S = R \cdot \lambda \cdot c \cdot T^{pulse} / (2 \cdot D \cdot \sin \theta)$. This is by definition the surface area of the raw data resolution cell. Knowing S , we deduce the SNR for an extended target before SAR synthesis:

$$SNR_{ext}^{raw} = P_e \cdot \left(\frac{\lambda}{4\pi R} \right)^3 \cdot G^2 \cdot \frac{1}{Loss} \cdot \frac{1}{k \cdot T_n \cdot B_n} \cdot \frac{\sigma^0 \cdot c \cdot T^{pulse}}{2 \cdot D \cdot \sin \theta} \quad (3.9)$$

3.5 Modifying the SNR during SAR synthesis

Apart from its effect on resolution (Sects. 2.2.4 and 2.2.5), SAR synthesis modifies the SNR significantly (at least for point targets, as we shall see). Its impact can be explained by means of the adaptive filter theory: by 'intelligently' summing N samples

of the *same measurement* (i.e. by giving them all the same phase, as is done in SAR synthesis, Sect. 2.5.1) with independent noise samples, the SNR is improved by a factor of N .

We can illustrate this geometrically. The first four sketches in Figure 3.3 show the breakdown of a radar pixel into a ‘useful’ component, represented by an unbroken line, whose length is 1, and an ‘unstable’ component represented by a dotted line with a mean of 1. The real and imaginary parts of the unstable components have a Gaussian distribution. The phase of the overall contribution inside a pixel (useful component plus unstable component, in broken line) differs from zero as a result of this random contribution. In the fifth sketch, we show the coherent addition of the vectors from the four previous sketches. The modulus of the ‘useful’ contribution has a value of four. Because of the inefficiency of the incoherent additions, the random contributions do not alter the phase of the coherent contributions as strongly as before the summing. It should be noted that though this is true for the phase, it is not true for the modulus, which is significantly altered by the summing. The summing of complex numbers gives very different results depending on whether the contributions are coherent or non-coherent. The modulus of the sum of N coherent vectors with the same magnitude is N times the modulus of one of them alone. The resulting power (the square of the modulus) in this case is amplified by a factor of N^2 . On the other hand, the power of the sum of N random vectors is N times the mean power of one of them, i.e. a gain of only \sqrt{N} on the modulus. The end result is an improvement of the final ratio between coherent and incoherent modulus of \sqrt{N} , and thus of N for the SNR. In our illustration, the improvement factor is 4. In practice, it is of the order of 1500, and it is called the compression rate (Sect. 2.2.5). We would need 1500 sketches to visualize what actually happens!

We now turn our attention towards the impact of SAR synthesis on both point and extended targets.

3.5.1 Point targets

3.5.1.1 The effect of pulse compression

Every $1/f_a$ the radar transmits a pulse $s^{inc}(t)$ of length T^{pulse} (Fig. 3.4(a)). f_a is the pulse repetition frequency (Sect. 2.2.3). The echo $s^{rec}(t)$, assumed to come from a single point target at range R , is received (and sampled) after a lapse of time $\Delta t^{prop} = 2 \cdot R/c$ giving:

$$s^{rec}(t) = \xi \cdot s^{inc}(t - \Delta t^{prop}) + n(t) \quad (3.10)$$

ξ is an attenuation coefficient which will be defined later (Sect. 3.10). The quantity $n(t)$ is a zero mean Gaussian thermal noise. The SNR at reception is thus

$$SNR^{raw} = \xi^2 \cdot \frac{\langle s^{inc}(t) \cdot s^{inc*}(t) \rangle}{\langle n(t) \cdot n(t)^* \rangle} \quad (3.11)$$

where $\langle \cdot \rangle$ denotes an averaging process over several samples. The quantity $s^{rec}(t)$ is then filtered with what is known as an ‘adaptive filter’:

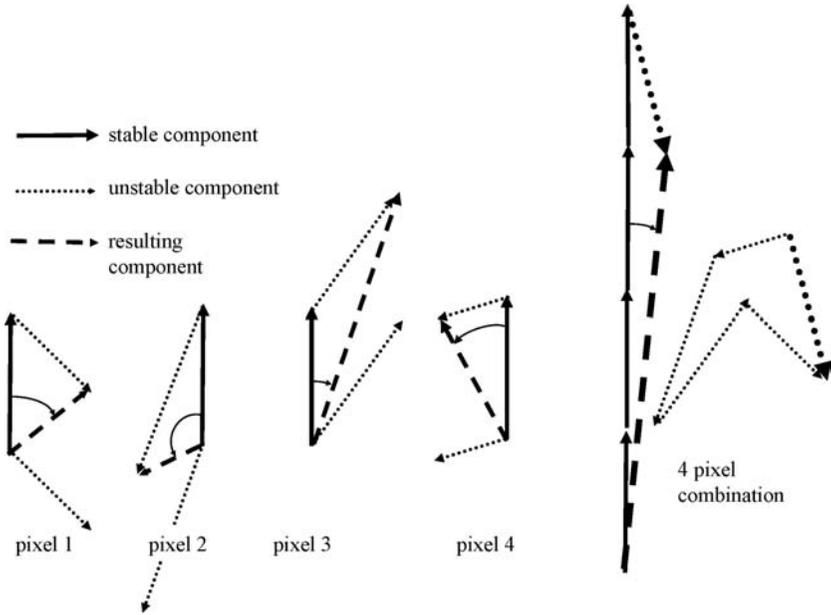


Fig. 3.3 The effect of adaptive filtering on the coherent and incoherent components of a radar pixel.

$$s^{out}(t) = \int_{-\infty}^{+\infty} h(t - \tau) \cdot s^{rec}(\tau) \cdot d\tau \tag{3.12}$$

with

$$h(t) = s^{inc*}(-t) \tag{3.13}$$

s^{out} is the filter output. Equations (3.12) and (3.13) express the correlation between the received signal and the replica of the transmitted one (this is discussed in Eqs. (3.12) and (3.13)). The integration (3.12) involves a weighted addition of the samples $s^{rec}(\tau)$ collected at regular intervals $d\tau$. This weighting has the effect of putting the respective useful parts of the samples in phase. Since the condition of independence for the corresponding noise samples requires: $d\tau \geq 1/B_n$, the maximum number of independent noise samples is equal to $B_n \cdot T^{pulse}$. This is also the gain in SNR:

$$SNR^{out} = B_n \cdot T^{pulse} \cdot SNR^{raw} \tag{3.14}$$

where SNR^{out} is the SNR at the filter output. Next, the receiving bandwidth B_n is assumed to be equal to the bandwidth B_d of the transmitted signal. Consequently, if the transmitted signal was not frequency modulated, then $B_d \approx 1/T^{pulse}$ and $SNR^{out} \approx SNR^{raw}$. However, in practice, it is frequency modulated (i.e., by a ‘chirp’, Sect. 2.3.7):

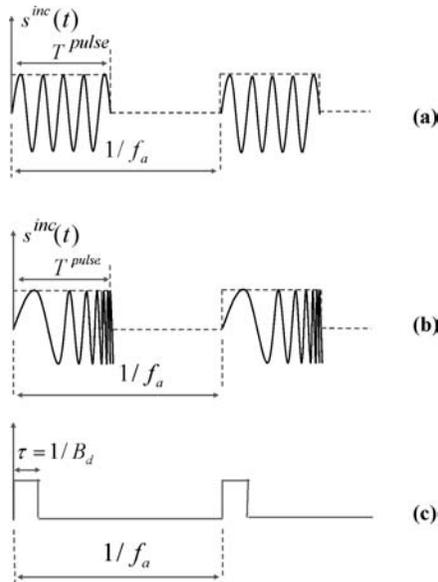


Fig. 3.4 The structure of the transmitted radar signal; (a) transmission of a carrier without modulation; (b) chirp modulation of the carrier (bandwidth B_d); (c) equivalent transmitted signal.

$$s^{inc}(t) = \Re \left(\exp \left[j \cdot 2\pi \cdot \left(f_c \cdot t + \frac{\rho \cdot t^2}{2} \right) \right] \right) = \Re(\exp[j \cdot \phi(t)]) \quad (3.15)$$

with $\rho = B_d/T^{pulse}$, f_c is the carrier frequency and $t \in [-T^{pulse}/2; T^{pulse}/2]$. We will assume here that transmitting a chirp and then using an adaptive filter on reception is equivalent to transmitting a monochromatic pulse whose duration is $\tau = 1/B_d$ (Fig. 3.4(b)-(c)). Since B_d is in practice much larger than $1/T^{pulse}$, the ‘compressed’ pulse is much shorter than the real one (called the long pulse). This leads to a significant improvement in range resolution and in SNR, hence the name ‘pulse compression’. The associated maximum gain (or compression rate)⁵ is given by $G^{range} = T^{pulse}/\tau = B_d \cdot T^{pulse}$. Please note: Applying strictly Shannon condition (Sects. 1.4.6 and 1.4.7) in range gives $B_d = f_d$ (where f_d is the range sampling frequency of the echoes, Sect. 2.2.3). We then obtain $G^{range} = N_d$ (Sect. 2.3.7).

3.5.1.2 The effect of SAR synthesis

In the same way, SAR synthesis is based on the weighted sum of independent echo samples located at the same range from the radar (after correction for migration effects, Sect. 2.3.6). These independent samples are acquired successively during the illumination period T^{ill} (Sect. 2.3.1). The number on independent samples is given by

⁵In the case of the ERS-1 satellite [Tab.I], G^{range} is approximately 575, which corresponds to a short (compressed) pulse duration of approximately $\tau \approx 65$ ns.

Table 3.1 Characteristics of past and present civilian SAR Space missions.

	SEASAT	ALMAZ	ERS-1/2	JERS-1	SIR-C/XSAR	RADARSAT-1	ASAR/ENVISAT
Origin	USA/JPL	Russia	ESA	Japan	USA/JPL	Canada	ESA
Launch year	1978	1991	1991/1995	1992	1994 (Space Shuttle)	1995	2002
Freq. band (Central freq. in GHz)	<i>L</i> (1.275)	<i>S</i> (3.125)	<i>C</i> (5.3)	<i>L</i> (1.275)	<i>L</i> (1.276) <i>C</i> (5.17) <i>X</i> (9.68)	<i>C</i> (5.3)	<i>C</i> (5.331)
Polarization	<i>HH</i>	<i>HH</i>	<i>VV</i>	<i>HH</i>	Full pol. <i>L</i> – <i>C</i> , <i>X</i> - <i>VV</i>	<i>HH</i>	<i>HH</i> , <i>VV</i> image and Global Monitoring mode <i>HH</i> / <i>VV</i> , <i>HH</i> / <i>HV</i> , <i>VV</i> / <i>VH</i> in AP mode
Incidence	23°	30°–60°	23°	35°	17°–63°	20°–50°	15°–45°
Swath width (km)	100	32–65	100	75	15–90 (<i>L</i> and <i>C</i>) 15–40 (<i>X</i>)	35 (fine) 100 (standard) 500 (SCANSAR)	100 image and AP mode 400 Global Monitoring mode
Resolution (m)	25 × 25, 4 looks	15 × 15	20 × 20, 3 looks	18 × 18, 3 looks	Typically 30 × 30	10 × 10 fine mode 25 × 28 standard 100 SCANSAR	20 × 20, 3 looks (image mode) 30 × 30 AP mode
$NE\sigma^0$ (dB·m ² /m ²)	–18	—	–21	–20	—	–21	–20
Bandwidth (MHz)	19	10–15	15.5	15	10, 20, 40	30 (fine mode) 11 (standard)	16
PRF (Hz)	1500	—	1679	1500, 1600	1395, 1736	1270 or 1390	1680 or 2100
f_d (MHz) (range)	19	28.8	18.96	17		32 (fine) / 14 (std)	

Continued

Table 3.1 Continued

	SEASAT	ALMAZ	ERS-1/2	JERS-1	SIR-C/XSAR	RADARSAT-1	ASAR/ENVISAT
Pulse length (μs)	33	0.5	37	35	33.8, 16.9, 8.5 (L) 33.8, 16.9, 8.5 (C) 40 (X)	42	22 to 41
Altitude (km) / Inclination ($^{\circ}$)	800 / 108	240–292 / 72.7	780 / 98	570 / 97.6	225 / 60	790 / 98.6	800 / 98
Antenna size (m)	10.7×2.16	15×1.5	10×1	11.9×2.4	12×4 (for the 3 frequencies)	15×1.5	10×1.3
Peak / Mean power transmitted	3 kW peak 1200 W mean	190 kW / -	5 kW peak 1245 W mean	1.3 kW peak	3 kW to 9 kW mean	5 kW peak 2500 W mean	320 modules with 8 W peaks 1200 W mean (SSPA)
TM data rate ($\text{MB} \cdot \text{s}^{-1}$)	110	—	105	60	90 (L, C) 45 (X)	85 to 105	100

$B^{dop} \cdot T^{ill}$, where B^{dop} is the azimuth Doppler bandwidth and T^{ill} the illumination time (Sect. 2.3.1). As before, the gain in SNR, G^{azi} , equals this number of independent samples. When the entire Doppler bandwidth is processed:⁶

$$G^{azi} = B^{dop} \cdot T^{ill} \approx \frac{2 \cdot R_0 \cdot \lambda}{D^2} \quad (3.16)$$

R_0 is the range for the point of closest approach (Sect. 2.2.5).

Please note: Applying strictly Shannon's condition in azimuth ($D/2 = v/f_a$), leads to $G^{azi} = \lambda \cdot R_0 \cdot f_a^2/2/v^2$: this is nothing other than the equation for the azimuth compression rate N_a (Sect. 2.2.5).

3.5.1.3 SNR on point targets after range compression and SAR synthesis

As a consequence of Eqns. (3.15) and (3.16), complete processing (in range and in azimuth) for a point target produces an overall increase in SNR of the quantity $G^{range} \cdot G^{azi}$ (≈ 58 dB for ERS-1, Table 3.1). Let SNR_{point}^{slc} be the SNR of the resolved image on a point target (SLC stands for *single look complex*). Then:

$$SNR_{point}^{slc} = SNR^{raw} \cdot B_d \cdot T^{pulse} \cdot \frac{2 \cdot R_0 \cdot \lambda}{D^2} \quad (3.17)$$

or, if we expand it further:

$$SNR_{point}^{slc} = P_e \cdot \frac{\lambda^3}{(4\pi)^3} \cdot \frac{G^2}{R_0^3} \cdot \frac{1}{Loss} \cdot \frac{1}{k \cdot T_n} \cdot T^{pulse} \cdot \frac{2}{D^2} \cdot \sigma \quad (3.18)$$

Applying Shannon's condition in azimuth ($v/f_a = D/2$), we finally obtain:

$$SNR_{point}^{slc} = P_{av} \cdot \left(\frac{\lambda}{4\pi R_0} \right)^3 \cdot G^2 \cdot \frac{1}{Loss} \cdot \frac{1}{k \cdot T_n} \cdot \frac{f_a}{2v^2} \cdot \sigma \quad (3.19)$$

where P_{av} is the mean power radiated by the radar.⁷

$$P_{av} = P_e \cdot T^{pulse} \cdot f_a = E \cdot F_a \quad (3.20)$$

and E is the pulse energy. Equation (3.19) is valid when the entire Doppler bandwidth is processed. If this is not the case, f_a in the equation $v/f_a = D/2$ is replaced by the actual value of the azimuth bandwidth processed. The SNR of point targets includes a term in $1/R_0^4$ before SAR synthesis and $1/R_0^3$ after, since the further a point is from the antenna, the longer it is observed.

Adaptive filtering on point targets has some side effects. They result from the fact that the imaging system is unable to focus all their contributions inside a single pixel because the range and azimuth bandwidths are finite. SNR_{point}^{slc} is consequently attenuated by a factor depending on the shape of the impulse response (Sect. 3.12).

⁶The calculations developed below will generally address the case where the entire Doppler bandwidth is considered. Similar calculations can nonetheless be applied to a portion of the entire Doppler bandwidth.

⁷For ERS-1, $P_{av} = 157$ W. Note the low value: the mean radiated power of a space imaging radar is hardly more than the power of a light bulb!

Please note: Because the Earth is a sphere, we ought, strictly speaking, to distinguish between the speed at which the radar spot travels along the ground (v_{ground}) and the velocity v . Generally, v and v_{ground} differ by about 10 %. For the sake of simplicity, we do not take this difference into account as it does not modify the link budget significantly.

3.5.2 Extended targets

Radar processing is not naturally capable of distinguishing between point targets and extended targets. Azimuth and range compression offer the same advantages in both cases.

However, the SNR of extended targets before processing, SNR_{ext}^{raw} , was determined with respect to the resolution cell of the raw data, namely $S = R_0 \cdot \lambda \cdot c \cdot T^{pulse} / (2 \cdot D \cdot \sin \theta)$. After synthesis, the same procedure as that described in Sect. 3.4 is applied, but to a different reference surface. This surface is now the resolution cell of the *SLC* data, i.e. the image pixel under Shannon's condition ($v/f_a = D/2$, $B_d = f_d$).

Consequently, after processing, the SNR improvement (proportional to $G^{range} \cdot G^{azi}$) is counterbalanced by a 'shrinking' of the reference surface which is inversely proportional to $G^{range} \cdot G^{azi}$! The SNR of an extended target, SNR_{ext}^{slc} , is thus unchanged after complete SAR synthesis. From Eqs (3.3)-(3.9), it can be seen that:

$$\begin{aligned} SNR_{ext}^{raw} &= SNR_{ext}^{slc} = P_{av} \cdot \left(\frac{\lambda}{4\pi R_0} \right)^3 \cdot \frac{G^2 \cdot \sigma^0 \cdot c}{Loss \cdot k \cdot T_n \cdot B_n \cdot 4v \sin \theta} \\ &= P_{av} \cdot \left(\frac{\lambda}{4\pi R_0} \right)^3 \cdot \frac{G^2 \cdot \sigma^0 \cdot r_d}{Loss \cdot k \cdot T_n \cdot 2v} \end{aligned} \quad (3.21)$$

where r_d is the ground range resolution ($c/2/B_d/\sin \theta$). Equation (3.21) is in accordance with the literature [Curlander, 1991].

In the case of ERS-1, for a backscatter coefficient of -7 dBm²/m² (e.g. for forest cover) and a loss coefficient of 5 dB, $SNR_{ext}^{raw} = SNR_{ext}^{slc} \approx 12$ dB. Using the values in Tab. 3.1a and 3.1b, we can perform similar numerical applications for other Space missions, using standard values for the unspecified parameters (the antenna gain G is calculated from the empirical formula in Sect. 1.2, loss is between 4 and 6 dB, the noise factor F is between 0.5 and 1.5 dB and satellite velocities are all close to $7 \text{ km} \cdot \text{s}^{-1}$).

Lastly, the dependency on R_0 and on λ after SAR synthesis (which is in fact the same as it was before the process) is identical to that described in the preceding section. The rest of this chapter will be devoted to extended targets.

Table 3.2 Characteristics of future civilian SAR Space missions.

	ALOS/PALSAR	RADARSAT-2	TERRASAR-X	COSMO-SKYMED	RISAT	SENTINEL-1 (specifications)
Origin	Japan	Canada	Germany	Italy	India	ESA
Launch date	09/2005	End 2007	06/2007	2007	2006–2008	2008 – 2009
Freq. band (Central freq. in GHz)	L (1.270)	C	X	X (9.6 GHz)	C	C
Polarization	<i>HH</i> or <i>VV</i> (fine, 28) <i>HH</i> + <i>HV</i> or <i>VV</i> + <i>VH</i> (fine, 14) <i>HH</i> or <i>VV</i> (SCANSAR) Full pol. (14)	Nominal <i>HH</i> Full polarization (experimental mode)	<i>HH, VV, HV, VH</i>	<i>HH, VV</i>	Single, Co+Cross, Quad	<i>HH</i> + <i>HV</i> or <i>VV</i> + <i>VH</i>
Incidence angle	8°–60°	20°–58°	20°–55°	20°–55°	20°–49°	20°–45° (IWS) 23° (SW1)–45° (SW6) 20°–48° (extra wide)
Swath width (km)	40–70 (fine mode) 250–350 (SCANSAR)	50 (fine) 170 (standard) 500 (SCANSAR)	10 (spotlight) 40–60 (standard) 100–200 (SCANSAR)	40 (fine) 100–200 (standard) 250 (scansar)	30 km (FRS-1) 30 km (FRS-2) 120 km (MRS) 240 km (CRS)	> 80 km (strip map) > 240 km (interferometric wide swath) > 400 km (extra wide)
Resolution (m)	7–44 (fine, 28 MHz) 14–88 (fine, 14 MHz) 100 (multi-look SCANSAR, odes) 24–89 (full polarization)	3m (ultra-fine) 10 m (fine) 100 m (SCANSAR) 25 × 28 (full polar.)	1–3m (spotlight) 3–15 m (strip map) 15–30 m (SCANSAR)	1 (spotlight) 3 (stripmap)	3 (FRS-1) 12 (FRS-2) 25 (MRS) 50 (CRS)	< 5m × 4m (1 view, strip map) < 20 m × 5 m (interferometric wide swath) < 80 m × 25 m (extra wide, 3 looks)

Continued

Table 3.2 Continued

	ALOS/PALSAR	RADARSAT-2	TERRASAR-X	COSMO-SKYMED	RISAT	SENTINEL-1 (specifications)
$NE\sigma^0$ (dBm ² /m ²)	<-23 (fine, swath<70) <-25 (fine, swath<70) <-29 (polarimetry)	-21	-19	-22	-	<-25 (objective) <-30 dB (wave mode)
Bandwidth (MHz)	14, 28 (fine) 14, 28 (SCANSAR) 14 (full polarization)	100	150 (nominal) 300 experimental mode)	> 100	18.75 (CRS) 37.5 (FRS-2) 75 (FRS-1) 225 (HRS)	100MHz
PRF (Hz)	—	—	2000-6000	—	3000-3500	—
f_d (MHz) (range scale)	—	—	—	—	21 (CRS) 42 (FRS-2) 83 (FRS-1) 250 (HRS)	—
Pulse length (μ s)	—	—	—	—	20	—
Altitude (km) / Inclination ($^\circ$)	691/98.1	798/98.6	660	620/98	609 / polar - sun-synchronous	693 / sun-synchronous
Antenna size (m)	8.9 \times 3.1	15 \times 1.37	4.8 \times 0.8	5.74 \times 1.4	6 \times 2	10 \times 1.33
Peak / Mean power transmitted	25 W / 1 W	10 W / 0.5 W	6W / 1.1 W	4 W / 1.2 W	2880 KW peak (active antenna)	5 kW peak
TM data rate (Mb \cdot s ⁻¹)	240 (fine) 120 or 240 (SCANSAR) 240 (polarimetry)	—	2 x 130	—	640 Mbits/s (X band) Storage : 240 Gbits	640 Mbits/s

3.6 Instrument Noise Equivalent $\sigma^0(NE\sigma^{0inst})$

The $NE\sigma^{0inst}$ is the value for σ^0 giving $SNR_{ext}^{slc} = 1$: this is the weakest of the signatures located above the image noise of thermal origin. According to Eq. (3.21), this gives:

$$NE\sigma^{0inst} = \left(\frac{4\pi R_0}{\lambda}\right)^3 \cdot \frac{Loss \cdot k \cdot T_n \cdot 2v}{P_{av} \cdot G^2 \cdot r_d} = \left(\frac{4\pi R_0}{\lambda}\right)^3 \cdot \frac{Loss \cdot k \cdot T_n \cdot L}{E \cdot G^2 \cdot r_d} \quad (3.22)$$

We can deduce from Shannon's condition that the second and third terms are equal. For ERS-1 for example, $NE\sigma^{0inst} \approx -21$ dBm²/m².

3.6.1 Energy cost for improving resolution

Equation (3.22) helps us to determine the energy required if we wish to improve the resolution while keeping the same $NE\sigma^{0inst}$. Under Shannon's condition, the improvement of the *range resolution* ($r_d \rightarrow r_d/2$), obtained by doubling B_d , also doubles the value of $NE\sigma^{0inst}$. To compensate for this degradation, it is necessary to double the energy content of a pulse ($E \rightarrow 2E$) and consequently to double the mean power consumption $P_{av} = E \cdot PRF$.

Improving *azimuth resolution* requires a reduction in the length D of the antenna by a factor of two. The Doppler bandwidth is doubled, which requires doubling f_a to satisfy Shannon's condition. To protect against range ambiguities (Sect. 3.7.1.), the swath width must be reduced by a factor of two, by doubling the antenna width W . The antenna gain (which is proportional to the product $W \cdot D$) thus remains unchanged. The end result is that the $NE\sigma^{0inst}$ is divided by 2! But at the same time, P_{av} is doubled. Keeping the same $NE\sigma^{0inst}$ reduces the pulse's energy E by half and so avoids increasing the mean power consumption (though at the cost of a swath width reduction).

In Spotlight mode (Sect. 2.5.4.1.), it is possible to avoid reducing the swath width, but the illumination duration must be increased. Azimuth resolution is improved by 'sticking together' adjacent Doppler bands whose width is f_a . This means that all key parameters for the design (antenna size, f_a and mean power) remain the same.

3.7 Impact of image ambiguities on the $NE\sigma^{0inst}$ - total image noise ($NE\sigma^{0tot}$)

The $NE\sigma^{0inst}$ defined in the preceding section refers to the noise in the image caused by the instrument itself (due to the receiving devices and algorithms used to compress the raw data). How much do range and azimuth radar image ambiguities increase the image noise? We shall review the origin of these ambiguities, quantify the increase in image noise and deduce the total image noise ($NE\sigma^{0tot}$). Visually, this noise reduces the contrast of the image, turning true black into grey in the darkest parts of an image.

Thus, it will be estimated on a zone assumed to backscatter zero power to the radar (such as airport runways, calm water, shaded areas, etc.).

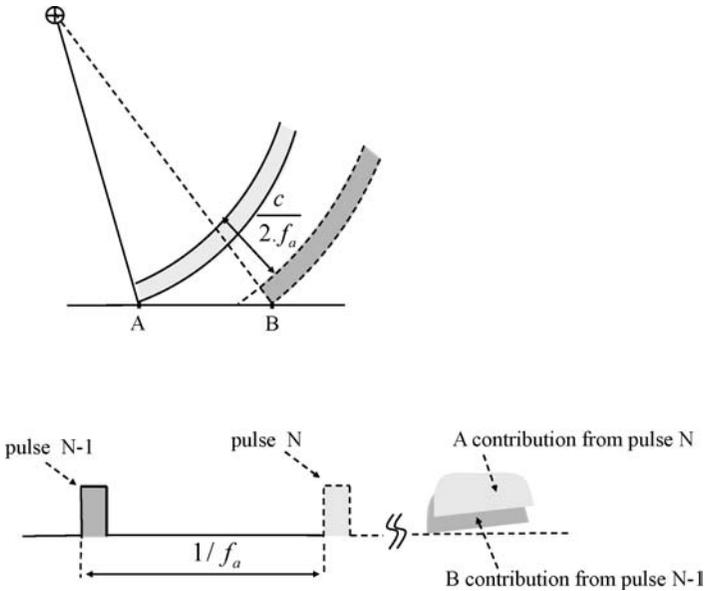


Fig. 3.5 Geometric configuration and timing analysis of range ambiguities.

3.7.1 Range ambiguities

Since the transmission is made up of pulses (at frequency f_a), two echoes arriving at the radar at the same instant, may be separated in range by $c/2 \cdot f_a$, which is called range ambiguity (Fig. 3.5). The antenna width W is chosen so that the main lobe cannot simultaneously contain two targets separated by the range ambiguity. In other words, A and B are not in the same swath⁸ S_w (Fig. 3.6(a)):

$$S_w \approx \frac{\lambda \cdot R_0}{W \cdot \cos \theta} < \frac{c}{2 \cdot f_a \cdot \sin \theta} \quad (3.23)$$

Even though the shape of the main lobe of the antenna protects the signal against ambiguity, the side lobes may favor ambiguous targets. The total range ambiguity noise is characterized by the ratio of the energy returned by the adjacent echoes (rank $N - 1$ and $N + 1$) to that of the useful signal (rank N). The time window of the latter is $2 \cdot S_w \cdot \sin \theta / c + T^{pulse}$.

In the timing analysis, particular care must be taken to avoid superposing the nadir echo onto the useful signal. This is done by adjusting f_a [Elachi, 1988]. In

⁸Equation (3.23) implies an “all or nothing” illumination law for the antenna. In reality, the illumination law is weighted in order to reduce the level of the side lobes, which in turn broadens the main lobe.

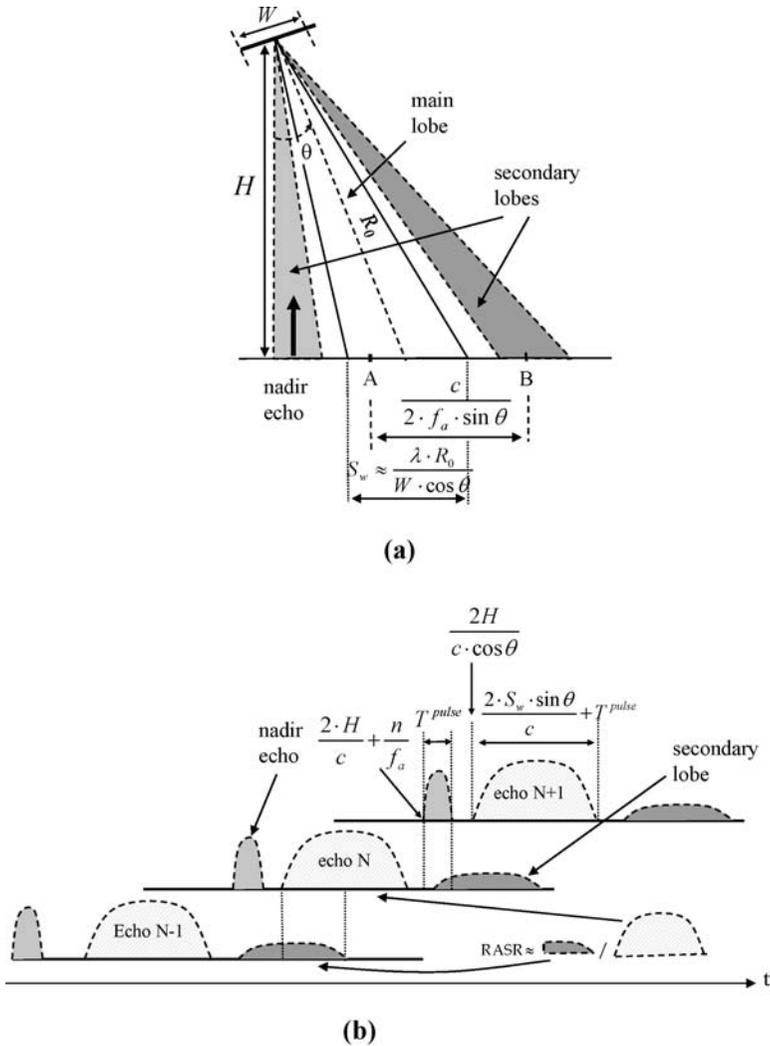


Fig. 3.6 (a) Configuration of the antenna pattern; (b) range ambiguities generated by the side lobes. Range to Ambiguity Side Lobe Ratio (RASR).

Fig. 3.6(b) H is the altitude of the satellite, and the frequency f_a is chosen so that the receiving window for the useful echo (spread between $2H/(c \cdot \cos \theta)$ and $2H/(c \cdot \cos \theta) + 2 \cdot S_w \cdot (\sin \theta)/c + T^{pulse}$) does not overlap that of the nadir echo (which is between $2H/c + n/f_a$ and $2H/c + n/f_a + T^{pulse}$), where n is an integer. The frequency f_a must also be chosen so that the radar does not transmit while the useful echo is being received by the antenna, that is, in an interval between n'/f_a and $n'/f_a + T^{pulse}$, where n' is an integer. These two conditions make it difficult to choose f_a particularly in situations where the imaging radar features incidence agility.

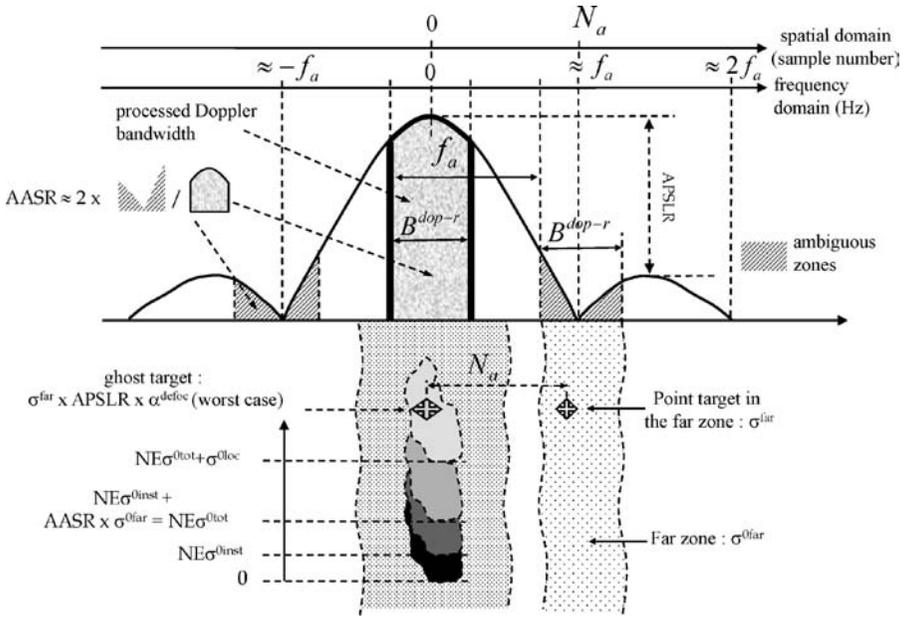


Fig. 3.7 Generation of azimuth ambiguities and the effect of distant radiometric pollution on the image.

The ratio between ambiguity power and useful power is called the *Range Ambiguity to Signal Ratio (RASR)* (Fig. 3.6(b)).

3.7.2 Azimuth ambiguities

Range variations between target and radar during the illumination time T^{ill} produce a frequency spread of the signals received whose maximum extent (Sect. 2.3.1) is:

$$B^{dop} = B^{dop-max} = \frac{2 \cdot v^2 \cdot T^{ill}}{\lambda \cdot R_0} \approx \frac{2 \cdot v}{D} \tag{3.24}$$

Pulse transmission at a rate of f_a here causes aliasing (Sect. 1.4.7): all elements of the signal separated in the frequency domain by an integer multiple of f_a clump together. As long as Shannon's condition is satisfied ($f_a \geq 2 \cdot v/D$), aliasing is avoided if we ignore the side lobes in azimuth. In reality, they widen the azimuth spectrum, consequently generating ambiguities. For processing that includes a pre-summing phase (Sect. 2.4.6) reducing the processed bandwidth to B^{dop-r} , the level of azimuth ambiguities is given by the Azimuth Ambiguity to Signal Ratio (AASR) (Fig. 3.7):⁹

⁹For simplicity, only the first order side lobes are shown.

$$AASR \approx \sum_{\substack{m = -\infty \\ m \neq 0}}^{+\infty} \int_{-B^{dop-r}/2}^{B^{dop-r}/2} G^2(f + m \cdot f_a) df / \int_{-B^{dop-r}/2}^{B^{dop-r}/2} G^2(f) df \quad (3.25)$$

The *AASR* expresses the cumulative effect on the *SNR* of the full range of ambiguous directions (separated in the time domain by an integer multiple of the compression rate N_a). They are weighted by the square of the radiation pattern.¹⁰ The entire azimuth antenna pattern is thus taken into account when determining the impact of the ambiguities. A value of less than -20 dB is generally targeted for *AASR*. To obtain this, we give a specific weighting to the azimuth antenna pattern in order to lower the level of the side lobes, a key contributor to the *AASR*, but this broadens the main lobe (Sect. 1.2.5). Another way of lowering the *AASR* is to increase the value of f_a , while taking the necessary precautions concerning range ambiguities.

3.7.3 Combined processing of range and azimuth ambiguities

It can be seen from Sects. 3.7.1 and 3.7.2 that reducing range and azimuth ambiguities together requires placing limits on the value of f_a (f_a must be reduced to limit range ambiguities and increased to limit azimuth ambiguities). Condition (3.23) together with the Shannon condition ($f_a \geq 2 \cdot v/D$) thus leads to a lower bound of the antenna's surface area:

$$W \cdot D \geq 4 \cdot \frac{v}{c} \cdot \lambda R_0 \cdot \tan \theta \quad (3.26)$$

This condition depends only on the orbit, the carrier frequency ($f_c = c/\lambda$) and the incidence angle. Once these preliminary choices have been specified, the designer still has some flexibility as to the choice of the antenna's actual shape (the swath depends on its width W and the azimuth resolution on its length D), as long as its surface area exceeds a critical threshold accordingly to Eq. (3.26). Most popular designs suggest choices for f_a , the illumination law and the W/D ratio in order to ensure that azimuth and range ambiguities remain at least 20 dB below the useful signals.

3.7.4 Total $NE\sigma^0(NE\sigma^{0tot})$

Total image noise $NE\sigma^{0tot}$ is made up of instrument image noise ($NE\sigma^{0inst}$), increased by range and azimuth ambiguities (Fig. 3.7). In particular, in azimuth, they have the effect of superimposing on useful signatures (σ^{0loc}) those coming from far zones (σ^{0far}), separated from the useful target by a number of azimuth cells equal to the compression rate N_a , and sharing essentially the same range cells. These ambiguous contributions are weighted by the *RASR* (range) and *AASR* (azimuth) rates as previously defined. The $NE\sigma^{0tot}$ is a fundamental parameter for assessing image quality. Instrument design uses standard profiles of radar signatures [Ulaby, 1989], which ensure that $NE\sigma^{0tot}$ is

¹⁰Equation (3.25) assumes that σ^0 is stationary between the useful zone and the ambiguous zones.

typically kept 10 to 15 dB below the expected radar signatures in standard conditions. However, ambiguities can always be increased locally by strong radiometric contrasts ($\sigma^{0far} / \sigma^{0loc} \gg 1$). This phenomenon can be observed for example in coastal areas with calm sea conditions where ghost images of building can be seen in the sea (Fig. 3.8).

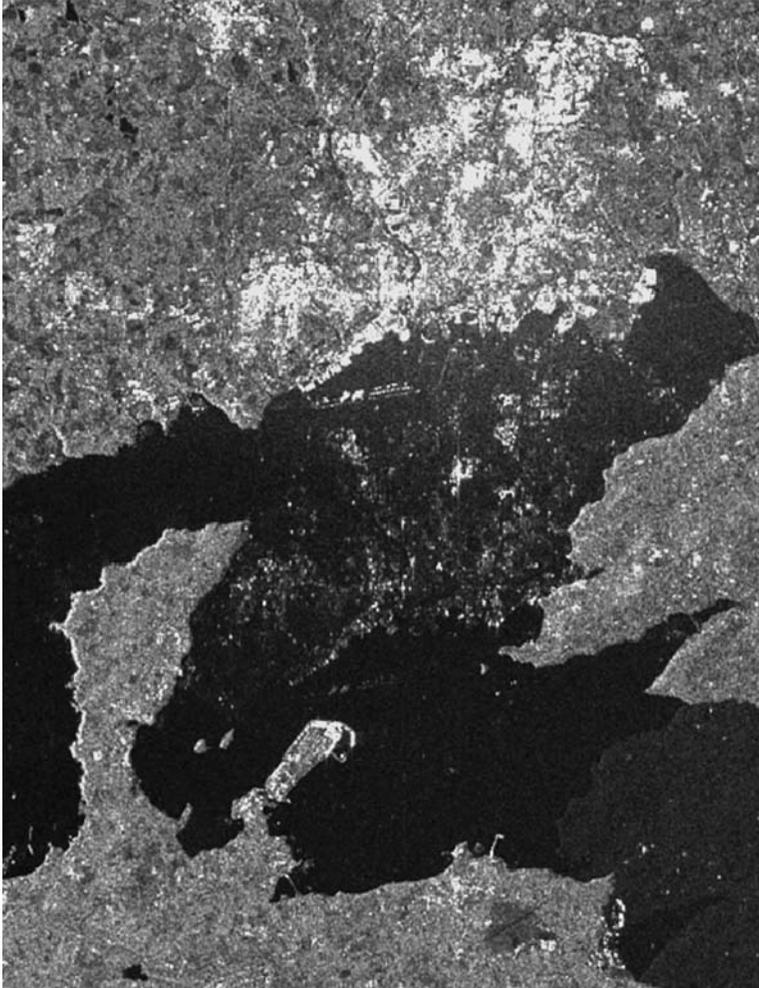


Fig. 3.8 An example of radiometric pollution linked to azimuth ambiguities, due to high radiometric contrast between a calm sea and an urban environment. Standard RADARSAT S6 image, Brest harbor, France. (NB: the impact of the azimuth ambiguities is amplified here because of problems occurring during SAR synthesis.)

We can see from Fig. 3.7 that azimuth ambiguity is reduced when the processed band B^{dop-r} is lower than the maximum Doppler band. In this case, spatial resolution is deteriorated but the radiometric image is ‘cleaner’. Once again there is a

geometry/radiometry trade-off.

There are other sources of noise contributing to $NE\sigma^{0tot}$ linked to the quantification of radar echoes and the bit error rate when transmitting data to the ground. The telemetry aspect of the link budget is addressed in Sect. 3.8, although we will not discuss its impact on the $NE\sigma^{0tot}$.

3.7.4.1 Behavior of point targets

What is the influence of point targets located in ambiguous zones? They produce an azimuth ambiguity located on the same range gate, but separated by precisely N_a azimuth lines (locating a bright point and its ambiguity is one way of estimating N_a accurately).¹¹ Such ambiguities are known as ‘ghost targets’ of which an example can be clearly seen in Fig. 3.18 (for $F_d = -2.5$). The peak amplitude of the ambiguity is reduced in two ways: (1) By the antenna pattern, through the *Peak Side Lobe Ratio (PSLR)*, Sect. 3.12.3; and (2) by resolution degradation due to badly compensated parabolic migration during SAR synthesis (Sect. 2.3.6). The resulting target spreading reduces the peak intensity of the ghost target proportionally. A specific ambiguity rate for point targets is designed by taking into account these effects. This rate is usually around -30 to -35 dB, and is generally less constraining than that for extended targets.

3.8 Volume of data generated onboard

The telemetry data rate is a critical consideration when designing a Spaceborne SAR. It is calculated from the time window of the radar echo (Fig. 3.6(b)):

$$\Delta t_s = T^{pulse} + \frac{2 \cdot S_w \cdot \sin \theta}{c} \quad (3.27)$$

The collected echo, sampled at a frequency f_d , produces an image line made up of $f_d \cdot (T^{pulse} + 2 \cdot S_w \cdot \sin \theta / c)$ pixels. The transmission of pulses at the rate f_a thus generates a number of pixels per second given by:

$$Nb^{pix/s} = f_d \cdot f_a \cdot \left(T^{pulse} + \frac{2 \cdot S_w \cdot \sin \theta}{c} \right) \quad (3.28)$$

Raw data are stored in the mass memory after conversion by an analog-to-digital converter (ADC) which quantifies the real parts (I) and imaginary parts (Q) of the collected signal onto N_q bits.¹² The resulting data output rate from the ADC (in bits/s) is given by:

$$Rate^{board} = 2 \cdot N_q \cdot f_d \cdot f_a \cdot \left(T^{pulse} + \frac{2 \cdot S_w \cdot \sin \theta}{c} \right) \quad (3.29)$$

¹¹ Strictly speaking, the real target and its azimuth ambiguity are very slightly shifted by range. This effect is ignored here.

¹² This quantification stage generates a noise proportional to the number of bits to be quantified. For a typical quantification on $2 \cdot 8$ bits, it is approximately -35 dB (and thus considerably below the usual value of $NE\sigma^{0tot}$).

3.9 Telemetry data rate

Receiving stations are only visible for relatively short periods (approximately 12 minutes for ERS-1, which is on a quasi-circular sun-synchronous orbit¹³ with an inclination of 98° relative to the equator). The receiving station network must be able to receive all the data that the satellite is capable of generating, which depends on its power through the maximum duration for image acquisition per orbit. The storage capacity onboard the satellite¹⁴ helps to manage this flux.

Fig. 3.9 shows the data acquisition and restitution chain. Coding is divided into two parts (source coding and channel coding). Source coding reduces the volume of data transmitted to the ground with the help of compression algorithms adapted to the physics of radar measurements. Channel coding, sometimes called commercial coding (we shall not discuss this here), produces an artificial redundancy which protects the signal from transmission errors or losses. For comparable efficiency (same resolution, same swath width) the telemetry data rates are generally more voluminous for radar systems ($105 \text{ Mbits}\cdot\text{s}^{-1}$ for ERS-1) than for optical systems ($48 \text{ Mbits}\cdot\text{s}^{-1}$ for SPOT4 after image compression). The new generations of satellite (TERRASAR-X and RADARSAT-2) have telemetry data rates of approximately $300 \text{ Mbits}\cdot\text{s}^{-1}$ (the value for TERRASAR-X) for resolutions ranging from 1 to 3 m and swath widths of approximately 30 km (Table 3.2).

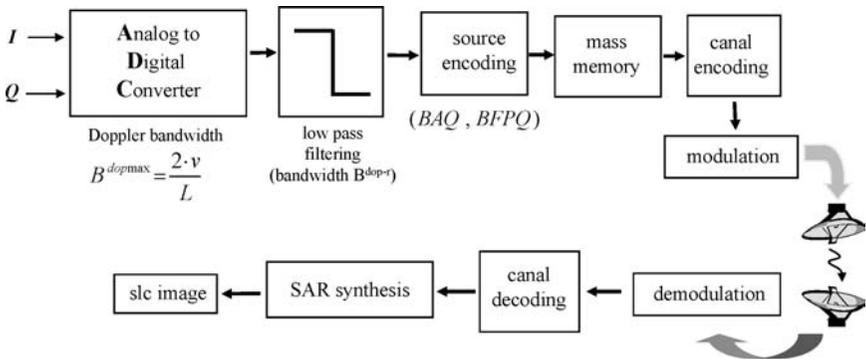


Fig. 3.9 Data processing chain, from onboard sampling of raw echoes to ground segment SAR synthesis processing.

¹³A ‘sun-synchronous’ orbit is one that carries a satellite over a given latitude at the same local time for each pass. Most optical observation missions use sun-synchronous orbits (10:30-22:30), which ensure optimum conditions for image acquisition in descending orbits (daytime). The orbit of the ERS-1 satellite is of this type and it has the same platform as the optical satellite SPOT. For radar missions however it seems preferable to use ‘dawn-dusk’ sun-synchronous orbits such as that of RADARSAT (6:00 - 18:00), which simplifies the management of solar panels (constant orientation) and limits the time in eclipse.

¹⁴Technical advances in solid state mass memories has removed practically all onboard data storage constraints.

3.9.1 Source coding

The object here is to compress the volume of raw data generated onboard the satellite while preserving the information contained. The choice of the coding algorithm depends on the physical nature of the signals. For radar raw data, the natural resolution is very poor (a few kilometers), and thus varies smoothly and exhibits a small dynamic range. They can thus be coded on a lower number of bits than optical data.

At their output from the ADC, radar signals are usually sampled on 2^*8 bits (the real part and the imaginary part). After source coding, the number of bits of the coded signal varies: 2^*5 bits for ERS, 2^*3 bits for JERS, 2^*4 bits for RADARSAT. The ENVISAT chain includes a Flexible Block Adaptive Quantization coder (*FBAQ*) which allows the number of bits for coding to vary between 3 and 8 depending on the SAR's acquisition mode [McLeod, 1998]. For the Magellan mission¹⁵, the telemetry and storage constraints imposed coding on 2^*2 bits: in this radical simplification, the raw signal is reduced to one sign bit indicating positive or negative and one value bit (giving four levels of restitution).

Source coding [Lebedeff, 1995] requires defining a quantifier Q at L levels, which attributes to each sample x found in a 'decision interval' $[t_{l-1}; t_l]$, a 'reproduction level' y_l chosen from a set of L values¹⁶:

$$Q(x) = y_l \quad \text{if: } t_{l-1} \leq x < t_l \quad \text{for: } l = 1, 2, \dots, L$$

with:

$$t_0 = -\infty \quad \text{and: } t_l = +\infty \quad (3.30)$$

The gap between t_{l-1} and t_l is the quantification step. Coding quality criteria, for example the *Root Mean Square Error (RMSE)* $\varepsilon(L)$, are based on various assessments of the difference between the initial image and the restituted image:

$$\varepsilon(L) = \sum_{l=1}^L \int_{C_l} (x - y_l)^2 \cdot f(x) \cdot dx \quad (3.31)$$

where C_l is the set of samples x attributed to the reproduction level y_l and $f(x)$ is the probability density of the variable x .

The coding parameters (decision intervals and reproduction levels) are usually adapted to the variations of the signal statistics, and updated with an appropriate time constant.

Coding involves a reduction of information, and thus contributes to the level of noise, which adds to the value of $NE\sigma^{0tot}$ defined previously. Nevertheless, apart from extreme situations such as Magellan (coding on 2^*2 bits), this contribution to overall noise usually remains negligible compared to the other contributions.

We shall now introduce two algorithms: the BAQ and the BFPQ algorithms.

¹⁵The Magellan mission produced a radar map of 70% of the planet Venus, with 150-m resolution, in 1990.

¹⁶Strictly speaking, source coding is applied to a signal which has already been sampled, but very precisely (2^*8 bits). To keep things clear, we have assumed here a signal whose input is analog.

3.9.1.1 The Block Adaptive Quantization (BAQ) algorithm

This algorithm was developed by the Jet Propulsion Laboratory [Kwok, 1989] for the Magellan mission. It adapts decision levels and quantification reproduction levels locally in order to minimize the RMSE. Under this constraint, the threshold and restitution levels satisfy the Max-Lloyd equation [Max, 1960]:

$$t_l = (y_{l+1} + y_l)/2 \quad (3.32-a)$$

and :

$$y_l = \int_{t_{l-1}}^{t_l} x \cdot f(x) \cdot dx / \int_{t_{l-1}}^{t_l} f(x) \cdot dx \quad (3.32-b)$$

for $l = 1, \dots, L - 1$. The optimum decision levels t_l are positioned at an equal distance from two consecutive restitution levels, the restitution level y_l being the barycenter of the interval $[t_{l-1}; t_l]$. The quantification steps are not equidistant from each other: the BAQ algorithm samples zones with a high density of data more precisely (Fig. 3.10(a)).

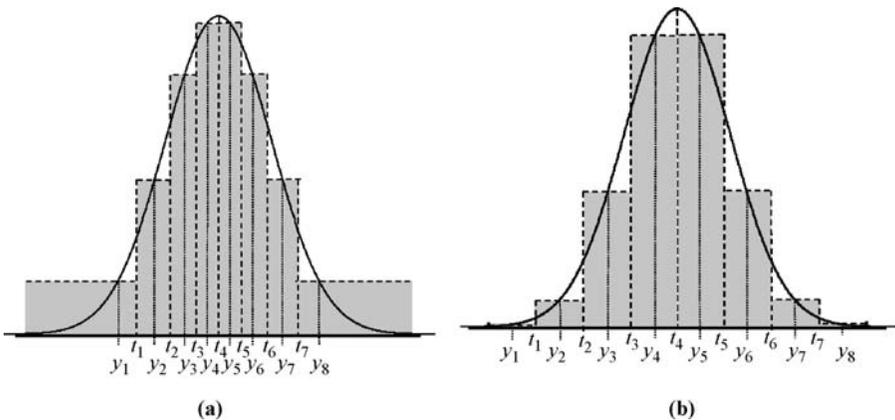


Fig. 3.10 (a) A representation of the decision levels of a 3 bit BAQ; (b) a representation of the decision levels of a 3 bit BFPQ.

Since the statistics for the real part I and the imaginary part Q of the radar signal are known (zero mean Gaussian variables, with variance $\sigma^2/2$), we can compile them into tables (t_l and y_l are given as a function of the number of coding bits (Table 3.2)). Peskova [1999] has evaluated the efficiency of BAQ coding in different cases, whether applied to raw or *slc* images, as well as for Cartesian representations (real part and imaginary part) or polar representation (modulus and phase) of the radar signal.

3.9.1.2 The Block Floating Point Quantization (BFPQ) algorithm

This algorithm (Fig. 3.10(b)) works with a uniform quantization step. For a given number of bits, $\varepsilon(L)$ is minimized by this constraint (Table 3.3). The principal advantage of the BFPQ algorithm is that it is easy to implement, as it requires no more than

Table 3.2 BAQ coding: decision level t_l and restitution level y_l (relative to the standard deviation σ) as a function of the number of coding bits (from [Lebedeff, 1995]).

1 bit / pixel (2 output levels)		2 bits / pixel (4 output levels)		3 bits / pixel (8 output levels)	
t_l	y_l	t_l	y_l	t_l	y_l
0.0000	$0.7979 \cdot \sigma$	0.0000	$0.4528 \cdot \sigma$	0.0000	$0.2451 \cdot \sigma$
		$0.9816 \cdot \sigma$	$1.5104 \cdot \sigma$	$0.5006 \cdot \sigma$	$0.7561 \cdot \sigma$
				$1.0500 \cdot \sigma$	$1.3440 \cdot \sigma$
				$1.7480 \cdot \sigma$	$2.1520 \cdot \sigma$

Table 3.3 BFPQ coding: value of the quantization step (relative to the standard deviation σ) as a function of the number of coding bits (from [Lebedeff, 1995]).

Number of coding bits	Optimum quantization step
1 bit / pixel (2 output levels)	$1.596 \cdot \sigma$
2 bits / pixel (4 output levels)	$1.224 \cdot \sigma$
3 bits / pixel (8 output levels)	$0.996 \cdot \sigma$

integer division (i.e. without taking the remainder into account) by the value of the optimum quantization step, usually rounded to the closest power of 2. The division is thus reduced to a simple bit shift. Since the decision and restitution levels do not conform to the terms of the Max-Lloyd equations, the BFPQ is less efficient than the BAQ.

3.10 Calibration and corresponding image quality requirements

The purpose of calibration is to associate each image pixel with an absolute radiometric content which is directly related to the measurement physics. Many applications (cartography, interferometry, photo-interpretation, etc.) do not however require prior calibration,¹⁷ unlike those which are specifically based on inverting radar measurements into bio-geophysical parameters (for estimating soil moisture, surface roughness, biomass rate, etc.).

The aim of calibration is to correct for spatial and temporal fluctuations in the link budget. To achieve this, using *in situ* standard reference targets with a known radar signature is the ideal situation. In addition, this facilitates later analysis of the impulse response. This procedure, called *external* calibration, completes the *internal* calibration, which compensates for the drift of the instrument transfer functions by using feedback loops applied to the transmitted signals. We shall briefly describe these two steps below before reviewing their requirements.

¹⁷The images from the first Spaceborne SAR mission SEASAT (1978) were never calibrated.

Ideally, each pixel is assigned a complex coefficient $s(x, y)$ characterizing its backscatter signature:

$$E^s = \frac{\exp[-j \cdot k \cdot R]}{k \cdot R} \cdot s(x, y) \cdot E^{inc} \quad (3.33)$$

E^{inc} and E^s are respectively the incident and backscattered fields¹⁸, k is the wave number (Sect. 1.1.2.2), R is the range. In reality, $s(x, y)$ is subject to the combined effects of the instrument's impulse response $h(x, y)$ (Sect. 3.12), the gain K_s and the phase shift ϕ_s of the acquisition chain and the measurement noise $n(x, y)$ (amplified by K_n). The pixel is assigned the quantity [Freeman, 1992]:

$$s^{rest}(x, y) = \sqrt{K_s} \cdot \exp[j \cdot \phi_s] \cdot s(x, y) * h(x, y) + \sqrt{K_n} \cdot n(x, y) \quad (3.34)$$

where $*$ is the convolution operator.

3.10.1 Internal calibration

After SAR synthesis, the power attached to a pixel is written as (Sect. 3.5):

$$P^{final} = P_e \cdot \frac{\lambda^2}{(4\pi)^3} \cdot \frac{G(\theta, \phi)^2}{R^4} \cdot \frac{1}{Loss} \cdot G^{azi} \cdot G^{range} \cdot G^{elec} \cdot \sigma \quad (3.35)$$

G^{elec} is an amplification gain applied to the received signal¹⁹. The purpose of internal calibration is to evaluate K_s , at any point (x, y) of the image. This assumes that we know the product of the transmitted power P_e and electronic gain G^{elec} , the antenna pattern $G(\theta, \phi)$, the range R between the target and the radar, the synthesis gain $G^{azi} \cdot G^{range}$, and the $Loss$ factor.

The product $P_e \cdot G^{elec}$ is controlled by a coupling device between the input circuit and the radar output, which is very stable with respect to temperature and ageing. However as it is difficult to estimate the term for losses $Loss = L^{sys} \cdot L^{fs}$, which combines the attenuating effect of the instrument hardware (L^{sys}) and the loss of propagation in free space (L^{fs}), the internal calibration must be complemented with external calibration.

3.10.2 External calibration

This requires a standard reference target in the image, with a known SER σ^{cal} . If P^{cal} is its pixel radiometry, the SER σ of any target with radiometry P is deduced from the rule of three ($\sigma^{tar} = P \cdot \sigma^{cal} / P^{cal}$). This implicitly assumes that the reference target (also called calibrator) and the actual target are located at the same range from the radar instrument. Rigorous calibration of the image would thus require deploying

¹⁸The corresponding SER σ (Sect. 3.2) is given by: $\sigma = 4\pi \cdot |s(x, y)|^2$.

¹⁹ G^{elec} does not appear in the link budget if we assume to first order that there is an identical amplification effect on the signal and on the noise.

reference targets across the swath width. When determining P^{cal} of the reference point $\sigma(x, y) = \sigma^{cal} \cdot \delta(x) \cdot \delta(y)$, we need to take into account the ‘diluting’ effect of the impulse response [Gray, 1990], which distributes this power over a cross-shaped area centered on the actual target location (Sect. 3.12).

The main factors to be taken into account when selecting a calibrator are: (1) the value of its SER, preferably as high as possible; and (2) the angular width of its radiation pattern—the wider the pattern, the lower the accuracy required for its orientation and the stability of its signature during radar illumination. As a result, the trihedral corner reflector (Fig. 3.2) is the most used standard target for calibration.

The ‘visibility’ of the calibrator is defined by the SCR (*Signal-to-Clutter Ratio*) [Freeman, 1992], which characterizes the power of the signal coming from the calibrator, with respect to that of its environment:

$$SCR = \frac{\sigma^{cal}}{\sigma_{clutter}^0 \cdot p_a \cdot p_g} \quad (3.36)$$

$\sigma_{clutter}^0$ is the clutter backscatter coefficient, p_a and p_g are the azimuth and ground projected range pixel sizes, respectively. A typical value of SCR greater than 20 dB is recommended for satisfactory extraction of the calibrator response. For ERS-1/2, and for a clutter of about $-12 \text{ dB m}^2/\text{m}^2$, a 1-m vertex trihedral corner reflector is sufficient ($SCR = 23 \text{ dB}$).

Please Note: The SCR strongly depends on the SAR observation mode. For RADARSAT (C band), a 1-m vertex trihedral corner reflector observed in fine mode F1 ($p_a \approx 4.5 \text{ m}$, $p_g \approx 10 \text{ m}$) produces an SCR of 24 dB. The same target observed in SCANSAR mode ($p_a \approx 100 \text{ m}$, $p_g \approx 100 \text{ m}$) gives an SCR of 1 dB. Maintaining the SCR at 24 dB requires a 3.8-m vertex trihedral corner reflector.

3.10.3 Calibration requirements and expected scientific results

The requirements for the calibration quality depend on the scientific application in question and the precision sought for the parameters of interest. The calibration coefficient in Eq. (3.34) depends on a series of factors which are subject to temporal and/or spatial fluctuations. There are temporal fluctuations both during an orbital revolution (due to cyclic temperature variations affecting electronic circuits), and throughout the lifetime of the satellite. Platform attitude variations, orbital drift, and drifting of the radiation pattern are all sources of instability. Therefore, calibration requires several space and time scales.

Ulaby [1998] emphasizes the progress made in the calibration field since the 1970s when a quality calibration then meant a relative measurement precision of the order of $\pm 2 \text{ dB}$, and an absolute precision of the order of $\pm 3 \text{ dB}$. To appreciate the progress made, we should note the exceptional radiometric stability of ERS-1/2 with an absolute precision of the order of $\pm 0.25 \text{ dB}$. Calibration is more critical for multi-mode imaging radars, due to uncertainty arising from incidence angle and resolution agility. For example, the standard RADARSAT-1 mode has the following calibration requirements:

relative measurement precision²⁰ within an image: $< \pm 1$ dB, relative precision on an orbit: $< \pm 1.5$ dB, relative precision over 3 days: $< \pm 2$ dB, absolute radiometric precision²¹: $< \pm 3$ dB. Feedback from experience indicates that performances are usually better than the requirements.

Finally, the Table 3.4 gives the calibration requirements for a series of scientific objectives relating to the continental biosphere. Most of them are extremely constraining both for absolute level (of the order of ± 1 dB), and relative level (from ± 0.5 to ± 1 dB for the short term, of the order of ± 1 dB for the long term): an imaging radar is still difficult to use as a quantitative measuring instrument.

Table 3.4 Calibration requirements for several continental biosphere scientific objectives (from [Freeman, 1992]).

Geophysical theme	Required precision	Absolute calibration	Relative calibration (short term)	Relative calibration (long term)
Classification of ice into three classes (winter)	Correct classification rate $> 80\%$	± 2.0 dB	± 0.5 dB	± 2.0 dB
Ice motion	Detection probability $> 95\%$	–	± 0.5 dB	–
Snow cover water equivalent	For layer < 20 cm and for snow mantle depth > 20 cm	± 1.0 dB	± 1.0 dB	± 1.0 dB
Ground moisture (5 levels)	20% of variation per level ($< \pm 0.3$ g/cm ³)	± 1 dB	$< \pm 0.5$ dB	± 1 dB
Surface roughness	–	± 1 dB	–	–
Cartography/vegetation monitoring	Biomass density to within 25%. $0 < LeafAreaIndex (LAI) < 2$ (to within 0.5)	± 1 dB	$< \pm 0.5$ dB	± 0.5 dB
Age of lava	Precise classification of 3 age classes	± 3 dB	± 1 dB	± 1 dB

3.11 Speckle noise and image statistics

3.11.1 Physical origin

The coherent nature of radar illumination (Sect. 1.1) causes the *speckle* effect, which gives the SAR image its noisy appearance, which is uncomfortable for an untrained

²⁰A relative precision of ± 1 dB means that the estimate of the differences of the σ^0 of two pixels will not differ by more than ± 1 dB from the actual value.

²¹The absolute radiometric accuracy characterizes the largest possible difference between the estimate of σ^0 and its actual value.

eye (Fig. 3.11). For example, let us imagine a natural fairly ‘regular’ surface so that the human eye would perceive it as a regular surface (Fig. 3.12(a)). Seen by a radar, each image pixel from this surface contains a large number of elementary scatterers, which add their contributions to the field radiated by the pixel in a coherent way. This addition accounts for the relative phase shift $\Delta\varphi$ arising from their respective ranges from the radar. The total pixel response in amplitude and in phase is the result of vector addition of these contributions in the complex plane.

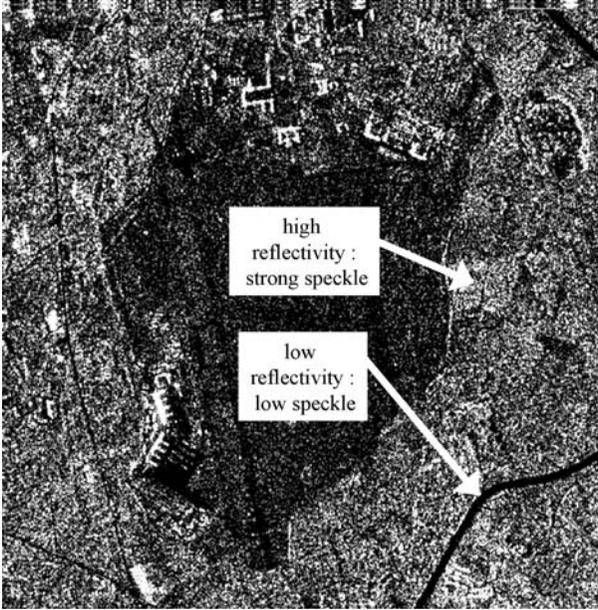


Fig. 3.11 Speckle in a single look radar image. Fine mode RADARSAT extract (C band) of an airport zone (resolution of 5 m in azimuth and 9 m in range). The multiplicative nature of speckle can be seen in the image.

From a statistical point of view, the complex amplitude \tilde{A}^{pix} of the pixel is modeled as the sum of a ‘large number’ N_{diff} of complex elementary contributions $\alpha_k \cdot e^{j\phi_k}$, where α_k is the amplitude of the elementary field coming from the k^{th} scatterer, ϕ_k is its intrinsic phase signature increased by $\Delta\varphi$:

$$\tilde{A}^{pix} = X + j \cdot Y = \sum_{k=1}^{N_{diff}} \alpha_k \cdot e^{j\phi_k} \quad (3.37)$$

Even though adjacent pixels are created by similar conditions at a macroscopic scale, their internal structures at the scale of a wavelength (a few centimeters), differ enough to result in independent phase rearrangements. Differences in range as small as $\lambda/4$, which are equivalent to phase shifts of π reshuffle the phases between scatterers. If, within the pixel, the contributions interfere constructively (e.g. in pixel No. 1), this

pixel will inherit a high radiometry. For the opposite case (destructive interference), the pixel will have a low radiometry (e.g. in pixel No. 4). These uncontrolled radiometric variations create the speckle effect (Fig. 3.12(b)). This mechanism is not random, but is unpredictable and will later be assumed to be a multiplicative noise.

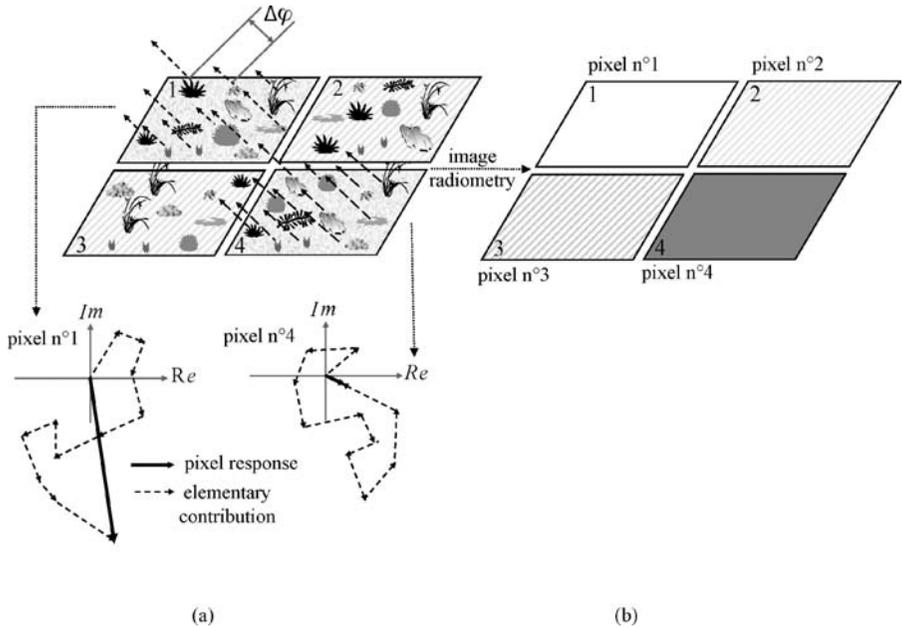


Fig. 3.12 (a) Physical origin of speckle: effect of the coherent nature of radar illumination. The combination of elementary contributions within each pixel; (b) Radiometric distribution of the resulting SAR image.

3.11.2 Statistics of fully developed speckle

The speckle is said to be fully developed [Goodman, 1976] when (1) the responses of each scatterer are independent of the others; (2) the amplitude α_k and the phase ϕ_k are independent; (3) the variables α_k are distributed according to the same probability density function (PDF), in other words, all scatterers produce comparable intensity responses (we are not considering here the specific case of a predominant scatterer within a cell); and (4) The phases ϕ_k are uniformly distributed between $-\pi$ and π . Given these hypotheses, the radiometric distribution of a surface with uniform texture is deduced from Eq. (3.37).

In terms of the central limit theorem, the real and imaginary parts, X and Y respectively, of the complex amplitude \tilde{A}^{pix} are independent random Gaussian variables with a zero mean and the same standard deviation σ . Their respective PDF are given by:

$$\begin{cases} p(X = x) &= \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{x^2}{2 \cdot \sigma^2}\right) \\ p(Y = y) &= \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{y^2}{2 \cdot \sigma^2}\right) \\ E(X \cdot Y) &= 0 \end{cases} \quad (3.38)$$

The PDF of the intensity $I = X^2 + Y^2$ is deduced from Eq.(3.38). Writing $E(I) = R = 2\sigma^2$, one then obtains the PDF of I , knowing R :

$$p(I = i|R) = \frac{1}{2\sigma^2} \cdot \exp\left(-\frac{i}{2\sigma^2}\right) = \frac{1}{R} \cdot \exp\left(-\frac{i}{R}\right) \quad i \geq 0 \quad (3.39)$$

This is an exponential distribution with moments of order m equal to $E(I^m) = m!(2 \cdot \sigma^2)^m$.

The PDF $p(A = a|R)$ of the amplitude $A = \sqrt{I}$ is obtained from the equation:

$$p(A = a|R) \cdot dA = p(I = i|R) \cdot dI = 2 \cdot p(I = i|R) \cdot A \cdot dA \quad (3.40)$$

i.e.,

$$p(A = a|R) = \frac{a}{\sigma^2} \cdot \exp\left(-\frac{a^2}{2 \cdot \sigma^2}\right) = \frac{2 \cdot a}{R} \cdot \exp\left(-\frac{a^2}{R}\right) \quad a \geq 0 \quad (3.41)$$

The amplitude A is distributed according to a Rayleigh distribution²², with moments: $E(A) = \sigma \cdot \sqrt{\pi/2}$ and $E(A^2) = 2 \cdot \sigma^2$, and standard deviation $\sigma_A = \sqrt{E(A^2) - [E(A)]^2} = \sigma \cdot \sqrt{2 - \pi/2}$.

3.11.3 Speckle noise: multiplicative nature and modeling

For a stationary zone of fixed reflectivity R , the speckle effect is created by the amplitude distribution as in Eq.(3.40). When $E(A)$ increases (Fig. 3.13), its RMS σ_A increases proportionally. This proportionality between mean and standard deviation gives speckle the behavior of a multiplicative noise. This can be seen in an extract of a RADARSAT image acquired over an airfield (Fig. 3.11): the calm river, with very low radiometry, has no speckle, unlike the higher radiometry zone, on the edge of the landing strips. Let

$$I = R \cdot n \quad (3.42)$$

where n is the random variable representing *speckle*. We assume in addition that n and R are independent variables and that n has a unit mean:

$$E(I) = E(R \cdot n) = E(R) \cdot E(n) = R \quad (3.43)$$

²²Following Eq. (3.39), the PDF of I is maximum at 0, whereas following Eq. (3.41), that of $A = \sqrt{I}$ is null at 0! This apparent paradox reminds us that we should not confuse a probability density and a probability... The differential form Eq. (3.40) expresses the equality of probabilities of events $A = a$ and $I = a^2$.

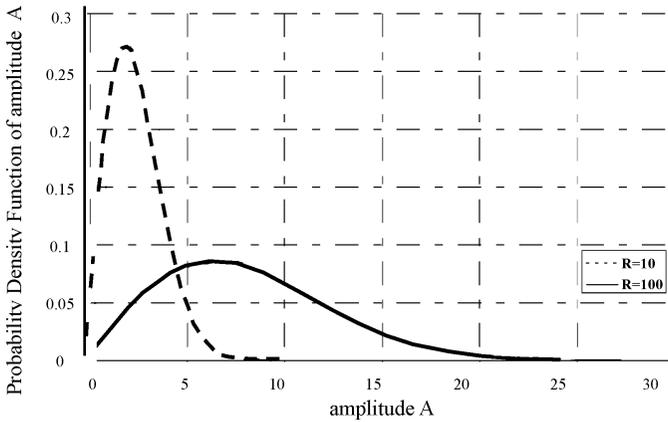


Fig. 3.13 Probability density function (PDF) of the amplitude of fully developed speckle. The mean radiometry value R is 10 or 100.

The equality between $E(I)$ and R indicates the possibility of reducing speckle by averaging nearby pixels, in a trade-off with spatial resolution.

The speckle ‘power’ is quantified by means of the intensity variation coefficient (or radiometric contrast):

$$C_I = \sigma_I / E(I) \quad (3.44)$$

where σ_I is the R.M.S. of the intensity. In the case of a *single look* complex image, $C_I = 1$ for a fully developed speckle. Speckle is reduced when C_I is lower than 1.

3.11.4 Texture effect

In addition to the radiometric dispersion due to *speckle*, the spatial evolution of the landscape causes its own radiometric variations which result in a texture effect. The image of a tropical environment (Fig. 3.14) shows variable texture zones. To account for radiometric variations due to different textures, we introduce a dedicated random variable (t) in the equation for I :

$$I = t \cdot E(R) \cdot n \quad (3.45)$$

We assume that t and n are independent variables and that t is of unit mean, as is n . Under these conditions, the variation coefficients of the intensity C_I , of the texture C_t , and of the speckle C_n are related by:

$$C_I^2 = C_t^2 + C_n^2 + C_t^2 \cdot C_n^2 \quad (3.46)$$

If there is no texture ($t = 1$ and $C_t = 0$), we then get $C_I = C_n$.

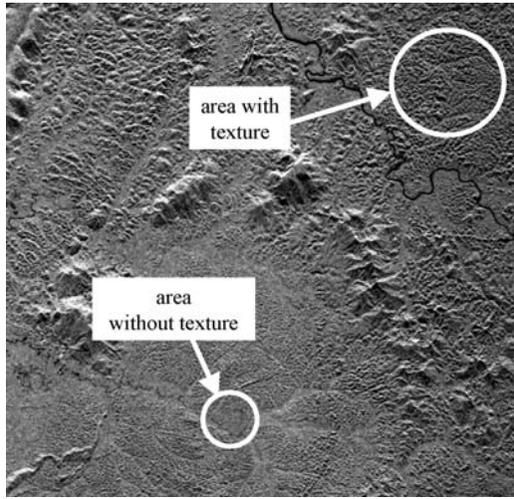


Fig. 3.14 Surface of variable texture. Tropical forest zone observed by ERS (C band, 23°, VV).

3.11.5 Speckle noise in multi-look images

The generation of multi-look images is based on incoherent sums of SLC image samples which reduces its speckle while degrading its spatial resolution. Moreover the phase information is lost during this operation due to the incoherent nature of these additions.

- The ‘spatial’ multi-look associates a unique pixel to any batch of n_a azimuth pixels and n_d range pixels. The intensity of this pixel:

$$I^{ml} = \frac{1}{n_a \cdot n_d} \cdot \sum_{k=1}^{n_a \cdot n_d} I_k \quad (3.47)$$

where I_k is the intensity of the k^{th} pixel. The number of looks is given by $L = n_a \cdot n_d$, if spatial correlation between nearby pixels is ignored. Yet spatial correlation does occur, as the radar signal is always oversampled to some extent.

The intensity variation coefficient for the multi-look image (also equal to the speckle variation coefficient C_n if we assume no texture) is given by:

$$C_{I^{ml}} = C_n = \frac{\sigma_{I^{ml}}}{E(I^{ml})} = \frac{1}{\sqrt{n_a \cdot n_d}} \cdot \frac{\sigma_{I_k}}{E(I_k)} = \frac{1}{\sqrt{L}} \cdot C_I = \frac{1}{\sqrt{L}} \quad (3.48)$$

where $\sigma_{I^{ml}}$ and σ_{I_k} are the RMS of I^{ml} and I_k , respectively. The decrease in the variation coefficient by a factor \sqrt{L} expresses the amount of speckle reduction. The visual effect of speckle reduction based on multi-look and the corresponding deterioration of spatial resolution are shown in [Figures 15\(a\)-15\(c\)](#).

- With equivalent radiometric performances, ‘frequency’ multi-looking reduces the computing time if applied during synthetic aperture processing. For the sake of simplicity, let us consider multi-look processing limited to the azimuth direction (Fig. 3.16). The variations in target-radar range during the illumination time T^{ill} produce a frequency spread in the response (Sect. 2.3.1, Fig. 3.16(a) and Fig. 3.16(b)). After applying a FFT in the azimuth direction, the frequency spectrum is divided into several parts (Fig. 3.16(c)). For each one, we then generate an image (or look) using inverse FFTs. Then the looks are summed incoherently (Fig. 3.16(d)). The effective number of looks equals the actual number of looks as long as their spectra do not overlap.

Please note: The frequency approach is appropriately called ‘multi-look’: the Doppler behavior (Fig. 3.16(a)) assigns to each look the energy in the radar echoes acquired under specific viewing conditions (e.g., backward vision, central vision, forward vision). The multi-look image is none other than a merger of the different looks of a given target.

3.11.5.1 Statistics for multi-look image

For an image made up of L independent looks, the image intensity is distributed according to a *gamma* distribution [Lee, 1994]:

$$p(I = i|R) = \frac{L^L \cdot I^{L-1}}{(L-1)! R^L} \cdot \exp\left(-\frac{L \cdot i}{R}\right) \quad i \geq 0 \quad (3.49)$$

The mean (equal to R) remains unchanged with regard to the *SLC* image, but its standard deviation is reduced by the factor \sqrt{L} .

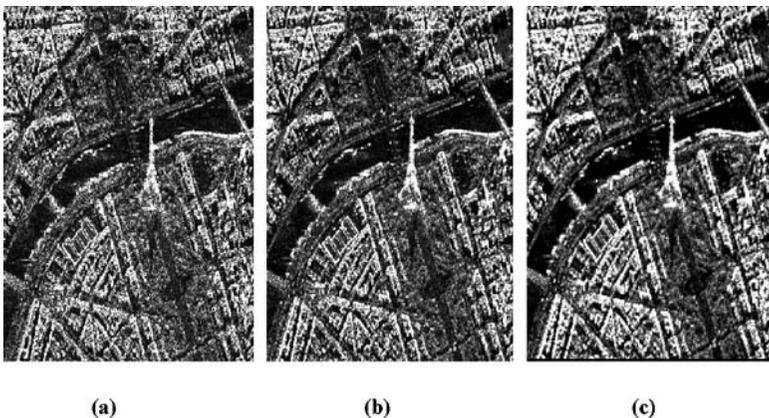


Fig. 3.15 Multi-look processing applied to an image acquired by the ONERA Sethi airborne radar (C band, 3-m resolution) of Paris (zone showing the Eiffel tower and the champ de Mars). (a) original one-look image; (b) 2×2 multi-look images (2 azimuth looks; 2 range looks); (c) 3×3 multi-look images.

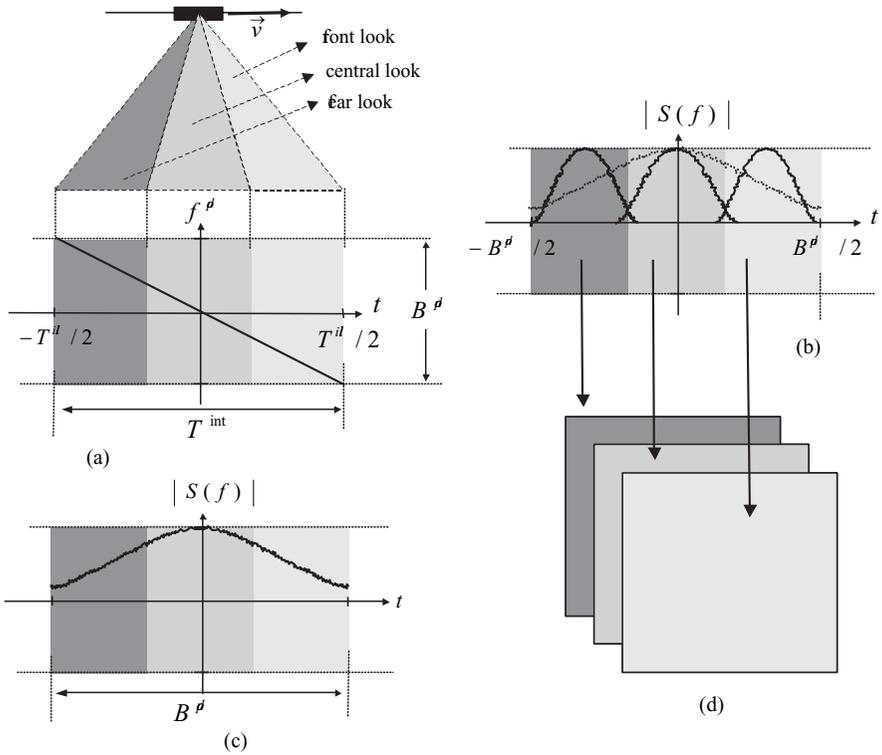


Fig. 3.16 Azimuth multi-look processing: (a) doppler spectrum in the time domain; (b) modulus of frequency azimuth spectrum (c) decomposition of azimuth spectrum into three sub-looks; (d) separate synthesis of the three sub-looks and incoherent summing of resulting images.

The amplitude is distributed according to a *Rayleigh-Nakagami* distribution (or *generalized gamma distribution*):

$$p(A = a|R) = \frac{2 \cdot L^L}{(L - 1)! R^L} a^{2L-1} \cdot \exp\left(-\frac{L \cdot a^2}{R}\right) \quad a \geq 0 \quad (3.50)$$

For a fully developed speckle, the PDF of a multi-look image amplitude (Fig. 3.17) shows that there is less speckle as the number of looks increases (narrowing of profiles around their mean value).

3.11.5.2 Estimating the number of looks in an image

The number of looks L is the image quality parameter which characterizes radiometric resolution Sect. 3.13. To estimate it, we select a homogenous part of the image with no apparent texture ($t = 1$ and $C_t = 0$). According to Eq. (3.48), the intensity variation coefficient gives an estimate of the number of equivalent looks. Subsequently, this estimate enables us to determine the texture coefficient for any part of the image using Eq. (3.46).

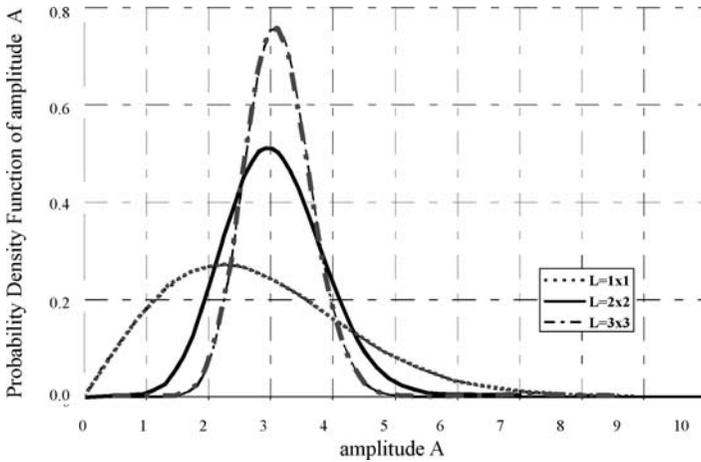


Fig. 3.17 Probability density function (PDF) of the amplitude of fully developed speckle for a multi-look image with a variable number of looks L . The mean radiometry R is equal to 10.

In spite of a simple formulation, the number of looks is a difficult parameter to estimate, particularly if it is low. We recommend renewing the estimation for several homogenous parts of the image. Moreover, the slight oversampling of the image (pixel size $<$ spatial resolution) leads to a spatial correlation between neighboring pixels, which results in a lower number of equivalent looks than expected.

Finally, we should keep in mind that the previous distributions were determined with the fully developed speckle hypothesis. With the advent of high-resolution sensors (≤ 1 m), which make it more likely to have a limited number of scatterers inside a resolution cell (and not the ‘large number’ required for satisfying the fully developed speckle conditions), the formalism described in the previous sections is becoming less adequate. Alternative distributions are then used for the statistical description of radiometry (K , *Weibull*, *Fisher* distributions), depending on the nature of the media observed [Tison, 2004].

3.11.6 Speckle reduction filters

Their function is to produce an estimate \hat{R} of the local reflectivity R , on the basis of a multiplicative noise model. The ideal adaptive filter smoothes most of the homogenous part of an image but on the other hand maintains its singularities, which reveal the presence of unusual structures (point targets, buildings, edges of roads, etc.). A comparison of filtered and unfiltered images invariably leads to a discussion on the relevance of filtering. Even though filtering improves the appearance of images (a very subjective criterion), it does not improve levels of detail seen in the original image: a trained eye is the best anti-speckle ‘filter’ for radar images as it changes neither the statistics nor the image resolution.

Below we shall give the characteristic equations for the most widely used speckle reduction filters without considering the many refinements described in the literature (contour detection, introduction of decision thresholds, etc.) [Lee, 1994], [Touzi, 2002].

- The box filter: The estimate \hat{R} of R is the mean of the intensity $E(I)$ estimated over an analysis window centered on the pixel under consideration:

$$\hat{R} = E(I) \quad (3.51)$$

This non-adaptive filter is optimal for fully developed speckle but cannot be used for point targets or fine structures.

- The Kuan and Lee filters: \hat{R} is modeled as a linear combination of the value I of the intensity at the center of the analysis window and of the mean $E(I)$ estimated over the entire analysis window:

$$\hat{R} = \alpha \cdot E(I) + \beta \cdot I \quad (3.52)$$

The coefficients α and β satisfy a maximum likelihood criterion (minimizing of $E[(R - \hat{R})^2]$) which leads to:

$$\hat{R} = E(I) + (I - E(I)) \cdot \frac{C_t^2}{C_t^2 + C_n^2 + C_t^2 \cdot C_n^2} \quad (3.53)$$

The Lee filter (which does not include the term $C_t^2 \cdot C_n^2$) is a variation of the Kuan²³ filter. Equation (3.53) expresses the adaptive properties of the filter: for fully developed speckle ($C_t = 0$), the filter reduces to a simple box filter. When the statistic becomes locally more complex (texture, point targets, etc.), C_t increases and \hat{R} moves closer to I : the local structures in the image are mostly preserved.

- The Frost filter: we assume that the auto-correlation function of the image satisfies a decreasing exponential distribution. In this context, R is estimated by filtering I with a Wiener filter:

$$\hat{R} = I * m(z) \quad (3.54)$$

The symbol $*$ is the convolution operator and z is a translation in the analysis window. The function $m(z)$ is designed so that \hat{R} is the best estimate of R using a least square criterion. After a few simplifications, this gives [Bruniquel, 1996]:

$$m(z) = k_1 \cdot \alpha \cdot \exp(-\alpha \cdot |z|) \quad (3.55a)$$

²³Elimination of the term $C_t^2 \cdot C_n^2$ implicitly assumes that $1 + C_n^2 \approx 1$, with C_n decreasing as the number of looks increases. In the extreme case of a single look image ($C_n = 1$), this hypothesis is no longer relevant. In practice, this simplification can be used only for a number of looks greater than 3 [Bruniquel, 1996].

where:

$$\alpha^2 = k \cdot C_I^2 \quad (3.55b)$$

k_1 is an fitting coefficient used to correct bias introduced by the filtering (it can be chosen equal to 1 if the data does not have to be calibrated). The weighting distribution envelope narrows as C_I increases, thus giving the central pixel an increased weight. Moreover, the coefficient k offers the possibility of ‘adjusting’ the ‘strength’ of the filtering.²⁴

- **The Maximum A Posteriori (MAP) filter:** Here, the estimate of R requires knowledge of the PDF of R , $p(R)$. It is moreover necessary to make an hypothesis about the PDF of the intensity given the reflectivity $p(I|R)$. According to Bayes law, we then derive the PDF of the reflectivity given the intensity $p(R|I)$:

$$p(R|I) = \frac{p(I|R) \cdot p(R)}{p(I)} \quad (3.56)$$

\hat{R} is the value of R which maximizes $p(R|I)$. The ‘statistics reversal’ (transition from $p(I|R)$ to $p(R|I)$), combined with the search for a maximum explains the name ‘Maximum *a posteriori*’. It is generally assumed that $p(I|R)$ satisfies a gamma distribution. The hypotheses concerning $p(R)$ are more varied. An alternative to the classic hypothesis of a Gaussian distribution is the gamma distribution which has the twofold advantage of being more realistic and which strangely enough simplifies the calculations for determining \hat{R} [Lopes, 1990]. This then leads to the ‘Gamma-Gamma MAP’ filter ($p(R)$ and $p(I|R)$ are both gamma distributed), which provides an analytical equation for \hat{R} :

$$\hat{R} = \frac{E(I) \cdot (\alpha - L - 1) + \sqrt{E^2(I) \cdot (\alpha - L - 1)^2 + 4\alpha \cdot L \cdot I \cdot E(I)}}{2 \cdot \alpha} \quad (3.57)$$

Unlike the previous filters, this estimator is biased (i.e., it does not converge towards the true value of R). L is the number of looks. $\alpha = 1/C_I^2$ is called the local heterogeneity coefficient. If the scene is very homogenous, $\alpha \rightarrow \infty$ and $\hat{R} \rightarrow E(I)$ and Gamma-Gamma MAP behaves as a box filter. Finally, we should note that the MAP filter is flexible enough to integrate various types of distribution laws $p(R)$ for radiometry (*Fisher*, ...) [Nicolas, 2003].

3.11.6.1 Use of speckle reduction filters

A sliding analysis window is necessary to estimate local statistics: $E(I)$, C_I^2 and C_I^4 . The two first quantities are estimated directly from the pixel radiometric contents inside

²⁴Equation (3.54) implicitly assumes that the speckle is uncorrelated which is incorrect as images are generally slightly oversampled. It should be noted here that there is a difference between the correlation of a scene (which is a physical concept) and the correlation of the speckle. The Frost filter takes the first into account while assuming that the second is zero.

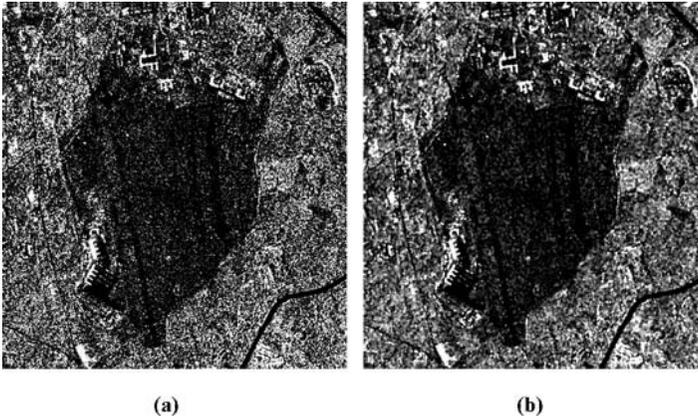


Fig. 3.18 Extract from an over-sampled, fine mode RADARSAT image (SGX) of the Salon de Provence air base. Resolution: 6 m (azimuth) \times 8.9 m (range). Pixel: 3.125 m (azimuth) \times 3.125 m (range). (a) unfiltered initial image (1 look); (b) image filtered with Frost filter (9×9 window). $k = 0.75$.

the window. Once the number of looks L has been estimated (Sect. 3.11.5.2), we then deduce the value of C_i^2 from Eq. (3.44) and Eq. (3.46).²⁵ All of the filter parameters are now known.

By comparing an extract from a RADARSAT image (Fig. 3.18(a)), and the output from a Frost filter (window 9×9 , $k = 0.75$) (Fig. 3.18(b)), we observe a strong speckle reduction over homogenous areas (on the right side of the image) but also a good restitution of the building contours (e.g., central area at the top of the image) thus confirming the adaptive nature of the Frost filter. The performances of the various filters are compared (Fig. 3.19(a) to 3.19(e)), for an extract of the previous image (buildings in the form of combs with variable orientations, at the bottom left). The box filter (Fig. 3.19(b)) flattens the high frequencies of the image (contours of the buildings). After comparison, the Frost filter, the Gamma-Gamma MAP filter, and to a lesser extent the Kuan filter, demonstrate their adaptive properties.

The quality of the filtering may be evaluated by means of various criteria. The *a posteriori* estimate of the number of equivalent looks quantifies the level of speckle reduction. We need to evaluate how much we paid for this reduction in terms of image quality. The trade-off between spatial resolution and radiometric resolution (Sect. 4.5.3) leads to the use of ‘medium’ size analysis windows (e.g., 15×5 for *ERS*). A filter generally comes with peripheral tools which are designed to minimize the negative effects: detectors of contours and structures, preliminary segmentation, higher order statistics, Markov processes, multi-scale analyses, etc. The ability of these filters to preserve textures must be evaluated, using most subjective criteria based on an operator’s perception of the scene observed. Finally, computer criteria such as the

²⁵This type of procedure may yield a negative value due to a limited number of independent samples within the analysis window. In this case, the filter is not applied to the point in question.

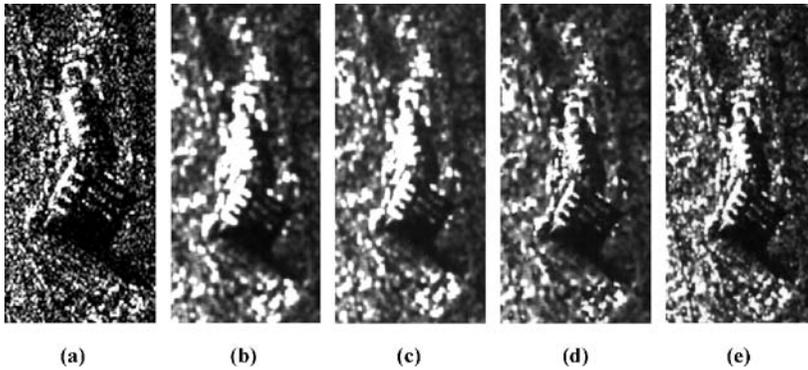


Fig. 3.19 Extract from an over-sampled, fine mode RADARSAT image (SGX) of Salon de Provence air base. Resolution: 6 m (azimuth) \times 8.9 m (range). Pixel: 3.125 m (azimuth) \times 3.125 m (range). (a) initial one-look image, without filtering; (b) medium filter, 7×7 window; (c) Kuan filter, 7×7 window; (d) Frost filter, 7×7 window, $k = 0.75$; (e) MAP Gamma-Gamma filter, 7×7 window.

complexity of the implementation or computing time required have to be taken into account. [Lee, 1994] has made a thorough comparative analysis of all these filters.

3.11.6.2 Speckle reduction by information diversity

The merging of radar information by incoherently summing images acquired at different times (or with different polarizations) leads to a synthetic product with improved radiometric resolution, while preserving the spatial resolution (unlike the multi-look product). This may be seen by comparing Fig. 3.20(a) (ERS PRI image on one date) and Fig. 3.20(b) (merging of eight ERS PRI dates). This approach implicitly assumes the stability in time of the information of interest (e.g. determining building contours, extracting networks, etc.) and *a contrario* the instability in time of the speckle.²⁶ The speckle will be reduced *pro rata* according to the number of available images if we assume that it is totally uncorrelated in time.

3.12 The impulse response (IR)

The impulse response (IR) is a key indicator of the image quality. It characterizes the ‘reaction’ of the radar instrument / processing algorithm system to a point target. The IR describes the dispersion of the energy coming from this single point in the image. In an ideal system, all its energy would be deposited on one pixel.

²⁶Inter-correlation of speckle from one time to another may be taken into account by using a weighted sum of the images (algorithm from [Bruniquel, 1996]).

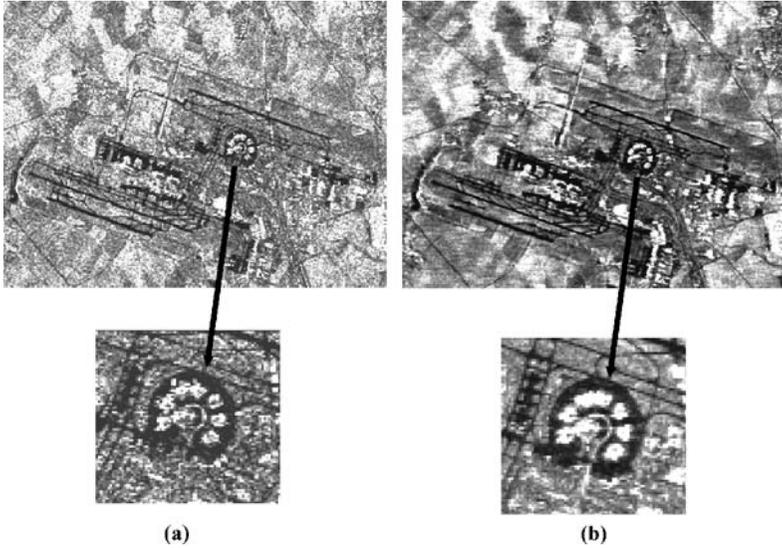


Fig. 3.20 Multi-temporal ERS analysis of Paris-Roissy airport: (a) ERS PRI image corresponding to one date, courtesy SERTIT; (b) Co-registration of 8 dates. In the sub-image, the airport and its terminals, courtesy SERTIT.

3.12.1 Range impulse response (RIR)

The RIR is ‘embedded’ in a single echo (i.e., a single image line). At the time of the transmission, the corresponding pulse is written (Sect. 3.5.1.1.)²⁷

$$s^{inc}(t) = \exp(j\pi B_d t^2 / T) \cdot \Pi_T(t) \quad (3.58)$$

B_d is the chirp modulation band, $\Pi_T(t)$ is gate function of width $T = T^{pulse}$ (uncompressed pulse duration, also called long pulse duration) and unit amplitude. By ignoring the receiving noise, the signal coming from a point target at a range $r_0 = c \cdot \Delta t / 2$ is written as

$$s^{rec}(t) = \alpha \cdot s^{inc}(t - \Delta t) \quad (3.59)$$

where α is a coefficient related to the propagation and to the wave interaction with the target. During reception, $s^{rec}(t)$ undergoes adaptive filtering described in Sect. 3.5.1.1:

$$s^{out}(t) = \alpha \cdot \int_{-\infty}^{+\infty} s^{inc*}(\tau - t) \cdot s^{inc}(\tau - \Delta t) \cdot d\tau = \int_{-\infty}^{+\infty} s^{inc*}(\tau) \cdot s^{inc}(\tau - u) \cdot d\tau$$

with:

²⁷ Such a representation is written here in complex form and does not include the term tied to the carrier, as shown in the expression of Sect. 3.5.1.1.

$$u = \Delta t - t \quad (3.60)$$

After substitution and expanding:

$$s^{out}(t) = \alpha \cdot \exp(j\pi B_d u^2 / T) \cdot \int_{-\infty}^{+\infty} \exp(2j\pi B_d u \tau / T) \cdot \Pi_T(t) \cdot \Pi_T(t - u) \cdot d\tau \quad (3.61)$$

Noting that the non-zero domains of $\Pi_T(t)$ and $\Pi_T(t - u)$ have a non-zero intersection only when $-T \leq u \leq T$, we then get

$$s^{out}(t) = \alpha \cdot T \cdot \exp(2j\pi B_d u^2 / T) \cdot \frac{\sin[\pi B_d |u|(1 - |u|/T)]}{\pi B_d |u|} \cdot \Pi_{2T}(t - u) \quad (3.62)$$

Let r be the range variable and r_0 the range of the point target ($|r - r_0| = c \cdot |u|/2$). Letting p_d be the size of the range pixel ($\approx c/2B_d$), we then get

$$|u|/T \approx \frac{1}{B_d \cdot T} \cdot \frac{|r - r_0|}{p_d} \quad (3.63)$$

As already said, the term $B_d \cdot T = B_d \cdot T^{pulse}$ is the range compression rate. It is much larger than 1. Under these conditions we can neglect the term $|u|/T$ in Eq. (3.62). The shape of the impulse response thus varies according to $\text{sinc}[\pi B_d |u|]$ ($\text{sinc } c$ represents the cardinal sine), or in an equivalent way according to $\text{sinc}[\pi |r - r_0|/p_d]$. The main lobe with a width of $2p_d$, determines the resolution (HPBW or 3 dB beamwidth) around p_d .

3.12.2 Azimuth impulse response (AIR)

Whereas the footprint of the range impulse response is included in one single echo, that of the azimuth impulse response spans N_a echoes (N_a is the azimuth compression rate, Sect. 2.2.5.) By setting the time origin at the point of closest approach to the target and ignoring the effect of range migrations, the part of the signal returned in a given range gate is written as

$$s^{rec}(t_{az}) = \alpha \cdot \text{weight}(t_{az}) \cdot \exp(j\pi B^{dop} t_{az}^2 / T^{ill}) \cdot \Pi_{T^{ill}}(t_{az}) \quad (3.64)$$

B^{dop} is the Doppler bandwidth generated during the illumination time T^{ill} . The function $\text{weight}(t_{az})$ is the weighting of the azimuth antenna pattern. This weighting prevents analytical development. By acting in the frequency domain, the response can be 'unweighted' in order to come up with an impulse response shape in $\text{sinc}[\pi B^{dop} t_{az}]$, or in an equivalent way in $\text{sinc}[\pi |x - x_0|/p_a]$, $p_a \approx D/2$, (where D is the length of the radar antenna) being the best achievable azimuth resolution.

More generally the shape of the IR is affected by natural weighting (such as that from the azimuth antenna pattern) or artificial weighting applied to azimuth and range frequency spectra. Their effects are quantified by means of two parameters, the *ISLR* and the *PSLR*, discussed in the following section.

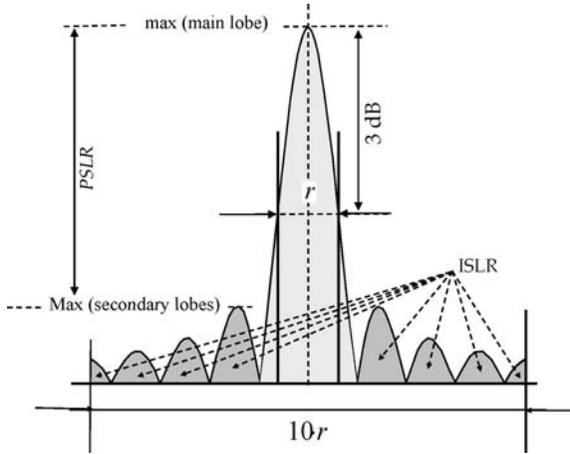


Fig. 3.21 Power distribution of the impulse response in a given direction (azimuth or range). The dark area represents the energy contained in the side lobes, the light area, that of the main lobe. The ratio of the two gives the ISLR. The PSLR is the ratio of the maximum signal level observed in the side lobes to that observed in the main lobe.

3.12.3 Complex image spectrum, ISLR, PSLR, weighting effect

The shape of the IR (expressed as power) is characterized by the Integrated Side Lobe Ratio (*ISLR*) and the Peak Side Lobe Ratio (*PSLR*) (Fig. 3.21). The *ISLR* is the ratio between two quantities: the energy contained in a width of ten resolutions excluding a central band with a width of two resolutions (which models the side lobes), and the energy contained in the central band (which models the main peak). The *PSLR* is the ratio between the maximum intensity peak observed in the side lobes over a width of 20 resolutions and the maximum peak of the central band. These parameters affect the level of radiometric interactions between nearby pixels.

What effect do the azimuth and range frequency spectra of the *SLC* image have on the IR? Analyzing these spectra is part of basic image quality analysis. One particularity of radar imagery is that it has two kind of frequency spectra features, one linked to its amplitude image alone, the other including the entire radar image (amplitude and phase). The latter does not yield landscape characteristics but information on the instrument itself. The envelope of this complex image spectral signature is the aliased antenna main lobe radiation pattern (Fig. 3.22(a)). This envelope can be further modified by applying additional weighting to the data in the frequency domain. This is the case for Fig. 3.22(a). In our example, the complex image frequency spectrum moreover reveals the oversampling applied to the radar signal. The central gap has a width linked to the rate of oversampling (about 60% in azimuth and 30% in range).

The additional weighting applied to the data changes the IR shape. The spatial resolution (see definition in Sect. 3.12) is favored by uniform weighting (Fig. 3.22(b)), for instance obtained by ‘unweighting’ the natural image frequency spectrum. The side lobes of the IR are higher for the ‘unweighted’ case, with the point targets ‘spilling’

more onto their environment. The weighted image offers in return better protection from *speckle* (Sect. 3.11). This type of analysis illustrates again the trade-offs between spatial and radiometric resolution through spectral weighting.

To conclude, the IQ parameters are, simply speaking, the result of a double weighting, one (physical) affecting the antenna pattern, the other (artificial) applied to the radar data. The ‘upstream’ weighting applied by the antenna illumination distribution, affects the radiometric pollution coming from the remote environment, via image ambiguities. The ‘downstream’ weighting, in addition to its effects on spatial resolution, sets the level of radiometric pollution from the nearby environment via the side lobes of the IR.

3.13 Radiometric elements of Image Quality

3.13.1 Estimating and analysing $NE\sigma^{0tot}$

Section 3.7.4 highlighted the joint contribution of the instrument and processing to total image noise $NE\sigma^{0tot}$. This key parameter in image quality is visually tied to the ‘more or less black’ nature of the darkest areas of the image. It is estimated over a surface which is assumed not to backscatter any energy towards the radar such as airport strips, calm water, shadowed areas, etc. However, the image noise is often measured together with a radar signature of the ground, which though weak, may dominate the noise. The estimate of $NE\sigma^{0tot}$ is thus by nature overestimated (hence biased). This bias may be removed by using multiple observations, thus providing independent noise samples and various background situations. An example will be given in the chapter on radar polarimetry (Sect. 5.8).

Fig. 3.23(a) to 3.23(c) illustrate the rendering of a series of urban radar images (X band, single look, 1 m) with varying image noise levels. This example, with a value typical for an airborne acquisition (Fig. 3.23(a), $NE\sigma^{0tot} = -35 \text{ dBm}^2/\text{m}^2$) shows a huge contrast between the brightest zones (buildings) and their shadows. In the case of satellite acquisition from a ‘standard’ platform (ERS type), the contrast is necessarily degraded (Fig. 3.23(b), $NE\sigma^{0tot} = -22 \text{ dBm}^2/\text{m}^2$). Finally preliminary studies conducted on ‘micro-satellite’ platforms (150 kg class) show that $NE\sigma^{0tot}$ increases to a value of approximately $-18 \text{ dBm}^2/\text{m}^2$ (Fig. 3.23(c)). For the last case, it would be difficult to work directly on a single-look image, hence prior multi-look filtering is desirable.

3.13.2 Estimating the ambiguity level

The ambiguity level for extended targets (e.g. *AASR* for azimuth, Sect. 3.7.2) and point targets (*APSLR*, Sect. 3.7.4.1) may be estimated only by taking advantage of specific conditions. For the *AASR*, this means for example finding an extended target whose ‘ghost’ falls in an area with very weak radiometry, and then computing the ratio of its intensity compared that of the real target (e.g., Fig. 3.8). For the *APSLR*, we would need to pair a target and its ambiguity. This particular situation is shown in

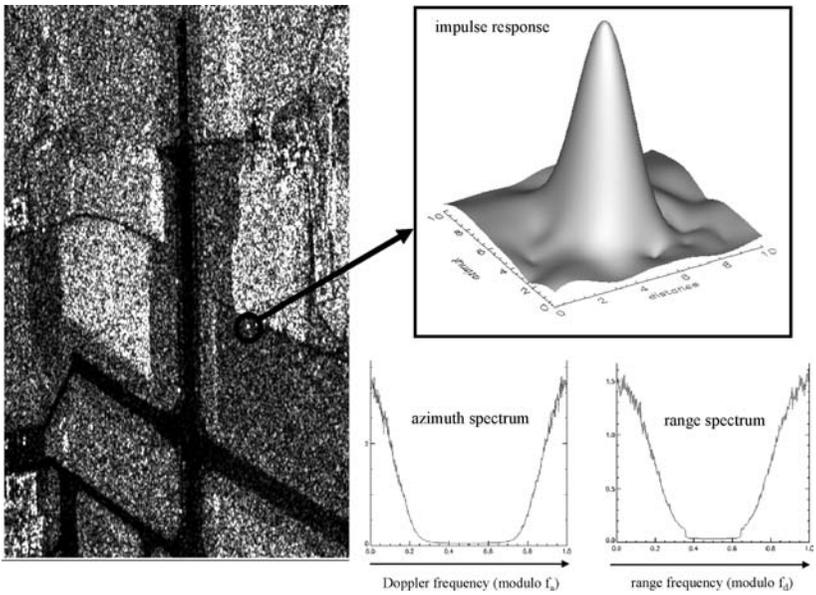


Fig. 3.22(a) One-look radar image. Weighted and over-sampled azimuth spectrum. Weighted and over-sampled range spectrum.

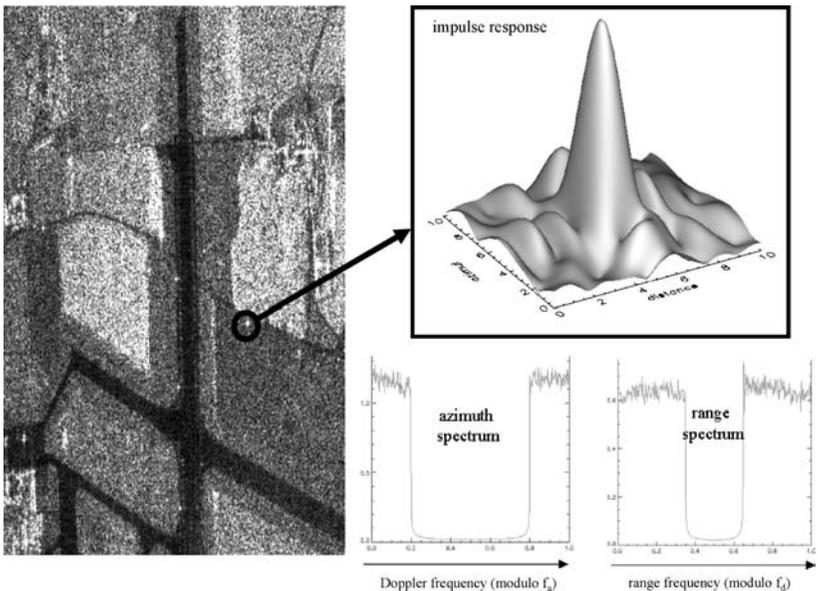


Fig. 3.22(b) One-look radar image. Unweighted (whitened) and over-sampled azimuth spectrum. Unweighted (whitened) and over-sampled range spectrum.

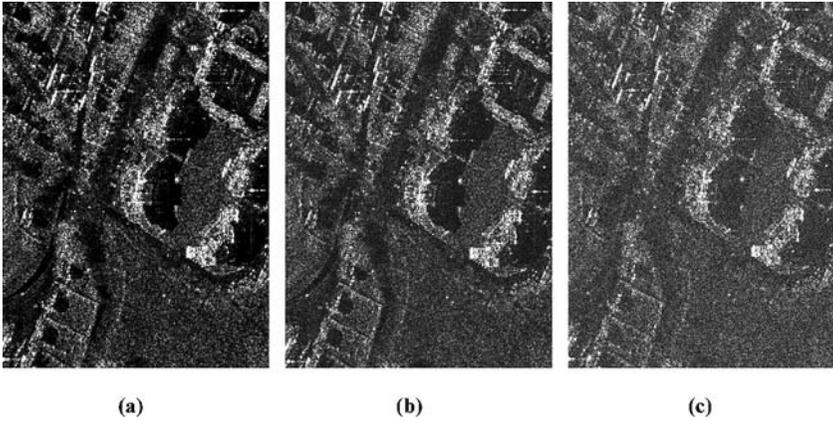


Fig. 3.23 ONERA/RAMESSES one-look radar image, X band, 1-m resolution, including variable levels of image noise ($NE\sigma^{0tot}$). (a) $NE\sigma^{0tot} = -35$ dB m^2/m^2 ; (b) $NE\sigma^{0tot} = -22$ dB m^2/m^2 ; (c) $NE\sigma^{0tot} = -18$ dB m^2/m^2

the example of an ASAR/ENVISAT image (Fig. 3.24) including a transponder (active calibration target) which backscatters enough energy to enable easy identification of its main ambiguities.

3.13.3 Radiometric resolution:

This characterizes the radiometric stability of a stationary part of the image (with constant σ^0 and $NE\sigma^{0tot}$). The radiometric resolution depends on the speckle intensity (Sect. 3.11), i.e. the number of looks L (Sect. 3.11.5.2) and the level of image noise. It is defined by:

$$\gamma = 10 \cdot \log_{10}(1 + SD(\hat{\sigma}^0)/\hat{\sigma}^0) \quad (3.65)$$

where SD is the standard deviation and $\hat{\sigma}^0$ the estimated value of σ^0 . Using a multiplicative speckle model, we show that:

$$\gamma = 10 \cdot \log_{10} \left(1 + \frac{1 + 1/SNR}{\sqrt{L}} \right) \quad (3.66)$$

where SNR is $\sigma^0/NE\sigma^{0tot}$. A one ‘look’ ($L = 1$) SLC image, for which speckle noise is maximum, has a radiometric resolution larger than the value of 3 dB (limiting case for $SNR = \infty$). As the number of looks increases γ also increases while the spatial resolution is degraded, leading to a trade-off.²⁸

²⁸For example, the leading product of the ERS mission, the Precision Image (PRI) is an azimuth 3-look image with a resolution of 20 m.



Fig. 3.24 Radar signature of a transponder and its two main azimuth ambiguities (ASAR/ENVISAT S4 *HH*, of Edam, Holland, Oct. 02). Copyright ESA.

3.14 Geometric elements of image quality

3.14.1 Spatial resolution and pixel size

We should first distinguish between resolution and pixel size. The pixel size (Sect. 2.2.3) only depends on the radar signal sampling. It does not change during standard SAR synthesis, unlike spatial resolution. The latter is defined by the 3 dB cut-off (of energy)

of the main lobe in the Impulse Response (IR) (Fig. 3.21). Due to the slight oversampling of the radar signal (typically 15 to 20% higher than the sampling frequency required by the Shannon condition), the geometric resolution will be greater than the pixel size in proportion. As for IR, it also depends on the frequency weighting applied to the data (Sect. 3.12.3).

3.14.2 Geometric distortion

Range vision, which is specific to radar imagery, causes specific artefacts which have been described in Sect. 2.6.1, i.e. the effects of foreshortening and layover. Fig. 3.25 shows foreshortening, layover and shadow effects observed over a mountainous area.

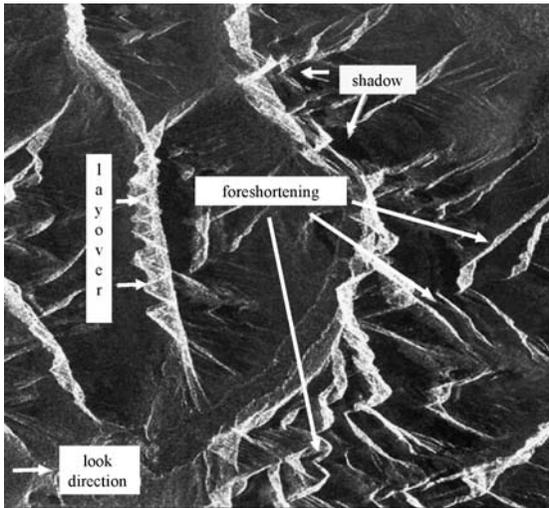


Fig. 3.25 Foreshortening, layover and shadow effect in mountainous area, Copyright RADARSAT International, from *Geology manual RADARSAT*.

Range vision produces another type of artifact, that occurs even for flat ground, which is tied to compression or expansion of an image line. In the range direction the pixel size p_d is given by $p_d = c/2 \cdot f_d$, where f_d is the sampling frequency of the received echoes. Once projected onto the ground, the pixel size depends on the local incidence angle θ_l , i.e., $p_g = c/2 \cdot f_d / \sin \theta_l$. This projection introduces a compression/expansion process for the image which gradually decreases the pixel size when moving along the image line from near range to far range,²⁹ even without significant relief in the topography.

As we have already seen, there is geometric distortion due to the “stop-and-go” assumption (Sect. 2.3.10), which is not as important as the previous effects. It remains

²⁹For ERS (central incidence angle of 23°) the incidence angle along the swath varies from 19° to 26° for flat ground, leading to an additional compression in the pixel size of about 26% in the far range, with respect to the near range. This effect will be greater for an airborne observation, typically ranging between 20° and 50° .

less than a half-pixel in relative value between near range and far range, due to the design constraints related to range ambiguities.

3.14.3 The image location

Due to the range vision principle, the SAR image location is independent of the orientation of the platform. Of course, it still depends on the platform's position. In the geometric positioning equation set developed in Sect. 2.6.2 the range measurement localizes a point on an iso-range curve (the range reference is the near range, associated with the echo sampling-window-start-time, SWST). For azimuth, location is based on time-tagging the transmitted pulses (the time reference is that of the point of closest approach). Ultimately, locating thus depends on two parameters (one for range and one for time) which are subject to various types of errors: imprecise orbit knowledge, lack of knowledge of the terrain's relief, unknown or varying electronic delays, and atmospheric effects. To attain the final location, the time-tagging and the range must be completed by introducing a reference surface for the ground, such as a plane, an ellipsoid, the local geoid, a digital elevation model, etc., which is assumed to contain all of the points to be located.³⁰ Finally we need to remove the residual right/left geometric ambiguity as there are generally two points that satisfy the conditions. However, only one is illuminated by the antenna. Fig. 3.26 illustrates the location problem.

3.14.3.1 Imprecision in range

The range location is determined from the satellite's position on its orbit (in the radial and cross-track directions), and the near range measurement. Orbital accuracy varies according to the direction in question but also depends on the type of orbital product (predicted orbit, instantaneous orbit or filtered orbit). The radial direction (between the satellite and the center of the Earth) has the greatest precision followed by the cross-track direction.³¹

Near range estimates consist of expressing the sampling window start time (SWST) as range information (Sect. 2.2.3). This conversion is affected by the delay of electronic systems during transmission (time elapsed between the transmission command and the effective transmission) and also for reception (time elapsed between reception of the wave and response of the system). These reaction times are cumulative and produce (in the case of ERS), a range offset of approximately 740 m, which is determined empirically during the calibration process. A residual dispersion ε_{NR} tied to atmospheric propagation (smaller than ± 10 m), subsists around this offset value and causes a planimetric location uncertainty (Fig. 3.27) given by:

$$\Delta x_{NR} = \varepsilon_{NR} / \sin \theta_l \quad (3.67)$$

where θ_l is the local incidence angle.

³⁰There are different ways of attaining absolute location (use of ground control points, crossing of ascending and descending orbits) which are not described here.

³¹For example, a filtered ERS orbit offers radial precision to within one centimeter.

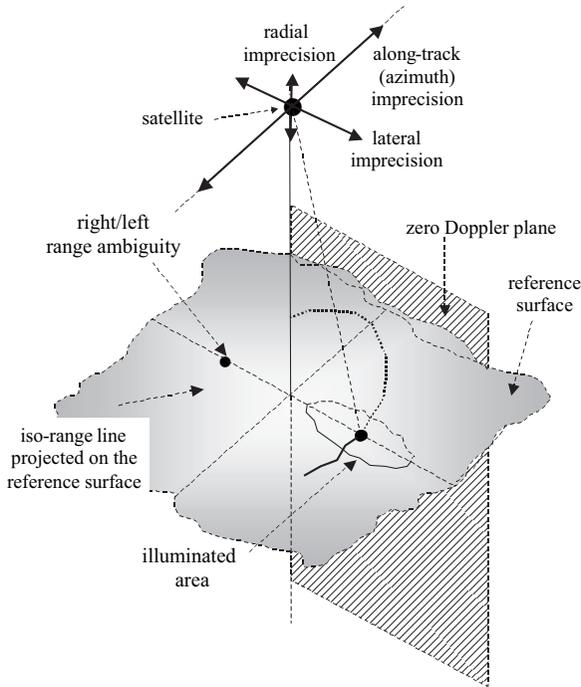


Fig. 3.26 Geometric localization of a point.

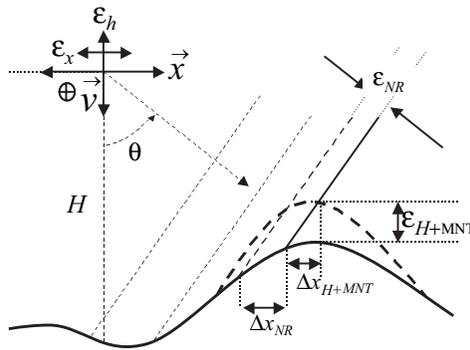


Fig. 3.27 Localization errors related to lack of precision in the orbit, terrain and near range measurements.

Uncertainties linked to the radial position of the satellite (altitude) and to the value of the Digital Elevation Model (DEM) in turn affect near range location on the ground (the wave ‘hits’ the Earth sooner or later than expected). The cumulative uncertainty

ε_{H+DEM} causes a planimetric location uncertainty (Fig. 3.27) given by:

$$\Delta x_{H+DEM} = \varepsilon_{H+DEM} / \tan \theta_l \quad (3.68)$$

For ERS, an altimetric uncertainty of 15 m leads to a mean location uncertainty of 35 m. This error is not uniform for the whole image as the incidence angle amplifies it less at higher incidence angles.

Finally the atmosphere is taken into account in the error budget as a range dilatation effect. Tropospheric effects³² predominate beyond the C band (X , K_u , etc.). Total dilatation while crossing the atmosphere consists of a ‘stable’ component (of the order of 2 m vertically in C band) and an unstable component (of about ± 0.2 m for C band), corresponding to an uncertainty.

3.14.3.2 Imprecision in azimuth

This is governed by knowledge of the satellite’s orbit (in the longitudinal direction, i.e. azimuth axis) and the time-tagging of the image data, from which the acquisition time of the first image line is determined. In addition to orbital accuracy which is critical in the longitudinal direction, offsets between clocks aboard the satellites may come into play. The orbit trajectory is referenced in relation to the platform clock while the radar data is referenced in relation to the payload clock. Finally data acquisition with a non-zero Doppler centroid does not affect the precision for azimuth location as long as one knows the value of the integer part of the Doppler centroid. In these conditions it is possible to reconstitute the image for a zero Doppler, i.e. for the point of closest approach. In this respect, the yaw control of ERS which makes it possible to maintain the Doppler centroid below $f_a/2$, is a good example of a safety margin.

3.15 Radar image interpretation

As we have already seen in previous sections, radar imaging systems must comply with very tight design constraints. The choice of a few parameters is sufficient to determine the whole system. These parameters also govern the particularities of radar images such as ambiguities, the level of saturation and the range and azimuth compression rates. We shall now analyze in detail a set of radar images by imagining that we are analyzing a radar about which we know nothing at all. The purpose of this section is to show to what extent these radar parameters affect image characteristics.

3.15.1 Description of data

The series of images in Fig. 3.28 was produced from a single acquisition by the E-SAR airborne radar belonging to the *Deutsche Zentrum für Luft- und Raumfahrt* (DLR), whose raw data was processed by CNES. The E-SAR radar, which flew on a Dornier

³²Below that, ionosphere effects become significant.

228, has a multi-frequency and multi-polarization capability as well as various possibilities for onboard processing. In this old image (1990) taken when this radar was still being developed, the data were taken in *C* band ($\lambda = 5.6$ cm) with a sampling frequency f_d of 100 MHz and horizontal polarization.

The site observed is close to the Oberpfaffenhoffen airfield in the suburbs of Munich; an important DLR center. The site is of interest due to the variety of natural and artificial targets found within a few kilometers. There are forests, fields, isolated trees, roads and highways, a parking lot which is partially filled, buildings, railway tracks, bridges, fences, an air strip and various infrastructures.

How were several different images obtained from the same radar acquisition? By varying the Doppler processing centroid and selecting only a part of the azimuth spectrum for each processing. Each of the images corresponds to processing one eighth of the available Doppler bandwidth. This radar has a small antenna, about $D = 0.2$ m along the flight path of the aircraft. As the aircraft typically flies at a speed of $v = 80$ m·s⁻¹, the signal is sampled at a frequency of about $f_a = 1000$ Hz (f_a is the pulse repetition frequency or PRF) so that the aircraft will not cover more than half of the length of the antenna between each pulse or, in other words, that Shannon's condition is respected ($v/f_a < D/2$). Here each pulse is transmitted every 8 cm under normal flight conditions, which corresponds to the width of an azimuth pixel. The best resolution that can be obtained from these data is thus of the order of 10 cm in azimuth (i.e. $D/2$) and 1.5 m in range, due to the $f_d = 100$ MHz range sampling. However it is unrealistic to expect this resolution since the antenna observes an angular sector of about 15° ($\approx \lambda/D$ in radians). Even if the observation range is small, for example $R = 5$ km, the width of processing will be greater than 1 km ($R \cdot \lambda/D$), and will thus include more than 10000 azimuth samples. Moreover, the aircraft will take about 15 s to cross the width of the processing zone. The high resolution processing thus requires an enormous amount of computing but more importantly, knowledge of the relative position of the aircraft to within one millimeter (we are working in *C* band with a wavelength of 56 mm) during 15 s ! This exceeds the capabilities of the navigation systems that were used on aircraft at the time. On the other hand we know that if we select an eighth of the azimuth spectrum we decrease the observation time by 8, and the width of processing by 64 using the equations in Chapter 2.³³ Then estimating the aircraft displacement for 1 or 2 s and focusing on a few hundreds of samples become more feasible. Thus, we can create 8 independent images which each correspond to one eighth of the spectrum. The images in Fig. 3.28 are respectively centered on $-2.5, -1.5, -0.5, 0.5, 1.5$ and 2.5 . These dimensionless figures represent the position of the center of the part of the bandwidth processed, expressed as eighths of the total bandwidth. In these units, the azimuth sampling frequency (PRF) f_a is 8 and the bandwidth of the simplified processing is 1. Fig. 3.28 does not include images centered on -3.5 and 3.5 whose quality is even more degraded than for images -2.5 and 2.5 .

The radar acquisition consists of 100,000 echoes, each one containing 2000 range samples. Given the range sampling frequency, the range swath width is 3 km. The

³³The width of the processing N_a is indeed equal to $\lambda R_0 f_a^2 / 2v^2$. By taking into account 1/8 of the Doppler bandwidth it is possible to reduce f_a in the same proportion and to reduce N_a by a factor of 64.

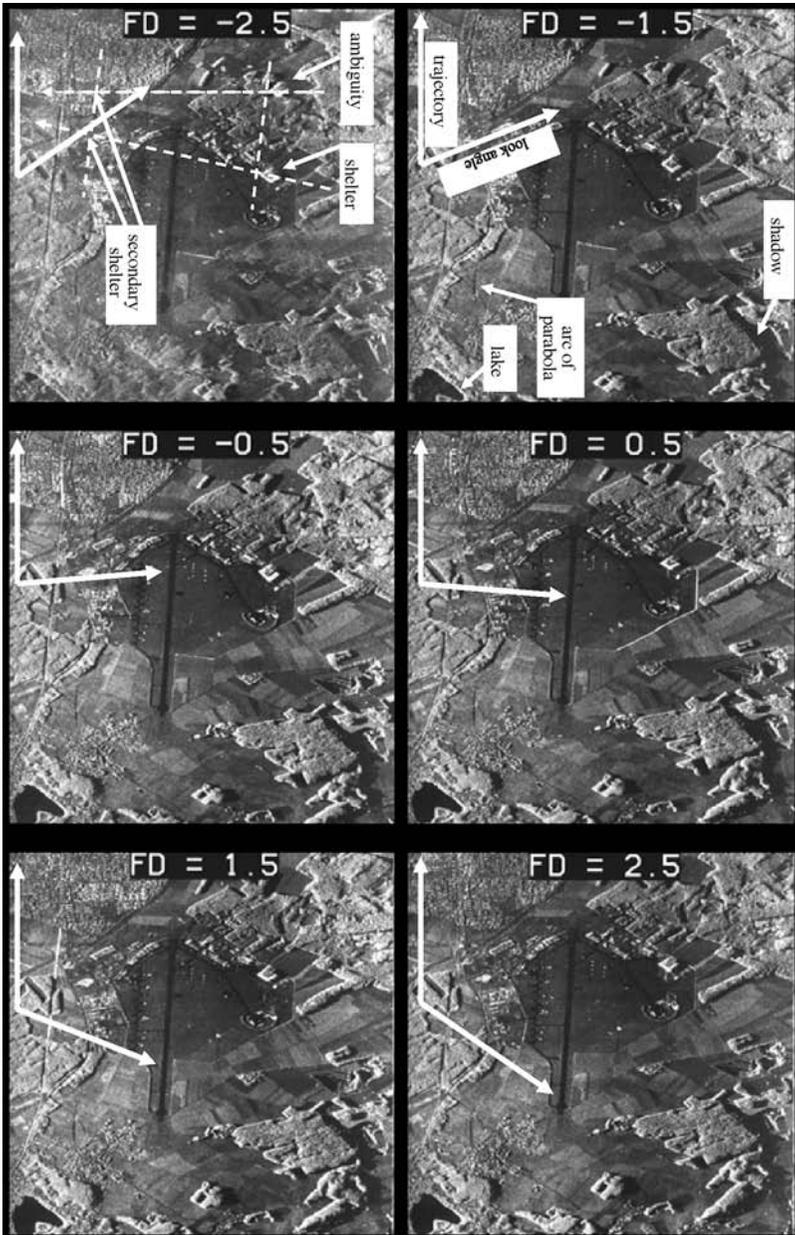


Fig. 3.28 Series of images acquired by the DLR E-SAR sensor over the Oberpfaffenhoffen airfield (Germany). The FD parameter is the Doppler centroid value for the sub-look in question, expressed as heights of the total azimuth bandwidth.

azimuth sampling frequency is actually 1200 Hz. The data thus correspond to about 80 s of imaging, i.e. a little more than 6 km given the aircraft's speed. To facilitate interpretation, each image was multi-looked (Sect. 3.2.5) by merging 2 range samples and 8 azimuth samples. The images, still represented in radar range, thus consist of pixels of 3 m in range and about 4 m in azimuth. Indeed, the elementary azimuth pixel equal to the speed v of the aircraft divided by f_a was first multiplied by 8 due to the frequency selection or 'pre-summing' (Sect. 2.4.6), which only keeps 150 Hz out of the initial 1200 Hz, then again by 8 due to the multi-look filtering. With comparable pixel dimensions in both directions, the images are quite readable, especially since the aircraft is flying at a relatively low altitude; hence the incidence angle is a grazing angle which will be confirmed by the image analysis. This minimizes the effects of geometric distortion in range. Finally, to complete the description of the processing applied to this series of images, the azimuth geometric representation does not use the zero Doppler convention. The points here are not focalized on a position corresponding to the point of closest approach of the satellite since images with variable Doppler centroids cannot be co-registered.

3.15.2 Assessment of the Doppler spectrum

Let us now look at these images in detail. They vary in quality. The indexed image ' $FD = -2.5$ ', i.e., that for which the processed Doppler centroid corresponds to -375 Hz, is the poorest one. It appears to be covered by a fuzzy veil. This loss of quality is related to the high level of azimuth ambiguity. The processing is applied at the same time to a band of 150 Hz centered on -375 Hz and to a band of 150 Hz centered on 825 Hz, which is the first ambiguity repetition when $f_a = 1200$ Hz. The loss of quality is related to the co-registration of distinct images, the real image and the ambiguous image, whose amplitudes are of the same order of magnitude. This co-registration prevents a precise analysis of the ambiguities, unless the target in the normal image is particularly weak, as is the case for the small lake, or if the target which produces the ambiguity is particularly strong as is the case with one of the airfield hangars, whose 'ambiguous' targets are found higher up in the image ('ghost targets'). In the first case, the ambiguity has no competition from the normal target, as a surface of calm water is a 'non-target' for the radar instrument. The landscape lower down in the image (which is outside the area in our images) is thus projected in place of the lake. The 'pollution' from the lake can already be seen in the image ' $FD = -1.5$ ' even though it is affected by a lower ambiguity ratio. In the second case, the enormous power of the normal target (the aviation hangar, a big building probably made of metal and with lots of sharp edges) results in an ambiguity level greater than the level of the normal target which means it can occult the latter. Why do we not find as great a level of ambiguity in the symmetrical image ' $FD = 2.5$ '? Because the aircraft antenna was not observing perpendicular to the velocity but rather towards the positive Doppler signals. If we want to deduce the Doppler centroid from this series of images, we need to identify two images whose levels of ambiguity appear to be similar. The 'hangar' ambiguity can still be seen slightly at ' $FD = -0.5$ '. We start seeing it at ' $FD = 2.5$ ', it is then naturally located on the other side of the hangar, close to the

road lined with trees. If we assume that the levels at ' $FD = -0.5$ ' and ' $FD = 2.5$ ' are equivalent, then the mean antenna Doppler should be located close to ' $FD = 1$ ' (another intermediate image which we did not produce), i.e. close to a value of 150 Hz, which we can of course also calculate from the raw data using (Sect. 2.3.4).

3.15.3 Platform position

More information can be deduced from the ambiguous targets. In Section 2.2.5, we saw that the separation in azimuth between the real target and the ambiguous target corresponds to the processing width³⁴ expressed as a number of samples, N_a . By using the ratio of the length separating the hangar from its ambiguity in the image, to the length of the image and knowing that the image was produced with 100,000 raw data lines, we find that N_a is equal to 24,000. This is the azimuth compression rate that we would have to apply if we had processed the image for high resolution. At the range of the hangar, the rate applied to the data pre-summed by a factor of eight is 64 times lower, i.e. 375. Curiously, the calculation of N_a enables us to determine the position of the aircraft! Indeed, if we identify another ambiguous target closer to the left edge of the image, we could evaluate the variation of the rate across the image, which is essentially proportional to the range, assuming that all other factors remain constant. We find another hangar for which the distance between the ambiguities is only 12% of the length of the image. Since in range the two hangars are located respectively at 21% and 66% of the swath width of the image, we conclude that the aircraft is located at -24% of the swath width of the image. That is where the rate is reduced to zero ... on the antenna!

We therefore know that the image was taken from an aircraft with a near range just equal to a quarter of the image swath width. In fact, the 2000 range pixels in the image correspond to 3 km at the sampling frequency of 100 MHz. The aircraft is thus approximately 700 m from the beginning of the image. Since the near range is not in free space (as we cannot see the ground echo), the aircraft was flying lower than 700 m. Consequently, the image was observed from a grazing angle as confirmed by the length of the trees shadows especially at far range.

3.15.4 Saturation effects

Another effect is shown by the series of images: the saturation of raw data. A dark parabolic arc can be seen in all the images in the series. It is roughly aligned with the flight path and we can see a bright point whose position on the arc varies according to the Doppler centroid selected for each image. This darkening is related to the elimination of all targets sharing raw data with the target represented by the bright point. The shape of the saturated zone indicates that the saturation took place after range compression of the data. Otherwise, the saturated zone would extend not only along the azimuth impulse response but also along the range impulse response. We would have had a vast dark zone extending over a typical number of elementary pixels of N_a in azimuth

³⁴The target and its azimuth ambiguity are also separated by a slight difference in range which is neglected here.

and N_d in range (Sect. 2.3.7). The saturation effect is thus more spectacular when it occurs before compression as it involves a very large surface. On the other hand, it is more difficult to obtain since the saturating target must itself have an amplitude much greater than the usual amplitude expected by the radar for the combination of all targets from this large surface. Space radars are sometimes saturated before compression. To do this purposefully for a satellite, we would have to build a corner reflector whose edges would be about five or six meters long. Of course, the length of the saturated arc corresponds to the length of the impulse response. It is almost identical to that of the second example of the ambiguous target (discussed just above). This is to be expected as the saturating target has the same range as the ambiguous target. The position of the bright point on the dark arc reveals the Doppler centroid used for processing the image. The bright point is located near one of the ends of the arc for extreme Doppler centroids.

3.15.5 Directional effects

Finally, another interesting effect is illustrated by the series of images. We can again observe the high sensitivity of the radar image to variations, albeit slight, in the observation angle. We can thus see that going from image ' $FD = 0.5$ ' to image ' $FD = 1.5$ ', part of the airfield fence which is very bright in one is darker in the other. Inversely, the track of the regional metro line, which can only be detected by the hole that it makes in the landscape in the first image (a cut in a forested area), becomes very bright in the second image. We can calculate the angular difference between the two images. Using, for example, the proportionality between the angular range $\Delta\alpha$ and the reduced azimuth frequency variation Δf_r we have (Sect. 2.3.3) :

$$\Delta\alpha = \frac{f_a \cdot \lambda}{2 \cdot v} \cdot \Delta f_r \quad (3.69)$$

As the reduced frequency in question represents one eighth of the bandwidth at a wavelength of 56 mm and a speed of $80 \text{ m} \cdot \text{s}^{-1}$, we obtain an angular range of only 3° . The image in which the railway line is very visible thus corresponds to an observation with a viewing angle of 4.5° between the normal to the flight path and the radar line of sight (i.e. $\Delta\alpha \cdot 1.5$, since in the image at ' $FD = 1.5$ ', the railway line exhibits the strongest response). This angle is very close to the angle between the flight path and the railway line. The brightness of the track is probably related to a dihedral effect formed by the rails and their supporting base. The angular sensitivity in these images illustrates the difficulty encountered when making a long synthesis (i.e. across a wide angular sector) in order to obtain high-resolution images. The assumption that the target's amplitude is stable during the synthesis is obviously wrong here. The bright contribution of the fence, which is effective for only one eighth of the bandwidth would indeed give an image whose resolution would not be better than eight times the best resolution. Some targets, in particular corner reflectors, might benefit from a complete synthesis because they have a stable response over a wide angular domain.

3.15.6 Is it possible to fully characterize a radar instrument only from an image?

Let us now imagine that we received these raw data without knowing anything about the radar which produced them. We would have understood that the azimuth compression rate needs to be very high and we would have used a massive pre-summing around the mean Doppler determined by applying the algorithm in Sect. 2.3.4. We would also have used an unfocussed processing described in Sect. 2.4.7, which would have provided us (at a range of 5 km and in C band) with a poor resolution of the order of 20 m, nevertheless sufficient for identifying the landscape. The number of samples to be summed in azimuth (about 200) can be deduced by tests. The range compression rate to be applied can be determined from tests or by applying the autofocus algorithm [Jakowatz, 1996]. In this case, the data were already compressed for range which is confirmed by the presence and shapes of the areas subjected to saturation. We could thus have processed the series of images to deduce that they were taken from an aircraft, and also all of their characteristics. Recognition of a few objects in the image whose size is approximately known (air strip, isolated trees, highway apron, parking lots...) would have enabled us to determine the scale of the image and hence the wavelength used as well as the observation range, etc.

References

- [1] J. Bruniquel, *Contribution des données multi-temporelles à l'amélioration radiométrique et à l'utilisation d'images de radars à synthèse d'ouverture*, thesis from Université Paul Sabatier, Toulouse, France, (1996).
- [2] J. C. Curlander, R. N. McDonough, *Synthetic Aperture Radar*, Wiley Series in Remote Sensing, (1991).
- [3] C. Elachi, *Spaceborne Radar Remote Sensing: Applications and Techniques*, IEEE Press, (1988).
- [4] A. Freeman, 'SAR calibration: an overview', *IEEE Transactions on Geoscience and Remote Sensing*, 30(6), (1992).
- [5] J. W. Goodman, 'Some fundamental properties of speckle', *Journal of Optical Society American*, 66(11), 1145–1150, (1976).
- [6] A. L. Gray, P. W. Vachon, C. E. Livingstone, T. I. Lukovski, 'Synthetic Aperture Radar calibration using reference reflectors', *IEEE Transactions on Geoscience and Remote Sensing*, 28(3), (1990).
- [7] C. V. J. Jakowatz, Daniel E. Wahl, Paul H. Eichel, Dennis C. Ghiglia, Paul A. Thompson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Springer, (1996).

- [8] R. Kwok, W. T. K. Johnson, 'Block Adaptive Quantization of Magellan SAR data', *IEEE Transactions on Geoscience and Remote Sensing*, 27(4), (1989).
- [9] D. Lebedeff, *Etude de la quantification vectorielle des données brutes issues d'un radar à synthèse d'ouverture*, thesis presented to the University of Nice-Sophia Antipolis, December 1995.
- [10] J. S. Lee, I. Jurkewich, P. Dewaele, P. Warnbacq, A. Costerlinck, 'Speckle filtering of synthetic aperture radar images', *Remote Sensing Reviews*, 8, 311–340, (1994).
- [11] Ian H. McLeod, I. G. Cumming, M. S. Seymour: 'ENVISAT ASAR data reduction: impact on SAR interferometry', *IEEE Transactions on Geoscience and Remote Sensing*, 36(2), 589–602, (1998).
- [12] J. Max, 'Quantizing for minimum distortion', *IEEE Transactions of Information Theory*, IT-6, 7–12, (1960).
- [13] W. C. Morchin, *Airborne Early Warning Radar*, pp. 75–76. Artech House, (1990).
- [14] J. M. Nicolas, 'La loi de Fisher and ses applications aux images de radar à synthèse d'ouverture', *Proceedings GRETSI 2003*, 3, 209–212, (2003).
- [15] S. Peskova, S. Vnotchenko, 'Analysis of complex SAR raw data compression', in *Proceedings CEOS SAR working group*, Toulouse, France, (1999).
- [16] G. Ruck, D. Barrick, W. Stuart, C. Krichbaum, *Radar Cross Section Handbook*, vol. 1, Plenum Press, (1970).
- [17] C. Tison, J. M. Nicolas, F. Tupin, H. Maître, 'A new statistical model of urban areas in high resolution SAR images for Markovian segmentation', *IEEE Transactions on Geoscience and Remote Sensing* (IGARSS '03 special issue), 42(10), 2046–57, (2004).
- [18] R. Touzi, 'A review of SAR image speckle filtering', *IEEE Transactions on Geoscience and Remote Sensing*, 40, 2392–2404, (2002).
- [19] F. T. Ulaby, M. C. Dobson, *Radar Scattering Statistics for Terrain: Software and User's Manual*, Artech House, (1989).
- [20] F. T. Ulaby, 'SAR biophysical retrievals: lessons learned and challenges to overcome'. *Proceedings of 2nd International workshop 'Retrieval of Biophysical parameters from SAR data for land applications'*, ESA/ESTEC, Noordwijk, Netherlands, October 1998.

CHAPTER 4

SAR INTERFEROMETRY: TOWARDS THE ULTIMATE RANGING ACCURACY

4.1 Principles and limitations of radar interferometry

4.1.1 A specific use for radar

Radar imaging can be used for applications for which optical imaging cannot, because of its two defining characteristics: amplitude and phase. These applications reveal however that, contrary to what we might expect from its traditional ‘all-weather’ reputation, radar imagery is somewhat affected by cloud cover, which is one of the main limitations in applying the technique we are going to discuss here: radar interferometry [Massonnet, 1997], [Massonnet, 1998].

This technique chiefly produces two types of information, usually combined in the same image: topographical information [Zebker, 1986] and information on ground displacement. The results can also include unwelcome contributions due to variations in the state of the atmosphere during image acquisition. It is important to be able to detect these in order to avoid interpretation errors.

We shall explain the principles of interferometry as a technique for comparing radar images, the limitations on its use, and the systematic contributions which have to be eliminated. We shall then look in more detail at topographical calculations using interferometry which, under optimal circumstances with data from current civilian satellites, can attain an altimetric accuracy of better than *one or two meters* on geographical cells with sides of typically 30 m. We shall then look in detail at the calculation of displacements of the Earth’s surface, for which accuracy on the order of 2 mm has been achieved.

We shall explain the practical aspects for obtaining interferometry products from radar data, whether or not the said data are already in image form. We shall also discuss the problem of ambiguity interpretation and how to resolve it, the availability of space-borne or airborne systems, radar-data archives and software-processing resources.

Lastly we will compare interferometry with competing techniques, listing its advantages and disadvantages before concluding with the perspectives for space missions designed specifically for this technology.

4.1.2 A brief history of interferometry

The use of radar signal phase for comparing two images acquired simultaneously goes back to the 1970s [Graham, 1974]. This practice was extended to pairs of images taken from space at different moments at the beginning of the 1980s [Caves, 2002]. The idea of using interferometry for measuring ground displacements was suggested in the middle of the 1980s. It was several years before the technique was actually demonstrated, with varying success, because of the slow rate at which new space data became available at that time.

Current radar data can be processed without loss of quality by standard computers in a very reasonable time, which will be further reduced as processing power increases. Unlike earlier implementation that used analogue devices (such as film), in digital processing, images with calculated phase differences can be used easily and reliably.

The number and quality of civilian radar satellite systems commissioned since the 1990s has allowed coherent techniques to be tested extensively, thanks to the intrinsic regularity of satellite trajectories. At the same time, several airborne systems have partly solved the problems due to the more irregular trajectories of aircraft, by acquiring two radar images simultaneously. With these images, interferometry can be used for topographical applications without worrying about atmospheric artifacts. Some of these systems are currently used operationally.

4.1.3 Interferometry and the physical properties of radar images

Let us quickly review the two techniques used for getting around the low resolving power of radar. The first is range imagery (Sect. 2.2.2), with which it is possible to distinguish separate points on the ground by comparing the time that the radar wave takes to complete the round trip. This wave is sampled at a rate of several tens of megahertz corresponding to a resolution of some tens of meters on the ground. Each sample includes both the amplitude and the phase of the wave. The second technique is Doppler imagery (Sect. 2.3.2) which can distinguish the points along the direction of the radar's flight path by detailed analysis of the phases of the successive radar signals.

The synthesized radar image has two levels of information: (1) amplitude which measures the strength of the radiation, or backscatter, returning from a given point on the ground (Sect. 1.3); and (2) phase which indicates the state of vibration of the wave at the instant that it is received by the radar. The wave comes back in the same state of vibration each time it completes a journey comprising an integer number of wavelengths. The phase therefore contains geometrical information. For example, with the satellite at a range of 1000 km from a bright point and working at a 5 cm wavelength (i.e. a frequency of 6 GHz), the phase would be unchanged after a round trip of 2000 km, covering a very large but integer number of wavelengths. If the point's

range is greater by a single centimeter, those two centimeters of the round trip would imply a phase offset amounting to 40 % of a cycle, which the radar could easily detect. In mathematical terms, radar data is made up of complex numbers (amplitude and phase or, to put it another way, a real part and an imaginary part). The complex nature of radar images is the key factor allowing high resolution images to be synthesized from the low resolution image naturally acquired by the radar (Sect. 2.2.5). What interests us here is the fact that after the synthesizing process, the high resolution image is still complex, meaning that it still possesses amplitude and phase.

The naive range measurement that we have just described is impossible in practice however. One pixel of a radar image usually represents a surface of several tens of square meters, containing numerous elementary targets (stones, branches, etc.). These targets all contribute to the signal and are located at different ranges from the radar. Since the wavelength is much smaller than the size of the pixel, the phase of a given target may be shifted by any value. The combination of these targets further randomizes the phase of the pixel. Clearly, no one value is statistically more significant than another. As far as amplitude is concerned, the contributions of two identical targets found in the same pixel can reinforce each other if their phases are identical or cancel each other out if they are opposite. The summing of the random phase values of these various targets produces the phenomenon known as ‘speckle’ (Sect. 3.11). If the same targets were arranged differently in another pixel they might produce a significantly different amplitude and an unpredictably different phase.

If we want to use the phase of the signal as a measuring technique, the trick is above all to fully comprehend what is meant by ‘random’. A pixel phase is random because we do not know where the elementary targets are located within it. But on a second pass over the same pixel in exactly the same conditions we would of course obtain the same phase. Interferometry depends on the idea that, instead of using the phase of a radar image to measure the ranges, we can use the difference of phase between two radar images to measure differences or geometric distortions in range between these two images. We therefore count on the fact that the complex contribution depending on the particular arrangement of elementary targets in each pixel will be cancelled by combining the two images.

4.1.4 Phase difference between radar images

The phase of a radar image is therefore only meaningful when compared with that of a second one. The two images which are merged point by point to form the image of their phase differences, called an *interferogram*, usually have different viewpoints because they were not taken from exactly the same place, and a time shift because they were not taken at the same moment. These two differences are almost always found in an interferogram. There may be no time difference in the case of systems with two radar antennas which can create two images simultaneously. In certain circumstances there may be no difference in viewpoint, if a satellite repasses at almost exactly the same point when acquiring the second image. These two differences are the source of the two types of information provided by interferometry. The difference of viewpoint provides the topographical information in the interferogram. The time difference

provides information on displacement. Each of these differences also limits the way this technique can be used.

In practice, if we assume that the landscape is described by a digital elevation model (with a regular grid of elevation values), we can determine the Cartesian coordinates for each point relative to the center of the Earth. This operation requires no particular knowledge because it is based on a geographic ‘black box’ (a sub-routine). Once the coordinates have been determined in this way, we can define the trajectory point and the time when a satellite is the closest to each point of the landscape. Here again we are using knowledge in the form of a sub-routine ‘borrowed’ from orbit determination specialists. For each landscape point we can then calculate the range of the round trip when the satellite is at its point of closest approach and scale it to the wavelength. We then consider the fraction part of the result as the phase that the signal ought to have. This phase difference is then predicted point by point. In parallel, we obtain the experimental value of this phase difference by calculating the difference of the phases of the radar images acquired by the satellites.

4.1.5 Robustness of the phase signal in interferometry

One of the strong points of interferometry is the way coherent signals can ultimately dominate incoherent signals. Phase basically acts as a sign change of the magnitude of a physical signal.

When an incoherent signal such as noise (phase ϕ_n uniformly distributed between 0 and 2π (Fig. 4.1(a)) is combined with a coherent signal with the same amplitude ($SNR = 1$), the standard deviation of the phase difference between the resulting signal and the coherent signal (i.e., the phase error ϕ_t) is only $\pm 13\%$ of a cycle. When the amplitude of the incoherent contribution is equal to one half of the coherent contribution (giving an advantage of only 6 dB), the standard deviation becomes $\pm 6\%$ (Fig. 4.1(b)). If the incoherent perturbation has an amplitude of 10%, the standard deviation on the error drops to $\pm 1\%$ (Fig. 4.1(c)). These results are characteristic of the noise effect at the scale of one image pixel: summing samples in which coherent and incoherent contributions are mixed always results in a reduction of this phase error.

In interferometry it is necessary to bear in mind the ultimate size of the pixel or geographic cell that we are aiming to acquire. If we wish to move from an elementary pixel of 30 m^2 to a horizontal sampling of the $30\text{ m} \times 30\text{ m}$ DEM, a surface 30 times larger, the ratio between coherent signal and incoherent signal will increase by a factor of $\sqrt{30}$, after coherent summing of the signals associated with each pixel. A geometric interpretation of this result can be found in Section 3.5.

4.1.6 Limitations due to geometric and surface changes

There should not be too great a difference between the viewpoints of two images. Let us take two distinct targets A and B (Fig. 4.2), located at opposite edges of the same pixel in a reference image (one near the radar, the second further away). Targets A and B are indistinguishable within the pixel. Any elementary target in the pixel is subject to phase variations when passing from one image to another. These two images are

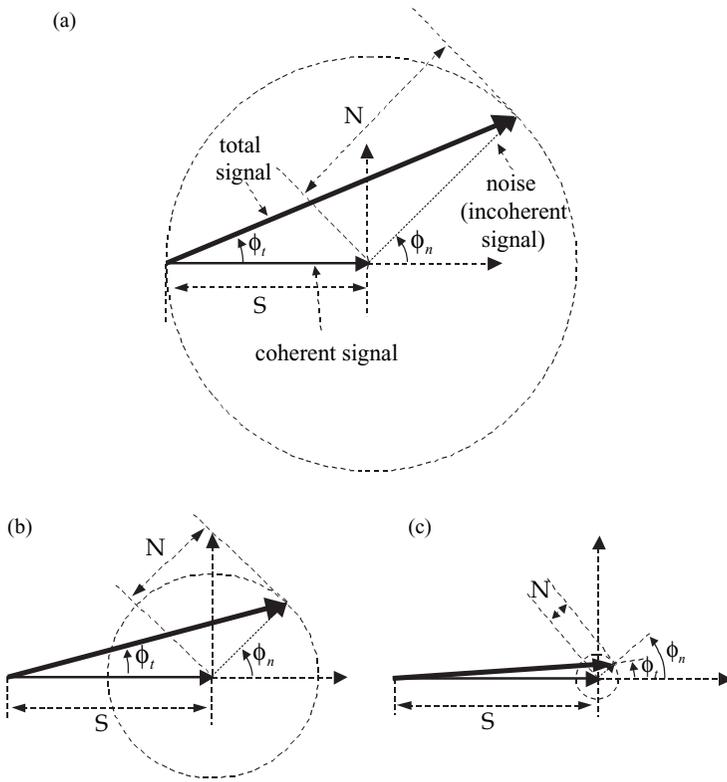


Fig. 4.1 Phase shift produced on a coherent signal by a noise. (a) $S/N=1$; (b) $S/N=0.5$; (c) $S/N=0.1$.

known respectively as the master (M) and slave (S). The phase difference of the same pixel in M and S should not depend too much on the position of the target within the pixel. For instance, the phase difference itself should vary by much less than a full cycle between A and B. The overall phase difference resulting from the mixing of points P at various locations in the pixel will be significant at the scale of the pixel as long as the difference in δ (called the horizontal baseline) remains less than a limiting value. If this is not the case, the phase difference between the two images will again be the result of contributions which are random since they can vary within the pixel itself by more than a cycle. It will then be impossible to exploit this phase difference.

For the limiting case where the targets are located at opposite edges of the pixel, the stability of the phase difference will be guaranteed by the stability of the incidence angle of the wave between the two images. Should this change too much, the path difference between the two targets in one image will differ from the path difference in the other image by more than a wavelength, resulting in a pure random difference of phase. For example, a 10-m ground pixel observed from an incidence angle of 30° implies a round trip path difference of 10 m between two targets at opposite edges of the pixel. If we wish to limit the variation of this path difference to a fraction of the

in incidence angle between images must be very small, because there are generally several hundred wave fronts in a pixel (and not just a few as drawn). Their number is equal to the factor Q (Sect. 2.1.3).

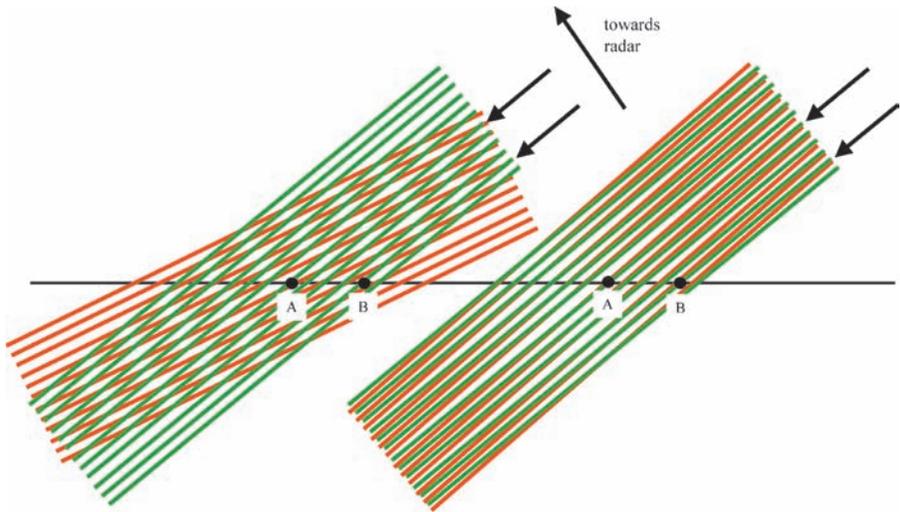


Fig. 4.3 Fringe effects at pixel level as a function of the difference in viewpoint.

4.1.7 Eliminating systematic contributions

Even if the interferometry condition described in Sect. 4.1.6 is satisfied, which results in less than a phase cycle being produced across a pixel, the slight difference in incidence angle between the two images will show a progressive shifting of many cycles across the swath width from one image to the other (Fig. 4.4). For example, even if this difference only creates one tenth of a phase cycle on one pixel, it will create an entire cycle every ten pixels and several hundreds of cycles across the swathwidth of an image. Interference fringes of this type, called ‘orbital fringes’, can be modeled and eliminated using knowledge of the trajectories. Uncertainties about these trajectories can still leave a few fringes even after this correction (Fig. 4.5). These fringes are used for refining knowledge of the trajectories in the same way that mirrors are aligned in an optical interferometer (Sect. 4.2.3).

4.2 Implementing interferometry processing

The processing stages are (1) synthesizing radar images from raw radar data (Sect. 2.2), alternatively, already-synthesized images (SLC) are available on the market; (2) point by point registration of the two images to be combined by appropriate resampling; (3) calculating the phase differences; and (4) finishing tasks which include fine adjustment to the orbit and projecting the product into the desired map geometry.

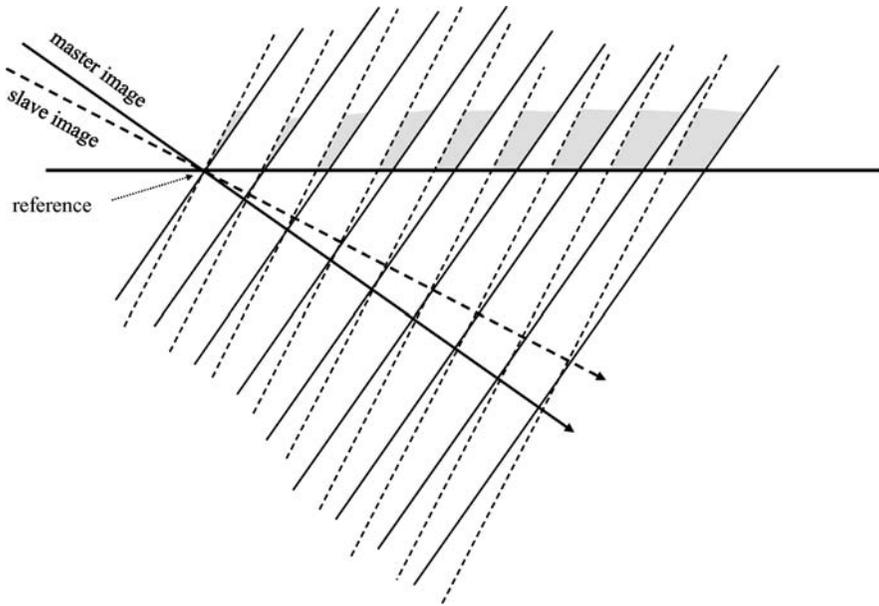


Fig. 4.4 Progressive shifting of a slave image relative to a master image. The grey areas represent the shifting in ground projection.

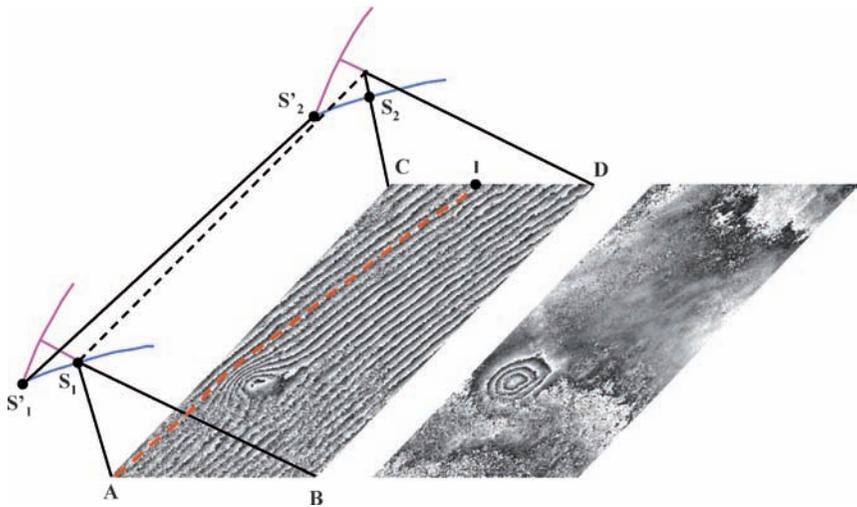


Fig. 4.5 Better trajectory determination using residual fringe count. The topology of the fringes in the left hand image (quasi-linear fringe array) is characteristic of orbital fringes.

4.2.1 Radar image co-registration

There are two steps in image co-registration. First comes the diagnosis: the geometrical difference between the images must be modeled very precisely, at the ‘sub-pixel’ level, which implies that matching between corresponding pixels must be almost perfect. Only the properly matched portion contributes to the coherent combination; the remainder behaves like noise. How important is the quality of the co-registration? Fig. 4.6 shows how interferometric coherence (Sect. 4.2.2.) varies as a function of the relative shift between the two images. Compared to optimum co-registration with maximum coherence, a registration error of 0.4 pixels causes coherence to drop by 10 % whereas an error of 0.8 pixels reduces coherence by half.

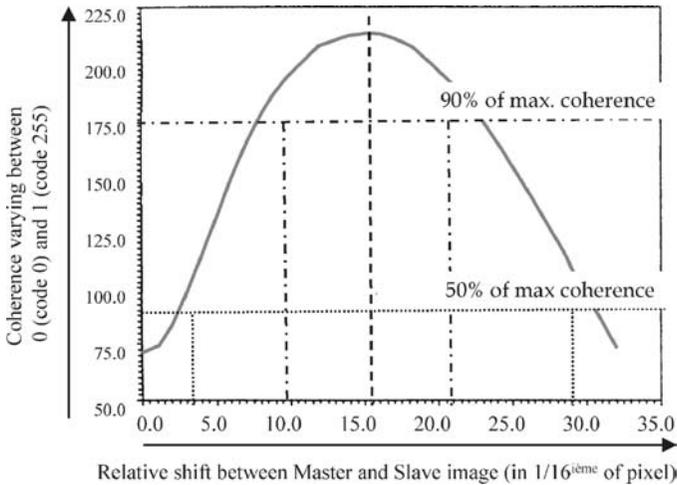


Fig. 4.6 Evolution of coherence as a function of the precision of co-registration. ERS images from the Dniepr basin (Ukraine).

It will then be necessary to stretch one of the images so that it can be registered on the other. This operation must preserve the phase content of the deformed image. Otherwise, bias would be introduced in the interferometric measurements.

4.2.2 Calculating phase differences between images

On co-registered images, calculating phase differences consists in forming the complex quantity $I(m, i)$:

$$I(m, i) = \frac{\sum_{cell} M(m, i) \cdot S^*(m, i) \cdot \exp(j2\pi QG(m, i))}{\left(\sum_{cell} |M(m, i)|^2\right)^{1/2} \cdot \left(\sum_{cell} |S(m, i)|^2\right)^{1/2}} \quad (4.1)$$

M (for Master) and S (for slave) represent the two radar images whose pixels are indexed m and i for line and column, while cell is the analysis window, that is, the number of adjacent pixels included in the averaging.

The phase image of $I(m, i)$ or the interferogram itself is usually coded on eight bits. The phases are multiplied by $256/2\pi$ in order to benefit from the entire dynamic range available. The function $G(m, i)$ summarizes what is already known about range stretching. It incorporates the model of the geometrical difference between the images, possibly including a digital elevation model (DEM) if one is available. It is expressed in phase by multiplying by $2\pi Q$ (with $Q = f_c/f_d$, Sect. 2.2.3). $G(m, i)$ is used to remove the majority of the phase variations in the image and to ensure that a complex mean can be properly calculated (i.e. the phases only rotate very slightly inside the analysis window).

A very useful by-product of this interferometric calculation is the coherence image, which is the modulus of the complex image $I(m, i)$, called the ‘interferometric coherence’. In practice this product is multiplied by 256 to take full advantage of the eight-bit coding given that this modulus is between 0 and 1. The coherence is a self-validating indicator of the phase measurement, which depends on the proportion of ‘useful’ signal to ‘non-useful’ signal. Intuitively, if neighbor pixels “agree” on an average phase value, it is proof that the phase is not random. In this case, the complex values add efficiently, and the coherence is high. If σ_s is the standard deviation of the signal’s amplitude and σ_n the standard deviation of the noise amplitude, the uncertainty ε of the range measurement by the phase is of the order of:

$$\varepsilon = \pm \frac{\lambda}{4\pi} \cdot \left(\frac{\sigma_n}{\sigma_s} \right) \quad (4.2)$$

assuming that σ_s is much larger than σ_n (λ represents the wavelength).

The coherence is mainly destroyed by the modifications of surfaces between acquisitions, but it may also indicate some amount of “volume scattering” (i.e., targets with a radar penetration depth such as forests), which is more demanding in terms of incidence angle variation between images. In this case, coherence may become a physical measurement by itself (Sect. 5.9.1.1).

4.2.2.1 Eliminating geometric effects using only radar images

We have discussed the elimination of geometric effects using auxiliary information, such as DEM. An alternative to this method has been developed [Gabriel, 1989], in which radar images alone are used to eliminate topographic contributions and reveal contributions related to ground displacements. At least three radar images are necessary. We will label them M (‘master’ image, used as a geometrical reference), S (‘slave’ image, geometrically dependant on the master image) and C (for another ‘slave’ image known as the complementary image). Letting the range pixel number in the master image be i , we have (Fig. 4.7):

$$H - e_i = R_m \cdot \cos(\gamma_i - \theta) \quad (4.3)$$

Applying Pythagoras’s theorem to the triangle SMT , we get:

$$R_e^2 = \delta_{SM}^2 + R_m^2 + 2\delta_{SM}R_m \sin \gamma_i \quad (4.4)$$

whence:

$$\sin \gamma_i = \frac{R_e^2 - \delta_{SM}^2 - R_m^2}{2\delta_{SM}R_m} \quad (4.5)$$

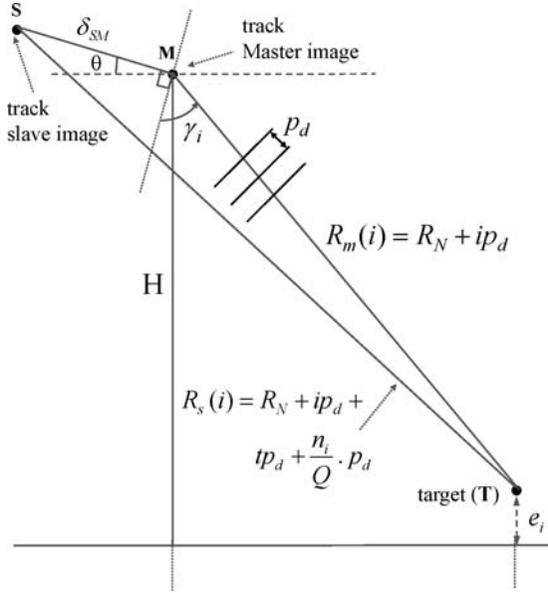


Fig. 4.7 Definition of a fringe image from knowledge of the relief and vice-versa.

We shall now express all the lengths by introducing the size of the range pixel p_d (Sect. 2.2.3). If R_N represents the near range of the master image, we can write:

$$R_m(i) = R_N + ip_d \quad (4.6)$$

$$R_s(i) = R_N + ip_d + tp_d + \frac{n_i}{Q} \cdot p_d \quad (4.7)$$

where t is the difference between the near ranges of the slave and master images expressed in units of range pixels (t was obtained previously by correlation of the images) and finally n_i is the number of fringes observed in the interferogram from the near range to pixel i . Therefore n_i/Q is the number of corresponding pixels (unlike t , n_i and γ_i are functions of i). For a single image line, Equation (4.5) becomes:

$$\sin \gamma_i = \frac{(R_N + ip_d + tp_d + (n_i/Q)p_d)^2 - \delta_{SM}^2 - (R_N + ip_d)^2}{2\delta_{SM}(R_N + ip_d)} \quad (4.8)$$

The altitude of ambiguity (Sect. 4.3.1) can be calculated from Eqs. (4.8) and (4.3): it is the elevation variation that causes a shift of one fringe. It is therefore equal to $e_a = de_i/dn_i$ (which is a less than satisfactory notation because actually all we are trying to calculate is the elevation variation when n_i varies by one unit) which can be expanded as:

$$\frac{de_i}{dn_i} = R_m \frac{d\gamma_i}{dn_i} \sin(\gamma_i - \theta) \quad (4.9)$$

or:

$$\frac{de_i}{dn_i} = R_m \frac{d(\sin \gamma_i)}{dn_i} \frac{\sin(\gamma_i - \theta)}{\cos \gamma_i} \quad (4.10)$$

or:

$$e_a \approx \frac{\lambda}{2} \cdot \frac{R_N + ip_d + tp_d + \frac{\lambda n_i}{2}}{\delta} (\tan \gamma_i \cos \theta - \sin \theta) \quad (4.11)$$

If the value of γ_i is known, we can deduce n_i from Eq. (4.8):

$$\frac{n_i}{Q} p_d = \sqrt{(R_N + ip_d)^2 + \delta_{SM}^2 + 2\delta_{SM}(R_N + ip_d) \sin \gamma_i} - (R_N + ip_d + tp_d) \quad (4.12)$$

which relates the fringe network (expressed as n_i/Q) to the topography (expressed through γ_i). This equation also makes it possible to predict the fringe network of one interferogram from that of a first interferogram (as long as the second one reflects only topographic effects) using the following procedure:

- (1) Using Eq. (4.8), we calculate $\gamma_s(i)$ as a function of $t_s, n_s(i)$ and δ_{SM} , which are the values for the first interferogram (Master-Slave). In the complementary interferogram, the look angle $\gamma_c(i) - \theta_c$ is equal to the look angle $\gamma_s(i) - \theta_s$, because both interferograms are based on the same master scene. In practice, it is only necessary to know the difference $\theta_c - \theta_s$ for this operation.
- (2) We reconstitute $\sin \gamma_c(i)$ and then, with the help of δ_c and t_c , we calculate $n_c(i)/Q$ which gives $n_c(i)$. We can then subtract the calculated fringes from the second interferogram which may reveal possible phase shifts due to a differential effect.

It is necessary to remove phase ambiguities from the first interferogram (phase unwrapping) before this operation because the number n_s must be absolute. The manipulation that we have just described often produces significant artifacts because of the difficulties of the unwrapping operation, which may generate errors, and also because an interferogram hardly ever contains purely topographical information (except for simultaneous acquisitions).

4.2.3 Finishing tasks

Raw phase differences between two images are chiefly caused by orbital effects, such as a slight closing in between two orbits during acquisition (a closing in of 28 mm is

enough to create a fringe in the case of ERS). We can remove most of this contribution from the interferogram by modeling the orbits. When the uncertainty of this differential knowledge is more than a few centimeters, unmodeled ‘orbital’ fringes remain which can be removed by ‘adjusting’ one of the orbits as follows: on the left-hand image in Fig. 4.5, 15 fringes can be counted from point A to point B, meaning that the distance between satellite S_1 and B must be increased (or reduced) by 15 times one half wavelength. Let us assume that the phase sign indicates that this distance should be increased. If the distance between A and S_1 is taken as a reference, the satellite’s corrected position is at the intersection of the two arcs. After removing the fringes between A and C, and keeping A as a reference, we find that the distance DS_2 needs to be increased by 7 cycles and the distance CS_2 shortened by 5 cycles, which gives the new position of the satellite on the other side of the image. These quantities may be conveniently determined by locating the point I where the fringe containing A intersects with the segment CD. The refined trajectory $S'_1S'_2$ now enables us to remove all the ‘orbital’ fringes, which may reveal the underlying phenomena, such as the effect of a small magnitude earthquake in the right-hand image. This process is similar to adjusting the parallelism of mirrors for optical interferometry except that the adjustment here is virtual, being obtained by calculation.

4.3 Application for topography

4.3.1 Set of equations and error budget

After orbital fringes have been removed ‘topographic fringes’ usually dominate the interferogram. In Fig. 4.2, we notice that if there is an elevation variation e for the terrain in a pixel, and keeping the range of the master image as a reference ($R_{mA} = R_{mC}$), at point C the range R_{sC} will be modified to $R_{sA} + r$. This modification will take the form of an additional R signal path difference revealing the relief. The path difference will be observed as a phase difference between two neighboring pixels. The entities e and r are linked by the following first order equation [Massonnet, 1993]:

$$e \approx \frac{rD_A R_{sA}}{H\delta} \quad (4.13)$$

To avoid having to manipulate complicated geometrical equations, a practical solution is to combine all the parameters involved in measuring the elevation such as orbital separation, wavelength or incidence angle in a single significant quantity, the *altitude of ambiguity*. This is defined by the elevation variation which causes one topographic fringe. It usually varies relatively little within a given interferogram which makes it a simple and practical quantity. If we assume a purely horizontal separation δ between viewpoints, an incidence angle of θ , an observation range of R_m and a wavelength of λ , the altitude of ambiguity e_a is approximated by [Massonnet, 1993]:

$$e_a \approx \frac{R_m \lambda \tan \theta}{2\delta} \quad (4.14)$$

The altitude of ambiguity is often expressed as a function of the perpendicular or orthogonal baseline B_{orth} , which is the separation between orbits projected onto a direction perpendicular to the viewing direction (Fig. 4.2). We then have:

$$e_a = \frac{R_m \lambda \sin \theta}{2B_{orth}} \quad (4.15)$$

Fig. 4.8(c) shows an example of a topographic interferogram acquired over Awaji Island (Japan) by the JERS-1 satellite (Table 3.1). The mean amplitude of the M and S images is also shown in Fig. 4.8(a), and Fig. 4.8(b) displays the interferogram before removal of the orbital fringes.

Please note:

- (1) As the baseline increases, the altitude of ambiguity decreases, which corresponds to an increase in the topographic sensitivity of the interferogram, such as in (Fig. 4.9).

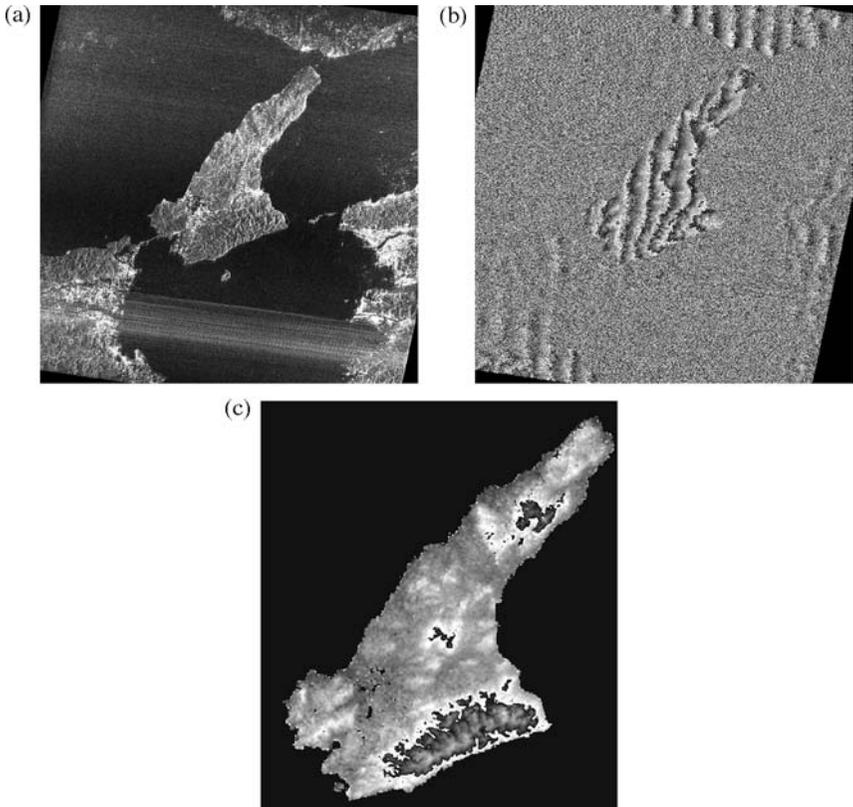


Fig. 4.8 Example of a topographic interferogram obtained from JERS over Awaji Island (Japan); (a) mean amplitude image; (b) interferogram before elimination of the orbital fringes; (c) topographic interferogram.

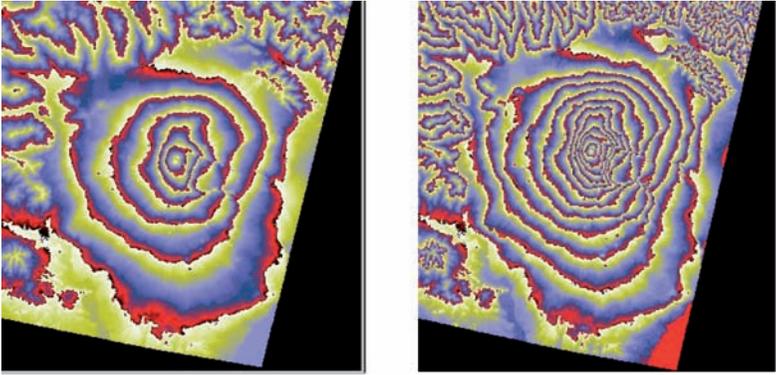


Fig. 4.9 Simulation of a topographic interferogram over Etna (Italy), in ERS configuration; (a) altitude of ambiguity: 500 m; (b) altitude of ambiguity: 250 m.

- (2) Passive interferometric systems, in which one antenna both transmits and receives and the other only receives, can also be formulated in this way. The *Shuttle Radar Topography Mission* (SRTM) was based on a similar concept [Jordan, 1996]. For such systems, the reasoning developed here applies but the condition concerning the critical baseline (defined in Sect 4.1.6 and calculated in Sect. 4.5) is multiplied by 2 because the signal path difference only applies to the return path. The 60 m extendable beam separating the two antennas carried by the Shuttle for the SRTM mission is therefore equivalent to only 30 m for an active system such as ERS.
- (3) For ERS, the altitude of ambiguity can vary from ten meters (corresponding to the greatest authorized horizontal separation of one kilometer) to infinity (in the unlikely case where both images are acquired from exactly the same point). It may seem strange that such a difference in altitude (at least ten meters) should be necessary to create a viewpoint difference equivalent to only one wavelength. The reason is that topographical error is observed in almost the same way on the two images, since their parallax is very low due to the interferometry condition.

4.3.2 Eliminating measurement ambiguity

In mountainous terrain, after eliminating orbital fringes, an interferogram will look like a series of interference fringes very similar to contour lines, separated in elevation by the local value of the altitude of ambiguity. In the example in Fig. 4.10, the interferogram was produced from two radar images of Mount Etna taken by the space shuttle in 1994 at a one-day interval. Fig. 4.10(a) shows the mean amplitude of image M-S and Fig. 4.10(b) the interferogram. The altitude of ambiguity was 500 m. We assume here that no ground displacement or changes in the atmosphere perturbed the topographical measurement. The essential difference between an interferogram like this one and a topographical map is that the ‘contour lines’ present in the form of fringes are not

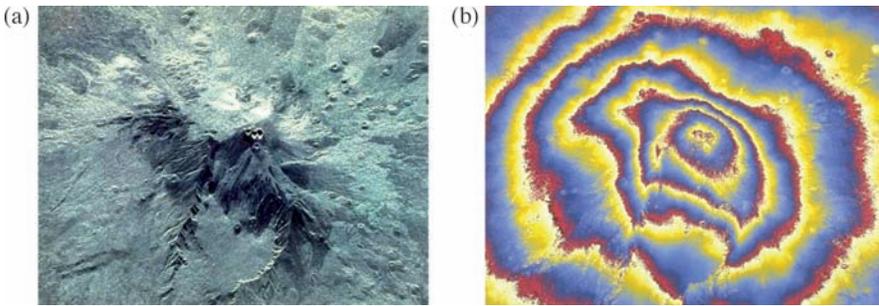


Fig. 4.10 Example of a topographic interferogram obtained from SRTM over Etna (Italy). The altitude of ambiguity is 500 m; (a) mean amplitude image M-S; (b) topographic interferogram after elimination of the fringes.

numbered. In fact the radar signal cannot assign any order to the fringes because each one corresponds to geometrical differences between the images which are generally less than one hundredth of the size of the pixel. These differences are so small that the correlation measurements cannot resolve the fringe rank ambiguity. The only way of numbering the fringes, an operation commonly called ‘phase unwrapping’, is to analyze the continuity of the interferogram. This is what we do instinctively when we analyze a topographic map. The noisy areas in the interferogram, and also areas where the topographical fringes are too close together, create serious problems for the automatic phase unwrapping algorithms, for which several designs have been suggested [Ghiglia, 1998]. Phase measurement provides a very good precision within λ . But ambiguity occurs whenever range offsets exceed a wavelength. Image correlation measurements are capable of taking over phase measurements for greater offsets of range. An important parameter that we have already discussed (Sect. 2.2.3) is the ratio Q between the carrier frequency f_c and the range sampling frequency f_d . This is also twice the number of wavelengths which can fit into a range pixel. For the ERS-1 radar, Q is equal to 279.4. If a precision of 3% of the size of a pixel can be obtained through correlation, a typical ambiguity would span approximately ten wavelengths (the ambiguity rank would then be limited to ten).

4.4 Application for displacement measurement

4.4.1 Set of equations and error budget

If we assume that orbital effects and relief effects have been compensated for, an interferogram will highlight any range variations occurring during the elapsed time between acquisitions. Although these variations can also be caused by atmospheric effects (Sect. 4.6.1), they are basically useful for measuring ground displacements. If part of the terrain observed in an image has shifted, the shift will show up as fringes in the interferogram. If the displacement at point A is represented by a vector \vec{D} and the position of the satellite is \vec{S} in a given coordinate system, the displacement will involve

a number of fringes n equal to:

$$n = \frac{2\vec{D} \cdot (\vec{S} - \vec{A})}{\lambda \|\vec{S} - \vec{A}\|} \quad (4.16)$$

where the factor 2 accounts for the round trip and where the operator ‘ \cdot ’ represents the scalar product. The precision of such a measurement will then depend on the direction of the displacement and on the smallest part of a phase cycle which remains significant with respect to noise. Displacements along the radar line of sight are most efficiently observed. From Eq. (4.16), we can deduce that they produce a fringe in the interferogram for a displacement of $\lambda/2$. For a displacement along the satellite track, however, where \vec{D} is perpendicular to $(\vec{S} - \vec{A})$, the interferogram is not modified. The vector, $(\vec{S} - \vec{A})$ once it has been normalized and expressed in geographical terms (for example: North, East and Vertical), gives the sensitivity of the measurement. For example, a typical value in the mid-latitudes for a satellite like ERS-1 would be $(-0.08, 0.35, 0.93)$. The signal-to-noise ratio in the interferogram determines what fraction of a phase cycle is significant. It depends on the power of the radar signal (Sect. 3.4), and even more on the way in which the configuration of the elementary targets (which creates speckle, Sect. 4.3.11) has been maintained during the time elapsed between the two acquisitions. From this point of view, a desert, rocky or urban landscape will be very favorable, water (ocean, river or marsh) will give no result and terrain with vegetation will give a medium result depending on the time elapsed and the extent to which the radar waves can pass through the vegetation and attain ‘solid’ targets on the ground. Longer wavelengths can do this more easily. An interferogram’s signal-to-noise ratio is therefore essentially experimental data. Displacement measurements with a precision less than 2 mm have been obtained in several cases.

The precision also depends on how well the other components of the interferometric information have been removed. For example, if the interferogram used for the measurement of the displacement has an altitude of ambiguity of 100 m and if the elevation model used for removing the topographic contribution can have errors of 20 m, one fifth of a topographic fringe can remain. For a wavelength of 5 cm, a ‘noise’ corresponding to half a centimeter (i.e. $\lambda/2 \times 1/5$) of a round trip, or a 5 mm error in range, may remain in the interferogram.

4.4.2 Examples of use

Interferometry has more to offer for the measurement of ground displacement than for topographical measurement because it really has no competitors (Sect. 4.8). The technique suffers from three limitations: (1) the ambiguous nature of the measurement, which requires unwrapping; (2) geometric sensibility which varies according to the direction of the displacement; and (3) the efficiency of the application depends on the kind of terrain observed. These limitations have not prevented spectacular developments when measuring deformations caused by earthquakes [Massonnet, 1993], glaciers [Rignot, 1997], volcanoes [Massonnet, 1995] and even phenomena with much

smaller amplitudes such as post-seismic displacements [Massonnet, 1996], [Peltzer, 1996] or continental drift [Vadon, 1997].

Fig. 4.11 shows a case study of the Gulf of Aqaba (Jordan). Fig. 4.11a shows the amplitude of one of the radar images. Fig. 4.11(c) shows the uncorrected interferogram including the orbital errors. Fig. 4.11(d) shows the interferogram after correcting the positions (using the method shown in Fig. 4.5). Here we can distinguish two ‘layers’ with very different meanings: the series of fringes alongside the Gulf of Aqaba was caused by an earthquake. Each fringe corresponds to approximately 3-cm offset in the line of sight. The remaining fringes represent the topography of the region and have not been removed here. There is only slight topographical sensitivity (the orbits are very close together and the altitude of ambiguity reaches several hundred meters). Fig. 4.11(b) shows the coherence (Sect. 4.2.2). This is found by mathematical measurement of the local similarity of the values in the interferogram. If there is strong similarity we can conclude that the measurement is valid (ground displacements or topography are generally continuous). If not, the measurement is dominated by noise, as on the Gulf of Aqaba itself, where it is purely noise. The coherence gives information about the local reliability of the measurements and also about the ‘lifetime’ of the targets and their volume signature. This constitutes a measurement by itself that we will not go into here.

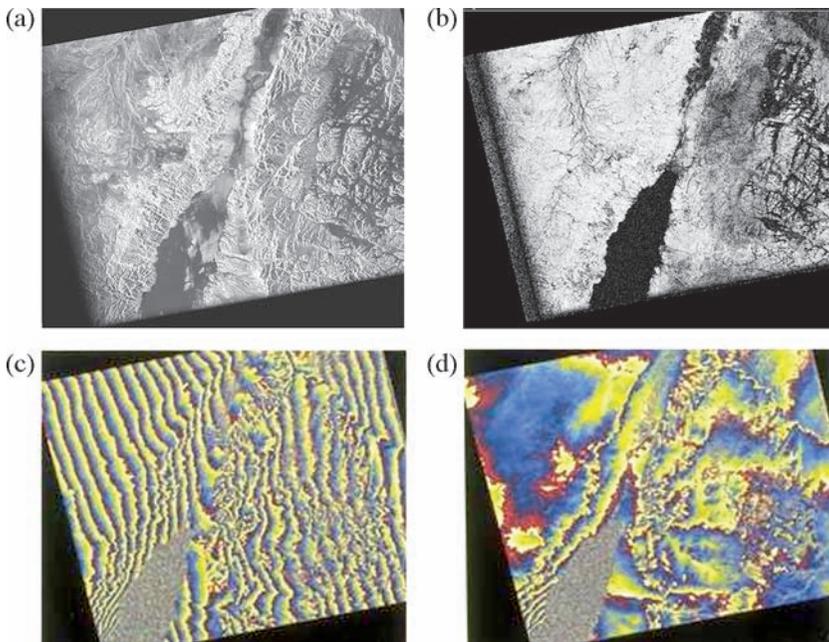


Fig. 4.11 Example of an interferogram obtained from ERS over the Gulf of Aqaba (Jordan); (a) mean amplitude image; (b) coherence; (c) interferogram before elimination of the orbital fringes; (d) interferogram.

Fig. 4.12 shows the distortion field of the earthquake at Landers, California (1992), observed by the European ERS-1 satellite [Massonnet, 1993]. In this case the topographic contribution has been removed from the interferogram leaving only the distortion field of the terrain created during the 18 months that passed between the acquisitions of the radar images composing the interferogram. The region represented covers approximately 60 km. Each color cycle corresponds to a difference in range between the images of one 56 mm wavelength. As the ground displacements are observed from the round trip of the wave however, 28 mm of deformation causes a complete phase cycle. In order to know how much displacement was caused by the earthquake at point A we simply count the number of fringes between a point where displacement is assumed to be zero (here point O) far from the epicenter, and point A. In this case there are 5 fringes representing 15 cm. Close to the rupture line, the fringes become too dense to be properly read. Actually, above the rate of one fringe per pixel, we exit the interferometric domain, not because is of a difference of viewpoint (Sect. 4.1.6), but because of the physical extension or contraction of the pixel area between acquisitions. During those 18 months several earthquakes left their mark on the region. The small earthquake indicated in the figure caused a depression of 12 cm in a small area of a few kilometers, producing one thousand times less energy than the big one.

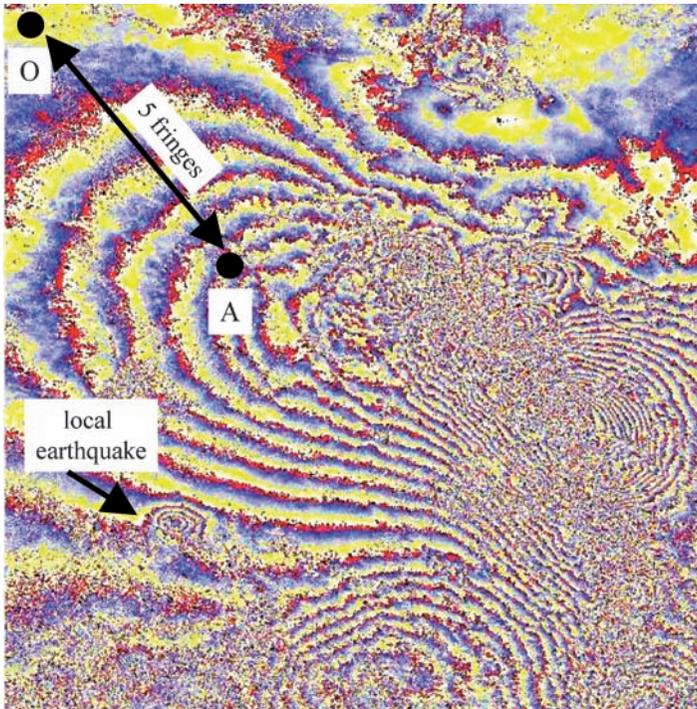


Fig. 4.12 Distortion field caused by the Landers earthquake in California (1992) observed by the European ERS-1 satellite.

The high precision of interferometry measurements is shown in the example in Fig. 4.13(b) where a fault has shifted 10 mm, i.e. one third of a fringe as can be seen from the color legend, over a distance of 22 km. It would almost seem that it is easier to detect such a displacement from 800 km away in space while traveling at 7.5 km per second, than when standing right on it. The two images used for this comparison were acquired by ERS-1 two years apart. As can be seen from the amplitude of one of the images (Fig. 4.13(a)), part of the landscape is used for agriculture. Over two years, this surface has obviously been profoundly remodeled and interferometry is of no use in this case.

In the field of hydrology (Fig. 4.14), interferometry has been used to measure the quantity of water pumped by a geothermal plant. This example (near Mesa, California, Fig. 4.14(a)) shows the ground subsidence due to a geothermal plant. In the course of the two years between the image acquisitions by ERS, this plant has pumped 100 million m^3 of hot water from aquifers located at a depth of 2200 m in order to generate electricity and has re-injected 95 millions m^3 . Fig. 4.14(c) shows a sunken area 18 km by 10 km, at most three fringes or 9 cm deep, representing a volume of 4 million m^3 , approximately the missing volume acknowledged by the plant. The central interferogram (Fig. 4.14(b)) shows that if the time lag is sufficiently short (here only 6 days), we can make legible interferograms of agricultural areas. We observe an example of the direct effect of irrigation on phase [Massonnet, 1998] which can be caused by either changing geometry (swelling) or changes in permittivity.

In the example in Fig. 4.15, researchers have used images from the ERS satellite to observe the hinge areas due to glacier bending caused by tidal displacements [Rignot, 1997]. The bending area is shown by a few visible fringes on three Greenland glaciers.

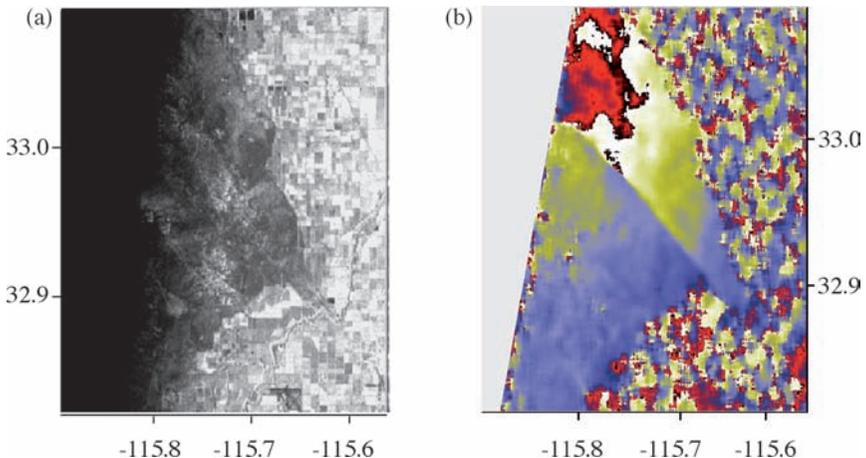


Fig. 4.13 Displacement of a fault line observed on an ERS interferogram. Superstition Hill, California, 1992. The second image was acquired two years after the first; (a) Mean amplitude image; (b) Interferogram.

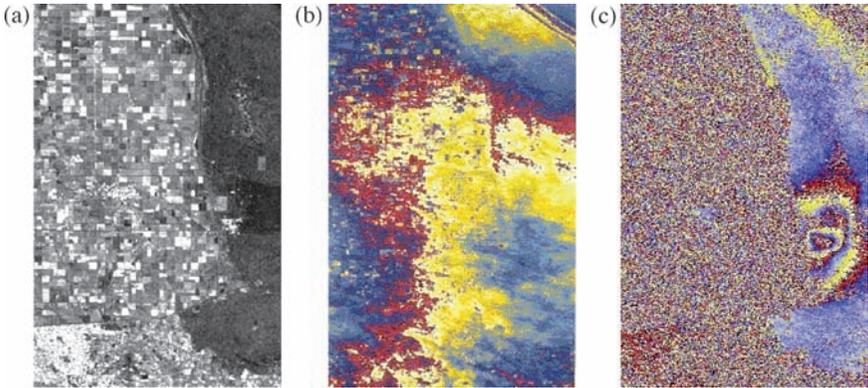


Fig. 4.14 (a) ERS radar image of the Mesa region, California; (b) 6-day interferogram; (c) 2-year interferogram.

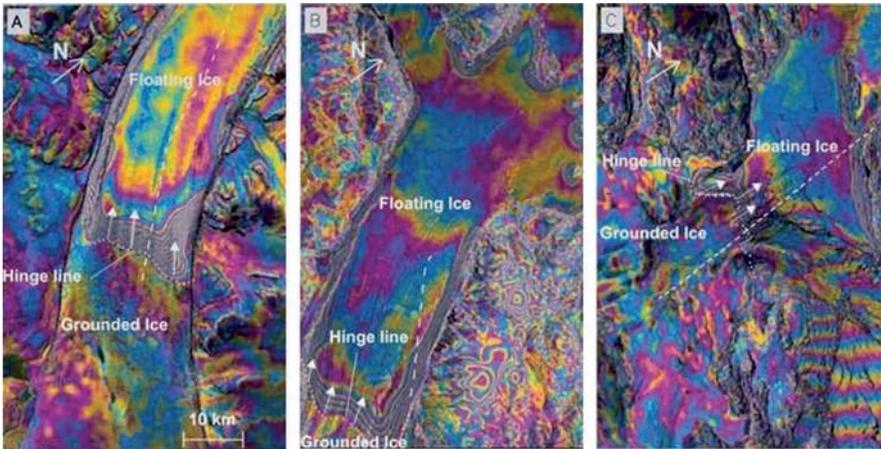


Fig. 4.15 Areas of bending on three Greenland glaciers caused by tidal displacement, observed by ERS interferometry from [Rignot, 1997]

The bending area is much farther from the mouth of the glacier than was previously thought and the sea, which is under the glacier for a considerable distance, contributes to its melting. This study has revealed that glaciers are melting at a faster rate than had been previously thought.

4.5 How slope effects limit interferometry

The effect of slopes in interferometry is shown in Fig. 4.2 as well as the corresponding relationships of ‘pixel distortion’. However, a small model can be useful for expressing the geometric limitations in a way that is easier to manipulate even though it is approximate (Fig. 4.16(a)).

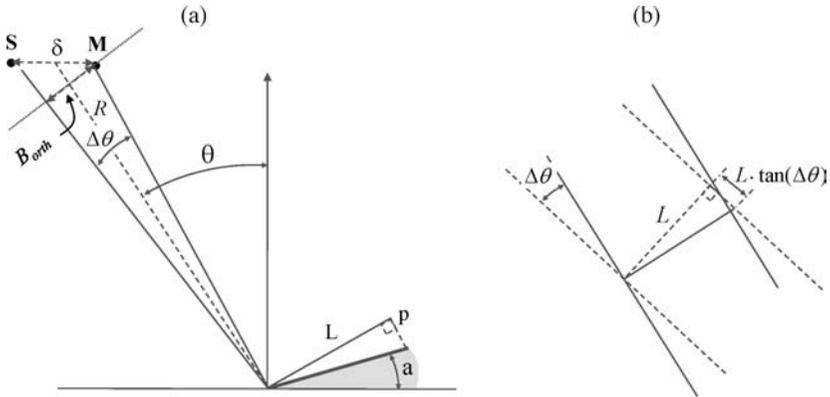


Fig. 4.16 (a) Interferometry limitations due to slopes; (b) Geometric limit conditions for interferometry.

Let us take two satellites separated by a baseline, which may or may not be orthogonal, with an angle $\Delta\theta$ between their two viewpoints. The scene is observed with an incidence angle θ and the slope of the terrain has an angle a which may be positive or negative as shown in Fig. 4.16(a).

The entity p indicates the range resolution cell (very close to the size of a range pixel p_d). The pixel’s ‘orthogonal width’ L (meaning the extent of the terrain covered by a given pixel) is equal to:

$$L = p \cdot \cot(\theta - a) \tag{4.17}$$

The ‘interferometry limit’ is reached when the variation between the one-way signal path differences for the two extremes of the pixel exceeds one half of the radar’s wavelength (the interferometry phase then evolves by more than one round trip between the two edges of the pixel) (Fig. 4.16(b)). Interferometry can therefore be used when:

$$L \cdot \tan(\Delta\theta) < \lambda/2 \tag{4.18}$$

Angle $\Delta\theta$, which is very small, can be approximated by its tangent. The condition governing the range pixel size can therefore be written as:

$$p \cdot \Delta\theta \cdot \cot(\theta - a) < \lambda/2 \tag{4.19}$$

$2p/\lambda$ is in fact the carrier to sampling frequency ratio Q (Sect. 2.2.3), of the radar. Our condition then becomes:

$$Q \cdot \Delta\theta < \tan(\theta - a) \tag{4.20(a)}$$

or:

$$a < \theta - \text{Arc tan}(Q \cdot \Delta\theta) \tag{4.20(b)}$$

which is the slope condition sought. From this equation we can find the limit on flat ground using $\theta = \text{Arc tan}(Q \cdot \Delta\theta)$, which gives:

$$\Delta\theta_{\max} = \frac{\tan \theta}{Q} \quad (4.21)$$

From Eq. 4.14, and noting that $\delta \approx R \cdot \Delta\theta / \cos \theta$, the topographic sensitivity of the image pair expressed using its altitude of ambiguity is equal to:

$$e_a = \frac{\lambda \sin \theta}{2\Delta\theta} \quad (4.22)$$

For $\Delta\theta_{\max}$, the altitude of ambiguity is therefore:

$$\frac{\lambda Q \cos \theta}{2} = p \cos \theta \quad (4.23)$$

This is the smallest altitude of ambiguity that is achievable.

Please note: From Eq. (4.20(b)) we can establish the equation for the **critical orthogonal baseline** $B_{\text{orth}}^{\text{crit}}$ as defined in Sect. 4.1.6. From Fig. 4.16(a) and using the value for $\Delta\theta_{\max}$ corresponding to a local slope a , we obtain:

$$B_{\text{orth}}^{\text{crit}} = \frac{B \cdot R \cdot \text{tg}(\theta - a)}{f_c} \quad (4.24)$$

Let us now give a few numerical applications. For ERS, $\theta = 23^\circ$. For an angle of $\Delta\theta$ of 0.02° , which corresponds to about one quarter of the maximum interferometry baseline allowed on flat ground, the steepest observable slope is 17.4° ($Q = 280$). If ERS had five times more resolution, Q would be 56 and the steepest observable slope would be 21.9° . It should be noted that in this last case, angle $\Delta\theta$ would represent a much smaller proportion of the maximum baseline for interferometry.

Let us now assume that $\theta = 45^\circ$ while $\Delta\theta$ keeps the same value. If ERS were to use this look angle, the steepest observable slope would be 40.4° whereas for a satellite with five times more resolution it would be 43.9° .

Let us now assume that we wish to create an elevation model with a precision of 1 m RMS on the ERS resolution cell (which we estimate to be 80 m^2). We only need to work with an altitude of ambiguity of 10 m as long as we can 'divide the phase into ten' meaning if the phase noise level is 10 % of a cycle. This can be obtained even if the ratio between the incoherent and coherent signals is close to 0 dB (Sect. 4.1.5). If the geographic cell being considered is 30 m by 30 m (i.e. 11 times the surface of the ERS cell), the added coherent gain in power would be 3.35, meaning that we could even work with an incoherent level higher than the coherent.

Working at three times the minimum value of the altitude of ambiguity, in the 'topographic' mission we are describing, means choosing:

$$\Delta\theta = \frac{\Delta\theta_{\max}}{3} \quad (4.25\text{-a})$$

therefore:

$$Q \cdot \Delta\theta = \frac{Q \cdot \Delta\theta_{\max}}{3} = \frac{\tan \theta}{3} = \frac{1}{3} \tag{4.25-b}$$

The steepest observable slope is then $\theta - \text{Arc tan}(1/3)$ or 26.5° . Fortunately, such slopes are fairly unusual.

4.5.1 Frequency interpretation of the slope effect

Conditions limiting the use of interferometry can also be expressed in the frequency domain. They relate to the equivalence between viewpoint diversity and frequency shift (Fig. 4.17): projected onto the ground, a monochromatic wave with frequency f_c transmitted at an incidence angle θ , produces the same wave fronts as a monochromatic wave at frequency $f_c + \delta f$ with incidence angle $\theta + \delta\theta$, if δf and $\delta\theta$ are sufficiently small. In other words, let us consider the frequency offset δf to be applied to the carrier frequency f_c which would keep the same phase difference between A and B. In this case, δf satisfies the equation:

$$\frac{c}{(f_c + \delta f) \cdot \sin(\theta + \delta\theta - a)} = \frac{c}{f_c \cdot \sin(\theta - a)} \tag{4.26}$$

which leads to:

$$\delta f \approx -\frac{f_c \cdot \delta\theta}{\tan(\theta - a)} \approx -\frac{f_c \cdot B_{\text{orth}}}{R \cdot \tan(\theta - a)} \tag{4.27}$$

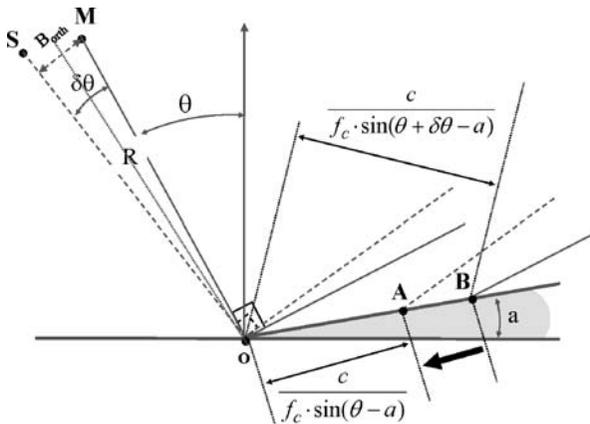


Fig. 4.17 Ground projection of the wavelengths associated with a monochromatic transmission whose frequency is f_c from Master (M) and Slave (S) images.

In reality, the wave transmitted is not monochromatic but is modulated by a *chirp* whose width is B_d (Sect. 3.3.7). The interferometry coherence between the images M and S will be lost if their respective frequency spectra become separated as a result of

the shift δf (the two waves' spectra no longer overlap; they are totally incoherent), i.e. when $|\delta f| > B_d$. This condition is equivalent to the equation for the critical orthogonal baseline in Eq. (4.24).

The equivalence principle that we have just described is known as the *wavenumber shift*. It was first formalized by Gatelli [Gatelli, 1994] in the context of use for super-resolution (Sect. 2.7).

4.6 Interpreting the results

4.6.1 Typical signatures of different contributions

Phase signatures may be logically classified for correct interpretation if several interferograms are available. Particular examples are:

- *Signatures related to orbital separation*: These are phase signatures found in several interferograms with an amplitude proportional to the topographic sensitivity of the pairs which itself depends on the separation of the orbits. This behavior is characteristic of uncompensated or badly corrected residual topographic data. Such topology is similar to that shown in [Figures 4.5 and 4.8\(b\)](#) before correction.
- *Atmospheric artifacts*: This term covers the signatures that occur with the same amplitude in all the interferometry combinations of a given image. This behavior is characteristic of heterogeneous conditions of propagation in the atmosphere at the time of this image acquisition. Since the responsible image has been identified, the sign of the artifact is known. Artifacts which lengthen the signal path can be attributed to humidity content, pressure waves or tropospheric turbulences. Those which shorten the path are due to local neutralizations in the ionosphere ([Fig. 4.18](#)).
- *Time-dependant signatures*: These are phase signatures observed with the same amplitude in several combinations of interferometry images, for which the acquisition times of the component images bracket a particular time. This behavior is found in geophysical events such as earthquakes.

We could continue this list, for example with phase signatures observed in interferograms whose amplitude is proportional to the time elapsed between the two component images. This behavior is found in geophysical deformations which are progressive over time.

4.6.2 Methods for improving interpretation

There are several simple methods for improving the appearance or content of interferograms. Several filtering methods have been suggested. The simplest consists in combining the phases of an interferogram with the corresponding amplitude of the source radar images and then averaging the complex numbers thus formed in a given

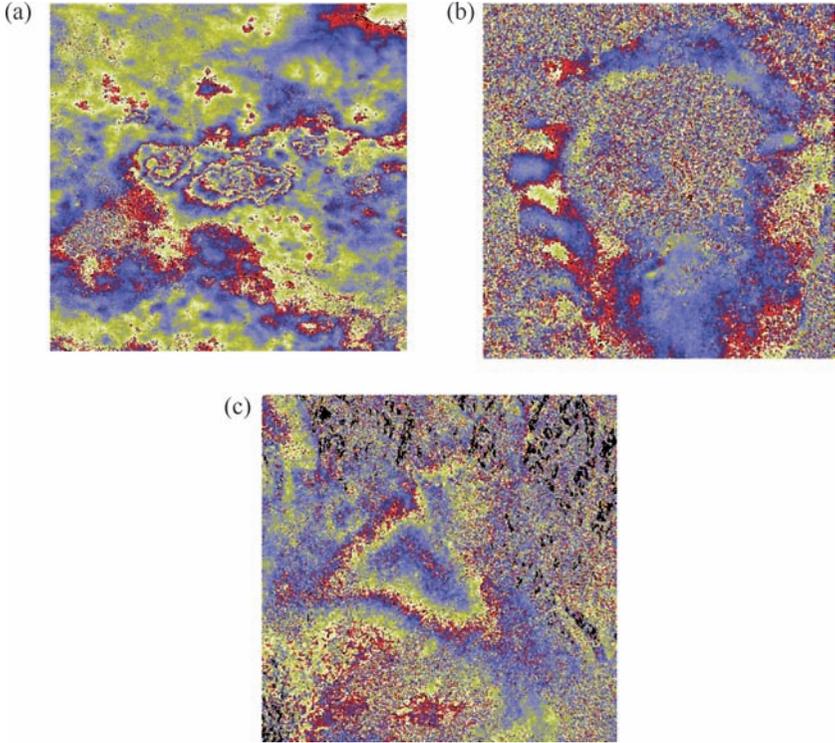


Fig. 4.18 Examples of atmospheric artifacts observed in interferograms; (a) clouds (cumulus); (b) cloud chains; (c) hole in the ionosphere.

neighborhood (Sect. 4.2.2). The phase of the mean complex number thus formed will have much less noise than the initial phases at the cost of a degradation of the image's spatial resolution. This method assumes that the densest fringes of the interferogram have previously been removed, in particular the orbital fringes because otherwise calculating the mean would spoil the results.

Hardly more sophisticated processing methods than the averaging described in the previous section are less intuitive. The human brain does not seem capable of fully grasping the images formed by series of fringes commonly found in interferograms. Fig. 4.19 is an example. Fig. 4.19(a) is an array of 'almost' horizontal fringes. It is therefore tempting to model it as an image of horizontal fringes of identical frequency, as in Fig. 4.19(b). But, if we subtract 19(a) from 19(b) we obtain a single vertical fringe Fig. 4.19(c). It is very difficult to conceive this as being the straightforward difference between the two first images. This counter-intuitive result is in fact very useful for the so-called integer combination technique.

An interferogram is an ambiguous signal because the fringes are not numbered. The real value of the interferogram is unknown as long as the integer number of

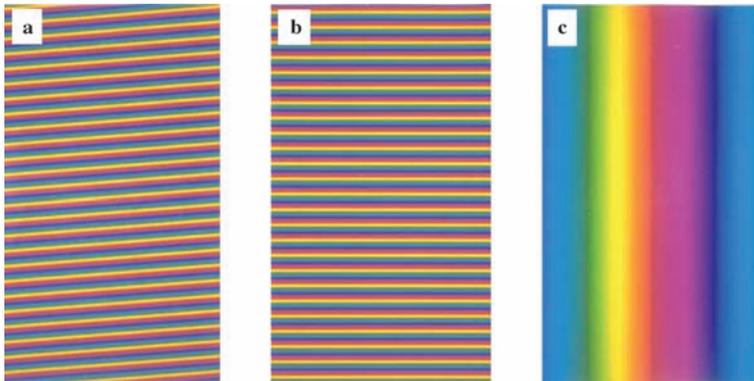


Fig. 4.19 Strange phenomenon observed in the ‘arithmetic’ of fringe arrays. Array (c) is the result of the difference between array (a) and array (b).

wavelengths concealed behind each fringe is unknown. This means that we cannot multiply an interferogram by a non-integer number, for example 1.3, because each fringe would then become 1.3 fringes (with new transitions located three quarters of the way across each fringe) and the unknown integer behind each fringe would become some *non-integer* unknown number. If we multiply an interferogram by an integer on the other hand, 2 for example, each fringe (or cycle) will become two, with new intermediate fringes in between. The integers hidden behind each fringe will become even numbers and the ‘new numbers’ behind the ‘new fringes’ will be odd numbers. An interferogram thus multiplied by an integer is always a ‘normal’ interferogram, except that the altitude of ambiguity is divided by 2, which means that each fringe retains only half of the initial geometric value.

Using this principle, we can also use the logical characterization of the content of interferograms to attenuate or amplify a given contribution. For example, let us assume that an interferogram, because of the orbital separation of the viewpoints that make it up, has a topographic sensitivity of 100 m of difference in elevation per fringe (in other terms an altitude of ambiguity of 100 m), whereas another interferogram shows only 50 m, then the combination of twice the first interferogram minus the second will remove all topographical effects. Interferograms can be combined by multiplying with integer coefficients, and this can be done before any phase unwrapping. Let us illustrate the efficiency of the method in a real case, where the altitude of ambiguity ratio is not necessary close to an integer. We subtract an interferogram with an altitude of ambiguity $e_{a2} = 98$ m (Fig. 4.20(a)) from another with an altitude of ambiguity $e_{a1} = 71$ m (Fig. 4.20(b)), in order to obtain a synthesized interferogram, which is practically devoid of topographical effects (Fig. 4.20(c)).¹ Subtracting one interferogram from

¹It can be seen that the resulting interferogram has an altitude of ambiguity e_a such that $1/e_a = 1/71 - 1/98$, i.e. $e_a = 258$ m. The amplitude of ground displacements along the radar line of sight is visible here through a series of fringes separated by $\lambda/4$ (and not $\lambda/2$), which is here 14 mm.

another produces a new interferogram whose altitude of ambiguity e_a is given by:

$$\frac{1}{e_a} = \frac{1}{e_{a1}} - \frac{1}{e_{a2}} \quad (4.28)$$

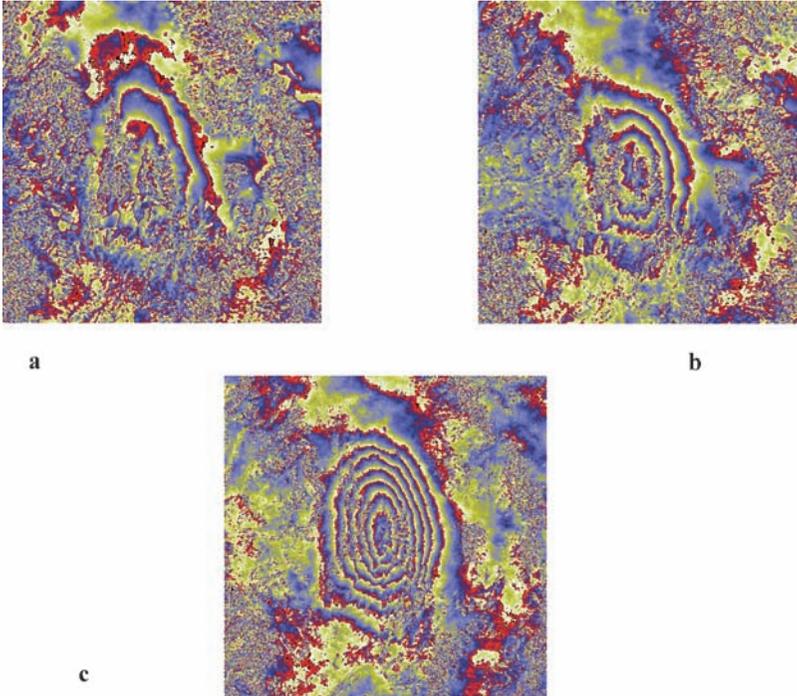


Fig. 4.20 Integer combinations of interferograms. Eureka Valley earthquake (California, 17 May 1993) observed by ERS-1; (a) interferogram S1, altitude of ambiguity: 98 m; (b) interferogram S2, altitude of ambiguity: 71 m; (c) interferogram by synthesis of S2-S1 (altitude of ambiguity equivalent to 258 m).

In the case under discussion, $e_a \approx 260$ m, which results in far lower sensitivity to topography. This very straightforward method can be used in a variety of situations, such as removing or reducing topographic contributions, subtracting absolute geophysical models from wrapped interferograms, etc. This method, the so-called interferometric integer combination, radically changes the use of phase unwrapping [Massonnet, 1996].

4.6.3 Interferometry interpretation in practice

One of the main limitations on interferometry measurement is that several types of geometric information are mixed in the same signal. This means that the precision of

the measurements does not depend only on the properties of the radar system (power, resolution, etc.). The physical nature of the ground observed and our ability to distinguish between the various contributions to the signal also play an important role. The raw capability of the measurements is dictated by the physical nature of the ground and its stability on the scale of the observation wavelength. The standard deviation of the measurement depends directly on the ratio between the mean amplitudes of the coherent and incoherent contributions to the signal. Finally, the potential precision of the data will only be meaningful in our final measurement if we can correctly account for the contributions from other types of geometric measurements, or at least place an upper bound on these effects, hopefully less than the precision desired for our measurement.

Precision of the data can be improved if filters are applied to the signal after interferometry processing. The most important of these filters is complex summing, which increases the ratio of the coherent to incoherent contributions to the signal proportional to the square root of the ratio of the integrated surface to the elementary surface (Sect. 4.1.5). It is possible to discriminate between different effects merged in the same measurement either through parametric error analysis, which depends on the viewing conditions or as a result of logical analysis of several interferograms of the same site. We shall now illustrate these mechanisms with the help of several scenarios based on the series of images in [Fig. 4.21](#).

This series is made up of the superposition of a signal with a circular shape and a background which is typical of the residues in an interferogram, coming from meteorological effects, poorly compensated topographic effects, or other sources of spurious effects. This background has been extracted from an actual interferogram. The artificial signal with a circular shape, however, has been added with variable amplitude to form the different interferograms in the series. A color representation of fringes has the advantage of being very easy to read, and of making the gradient of levels stand out more clearly than in black and white because of the eye's greater sensitivity to colors than to shades of grey. We should note however, that whereas the sign of the representation in black and white is not ambiguous, the color representation has an arbitrary sign depending on the color table used. In the last image of the series ([Fig. 4.21\(h\)](#)), which is identical to the first but in black and white, we can clearly see that the phase increases from the exterior to the interior. The boundary between one ambiguity and the next is shown by the sudden transition from a maximum value (in white) to a minimum value (in black). This is not as easy to see on the color image because the table could just as easily be organized with the sequence 'Red-Yellow-Blue' as 'Red-Blue-Yellow'. We therefore have to compare the phase variation either with the color table, or with phenomena whose sign is known because they have been processed with an identical procedure. This last method is in fact the most reliable, for several phenomena can change the sign of an interferogram. One of the phenomena that can cause such a change is the inversion of the real and imaginary in raw or processed data. The inversion of the pair formed by the master and slave images when the interferogram is created will also change the sign of the result.

As can be seen on this series of images, it is not always a simple matter to count the number of fringes which make up a signature, even if the latter is as elementary

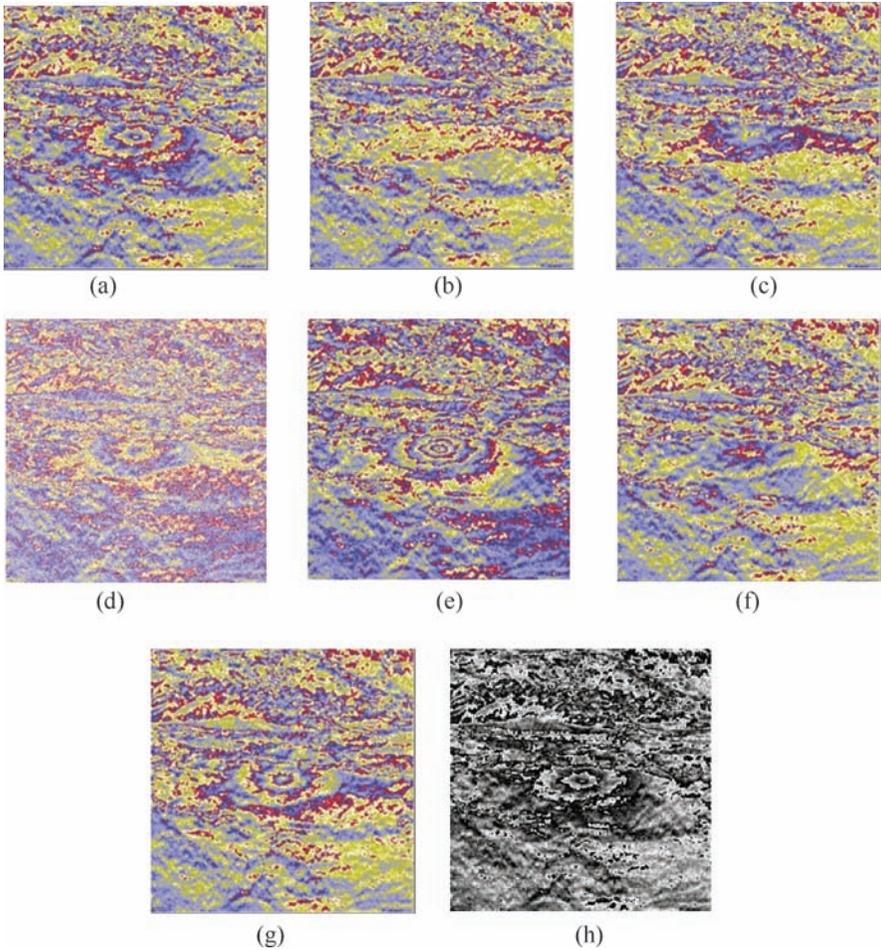


Fig. 4.21 Series of interferograms ‘Practical interpretation of interferometry’.

as our example with concentric fringes. Fig. 4.21(a) shows *three circular fringes*. From the yellow point at the center we pass through two other yellow circles before arriving at the background yellow. It is easy to see that the background yellow is ‘the same’ throughout and that there is no phase transition. Fig. 4.21(b) shows *no fringe*. Fig. 4.21(c) shows a single fringe but the progression of colors is inverted relative to that in the first image. If the first image contains three fringes we should interpret the third as containing ‘*minus one fringe*’. Fig. 4.21(d) contains *two fringes* and Fig. 4.21(e) has *minus five* concentric fringes. Fig. 4.21(f) has a *single fringe* and Fig. 4.21(g) ‘*minus three fringes*’. These examples show how difficult it can be to recognize signatures when there is background noise. The artificial signal introduced into the interferograms is strictly circular in shape with no noise. Only its amplitude varies.

We shall now discuss some fictional scenarios whose measurement results can be found in the interferograms in the series. To interpret them, we shall restrict ourselves to the following phenomena that we shall attempt to characterize:

- (1) Atmospheric artifacts, which can be recognized because they are tied to a given acquisition time (thus a specific image). They create the same number of fringes in each interferogram using this image, although possibly with a change of sign depending on whether the ‘faulty’ image is used as master or as slave,
- (2) Poorly compensated topographic contributions, characterized in each interferogram by a number of fringes that is inversely proportional to the altitude of ambiguity of the said interferogram
- (3) Sudden ground deformation characterized by a date, for example an earthquake. Each interferogram whose dates for the images bracket the date of the event will have the same number of fringes, although possibly with a change of sign depending on whether or not the master image is the first of the pair in time. Otherwise, there will not be any fringes.
- (4) We shall consider deformations which evolve at a constant rate over time, whose number of fringes in each interferogram will be proportional to the time elapsed between the acquisitions of the two images.

This is not an exhaustive list of such phenomena. We could add deformations whose behavior is more complex over time, and even phenomena which are partially reversible (swelling and subsiding of volcanoes and aquifers, etc.).

- *Scenario 1:* In this first scenario we have combined images taken from orbits $n^{\circ}5222$, 10232 and 11234 of the ERS-1 satellite. We should bear in mind that an Earth observation satellite typically performs about five thousand orbits per year. The differences between the numbers of the orbits are multiples of 501, which is the number of orbits performed by ERS-1 in the course of its 35-day orbital cycle. The combination of the orbits 10232 and 11234 produced the interferogram in Fig. 4.21(g) (–3 fringes). The analysis of these orbits indicates an altitude of ambiguity e_a of 70 m for this interferogram. The combination of orbits 11234 and 5222 produces the interferogram of Fig. 4.21(a) (3 fringes). The altitude of ambiguity is here equal to –120 m.

When it comes to interpretation, it is usually practical to proceed by elimination. If the phenomenon observed results from faulty topographic correction, the result will be a different number of fringes in the two interferograms, because of the significant difference (almost double) between the values of the altitudes of ambiguity. A regular displacement over time is also excluded, because the time elapsed corresponds respectively to 1000 orbits (almost two months) and 6000 orbits (more than a year). The number of fringes of the signatures is, however, identical. Better still, the period covered by the first interferogram is completely included in the period covered by the second. The change of sign in the number of

fringes can be explained by the ‘time inversion’ of the second interferogram, the first image of which is not the oldest. This scenario is therefore compatible with a distortion of ‘ -3 fringes’ created between the dates of orbits 10232 and 11234, which will obviously also be true of the pair 11234 and 5222, with inversion of the sign. But is our explanation the only possible explanation? The set of measurements are also compatible with an atmospheric effect on the image which is common to both interferograms (i.e. the image of the orbit 11234), characterized by three fringes. It will create three fringes in the interferogram 11234 and 5222 (the phases of 11234 minus the phase of 5222) and -3 fringes in interferogram 10232 and 11234 (the phases of 10232 minus the phase of 11234). The three orbits available to us do not allow us to reach a firm conclusion as to the nature of the phenomenon.

- *Scenario 2:* In a second scenario, we combine images taken from orbits number 5044 and 8050 of the ERS-1 satellite in an interferogram represented by Fig. 4.21(g) (-3 fringes), whose topographic sensitivity e_a is 30 m. Two other orbits from the same satellite, 9052 and 7549, produce the interferogram represented in Fig. 4.21(a) (3 fringes), with a topographic sensitivity e_a of -250 m.

In this case, we cannot invoke atmospheric effects because the interferograms do not share a single image and it is most unlikely that a signature of atmospheric origin could appear twice in identical form. The very great difference in altitude of ambiguity rules out topographical interpretation, as the two interferograms have the same number of fringes in absolute value. A constant rate of deformation over time is unacceptable because the number of fringes in the interferograms is the same whereas one of the time intervals is twice as long as the other. The only remaining hypothesis is that of a sudden displacement, with an amplitude of -3 fringes, which must have taken place during the time interval common to both interferograms, i.e. between the dates of the orbits 7549 and 8050. The sign inversion can be explained by the time inversion of the second interferogram. The four available orbits, which led to the same interferograms as for the first scenario, now allow us to draw firm conclusions about the nature of the phenomenon. It is worth noting that the timing of the event can be determined more accurately than from any of the interferograms taken separately.

- *Scenario 3:* In this scenario, we shall use the five ERS-1 orbits numbered 5001, 5502, 6003, 7005 and 7506, which we shall combine as follows: 5001 and 7506 give the interferogram in Fig. 4.21(e) (-5 fringes), for which $e_a = 100$ m; 7005 and 6003 give Fig. 4.21(d) (2 fringes), for which $e_a = 90$ m; and 5502 and 6003 give the interferogram in Fig. 4.21(c) (-1 fringe), for which $e_a = 50$ m.

If we reason in the same way as previously, we can deduce that we are dealing here with a displacement which is constant over time and which will create approximately one fringe per month, or more exactly every 35 days.

- *Scenario 4:* The ERS-1 images used in this scenario are taken from seven orbits (numbers 4530, 5031, 5532, 6033, 7035, 8037 and 12546). The first five correspond to the minimum time interval between images in interferometric conditions, i.e. one orbit every 501 orbits, or 35 days. Naturally, the site could also be observed by a different interferometric series, either from a different direction (ascending instead of descending for example), or more generally from an image with a non-zero geographic intersection with ours. Such images could be interleaved, over time, with our chosen images, but they would still only give measurements for 35 day intervals. The use of another time series may better constrain the data of a single date event. The interferograms used are represented in Fig. 4.21(b) (no fringe) for the pair 4530 and 12546, with $e_a = 1000$ m, Fig. 4.21(d) (2 fringes) for the pair 5031 and 5532, with $e_a = 60$ m, Fig. 4.21(g) (−3 fringes) for the pair 7035 and 8037, with $e_a = -40$ m, Fig. 4.21(f) (1 fringe) for the pair 6033 and 4530, with $e_a = 120$ m.

In this scenario, the hypothesis of a geophysical ground deformation was excluded from the outset because the first interferogram did not reveal one, even though its dates bracket the entire series. For the same reason, it cannot be due to a constant displacement over time. Since the interferograms of these series do not share images (they are made up of independent images), it cannot be a question of atmospheric effects. We may note that the number of fringes is inversely proportional to the altitude of ambiguity, which indicates a topographic error of 120 meters. Why is this not visible in the first interferogram? Because its amplitude of only one tenth of a fringe is practically undetectable against a background of such poor quality as in these examples.

- *Scenario 5:* In this last scenario we use the JERS satellite (Table 3.1), whose orbits numbered 10032 and 10691 provide a single interferogram shown in Fig. 4.21(e) (−5 concentric fringes). We may note that the difference between the orbit numbers is 659, which is the number of orbits covered by JERS in the course of one 44-day orbital cycle. The interferogram has an altitude of ambiguity of 250 meters.

Although we only have a single interferogram to work with, it can still be fully interpreted. If there were a topographic error, it would reveal a hole 1250-m deep. The existence of any such hole, not filled with water, would be known (a giant open-pit mine for example). With JERS, we are in L band ($\lambda = 23.5$ cm), so that the five fringes represent about 60 cm. There is no tropospheric phenomenon that can create heterogeneity equal to one quarter of the propagation delay effect of the atmospheric column (slightly more than two meters). Although ionosphere phenomena are capable of attaining this amplitude in L band, the effect would not be as localized as these circular fringes. If it were a case of regular displacement over time that would also imply extraordinary orders of magnitude (40 cm of swelling per month). It can therefore only be an earthquake or a volcano during a very active phase.

These five scenarios have helped us understand that there is neither absolute rule nor a minimum or maximum number of images for correctly interpreting an interferometric sequence -though we do need at least two radar images! As a general rule, four images capable of forming two completely independent interferograms can provide matter for useful discussion concerning atmospheric artifacts. The other combinations of these four images (there are six paired combinations in all) usually allow conclusions to be drawn as long as the kinematics of the displacements are not too complicated.

4.7 Availability and mission perspectives

4.7.1 Availability of archived radar data

Space radar images that can be combined must come from the same satellite, or from two identical satellites occupying approximately the same position (typically within one kilometer) at different times. The satellites repeat their position after covering a complete orbital cycle. Among the available satellites, we have ERS-1 (1991-2000), which followed two distinct 3-day orbital cycles for specific mission phases and a 35-day cycle which took up most of the rest of its mission. The ERS-2 satellite has an identical radar system to that of ERS-1 and is therefore compatible. It follows the same 35-day cycle but 24 hours later (an orbital combination called 'tandem mission'). By combining data from ERS-1 and ERS-2, we can therefore obtain intervals of 35, 34 or 1 day. The data are in *C* band (wavelength of 5.6 cm). The ENVISAT satellite (launched in 2002) also gives excellent results in *C* band, which are partially compatible with the ERS archive (the compatibility is not complete, because of the slightly different central frequency, 5.331 GHz instead of 5.300 GHz for ERS). The orbit is the same as that for ERS but ENVISAT's many modes of image acquisition means that it can hardly create a homogenous archive such as the ERS-like archive (23° , *VV* polarization). Also in *C* band, but incompatible with ERS because of orbital and instrument differences, we have data from RADARSAT-1 (1995-present), a Canadian satellite with a 24-day cycle, which can acquire images from different incidence angles. In *L* band ($\lambda = 23.5$ cm), we can use data from the Japanese JERS satellite (1992-1997), which describes an orbit with a cycle of 44 days, or ALOS, also in *L* band (launched in 2006). Although it is intrinsically less precise than *C* band, *L* band can obtain better results over terrain with vegetation cover, because the longer wavelength can penetrate vegetation and reach the ground.

One example of this is the Northridge earthquake, which caused damage to Northern Los Angeles in 1994, and gave an opportunity to compare ERS and JERS. The Japanese satellite produced few fringes (Fig. 4.22(a)) because each one represents about half a wavelength, or about 11.5 cm. There is also a noticeable projection effect, as the two satellites were observing from different sides of the image. We can see that the fringes remain clearly visible even though the quality is lower in the Northern part of the image. More fringes were produced by ERS Fig. 4.22(b), as they each represent about 3 cm. They are illegible in the Northern part of the image. This results in the fringe-count ratio between two given points in the images being different from the ratio of four expected from the wavelengths. There are three possible explanations for

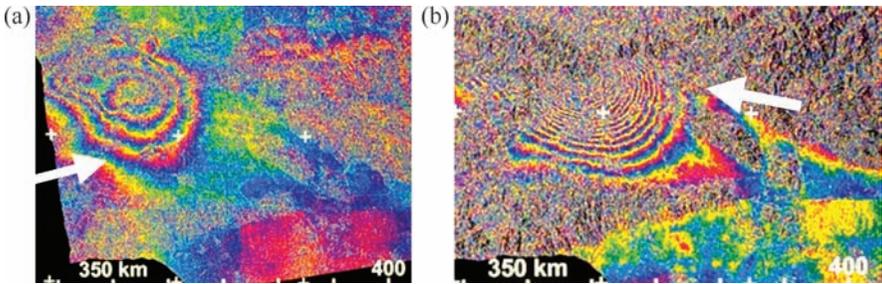


Fig. 4.22 Northridge earthquake, California. The white arrows show the pointing direction of the radar; (a) Distortion field mapped by JERS -1 (*L* band, NASDA) between 30 April 1993 and 14 July 1995 (one fringe: 115 mm of line-of-sight displacement); (b) Distortion field mapped by ERS-1 (*C* band, ESA) between 4 October 1993 and 5 April 1995 (one fringe: 28 mm of line-of-sight displacement).

this, as the time interval between the image acquisitions was similar in each case (and longer than a year): (1) vegetation which would have changed during the interval but which the *L* band could have crossed to reach the hard ground while the *C* band could not; (2) A degree of soil erosion at a scale that *C* band can detect but too small to affect *L* band; and (3) The effect of slopes which are significant in this region and which would have interfered with the ERS images taken at an incidence angle of 23° because of the layover effect (Sect. 2.6.1), but not the JERS images taken at 35° . This last explanation seems the most probable though we cannot exclude the others. Experiments over shorter periods have also been conducted with radars onboard the Space Shuttle (SIR-A, SIR-B and SIR-C missions).

4.7.2 Availability of processing resources

The use of radar images for calculating topography and ground displacements is widespread and not restricted just to radar specialists. So it is important that non-specialist users with ambitious measurement goals, or in search of subtle effects, be able to master the different processing stages involved in radar imagery.

Such processing has become accessible to non-radar specialists through the use of scientific or commercial software. These can work on standard workstations or even personal computers. It is recommended that radar data be processed from the raw data. The disadvantage is the large volume of calculations to be performed, but although these take a long time, they are controlled by simple parameters and do not cause any real difficulties. On the other hand, raw data are easier to concatenate than processed data (in the event that a site should extend beyond the length of a single image). Interferometry works best when the processing parameters of the scenes to be combined are identical rather than optimized for each scene, which would be the case for scenes processed by the usual method. Anyone using raw data for interferometry can impose a compromise for the processing parameters, taken from the optimal values recommended for each scene. Lastly it may be less costly to purchase scenes in raw data format.

4.7.3 Principles of data selection

The choice of radar images to be combined by interferometry depends on the measurement objectives and the nature of the terrain to be observed. For studying topography, it is best to have very short time intervals (simultaneous images are ideal) and altitudes of ambiguity which are compatible with the type of relief observed. For example, if the scene includes elevation differences of a thousand meters, it is preferable to choose an altitude of ambiguity of 100 m rather than an altitude of ambiguity of 10 m, even if this means starting again with a lower altitude of ambiguity after the first pair of images has ‘approximated’ the topography. It is also always necessary to confirm the elevation map with several sets of data, in order to identify atmospheric artifacts. In the case of measurements of displacements that are regular over time, it may be necessary to wait until the displacement has reached a measurable amplitude. In the event of instantaneous displacements (such as an earthquake), it is only necessary that the images be acquired at dates that bracket the date of the event. For measuring displacement, it is preferable to choose data with a high altitude of ambiguity, in order to make it easier to eliminate topographic contributions. Climatic conditions must always be considered. For work over Iceland, the images must always be acquired in summer to avoid the snow [Vadon, 1997]. It may seem like a paradox, but it is sometimes preferable to work with images acquired at an interval of a year rather than an interval of six months because of seasonal variations. We can take advantage of radar’s night-time capability. Being an active instrument, it is its own source of illumination and night offers more stable atmospheric and vegetation conditions [Massonnet, 1995], [Massonnet, 1998]. With dense vegetation cover, it is preferable to use data acquired in L band. The best choice of data therefore depends on circumstances and on the know-how of the specialists for whom the measurements are intended and on the knowledge of the sites.

4.7.4 Possibilities for future dedicated space missions

Interferometry has produced some spectacular results using data from satellite radar systems which were not designed for this purpose. It has developed very rapidly since the launch of the ERS satellite, although isolated experiments had taken place in the United States with SEASAT data or data from radar instruments carried on the Space Shuttle. Development so far has been largely fortuitous and initially, no radar satellite had been designed specifically to exploit the functions of this technology. Interferometry has however benefited from the excellent orbital stability of radar satellites, which revisit in good conditions for interferometry (typically at less than one kilometer from the planned position) as a matter of routine. Also interferometry can be described as a robust technology because it can exploit data with less than perfect image quality.

Fig. 4.23, however, shows a circumstance for which the capacities of current radar satellites seem inadequate. In this example, a segment over two thousand kilometers long was processed in interferometry mode. The terrain is not very mountainous, so we should not see any fringes after correction for orbital uncertainties. Some of the fringes visible here can be attributed to the lack of stability of the radar’s oscillator,

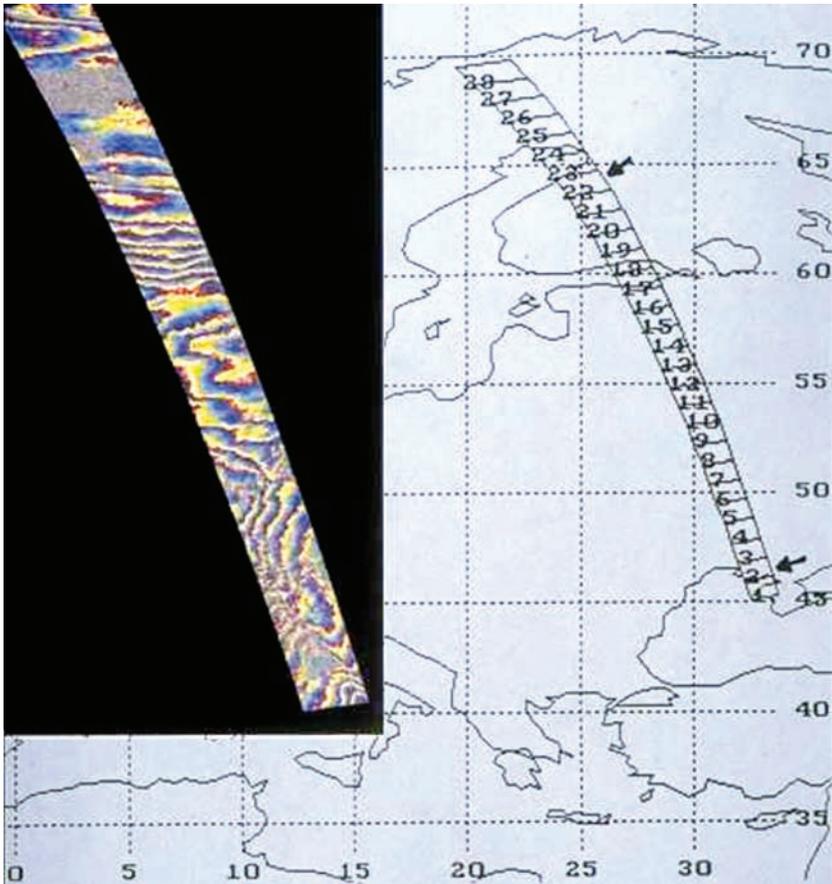


Fig. 4.23 Interferometry processing of a long ERS segment (2000 km) showing clock drift.

i.e. the clock that beats out the wavelength inside the radar instrument. This clock is theoretically designed to maintain a stable rhythm during the image synthesis, i.e. approximately one second. For this segment, acquired over 300 seconds, the clock has drifted (with an amplitude considerably less than what it is supposed to provide, so this is not a quality issue) actually interfering with itself by creating beats, several days later, when the second image was acquired.

Interferometry has been so successful that we can expect any future radar mission to take its particular requirements into account. Before planning one or more space missions dedicated to interferometry however, they must first offer definite advantages. We shall now distinguish between missions whose objective is topography mapping and those intended for monitoring displacements of land over time.

4.7.4.1 Topography

Radar interferometry has aroused great interest for calculating topography. Interferograms are so sensitive to other factors, such as variations in the thickness of the atmosphere as discussed in Section 4.6.1, that the preferred systems take simultaneous images, which allow such contributions to the two images to be cancelled out.

The principal limitation when interpreting the interferometric signal is the ambiguity caused by the different 'layers' of information that make it up. Several systems have been designed to simplify this interpretation. One way is to eliminate the 'point of view' components by keeping the orbit within a small distance to the average orbit, thus canceling the stereoscopic baseline almost entirely. Another way is to make it easier to eliminate the atmospheric component (the ionosphere effect) by integer combination using two frequencies whose ratio is either an integer or a simple fraction, like $2/3$ [Massonnet, 1999]. Most of the work has involved topographic restitution however, which benefits greatly from the elimination of the 'time' factor and of the corresponding effects, in particular the influence of the atmosphere.

Spaceborne radar systems could conceivably carry two antennas but this has mostly been implemented on airborne systems. Some of these systems offer commercial resources for topography computation and advantageous operational characteristics. When an aircraft is used, since it is much closer to its target than a satellite, the critical distance between the antennas is a few meters at most, and not a kilometer. It is therefore possible to install the two antennas on the same aircraft, separated either horizontally or vertically. The separation between the two antennas is thus very accurately known. The principal difficulty is how to precisely take into account the roll experienced by the aircraft, which can 'tip' the observation baseline.

The SRTM mission

As a general rule, it is difficult to persuade decision makers of the need to invest in optimized interferometry systems: why should they spend a lot to do a perfect job when you can do an approximate job with radar systems which are neither optimized nor even designed for interferometry?

A two-antenna space system is only of use specifically for interferometry, and its precision is limited by the difficulty in constructing a mast long enough to separate the two antennas sufficiently. However, when strategic interest and political will coincide, all obstacles are removed and this was how the Space Shuttle mission was able to map a large part of the Earth in February 2000 for the United States Department of Defense as part of the Shuttle Radar Topography Mission (SRTM) [Jordan, 1996]. A degraded version of this map has been made available to civilians throughout the world (Fig. 4.24), although access to the most precise data with typical vertical accuracy of 10 m remains restricted.

The American authorities were therefore able to have their cake and eat it too by having an efficient product available for their own security with a vertical precision of about ten meters while offering the world at large a lower resolution product that was nonetheless considerably better than existing maps and also entirely free. This ambitious mission also involved risk and boldness, as the Shuttle can only fly for about



Fig. 4.24 Example of low resolution SRTM DEM, Turkey. The horizontal posting is 90 m and the relative vertical resolution of the order of 10 m RMS.

ten days and it was necessary to cover the entire Earth in that period of time. The mission required that a sixty-meter mast be deployed so that the two radar images could be acquired simultaneously: one of the antennas was installed in the hold and the other at the end of the mast. These operational constraints affected the results. The Shuttle's orbit covered latitudes up to 60° . The coverage requirements made it necessary to use the SCANSAR mode (Sect. 2.5.4.2) and a wide range of angles, which compromised the angular homogeneity of the product and its radiometric quality. The mechanical link between the two antennas was subject to complex oscillations which had to be analyzed in order to eliminate their effects. The mission was nonetheless very successful. In addition, the German and Italian space agencies had installed a pair of radar antennas in X band (whereas the principal mission was operating in C band). These X band data had more limited coverage because they did not use SCANSAR, but they produced more accurate elevation results.

Solutions with two satellites and the interferometric cartwheel

In the case of a solution using two satellites, the system itself may not be dedicated exclusively to interferometry but may simply take advantage of two compatible satellites. In order to eliminate atmospheric effects, one of the satellites must follow the other within a few seconds so that the atmosphere does not have time to change. The ESA experiment which consisted in flying the ERS-1 and ERS-2 satellites at an interval of one day did not satisfy this condition. It would also be possible to make one of the two satellites passive so that it only needed to receive the signal which would make it cheaper, but would restrict it to the interferometry experiment.

Europe is certainly a strong contender in the global topography stakes which are of such importance in every field; by taking part in the Shuttle mission, Germany and Italy were able to gain access to a part of global coverage. The SPOT5's HRS instrument [Baudoin, 1999], [Baudoin, 2003] also has a very high operational capability for topographic calculation, using optical stereoscopy.

More ambitious projects are being designed in Europe to make topography more precise, more global and less costly. This is the case with the Interferometric Cartwheel, based on three receiver satellites which only have to receive the signal from a conventional satellite (Fig. 4.25), which calls for less expensive resources (no transmitter) and high precision: the baseline is not limited by the length of a physical mast.

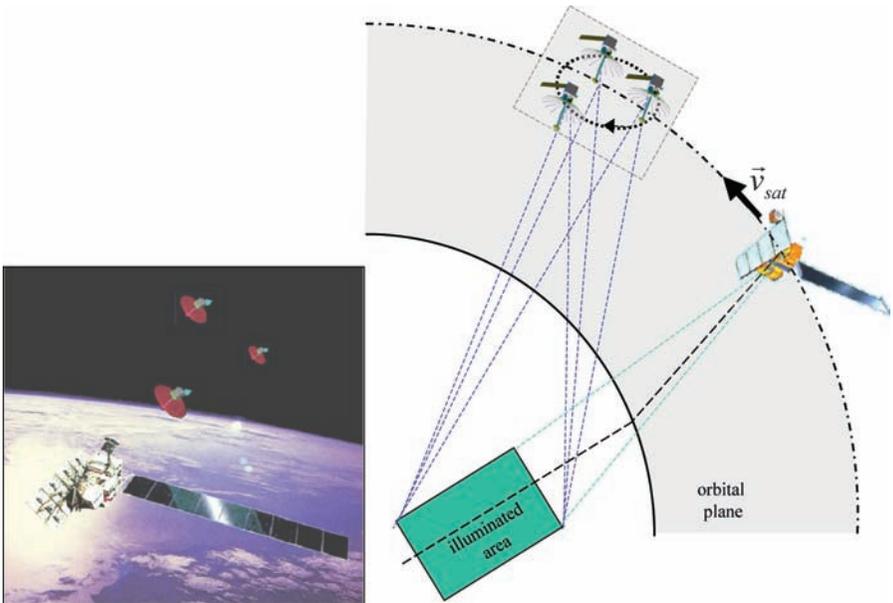


Fig. 4.25 Artist's impression of the Interferometry Cartwheel concept.

When a mission involves more than one satellite, they can be flown with a time interval such that the atmospheric effect is negligible. This was considered very seriously in 1994 when the launch of ERS-2 was being prepared while ERS-1 was in operation and in the early 2000s between the planned Radarsat-2 and Radarsat-3 satellites [Caves, 2002]. Although these projects were finally dropped, this idea is currently embodied by the Terrasar tandem [Moreira, 2004]. There are also some disadvantages however, which up to now have prevented such missions being put into practice. We shall list some of them here:

- Thematic risk: what is the maximum time interval that guarantees the absence of atmospheric effects? A cloud can travel at 100 km/h and a few tens of seconds can be enough to create a differential effect.

- Loss of lifetime: a second satellite is usually intended to prolong a mission's lifetime. A tandem mission may involve the second satellite being launched earlier, thus reducing the overall mission duration.
- Lower rate of observation: having two satellites in flight usually means doubling the number of passes and halving the revisit time for a given zone. If the satellites are acting in tandem, this advantage is lost.
- Insufficient reception capacity without significant ground investments: it is often not possible to use a second satellite because the ground stations can neither receive telemetry from two satellites simultaneously without halving the data rate, nor aim at two satellites at the same time with a single antenna. Besides, interference problems often lead to making one satellite passive during tandem operations, which means defining a specific operating mode for it.
- Slight heterogeneity of the products: it is impossible to guarantee a stable interferometry baseline throughout the orbit, which is one of the obvious advantages of solutions using a mast like SRTM.

Nevertheless, at the time of printing a new mission is being launched in X-band based on this concept (the German Terrasar X tandem mission).

Alternatively, the concept of the Interferometric Cartwheel [Massonnet, 2001], a set of receive-only microsatellites, has been suggested as a compromise between economy and efficiency. The economy is realized because the system does not need to illuminate the target; this is left to an existing standard radar satellite called the partner. Each individual receiver only requires low power which makes it possible for each one to be carried on a microsatellite with a typical mass of 100 kg. Moreover, it is not necessary to fly this microsatellite in tandem with the partner, which would entail the risk of collision and of degraded product quality due to the difference in receiving systems. The Interferometric Cartwheel involves flying a set of microsatellites at a safe distance from the partner and creating the interferometry products by combining only the images from the microsatellites. This scheme would work even if the partner satellite does not actually record any data.

One of the most important aspects of the Cartwheel concept is the stability of the interferometry baselines throughout the orbit. The trick is to use three microsatellites placed on orbits that are synchronized with each other and with the partner's orbit. Each of the receivers has the same degree of additional eccentricity with respect to the partner's orbit, and their perigees are evenly distributed throughout the orbit. Relative to the partner's orbit, the three microsatellites could be said to dance a ballet (Fig. 4.26) and to describe a relative ellipse during their orbital period, whose semi-major axis is twice the semi-minor axis (r). It can then be shown that whatever the phase in this rotation, the choice of the two best-placed microsatellites (S_1 and S_3 in the case of Fig. 4.26) results in an interferometric baseline which is never more than 7.5 % from a mean value determined by the size of the ellipse, which is itself a function of the additional eccentricity [Massonnet, 2001]. We have in fact created a virtual mast, whose length is not strictly fixed, but which does not oscillate like the one on the Shuttle. Its length

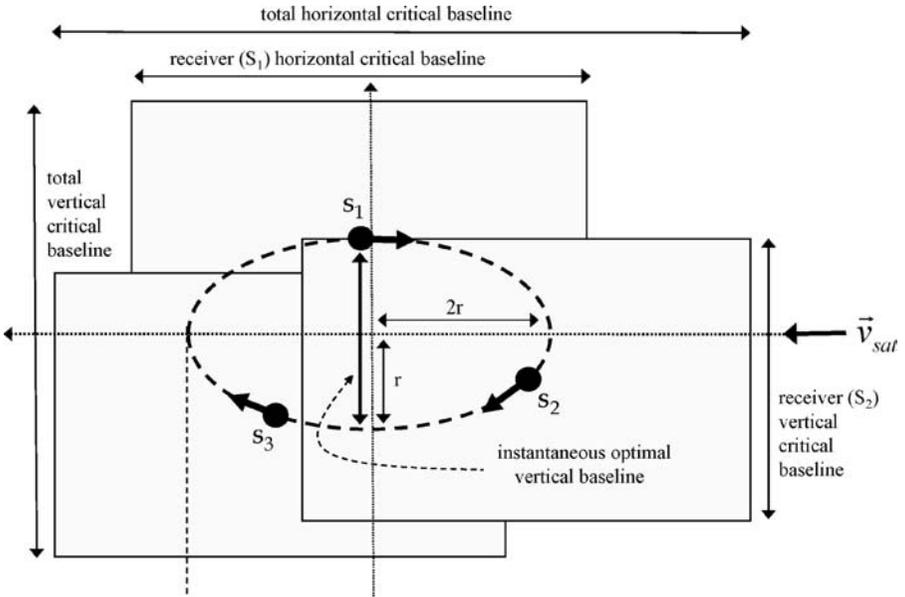


Fig. 4.26 Configuration of the 3 receivers of the wheel in the orbital plane.

is not limited, which means it can be adjusted to the optimum length for topography calculation. The figure shows the rectangles within which each microsatellite can move without exceeding the critical separation beyond which the interferometric effect would disappear either due to the difference in incidence angle in a vertical direction or to a difference in Doppler in the horizontal direction. The rectangles must therefore always overlap to a reasonable extent, so that the microsatellites, at their center, are interferometrically compatible.

In terms of quality, the trade-off chiefly concerns the microsatellite's antenna, whose azimuth footprint covers more ground than the partner's antenna does, and therefore cannot properly eliminate ambiguities (ghost images, Sect. 3.15.2). We thus accept that ambiguities will only be attenuated by the partner's antenna and not on the return trip. It can be shown however that the ambiguities, even though they are higher in Cartwheel images than in partner images, do not contribute coherently to the interferograms. They behave like a noise which does not bias the measurement (Sect. 4.9).

From the point of view of cooperation, this concept is quite innovative: cooperation is very real because the Cartwheel can produce nothing without its partner, and the partner cannot achieve the same quality of products without the Cartwheel. The necessary close cooperation is very simple to implement however, because the partner simply has to warn the Cartwheel in advance of its acquisition programme. Even closer cooperation programmes are possible of course, such as illumination modes specifically designed for Cartwheel operations, joint launches with the partner, all telemetry data being first transmitted and recorded in the partner's memory, etc.

Several studies have shown that the concept of the Interferometry Cartwheel is capable of producing a DEM of greater precision than one meter for the entire land surface observable by the partner satellite, i.e. almost the entire globe in the general case of a sun-synchronous radar [Amiot, 2004]. The concept would also seem to open up a wide range of other possibilities, some of which we list here:

- A horizontal baseline is available in the same way as the vertical baseline with the same stability conditions (it is simply twice as long). It should therefore be possible to perform interferometry along the satellite ground track for mapping ocean currents for example.
- The production of images from different ranges would allow new possibilities for characterizing mobile targets, whatever their velocity, by correlating the receivers' images.
- The existence of two other interferograms produced with the help of the 'third satellite', i.e. the one which is not part of the stable baseline currently being used, would help to eliminate elevation ambiguities in urban terrain characterized by sudden variations in elevation, through multi-baseline processing.
- As there is no time interval between acquisitions, coherence measurement is directly related to the volume properties of the targets (Sect. 5.14). The Cartwheel makes it possible to use the coherence as a new measurement of these properties which is particularly suitable for studying the biomass, usually from three simultaneous observation baselines.
- Lastly, the Cartwheel concept is perhaps a precursor of new designs for radar systems. Fig. 4.26 shows the critical vertical and horizontal baselines which can also be represented in frequency terms. For example, the critical vertical baselines are the frequency equivalent of the bandwidth of the chirp while the horizontal baselines are the equivalent of the Doppler bandwidth in azimuth. To use the vocabulary of frequency specialists, these two baselines are inversely proportional to the range and azimuth resolutions respectively of the individual receivers and also the partner's. In this way, the size of the smaller rectangle created by the intersection of two of the rectangles surrounding the receivers is related to the resolution of the interferometry product that the receivers can produce. Inversely, if we consider the 'frequency surface' covered by all the rectangles of the three receivers, it is possible to create an image product with a better resolution than that of the partner. The Cartwheel makes it possible to test these principles so that afterwards plans might be made for embedded concentric wheels by means of which numerous microsattellites could 'pave' the frequency plane and consequently considerably increase the nominal resolution of the partner satellite in both azimuth and range. Ultimately, the partner could cease recording altogether!

4.7.4.2 Ground displacements

In applications related to ground displacements, the added value of a specific system lies mainly in its ability to reuse old radar archives, so as to reveal slow displacements through pairs of images whose time interval is very large. Other technical elements are significant, but less specific: precise trajectory repetition so that topographic contributions are minimized, auxiliary instruments to give simultaneous information on the state of the atmosphere, pointing capability at multiple angles, capability for using various wavelengths. Archives can only be used if strict technical specifications are imposed on the future satellite, which must follow the same orbit as its predecessor, use the same wavelength and maintain the same orientation. Several mission concepts have been put forward. Their selection will depend on their being significantly more adapted to the task than future conventional radar satellites, which will all have an interferometric capability.

4.8 Comparison of interferometry with other methods

4.8.1 Comparison with optical stereoscopy for topographic measurement

In optical stereoscopy, relief can be calculated using the distortion caused by the difference of viewpoint between images. The critical parameters of the method are (1) the size of the pixel and (2) the ‘base to height’ ratio. The smaller the pixel, the better the distortion measurement. Depending on the measurement methods used, it is possible to detect distortion with a precision equal to a fraction of the size of the pixel, depending on the signal’s characteristics. Over given terrain, the distortion is a function of the ‘base to height’ ratio. In optical stereoscopy, it is possible to have a ratio of one, meaning that the distance between the points from where the images were taken is equal to the altitude of the image acquisition. Under these conditions, a difference of elevation in the terrain observed reveals a distortion of the same magnitude between the images. The elevation precision is then equal to a fraction of the size of the pixel, typically a few meters.

In radar interferometry, the base to height ratio is necessarily very low, as we have seen in Section 4.1.6, and is unlikely to exceed one thousandth. Even though the equations for geometric sensitivity to altitude difference are different from those used in optics, we may conclude that for a given topography, the distortion between two radar images that can be combined by interferometry is typically one thousand times less than that for stereoscopic images. On the other hand, the measurement is based on the phase difference which, depending on the wavelength and the signal power, can allow precision of the order of a centimeter. This precision compensates for the poor baseline and helps us to understand why the results of the two methods are comparable in quantitative terms. They both achieve precision of a few meters when measuring elevation. Qualitatively however, they remain very different: interferometry allows all-weather access and therefore guarantees a result. The result must be confirmed by several attempts however in order to eliminate the contribution of atmospheric effects.

Besides, certain surfaces like forests will only give mediocre results. In comparison, stereoscopy performed with an optical satellite using lateral off-pointing will have a low success probability in the event of heavy cloud cover. For example, if a region is cloud-free only 10 % of the time, the probability of obtaining a pair of cloudless images is only 1 %. If the satellite can acquire the two images when observing first forwards and then backwards (stereoscopy along the ground track), the probability will remain close to 10 % in our example. This is the case for the HRS instrument carried on SPOT-5.

Interferometry is more efficient on flat terrain with little contrast, for which it is difficult to correlate optical images. It is also more efficient with less noisy background: for weak topographic patterns, interferometry produces relatively smooth results whereas correlation residuals are noisier. Optical stereoscopy on the other hand is more efficient in mountainous terrain, where the different types of surface and illumination provide good contrast. In these regions, the geometrical disadvantages of radar such as overlapping and foreshortening have a major impact. It is also difficult to count and unwrap the fringes in these zones, as they will be numerous and possibly broken.

Each of these techniques leads to very different types of errors: for optics, atmosphere is 'all or nothing'. If there are clouds, no measurement is possible. For radar interferometry, atmosphere introduces a measurement bias which can only be detected by comparing several results for the same site. In radar, error is therefore trickier to handle.

4.8.2 Comparison with GPS for measuring displacement

At the time of writing, no other technology can compare with interferometry for measuring ground displacement from space. Although displacement can be measured by comparing optical images acquired before and after events, it is unlikely that measurements could be taken with better resolution than the decimeter level using current civilian optical satellites. High precision instantaneous measurements can be obtained using specific ground equipment, particularly differential GPS technology.

The advantages of GPS technology or positioning systems in general are (1) it depends very little on the nature of the ground on which the receiver is installed, as long as the ground is stable and the site is relatively clear, and (2) it is capable of taking measurements in three dimensions, even if the vertical component gives a less precise measurement than the other components. Depending on conditions and the time available for integration, measurement accuracies can be much better than a centimeter. A final advantage worth mentioning is since measurements are continuous, atmospheric effects harmful to interferometry can be filtered out.

The advantages of interferometry are (1) lower cost (in both cases, GPS and radar, we can exclude satellite costs, because missions involving each of these space systems go beyond geodesy studies), and (2) very high spatial density for the measurements, generally at least one point per hectare, allowing for greater understanding of phenomena with complex morphology.

An analysis of post-seismic distortion at Landers in California (Fig. 4.27) clearly shows how interferometry can complement GPS-type measurements. The 5-km scale

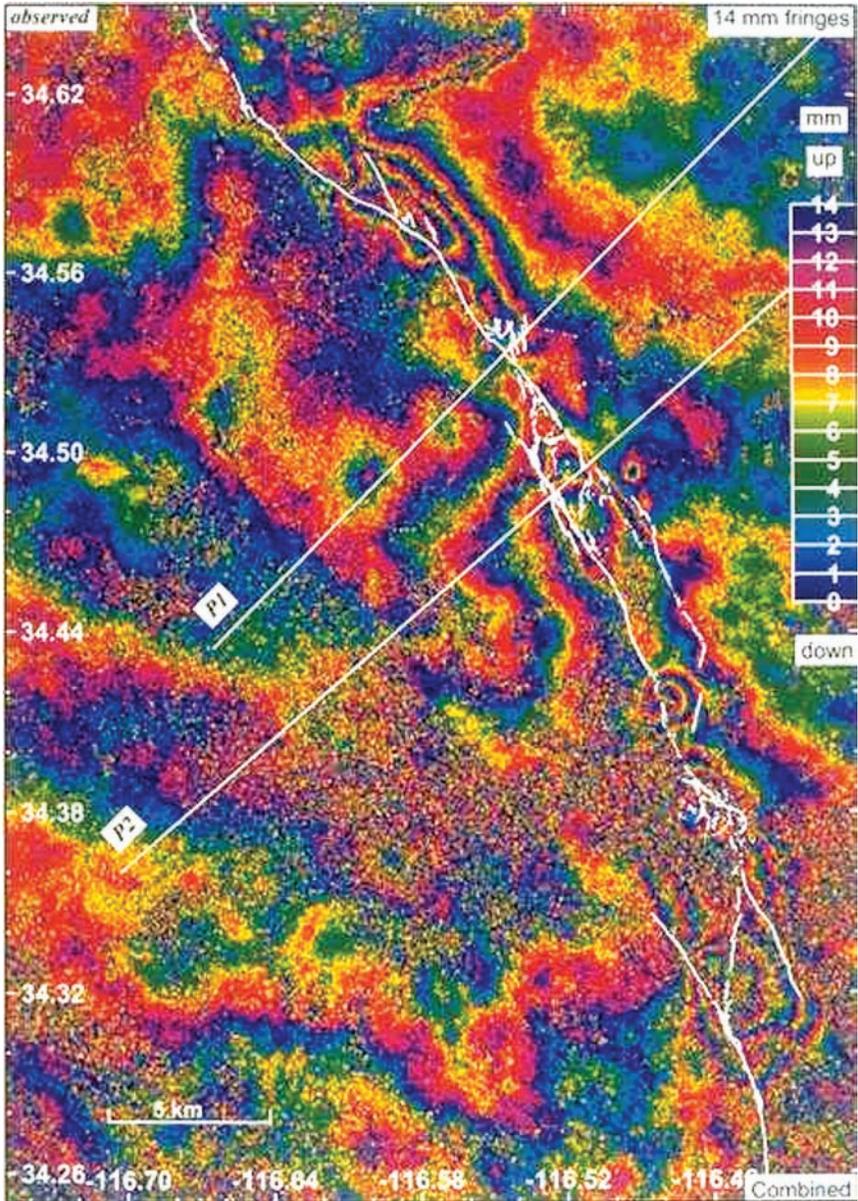


Fig. 4.27 Analysis of post-seismic displacement, on the Landers site (California), based on complementarity between interferometry and GPS.

shows that the typical measurement density of which interferometry is capable is completely inaccessible to GPS, because it would be necessary to cover the region with GPS receivers to give a full account of the complexity of the displacements which have

affected the terrain. However, if GPS measurements are performed correctly, they should not be affected by changes in ground surface features. Most of these displacements occur during the month following the earthquake and interferometry is the only method that records them correctly. In this example, two post-seismic interferograms have been added, which means that each fringe represents 14 mm, and not the 28 mm that are usual for ERS.

4.9 Robustness of coherent processing when faced with ambiguities

We now take a look at the way interferometry processing handles image ambiguities (Sects. 2.2.4, 2.52, and 3.7). Protecting images from radar ambiguities in both range and azimuth is a critical parameter for radar image quality, which strongly influences the design of radar systems. Purely as an intellectual exercise combining both radar processing characteristics and coherent combination techniques, we will now show that under certain conditions, ambiguities do not contribute to coherent combination. After all, coherent image combination assumes the co-registration of the images. The parameters of the co-registration are selected over the entire image, and therefore are optimized for the dominant, non-ambiguous targets.

‘Ghost’ images corresponding to *range ambiguities* are not properly co-registered. They come either from the echo of the preceding pulse (further away perpendicular to the satellite’s ground track) or from the beginning of the echo of the following pulse (on targets closer to the satellite’s ground track). In order to produce topographical information, both radar passes must observe from slightly offset viewpoints. The variation in incidence angle resulting from this offset must not exceed a critical value such that the progressive shift of the two images remains less than one wavelength per pixel (Sect 4.5) as a rule. The relative shift must therefore remain inferior to the ratio between the sampling frequency and the carrier frequency ($1/Q$), i.e. usually between 0.3 % and 1 % depending on the satellite. The lower limit of the offset depends on the necessity of creating a sufficient topographic effect. Assuming that this requires a baseline that is at least one tenth of the critical perpendicular baseline, we conclude that the relative shift between the images is at least 0.03 % to 0.1 %. In geometrical terms, this is the same as equating the base to a tenth of the ‘critical perpendicular baseline’. What is the effect of this progressive shift on images containing range ambiguities? The number of range pixels separating the real image from the ambiguous image is equal to the ratio of the range sampling frequency f_d to the pulse repetition frequency f_a (Sect. 2.5.2). This ratio is typically equal to 10, 000 (which is the case for ERS-1). If we take realistic lower values of the relative shift, the ambiguous images remain shifted by several tens of pixels in range when the real images coincide. Range ambiguities do therefore not usually give coherent combinations.

Azimuth ambiguities are characterized by degraded resolution relative to the nominal images because their parabolic migrations have been improperly compensated for. Even in the case of a satellite such as ERS, which produces particularly small parabolic migrations, the range spread corresponds to 2.5 pixels. This spread in range causes a

degradation of a similar magnitude in azimuth (see below). Since the resolution cell is larger, the critical perpendicular and horizontal baselines are reduced (by a factor of 2.5 in the case of ERS). Hence, if the baseline reaches 40 % of the critical perpendicular value, azimuth ambiguities can no longer be coherently combined, as their perceived baseline exceeds the critical baseline.

Under these conditions, the degradation of the interferograms is only the result of the *incoherent* contribution of ambiguities. Even if these ambiguities are as high as -10 dB of the real image (Sect. 3.7.2), they cannot corrupt the phase of the real image with a standard deviation of more than 5 % of a cycle. Under no circumstances can they bias the result. They can only add noise.

Let us review the underlying mathematics using the radar *carrier frequency* f_c , *the range sampling frequency* f_d and *the pulse repetition frequency* f_a . Referring to the range ambiguity shift, let us assume we have a configuration that creates α times the critical perpendicular baseline ($0 < \alpha < 1$). We therefore have a shift in range of one pixel for every n pixels with: $n = f_c/(\alpha f_d)$. The interval in pixels between two successive pulses is equal to $m = f_d/f_a$. If we have a perfect co-registration, the range offset remaining between the ambiguous contributions is a number of pixels, which depends on the three typical frequencies of SAR: $m/n = \alpha f_d^2/(f_c f_a)$.

In the case of ERS, this term has a value of 40 and therefore, even if the baseline is one tenth of the critical baseline, the shift is as much as 4 pixels for the ambiguous contributions. For a satellite using *L* band, this term can easily reach 150.

Now let us look again at the azimuth ambiguity shift. A radar observing at zero Doppler (perpendicular to the ground track), analyzes an azimuth frequency range from $-f_a/2$ to $f_a/2$. This range of observations corresponds to the increasing distance, relative to the point of closest approach, proportional to the square of the frequency, namely $\beta f_a^2/4$, where β is a geometric factor specific to the radar system for expressing increased distances (i.e. range migrations) in units of range pixel.² The azimuth ambiguity, produced by the same frequency bandwidth shifted by f_a , i.e. from $f_a/2$ to $3f_a/2$, spreads between the increased ranges $\beta f_a^2/4$ and $9\beta f_a^2/4$. Since the ambiguity is processed in the same way as the real image, the real image's migrations are compensated, up to $\beta f_a^2/4$. The ambiguous targets thus experience uncompensated migrations between 0 and $2\beta f_a^2$, resulting in a 'range spread' of the ambiguous target ($8\beta f_a^2/4$). This corresponds to a degradation of the range resolution by a factor which is eight times the maximum value of the parabolic migration, which in turn decreases the critical interferometric baseline by the same factor. For the ERS satellite, the amplitude of the migrations is 2.5 range pixels. For a satellite working in *L* band, such as JERS, the spread can be as high as 150 pixels. Range spread leads to the same order of spread in azimuth, measured in numbers of pixels (Sect. 2.3.6). The ambiguous target therefore shows a much wider impulse response (the widening factor is 2.5 for ERS and 150 for JERS) which rapidly makes it unsuitable for interferometry as it narrows the critical baseline by the same factor. For example, the azimuth ambiguities of a satellite using *L* band become incoherent before the perpendicular

²We have seen (Sect. 2.3.6) that $\mu(f_r) = \frac{N_a f_r^2}{2Q} = \frac{\lambda R_0}{2v^2} \cdot \frac{1}{2Q} \cdot f_r^2 f_a^2 = \beta f_r^2 f_a^2$. For a non-ambiguous signal, $-1/2 \leq f_r \leq 1/2$, the maximum migration is therefore indeed $\beta f_a^2/4$.

baseline reaches 1 % of the critical perpendicular baseline. In the same way, the degradation of the azimuth resolution prohibits the ambiguous targets from taking part in the interferometric processing because they ‘see’ a much reduced horizontal critical baseline.

The explanations in Section 4.9 are further developed in [Massonnet, 2001] and [Fjørtoft, 2004].

4.10 Permanent reflectors

One of the most difficult aspects of interferometry is maintaining the ground surface characteristics between acquisitions separated by a time interval. There are several kinds of terrain that produce interferograms that are partially or totally incoherent. However, while we understand why a field, a forest or a marsh gives incoherent results, we also know that within this instability there are always stable targets to be found (rocky outcrops in forests, separation walls between agricultural fields, etc.). The problem is that the quality of the fringes always depends on the continuity of the results, which is also required for reconstructing the number of fringes by phase unwrapping. Coherence is no more or less than a measure of the continuity. Without continuity, how can we know whether a pixel has a valid phase value when its neighbors are giving random results?

It is very important to identify these stable pixels or ‘permanent reflectors’ because in the absence of a clear interferogram, a few dozens of these points may be sufficient to characterize for example the evolution of regional ground displacements. The chief difficulty is to identify them because once that has been done, the very high consistency of the geometry of radar images will make it easy to find them again from one image to the next.

Several techniques have been suggested to find and exploit these reflectors. The earliest and most successful [Elachi, 1982], [Ferretti, 2000] have been mostly applied to urban landscapes which are intrinsically stable and therefore not typical of the areas where such techniques are the most wanted. However in these conditions the technique called ‘permanent scatterers’ has led to very interesting results and, in particular, very accurate down to the millimeter scale or even below. The method relies on the analysis of a very long series of images, typically a few dozens. All the images are “slaved” and combined to the same master, which limits the number of interferograms to the number of images minus one. Since the stable ‘permanent scatterers’ (PS) are likely to be point-like features not necessarily subject to speckle, interferograms with baselines higher than the critical baseline are kept in the process. However with the largest baselines, the more difficult geometric conditions are very demanding for permanent scatterers candidates, which reduces the number of appropriate points. A typical collection of PS is generally sparse and not evenly distributed.

The first selection criterion is the consistency. Then the conditions to be observed by the would-be PS are expressed as a non-linear equation linking the observed phases in the series with the conditions (such as the baseline) and several unknowns (such as a constant phase offset linked to each image, phase slopes along azimuth and range, or

geophysical hypotheses such as a sinking rate). The non-linear nature of the equations is linked to the fact that, at this stage, the phases are still wrapped (i.e. 2π ambiguous). The equations can then be solved iteratively if the signal to noise ratio is sufficient and assuming some additional and likely hypotheses, such as modeling the impact of the atmosphere by an “atmospheric phase screen” consisting of a phase plane associated with each interferogram. A key aspect of the process is the exploitation of the specific behavior of the components of the signal: the atmosphere is characterized by low spatial frequencies and high time frequencies (at the scale of satellite passes). Using adapted filtering, residuals end up by being attributed to atmosphere, topographic errors, orbital offsets etc.

Although this method lead to very spectacular results and raised hopes for further improvements, the discovery of a method for detecting permanent reflectors in natural landscapes is of critical importance for the future of differential interferometry. Possible methodological improvements may include the use of all the interferograms possible from a series of images, or only part of them but with a different selection scheme, for example making sure the same atmospheric contribution is not seen twice in the series [Mouelic, 2005]. These improvements are likely to make a more thorough exploitation of auxiliary data on a given site (geophysical models, atmospheric measurements assimilated in a model, etc).

References

- [1] T. Amiot, G. Krieger, F. Douchin, M. Werner, R. Fjortoft, ‘TERRASAR-L interferometric cartwheel: digital elevation model performance estimation’, in *Proceedings Radar’04 Symposium*, Toulouse, (2004).
- [2] A. Baudoin, ‘The current and future SPOT program’, *ISPRS Workshop*, Hanover, (1999).
- [3] A. Baudoin, M. Schroeder, C. Valorge, M. Bernard, V. Rudowski, ‘The HRS-SAP initiative: a scientific assessment of the HRS instrument on board of SPOT-5 by ISPRS investigators’, *ISPRS Workshop*, Hanover, (2003).
- [4] R. Caves, A. P. Luscombe, P. K. Lee, K. James, ‘Topographic performances evaluation of the RADARSAT-2/3 tandem mission’, in *Proceedings of the IGARSS’02 Conference*, Toronto, Canada, (2002).
- [5] C. Elachi, ‘Radar Images of the Earth from Space’, *Scientific American*, 274 (12), 46–53, (1982).
- [6] A. Ferretti, C. Pratti, F. Rocca, ‘Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry’, *IEEE Transactions on Geoscience and Remote Sensing*, 38(5), (2000).
- [7] A. Ferretti, C. Pratti, F. Rocca, ‘Permanent scatterers in SAR interferometry’, *IEEE Transactions on Geoscience and Remote Sensing*, 38(5), (2001).

-
- [8] R. Fjørtoft, J. C. Souyris, J. M. Gaudin, Ph. Durand, D. Massonnet, 'Impact of Ambiguities in Multistatic SAR - Some Specificities of an L-band Interferometric Cartwheel', *Proc. International Geoscience and Remote Sensing Symposium (IGARSS'04)*, Anchorage, Alaska, (2004).
- [9] A. K. Gabriel, R. M. Goldstein, H. A. Zebker, 'Mapping small elevation changes over large areas: differential radar interferometry', *Journal of Geophysical Research*, 94 (B7), 9183–9191, (1989).
- [10] F. Gatelli, A. M. Guarnieri, F. Parizzi, P. Pasquali, C. Prati, F. Rocca, R. L. Jordan and al., 'The wavenumber shift in SAR interferometry', *IEEE Transactions on Geoscience and Remote Sensing*, 32(4) (1994).
- [11] D. C. Ghiglia and M. D. Pritt, *Two-dimensional Phase Unwrapping*, John Wiley and Sons, (1998).
- [12] L. C. Graham 'Synthetic interferometer radar for topographic mapping' *Proc. IEEE*, 62, 763–768, (1974).
- [13] Goldstein, Engelhardt, Kamb, Frolich, 'Satellite Radar Interferometry for Monitoring Ice Sheet Motion: Application to an Arctic Ice Stream'. *Science*, 262, 1282–1285, (1993).
- [14] R. L. Jordan et al., 'Shuttle Radar Topography Mapper (SRTM)', *Microwave Sensing and Synthetic Aperture Radar* (proc. SPIE), 412–422, (1996).
- [15] D. Massonnet, T. Rabaute, 'Radar Interferometry: Limits and Potential', *IEEE Transactions on Geoscience and Remote Sensing*, 31(2) (1993).
- [16] D. Massonnet, M. Rossi, C. Carmona, F. Adragna, G. Peltzer, K. Feigl, T. Rabaute, 'The displacement field of the Landers earthquake mapped by radar interferometry', *Nature*, 364, 138–142, (1993).
- [17] D. Massonnet, P. Briole, A. Arnaud, 'Deflation of mount Etna monitored by spaceborne radar interferometry', *Nature*, 375, 567–570, (1995).
- [18] D. Massonnet, W. Thatcher and H. Vadon, 'Detection of postseismic fault-zone collapse following the Landers earthquake', *Nature*, 382, 612–616, (1996).
- [19] D. Massonnet, H. Vadon, M. Rossi, 'Reduction of the Need for Phase Unwrapping in Radar Interferometry', *IEEE Transactions on Geoscience and Remote Sensing*, 34(2) 489–497, (1996).
- [20] D. Massonnet, K. L. Feigl, 'Radar Interferometry and its Applications to Changes in the Earth's Surface'. *Reviews of Geophysics*, 36(4), 441–500, 1998
- [21] D. Massonnet, 'Satellite radar interferometry', *Scientific American*, February 1997, traduit dans *Pour la Science* (avril 1997) 'L'interférométrie par radar'.

- [22] D. Massonnet, 'Radar imagery: looking from space with wavelengths 100000 times larger than the ones of visible light', *Journal of Visualization*, The Visualization Society of Japan, 3(1) (1999).
- [23] D. Massonnet, 'Capabilities and limitations of the interferometric cartwheel', *IEEE Trans. Geosci. Remote Sensing*, 39(3) 506–520, (2001).
- [24] A. Moreira, G. Krieger, I. Hajnsek, D. Hounam, M. Werner, S. Riegger, E. Settlemyer, 'TanDEM-X: A TerraSAR-X Add-On Satellite for Single-Pass SAR Interferometry', *Proceedings of IGARSS 2004*, Anchorage, Anchorage, USA, (2004).
- [25] S. Mouelic, D. Raucoules, C. Carnec and C. King, 'A least Square Adjustment of Multi-temporal InSAR Data : Application to the Ground Deformation of Paris', *PE&RS*, (41) 2243–2253, (2005).
- [26] G. Peltzer, P. Rosen, F. Rogez, and K. Hudnut, 'Postseismic rebound in fault step-overs caused by pore fluid flow', *Science*, 272, 1202–1204, (1996).
- [27] E. J. Rignot, S. P. Gogineni, W. B. Krabill, and S. Ekholm, 'North and northeast Greenland ice discharge from satellite radar interferometry', *Science*, 276, 934–937, (1997).
- [28] H. Vadon, and F. Sigmundsson, 'Crustal distortion from 1992 to 1995 at the Mid-Atlantic ridge, southwest Iceland, mapped by satellite radar interferometry', *Nature*, 275, 193–197, (1997).
- [29] H. Zebker, R. Goldstein, 'Topographic mapping from interferometric SAR observation', *J. Geophys. Res.*, 91, 4993–5001, (1986).

CHAPTER 5

SAR POLARIMETRY: TOWARDS THE ULTIMATE CHARACTERIZATION OF TARGETS

5.1 Introduction

Polarization diversity has been uncommon in spaceborne SAR of the last century. Most civilian systems were designed to transmit a single waveform in terms of polarization, such as the ERS-1 and ERS-2 satellites (vertical linear polarization), or RADARSAT-1 and JERS (horizontal linear polarization). New space missions however all include experimental polarimetric modes to various extents. For the first time, ASAR/ENVISAT (launched in March 2002) was given a limited polarization diversity. The current situation has changed rapidly with the launch of full polarimetry payloads (RADARSAT-2, TERRASAR-X and ALOS), using *C* band, *X* band and *L* band respectively.

The earliest work in radar polarimetry is found in the early 1950s [Sinclair, 1950]. The theory was inspired by the work of Stokes [Stokes, 1852] and of the ‘last universal genius’ Poincaré [1989], both of whom laid the basis for a unified formalism for electromagnetic waves, regardless of their state of polarization. Sinclair defined the *scattering matrix*, which expresses the ‘ability’ of a target to modify the polarization of the waves that illuminate it. Over the following years more scientists reinforced the theory underlying radar polarimetry, for example [Deschamps, 1951], [Kennaugh, 1952] and [Graves, 1956].

During the 1960s, interest in radar polarimetry waned. Then in 1970 the publication of *Phenomenological Theory of Radar Targets* [Huynen, 1970], in conjunction with rapid technical progress, renewed interest in such research. In the 1980s and 1990s several airborne radar instruments were deployed (JPL/AIRSAR, DLR/ESAR, EMISAR, ONERA/RAMSES, TNO/PHARUS, CCRS/CONVAIR, JAXA/PISAR etc.). Many applications were developed from these exploratory campaigns, further enriched by the SIR-C/XSAR mission on the Space Shuttle. Among these are the classification of land surface [Lee, 2001], forest parameters [Le Toan, 1992], soil moisture [Dubois, 1995] surface roughness [Mattia, 1997], lava [Zebker, 1987], sea ice [Drinkwater, 1993] snow [Shi, 2000], topographical measurements [Schuler, 1996], and target detection

[Novak, 1990], [Souyris, 2003]. More recently, the combination of interferometry and polarimetry techniques (PolInSAR) [Cloude, 1998] has opened up a new range of applications. The contributions described in references [Ulaby, 1990], [Cloude, 1996], [Boerner] and [Touzi, 2004] exhaustively review radar polarimetry techniques and applications. In parallel, polarimetry toolbox became available, such as the one of the University of Rennes [Pottier, 2002]. Some of the results developed in this chapter were obtained using it.

The purpose of this chapter is to give the necessary keys for understanding polarimetric measurements. This choice forces us to address a number of different topics, without necessary detailing them as much as they would deserve. For a full understanding it is strongly recommended to first read Sect. 1.1 of this book on the propagation and polarization of electromagnetic waves. Here, we shall start with a discussion of the operating principles of polarimetric radar and its very specific timing analysis (Sect. 5.2). We introduce the *scattering matrix* in Sect. 5.3. Its expression in the monostatic case (in which transmitting and receiving antennas share the same location) is developed for single-bounce scattering, double-bounce scattering and diffractions in Sect. 5.4. The rest of the chapter is restricted to this monostatic geometry. In this case, the scattering matrix becomes the *backscattering matrix*. The principles of polarization synthesis are developed in Sect. 5.5, characteristic polarizations and Euler parameters in (Sect. 5.6).

We shall then see that there are basically two approaches for decomposing polarimetric measurements, i.e. coherent and incoherent. Coherent decompositions (Sect. 5.7) ignore the effects of depolarization caused by wave-surface interactions. When these are taken into account, the *Stokes* formalism, the *Mueller* matrix (Sect. 5.8), and the *covariance* or *coherence* matrices (Sect. 5.9) are used instead of the backscattering matrix formalism. In this context, the polarimetric measurements are decomposed using so-called ‘incoherent’ techniques (Sect. 5.10). In Sect. 5.11 examples of incoherent analyses of satellite- and airborne images are used to illustrate the theoretical development from the previous section and to underline certain of its limitations. An original method for visualizing polarimetric information is presented in Sect. 5.12.

We evaluate various architectures for compact polarimetry instruments in Sect. 5.13, which could facilitate the development of polarimetry for spaceborne systems on smaller and cheaper platforms. In the final section (Sect. 5.14), we address the merging of polarimetry and interferometry (PolInSAR). We draw conclusions in Sect. 5.15.

5.2 Radar polarimetry: operating principle

Its operating principle (Fig. 5.1) is based on the quasi-simultaneous transmission of two linear orthogonal waves (Sect. 1.1.2) (quasi-simultaneous with respect to the coherence time of the observed targets). In reception, the collected echoes are received on two linear orthogonal polarizations which have the same phase reference.

More precisely, let us assume that at a given instant the radar transmits along direction \hat{k}^{inc} a pulse whose polarization (Jones vector, Sect. 1.1.2) $E_0^{\hat{k}^{inc}}$ is directed

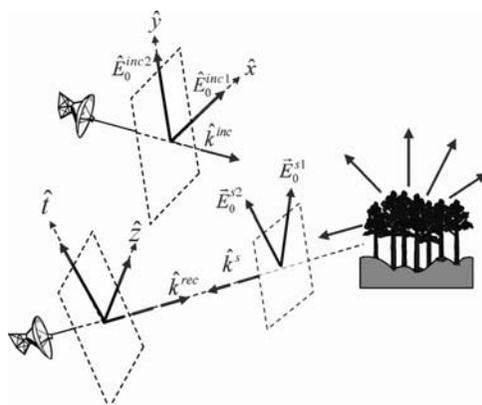


Fig. 5.1 Principle of radar polarimetry.

along the \hat{x} axis. The target backscatters along direction \hat{k}^s an electric field \vec{E}_0^{s1} whose direction¹ is unknown. However, two components of this scattered field, namely $\hat{z} \cdot \vec{E}_0^{s1}$ and $\hat{t} \cdot \vec{E}_0^{s1}$ (where " \cdot " represents the dot product), will be collected on the two receiving polarizations directed along the \hat{z} and \hat{t} directions. A moment later, a pulse with Jones vector \vec{E}_0^{inc2} oriented along the \hat{y} axis is transmitted. Similarly, the components $\hat{z} \cdot \vec{E}_0^{s2}$ and $\hat{t} \cdot \vec{E}_0^{s2}$ are collected.

5.2.1 Timing analysis - impact on system design

The main constraint in polarimetric measurements comes from the need to transmit two phase-locked waveforms with orthogonal polarizations. It requires to transmit interleaved pulses, alternating the polarization after each transmitted pulse (\hat{x} , \hat{y} , \hat{x} , \hat{y} etc). Each echo collected requires the recording of two signals, each one associated with one of the receiving polarizations. What are the consequences of this double polarization agility (both in transmission and in reception)? Regarding the timing analysis, if we want to maintain, on each polarization channel, performances comparable to that of a single polarized radar with an azimuth sampling frequency f_a (Sect. 2.2.3), the polarimetric radar must operate at $2f_a$. As a consequence of the doubling of both sampling frequency and reception modes, the volume of data generated per image pixel is increased by a factor of four. Another consequence of doubling f_a is that the radar's swath width is reduced by a factor of two, in order to protect the receiver against range ambiguities (Sects. 2.5.2 and 3.7.1). The antenna width is multiplied by a factor of two (and thus so is its surface), while the total transmitted power may remain the same. In conclusion, the polarimetry option involves multiplying the volume of data per image pixel by four and halving the swath width, which are heavy constraints that tend to

¹For the sake of clarity, we assume here that the target is not depolarizing. It is important to be aware of the difference between 'polarization change' (i.e., modifying the orientation of the electric field) and 'depolarization' (i.e., creating an incoherent component in the scattering). Depolarization mechanisms are not introduced until the second part of this chapter (Sect. 5.8).

argue against polarimetry. We shall see later that compact polarimetric architectures can relieve these intrinsic constraints (Sect. 5.13).

5.3 The scattering matrix

The acquisition strategy described above associates a batch of information stored in matrix form for each pixel according to Eq. (5.1):

$$\left[\bar{S} \right] = \begin{bmatrix} S_{xz} & S_{yz} \\ S_{xt} & S_{yt} \end{bmatrix} = \begin{bmatrix} \hat{z} \cdot \vec{E}_0^{s1} & \hat{z} \cdot \vec{E}_0^{s2} \\ \hat{t} \cdot \vec{E}_0^{s1} & \hat{t} \cdot \vec{E}_0^{s2} \end{bmatrix} \quad (5.1)$$

where \bar{S} is called the target *scattering matrix*. The indices i and j of the complex coefficients S_{ij} refer respectively to the transmission modes (x or y) and reception modes (z or t).^{2,3} The polarization basis in the transverse planes (\hat{x}, \hat{y}) and (\hat{z}, \hat{t}) are oriented according to the Back Scattering Alignment convention (BSA), which makes the reception vector triplet $(\hat{z}, \hat{t}, \hat{k}^s)$ left-handed (Fig. 5.1). The advantage of BSA is that (\hat{x}, \hat{y}) and (\hat{z}, \hat{t}) are identical in a monostatic configuration (Sect. 5.3.1).

From the matrix \bar{S} , the response of the target to any elliptical polarization of Jones vector \hat{E}_0^{inc} can be computed immediately. \hat{E}_0^{inc} is indeed the linear combination of two waves with linear orthogonal polarizations \hat{x} and \hat{y} (Sect. 1.1.2.2)⁴:

$$\hat{E}_0^{inc} = (\hat{x} \cdot \hat{E}_0^{inc}) \cdot \hat{x} + (\hat{y} \cdot \hat{E}_0^{inc}) \cdot \hat{y} = E_{0x}^{inc} \cdot \hat{x} + E_{0y}^{inc} \cdot \hat{y} \quad (5.2)$$

from which, assuming the linearity of the measurement process described in Sect. 5.2:

$$\begin{bmatrix} E_{0z}^s \\ E_{0t}^s \end{bmatrix} = \begin{bmatrix} S_{xz} & S_{yz} \\ S_{xt} & S_{yt} \end{bmatrix} \cdot \begin{bmatrix} E_{0x}^{inc} \\ E_{0y}^{inc} \end{bmatrix} \quad (5.3)$$

E_{0x}^{inc} and E_{0y}^{inc} are the components of the incident field (i.e., the incident Jones vector) \hat{E}_0^{inc} expressed in the transmission basis (\hat{x}, \hat{y}) ; E_{0z}^s et E_{0t}^s are those of the scattered field (i.e. the scattered Jones vector) \vec{E}_{0x}^s expressed in the reception basis (\hat{z}, \hat{t}) (the propagation terms in $\exp(-jkr)/r$ are omitted here).

Equation (5.3) thus links the Jones vectors of the transmitted and scattered waves, meaning that the target is characterized here by its ‘ability’ to modify the polarization of the illumination wave. What we have just described very broadly is the so-called *polarization synthesis* technique, which we will look at in more detail in Sects. 5.5 and 8.1.3.

²The coupling of the reception polarization with the scattered field is explained in Sect. 1.1.3.1

³For a single polarized radar, (such as ERS-1 (VV) or RADARSAT-1 (HH)) only one term of the scattering matrix is measured, for example $\hat{z} \cdot \vec{E}_0^{s1}$ when transmission is along \hat{x} and reception is along \hat{z} .

⁴Our choice to represent polarization \hat{E}_0^{inc} with a magnitude of 1 implies that $(E_{0x}^{inc})^2 + (E_{0y}^{inc})^2 = 1$.

5.3.1 The monostatic singularity - the backscattering matrix

Although references to multistatic SAR systems are beginning to appear, current systems function in a monostatic configuration, in which the same antenna transmits and receives. As a result, we can write (Fig. 5.1):

$$\hat{z} = \hat{x} \quad \hat{t} = \hat{y} \quad \hat{k}^{rec} = \hat{k}^{inc} = -\hat{k}^s \quad (5.4)$$

where \hat{k}^{rec} is the wave vector (Sect. 1.1.2) of the receiving polarization. If the observed media do not contain non-reciprocal material and if the wave propagation between the radar and the ground does not involve non-reciprocal phenomena (such as when low frequencies cross the ionosphere), then applying the reciprocity theorem [Kong, 1990] to the target-radar system implies that the off-diagonal terms of $[\bar{\bar{S}}]$, renamed here the *backscattering matrix*, are equal:

$$S_{yx} = S_{xy} \quad (5.5)$$

The backscattering matrix is consequently a complex diagonal square matrix, defined by six parameters, three amplitude terms and three phase terms. If we factor out a term of absolute phase (for example that of S_{xx}) the number of independent parameters is reduced to five, namely three amplitude terms and two phase difference terms:

$$\begin{aligned} [\bar{\bar{S}}] &= \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix} \\ &= \exp[j \cdot \phi_{xx}] \cdot \begin{bmatrix} |S_{xx}| & |S_{xy}| \cdot \exp[j(\phi_{xy} - \phi_{xx})] \\ |S_{xy}| \cdot \exp[j(\phi_{xy} - \phi_{xx})] & |S_{yy}| \cdot \exp[j(\phi_{yy} - \phi_{xx})] \end{bmatrix} \end{aligned} \quad (5.6)$$

This is the configuration we shall now retain. The term \hat{x} is replaced by \hat{h} (for horizontal direction) and \hat{y} by \hat{v} (for vertical direction): $S_{xx} = S_{hh}$, $S_{yy} = S_{vv}$, $S_{xy} = S_{hv}$.

5.3.2 Target vector

It may sometimes be useful to ‘vectorize’ the matrix $[\bar{\bar{S}}]$, by transforming it into a vector $\vec{\kappa} = (\kappa_0, \kappa_1, \kappa_2)^T$ by:

$$\kappa_i = Tr([\bar{\bar{S}}] \cdot [\bar{\bar{\Psi}}]_i) \quad (5.7)$$

The entity T is the transpose operator, and the operator $Tr([\bar{\bar{A}}])$ sums the diagonal terms of the matrix $[\bar{\bar{A}}]$. There are two options for choosing the structure of $[\bar{\bar{\Psi}}]_i$, from either the Ψ_L or the Ψ_P families:

$$\Psi_L : \quad \Psi_{L0} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}; \quad \Psi_{L1} = \begin{bmatrix} 0 & 0 \\ \sqrt{2} & 0 \end{bmatrix}; \quad \Psi_{L2} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (5.8)$$

$$\Psi_P : \quad \Psi_{P0} = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \Psi_{P1} = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix};$$

$$\Psi_{P2} = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (5.9)$$

Equation (5.9) has a structure similar to the Pauli matrix group (Sect. 1.1.3.3). Ψ_L and Ψ_P respectively lead to:

$$\vec{\kappa}_L = [S_{hh}, \sqrt{2} \cdot S_{hv}, S_{vv}]^T \quad (5.10)$$

$$\vec{\kappa}_P = \frac{1}{\sqrt{2}} \cdot [S_{hh} + S_{vv}, S_{hh} - S_{vv}, 2 \cdot S_{hv}]^T \quad (5.11)$$

5.4 Standard forms of backscatter

We shall now describe three examples of backscattering matrices representing basic non-depolarizing mechanisms: odd-bounce scattering, even- and dipole diffraction.

5.4.1 Odd-bounce (single, triple) scattering

Single-bounce scattering (or single scattering) corresponds to a unique interaction between a wave and a dielectric or conducting surface (Fig. 5.2(a)). This surface is locally approximated by its tangent plane containing the tangent \hat{t} and cotangent \hat{c} . At first order, this surface has a polarimetric behavior similar to the one of an infinite perfectly conducting plane illuminated under normal incidence (Fig. 5.2(b)). When the transmitted electric fields reflected⁵ from this plane they undergo a phase offset of π : $\vec{E}_0^{s1} = -\hat{E}_0^{inc1}$ and $\vec{E}_0^{s2} = -\hat{E}_0^{inc2}$. When the phase term of the coefficient $S_{hh} = -1$ is factored out, the normalized backscattering matrix $[\bar{S}]_{SS}$ can be written, as Eq. (5.12):

$$[\bar{S}]_{SS} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (5.12)$$

Single-bounce scattering has two features: (1) it preserves the polarization orientation when linear polarizations are transmitted, resulting in the off diagonal terms of $[\bar{S}]_{SS}$ being zero; and (2) the terms S_{hh} and S_{vv} are equal. In practice, they can differ because of incidence effects (the Fresnel coefficients) and surface irregularities

⁵In the theoretical case shown in Fig. 5.2(b), reflection is total and the backscattering coefficient has a modulus of 1. This is not the case when the incidence angle is oblique. However, the behavior in both cases is similar as far as their ability to modify the polarization of the transmitted wave is concerned.

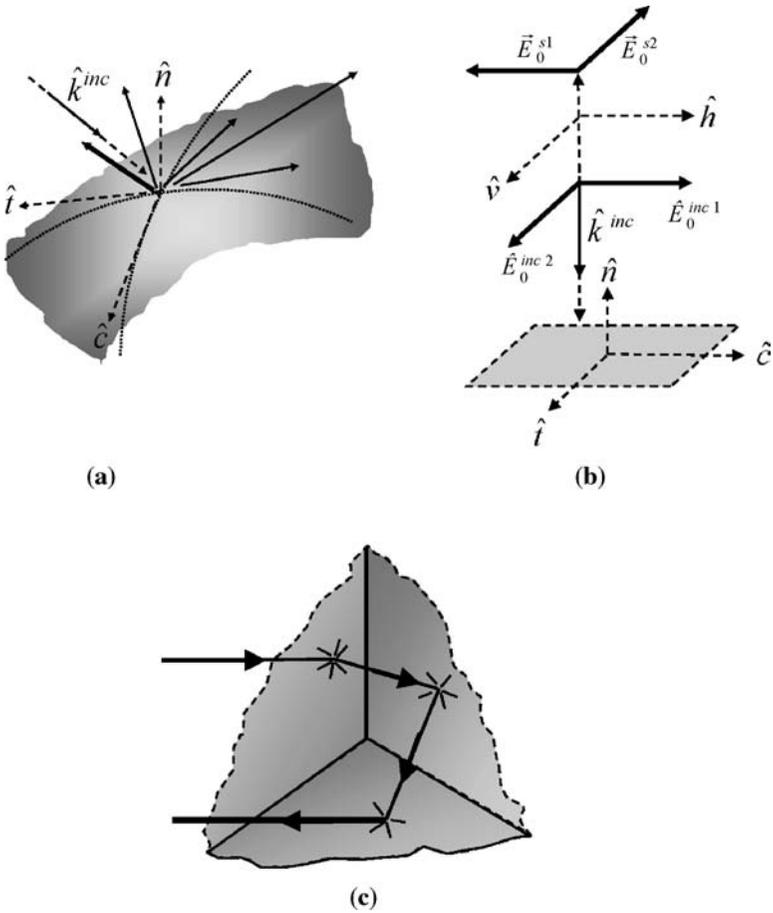


Fig. 5.2 (a) Single-bounce scattering mechanisms: general case; (b) special case of normal incidence on a perfectly conducting plate; (c) backscattering on a trihedral corner reflector.

(soil roughness and moisture, Sect. 1.3.2). Although these effects may lead to slight modifications, they do not alter the main features of the polarimetric signature.

The behavior described in points (1) and (2) will be the same for any kind of odd-bounce scattering, such as third order scattering (triple-bounce) occurring on a trihedral corner reflector, (Fig. 5.2(c)).

Please note: The Jones vector of a circularly polarized incident wave (e.g. right-handed) expressed in the basis $[\hat{h}, \hat{v}, \hat{k}^{inc}]$ is $(1, -j)^T$ (Sect. 1.1.1.2). If we apply Eq. (5.12), the Jones vector of the backscattered wave is also $(1, -j)^T$ when expressed in the basis $[\hat{x}, \hat{y}, \hat{k}^s]$ with $\hat{k}^s = -\hat{k}^{inc}$ (BSA convention). Beware however! Despite appearances, the polarization of the backscattered wave is now orthogonal to that of the transmitted wave: since they are propagating in opposite directions (Sect. 1.1.2.2), the interaction of the two waves is given by the Hermitian product $\langle \vec{E}_0^{inc} \cdot \vec{E}_0^{s*} \rangle =$

$(1, -j) \cdot (1, -j)^{*T}$, which is zero. The sphere or the perfectly conducting plane therefore backscatters all the energy of an incident circularly polarized wave in its orthogonal polarization: this shows us that polarization change does not only depend on the target, but also on the polarization of the transmitted wave.

5.4.2 Even-bounce (double) scattering

Double-bounce scattering (or double scattering) is the result of two successive single scattering events (Fig. 5.3(a)). If the two surfaces are locally orthogonal, they will create two successive specular reflections, which scatters the incident energy back to the transmitter (this is known as the *dihedral* or *double-bounce* effect). Double-bounce effects are extremely common in urban areas, which explains the high level of urban backscatter.

Even-bounce scattering occurs in particular when two orthogonal perfectly conducting planes are illuminated by an incident wave perpendicular to their vertex line.

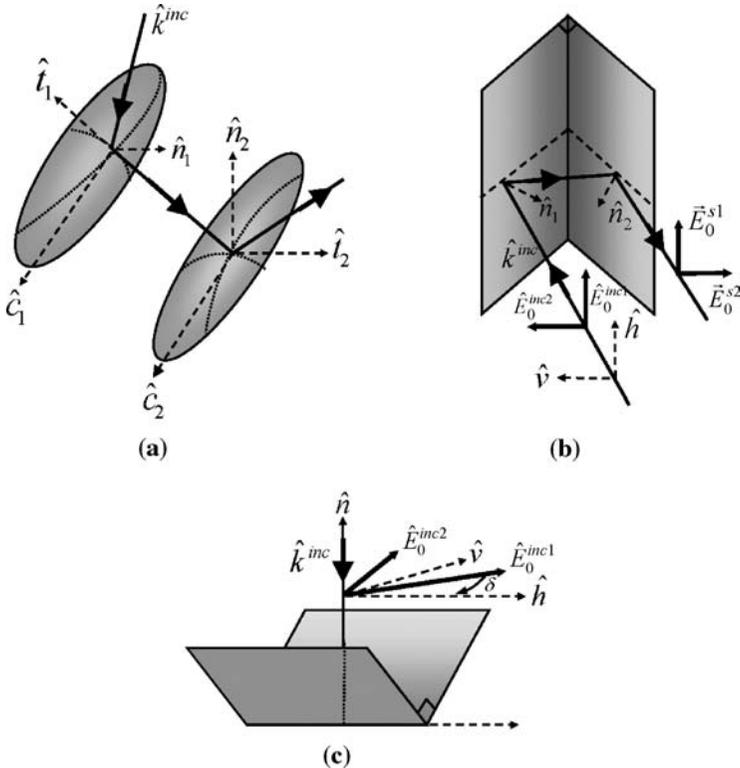


Fig. 5.3 (a) Double-bounce scattering mechanisms: general case; (b) special case of an illuminated dihedral corner reflector made up of two perfectly conducting perpendicular plates (orthogonal illumination); (c) double-bounce scattering mechanisms in a configuration where the vertex of the dihedral corner reflector makes an angle δ with the axis \hat{h} of the polarization basis.

Let us look at the situations where either the electric field or the magnetic field of the incident wave is parallel to the vertex (Fig. 5.3(b)). In the first case, the electric field \hat{E}_0^{inc1} parallel to the vertex is affected twice by a scattering coefficient equal to 1, so that the double-bounce mechanism has no effect on the phase of the electric field. When \hat{E}_0^{inc2} is transmitted, it is easier to deal with the corresponding magnetic field, which remains parallel to the vertex. This is affected twice by a scattering coefficient of 1 (since a perfectly conducting surface causes no phase offset on a magnetic field parallel to it). Therefore the phase of the corresponding electric field will be offset by π , since the structure of an electromagnetic wave forces the vector triplet made up of the electric field, the magnetic field and the propagation direction to be right-handed (Sect. 1.1.2.1). The double-bounce therefore has different effects on the phase in the two situations described above. In other words: $\vec{E}_0^{s1} = \hat{E}_0^{inc1}$ and $\vec{E}_0^{s2} = -\hat{E}_0^{inc2}$. The resulting normalized backscattering matrix $[\bar{S}]_{DS}$ is:

$$[\bar{S}]_{DS} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (5.13)$$

Double-bounce scattering causes a phase offset of π between the backscattered components of two orthogonal incident linear polarizations. From $[\bar{S}]_{DS}$, we can show that when the vertex of the dihedral corner reflector forms an angle δ with the \hat{x} axis of the polarization basis (Fig. 5.3(c)), the expression for $[\bar{S}]_{DS}^\delta$ becomes:

$$[\bar{S}]_{DS}^\delta = \begin{bmatrix} \cos 2\delta & \sin 2\delta \\ \sin 2\delta & -\cos 2\delta \end{bmatrix} \quad (5.14)$$

This configuration retains the phase-offset property highlighted in Eq. (5.13), but distributes the backscattered energy between the direct polarization (also called co-polarized) channels (S_{hh} , S_{vv} and) and the cross-polarized (S_{hv}) channel. When $\delta = 45^\circ$, all the energy is backscattered on the cross-polarized channel. This '45°' technique is sometimes used for calibrating cross-polarization channels in polarimetric systems.

5.4.3 Diffraction or dipole mechanisms

Diffraction effects are usually caused by sharp edges (Fig. 5.4(a)), which in the case of polarimetric behavior is similar to infinite straight wires or dipoles (Fig. 5.4(b)). If we assume that these wires are only sensitive to the electric field component parallel to them, we then get:

$$[\bar{S}]_{Dip} = \begin{bmatrix} \cos^2 \alpha & \cos \alpha \cdot \sin \alpha \\ \cos \alpha \cdot \sin \alpha & \sin^2 \alpha \end{bmatrix} \quad (5.15)$$

The main feature of dipole backscattering is that it filters out part of the transmitted polarization (to that extent, it is considered to be a polarizer). The special case $\alpha = 0^\circ$ leads to:

$$[\bar{S}]_{Dip}^{\alpha=0^\circ} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad (5.16)$$

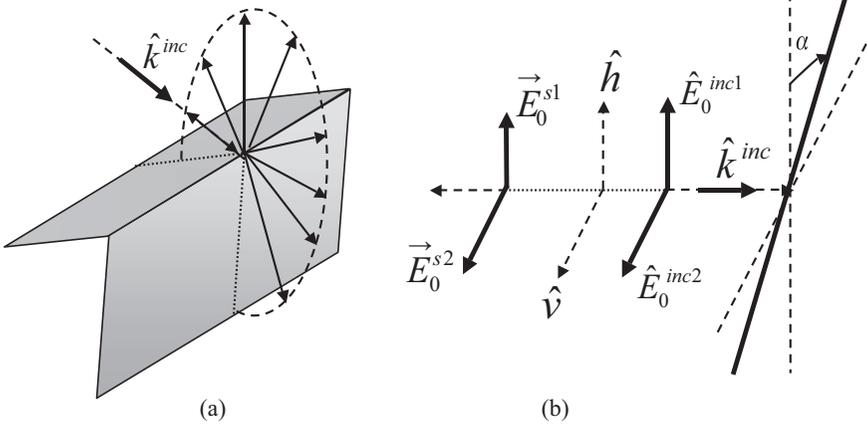


Fig. 5.4 (a) Diffraction mechanisms: general case; (b) special case of diffraction on infinitely long perfectly conducting wire.

5.5 Polarization synthesis

One major advantage of the polarimetric technique is that it enables the determination of the target response to any transmitted elliptical polarization, from the knowledge of its response to two orthogonal linear polarizations. This principle is already expressed by Eq. (5.3).

We can however establish a general expression for the backscattering matrix in any polarization basis (\hat{a}, \hat{b}) . From Sect. 1.1.2.4, we have the relationship between the Jones vector expressed in (\hat{a}, \hat{b}) and its expression in a reference basis (\hat{h}, \hat{v}) .

If we apply this relationship to the incident wave, we obtain:

$$\hat{E}_0^{inc}(a, b) = \left[\bar{\bar{P}}(\Psi, \tau, \alpha) \right]^{-1} \cdot \hat{E}_0^{inc}(h, v) \tag{5.17}$$

where “ -1 ” is the inverse operator, Ψ, τ, α are the characteristic parameters of the basis (\hat{a}, \hat{b}) (the matrix $\left[\bar{\bar{P}}(\Psi, \tau, \alpha) \right]$ of the group SU2 was defined in Sect. 1.1.2.4). Can we use the same equation for the backscattered field? Not really, because of the need for conjugating the Jones vector of a wave propagating in the opposite direction (Sect. 1.1.2.2.). We must therefore conjugate the transition matrix linking the fields $\vec{E}_0^s(a, b)$ and $\vec{E}_0^s(h, v)$:

$$\begin{aligned} \vec{E}_0^s(a, b) &= \left[\bar{\bar{P}}(\Psi, \tau, \alpha) \right]^{-1*} \cdot \vec{E}_0^s(h, v) = \left[\bar{\bar{P}}(\Psi, \tau, \alpha) \right]^T \cdot \vec{E}_0^s(h, v) \\ &= \left[\bar{\bar{P}}(\Psi, \tau, \alpha) \right]^T \cdot \left[\bar{\bar{S}} \right] \cdot \vec{E}_0^{inc}(h, v) \end{aligned} \tag{5.18}$$

This is a result of the relations that apply to the SU2 matrices (Sect. 1.1.2.2.). Under these conditions:

$$\vec{E}_0^s(a, b) = \left[\bar{\bar{P}}(\Psi, \tau, \alpha) \right]^T \cdot \left[\bar{\bar{S}} \right] \cdot \left[\bar{\bar{P}}(\Psi, \tau, \alpha) \right] \cdot \vec{E}_0^{inc}(a, b) \quad (5.19)$$

Finally, for the backscattering matrix expressed in the basis (\hat{a}, \hat{b}) we get:

$$\left[\bar{\bar{S}} \right]_{(\hat{a}\hat{b})} = \left[\bar{\bar{P}}(\Psi, \tau, \alpha) \right]^T \cdot \left[\bar{\bar{S}} \right] \cdot \left[\bar{\bar{P}}(\Psi, \tau, \alpha) \right] \quad (5.20)$$

A matrix transform of the type shown in Eq. (5.20) was first introduced to radar polarimetry by Ernst Luneburg [Luneburg, 1996]. It is called a con-similarity transform.

For example, let us now express the backscattering matrix in a right-hand (\hat{r}) - left-hand (\hat{l}) circular polarization basis. From (Sect. 1.1.2.2), the transition matrix from a linear polarization basis (\hat{h}, \hat{v}) to a circular polarization basis (\hat{r}, \hat{l}) is defined by $\tau = -\pi/4$ (to simplify the calculations we will chose $\Psi = \alpha = 0$) and a multiplicative coefficient A such that its determinant is equal to 1, leading to:

$$\left[\bar{\bar{P}}(\Psi = 0, \tau = -\pi/4, \alpha = 0) \right] = \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & -j \\ -j & 1 \end{pmatrix} \quad (5.21)$$

And consequently, applying Eq. (5.20):

$$\begin{pmatrix} S_{rr} & S_{rl} \\ S_{rl} & S_{ll} \end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix} S_{hh} - S_{vv} - 2jS_{hv} & -j(S_{hh} + S_{vv}) \\ -j(S_{hh} + S_{vv}) & S_{vv} - S_{hh} - 2jS_{hv} \end{pmatrix} \quad (5.22)$$

An example of circular polarization synthesis is given in Sect. 5.8.1.

5.5.1 Polarimetric signatures

In addition to matrix representation, we can also represent polarimetric behavior ‘visually’. For any polarization basis, the polarimetric signature displays the variations in intensity collected by a virtual receiving antenna whose polarization is either identical (co-polarized response) or orthogonal (cross-polarization response) to that of the receiving antenna. The mean detected power on reception is proportional to:

$$P(\Psi, \tau) = V(\Psi, \tau) \cdot V(\Psi, \tau)^* \quad (5.23)$$

where $V(\Psi, \tau)$ (analogous to a detected voltage) equals $\hat{E}_0^{incT} \cdot \left[\bar{\bar{S}} \right] \cdot \hat{E}_0^{inc}$ for a co-polarized response and $\hat{E}_0^{inc\perp T} \cdot \left[\bar{\bar{S}} \right] \cdot \hat{E}_0^{inc}$ for a cross-polarized one ($\hat{E}_0^{inc\perp}$ represents the orthogonal polarization of \hat{E}_0^{inc} , that propagates in the same direction).

The expression $V(\Psi, \tau)$ is the result of the following process: first, \hat{E}_0^{inc} is transmitted, then the target backscatters a wave with a polarization $\left[\bar{\bar{S}} \right] \cdot \hat{E}_0^{inc}$. We then apply the coupling principle for two waves propagating in opposite directions (Sect. 1.1.2.2. and Sect. 1.1.3.1); finally we get (for example for $\hat{T}_0^{rec} = \hat{T}_0^{inc}$) the scalar quantity $\hat{E}_0^{incT} \cdot \left[\bar{\bar{S}} \right] \cdot \hat{E}_0^{inc}$. Fig. 5.5(a) and Fig. 5.5(b) show, respectively, the variations of $P(\Psi, \tau)$ in co- and cross-polarization for a dihedral corner reflector. We identify the maximization of the cross-polarized response for $\Psi = 45^\circ$ or $\Psi = 135^\circ$.

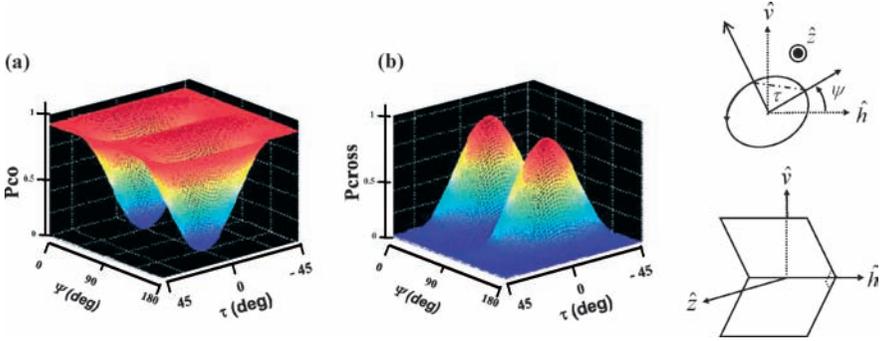


Fig. 5.5 (a) Polarimetric signature of co-polarized double-bounce scattering; (b) Polarimetric signature of cross-polarized double-bounce scattering.

5.6 Characteristic polarization and Euler parameters

For any matrix $[\bar{S}]$, there are special polarizations for which the co- or cross-polarized signals defined in the previous section reach minima or maxima [Huynen, 1970].

To identify these specific polarizations we must first look for so-called *pseudo-eigen* transmission polarizations, for which the polarization is not modified by the backscattering process in other words, for which the cross-polarization signature is null. The Jones vectors of these special polarizations satisfy:

$$\hat{E}_0^{inc\perp} \cdot [\bar{S}] \cdot \hat{E}_0^{inc} = 0 \tag{5.24}$$

Expressed differently, there is a complex scalar λ such that $[\bar{S}] \cdot \hat{E}_0^{inc} = \lambda \cdot \hat{E}_0^{inc*}$. The conjugation of the right-hand term expresses the fact that the two waves under consideration propagate in opposite directions (which explains why we speak of *pseudo-eigenvectors* rather than *eigenvectors*).

Equation (5.24) gives two solutions (noted *X-POL Null*), \hat{E}_0^{XNP1} and \hat{E}_0^{XNP2} , which are associated with the pseudo-eigenvalues v_1 and v_2 respectively. They form an orthogonal basis of the polarization space. When expressed in the basis $(\hat{E}_0^{XNP1}, \hat{E}_0^{XNP2})$, the backscattering matrix is diagonal:

$$[\bar{S}]_{(\hat{E}_0^{XNP1}, \hat{E}_0^{XNP2})} = \begin{bmatrix} v_1 & 0 \\ 0 & v_2 \end{bmatrix} = m \cdot \begin{bmatrix} \exp[j2v] & 0 \\ 0 & tg^2(\gamma) \cdot \exp[-j2v] \end{bmatrix} \tag{5.25}$$

where:

- m is an amplitude parameter (strength of the target).
- γ is the characteristic angle of the target. Its values are between 0 and $\pi/4$ (\hat{E}_0^{XNP1} is the eigenvector associated with the eigenvalue of greatest amplitude): $\gamma = 0$ for an infinite straight wire and $\gamma = \pi/4$ for single-bounce and double-bounce scattering.
- v defines the degree of double-bounce. It has a maximum value of $\pi/4$ for pure double bounce scattering.

It can be shown that the polarization \hat{E}_0^{XNP1} (associated with the characteristic parameters Ψ_{XNP1} and τ_{XNP1} , Sect. 1.1.2.2) leads to a maximum co-polarized backscatter signal. It is therefore an *X-POL Null* but also a '*CO-POL Max*'. The set of parameters $[\gamma, \nu, \tau_{XNP1}, \Psi_{XNP1}, m]$ form the Euler parameters.

5.6.1 The Huynen fork

Apart from *X-POL Null*, other specific polarizations can be defined for the target:

- The *CO-POL Max* and *CO-POL Null* polarizations which lead to maximum (respectively null) co-polarized backscattering;
- *X-POL Max* polarizations which lead to maximum cross-polarized backscattering.

J.R. Huynen provides a graphic method for visualizing these specific polarizations, by plotting the points representing the *X-POL Null* and *CO-POL Null* pairs on the Poincaré sphere (Sect. 1.1.3.3). We can then draw the four lines through the center of the sphere from each of these points obtaining a shape resembling a fork. Its aperture and its orientation on the Poincaré sphere are tied to the Euler parameters. Neither the Huynen fork nor Euler parameters have been used very much by the polarimetric community for the past two decades. However, this may change in the future with the advent of high resolution polarimetric systems.

5.7 Coherent decomposition of the polarimetric measurement

Until now the matrix $[\bar{\bar{S}}]$ has been considered as an 'indivisible' entity, revealing the entire backscattering process. The decomposition algorithms consider the backscattering process as resulting from the coherent summation of elementary mechanisms:

$$[\bar{\bar{S}}] = \sum_{i=1}^k \alpha_i \cdot [\bar{\bar{S}}]_i \quad (5.26)$$

Each of these mechanisms is characterized by a backscattering matrix $[\bar{\bar{S}}]_i$, weighted by a complex coefficient α_i . This decomposition is not unique, which means that there are in fact a series of algorithms, divided into two categories: those that perform decomposition into standard target contributions (planes, dihedral corner reflectors, helices etc.), and those which extract components with particular algebraic properties (reciprocity, symmetry etc.).

5.7.1 Decomposition into standard mechanisms - the group of Pauli matrices

We have already encountered the group of Pauli matrices ($[\bar{\bar{\sigma}}_0], [\bar{\bar{\sigma}}_1], [\bar{\bar{\sigma}}_2], [\bar{\bar{\sigma}}_3]$) (Sect. 1.1.3.3). We shall see here how they are used in the Pauli decomposition of the

matrix $[\bar{S}]$:

$$[\bar{S}] = \sum_{i=0}^3 \alpha_i \cdot [\bar{\sigma}_i] = \frac{\alpha_0}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\alpha_1}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ + \frac{\alpha_2}{\sqrt{2}} \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \frac{\alpha_3}{\sqrt{2}} \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad (5.27)$$

If we assume a monostatic configuration with reciprocal interaction mechanisms, then $\alpha_3 = 0$, as the component of $[\bar{S}]$ along $[\bar{\sigma}_3]$ is the only non reciprocal one. The other parameters are:

$$\alpha_0 = \frac{S_{hh} + S_{vv}}{\sqrt{2}} \quad \alpha_1 = \frac{S_{hh} - S_{vv}}{\sqrt{2}} \quad \alpha_2 = \sqrt{2} \cdot S_{hv} \quad (5.28)$$

The $[\bar{\sigma}_0]$ matrix characterizes odd-bounce scattering (Sect. 5.4.1) from smooth surface, plane, sphere or trihedral reflectors. The $[\bar{\sigma}_1]$ matrix involves double-bounce scattering (Sect. 5.4.2) from a dihedral corner reflector whose vertex is either parallel or perpendicular to the incident wave ($\delta = 0$ or $\delta = \pi/2$ in Eq. (5.14)). $[\bar{\sigma}_2]$ also refers to double-bounce scattering, but from a dihedral corner reflector with a 45° vertex orientation ($\delta = 45^\circ$).

The physical interpretation of Pauli decomposition is as follows: the backscatter process is the result of the coexistence of 3 coherent elementary mechanisms: single scattering; double-bounce scattering with no polarization change; and double-bounce scattering ‘at 45° ’, which backscatters the entire signal on the cross-polarized channel. This latter component will be dominant for any interaction process that creates significant cross-polarization, such as volume scattering (Sect. 5.10.2).

Pauli decomposition has been applied to the *L* band SIR-C image of the Flevoland area, the Netherlands. Fig. 5.6(a) displays a color composite of the radiometric channels⁶ *HH* (red), *HV* (green) and *VV* (blue), primarily filtered by a Lee polarimetric speckle filter [Lee, 1999]. Fig. 5.6(b) displays the corresponding Pauli decomposition (averaged over neighbor pixels to reduce speckle effects). The dominance of blue and red tones expresses the overwhelming presence of single and double scattering (horizontally or vertically oriented). In the latter case, we can interpret double-bounce effects as interactions between the ground and the vertical stalks of certain plants. The local greenish tones suggest volume effects.

5.7.1.1 Krogager’s approach

The algorithm of Krogager [Krogager, 1990] is based on an approach similar to the Pauli approach (decomposition into elementary mechanisms) but here the decomposition is into single scattering, oriented double scattering (for any δ) and ‘helix’ scattering.

⁶For sake of clarity, the $(|S_{pq}|^2)^{1/2}$ quantities, where $\langle \cdot \rangle$ is the average operator over neighbor pixels, and $p, q = h$ or v are hereafter denoted “*PQ*” (e.g. $\langle |S_{hh}|^2 \rangle^{1/2} = HH$).

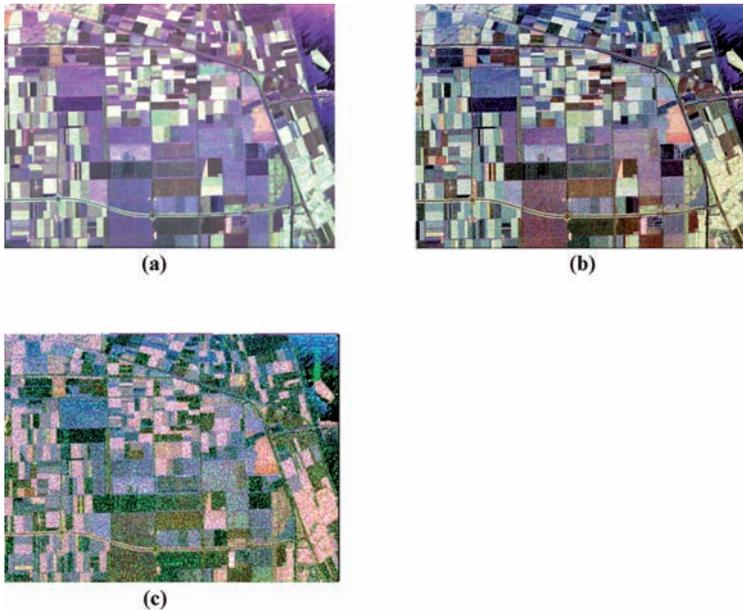


Fig. 5.6 (a) Color composite polarimetric image by JPL/AIRSAR, L band, Flevoland, Netherlands: HH (R), HV (G), VV (B), image after Lee polarimetric speckle filtering (courtesy J.S. Lee, US Naval Research Laboratory); (b) Pauli decomposition: single-bounce scattering (B), double-bounce scattering 0° (R), double-bounce scattering 45° (G), averaged over a 3×3 window (POLARSAR-PRO result); (c) Krogager decomposition, using POLARSAR-PRO: $|S_{ri}|$ (B), $\max(|S_{ll}|, |S_{rr}|)$ (R), $(|S_{ll} - S_{rr}|)$ (G), averaged on a 3×3 window.

This latter type of scattering creates non-symmetrical behavior according to whether the illumination is left- or right-handed circular:

$$\left[\bar{S} \right]_{helix} = \begin{bmatrix} 1 & \pm j \\ \pm j & -1 \end{bmatrix} \quad (5.29)$$

The \pm sign expresses the handedness of the helix. We shall not delve any further into this method. Note nevertheless that this decomposition is usually expressed in a circular polarization basis, which simplifies the formalism. Fig. 5.6(c) shows Krogager's decomposition of the Flevoland image, displaying the signature $|S_{ri}|$ (see Eq. (5.22)) on the blue channel, the maximum of $|S_{ll}|$ and of $|S_{rr}|$ on the red channel and $|S_{ll} - S_{rr}|$ on the green channel. As expected, surface scatter (in blue) strongly modifies the polarization of the incident waves when expressed in a circular polarization basis (Sect. 5.4.1).

5.7.2 Algebraic decomposition: the Cameron approach

We shall now look at a radically different approach. The decomposition is first considered geometrically and then expressed in algebraic terms.

The Cameron approach [Cameron, 1996] ‘tracks down’ behavior suggesting particular properties of reciprocity and symmetry in the backscattering process. It is a subtle approach and its mathematical meaning may be difficult to grasp.

As in the previous cases, we eliminate the non-reciprocity hypothesis (we assume $S_{vh} = S_{hv}$, Sect. 5.3.1). The projection of $[\bar{S}]$ onto the group of Pauli matrices therefore leads to $\alpha_3 = 0$ as shown in Eq. (5.27); this can be rewritten as

$$[\bar{S}] = \frac{\alpha_0}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & -\alpha_1 \end{bmatrix} \quad (5.30)$$

The parameters $\alpha_0, \alpha_1, \alpha_2$ are calculated from Eq. (5.28).

We can now focus on symmetry properties. What exactly do we mean by this? A backscattering process is called ‘symmetrical’ when the target that creates it has an axis of symmetry contained in the plane perpendicular to the radar-target line of sight [Nghiem, 1992]. If this is the case, let us imagine that the target is rotated in this plane, so that its axis of symmetry is brought into the incident plane (e.g. a dihedral corner reflector ‘at 45°’ rotated so that it is located ‘at 0°’ or ‘at 90°’). In its new position, the target can no longer return a cross-polarized component. (Why should it? This would only ruin the symmetry of the situation!) The backscattering matrix of the target in its new position is therefore diagonal.

However, when the target has no particular symmetry, there is no reason to suspect that a certain rotation could completely eliminate the cross-polarized term in $[\bar{S}]$. Rather than wiping it out completely we can try to minimize it, creating a sort of ‘pseudo-diagonalization’ for $[\bar{S}]$. Since the component of $[\bar{S}]$ along $[\bar{\sigma}_0]$ is unaffected by rotation, there is no point in maintaining it in the calculations. Once it has been removed, the off diagonal term obtained after applying Eq. (5.20) to the remaining matrix components for $\tau = 0$ and $\alpha = 0$ and for any angle Ψ is given by:

$$\begin{aligned} S_{hh}(\Psi) &= 1/\sqrt{2} \cdot (\alpha_1 \cdot \cos 2\Psi + \alpha_2 \cdot \sin 2\Psi) \\ S_{vv}(\Psi) &= 1/\sqrt{2} \cdot (-\alpha_1 \cdot \cos 2\Psi - \alpha_2 \cdot \sin 2\Psi) \\ S_{hv}(\Psi) &= 1\sqrt{2} \cdot (-\alpha_1 \cdot \sin 2\Psi + \alpha_2 \cdot \cos 2\Psi) \end{aligned} \quad (5.31)$$

The value of Ψ which minimizes the modulus of $S_{hv}(\Psi)$ satisfies:

$$\frac{d}{d\Psi} |S_{hv}(\Psi)|^2 = 0 \quad (5.32)$$

A straightforward calculation leads to:

$$\tan(4\Psi) = \frac{\alpha_1^* \cdot \alpha_2 + \alpha_2^* \cdot \alpha_1}{|\alpha_1|^2 - |\alpha_2|^2} \quad (5.33)$$

giving the angle Ψ that limits the production of cross-polarized energy. Using $[\bar{S}] = \alpha_0 \cdot [\bar{\sigma}_0] + \alpha_1 \cdot [\bar{\sigma}_1] + \alpha_2 \cdot [\bar{\sigma}_2]$, and applying Eq. (5.31), we obtain the full expression

of the matrix $[\bar{S}]$ in a basis (\hat{a}, \hat{b}) which has been rotated by the angle Ψ with respect to the reference basis (\hat{h}, \hat{v}) . Thus:

$$[\bar{S}] = \alpha_0 \cdot [\bar{\sigma}_0] + [\alpha_1 \cdot \cos(2\Psi) + \alpha_2 \cdot \sin(2\Psi)] \cdot [\bar{\sigma}_1] - [\alpha_1 \cdot \sin(2\Psi) - \alpha_2 \cdot \cos(2\Psi)] \cdot [\bar{\sigma}_2] \quad (5.34)$$

The first two terms of $[\bar{S}]$ indicate symmetrical behavior. Let us note:

$$\begin{aligned} [\bar{S}]_{sym}^{\max} &= \alpha_0 \cdot [\bar{\sigma}_0] + [\alpha_1 \cdot \cos(2\Psi) + \alpha_2 \cdot \sin(2\Psi)] \cdot [\bar{\sigma}_1] \\ &= \alpha_0 \cdot [\bar{\sigma}_0] + \varepsilon \cdot [\bar{\sigma}_1] \end{aligned} \quad (5.35)$$

which is by construction a diagonal matrix. The term in $[\bar{\sigma}_2]$ expresses the ‘stretching of $[\bar{S}]$ ’ with respect to completely symmetrical behavior. Cameron quantifies this ‘stretching’ using an angle parameter τ , once the backscatter matrices have been vectorized (Sect. 5.3.2).

5.7.2.1 Projection onto standard symmetry mechanisms

We shall now build a classification algorithm from the $[\bar{S}]_{sym}^{\max}$ component of $[\bar{S}]$. Since it is diagonal, we rewrite it as follows:

$$[\bar{S}]_{sym}^{\max} = a \cdot \exp(j\varphi) \cdot \begin{bmatrix} 1 & 0 \\ 0 & z \end{bmatrix} \quad (5.36)$$

To within the amplitude term a , an equivalent representation can be given by a unit vector:

$$\hat{\Lambda}(z) = \frac{1}{\sqrt{1+|z|^2}} \cdot \begin{bmatrix} 1 \\ z \end{bmatrix} \quad (5.37)$$

Finally, Cameron introduces a distance value based on the scalar product between $\hat{\Lambda}$ and a series of reference vectors $\hat{\Lambda}(z^{ref}) = (1, z^{ref})^T$ associated with standard symmetry mechanisms:

$$d(z, z^{ref}) = \frac{1}{\sqrt{1+|z|^2}} \cdot \frac{1}{\sqrt{1+|z^{ref}|^2}} \cdot |1 + z^* \cdot z^{ref}| \quad (5.38)$$

The following standard behaviors are considered: single scattering (planes, spheres, trihedral corner reflectors) with $z^{ref} = 1$; dihedral corner reflectors with $z^{ref} = -1$; dipoles with $z^{ref} = 0$; cylinders with $z^{ref} = 1/2$; narrow dihedral corner reflectors with $z^{ref} = -1/2$; and quarter wave dipole with $z^{ref} = j$. The nature of the target being studied is identified by determining the shortest distance, determined by maximizing the scalar product $d(z, z^{ref})$.

Cameron decomposition may provide rather unsatisfactory results, giving an impression of disorder, when applied to an image ‘pixel by pixel’. The algorithm probably

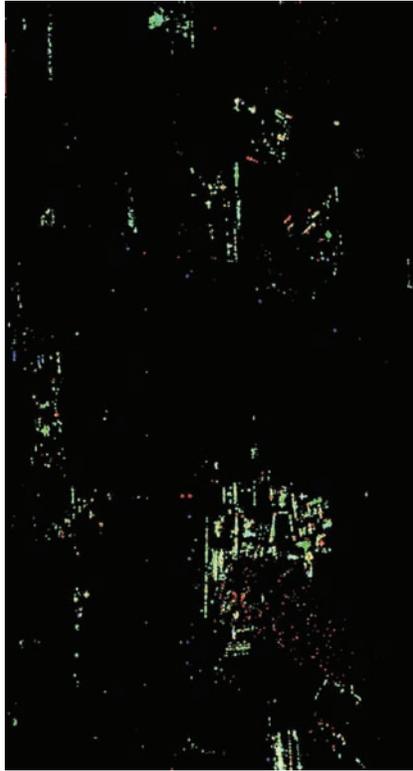


Fig. 5.7 Cameron decomposition using a pre-filtering of point targets: single-bounce scattering (blue), double-bounce scattering (red), diffractions (green), and quarter wavelength dipole (cyan).

fails as a result of excessively fine ‘granularity’, due to insufficient resolution of the available images. The fact that depolarization effects are not taken into account is also prejudicial (even though this is also true for the other coherent approaches). To overcome this failing, Touzi suggests the *Symmetric Scattering Characterization Method* (SSCM), a sort of ‘adjustable’ Cameron decomposition which takes full account of the polarization state of the backscattered wave [Touzi, 2004]. In Fig. 5.7 we illustrate an approach similar to that of Touzi, in which a mask for detecting coherent pixels is applied before Cameron analysis (the criterion used here is the invariability of the backscattered signal during illumination, [Souyris, 2003]). The example given here uses *L* band and 5 m resolution in an airport environment (Sect. 5.11). The result shows the dominance of diffraction effects (green) on the corners of buildings, and also the local effects of single and double scattering.

Until now, coherent decomposition methods have been little used by the scientific community, principally because of the lack of high resolution polarimetric data. Improvements in sensor design technology will undoubtedly increase the efficiency of this type of approach. Until recently, data processing has concentrated

principally on incoherent decompositions and their aptitude for taking depolarization into account.

5.8 Taking depolarization into account

The interaction process described by the backscattering matrix $[\bar{S}]$ has no way of revealing the possible creation of depolarization during interactions between the incident wave and the environment (Sect. 1.1.3). However, since it is likely that some scatterers will fluctuate in time or space, we may assume that a change will occur in the polarization state. These fluctuations also produce speckle (Sect. 3.11), which must be reduced.

The most common speckle reduction technique consists in ‘averaging’ the information of nearby pixels, resulting in a multi-look polarimetric image. For this, we calculate an ‘incoherent’ average of the target vectors $\vec{\kappa}_L$ or $\vec{\kappa}_P$ (Sect. 5.3.2). This incoherent average is the source of the depolarization mentioned above. The complete wave/surface interaction (including this average) cannot be described without the formalism of the Mueller matrix and the Stokes vector.

5.8.1 Stokes formalism and Mueller matrix

The Stokes vector provides a complete description of the polarization state of a partially polarized wave (Sect. 1.1.3.3.). When this description replaces the Jones vector, the backscatter process is expressed using the Mueller matrix $[\bar{M}]$, which describes the transition from the Stokes vector of the fully polarized incident wave \hat{S}^{inc} to that of the partially polarized backscattered wave \tilde{S}^s :⁷

$$\tilde{S}^s = [\bar{M}] \cdot \hat{S}^{inc} \quad (5.39)$$

$[\bar{M}]$ is a 4×4 real symmetric matrix. It is defined *a priori* by 10 parameters which have the same units as energy terms, but they are linked by a ‘trace condition’:

$$M_{11} = M_{22} + M_{33} + M_{44} \quad (5.40)$$

thus reducing the number of independent parameters from 10 to 9.

Hence in place of the 5 parameters characteristic of polarimetric backscattering from a pure target, we now have 9 parameters to account for possible depolarization during the backscatter. In the special case of a non-depolarizing target (called a pure target), these nine parameters are linked by 4 relationships (called target equations, Sect. 5.10.1), from which we can reconstruct an equivalent $[\bar{S}]$ matrix.

⁷The little hat over \hat{S}^{inc} indicates that the Stokes vector of the incident wave has a norm of 1. Moreover, as it is a fully polarized wave: $S_0^2 = S_1^2 + S_2^2 + S_3^2 = 1$, where S_0, S_1, S_2, S_3 are the components of \hat{S}^{inc} .

5.8.1.1 Transition from the scattering matrix to the Muller matrix

Before any time or space averaging is performed, each image sample remains tied to its own backscattering matrix $[\bar{S}]$. If we generalize the result obtained from Eq. (5.50) in Sect. 1.1.4, the equivalent $[\bar{M}]$ matrix associated with this $[\bar{S}]$ matrix (without any depolarizing effect) is given by :

$$[\bar{M}]_{(p,q)} = Tr \left\{ [\bar{S}]^{*T} \cdot [\bar{\sigma}_p] \cdot [\bar{S}] \cdot [\bar{\sigma}_q] \right\} \quad (5.41)$$

The target equations mentioned in the previous paragraph (and detailed in Sect. 5.10.1) come from this equation. We can also calculate the $[\bar{M}]$ matrix, which reveals the average polarimetric behavior of a group of nearby pixels, by averaging the equivalent $[\bar{M}]$ matrices. Once the averaging has been performed, generally we can no longer return to the formalism of the backscattering matrix to describe the behavior of this group of pixels.

5.8.1.2 Parametric form of Huynen

Huynen's parametric form of $[\bar{M}]$ is written as [Huynen, 1970]:

$$[\bar{M}] = \begin{bmatrix} A_0 + B_0 & C_\Psi & H_\Psi & F_\Psi \\ C_\Psi & A_0 + B_\Psi & E_\Psi & G_\Psi \\ H_\Psi & E_\Psi & A_0 - B_\Psi & D_\Psi \\ F_\Psi & G_\Psi & D_\Psi & -A_0 + B_0 \end{bmatrix} \quad (5.42)$$

The terms with index Ψ depend on the target's orientation about the radar-target line of sight (Fig. 5.8). The other terms (with index 0) are independent of this orientation, which leads to some interesting possibilities for analysis. For example, as a boat rides the waves, its orientation fluctuates about the radar-target line of sight. This fluctuation has no effect on A_0 and B_0 . For this reason, they are called polarimetric roll-invariants. The concept of polarimetric invariance will be seen again in incoherent algebraic decompositions.

The Mueller matrix is considered to be 'oriented' with respect to an arbitrary reference orientation $\Psi = 0$, from which nine Huynen parameters are extracted, namely A_0 , B_0 , B , C , D , E , F , G , and H :

- A_0 is the symmetry generator. For single-bounce scattering, only A_0 is non-zero.
- $B_0 - B$ is the *non-symmetry* generator (correlated with the signature in cross-polarization).
- $B_0 + B$ is an indicator of *irregularity*: $B_0 + B = 0$ for single-bounce scattering.
- C is a *shape factor* for targets that are predominantly symmetrical. Its value is maximum for dipole-like targets.
- D is a local *shape indicator*, depending on the difference in the radii of curvature for convex surfaces. It is zero for single scattering.

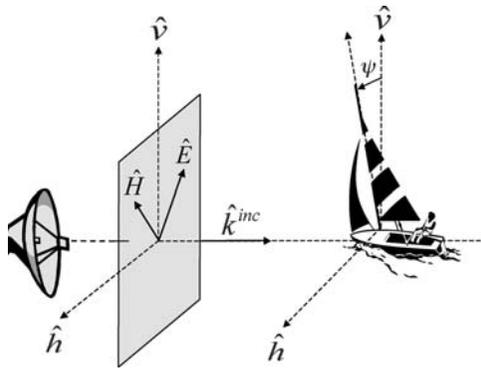


Fig. 5.8 Rotating the target around the radar-target line of sight.

- E is a *torsion* parameter.
- F is a *helicity* parameter.
- G is an indicator of coupling between contributions from symmetrical and non-symmetrical parts of the target.

5.8.1.3 Polarization synthesis in the Stokes formalism

The polarization of the fully polarized virtual incident wave is expressed using the Stokes formalism (Sect. 1.1.3.4):

$$\hat{S}^{incT} = [1, \cos 2\tau \cdot \cos 2\Psi, \cos 2\tau \cdot \sin 2\Psi, \sin 2\tau] \quad (5.43)$$

For a receive polarization \hat{S}^{rec} , the captured power density is proportional to:

$$P^{rec}(\Psi, \tau) = \hat{S}^{recT} \cdot [\bar{\bar{M}}] \cdot \hat{S}^{inc} \quad (5.44)$$

There are morphological similarities between the co- and cross-polarized polarimetric signatures of a built-up area (Fig. 5.9(a) and (b)), and those displayed in Fig. 5.5(a) and (b). These similarities are explained by the preponderance of double-bounce effects in the backscatter from the built-up area. However, a pedestal is present in Fig. 5.9(a) and Fig. 5.9(b), revealing the contribution of unpolarized energy (which is independent of the receive polarization). The amplitude of this pedestal is correlated with the degree of polarization of the wave (Sect. 1.1.3.2).

5.8.1.4 Polarimetric contrast enhancement

The search for transmit and receive polarizations giving maximum polarimetric contrast between two targets with respecting Mueller matrixes $[\bar{\bar{M}}_1]$ and $[\bar{\bar{M}}_2]$ can be broken down into two steps [Kostinski, 1988]: (1) The incident polarization \hat{S}^{inc} is selected in

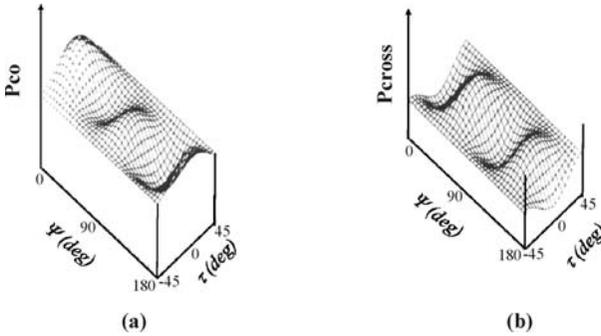


Fig. 5.9 Polarimetric signatures of co-polarization (a) and cross-polarization (b) of an urban area (extracted from [Zebker, 1987]).

order to give the largest ratio of free-space backscatter power densities coming from each target; and (2) the receive polarization \hat{S}^{rec} is then adjusted so as to be orthogonal to the polarized part of the free-space power density backscattered by one of these targets. An alternative method is found in [Van Zyl, 1987], leading in practice to analogous results.

For example, the projection of polarimetric measurements on to optimized polarizations can reinforce the radar signal sensitivity to the bio-geophysical parameters under study. For example, circular polarizations increase radar-signal sensitivity to surface roughness (e.g. for bare ground, lava fields, etc.) [Zebker, 1987], [Mattia, 1997], while orthogonal linear polarizations (horizontal transmission, vertical reception) are the most sensitive to forest biomass, as shown in Fig. 5.10(a) and Fig. 5.10(b). The areas outlined in full line and dash line are respectively bare ground with variable roughness and patches of forest at different stages of development. The polarimetric measurement projected onto circular polarization (right hand circular transmission, right hand circular reception) (Fig. 5.10(a)) reveals a very marked texture over bare ground. On the other hand, projection onto linear polarization (horizontal transmission, vertical reception) (Fig. 5.10(b)), clearly separates the wooded area into two distinct stands.

Please note: Polarimetric contrast enhancement can also be applied to single-look images (using the backscattering matrix formalism), but it is not yet very common, due to the disturbing effect of speckle.

5.9 Covariance matrix - coherence matrix

Although in the 1970s the Mueller formalism was central to the work of Huynen, its use is now rare. Today, expressions more closely related to radar measurements are preferred.

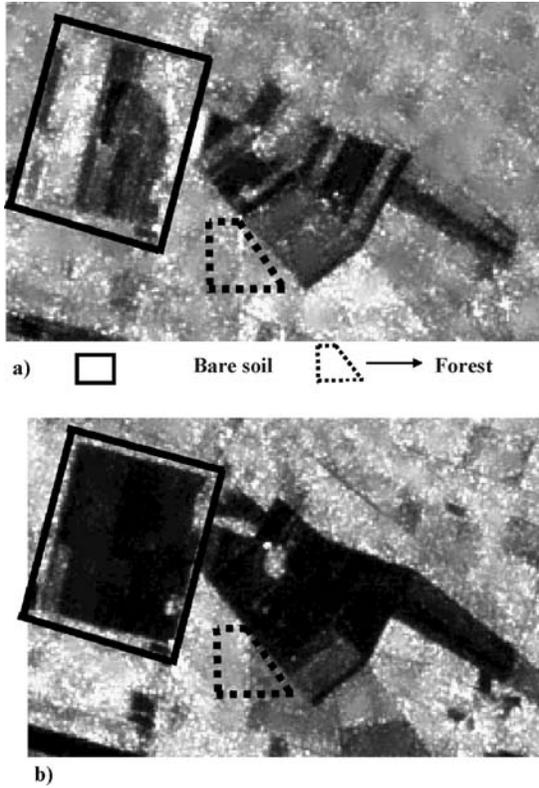


Fig. 5.10 Effect of polarization synthesis on wooded zones and bare soil. SIR-C image, Landes, France, L band, 24° : (a) projection of polarimetric data onto a circular polarization basis (transmission: right-hand circular; reception: right-hand circular); (b) projection of polarimetric data onto a linear polarization basis (transmission: horizontal reception: vertical).

5.9.1 Covariance matrix

The covariance matrix is constructed from the average of \vec{k}_L over nearby pixels (Sect. 5.3.2):

$$[\bar{C}] = \langle \vec{k}_L \cdot \vec{k}_L^{*T} \rangle = \begin{bmatrix} \langle S_{hh} \cdot S_{hh}^* \rangle & \sqrt{2} \cdot \langle S_{hh} \cdot S_{hv}^* \rangle & \langle S_{hh} \cdot S_{vv}^* \rangle \\ \sqrt{2} \cdot \langle S_{hv} \cdot S_{hh}^* \rangle & 2 \cdot \langle S_{hv} \cdot S_{hv}^* \rangle & \sqrt{2} \langle S_{hv} \cdot S_{vv}^* \rangle \\ \langle S_{vv} \cdot S_{hh}^* \rangle & \sqrt{2} \cdot \langle S_{vv} \cdot S_{hv}^* \rangle & \langle S_{vv} \cdot S_{vv}^* \rangle \end{bmatrix} \quad (5.45)$$

$$= \sigma_{hh} \cdot \begin{bmatrix} 1 & \sqrt{2} \cdot \eta \cdot \sqrt{e} & \rho \cdot \sqrt{g} \\ \sqrt{2} \cdot \eta^* \cdot \sqrt{e} & 2 \cdot e & \sqrt{2} \cdot \xi \cdot \sqrt{e \cdot g} \\ \rho^* \cdot \sqrt{g} & \sqrt{2} \cdot \xi^* \cdot \sqrt{e \cdot g} & g \end{bmatrix} \quad (5.46)$$

where $\langle \cdot \rangle$ is the averaging operator over neighbor pixels. The parametric form Eq. (5.46) includes 9 parameters: an energy parameter ($\sigma_{hh} = \langle S_{hh} \cdot S_{hh}^* \rangle$), two real parameters (e and g) and three complex numbers (thus six real parameters) ρ , η and ξ :

$$\begin{aligned}
 g &= \frac{\langle S_{vv} \cdot S_{vv}^* \rangle}{\langle S_{hh} \cdot S_{hh}^* \rangle} & e &= \frac{\langle S_{hv} \cdot S_{hv}^* \rangle}{\langle S_{hh} \cdot S_{hh}^* \rangle} \\
 \rho &= \frac{\langle S_{hh} \cdot S_{vv}^* \rangle}{\sqrt{\langle S_{hh} \cdot S_{hh}^* \rangle \cdot \langle S_{vv} \cdot S_{vv}^* \rangle}} & \eta &= \frac{\langle S_{hh} \cdot S_{hv}^* \rangle}{\sqrt{\langle S_{hh} \cdot S_{hh}^* \rangle \cdot \langle S_{hv} \cdot S_{hv}^* \rangle}} \\
 \xi &= \frac{\langle S_{vv} \cdot S_{hv}^* \rangle}{\sqrt{\langle S_{vv} \cdot S_{vv}^* \rangle \cdot \langle S_{hv} \cdot S_{hv}^* \rangle}}
 \end{aligned} \tag{5.47}$$

where e and g characterize the ratios of co-polarized energies. The entities ρ , η , and ξ are complex degrees of coherence, with modulus between 0 and 1 due to the Schwartz inequality. They quantify the level of similarity of responses backscattered by a target under variable transmission and reception conditions (horizontal or vertical).

From a theoretical point of view, the references [Borgeaud, 1987] and [Nghiem, 1992] have shown that the degree of coherence between co and cross-polarization terms (i.e., parameters η and ξ) are zero for backscattering from layers made up of randomly oriented ovoid particles. These layers realistically model most natural environments⁸ in interaction with an impinging electromagnetic wave. The \bar{C} matrix can then be expressed as:

$$\bar{C} = \sigma_{hh} \cdot \begin{bmatrix} 1 & 0 & \rho \cdot \sqrt{g} \\ 0 & 2 \cdot e & 0 \\ \rho^* \cdot \sqrt{g} & 0 & g \end{bmatrix} \tag{5.48}$$

The only remaining correlation term is ρ , which expresses the degree of coherence between the co-polarized signals S_{hh} and S_{vv} .

5.9.1.1 Behavior of the degree of coherence between co-polarized signals

Plotting the modulus of the degree of coherence between S_{hh} and S_{vv} ($|\rho|$) (Fig. 5.11(a)) reveals a sharp contrast between wooded areas and other surfaces ($|\rho|$ can be used for distinguishing between forest and non-forest cover). Radiometric data (Fig. 5.11(b)) (such as the C band backscattering coefficient σ_{vv}^0) are subject to strong variations due to varying ground moisture and roughness (see the area at top right of Fig. 5.11(b)). In areas with no ground cover, $|\rho|$ is close to 1. The wooded area in the center of the image also shows large radiometric variations, and a degree of coherence stabilized around $|\rho| \approx 0.35$. Also, note the high level of $|\rho|$ returned from water (bottom center of image).

⁸Most polarimetric calibration procedures rely on the hypothesis that η and ξ are zero (known as the reflection symmetry hypothesis) for specific areas of the image (for example dense vegetation) [Zyl, 1990].

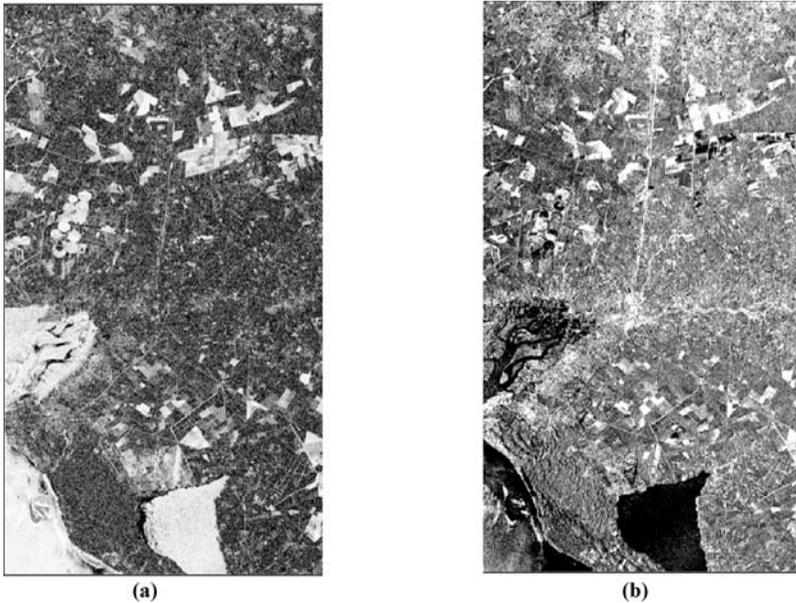


Fig. 5.11 (a) Coherence between S_{hh} and S_{vv} ($|\rho|$), C band, 26.4° , Landes forest, France. Estimated coherence on a sliding 3×3 window from an 8-look image (4 range looks $\times 2$ azimuth looks); (b) simulation of an ERS image (C band, polarization VV , 23°). SIR-C data, C band, 26.4° . Backscattering coefficient estimated using sliding 3×3 window on an 8-look image (4 range looks $\times 2$ azimuth looks).

The degree of coherence mainly depends on the nature of the dominant backscattering mechanisms (surface backscattering or volume backscattering) which is not very sensitive to changes in roughness and moisture. In addition, the fact that it is a normalized parameter makes it robust with respect to possible drift due to calibration errors and slope effects.

5.9.2 Coherence matrix

The Coherence matrix is constructed from the spatial average of \vec{k}_P :

$$[\bar{\bar{T}}] = \langle \vec{k}_P \cdot \vec{k}_P^{*T} \rangle = \frac{1}{2} \begin{bmatrix} \langle (S_{hh} + S_{vv}) \cdot (S_{hh} + S_{vv})^* \rangle & \langle (S_{hh} + S_{vv}) \cdot (S_{hh} - S_{vv})^* \rangle & 2\langle (S_{hh} + S_{vv}) \cdot S_{hv}^* \rangle \\ \langle (S_{hh} - S_{vv}) \cdot (S_{hh} + S_{vv})^* \rangle & \langle (S_{hh} - S_{vv}) \cdot (S_{hh} - S_{vv})^* \rangle & 2\langle (S_{hh} - S_{vv}) \cdot S_{hv}^* \rangle \\ 2\langle S_{hv} \cdot (S_{hh} + S_{vv})^* \rangle & 2\langle S_{hv} \cdot (S_{hh} - S_{vv})^* \rangle & 4\langle S_{hv} \cdot S_{hv}^* \rangle \end{bmatrix} \quad (5.49)$$

The covariance matrix $[\bar{\bar{C}}]$ and the coherence matrix $[\bar{\bar{T}}]$ are both semi-definite positive Hermitian matrices (i.e., they are equal to their transpose conjugate), thus their eigenvalues are positive and their eigenvectors are orthogonal. $[\bar{\bar{C}}]$ and $[\bar{\bar{T}}]$ have identical eigenvalues. However, the structure of the $[\bar{\bar{T}}]$ eigenvectors is more appropriate

for polarimetric analysis than that of the $[\bar{C}]$ eigenvectors. From now on, we will use the $[\bar{T}]$ formalism unless otherwise specified.⁹

5.10 Incoherent decomposition of polarimetric measurements

Incoherent polarimetric measurements are decomposed for the same reasons as coherent ones (Sect. 5.7), i.e., to represent the interaction as the sum of elementary effects. However, unlike the coherent case, here the summing is incoherent and is conducted using matrices representing energy rather than electric fields:

$$[\bar{T}] = \sum_{i=1}^I t_i \cdot [\bar{T}]_i \quad (5.50)$$

The polarization state of the backscattered wave, which was not used in coherent decomposition algorithms, is now taken into account in incoherent decomposition algorithms. We now have three categories of algorithms to ‘play with’: those that separate the polarized from the unpolarized parts of the backscatter wave, those that decompose it into standard backscattering mechanisms (simple, double and volume scattering) and lastly algebraic approaches, based on eigenvalue and eigenvector analysis of the coherence matrix. A review of these algorithms can be found in [Cloude, 1996]¹⁰ and [Touzi, 2004].

5.10.1 Decomposition into polarization states: the Huynen approach

Any partially polarized wave can be decomposed into the incoherent sum of a fully polarized wave and a fully unpolarized wave (Sect. 1.1.3.2). This dichotomy [Huynen, 1970] is transposed onto the Mueller matrices (the same reasoning can be applied to the covariance and coherence matrices):

$$[\bar{M}] = [\bar{M}^{pol}] + [\bar{M}^{upol}] \quad (5.51)$$

The polarized part is characteristic of a dominant fully polarized mechanism created by a ‘pure’ target. The unpolarized part reveals the contribution of residual targets that Huynen calls *symmetrical*. On reception, the unpolarized scattering generates a signal whose mean power is insensitive to the receiving polarization (Sect. 1.1.3.1). In particular, its behavior does not depend on the rotation around the radar-target line of sight, thus justifying the term ‘symmetrical’.¹¹

⁹Warning ! The “coherence matrix” $[\bar{T}]$ is not the “wave coherence matrix” found in Sect. 1.3.3.1.

¹⁰[Cloude, 1996] credits J.R. Huynen as the pioneer who developed target decomposition theorems for radar, while stressing that his approach was inspired by the work of Chandrasekhar (Sect. 1.3.3) on the scattering of light off of anisotropic particles.

¹¹Cameron uses the word ‘symmetrical’ to mean something different (Sect. 5.7.2).

The polarized part $[\bar{M}^{pol}]$ is constrained by the four relationships characteristic of pure targets (which reduces the number of independent parameters to 5 and consequently allows us to go back to the backscattering matrix formalism if necessary):

$$\begin{cases} 2 \cdot A_0 \cdot (B_0 + B) = C^2 + D^2 \\ 2 \cdot A_0 \cdot E = C \cdot H - D \cdot G \\ 2 \cdot A_0 \cdot (B_0 - B) = G^2 + H^2 \\ 2 \cdot A_0 \cdot F = C \cdot G + D \cdot H \end{cases} \quad (5.52)$$

In algebraic terms, the rank¹² of $[\bar{M}^{pol}]$ is 1, while that of $[\bar{M}^{upol}]$ is 3.

5.10.2 Decomposition into standard mechanisms - the Freeman approach

The backscatter is considered to be generated by three incoherent elementary mechanisms [Freeman, 1998]: first-order Bragg scattering (i.e., single-bounce scattering), double-bounce scattering and volume scattering.

The backscattering matrix for first-order Bragg scattering (an extension of the odd-bounce scattering discussed in Sect. 5.4.1) is given by:

$$[\bar{S}]_s = \begin{bmatrix} S_{hh}^s & 0 \\ 0 & S_{vv}^s \end{bmatrix} \quad (5.53)$$

If we apply Eq. (5.45), the corresponding covariance matrix is written as follows:

$$[\bar{C}]_s = f_s \cdot \begin{bmatrix} \beta^2 & 0 & \beta \\ 0 & 0 & 0 \\ \beta & 0 & 1 \end{bmatrix} \quad (5.54)$$

with $f_s = \langle |S_{vv}^s|^2 \rangle$ and $\beta^2 = \langle |S_{hh}^s|^2 \rangle / f_s$. Here, the term $\langle S_{hh}^s \cdot S_{vv}^{s*} \rangle = \beta \cdot f_s$ is actually a real number. The fact that $[\bar{C}]$ has rank 1 means that the backscattering mechanism can be considered as created by a pure target.

The backscattering matrix of a double-bounce response produced by a dielectric dihedral corner reflector (an extension of the metal case described in Sect. 5.4.2.) is given by:

$$[\bar{S}]_d = \begin{bmatrix} S_{hh}^d & 0 \\ 0 & S_{vv}^d \end{bmatrix} = \begin{bmatrix} \exp(j \cdot 2\phi_{hh}) \cdot R_{ghh} \cdot R_{thh} & 0 \\ 0 & \exp(j \cdot 2\phi_{vv}) \cdot R_{gvv} \cdot R_{rvv} \end{bmatrix} \quad (5.55)$$

¹²The rank of a matrix $[\bar{A}]$ is the maximal number of linearly independent columns (or rows) $[\bar{A}]$. For a square matrix, the rank is also the number its non-null eigenvalues.

The g and t indices refer to echoes reflected by horizontal surfaces (ground) and vertical surfaces (like tree trunks). ϕ_{hh} and ϕ_{vv} are phase terms, the R reflection coefficients are real. The corresponding covariance matrix is given by:

$$[\bar{C}]_d = f_d \cdot \begin{bmatrix} |\alpha|^2 & 0 & \alpha \\ 0 & 0 & 0 \\ \alpha^* & 0 & 1 \end{bmatrix} \quad (5.56)$$

with $f_d = \langle R_{gvv}^2 \cdot R_{tvv}^2 \rangle$ and $\alpha = \langle (R_{ghh} \cdot R_{thh} / R_{gvv} \cdot R_{tvv}) \cdot \exp(2j(\phi_{hh} - \phi_{vv})) \rangle$. The rank 1 of $[\bar{C}]_d$ recalls the conclusions of the previous case. Finally, if we consider the backscattering matrix of a thin infinite dipole (Sect. 5.4.3), the covariance matrix for volume scattering, considered as a collection of isotropically directed dipoles is given by:

$$[\bar{C}]_v = f_v \cdot \begin{bmatrix} 1 & 0 & 1/3 \\ 0 & 2/3 & 0 \\ 1/3 & 0 & 1 \end{bmatrix} \quad (5.57)$$

with $f_v = \langle |S_{vv}^v|^2 \rangle = \langle |S_{hh}^v|^2 \rangle$. Unlike the previous cases, the covariance matrix now has rank 3, which prevents us from treating the backscattering process as similar to that of a pure target. It is also the only one of the three mechanisms that creates a cross-polarized signal.

Freeman decomposition considers the covariance matrix as the sum of the three independent mechanisms just described:

$$[\bar{C}] = [\bar{C}]_s + [\bar{C}]_d + [\bar{C}]_v \quad (5.58)$$

or, by substitution:

$$\begin{aligned} \langle |S_{hh}|^2 \rangle &= f_s \cdot \beta^2 + f_d \cdot |\alpha|^2 + f_v \\ \langle |S_{vv}|^2 \rangle &= f_s + f_d + f_v \\ \langle S_{hh} \cdot S_{vv}^* \rangle &= f_s \cdot \beta + f_d \cdot \alpha + f_v/3 \\ \langle |S_{hv}|^2 \rangle &= f_v/3 \end{aligned} \quad (5.59)$$

Unfortunately this leads us to an under-determined system, consisting of 4 equations with five unknowns. Looking at the sign of the real part of $\{\langle S_{hh} \cdot S_{vv}^* \rangle - \langle |S_{hv}|^2 \rangle\}$ resolves this ambiguity. When it is positive, we assume that surface backscattering predominates over double bounce and thus $\alpha = -1$. System (5.59) can therefore be solved with f_s , f_d , f_v and β . When the sign is negative, $\beta = 1$ and the system is solved with f_s , f_d , f_v and α . We can now determine the power contributions of the different mechanisms considered, by substituting into $[\bar{C}]_s$, $[\bar{C}]_d$ and $[\bar{C}]_v$:

$$\begin{aligned} P_s &= f_s \cdot (1 + \beta^2) \\ P_d &= f_d \cdot (1 + |\alpha|^2) \\ P_v &= 8 \cdot f_v/3 \end{aligned} \quad (5.60)$$



Fig. 5.12 Freeman decomposition, Flevoland, L band. P_s (surface) in blue; P_v (volume) in green; P_d (double-bounce scattering) in red (POLoSAR-PRO result).

Freeman decomposition applied to the Flevoland example (Fig. 5.12) shows clear segmentation of the regions according to the mechanisms given in Eq. (5.60). Plotting $|\rho|$ the coherence (Fig. 5.13(a)) reveals high values for surface scatter and low values for volume scatter. Areas colored in violet also present low values of $|\rho|$ indicating a mixture of surface and volume effects.

5.10.3 Algebraic decomposition - The (H, α) approach

Algebraic decomposition of the covariance and coherence matrices is based on the fact that these matrices are Hermitian. Thus, they can be diagonalized, their eigenvalues are positive and their eigenvectors are orthogonal to each other. Hence, when expressed in the basis of the eigenvectors, the backscattering process has a diagonal covariance or coherence matrix, expressing the coexistence of three independent backscattering mechanisms. In the next section, we will expand this decomposition better known as (H, α) ‘decomposition’, when using the coherence matrix [Cloude, 1997]. It has become the standard method for incoherent decomposition over the last few years.

5.10.3.1 Diagonalization of the coherence matrix

Since the coherence matrix $[\bar{T}]$ is Hermitian, it can be diagonalized via a unitary¹³ transition matrix $[\bar{U}]_3$:

¹³The definition of a unitary matrix given in Sect 1.1.2.2 in the 2D case, can be generalized to 3D unitary matrices.

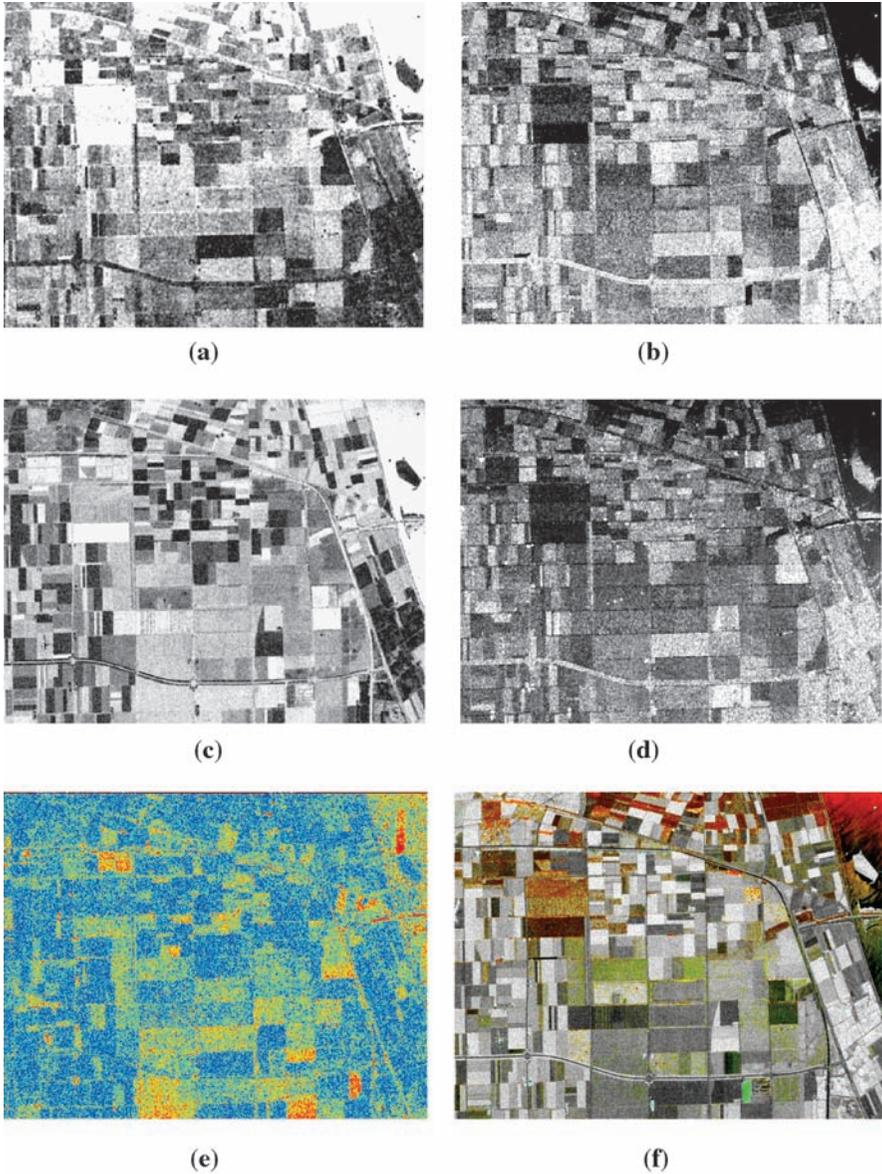


Fig. 5.13 (a) Coherence between S_{hh} and S_{vv} (ρ), Flevoland, L band; (b) entropy H ; (c) coherence between S_{rr} and S_{ll} (right-hand circular, left-hand circular); (d) average α coefficient; (e) anisotropy A , Flevoland, using POLSAR-PRO. Low values on the blue, high values on the red; (f) IHS decomposition, Flevoland.

$$[\bar{\bar{T}}] = [\bar{\bar{U}}_3] \cdot \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \cdot [\bar{\bar{U}}_3]^{*T} \quad (5.61)$$

The orthonormal eigenvectors $[\hat{v}_1 \quad \hat{v}_2 \quad \hat{v}_3]$ making up the transition matrix $[\bar{\bar{U}}_3]$ are written as:

$$\begin{aligned} [\bar{\bar{U}}_3] &= [\hat{v}_1 \quad \hat{v}_2 \quad \hat{v}_3] \\ &= \begin{bmatrix} \cos \alpha_1 & \cos \alpha_2 & \cos \alpha_3 \\ \sin \alpha_1 \cdot \cos \beta_1 \cdot e^{j \cdot \delta_1} & \sin \alpha_2 \cdot \cos \beta_2 \cdot e^{j \cdot \delta_2} & \sin \alpha_3 \cdot \cos \beta_3 \cdot e^{j \cdot \delta_3} \\ \sin \alpha_1 \cdot \sin \beta_1 \cdot e^{j \cdot \gamma_1} & \sin \alpha_2 \cdot \sin \beta_2 \cdot e^{j \cdot \gamma_2} & \sin \alpha_3 \cdot \sin \beta_3 \cdot e^{j \cdot \gamma_3} \end{bmatrix} \end{aligned} \quad (5.62)$$

$[\bar{\bar{T}}]$ still has only 9 independent parameters (as before) instead of 12, as suggested by the parametrization (5.62). The dependency relationships linking the real parameters α_i , β_i , δ_i , and γ_i result from the orthogonality of vectors \hat{v}_1 , \hat{v}_2 and \hat{v}_3 .

In the basis $[\hat{v}_1 \quad \hat{v}_2 \quad \hat{v}_3]$ the decomposition of $[\bar{\bar{T}}]$ becomes:

$$[\bar{\bar{T}}] = \lambda_1 \cdot \hat{v}_1 \cdot \hat{v}_1^{*T} + \lambda_2 \cdot \hat{v}_2 \cdot \hat{v}_2^{*T} + \lambda_3 \cdot \hat{v}_3 \cdot \hat{v}_3^{*T} = \lambda_1 \cdot [\bar{\bar{T}}_1] + \lambda_2 \cdot [\bar{\bar{T}}_2] + \lambda_3 \cdot [\bar{\bar{T}}_3] \quad (5.63)$$

The backscattering mechanism is here considered as the incoherent sum of three elementary independent mechanisms ($[\bar{\bar{T}}_1]$, $[\bar{\bar{T}}_2]$ and $[\bar{\bar{T}}_3]$). They are fully polarized (because each of the coherence matrices is defined from a pure target vector, respectively \hat{v}_1 , \hat{v}_2 and \hat{v}_3). From a mathematical point of view, the fully polarized nature of these elementary mechanisms is expressed by the rank 1 of $[\bar{\bar{T}}_1]$, $[\bar{\bar{T}}_2]$ and $[\bar{\bar{T}}_3]$, i.e. each of these has only one non-zero eigenvalue (respectively λ_1 , λ_2 and λ_3).

5.10.3.2 Entropy

The entropy H is an indicator of the respective weights of the polarized and unpolarized components in the make-up of the backscattered wave. H in fact indicates the local ‘degree of disorder’ (entropy) of the polarimetric response. H is defined as the logarithmic sum of the normalized eigenvalues of $[\bar{\bar{T}}]$:

$$H = - \sum_{i=1}^3 P_i \cdot \log_3(P_i) \quad (5.64)$$

where P_i is given by:

$$P_i = \frac{\lambda_i}{\sum_{j=1}^3 \lambda_j} \quad (5.65)$$

Entropy H lies between 0 and 1. There are two extreme cases:

(1) The backscattering mechanisms are fully polarized (pure target): only one of the eigenvalues is non-zero (for example λ_1), $P_1 = 1$ and $P_2 = P_3 = 0$. Consequently, $H = 0$ (since: $x \cdot \log_3(x) \rightarrow 0$ as $x \rightarrow 0$). *The entropy of a fully polarized backscattering mechanism is zero.*

(2) The backscattering mechanisms are fully unpolarized (polarimetric noise): in this case, it is impossible to extract a dominant fully polarized signature from the measurements. The three contributions of equal energy merge randomly, $P_1 = P_2 = P_3 = 1/3$, which results in $H = 1$. *The entropy of a fully unpolarized backscattering mechanism is equal to one.*

Between these two extreme cases, the target behaves like a mixture of three pure targets with comparable intensities, whose effects are summed incoherently, and whose probability for existence is P_i .

Please note: The Entropy H is to some extent a measure of the ‘polarimetric quality’, providing information about how efficient polarimetric analysis can be locally. There is a clear analogy with the interferometric degree of coherence, which gives information about the quality of phase measurements (Sect. 4.2.2).

Figure 5.13(b) displays the entropy of the Flevoland area. As a first approximation, entropy relates inversely to $|\rho|$ coherence (Fig. 5.13(a)) (low entropy \approx high coherence) and as such conveys the same type of information (also related to the ratio between the fully polarized and unpolarized backscattered energies). However the entropy indicates a wider contrast than coherence. The entropy has algebraic properties that make it insensitive to the wave’s decomposition basis (in particular the angle Ψ of Fig. 5.8). This is also true for the degree of polarization P_w (Sect. 1.1.3.2), and also for the Huynen parameters A_0 and B_0 . The coherence $|\rho|$ does not fall into this category of so-called “polarimetric invariant”. To demonstrate this dependency of the degree of coherence on the choice of the polarization basis, let us for example show the degree of coherence $S_{rr} - S_{ll}$ between the co-polarization signals expressed in a basis of circular polarizations, S_{rr} and S_{ll} (Fig. 5.13(c)). The result is very different to that given by the $|\rho|$ coherence (and nicer!). Different fields which could not be distinguished in Fig. 5.13(a) stand out clearly on, Fig. 5.13(c) and vice versa. In addition, circular coherence gives better textural information than $|\rho|$ coherence, in accordance with [Mattia, 1997].

5.10.3.3 Dominant/average backscattering mechanism

The dominant pure backscattering mechanism is determined from the eigenvector associated with the highest eigenvalue of $[\bar{T}]$. Its coherence matrix is:

$$[\bar{T}]^{dom} = \lambda_1 \cdot \hat{v}_1 \cdot \hat{v}_1^{*T} \quad (5.66)$$

This recalls the Huynen approach (which also filters out a dominant pure mechanism). However, the reduction of backscatter to its dominant mechanism may be deceiving if its level is hardly above noise. For example, if entropy is close to $H = 1$ (the three eigenvalues are almost equal), the dominant mechanism derived in this way is not representative.

[Cloude, 1997] overcame this objection by introducing the *average backscattering mechanism*: each of the parameters of $[\bar{U}_3]$ expressed using Eq. (5.62) (α for instance) is considered to be the result of a random sequence such as $\alpha_1, \alpha_3, \alpha_2, \alpha_2, \alpha_1, \alpha_2, \alpha_3, \alpha_3, \alpha_1 \dots$ for which the probabilities of $\alpha_1, \alpha_2, \alpha_3$ are P_1, P_2 and P_3 , respectively. The maximum likelihood of α is then:

$$\bar{\alpha} = P_1 \cdot \alpha_1 + P_2 \cdot \alpha_2 + P_3 \cdot \alpha_3 \quad (5.67)$$

This principle is extended to the full set of parameters, so that the target vector associated with the average backscattering mechanism is given by:

$$\hat{v}^{av} = [\cos \bar{\alpha}, \sin \bar{\alpha} \cdot \cos \bar{\beta} \cdot e^{j\bar{\delta}}, \sin \bar{\alpha} \cdot \sin \bar{\beta} \cdot e^{j\bar{\gamma}}]^T \quad (5.68)$$

The parameter $\bar{\alpha}$ (Fig. 5.13(d)) is of particular interest. It varies between 0 and $\pi/2$. $\bar{\alpha} \approx 0$ characterizes single-bounce scattering, $\bar{\alpha} \approx \pi/2$ double-bounce scattering, $\bar{\alpha} \approx \pi/4$ dipole-like scattering (often dominant in volume scattering). A classification algorithm is then derived using the ‘double identity’ ($H, \bar{\alpha}$) to classify various kinds of backscatters. This extends the family of standard mechanisms listed in Sect. 5.4. A physical interpretation of the other parameters of \hat{v}^{av} is given in [Cloude, 1997]. We should also mention that the parameters $\bar{\alpha}, \bar{\beta}, \bar{\delta}$ and $\bar{\gamma}$, are practically insensitive to perturbing parameters such as surface roughness, the moisture of vegetation cover or certain calibration artifacts. The quality of their estimation is correlated with the local value of entropy.

5.10.3.4 Anisotropy

For classification purposes an additional parameter is added to the pair ($H, \bar{\alpha}$), the anisotropy A :

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \quad (5.69)$$

The entities λ_2 and λ_3 are the second and third eigenvalues of $[\bar{T}]$, in decreasing order. A expresses the dissimilarity between the secondary eigenvalues of $[\bar{T}]$. For an unpolarized wave, $H = 1$ and $A = 0$. For a fully polarized wave, $H = 0$ and A is undetermined. Elsewhere, for low values of H , λ_2 and λ_3 are noisy and the usefulness of anisotropy is reduced. Anisotropy is most appropriate for entropy values between 0.7 and 0.9, i.e. when two backscattering mechanisms are prevalent. Fig. 5.13(e) gives some examples of these unusual situations. Overall, the anisotropy is low and noisy except for certain plots, which return a mixture of single- and double-bounce scattering, as shown by the corresponding Freeman analysis (Fig. 5.12).

5.10.3.5 (H, α) decomposition

Of all the parameters of \hat{v}^{av} , only $\bar{\alpha}$ can always be determined. It can also be used to discriminate among standard backscattering mechanisms discussed previously. Cloude and Pottier [Cloude, 1997] propose a classification algorithm based on partitioning the

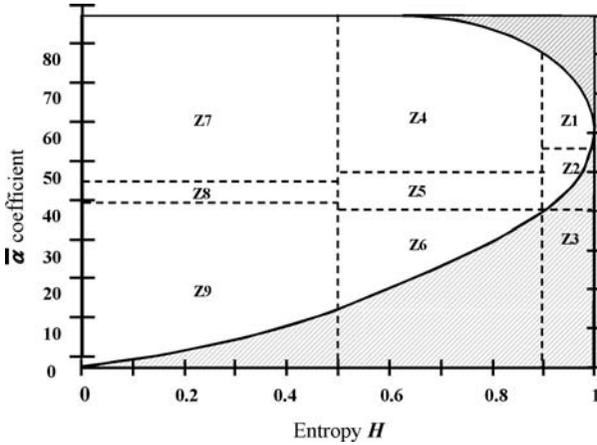


Fig. 5.14 Partition of the $[H, \bar{\alpha}]$ plane into backscattering mechanisms

$[H, \bar{\alpha}]$ plane into nine zones named z_1 to z_9 (Fig. 5.14). However, all regions of this plane are not equally accessible. We may grasp intuitively that as the entropy increases, the normalized eigenvalues of $[\bar{T}]$ tend towards the value $1/3$. The result of this is that $\bar{\alpha}$ tends towards a value close to 60° (the limit value is reached when the wave is unpolarized).

In the opposite case, for zero entropy (fully polarized wave), $\bar{\alpha}$ can take any value between 0 and $\pi/2$. We have thus described a ‘forbidden’ area (shaded gray in the Fig. 5.14) which cannot contain any measurement point.

Within the authorized area, two main sub-areas can be distinguished: (z_9, z_8, z_7) and $(z_6, z_5, z_4, z_3, z_2, z_1)$ separated by the boundary $H = 0.5$. The average backscattering mechanisms associated with each of these zones are discussed in reference [Cloude, 1997].

5.11 Practical cases of polarimetric analysis

The following analyses are conducted on an image acquired by the experimental, polarimetric, multi-frequency (L and C band) SIR-C instrument onboard the Space Shuttle Endeavour (April and October 1994), with 10 m resolution and on a data-take from the airborne RAMSES/ONERA L band instrument with 5-m resolution.

5.11.1 Radiometric analysis

Figures 5.15(a) to 5.15(c) display HH , HV and VV radiometric images (SIR-C L band, incidence angle 24°) from the Ulan-Ude region, a vast agricultural valley on the edge of the Bouryates mountain range, close to Lake Baikal (Southeast Russia). The relative gain of the HV channel has been enhanced for display purposes, because

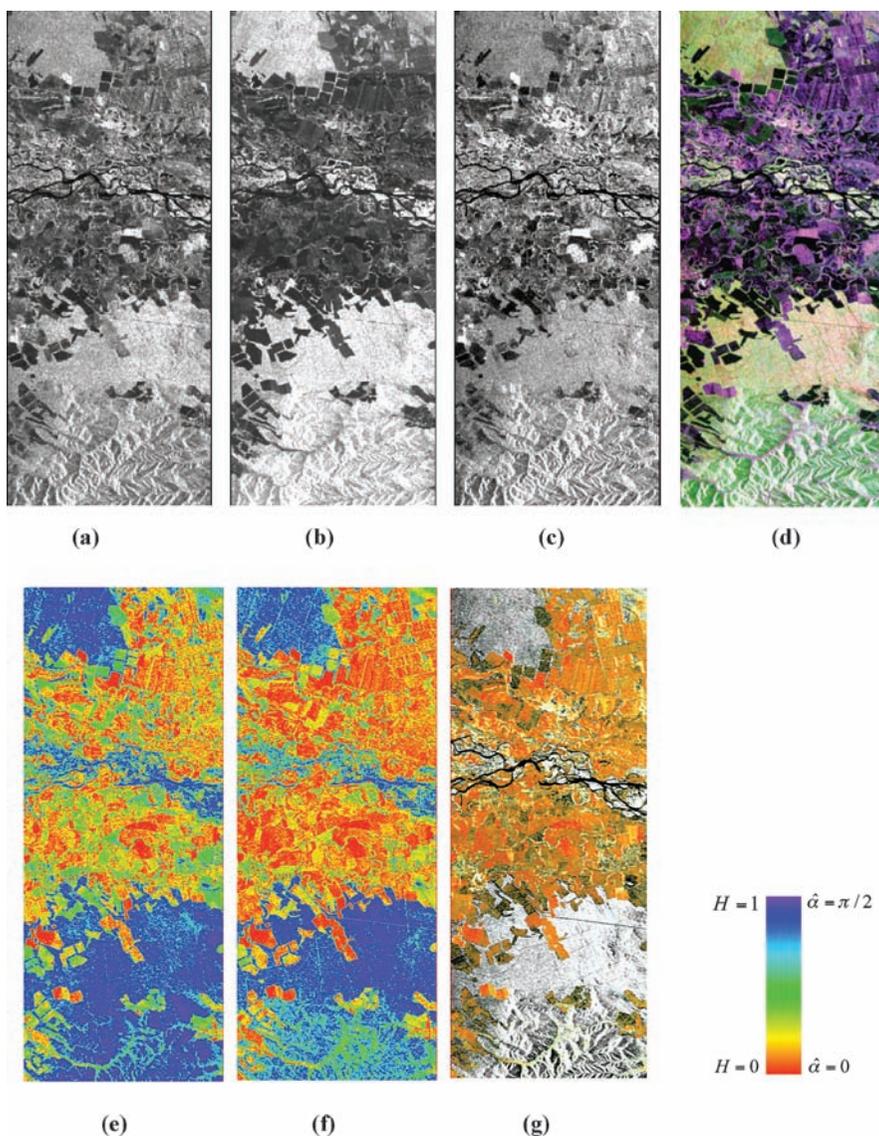


Fig. 5.15 SIR-C image (L band, 24°), over Ulan-Ude, Russia, 1994. (a) HH ; (b) HV ; (c) VV ; (d) Color composite HH (R), HV (G), VV (B); (e) entropy H ; (f) Average $\bar{\alpha}$ coefficient; (g) IHS composite, intensity (span), hue (average $\bar{\alpha}$ mechanism), saturation ($1-H$).

its mean radiometry is actually about 8 dB below that of co-polarized signatures. It is difficult to grasp the complementarity of the different radiometric channels when they are examined separately. This complementarity is obvious in the color composite obtained by overlaying the three images (Fig. 5.15(d)). Using selected gains, the shades

of green (e.g. the range of mountains at the bottom at the image) indicate comparable radiometric levels between the channels. A broader palette of colors is visible in the central part of the image, which is irrigated by the Selenga River; this chromatic richness is evidence of high sensitivity to polarization. The areas where blue dominates indicate a deficit in the HH signal relative to VV , linked to dominant Bragg scattering (surface scattering on bare ground). We can also recognize certain saturated portions in the image, caused principally by slopes facing the radar.

The *RAMSES* image from the airborne instrument (Fig. 5.16(a)) shows an airport environment, made up of buildings which saturate the radar signal, vegetation that is generally close-cropped and strips of tarmac with weak radiometry.

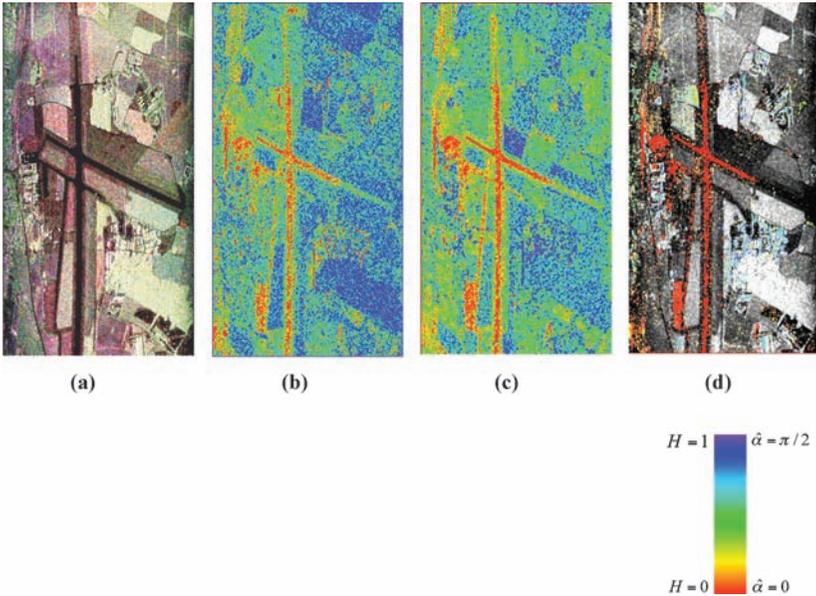


Fig. 5.16 ONERA/RAMESSES image L band, resolution 5 m, acquired over an airport; (a) Color composite HH (R), HV (G), VV (B); (b) Entropy H ; (c) Average $\bar{\alpha}$ mechanism; (d) IHS composite: intensity (span), hue (average $\bar{\alpha}$ mechanism), saturation $(1-H)$.

5.11.2 Entropy analysis

Fig. 5.15(e) shows entropy of the SIR-C image estimated on a sliding window of 7×7 pixels. The high-entropy areas (e.g., the mountain at the bottom of the image) backscatter almost fully unpolarized waves. These high-entropy values generally reflect dominant volume scattering. This hypothesis is confirmed by the high values of the HV signal. Moreover the fact that the HH , VV and HV radiometric values are evenly distributed (Fig. 5.15(d)) indicates that the signal received is practically insensitive to transmission and reception polarizations. Similar results can be observed at C band, although entropy is in average slightly higher than at L band, which makes C band polarimetric analysis less fruitful.

The lowest levels of entropy in the airborne image (Fig. 5.16(b)) are mainly found in a portion of the runway and in built-up areas. An entropy gradient, due to significant variations in incidence angle over the swath, stretches across the image from left to right (i.e., from the near range to the far range): low incidence angles maintain the coherence of the transmitted wave better than high incidence angles, which produce a more diffuse and less polarized signal. A break in the entropy cuts across part of the diagonal runway (it is invisible in the radiometric data of Fig. 5.16(a)). It is likely that this does not have a physical origin but is the result of an incidence angle effect: in near range, the signal returned by the tarmac, although weak, remains above the threshold of image noise $NE\sigma^{0tot}$ (Sect. 3.7.4) and can therefore imprint its own polarimetric signature, i.e. a dominant polarized component (weak H). The intensity of the tarmac's signature decreases as we leave the near range until it reaches the level of image noise. Beyond this limit, the polarimetric analysis applies to the image noise itself, which by nature is unpolarized and therefore highly entropic. It is therefore possible to evaluate image noise $NE\sigma^{0tot}$ using polarimetry!

5.11.3 Average backscattering mechanism

The coefficient $\bar{\alpha}$ of the SIR-C image (Fig. 5.15(f)) is strongly correlated with entropy. For low entropy, the backscattered signal is highly polarized, with significant complementarity between the different polarization channels HH , VV and HV (see the range of colors at the center of the image in Fig. 5.15(d)). The areas concerned correspond principally to surfaces that produce a dominant single-bounce scattering, associated with a low value of $\bar{\alpha}$.

In the airborne image, the coefficient $\bar{\alpha}$ (Fig. 5.16(c)) also shows strong correlation with the entropy H . The low values of H are also linked to low values of $\bar{\alpha}$ (surface scattering). When there is a volume component in the backscattering this causes an increase in H , but also in $\bar{\alpha}$. At the boundary ($H = 1$), $\bar{\alpha}$ converges towards an attractor (close to 60°) whose physical significance is difficult to explain (Sect. 5.10.3.5). A notable exception to the identical behaviors of H and $\bar{\alpha}$ concerns strongly polarizing mechanisms other than single-bounce scattering, such as double-bounce scattering on buildings ($\bar{\alpha} \approx \pi/2$). However, there are not enough of them in statistical terms to counter the impression of similarity between H and $\bar{\alpha}$ observed on numerous images.

5.12 Synoptic representation of polarimetric information

Since polarimetric information is multi-dimensional, with more than three dimensions, it is impossible to display a complete representation of it. However, for high entropy, complete representation of polarimetric information is superfluous (i.e., the phase information is useless). An adaptive display inspired by interferometry proposed in [Imbo, 1999] is based on a decomposition into *Intensity-Hue-Saturation (IHS)*.¹⁴

¹⁴The *intensity* of an image pixel is related to its radiometric content (i.e. the 'black and white' component), *saturation* refers to its level of coloring and *hue* to the color itself. *IHS* representation is an alternative to

This system automatically reduces the representation of the polarimetric information to its radiometric part whenever the signal is strongly unpolarized.

- The *Intensity* channel carries a layer of radiometric information, for example the image *span*, i.e., the incoherent sum of the radiometric signatures $|S_{hh}|^2 + |S_{vv}|^2 + 2|S_{hv}|^2 = HH + VV + 2HV$, or possibly the output from a polarimetric whitening filter [Novak, 1990]. The image background is consequently in black and white and with reduced speckle thanks to incoherent summing.
- The *Saturation* channel is controlled by the local entropy H (Sect. 5.10.3.2.). The law governing saturation is usually linear ($S = 1 - H$), although refined relationships may be involved. High entropy leads to low saturation (and vice versa); in other words, it results in an image which is locally black and white, which displays only radiometric information.
- Lastly, since the decreasing entropy colors the image gradually, the *Hue* channel will attribute to each pixel a color related to the local polarimetric behavior (e.g. via the coefficient $\bar{\alpha}$), but only when this is meaningful, i.e. when the backscatter wave is properly polarized.

This method makes it easier to distinguish between polarized and unpolarized areas. In the SIR-C image (Fig. 5.15(g)) in the heart of the unpolarized maelstrom at the bottom of the image, only the peaks of the mountains retain some wave coherence. The central part of the image, on the contrary, includes mostly polarized surface scattering. In the airborne case (Fig. 5.13(f)) and (Fig. 5.16(d)), the *IHS* representation emphasizes the incidence angle effect on the entropy of backscattered waves. The blue-tinted points of Fig. 5.16(d) show behavior that is both low in entropy and high in $\bar{\alpha}$, in other words polarized double-bounce backscattering created by built-up areas. Lastly, the *IHS* representation of an *X* band RAMSES image with 1.5 m resolution (Fig. 5.17) including a strip of tarmac marked out with artificial point targets (dihedral and trihedral corner reflectors) shows to what extent high frequencies (even under high resolution conditions) can prejudice the overall degree of polarization of backscattered waves. Only a few points resist ‘entropic overflow’, including the point targets (whose nature is determined by their hue value, i.e. their color) and the buildings.

Please note: A similar representation is useful for interferometry (intensity is the amplitude image, hue is phase, and coherence is saturation). The advantage is if the phases are meaningless (i.e., low coherence), the corresponding landscape appears in black and white.

5.13 Future compact polarimetric systems

Although current spaceborne programmes seem favorable to polarimetry (Sect. 5.1), it is still difficult to implement full polarimetry in space: antenna technology, telemetry

the trichromatic *red-green-blue* decomposition in a color image.

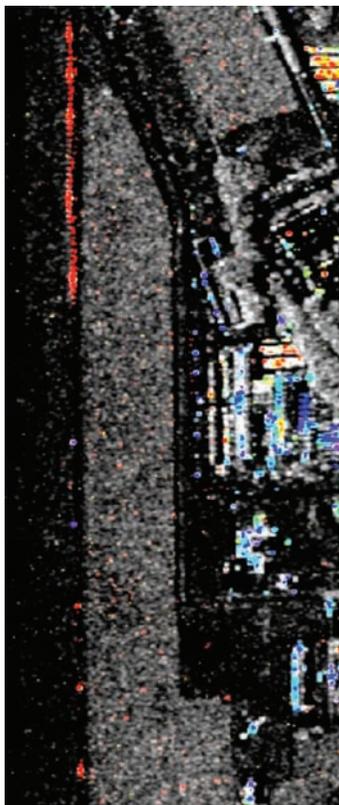


Fig. 5.17 Extract from ONERA/RAMESIS image, X band, resolution of 1.5 m, acquired over an airport. IHS composition: Intensity (span), Hue (average mechanism), Saturation (1-H).

data rates, swath sizes and power consumption are all critical factors. It is interesting to evaluate strategies for partial acquisition of polarimetric information. This approach also seems attractive due to the above-mentioned partial redundancy, of the information coming from the different channels (redundancy between HH , VV and HV intensities, between H entropy and ρ coherence, etc.).

There are two usual areas of investigation, which have been studied in the past [Imbo, 2000], [Lee, 2001]: (1) measurements reduced to incoherent acquisition from co- and cross-polarized channels e.g., $|S_{hh}|$ and $|S_{hv}|$; and (2) measurements reduced to coherent acquisition from only co-polarized channels (e.g. S_{hh} and S_{vv}). Each of these acquisition modes produces three real values per pixel, reducing the volume of useful data to 3/5 of what is necessary for full polari-break metry.

The first strategy requires only one single transmission polarization for reception on two orthogonal channels (the co- and cross-polarized transmission channels). ASAR/ENVISAT's *Alternating Polarization* (AP) mode falls into this category. The use of a single polarization considerably simplifies system design. The pulse repeti-

tion frequency f_a remains at its standard value, as does the instrument swath width (Sect. 5.2.1). However, in order to guarantee acceptable signal-to-noise ratio (Sect. 3.7.4) for measurement of the cross-polarization signal, the transmission power must be increased by a few dB relative to the power budget for co-polarization acquisition. The efficiency of this mode was shown to be limited for thematic applications.

The second strategy relies on the coherent acquisition of complex backscattering coefficients S_{hh} and S_{vv} . There is no cross-polarization measurement here, which relaxes the constraint on the instrument power budget. This mode gives access to the measurement of the degree of coherence between the S_{hh} and S_{vv} signals (Sect. 5.9.1). The polarization agility in transmission leads to a choice of f_a and a swath width reduction as described in Sect. 5.2.1. The efficiency of this mode for thematic applications is excellent, but it requires instrument design of a complexity comparable to that needed for full polarimetry.

5.13.1 Another idea: compact polarimetry and the $\pi/4$ mode

In order to take advantage of the previous options without suffering from their drawbacks, a compact polarization design based on mixed transmission and reception polarization measurements has been proposed [Souyris, 2005]. The single transmission polarization is chosen either circular or linear but oriented at $\pi/4$ with respect to the horizontal. Reception is made on two linear orthogonal polarizations, horizontal and vertical. It is then possible to reconstruct full polarimetric information with some additional physical hypotheses. One of them involves the reflection symmetry of the media observed [Nghiem, 1992]; the other is an empirical relationship linking backscattering coefficients and the degree of coherence ρ (Sect. 5.9.1). This empirical relationship obeys the theoretical developments of the Freeman approach (Sect. 5.10.2). The $\pi/4$ mode efficiency is assessed at two levels: first by considering how well the polarimetric information (for point and distributed targets) is preserved relative to full polarimetry; and second by considering system parameters (swath width, telemetry, link budget). Since this mode (known as the $\pi/4$ mode) uses only one transmission polarization, the azimuth sampling frequency f_a and the swath width can be maintained at their nominal value, while the telemetry data rate is halved.

The $\pi/4$ mode acquisition can be totally and rigorously simulated from a full polarimetric data set. It led to encouraging results in the L band, in the field of land cover classification and target analysis. This is shown by comparing a Landes forest color composite (Fig. 5.18(a)) taken from SIR-C full polarimetry, and the corresponding $\pi/4$ simulation (Fig. 5.18(b)). More detailed analysis of the $\pi/4$ mode has shown its ability to return the ρ , H and α parameters as cleanly [Souyris, 2005].

Bi-static missions using constellations of micro-satellites would be suitable for implementing the $\pi/4$ mode. As an example, a radar constellation with two micro-satellites (for its most economical configuration) can use this principle to reconstruct polarimetric information and therefore open up the possibility of PolInSAR processing [Sect. 5.14]. In this case one of the micro-satellites transmits and receives with circular

polarization, while the other micro-satellite is passive and collects two linear polarizations (horizontal and vertical). The interferometric cartwheel (Sect. 5.7.4) could provide a framework for this kind of implementation. Finally, low frequency mission concepts (P band) could also take advantage of the $\pi/4$ mode under the circular polarization transmission option. The circular option is actually preferred in this case because it partially protects from Faraday effects.

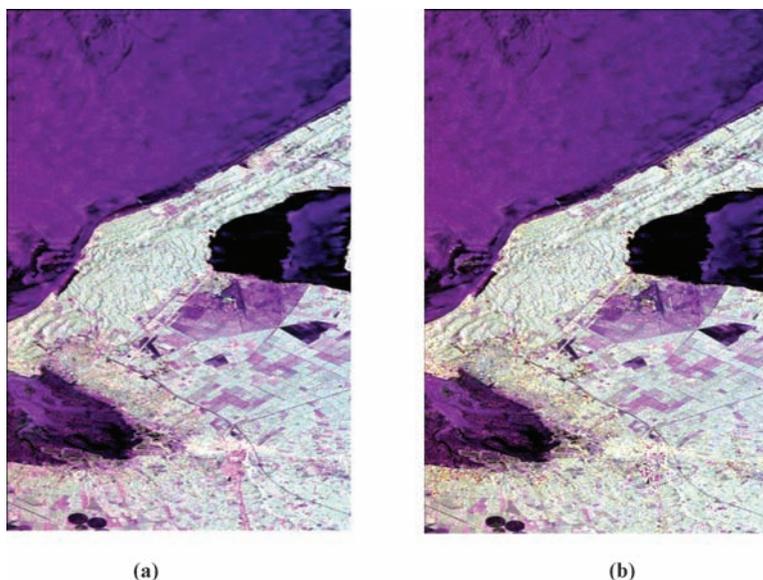


Fig. 5.18 (a) Multi-look polarimetric image, Landes forest, France, L band: color composite HH (R), HV (G), VV (B); (b) Multi-look $\pi/4$ image, same color composite.

5.14 Merging polarimetry and interferometry: PolInSAR

The merging of polarimetry and interferometry, usually called *PolInSAR*, is a major factor behind the growing interest of the radar community towards polarimetric techniques. PolInSAR consists in acquiring polarimetric data under interferometric conditions (Chap. 4), i.e. under slightly different viewing configurations. The advantages of PolInSAR are two-fold: (1) thanks to polarization tuning, the polarimetric acquisitions can be optimally combined so as to maximize their interferometric degree of coherence (Sect. 4.4.2), and consequently to improve the quality of the interferograms produced [Cloude, 1998]; and (2) using a vegetation height inversion algorithm [Cloude, 2003], the extraction of a topographic information becomes feasible, even in the presence of a vegetation cover. Hereafter, we briefly summarize the main aspects of PolInSAR with respects to the above-mentioned assets, before discussing some of its current extensions.

5.14.1 Interferometric coherence optimization

Let us consider the targets vectors \vec{k}_{P1} and \vec{k}_{P2} (Sect. 5.3.2) of two polarimetric data takes acquired under interferometric conditions. We wish to calculate an interferogram between \vec{k}_{P1} and \vec{k}_{P2} . In the formulation (1) of Sect. 4.2.2, the interferogram $I(m, i)$ is calculated from complex scalar value products. In order to extend this formulation to vector combinations, two 3D normalized complex vectors \hat{w}_1 and \hat{w}_2 are introduced. They are interpreted as being the target vectors of unknown scattering mechanisms. The scattering coefficients $M(m, i)$ and $S(m, i)$ introduced in Sect. 4.4.2, are now replaced by the scattering coefficients μ_1 and μ_2 , defined as the projections of the scattering vectors \vec{k}_{P1} and \vec{k}_{P2} onto \hat{w}_1 and \hat{w}_2 , respectively :

$$\mu_1 = \hat{w}_1^+ \cdot \vec{k}_{P1} \quad (5.70\text{-a})$$

$$\mu_2 = \hat{w}_2^+ \cdot \vec{k}_{P2} \quad (5.70\text{-b})$$

where the operator ‘+’ stands for ‘transpose conjugate’. Applying the standard interferogram formulation to μ_1 and μ_2 , we have :

$$I_{\hat{w}_1, \hat{w}_2}(m, i) = \frac{\langle \mu_1 \cdot \mu_2^* \rangle}{\sqrt{\langle \mu_1 \cdot \mu_1^* \rangle} \cdot \sqrt{\langle \mu_2 \cdot \mu_2^* \rangle}} \quad (5.71)$$

where $I_{\hat{w}_1, \hat{w}_2}(m, i)$ is the interferogram obtained at line m and column i . It develops into:

$$I_{\hat{w}_1, \hat{w}_2}(m, i) = \frac{\hat{w}_1^+ \cdot [\bar{\bar{\Omega}}_{12}] \cdot \hat{w}_2}{(\hat{w}_1^+ \cdot [\bar{\bar{\Omega}}_{11}] \cdot \hat{w}_1)^{1/2} \cdot (\hat{w}_2^+ \cdot [\bar{\bar{\Omega}}_{22}] \cdot \hat{w}_2)^{1/2}} \quad (5.72)$$

where $[\bar{\bar{\Omega}}_{11}] = \langle \vec{k}_{P1} \cdot \vec{k}_{P1}^+ \rangle$, $[\bar{\bar{\Omega}}_{22}] = \langle \vec{k}_{P2} \cdot \vec{k}_{P2}^+ \rangle$, $[\bar{\bar{\Omega}}_{12}] = \langle \vec{k}_{P1} \cdot \vec{k}_{P2}^+ \rangle$. The matrix $[\bar{\bar{\Omega}}_{12}]$ is a non-Hermitian 3×3 inter-correlation matrix. Alternatively, $[\bar{\bar{\Omega}}_{11}]$ and $[\bar{\bar{\Omega}}_{22}]$ represent the Hermitian coherence matrices (Sect. 5.9) for each data take.

The magnitude of $I_{\hat{w}_1, \hat{w}_2}(m, i)$, i.e. the coherence of $I_{\hat{w}_1, \hat{w}_2}(m, i)$ can be maximized by tuning \hat{w}_1 and \hat{w}_2 locally. It is performed by introducing a Lagrangian operator [Cloude, 1998]. The phase of $I_{\hat{w}_1, \hat{w}_2}(m, i)$ is given by Eq. (5.73):

$$\Phi_{\hat{w}_1, \hat{w}_2}(m, i) = \text{Arg}(\hat{w}_1^+ \cdot [\bar{\bar{\Omega}}_{12}] \cdot \hat{w}_2) \quad (5.73)$$

In the particular case where $\hat{w}_1 = \hat{w}_2$, i.e. the two data takes forming the interferogram are projected onto the same scattering mechanism, the resulting phase will be closely connected to a topographic information. Therefore, the search for the optimal value $\hat{w}_{1\text{max}}$ that maximizes $|I_{\hat{w}_1, \hat{w}_1}(m, i)|$ leads to the best quality topographical interferogram (Sect. 4.4.2). When $\hat{w}_1 \neq \hat{w}_2$, the phase behavior is more difficult to interpret, as $\Phi_{\hat{w}_1, \hat{w}_2}(m, i)$ jointly depends on an interferometric and on a polarimetric correlation.

5.14.2 Application to the inversion of vegetation height

One of the most encouraging applications of PolInSAR is the inversion of vegetation height using the so-called *Random Volume over Ground model (RVoG)* [Cloude, 2003], which removes the uncertainty in the phase center positioning due to the wave penetration in the volumetric observed media (Sect. 4.4.1).

In the *RVoG* model, the signal backscattered from the forest is modelled as the combination of a ground (surface or double-bounce) contribution and a random volume contribution. At first order (i.e., by neglecting the image noise effect), the ground contribution is assumed to have a coherence value of 1 (i.e., a perfect surface as far as the interferometric behaviour is concerned) regardless of the values of \hat{w}_1 and \hat{w}_2 selected, and a pure topographic phase φ_g linked to the local ground surface elevation. Moreover the propagation and extinction through the random volume are assumed to be polarisation independent. This independence leads to a coherence value for the volume contribution which is also free of polarisation effects.

When both volume and ground contributions participate in the backscatter (and that they are independent from each other), the interferogram $I_{\hat{w}_1, \hat{w}_2}$ is assumed to be a linear combination of a pure ‘ground’ interferogram I^g (i.e. an interferogram between pure ground contributions embedded in each data take) and a pure ‘volume’ interferogram I^v (with a similar interpretation):

$$I_{\hat{w}_1, \hat{w}_2} \approx \frac{\sigma_0^g(\hat{w}_1, \hat{w}_2)}{\sigma_0^g(\hat{w}_1, \hat{w}_2) + \sigma_0^v(\hat{w}_1, \hat{w}_2)} \cdot I^g + \frac{\sigma_0^v(\hat{w}_1, \hat{w}_2)}{\sigma_0^g(\hat{w}_1, \hat{w}_2) + \sigma_0^v(\hat{w}_1, \hat{w}_2)} \cdot I^v \quad (5.74)$$

where both I^g and I^v are independent of \hat{w}_1 and \hat{w}_2 . Reversely, the ground and volume backscattering coefficients (Sect. 3.3), $\sigma_0^g(\hat{w}_1, \hat{w}_2)$ and $\sigma_0^v(\hat{w}_1, \hat{w}_2)$, indicate a

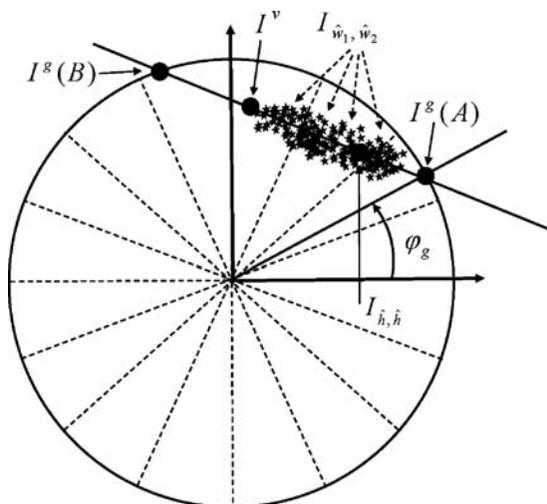


Fig. 5.19 Locus of PolInSAR interferogram in the polar representation plane.

dependency on polarization (Sects.3.2, and 3.3). As a consequence, the locus of $I_{\hat{w}_1, \hat{w}_2}$ in the polar representation plane (Fig. 5.19) is a line segment joining the two points characteristic of I^s (for $\sigma_0^v(\hat{w}_1, \hat{w}_2) \approx 0$) and I^v ($\sigma_0^s(\hat{w}_1, \hat{w}_2) \approx 0$).

Assuming the polarization behavior described in Eq. (5.74), a PolInSAR measurement inversion scheme has been proposed [Cloude, 2003]. Using a radiative transfer approach (Sect. 1.3.3), *RVoG* allows to infer the ground topography, the vegetation height and the attenuation through the canopy from the geometric parameters of the line segment obtained when \hat{w}_1 and \hat{w}_2 vary. The inversion scheme can be decomposed into four steps: (1) synthesize many interferometric coherences by varying the projections onto the scattering mechanisms \hat{w}_1 and \hat{w}_2 in order to determine the coherence line segment; (2) identify the two intersection points *A* and *B* between the coherence line and the circle, then select the one which is the closest to the point representative of $I_{\hat{h}, \hat{h}}$, i.e. the interferogram obtained when the *HH* channels of each data takes are combined (the reason is that ground-trunk interaction is generally strongest in *HH* than in *VV*). The ground phase is then estimated by φ_g (Fig. 5.19); (3) select the interferometric coherence furthest from the ground coherence as I^v . This assumes that there is a polarisation state of the transmit and receive antenna such that no contribution from the ground occurs; and (4) from the estimated I^v and the line segment parameters, the vegetation height and the attenuation are finally inverted by using the radiative transfer equation of the ‘ground + vegetation’ backscatter [Cloude, 2003].

5.14.3 PolInSAR extensions

Since the arrival of PolInSAR, several extensions were proposed in the literature.

- The first one concerns the frequency band under use. So far, the *L* band has been preferred, as it allows a good balance between ground and volume contributions (see the typical behaviour of backscatter in Fig. 5.19(a), Sect. 1.3). This good balance insures that the extreme points of the line segments generated in the polar representation plane by varying \hat{w}_1 and \hat{w}_2 are reached and properly separated, which then allows for an accurate determination of the coherence line segment, and subsequently of the vegetation canopy parameters.

Concurrent frequencies can be considered, such as *X* and *P* band, although the balance between ground and volume contributions may become more questionable. The reduced penetration capabilities (Sect. 1.3.4) of *X* band may lead to a more critical estimation of the coherence line segment, due to the trouble in capturing a pure ground contribution. The locus of coherence points will be less extended than for the *L* band reference case. Reversely, the increased penetration capabilities of *P* band will presumably lead to an opposite conclusion, i.e. the difficulty to capture a pure volume contribution.

- The PolInSAR algorithm can be applied between sub-looks of a polarimetric image (giving the so-called ‘internal PolInSAR’), with applications in target detection and analysis [Souyris, 2003].

- Finally, compact polarimetry leads naturally to the development of compact PolInSAR algorithms in order to ease the compliance of PolInSAR missions with spaceborne mission constraints [Dubois], especially at low frequency.

5.15 Conclusion

Polarimetry is now considered systematically in designs for space SARs, but to different degrees. ASAR/ENVISAT features an *Alternating Polarization (AP)* mode with 30-m resolution in *C* band, but simultaneous acquisition of only two of the three radiometry channels *HH*, *HV* and *VV*. In *C* band RADARSAT II's 10-m full polarimetric mode was launched in December 2007. PALSAR/ALOS, launched by the Japanese Space Agency, JAXA, in January 2006 is applying polarimetry in *L* band, while TERRASAR-X, launched in June 2007 does the same in *X* band. In such a favorable context, it is increasingly important to understand polarimetry, to be familiar with its workings and to be aware of its limitations. The intention behind this chapter was to help the reader grasp all of these aspects.

The physics behind SAR polarimetry can chiefly be broken down into two stages: first, a fixed and isolated point scatterer returns an echo that modifies the polarization of the wave that illuminates it. This polarization change occurs without creating any depolarization. Next, we can derive average polarimetric behavior when a variable number of scatterers combine locally. The averaging process causes depolarization and consequently reduces the efficiency of polarimetry, by increasing the redundancy of information carried by the different polarization channels. To reduce these negative effects we must reduce the number of scatterers inside a radar image pixel. Low frequencies (*L* or *P* band) and high resolutions (in the meter range) both help reduce this number, which means that the spaceborne missions discussed above do not necessarily possess the features needed for high quality polarimetry.

It is difficult to estimate the extent to which the efficiency of polarimetric systems decreases when polarimetric measurements are only partially acquired. A great deal of work remains to be done to finely evaluate the potential of compact polarimetry in specific thematic applications, such as biosphere studies, planetology, detection and identification of point targets, etc. and also the necessary trade-offs between system costs and performance.

To conclude, although polarimetry—whether full or compact—does not replace high resolution, it nonetheless complements it. Far from competing with interferometry, it promises to improve it throughout PolInSAR.

References

- [1] M. Borgeaud, R. T. Shin, J. A. Kong, 'Theoretical models for polarimetric radar clutter', *Journal of Electromagnetic Waves and Applications*, 1(1), (1987).

- [2] W. M. Boerner, H. Mott, E. Lüneberg, C. Linvington, C. Brisco, R. J. Brown, 'Polarimetry in radar remote sensing: basic and applied concepts', Chapter 5 (pp. 271–357) in F. M. Henderson and A. J. Lewis (eds.) *Principles and applications of imaging radar*, Vol. 2 of *Manual of remote sensing* (R. A. Reyerson, 3rd edition, John Wiley & Sons, New-York, ISBN: 0–471–29406–03.
- [3] W. L. Cameron, N. Youssef, L. K. Leung, 'Simulated polarimetric signatures of primitive geometrical shapes', *IEEE Transactions on Geoscience and Remote Sensing*, 34(3), 793–803, (1996).
- [4] S. R. Cloude, E. Pottier, 'A review of target decomposition theorems in radar polarimetry', *IEEE Transactions of Geoscience and Remote Sensing*, 34(2), (1996).
- [5] S. R. Cloude, E. Pottier, 'An entropy based classification scheme for land applications of polarimetric SAR', *IEEE Transactions of Geoscience and Remote Sensing*, 35(1), 68–78, (1997).
- [6] S. R. Cloude, K. P. Papathanassiou, 'Polarimetric SAR interferometry', *IEEE Transactions of Geoscience and Remote Sensing*, 36(5), (1998).
- [7] S. R. Cloude, K. P. Papathanassiou, 'A three stage inversion process for polarimetric SAR interferometry', *IEE proceedings, Radar, sonar and Navigation*, 150(3), 125–134, (2003).
- [8] G. A. Deschamps, 'Geometrical representation of the polarisation of a plane electromagnetic wave', *Proceedings of the I.R.E.*, 39, pp. 540–544, (1951).
- [9] P. C. Dubois, J. J. Van Zyl, T. Engman, 'Measuring soil moisture with imaging radars', *IEEE Transactions on Geoscience and Remote Sensing*, 33(4), 915–926, (1995).
- [10] P. Dubois, J. C. Souyris, S. Angelliaume, F. Garestier, 'The compact polarimetry alternative for spaceborn SAR at low frequency', *IEEE Transactions in Geoscience and Remote Sensing*, in press.
- [11] M. Drinkwater, R. Kwok, D. Winnebrener, E. Rignot, 'Multi-frequency Polarimetric SAR Observations of Sea-ice', *J. Geophys. Res.*, 96(C11), 20679–20698, (1993).
- [12] A. Freeman, S. L. Durden, 'A three-component scattering model for polarimetric SAR data', *IEEE Transactions on Geoscience and Remote Sensing*, 36(3), 963–973, (1998).
- [13] C. D. Graves, 'Radar polarization power scattering matrix', *Proc. IRE*, 44(2), 248–252, (1956).

-
- [14] P. Imbo, J. C. Souyris, A. Lopes, P. Marthon, 'Synoptic representation of polarimetric information', *Proceedings CEOS SAR calibration workshop*, (1999).
- [15] P. Imbo, J. C. Souyris, 'Visualization of the polarimetric information: comparison between partial versus full polarimetry architectures', *Proceedings IGARSS'00*, (2000).
- [16] J. R. Huynen, 'Phenomenological theory of radar targets'. Doctoral thesis, University of Delft, Netherlands, (1970).
- [17] E. M. Kennaugh, *Polarization properties of radar reflectors*, M. of Sc. Thesis, Dept. of Electrical Engineering, Ohio State University, Columbus (1952).
- [18] E. Krogager, 'New decomposition of the radar target scattering matrix', *Electronic Letters*, 26(18), 1525–1527, (1990).
- [19] J. A. Kong, *Electromagnetic Wave Theory*, Second Edition, John Wiley & Sons, Inc., (1990).
- [20] A. B. Kostinski, B. D. James, 'Optimal reception of partially polarised waves', *Journal of Optical Society of America A*, 5(1), (1988).
- [21] T. Le Toan, A. Beaudoin, J. Riom and D. Guyon, 'Relating forest biomass to SAR data', *IEEE Trans. on Geoscience and Remote Sensing*, 30(2), (1992).
- [22] J. S. Lee, M. R. Grunes, G. De Grandi, 'Polarimetric SAR speckle filtering and its implication for classification', *IEEE Trans. on Geoscience and Remote Sensing*, 37(5), 2363–2373, (1999).
- [23] J. S. Lee et al. 'Quantitative comparison of classification capability: fully-polarimetric versus partially polarimetric SAR', *IEEE Trans. on Geoscience and Remote Sensing*, 39(11), 2343–2351, (2001).
- [24] E. Luneburg 'Polarimetry: A Revision of Basic Concepts', in Cloude S R, A. H. Serbest (eds.) *Direct and Inverse Electromagnetic Scattering*, Pitman Research Notes in Mathematics Vol. 361, Longman Scientific and Technical, ISBN 0–582-29964–0, 257–275, (1996).
- [25] F. Mattia, T. Le Toan, J. C. Souyris, G. De Carolis, N. Floury, F. Posa, G. Pasquariello, 'The effect of surface roughness on multifrequency polarimetric SAR data', *IEEE Trans. on Geoscience and Remote Sensing*, 35(4), 954–966, (1997).
- [26] S. V. Nghiem, S. H. Yueh, R. Kwok, F. K. Li, 'Symmetry properties in polarimetric remote sensing', *Radio Science*, 27(5), 693–711, (1992).
- [27] L. M. Novak, M. C. Burl, 'Optimal speckle reduction in polarimetric SAR imagery,' *IEEE Trans. Aerosp Electron. Syst.*, 26(2), 293–305, (1990).
- [28] H. Poincaré, *Théorie mathématique de la lumière*, Georges Carre, Paris (1889).

- [29] *POLSAR-PRO* v2.0 educational and processing tool, © Eric Pottier & Laurent Ferro-Famil; ESA contract C 15601/01/I-LG, (2002).
- [30] D. L. Schuler, J. S. Lee, G. De Grandi, 'Measurement of topography using polarimetric SAR images', *IEEE Transactions on Geoscience and Remote Sensing*, 34(5), 1210–1221, (1996).
- [31] J. Shi, J. Dozier, 'Estimation of snow water equivalence using SIR-C/XSAR, Part I & II', *IEEE Transactions on Geoscience and Remote Sensing*, 38(6), 2465–247, (2000).
- [32] G. Sinclair, 'The transmission and reception of elliptically polarized waves'. *Proceedings I.R.E.*, 38, 148–151, (1950).
- [33] J. C. Souyris, C. Henry, F. Adragna, 'On the use of complex SAR image spectral analysis for target detection: assessment of polarimetry', *IEEE Transactions on Geoscience and Remote Sensing*, 41(12), (2003).
- [34] J. C. Souyris, P. Imbo, R. Fjortoft, S. Mingot, J. S. Lee, 'Compact polarimetry based on symmetry properties of geophysical media: the $\pi/4$ mode', *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), 634–646, (2005).
- [35] G. Stokes, 'On the composition and resolution of streams of polarized light from different sources'. *Proceedings of the Cambridge Philosophical Society*, 1, 140–147, (1852).
- [36] R. Touzi, W. N. Boerner, J. S. Lee, E. Lüneberg, 'A review of polarimetry in the context of synthetic aperture radar: concepts and information extraction', *Canadian Journal of Remote Sensing*, 30(3), 380–407, (2004).
- [37] R. Touzi, R. K. Raney, F. Charbonneau, 'On the use of permanent symmetric scatterers for ship characterization', *IEEE Transactions on Geoscience and Remote Sensing*, 42(10), 2039–2045, (2004).
- [38] F. Ulaby, C. Elachi (Eds.), *Radar polarimetry for geoscience applications*, Artech House, 364 pp., (1990).
- [39] J. J. Van Zyl, C. H. Papas, C. Elachi, 'On the optimal polarisations of incoherently reflected waves', *IEEE Transactions on Antenna and Propagation*, AP-35(7), (1987).
- [40] J. J. Van Zyl, 'Calibration of polarimetric radar images using only image parameters and trihedral corner reflector responses', *IEEE Transactions on Geoscience and Remote Sensing*, 20(3), 337–348, (1990).
- [41] H. A. Zebker, J. J. Van Zyl, D. N. Held, 'Imaging radar polarimetry from wave synthesis' *Journal of Geophysical Research*, 92(B1), 683–701, (1987).