

Bayesian Information Extraction from SAR Images

Vom Fachbereich Elektrotechnik und Informatik
der Universität Siegen zur Erlangung
des akademischen Grades

Doktor der Ingenieurwissenschaften
(Dr.-Ing.)

genehmigte Dissertation

vorgelegt von

Dipl.-Ing. Marc Walessa

geboren am 12. Juli 1970 in Köln

1. Gutachter: Prof. Dr.-Ing. habil. Otmar Loffeld
2. Gutachter: Prof. Dr.-Ing. Mihai Datcu

Vorsitzender der Prüfungskommission: Prof. Dr.-Ing. habil. Bernd Freisleben

Tag der mündlichen Prüfung: 25. Oktober 2001

Zusammenfassung

Bayessche Informationsextraktion aus SAR-Bildern

Synthetisches Apertur-Radar (SAR) stellt ein mächtiges Werkzeug zur Erdbeobachtung dar, da es von Wetterbedingungen und Tageszeiten weitgehend unbeeinflusst ist. Die automatische Interpretation der Intensitätsinformation der SAR-Daten ist allerdings auf Grund des Specklephänomens, das einem starken multiplikativen Rauschen entspricht und bei allen kohärenten Bildgebungsverfahren auftritt, extrem aufwendig und schwierig.

In dieser Dissertation wird ein Bayesscher Ansatz zur Informationsextraktion aus SAR-Daten vorgestellt. Der Schwerpunkt der Arbeit liegt auf der Bildfilterung und Schätzung der Radarrückstreufläche zur Erleichterung einer nachfolgenden Bildinterpretation. Von besonderer Wichtigkeit für das vorgestellte Verfahren sind dabei Methoden zur Segmentierung und Merkmalsextraktion, d.h. zur Extraktion zusätzlicher Information aus SAR-Bildern.

Die Filterung, insbesondere von hochaufgelösten SAR-Bildern, erfordert die Erhaltung von wichtigen Bildmerkmalen, wie Texturen, Kanten und Punktstreuern. In dem hier vorgestellten Ansatz, der auf den Grundlagen der Bayesschen Bild- und Datenanalyse basiert, wird Information bezüglich dieser Merkmale extrahiert und zur Rekonstruktion der Radarrückstreufläche verwendet. Im Gegensatz zu konventionellen Verfahren wird damit nicht nur ein einziges, sondern es werden mehrere Modelle zur Beschreibung des Bildinhaltes genutzt.

Zur Darstellung von Textureigenschaften dienen uns Gauß-Markov-Zufallsfelder. Deren texturbeschreibende Parameter werden iterativ unter Berücksichtigung der "Likelihood"-Funktion des Specklerauschens geschätzt. Zur Kantenfindung und zur Anpassung der Nachbarschaft des Gauß-Markov-Modells, d.h. zur Berücksichtigung von Instationaritäten bezüglich der mittleren Rückstreuintensität, wird ein spezielles Segmentierungsverfahren verwendet. Als letztes Merkmal werden Punktstreuer durch eine Analyse des Verhältnisses von Original- und gefiltertem Bild detektiert. Durch geeignete Kombination der so extrahierten Information wird auf Grund der verbesserten Modellbildung im Vergleich zu anderen Verfahren eine genauere Maximum-A-Posteriori-Schätzung der Rückstreufläche erzielt. Zusätzlich zur geschätzten Radarrückstreufläche können die extrahierten Features, d.h. die Texturparameter, die detektierten Kanten und die gefundenen Punktstreuer, zur weiteren Bildinterpretation herangezogen werden.

Abstract

Bayesian Information Extraction from SAR Images

Allowing the acquisition of high-resolution images of the Earth under all weather conditions and both day and night, synthetic aperture radar (SAR) systems represent a very powerful observation tool. However, an automatic interpretation of the information which is contained in the reflected intensity of the SAR data is extremely difficult. These difficulties are due to the speckle phenomenon that can be regarded as a strong multiplicative noise affecting all coherent imaging systems.

In this thesis, a Bayesian approach for information extraction from SAR images is presented. The emphasis of this work lies on speckle removal and estimation of the radar cross-section to obtain images easier to analyze with standard image interpretation tools. Strongly related to this task are methods for feature extraction and segmentation, i.e. methods for the extraction of additional information from SAR images.

Filtering SAR images, especially images of high resolution, requires a good preservation of important features, such as texture, edges and targets. In the discussed approach, which is based on the principles of Bayesian data and image analysis, information about these features is extracted and used for the restoration of the radar cross-section. Not only relying on one single assumption as most techniques do, our algorithm uses multiple models to describe the image content.

To model textural properties, Gauss-Markov random fields are used. Their parameters, which characterize the texture of the image, are iteratively estimated taking into account the likelihood function, i.e. the speckle noise. A region-growing segmentation algorithm is employed to detect edges and adapt the neighborhood of the Gauss-Markov model to preserve non-stationarities in mean backscatter. As a last feature, targets are extracted by analyzing the ratio of original and despeckled data. By suitably combining the extracted information, an improved maximum a posteriori estimate of the cross-section is obtained, which is due to the increased modeling in comparison with other approaches. The whole set of extracted information, such as cross-section, texture parameters, detected edges and point targets, is valuable for further image interpretation.

Acknowledgments

I am indebted to my supervisor at DLR Prof. Dr. Mihai Datcu for his scientific guidance and his encouragement throughout these years of work. I am also grateful to Prof. Dr. Otmar Loffeld of ZESS at the University of Siegen for giving me the opportunity to carry out this thesis. Further thanks go to Prof. Dr. Bernd Freisleben from the University of Siegen for being the chairman of the thesis committee.

The favorable working environment I have encountered was due to my colleagues Dr. Gintautas Palubinskas and Dr. Sorel Stan from the Image Science group at DLR, with whom I had stimulating discussions about my thesis and the algorithms I developed. Further thanks go to Dr. Michael Schröder, Dr. Hubert Rehrauer and Prof.-Dr. Klaus Seidel from the ETH Zürich for their collaboration and fruitful remarks concerning our joint project.

For their friendly cooperation not only in professional matters I would also like to thank my Ph.D. student colleagues Johannes Schulz-Stellenfleth and Tilman Steck. Additional thanks go to my colleague Dr. Susanne Lehner for her enthusiasm and her never-ending source of ideas.

I am especially grateful to Flore Faille for her strong encouragement during the writing of this thesis and for her help concerning the correction of the manuscript. Finally, I want to thank my family for their not only financial support during these three years of studying.

Contents

Zusammenfassung	i
Abstract	iii
Acknowledgments	v
Contents	vii
List of Figures	xv
List of Tables	xvii
1 Introduction	1
1.1 Background	1
1.2 Goal and Motivation	2
1.3 Outline of the Thesis	3
I Theoretical Background	7
2 Synthetic Aperture Radar	9
2.1 SAR Principle	9
2.1.1 Basic Scattering Mechanisms	10
2.1.2 Basic Properties of Side-Looking Radars	11
2.1.3 Pulse Compression	12
2.1.4 Synthetic Aperture	13
2.1.5 Spectral Properties of SAR Images	14

2.2	Radiometrical and Geometrical Effects	15
2.2.1	Foreshortening, Layover and Shadows in SAR Imagery	15
2.2.2	Radiometric Correction	17
2.2.3	Geometric Correction	17
2.3	SAR Image Statistics	17
2.3.1	Coherent Imaging	18
2.3.2	Statistical Properties of the Complex Signal	18
2.3.3	Statistical Properties of Multi-Look Data	23
2.3.4	Correlation Properties of Speckle	29
2.3.5	Statistical Properties of SAR Data with Non-Uniform Cross-Section	29
2.4	Summary	31
3	Bayesian Image Analysis	33
3.1	Bayesian Estimation Theory	33
3.1.1	Probability	35
3.1.2	Parameter Estimation	37
3.1.3	Model Selection	40
3.1.4	Non-Informative Priors and the Principle of Maximum Entropy . . .	43
3.1.5	Analogy to Tikhonov Regularization	45
3.2	Markov Random Fields	46
3.2.1	Neighborhood Systems and Cliques	46
3.2.2	Markov-Gibbs Equivalence	49
3.2.3	Gibbs Models	50
3.3	Summary	57
II	SAR Image Interpretation	59
4	Bayesian Image Despeckling	61
4.1	Overview of Existing Approaches	61
4.1.1	Simple Filters	62
4.1.2	Statistical Filters	64
4.1.3	Multi-Scale Approaches for Despeckling	65

4.1.4	Wiener Filters	66
4.1.5	Bayesian Filters	67
4.2	SAR Image Restoration with Markov Random Fields	69
4.2.1	Markov Random Fields in a Bayesian Framework	69
4.2.2	SAR Image Despeckling with Parametric Gibbs Priors	70
4.3	Model Parameter Estimation Techniques	75
4.3.1	Model Parameter Estimation from Complete Data	75
4.3.2	Model Parameter Estimation from Incomplete Data	81
4.4	Estimation of Hidden Gauss-Markov Model Parameters	84
4.4.1	Gauss-Markov Model Parameter Estimation from Speckled Data	84
4.4.2	Evaluation of Estimation Techniques for Gauss-Markov Parameters	86
4.5	Summary	91
5	Model-Based Despeckling with Gauss-Markov Random Fields	93
5.1	MAP SAR Image Estimation with a Gauss-Markov Prior	93
5.1.1	Fast Computation of the MAP Estimate	94
5.1.2	Error Evaluation for the Square-Root Gamma Approximation	97
5.2	Model Parameter Estimation by Iterative Evidence Maximization	99
5.2.1	Approximation of the Evidence Integral	100
5.2.2	Iterative Model Parameter Estimation	103
5.2.3	Evaluation of the Proposed Optimization Approaches	106
5.3	Properties of the Gauss-Markov Texture Estimator	107
5.3.1	Bias of the Gauss-Markov Texture Estimator	107
5.3.2	Quality of Model Parameter Estimation by IEM	108
5.3.3	Lower Bound of the Mean Square Error	110
5.4	Optimization of the Filtering Quality	112
5.4.1	Model Order Selection	113
5.4.2	Local Parameter Estimation and Spatially Adaptive Filtering	115
5.4.3	Gauss-Markov MAP Filtering in Different Signal Spaces	119
5.5	Summary	122
6	Extraction of Non-Linear Features	123
6.1	Edge-Detection for Model-Based Despeckling	123

6.1.1	Region-Growing Backscatter Segmentation	124
6.1.2	Detection of False Alarms and Region-Merging	128
6.2	Edge Information for Gauss-Markov Despeckling	128
6.2.1	Adaptation of the Gauss-Markov Neighborhood System	129
6.2.2	Introduction of Edge Information	130
6.2.3	Fusion of Texture and Edge Information	131
6.3	Extraction and Preservation of Strong Scatterers	133
6.3.1	Pre-Processing of Targets for Parameter Estimation	133
6.3.2	Post-Processing for the Detection of Blurred Targets	135
6.4	Outline of the Model-Based Despeckling Algorithm	136
6.4.1	Algorithmic Flowchart	137
6.4.2	Model-Based Despeckling with Structural Information	138
6.5	Summary	138
7	Illustration and Study of Results	141
7.1	Quantitative Filter Evaluation based on Synthetic Data	141
7.1.1	Quality Measures	142
7.1.2	Filtering of Artificially Speckled Test Images	143
7.1.3	Conclusions	145
7.2	Qualitative Filter Evaluation based on SAR Data	148
7.2.1	Filtering of Synthetic Aperture Radar Images	148
7.2.2	Conclusions	156
7.3	Full Scene Processing for Information Extraction	156
7.3.1	Information Extraction from SAR Data	157
7.3.2	Conclusions	161
7.4	Summary	161
8	Conclusions	163
8.1	Summary of the Thesis	164
8.2	Evaluation of Obtained Results and Outlook	166
A	Overview of SAR Sensors	169
A.1	Spaceborne SAR Sensors	169

A.2 Airborne SAR Sensors	169
B Optimization Methods	171
B.1 Simulated Annealing	171
B.2 Iterated Conditional Modes	172
B.3 Newton-Raphson Algorithm	173
C Restoration of Blurred Images	175
C.1 Deblurring of Noisy Images	175
C.2 Filtering of SAR Images with Correlated Speckle	176
D Generalized Model-Based Noise Filtering	177
D.1 Model-Based Filtering of Non-Stationary Noise	177
D.2 Model-Based DEM Filtering during the SRTM Mission	179
E Symbols and Acronyms	183
E.1 List of Symbols	183
E.2 List of Acronyms	187
Bibliography	189

List of Figures

1.1	Example of advanced SAR image processing.	4
2.1	Geometry of a side-looking radar system used for surface imaging.	10
2.2	Different backscattering mechanisms.	11
2.3	Resolution of SAR in range and azimuth.	12
2.4	Spectral and correlation properties of complex SAR data.	14
2.5	Slant range illustration of topographically induced radiometrical and geometrical distortions.	16
2.6	Random walk in the complex plane.	19
2.7	Analytically and experimentally derived SAR image distributions.	21
2.8	Complex X-SAR scene.	22
2.9	SAR image pdfs for different values of L and different mean values.	25
2.10	Coefficient of variation for different number of looks L	27
2.11	Qualitative illustration of the speckle auto-covariance function with different resolutions in range and azimuth.	30
3.1	Deductive and inductive logic.	34
3.2	Approximation of the posterior by a Gaussian distribution.	42
3.3	Cliques and neighborhoods.	48
3.4	Realizations of second order neighborhood Gibbs random fields using different potential functions and parameters.	53
3.5	Potential functions and their first derivatives.	55
4.1	Illustration of the despeckling performance for different classes of filters applied to a synthetic image.	63
4.2	Influence of the model and its parameters on the texture restoration ("beach sand").	72
4.3	Influence of the model and its parameters on the texture restoration ("straw").	73
4.4	Approximation of non-linear models by a combination of Gaussians.	77

4.5	Analysis and synthesis using Gauss-Markov random fields of fifth neighborhood order.	79
4.6	Texture restoration using Gibbs random fields.	83
4.7	Plots of estimated directional parameters extracted with different algorithms.	87
4.8	Restoration of speckled textures using Gauss-Markov random fields and different parameter estimation techniques.	88
4.9	Despeckling results from a synthetically speckled image using different techniques for local parameter estimation of a fifth order neighborhood.	89
5.1	Convergence of the ICM (Iterative Conditional Modes) algorithm using the Gauss-Markov prior.	95
5.2	Relative estimation error resulting from the square-root Gamma approximation.	98
5.3	A Gaussian and its square-root Gamma approximation.	98
5.4	Relative errors of the approximated evidence for a single pixel.	103
5.5	Multiplicative bias of the Gauss-Markov texture estimator.	108
5.6	Local model parameter estimation from complete and incomplete data.	109
5.7	Images of the lower bound of the mean square error for the four Brodatz textures.	112
5.8	Model order selection for the Gauss-Markov model	114
5.9	Influence of the parameter estimation window size on image restoration.	117
5.10	Model-based despeckling of an X-SAR image using the Gauss-Markov model.	118
5.11	Restoration of the "sand" texture in different signal spaces.	121
6.1	Region-growing segmentation for edge-detection.	126
6.2	Evolution of the edge-detection process.	126
6.3	Histograms of the noise-free and the speckled test image.	127
6.4	Adaptation of the Gauss-Markov neighborhood system.	130
6.5	Restoration under two different assumptions ("chess-board" image).	132
6.6	Restoration under two different assumptions ("straw" image).	132
6.7	Detection and restoration of strong targets.	134
6.8	False alarm rates as a function of the thresholds for pre- and post-processing.	135
6.9	Cross-sections of the target images.	136
6.10	Flowchart of the full Model-Based Despeckling algorithm.	137
6.11	Model-Based Despeckling (MBD) with structural information.	139
7.1	Artificially speckled test images and MBD filtered results.	143

7.2	Detailed cutout of the "Wessling" test image.	146
7.3	Detailed cutout of the "Brodatz" test image.	146
7.4	Detailed cutout of the "synthetic" test image.	147
7.5	Detailed cutout of the "runway" test image.	147
7.6	X-SAR image over Ukraine.	150
7.7	First detailed cutout of the X-SAR image.	151
7.8	Second detailed cutout of the X-SAR image.	151
7.9	FGAN airborne SAR image over Germany.	152
7.10	First detailed cutout of the FGAN SAR image.	153
7.11	Second detailed cutout of the FGAN SAR image.	153
7.12	Original SRTM (Shuttle Radar Topography Mission) image.	154
7.13	Mean filtered SRTM image.	154
7.14	MBD filtered SRTM image.	155
7.15	GGMAP filtered SRTM image.	155
7.16	SAR image of DLR/Oberpfaffenhofen.	158
7.17	Extracted edge information of the DLR/Oberpfaffenhofen SAR image.	158
7.18	MBD filtered DLR/Oberpfaffenhofen SAR image.	159
7.19	GGMAP filtered DLR/Oberpfaffenhofen SAR image.	159
7.20	Texture information extracted from the DLR/Oberpfaffenhofen SAR image.	160
7.21	Unsupervised clustering of texture and backscatter information from the DLR/Oberpfaffenhofen SAR image.	160
B.1	Convergence of simulated annealing, iterated conditional modes and Newton-Raphson.	173
D.1	Generalized Gauss-Markov noise filtering.	178
D.2	Model-Based DEM (Digital Elevation Model) filtering.	180
D.3	Gradient images of the original and filtered unwrapped DEM.	180
D.4	Perspective view of original and filtered DEM.	181
D.5	Perspective view of original and filtered landscape.	182

List of Tables

3.1	Model parameters used for texture synthesis.	54
4.1	Mean square errors of filtered images obtained from different speckle filters.	64
4.2	Model parameters used for despeckling and resulting mean square errors.	71
4.3	Gauss-Markov random field parameters estimated for the four Brodatz textures.	86
4.4	Quality of image restoration and texture parameters.	90
4.5	Gauss-Markov random field parameters estimated for a uniform cross-section degraded by speckle.	91
5.1	Mean square errors for the restoration of the four Brodatz textures.	99
5.2	Log-evidences E and mean square errors for the restoration of the four Brodatz textures with a Gauss-Markov model of fifth neighborhood order.	107
5.3	Bias of the Gauss-Markov texture estimator.	108
5.4	Gauss-Markov random field parameters estimated for the four noise-free Brodatz textures with $L \rightarrow \infty$	111
5.5	Measured mean square error and its theoretical lower bound.	112
5.6	Comparison of local and global model parameter estimation.	117
5.7	Mean square errors for the restoration of the four Brodatz textures in different signal spaces.	121
6.1	Original and estimated class mean values obtained by the region-growing segmentation.	127
6.2	Edge-detection accuracy of the region-growing segmentation.	127
7.1	Quantitative filter evaluation for the four synthetic test images.	144
A.1	Selection of important spaceborne SAR sensors.	169
A.2	Selection of important airborne SAR sensors.	169

1 Introduction

During the last decade, high-quality images of the Earth obtained by synthetic aperture radar (SAR) systems, both on airborne and spaceborne platforms, have become more and more available. ERS-1 was the first instrument in a series of orbital SARs planned to have long lifetimes and semi-operational capabilities. Currently in orbit are also the ERS-2, JERS-1, JERS-2 and RADARSAT satellite systems. ENVISAT is planned to be launched in 2001. By providing a long series of accurate measurements of the backscatter coefficient, these satellites allow dynamic processes to be observed over most of the Earth's surface. This will have a significant impact on many scientific domains, such as vegetation mapping and monitoring, hydrology, sea-ice mapping and geology.

Spaceborne radar systems are usually simple in the sense that they use single frequencies and polarizations with modest resolution, because of the constraints imposed by their deployment in space. On the contrary, more complex airborne systems have demonstrated the advantages of multiple frequencies and polarizations. An overview on SAR sensors is given in appendix A. These advantages were and will further be proved from space by the SIR-C/X-SAR [XSAR] and the SRTM [SRTM] missions of the Space Shuttle. As a consequence, much longer wavelength systems are now in operation and promise to provide effective methods for reconnaissance and remote sensing over heavily vegetated areas, due to their capability to deeply penetrate the underlying soil.

An enormous amount of effort has been expended on the development of SAR hardware, but comparatively little has been done to fully exploit the information content of the sensed data. To achieve this goal, the actual properties of the data must be examined in light of what is known about the physics of the imaging process and the envisaged applications. In this way, we are able to identify and extract optimal estimates of the information we seek in the data. This bridging between what we know about the data and how we get at the information they contain is the subject of this thesis.

1.1 Background

Radar systems are capable of producing very high quality images of the Earth. For these data to have value, they must be interpreted so as to yield information about the imaged area. However, the task of information extraction and the detection of significant features is very time-consuming if performed manually by an image analyst, and especially when

large amounts of remote sensing data must be examined. Moreover, consistency between the results achieved from different data and different analysts cannot be guaranteed. These limitations of manual data analysis motivate the development of automatic algorithms to extract relevant information more quickly and in a reproducible way. The need for automatic or semi-automatic methods is even larger when polarimetric, multi-frequency and multi-temporal images are used, whose analysis also requires the fusion of all available information.

To be valuable, an automatic data interpretation has to be based on a solid understanding of how the desired information is encapsulated in the data and is generally performed on two different scales [Oliver98]:

- A low-level analysis identifies and quantifies details of local image structure.
- A high-level analysis uses these details to construct a global structural description of the scene required by the image analyst.

Low-level algorithms are usually not sufficient because local details must be fused into some global high-level form of understanding. Examples are the joining of edges across gaps in edge-detection applications or the merging of regions in segmentation algorithms [Cook94]. On the other hand, sophisticated high-level techniques are of limited use if the low-level algorithms are incapable of extracting the information with sufficient accuracy. As a consequence, both levels must be combined in an optimal way in order to yield the information of interest for the user.

1.2 Goal and Motivation

The goal of this thesis is to present a Bayesian approach for high- and low-level information extraction from single channel intensity SAR data [Walessa99a, Walessa99b, Walessa00]. Due to its advantages compared to optical sensors and its steadily increasing resolution, SAR imagery becomes more and more important as a source of remote sensing information. Basic features of potential interest which can be observed in SAR data are homogeneous regions of different mean backscatter, i.e. of different cross-section or backscatter coefficient, textured regions such as forests, edges giving information for data classification and strong scatterers, that are mostly related to man-made targets. However, appropriate means for automated interpretation and extraction of such features from SAR data are still needed. This lack of interpretation tools is mostly due to the presence of speckle noise [Goodman], which makes existing algorithms designed for non-coherent data useless for SAR images. Many potential users do not know how to deal with this kind of imagery and consider it as not suitable for their purposes. To solve this misery, two basic approaches are possible and corresponding attempts have been made in the past:

- Since the desired information is contained in the radar cross-section, filters have been developed to reduce or filter out speckle, enabling the user to apply existing image analysis tools often designed for optical data.

- Special SAR data interpretation tools have been designed to extract the desired information directly from the speckled data without using speckle filters as a pre-processing step. Existing tools for optical data usually cannot be directly applied to SAR images.

In principle, the latter approach should be preferred, since it disposes of the full information and does not suffer from filter-induced artefacts. As far as the first approach is concerned, available filters rely on rather simple assumptions and are usually not capable of modeling and accurately restoring all SAR image features. Typical approaches rely on a local analysis of mean and variance in order to more or less strongly smooth the noisy image [Frost82a, Hagg, Kuan87, Lee83]. Other techniques incorporate additional structural low-level information to better preserve edges [Crimmins, Lopes90b] or are based on wavelet-shrinkage methods. None of these methods employs a sophisticated modeling of image features, which explains their moderate performance. Unfortunately, the second approach is often not feasible, due to its mathematical complexity. The extraction of edges [Desnos, Frost82b, Lopes93, Touzi88] between regions of homogeneous backscatter already requires a good understanding of the statistical properties of SAR data. The same applies for line-extraction techniques. Concerning simple parametric models used for information extraction, e.g. for texture description, the estimation of their parameters usually cannot be done analytically. As of this, the simple ratio of empirical mean to standard deviation known as coefficient of variation is still a widely used texture measure in the SAR community.

To overcome these deficiencies, we propose a new approach combining different kinds of low-level information to obtain improved high-level results. As low-level information, basic SAR image features, such as texture, edges, uniform areas and dominant scatterers, are extracted. This information is not only combined to generate an estimate of the higher-level radar cross-section but is also available for further image analysis. Hence, instead of relying only on one single assumption, as most available despeckling filters do, we use multiple models and estimate their parameters by taking into account properties of the speckle noise. This low-level information is employed to generate an improved restoration of the cross-section, as illustrated in Fig. 1.1, yielding consistent outputs for the cross-section and the extracted features. As a result, a vector of dominant image features and a high-quality estimate of the cross-section characterized by the extracted low-level information are obtained. The latter may be subject to further standard information extraction techniques if the information desired by the user is not contained in the low-level feature vector.

1.3 Outline of the Thesis

The aim of this thesis is to propose a Bayesian approach for information extraction from SAR data, representing an intuitive way to combine physical properties of SAR systems with user-dependent data models. Although a general technique for SAR data analysis is desirable, we concentrate on the interpretation of single channel intensity or amplitude data. Properties of polarimetric, multi-temporal and interferometric data are not discussed. In this work, the filtering of images affected by speckle noise plays the dominant role. We describe how information can be retrieved from under this noise and how speckle can



Figure 1.1: Example of advanced SAR image processing. Left: Original high-resolution airborne SAR data (400×400 pixels, copyright by N. A. Software Ltd.) exhibiting speckle. Right: Restored SAR image of almost optical quality. Speckle has been filtered out with the Model-Based Despeckling algorithm (MBD), which is the subject of this thesis.

be efficiently removed without destroying information that may possibly be extracted by subsequent image interpretation tools.

To be able to make effective use of SAR, it is important to understand its basic properties and the system's image formation process. In chapter two of this thesis, we describe part of the signal processing [Bamler, Barber], geometrical and radiometrical properties [Schreier] and statistical characteristics of SAR image data. The latter are of particular interest for image understanding and information extraction since data interpretation relies on assumptions about the statistics of the data to be investigated. Since we envisage a statistical analysis of SAR data, the basic principles of Bayesian estimation theory [Bayes, Bretthorst, Jaynes, Sivia] and image interpretation are outlined in chapter three. The denoising and the information extraction will be performed completely within the framework of Bayesian inference, which is a very powerful tool comprising other approaches like regularized data inversion. A description of the necessary models for Bayesian image interpretation, which we choose to be in the form of Markov Random Fields being of particular interest for image processing [Li, Winkler], closes the first part of the thesis.

In the second part, we put emphasis on SAR image interpretation. Commonly used algorithms for SAR image despeckling are reviewed and discussed with regard to the employed modeling in chapter four. We illustrate the problem of information extraction, i.e. parameter estimation, for despeckling with Markov random fields and, at the same time, demonstrate the need for more sophisticated solutions. Thus, chapter four prepares the

main part of this work, which is presented in chapters five and six. Here, the full Bayesian algorithm for SAR image despeckling and information extraction using Gauss-Markov models is introduced [Walessa00]. In chapter five the different steps of the algorithm, such as the computation of the MAP (maximum a posteriori) estimate and an iterative texture parameter estimation technique, as well as the used model and computationally necessary approximations are discussed and evaluated. The extraction of additional non-linear features, which cannot be described by the employed texture model, and the fusion of this information into the final filtering result are the subject of chapter six.

Before concluding this work, an extra chapter giving several examples in order to illustrate the performance of the developed algorithm and to compare it to other despeckling techniques follows. Besides presenting filtered images, we also show several examples of the extracted features, such as texture, edges and targets. Additional information completing the thesis can be found in the appendices.

Part I

Theoretical Background

2 Synthetic Aperture Radar

In the first part of this chapter, the basic theory of how a side-looking radar, flown on an airborne or spaceborne platform, uses a synthetic aperture together with pulse compression to generate high-resolution imagery [Bamler, Barber, Oliver98] is presented. For the imaging geometry, a flat Earth model and a straight flight path are assumed. However, for spaceborne SAR systems the imaging geometry must take into account the Earth's curvature and rotation and the satellite's orbit. Because these factors affect the details of both azimuth and range processing rather than the general theory, they are not further considered. The effect of range curvature will be ignored for the same reason.

In the second section, we shortly deal with geometrically induced effects in SAR imagery [Gelautz]. These effects are observed in presence of rough terrain and affect both the radiometric information of the sensed images and their geometry [Schreier]. As of this, SAR images usually require geometric and radiometric correction based on the topography and radiometric calibration in range direction due to a varying incidence angle. Since these issues do not play an important role for the main part of this thesis, we only present these effects for the sake of understanding of the SAR image formation process.

The last section of this chapter is most important for all following results of this work. All ensuing algorithms concerning the interpretation of SAR data are based on the conclusions presented here: The statistical properties of the sensed images [Goodman]. The speckle phenomenon, which gives the motivation for this work, since it hinders standard image interpretation techniques, is explained and statistically described. This knowledge is necessary for Bayesian estimation theory, which represents the basis of the developed and discussed algorithms for filtering and information extraction from SAR images. We close this and all following chapters with a short summary of what has been presented.

2.1 SAR Principle

The basic geometry of a SAR is shown in Fig. 2.1. A platform moving with velocity v at altitude H carries a side-looking radar antenna that illuminates the Earth's surface with pulses of electromagnetic radiation. The direction of travel of the platform is known as the azimuth direction and is denoted by y . The distance from the radar track is measured in the range direction x . For simplicity, we deal only with the case where the antenna is oriented parallel to the flight path, i.e. not squinted. In the following, we point out only the most

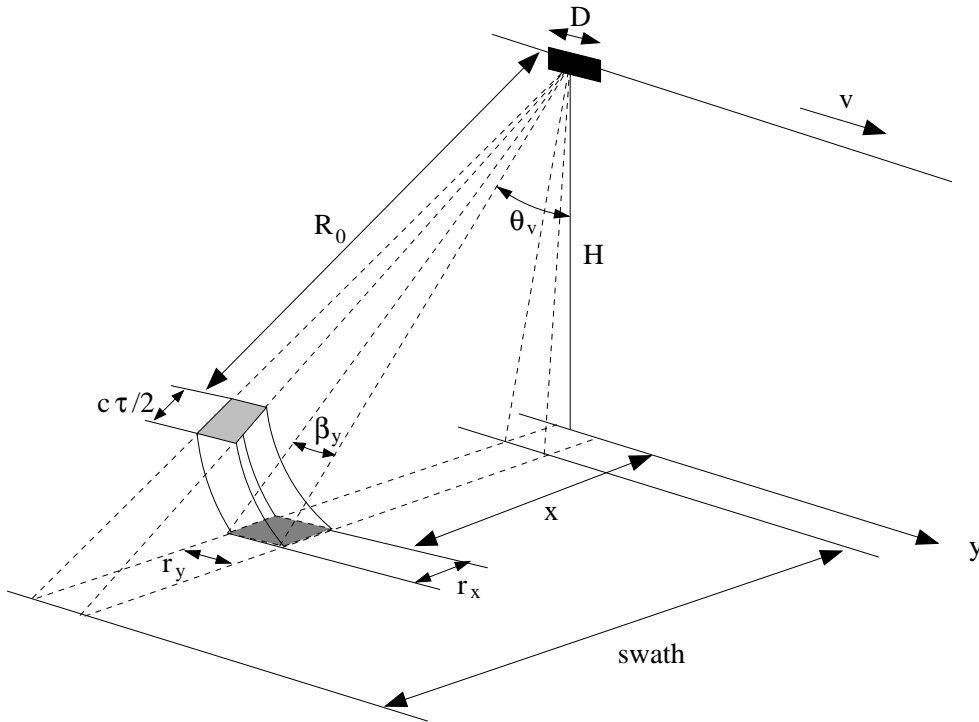


Figure 2.1: Geometry of a side-looking radar system used for surface imaging. The radar with an antenna of width D is flying at altitude H with speed v . The incidence angle for the considered resolution cell is θ_v . The resolutions in range and azimuth are denoted by r_x and r_y , respectively.

important features that characterize a SAR system and make this kind of sensor special compared to other, e.g. optical or non-coherent sensors. A more thorough understanding of the whole SAR processing procedure [Bamler, Barber] is not necessary for the ensuing chapters.

2.1.1 Basic Scattering Mechanisms

The principle of radar systems is to emit electromagnetic waves with wavelengths ranging from a few centimeters up to one decimeter and to receive the backscattered reflection from the imaged surface. The measured time between emission and reception of the reflected wave is used to localize the scatterer or target. Here, the term target denotes both single objects and distributed scatterers. Several basic backscattering mechanisms [Fjørtoft99a] depending on the micro- and macroscopic properties of the scatterer are illustrated in Fig. 2.2.

- **Reflection from smooth surfaces and double bouncing:** Unless the incidence angle θ_v between sensor and surface is zero and if no double bouncing occurs, a smooth surface reflects very few of the incoming energy back to the emitting antenna. For

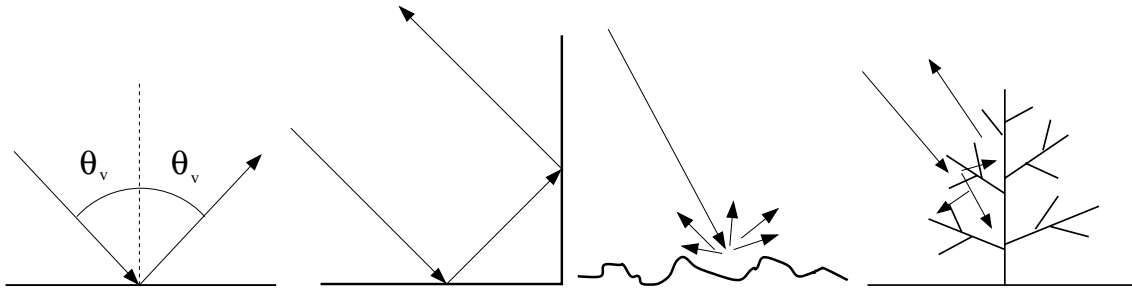


Figure 2.2: Different backscattering mechanisms. From left to right: Scattering from a smooth surface, double bouncing reflection, diffuse scattering from rough surfaces, volume scattering.

this reason, highways and calm lakes appear rather dark in SAR images. Buildings usually appear bright due to the double bouncing effect.

- **Reflection from rough surfaces and volume scattering:** Rough surfaces, i.e. non-specular reflectors, reflect the wavefront in multiple directions. Only a part of the emitted energy is received by the sensor. This effect is called diffuse scattering. The same effect occurs for volume scattering, e.g. in forests, where the wavefront partly penetrates the scatterers. The characteristics of volume scattering and the depth of penetration are a function of the employed wavelength of the system.

By measuring the reflected energy and the time delay between emission and reflection and by doing an appropriate processing, SAR images can be computed that are characteristic of the electromagnetic properties of the imaged surface. Thus, the obtained measurements allow for subsequent data interpretation to recover useful information.

2.1.2 Basic Properties of Side-Looking Radars

To illustrate the special characteristics of side-looking SAR systems, we consider a radar system mounted on a platform, either airborne or spaceborne, as depicted in Fig. 2.1. The sensor is flying at altitude H above ground, its distance from the target to be imaged is R_0 and the incidence angle is denoted by θ_v . Hence, an emitted pulse is received after $\Delta t = 2R_0/c$, where $c \approx 3.0 \cdot 10^8 \text{m/s}$ is the speed of light. We assume that the emitted waves can be described by rectangular pulses of duration τ that are repeatedly sent with the pulse repetition frequency $f_T = 1/T$, with $T \gg \tau$. The minimum distance between two distinguishable different objects, i.e. the sensor resolution, can then be calculated to be $\Delta R = c\tau/2$. For the ground resolution this results in

$$r_x = \frac{c\tau}{2 \sin \theta_v}, \quad (2.1)$$

as shown in Fig. 2.3. We remark that r_x is independent of H and that better resolutions are achieved by larger values of θ_v . Hence, resolution is improved with growing range distance

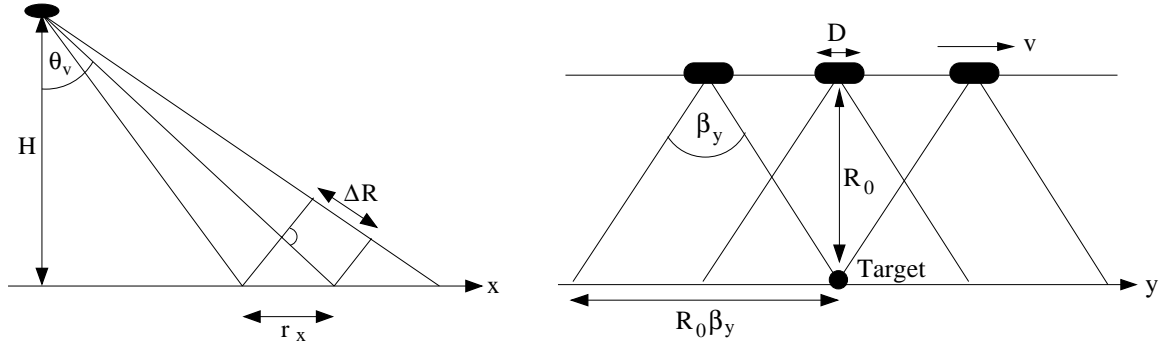


Figure 2.3: Resolution of SAR in range and azimuth. Left: Imaging geometry in range direction. Right: Target illumination by the moving sensor used to generate the synthetic aperture.

x . For RADARSAT, θ_v varies between 20° and 59° depending on the used mode. Since θ_v is constrained to a certain range, e.g. due to the signal-to-noise ratio, the main parameter to increase resolution remains the pulse length τ . The emitted pulses must be as short as possible, but at the same time, they must contain enough energy to guarantee a sufficient amount of reflected intensity at the receiver. However, the theoretical optimal signal, a Dirac pulse, is technically not realizable.

The resolution in azimuth denoted by r_y depends on the distance R_0 and on the antenna aperture in azimuth β_y . The width of the main antenna lobe at -3dB is given by

$$\beta_y = \frac{\lambda}{D}, \quad (2.2)$$

where λ is the wavelength of the emitted signal and D denotes the antenna length in azimuth. The resolution in azimuth direction can then shown to be

$$r_y = \beta_y R_0 = \frac{\lambda R_0}{D} = \frac{\lambda H}{D \cos \theta_v} \quad (2.3)$$

and depends mainly on the altitude H , the wavelength λ and the antenna size D . Variations of H and λ are rather limited for physical reasons, so that the antenna size D remains the only means to increase resolution. Naturally, D is constrained for practical reasons.

Considering the ERS satellites for which $\tau = 37.1\mu\text{s}$, $\theta_v = 23^\circ$, $\lambda = 5.3\text{cm}$, $D = 10\text{m}$ and $H = 780\text{km}$, this yields resolutions of $r_x = 14.2\text{km}$ and $r_y = 4.5\text{km}$. For practical applications this result is not acceptable and a special technique is used to considerably increase resolution in range and azimuth.

2.1.3 Pulse Compression

The range resolution can be dramatically increased by a technique called pulse compression. In this case, the radar does not emit a rectangular signal modulated by a carrier frequency

f_c , but a signal of duration τ with a linearly modulated frequency called chirp [Lüke92]. The phase of the emitted signal which is limited in time by $\Delta t = \tau$ is given by

$$\Phi_x(t) = 2\pi \left(f_c t + \frac{K t^2}{2} \right), \quad (2.4)$$

with the signal bandwidth $B = K\tau$. The received signal is passed through a matched filter [Lüke90] which is equivalent to a convolution with an ideal chirp. It can be demonstrated [Bamler] that the range resolution is then determined by

$$r_x = \frac{c}{2B \sin \theta_v}. \quad (2.5)$$

Compared to the previous case, this is equivalent to emitting a rectangular pulse of duration $\tau' = 1/B$. For the case of ERS which uses a bandwidth of $B = 15.5\text{MHz}$ this results in a range resolution of $r_x = 25\text{m}$ instead of $r_x = 14.2\text{km}$.

2.1.4 Synthetic Aperture

A frequency modulation similar to the one of the pulse compression in range can be obtained by using the Doppler effect in the azimuth direction, which again allows to improve the resolution. Flying at speed v , the sensor travels the distance $\Delta y = v\Delta t$ within time Δt . Furthermore, we assume that at time $t = 0$ the platform is at $y = 0$, where the radial distance R_0 to a given target is minimal. Since the traveled distance Δy is small compared to R_0 , we can approximate the distance $R(t)$ to the target as a function of time $t = y/v$ by

$$R(t) \approx R_0 + \frac{v^2 t^2}{2R_0}, \quad (2.6)$$

which results in a phase shift of

$$\Phi_y(t) = 2\pi f_c \frac{2R(t)}{c} = \frac{2\pi}{\lambda} \left(2R_0 + \frac{v^2 t^2}{R_0} \right) \quad (2.7)$$

between emitted and received signal.

This can be interpreted as another linear modulation of the frequency as a function of time t and is processed in the same way as the pulse compression by employing a matched filter. The azimuthal resolution is now derived using this equivalence and the fact that a given target stays within the antenna beam of width $R_0\beta_y$ on ground for the time $\Delta t = R_0\beta_y/v$, as illustrated in Fig. 2.3. In this way, the platform movement simulates a larger antenna aperture, also called synthetic aperture, with increased resolution of

$$r_y = \frac{D}{2}. \quad (2.8)$$

Consequently, the synthetic aperture neither depends on the wavelength, nor on the distance to the target, which is of particular interest for spaceborne radar systems. In contrast to $r_y = 4.5\text{km}$, now a resolution of $r_y = 5\text{m}$ is achieved for the ERS satellites with $D = 10\text{m}$.

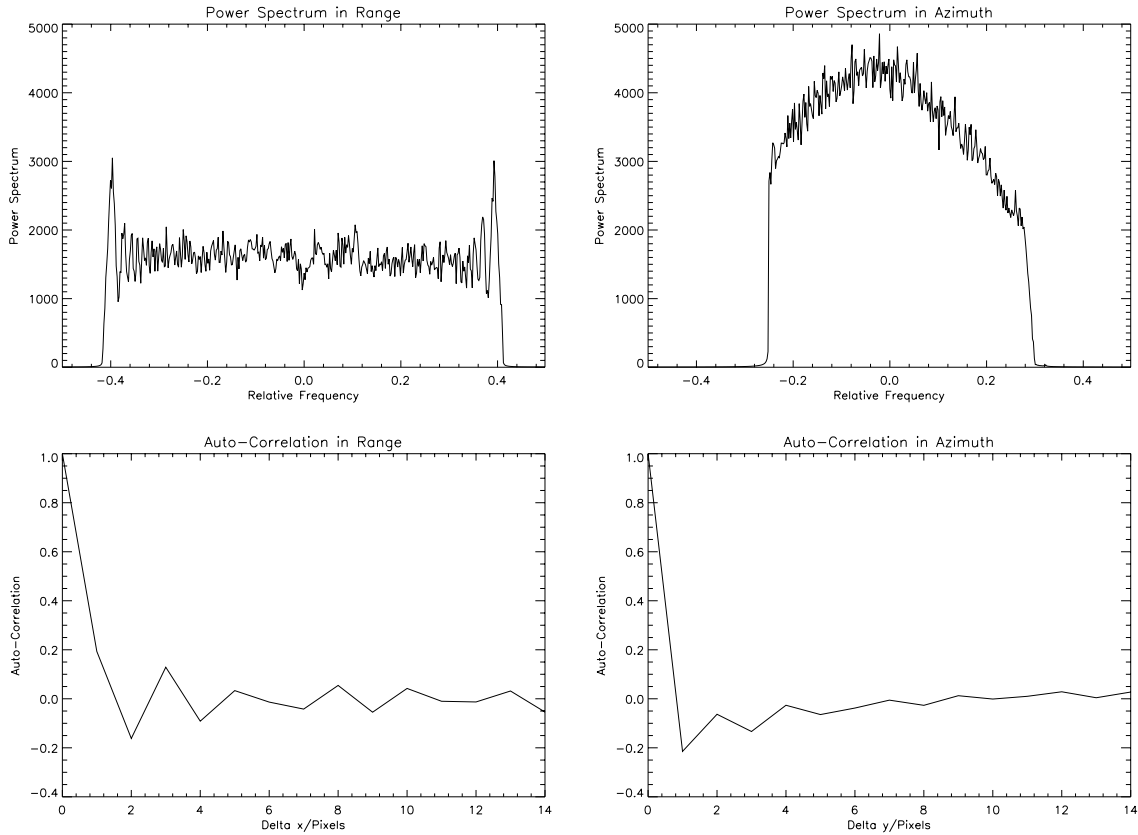


Figure 2.4: Spectral and correlation properties of complex SAR data. Clockwise: Incoherently averaged power spectrum in range direction, incoherently averaged power spectrum in azimuth direction, corresponding auto-correlation functions in azimuth and range. The plots are derived from real data explaining the skewness of the power spectra, in contrast to theoretically expected results.

We only shortly mention that several advanced signal processing techniques are needed to reach this theoretical resolution. Methods to compensate effects like range migration of the target and a good focusing are necessary. This especially applies for airborne SAR systems, since altitude and speed of the platform are not perfectly constant.

2.1.5 Spectral Properties of SAR Images

In practice, the matched filters used for SAR processing in range and azimuth are weighted by windowing functions, like Hamming windows, in order to attenuate secondary side-lobes in the spatial domain. The incoherently averaged power spectra in range and azimuth for a single look complex SAR image are displayed in Fig. 2.4. In addition to the used weighting function, the form of the azimuth spectrum also depends on the antenna gain. We note that the spectra do not extend over the full bandwidth, which is equivalent to a slight

oversampling of the signal. Thus, without losing information, the signal can be resampled to its bandwidth, reducing the amount of data to be stored.

The non-whiteness of the displayed spectra implies a spatial correlation between neighboring image pixels [Raney], which is introduced by the point-spread-function of the system, i.e. by the SAR processing. The resulting auto-correlation functions in both directions are also displayed in Fig. 2.4. It can be seen that the spatial correlation extends only to a few pixels and quickly diminishes.

The phenomenon of non-white spectra is highly disturbing for SAR image interpretation and filtering techniques: It is either difficult to take into account or considerably slows down computations. This becomes even more disturbing if the data are further oversampled. A simple solution to this problem consists in subsampling the data to their bandwidth or even more in order to reduce the effect of the system's point-spread-function.

2.2 Radiometrical and Geometrical Effects

In this section, we shortly review the radiometrical and geometrical characteristics of SAR images focusing on topographically induced distortions. The basic geometry that causes SAR layover and shadows is explained and the resulting effects on the radiometrical information are discussed. We also comment on terrain-induced geometrical distortions. Although the full knowledge of these effects is not crucial for the main part of this thesis, we want at least to hint at them, since they are of interest for the general understanding of SAR images.

2.2.1 Foreshortening, Layover and Shadows in SAR Imagery

As opposed to optical imaging devices, SAR systems actually detect the distance between the sensor emitting a microwave pulse and the target reflecting the energy back to the receiving antenna. This range measurement principle leads to specific geometric distortions in the processed SAR image, which make SAR images more difficult to interpret than optical images, especially for unexperienced users. The effects that will be described here mainly occur in mountainous regions, i.e. regions of rough terrain. The imaging of flat, e.g. agricultural, areas is not affected by this kind of degradation.

The most relevant SAR specific topographically induced distortions are illustrated in Fig. 2.5 and can be summarized as follows [Gelautz]:

- **Foreshortening:** For slopes facing the sensor, the area on the ground that is mapped onto one SAR resolution cell is larger than in the case of flat terrain. This so-called foreshortening situation has two consequences: 1) Due to the change in ground resolution, foreshortening areas appear compressed in SAR images, i.e. their extension in range direction is reduced. 2) Foreshortening areas are characterized by brighter image gray values. The received energy within a resolution cell is higher due to the larger imaged area.

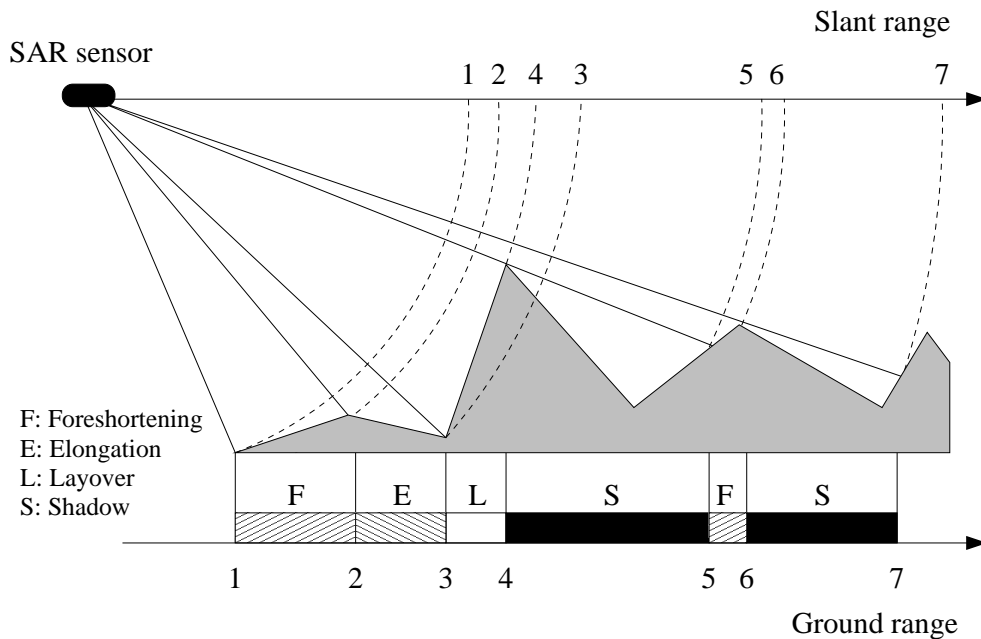


Figure 2.5: Slant range illustration of topographically induced radiometrical and geometrical distortions. Both the received energy at the antenna and the slant range resolution are affected by the imaged terrain. The different resolutions in slant and ground range are visible. For a proper interpretation of SAR data of rough terrain these effects have to be compensated.

- **Layover:** Layover is an extreme case of foreshortening and occurs where, due to steep terrain slopes, the top of a mountain is closer to the sensor than its bottom. Layover areas appear as particularly bright regions in the image with an inverted geometrical order. In Fig. 2.5, this is depicted by points three and four: Note the reversed position of these points in the slant range image compared to their actual location in ground coordinates.
- **Elongation:** In contrast to the foreshortening effect, slopes facing away from the sensor lead to rather dark, elongated regions in SAR images. The explanation for this behavior is analogous to the one of foreshortening.
- **Shadows:** Similarly to optical images, areas, which are not illuminated by the radar beam, are called radar shadows. In the image, shadow areas appear as dark regions corrupted only by thermal noise.

We note that the topography of the sensed area not only affects the radiometry of the data (bright layover and foreshortening areas, dark shadow areas), but also their geometry by changing the image resolution, called slant range resolution, in comparison with the ground range resolution. In practical SAR systems the slant range resolution is constant yielding a varying ground range resolution which depends on the topography, i.e. the terrain slopes.

2.2.2 Radiometric Correction

For radiometric correction of SAR layover areas, the energy accumulated in one image pixel has to be redistributed among those ground resolution cells, which are mapped onto that particular pixel. This problem of energy redistribution is also encountered in foreshortening areas, where each image pixel contains the energy from a larger area, comprising generally more than one ground resolution cell. Concerned terrain facets on a digital elevation model (DEM) are adjacent to each other and can be approximated to have identical local incidence angles and reflectivity. Basically, the radiometric correction [Pairman, Ulander] can therefore be performed by a division by the size of the imaged area, which, as a first approximation, is a function of the local incidence angle. In general, the local incidence angle is calculated by using DEMs of similar or, if available, higher resolution.

2.2.3 Geometric Correction

Geometrically induced effects have to be corrected as well: The slant range image must be resampled to its nominal ground resolution. This procedure of minimizing geometrical distortions and resampling the image to a homogeneous, predefined map grid, e.g. to the Universal Transversal Mercator (UTM) grid, is called geocoding [Schreier, Small]. To perform this operation the exact elevation of each pixel must be known which is usually taken from a DEM. It is clear that the success of accurate geocoding, as for radiometric correction, depends on the DEM quality and its resolution. The purpose of geocoding is to generate a map-like representation of the satellite image, where the SAR image is aligned with a Cartesian map projection grid. Otherwise, uncorrected data can hardly be interpreted due to effects in slant range images, such as e.g. mountains which appear to be leaning towards the sensor.

2.3 SAR Image Statistics

SAR images are subject to a phenomenon called speckle that affects all coherent imaging systems and, therefore, can be observed in laser [Goodman], acoustic and radar images. Basically, this usually disturbing effect is caused by random interferences, either constructive or destructive, between the electromagnetic waves which are reflected from different scatterers present in the imaged area. Comparing SAR images to optical data, a clear difference becomes visible: In contrast to incoherently imaged scenes, SAR images appear to be affected by a granular and rather strong noise named speckle. This effect is well-known since many years and has been studied intensively in laser imaging. Speckle becomes visible only in the detected amplitude or intensity signal. The complex signal by itself is distorted by thermal noise and signal processing induced effects only. As a consequence of the speckle phenomenon, the interpretation of detected SAR images is highly disturbed and cannot be done with standard tools developed for non-coherent imagery.

In the following, we shortly outline the causes for speckle noise and present its statistical properties, i.e. the SAR image statistics, which are necessary for understanding SAR imagery and are the basis for a further statistical SAR image analysis.

2.3.1 Coherent Imaging

Since the microscopic structures of the imaged reflecting surface, which cause interferences leading to speckle, are unknown, the image characteristics have to be described statistically. Within a rough, i.e. not specular, imaged area we assume an ensemble of scatterers possessing identical macroscopic but varying microscopic properties. In this case, the backscattered field E_i of one of these microscopic objects resulting from an incoming fully polarized wave is described by

$$E_i = E_{i0} \exp(j(\omega t - 2kR_i + \zeta_i)) = E_{i0} \exp(j\phi_i), \quad (2.9)$$

where E_{i0} is the magnitude of scatterer i , k is the wave number, ω denotes the corresponding angular frequency, R_i is the distance between the antenna and scatterer i and ζ_i is a phase shift. The resulting received complex wave field E can now be interpreted as the sum of all N individual scatterers i with different magnitudes E_{i0} and phases ϕ_i :

$$E = \sum_{i=1}^N E_{i0} \exp(j\phi_i) = |E| \exp(j\phi) \quad (2.10)$$

This process known as random walk is illustrated in Fig. 2.6. The recorded signal can be interpreted as a superposition of the reflections of a large number of individual scatterers with different uniformly distributed phases and different statistically independent magnitudes.

2.3.2 Statistical Properties of the Complex Signal

By employing the random walk model, we are able to derive several properties of the received signal. Before presenting the resulting probability density functions of the backscattered wave E , we shortly enumerate the necessary assumptions under which they are found:

- Magnitude and phase of the scatterers are statistically independent, allowing to obtain the received signal by a simple summation of the individual contributions. Interactions between scatterers are neglected.
- The phase of the scatterers is uniformly distributed between 0 and 2π , i.e. speckle is assumed to be fully developed [Tur]. This assumption does not apply for reflections from specular scatterers.
- The number of individual scatterers N is high, implying real and imaginary parts of E to be Gaussian distributed with zero mean and to be statistically independent.
- The resulting amplitude $|E|$ must be different from any other single contributing amplitude $|E_i|$.

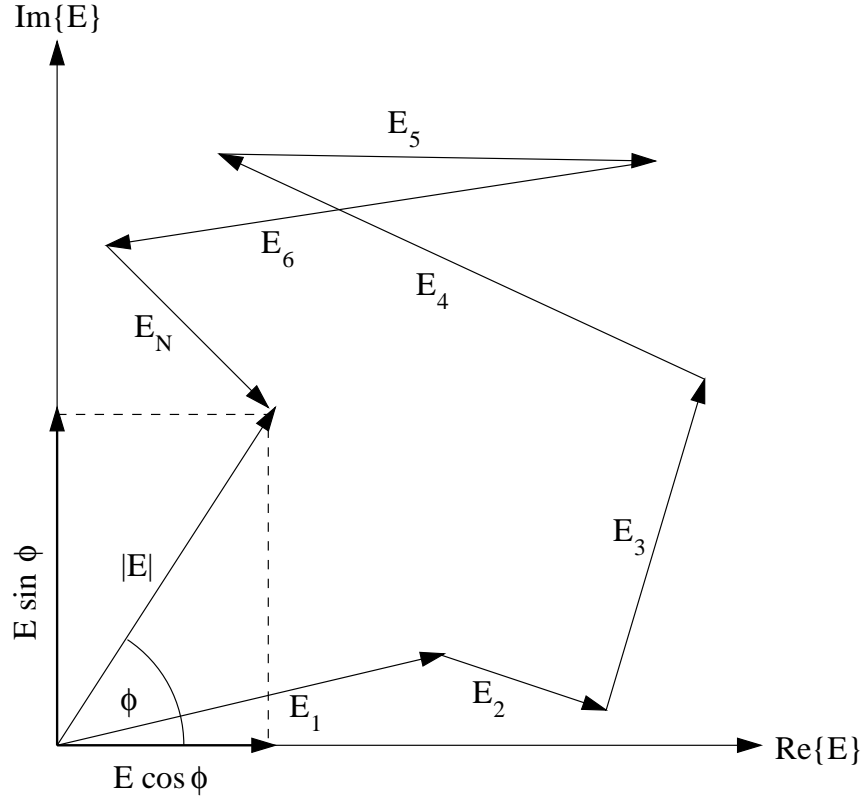


Figure 2.6: Random walk in the complex plane. The complex wave E is composed of a superposition of N contributing backscattered waves with random phases. Statistically, real and imaginary parts of E have zero mean, are Gaussian distributed and determine the magnitude of E by means of their variance.

Under these assumptions, we find for the first and second order moments of the real and imaginary parts of E :

$$E\{\text{Re}(E)\} = \sum_{i=1}^N E\{E_{i0}\}E\{\cos \phi_i\} = 0, \quad (2.11)$$

$$E\{\text{Im}(E)\} = \sum_{i=1}^N E\{E_{i0}\}E\{\sin \phi_i\} = 0, \quad (2.12)$$

$$E\{\text{Re}(E)^2\} = \sum_{i=1}^N \sum_{j=1}^N E\{E_{i0}E_{j0}\}E\{\cos \phi_i \cos \phi_j\} = \frac{N}{2}E\{E_{i0}^2\} = s^2 \quad (2.13)$$

and

$$E\{\text{Im}(E)^2\} = \sum_{i=1}^N \sum_{j=1}^N E\{E_{i0}E_{j0}\}E\{\sin \phi_i \sin \phi_j\} = \frac{N}{2}E\{E_{i0}^2\} = s^2. \quad (2.14)$$

Real and imaginary parts can be shown to be uncorrelated by computing

$$E\{\operatorname{Re}(E)\operatorname{Im}(E)\} = \sum_{i=1}^N \sum_{j=1}^N E\{E_{i0}E_{j0}\}E\{\cos\phi_i \sin\phi_j\} = E\{\operatorname{Re}(E)\}E\{\operatorname{Im}(E)\} = 0 \quad (2.15)$$

and obey a Gaussian distribution with $\mathcal{N}(0, s^2)$ according to the central limit theorem, since the number of scatterers N is assumed to be large. We remark that $2s^2$ is equivalent to the so-called radar reflectivity of the considered pixel. Neglecting sensor noise, this quantity is proportional to the radar backscatter coefficient or cross-section. The theoretical and experimentally measured distributions for the real and imaginary parts of E , as well as for the following distributions derived below, are displayed in Fig. 2.7.

Distributions of amplitude and phase

Using these assumptions, it is straightforward to derive the probability density functions (pdfs) for amplitude and phase of E :

$$p_{|E|}(E_0) = \frac{E_0}{s^2} \exp\left(-\frac{E_0^2}{2s^2}\right) \quad (2.16)$$

and

$$p_{\phi}(\phi_0) = \frac{1}{2\pi}, \quad (2.17)$$

for $E_0 \geq 0$ and $\phi_0 \in (-\pi, \pi]$ with s the standard deviation of the real and imaginary parts. The first two moments of this resulting Rayleigh distribution for the amplitude are given by

$$E\{|E|\} = \sqrt{\frac{\pi}{2}}s, \quad E\{|E|^2\} = 2s^2 \quad (2.18)$$

yielding for the variance

$$\operatorname{Var}(|E|) = \sigma_{|E|}^2 = \left(2 - \frac{\pi}{2}\right)s^2. \quad (2.19)$$

Distribution of the intensity

In the radar community, it is customary to consider not the amplitude but the measured intensity P of the received wave. After having derived $p_{|E|}$ in Eq. 2.16, we set $P = |E|^2$ in order to find the χ^2 -distribution with two degrees of freedom, i.e. the exponential distribution

$$p_P(P_0) = \frac{1}{2s^2} \exp\left(-\frac{P_0}{2s^2}\right), \quad (2.20)$$

for $P_0 \geq 0$ with its moments

$$E\{P\} = 2s^2, \quad E\{P^2\} = 2 \cdot E\{P\}^2 = 8s^4 \quad (2.21)$$

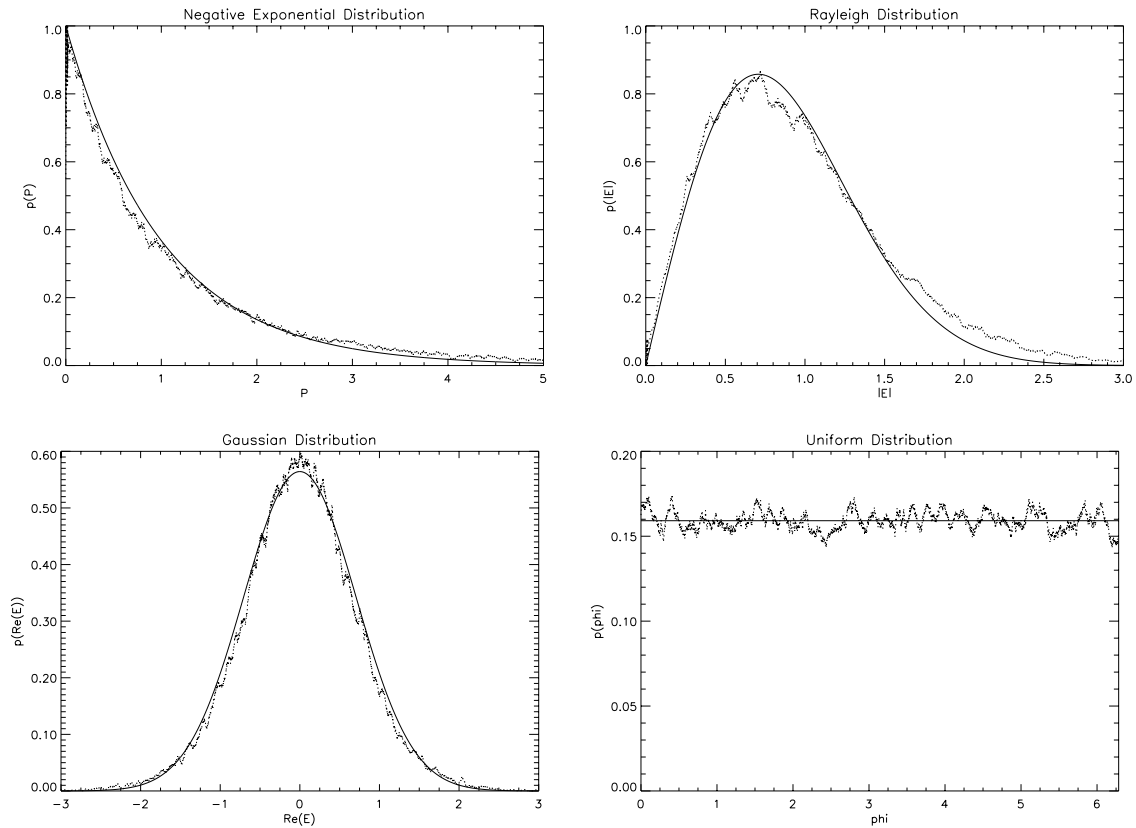


Figure 2.7: Analytically and experimentally derived SAR image distributions. Clockwise: Theoretical distributions (solid) for intensity, amplitude, phase and real/imaginary part of a single-look SAR image for $\mu_I = \mu_S = 2s^2 = 1$. These distributions are in good accordance with normalized histograms (dotted) of small stationary parts of the images shown in Fig. 2.8.

and the variance

$$\text{Var}(P) = \sigma_P^2 = E\{P\}^2 = 4s^4. \quad (2.22)$$

We note that for the exponential distribution, standard deviation and mean of the reflected intensity are equal. For a widely used measure in the SAR community, known as coefficient of variation CV_P , which is the ratio of these two values, this results in

$$CV_P^2 = \frac{\text{Var}(P)}{E\{P\}^2} = 1 \quad (2.23)$$

showing the high statistical uncertainty of the measured intensity. This uncertainty is directly reflected in the noisy appearance of detected SAR images and is interpreted as speckle noise. The measure given here is also known as noise coefficient of variation when the radar reflectivity is assumed to be uniform, i.e. variations are caused by speckle alone. It is clear that fluctuations of the reflectivity, not yet considered here, also influence the coefficient of variation.

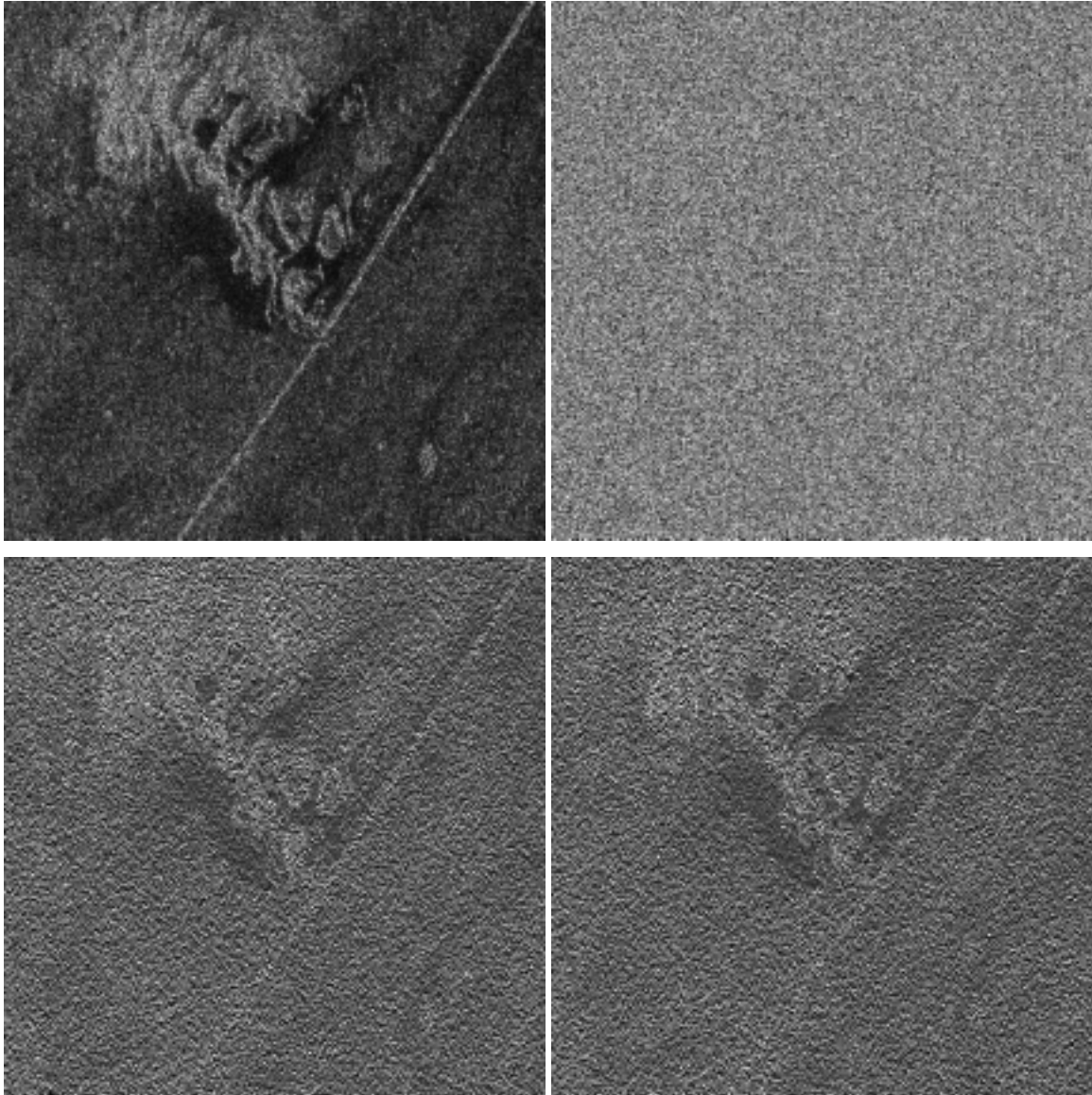


Figure 2.8: Complex X-SAR scene (512×512 pixels). Clockwise: Amplitude, phase, imaginary and real part of the complex data. The effect of speckle is clearly visible throughout the whole amplitude image. The uniformly distributed phase does not contain any useful information, unless used for interferometric purposes.

To summarize, imaging a slightly rough area of uniform backscatter results in an unpredictable, only statistically describable value of $|E|$ varying around a certain mean due to different phase shifts of the contributing scatterers. Examples for the measured signal in its different representations and the corresponding pdfs are illustrated in Figs. 2.7 and 2.8. The intensity image not shown here is similar to the amplitude image but possesses a much higher dynamic range. We want to emphasize the fact that all these relations only hold in

the case of fully developed speckle [Tur]. This implies that all scatterers contribute in an equal way to the reflected energy. In the case of one or more dominant scatterers within the imaged area the presented results are at best of limited use.

2.3.3 Statistical Properties of Multi-Look Data

SAR images are often processed in a way that reduces the speckle variance, i.e. the uncertainty of the measured data, and, as a consequence, their noisy appearance. This is called multi-looking and can either be performed in the frequency domain by incoherently averaging different parts of the spectrum (looks) or, in the spatial domain, by averaging neighboring pixels. The variance of speckle is reduced by L , the number of incoherently summed looks or pixels, i.e. $Var(P) \propto 1/L$. This increase in radiometric resolution is gained at the expense of spatial resolution, which is reduced by the same factor. So, a compromise between spatial and radiometric resolution has to be found. Among standard spaceborne SAR image products, $L = 3$ is an often used number.

Basically, three different representations of multi-look images exist depending on the application and the kind of processing applied to the data. Two additional transformations sometimes prove to be useful, either to reduce the high dynamics of SAR images or to change the statistical characteristics of the data. The most important pdfs of speckle are displayed in Fig. 2.9.

Intensity images

Multi-look intensity images are calculated by averaging L uncorrelated pixels with power P_i . As mentioned, this modifies the statistical properties of the resulting signal by reducing its variance. For L -look data it can be shown that the resulting observed intensity

$$I = \frac{1}{L} \sum_{i=1}^L P_i \quad (2.24)$$

is Gamma distributed with order parameter L and $I_0 \geq 0$ according to

$$p_I(I_0) = \frac{L^L I_0^{L-1}}{\Gamma(L) s^{2L}} \exp\left(-\frac{L I_0}{2s^2}\right). \quad (2.25)$$

Its mean is given by

$$E\{I\} = 2s^2 \quad (2.26)$$

and its variance can be shown to be

$$Var(I) = \sigma_I^2 = \frac{4s^4}{L}. \quad (2.27)$$

In order to bring this distribution into a more suitable form for Bayesian data interpretation, we perform a change of variables and set $E\{I\} = 2s^2 = \mu_I$. Hence, we obtain the conditional probability density function, also known as likelihood function of I given its mean value μ_I

$$p_I(I_0|\mu_I) = \frac{L^L I_0^{L-1}}{\mu_I^L \Gamma(L)} \exp\left(-\frac{L I_0}{\mu_I}\right) \quad (2.28)$$

with conditional mean

$$E\{I|\mu_I\} = \mu_I \quad (2.29)$$

and variance

$$Var(I|\mu_I) = \sigma_I^2 = \frac{\mu_I^2}{L}. \quad (2.30)$$

Calculating the coefficient of variation for intensity images yields

$$CV_I^2 = \frac{Var(I|\mu_I)}{\mu_I^2} = \frac{1}{L}. \quad (2.31)$$

An interesting property of this pdf around a given mean value μ_I is its interpretation as a multiplicative noise. It can be seen that the distribution of I with mean μ_I is identical to the one obtained by multiplying a fixed cross-section μ_I with a noise process n_I that is distributed according to $p_I(I_0|\mu_I = 1)$:

$$I = \mu_I \cdot n_I = \mu_I \cdot I_{\mu=1} \quad (2.32)$$

Due to this property, speckle is considered to be a multiplicative noise with

$$p_{n_I}(I_0) = p_I(I_0|\mu_I = 1), \quad (2.33)$$

which manifests itself only in the amplitude or intensity signal.

Square-root intensity images

Multi-look square-root intensity images are calculated by averaging L uncorrelated pixels with power P_i and taking the square-root, which is simply equivalent to taking the square-root of a multi-look intensity image. Therefore, the effect of reducing the variance is equivalent.

$$S = \sqrt{\frac{1}{L} \sum_{i=1}^L P_i} \quad (2.34)$$

The corresponding distribution of S is easily derived from the preceding equations by a change of variables with

$$S = \sqrt{I} \quad \text{and} \quad \mu_S = \sqrt{\mu_I} \quad (2.35)$$

and results in the conditional square-root Gamma distribution for $S_0 \geq 0$

$$p_S(S_0|\mu_S) = 2 \frac{S_0^{2L-1} L^L}{\mu_S^{2L} \Gamma(L)} \exp\left(-L \frac{S_0^2}{\mu_S^2}\right), \quad (2.36)$$

which is the Rayleigh distribution for $L = 1$ and $\mu_S^2 = 2s^2 = \mu_I$. The conditional mean and variance are found to be

$$E\{S|\mu_S\} = \frac{\Gamma(L + 0.5)}{\Gamma(L)\sqrt{L}} \mu_S \neq \mu_S \quad (2.37)$$

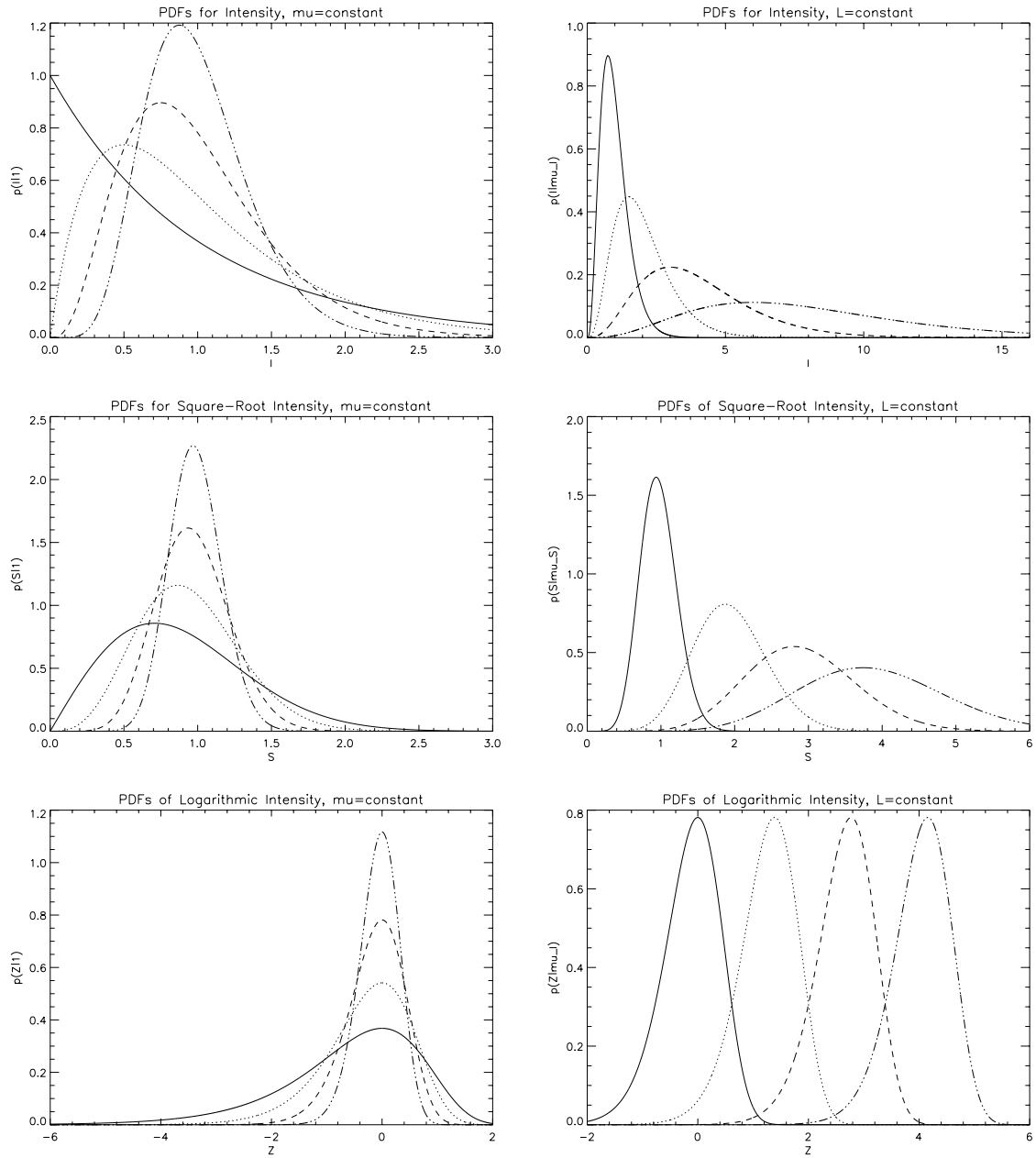


Figure 2.9: SAR image pdfs for different values of L and different mean values. Left: Pdfs for intensity (top), square-root intensity (center) and log-intensity (bottom) for $L = 1$ (solid), $L = 2$ (dotted), $L = 4$ (dashed) and $L = 8$ (dot-dashed) with $\mu_I = \mu_S = 1$, $\mu_Z = \log 1$. Right: Pdfs for intensity (top), square-root intensity (center) and log-intensity (bottom) for $\mu_I = \mu_S = 1$, $\mu_Z = \log 1$ (solid), $\mu_I = \mu_S = 2$, $\mu_Z = \log 4$ (dotted), $\mu_I = \mu_S = 4$, $\mu_Z = \log 16$ (dashed) and $\mu_I = \mu_S = 8$, $\mu_Z = \log 64$ (dot-dashed) with $L = 4$.

and

$$\text{Var}(S|\mu_S) = \sigma_S^2 = \mu_S^2 - E\{S|\mu_S\}^2 = \mu_S^2 \left(1 - \frac{\Gamma(L+0.5)^2}{\Gamma(L)^2 L}\right). \quad (2.38)$$

This yields for the coefficient of variation

$$CV_S^2 = \frac{\sigma_S^2}{E\{S|\mu_S\}^2} = \frac{\Gamma(L)^2 L}{\Gamma(L+0.5)^2} - 1. \quad (2.39)$$

By analogy to the intensity case considered before, this can be interpreted again as a multiplicative noise n_S affecting the mean value μ_S with

$$S = \sqrt{I} = \sqrt{\mu_I \cdot n_I} = \mu_S \cdot n_S = \mu_S \cdot S_{\mu_S=1} \quad (2.40)$$

and

$$p_{n_S}(S_0) = p_S(S_0|\mu_S = 1). \quad (2.41)$$

Amplitude images

Another means to reduce speckle variance and to take advantage of a reduced dynamic range is the calculation of multi-look amplitude images. In contrast to square-root images, the resulting distribution cannot be derived in a closed form. Since the multi-looked amplitude is computed by

$$A = \frac{1}{L} \sum_{i=1}^L \sqrt{P_i} \quad (2.42)$$

the pdf of A is determined by a convolution of L Rayleigh distributions.

As a consequence, images of this kind are seldom used, statistical conclusions are difficult to be drawn, and accurate interpretation cannot be easily done. However, amplitude images exhibit one nice property: In contrast to square-root images, they easily permit the computation of L out of the measured coefficient of variation CV_A , which is given by

$$CV_A^2 = \frac{\sigma_A^2}{\mu_A^2} = \frac{0.5227^2}{L}. \quad (2.43)$$

By σ_A we denote the standard deviation and by μ_A the mean of A . Since the coefficients of variation for amplitude and square-root images are almost identical as shown in Fig. 2.10, i.e. the ratio of standard deviation to mean is rather similar, this allows for square-root intensity images to analytically calculate L from an analysis of mean and variance without having to deal with the Gamma functions of Eq. 2.39. Of course, it is also possible to directly estimate L from the intensity image.

Log-intensity images

As we already pointed out, speckle found in SAR intensity or amplitude images can be regarded as a multiplicative process. However, many image processing techniques assume

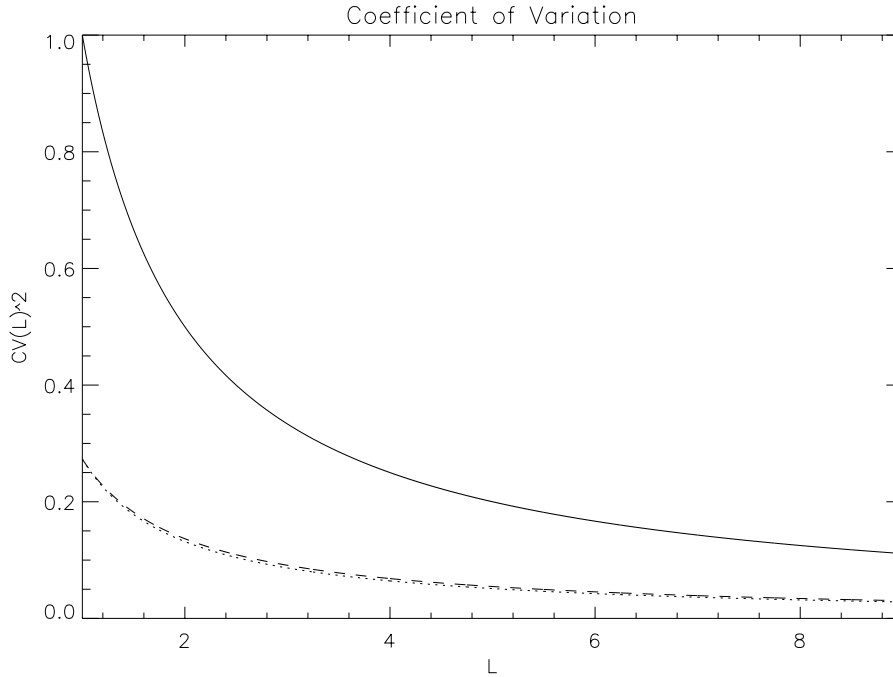


Figure 2.10: Coefficient of variation for different number of looks L . CV^2 is shown for intensity data (solid), for square-root intensity data (dotted) and amplitude data (dashed). The difference between the two latter is negligible.

that an image is affected only by additive noise and perform poorly otherwise. This problem, which hinders the analysis of SAR images containing multiplicative speckle, can be partly overcome by taking the logarithm of the image prior to interpretation. The log-operation transforms the speckle to being additive [Arsenault, Hoekman]:

$$Z = \log I = \log \mu_I + \log n_I = \mu_Z + n_Z \quad (2.44)$$

with its corresponding pdf resulting in

$$p_Z(Z_0|\mu_Z) = \frac{\exp\left(L\left(Z_0 - \mu_Z + \log L\right) - \exp\left(Z_0 - \mu_Z + \log L\right)\right)}{\Gamma(L)}, \quad (2.45)$$

which is the Fisher-Tippett distribution for $L = 1$, with the following properties for mean, variance and coefficient of variation [Caves]:

$$E\{Z|\mu_Z\} = \mu_Z - \log L + \psi(L), \quad (2.46)$$

$$\text{Var}(Z|\mu_Z) = \psi'(L) \quad (2.47)$$

and

$$CV_Z^2 = \frac{\psi'(L)}{\left(\mu_Z - \log L + \psi(L)\right)^2}, \quad (2.48)$$

where $\psi(L)$ is the Digamma function, $\psi'(L)$ its first derivative and $\mu_Z = \log \mu_I$. Because the disturbing noise is now additive, the shape of the distribution is the same regardless of its mean value. The mean μ_Z only causes a shift of the distribution, a characteristic property of additive noise.

Sometimes it is customary to represent the SAR data in decibels (dB). The image pixel values are then transformed by

$$Z_{dB} = 10 \log_{10} I. \quad (2.49)$$

This does not give any advantages for a machine-based interpretation of the image but is useful for a more physical analysis of the data.

Log-amplitude images

Obviously, speckle can also be transformed to being additive by taking the logarithm of the square-root intensity. However, this does not give any new insight, since this results in a simple dilatation of the distribution compared to the log-intensity case. With

$$Z_S = \log S = \frac{1}{2} \log I = \frac{1}{2} Z, \quad (2.50)$$

one obtains

$$p_{Z_S}(Z_0 | \mu_{Z_S}) = 2p_Z(2Z_0 | \mu_{Z_S}). \quad (2.51)$$

Interpretation of the derived speckle distributions

We have presented the SAR image statistics under the assumption of a uniform reflectivity. The resulting distributions, that are illustrated in Fig. 2.9, confirm the rather noisy visual appearance of detected images. The measured data are distributed around their mean value in a non-Gaussian way, and are either Gamma, square-root Gamma or Fisher-Tippett-like distributed, depending on their representation.

If the logarithm is not applied, the observed noise called speckle shows a multiplicative behavior, significantly reducing the performance of commonly used image analysis and interpretation tools. To circumvent this problem, SAR data may be transformed to the log-space making the noise additive. However, the dynamics of the data are significantly and non-linearly modified as shown in the lower right plot of Fig. 2.9, where the almost equally spaced distributions are given for mean values μ_I of 1, 4, 16 and 64.

Another drawback for SAR image interpretation is the fact that the data do not possess Gaussian properties. This especially applies for low values of L , which are desirable to preserve spatial resolution. If data are multi-looked a good number of times, e.g. $L \approx 8$, the resulting distributions in the log-space can be well approximated by a Gaussian for both intensity and amplitude, as illustrated. In this case, the log-signal is affected by an approximately Gaussian distributed additive noise.

On the contrary, it is highly desirable to preserve spatial resolution and the high dynamics of the data. In that case, special techniques are necessary to deal with the strong multiplicative non-Gaussian speckle noise.

2.3.4 Correlation Properties of Speckle

Based on the first order statistics with mean and variance, several statistical properties of speckle assuming a homogeneous reflectivity have been derived. Nonetheless, this does not explain the granular visual appearance of spatial structures in SAR images which is due to the correlation of speckle. For that reason, we now consider in more detail the second order statistics of speckle taking into account the correlation properties between neighboring pixels.

Assuming a stationary process and statistical independence between the mean intensity μ_I and the speckle noise n_I we find for the two-dimensional auto-correlation function of I

$$\begin{aligned} R_I(\Delta x, \Delta y) &= E\{I(x, y) \cdot I(x + \Delta x, y + \Delta y)\} \\ &= E\{\mu_I(x, y)n_I(x, y) \cdot \mu_I(x + \Delta x, y + \Delta y)n_I(x + \Delta x, y + \Delta y)\} \\ &= \mu_I^2 \cdot R_{n_I}(\Delta x, \Delta y), \end{aligned} \quad (2.52)$$

where $R_{n_I}(\Delta x, \Delta y)$ is the auto-correlation function for speckle, which is given by

$$R_{n_I}(\Delta x, \Delta y) = 1 + \frac{1}{L} \operatorname{sinc}^2\left(\frac{\Delta x}{r_x}\right) \operatorname{sinc}^2\left(\frac{\Delta y}{r_y}\right). \quad (2.53)$$

Here r_x and r_y denote the system's resolution in the x and y direction, i.e. in range and azimuth. Hence, this yields the auto-covariance function as qualitatively illustrated in Fig. 2.11

$$C_{n_I}(\Delta x, \Delta y) = \frac{1}{L} \operatorname{sinc}^2\left(\frac{\Delta x}{r_x}\right) \operatorname{sinc}^2\left(\frac{\Delta y}{r_y}\right). \quad (2.54)$$

This correlation, which results from the SAR processing, i.e. the system's point-spread-function, is responsible for the granular appearance of speckle. We also see that C_{n_I} converges towards zero with L growing, i.e. multi-looking reduces the speckle correlation. In fact, speckle correlation is an important problem for automated image interpretation. Since it is usually not considered, which is often the only means to keep algorithms tractable, some useful information is lost.

Besides multi-looking the data in the spatial or the frequency domain, several other approaches for reducing the speckle correlation are known [Oliver98, Quegan]. One example are interpolation techniques in the Fourier domain. Another often used method consists in subsampling the data, which reduces the correlation much faster compared to multi-looking, i.e. with less loss of resolution, but, on the contrary, does not increase the radiometric resolution.

2.3.5 Statistical Properties of SAR Data with Non-Uniform Cross-Section

Until now, we have only considered the case of a uniform cross-section to determine the statistics of SAR images and the statistical properties of the speckle phenomenon. The statistics of real images, however, do not obey the above derived distributions, since either speckle is not fully developed, as in urban areas [Tur], or the underlying cross-section is not

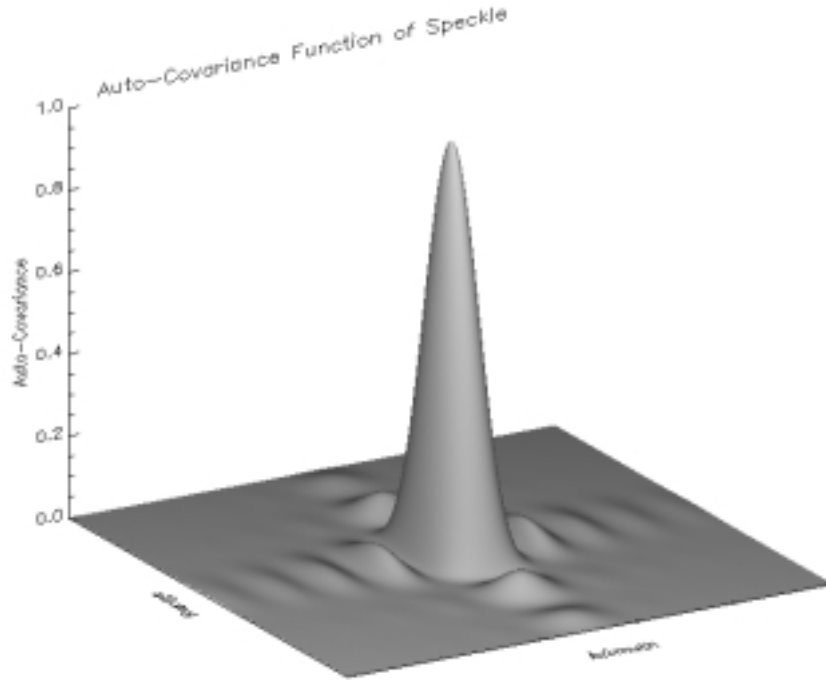


Figure 2.11: Qualitative illustration of the speckle auto-covariance function with different resolutions in range and azimuth.

uniform but shows some textural activity. It is clear, that for the latter case the resulting SAR image distribution depends both on the statistical properties of the backscatter itself and on those of speckle. The following description of the cross-section properties already belongs to scene-modeling, as we use prior assumptions for the noise-free scene, which will be intensively discussed in the following chapters.

Many areas of natural clutter in SAR images have been shown to have a cross-section σ that can be described by a Gamma distribution [Oliver91]

$$p_{\sigma}(\sigma_0|\mu_{\sigma}) = \frac{\nu^{\nu} \sigma_0^{\nu-1}}{\mu_{\sigma}^{\nu} \Gamma(\nu)} \exp\left(-\frac{\nu \sigma_0}{\mu_{\sigma}}\right) \quad (2.55)$$

for $\sigma_0 \geq 0$, where μ_{σ} is the expected mean and ν is the order parameter with $\nu = \mu_{\sigma}^2 / \text{Var}(\sigma)$. It can be demonstrated [Caves] that, together with the speckle distribution, which is also Gamma distributed according to Eq. 2.28, this results in the K-distribution for the intensity I of the observed image with $I_0 \geq 0$:

$$p_I(I_0|\mu_{\sigma}) = \frac{2}{I_0} \left(\frac{L\nu I_0}{\mu_{\sigma}}\right)^{(L+\nu)/2} \frac{1}{\Gamma(L)\Gamma(\nu)} K_{\nu-L} \left(2\sqrt{\frac{L\nu I_0}{\mu_{\sigma}}}\right) \quad (2.56)$$

Here, $K_{\nu-L}$ denotes the modified Bessel function of order $(\nu - L)$. The appearance of the Bessel function makes the derivation of exact analytical estimates for ν and μ_{σ} from

the data difficult. Nevertheless, the estimated parameters of the K-distribution can be used to discriminate between textures of different roughness [Oliver93, Oliver94, Oliver98], i.e. different variances of the underlying cross-section, within the limits of the estimation accuracy.

2.4 Summary

In this chapter, the following points have been discussed:

- SAR is a time-ranging system, which allows to generate images from backscattered intensity of electromagnetic waves. To obtain an adequate resolution, techniques like pulse compression and synthetic aperture are used.
- Radiometrical and geometrical information in SAR images are influenced by the topography of the imaged scene. For that reason, SAR data usually require geometrical correction and radiometrical calibration in combination with terrain correction to be of use in cartography, for example.
- A phenomenon called speckle highly degrades the radiometric resolution of SAR images and hinders their interpretation. Speckle is considered as a multiplicative granular noise only visible in detected SAR data.
- In general, speckle noise shows some degree of correlation, which is due to the SAR system's point-spread-function. However, this correlation is not considered in commonly used image processing techniques since it is difficult to be taken into account. Consequently, the effect is often reduced by appropriate techniques, like subsampling.
- The statistics of speckle noise have been described because they are necessary to be able to deal with this effect within a Bayesian framework. In the following, methods will be discussed that rely on a probabilistic description of the speckle noise and of the parameters to be estimated from the noisy data.

3 Bayesian Image Analysis

In the last chapter, we have discussed the statistical properties of SAR images. Why have we done this? To have a theoretically founded interpretation of data rather than a heuristical and usually inaccurate approach, it is necessary to have some knowledge about the data to be analyzed: What is the origin of the data, how are they processed, what are the characteristic effects that have to be considered? These questions have been answered in the last chapter by a statistical description of SAR images and, mainly, by illustrating the disturbing effect of speckle. Since we have the statistical description of the SAR image degradation at hand, a probabilistic data analysis allowing the extraction of different kinds of information, which may be encapsulated by stochastic or deterministic parametric models, can be performed. Here, stochastic models better adapt to the data than deterministic ones and have a better chance to extract information from unexpected data where deterministic models will fail.

The Bayesian approach allows a probabilistic description of the problem we are faced with, i.e. estimation of the SAR cross-section and parameter estimation for parametric models out of speckled data. In the first section of this chapter, we give a short review of general Bayesian data analysis and parameter estimation and, in the second section, we concentrate on its application for image processing in combination with Gibbs and Markov random fields, which serve as prior information, and discuss several of their properties.

3.1 Bayesian Estimation Theory

To introduce the problem of parameter estimation from any given data, we begin with the opposite problem. Given a certain cause, its consequences can be described by means of an existing model, either a deterministic or a stochastic one. This description is called the forward model of the data. Consequently, a cause may have a number of possible outcomes, that can be deduced if the model is known. We call this "deductive logic". In contrast to this case, most scientists are faced with the reverse problem: Given certain outcomes of an experiment, e.g a number of measurements or observations, the underlying cause must be inferred. This is usually named "inductive logic" and is a much harder problem since even for fully deterministic forward models, an inverse transformation may not exist. The filtering of a signal with an ideal low-pass is a good example: The filtering (deductive logic) yields a fully deterministic solution, but a unique solution for the inverse filtering (inductive logic) does not exist. A general illustration for induction and deduction is shown in Fig. 3.1.

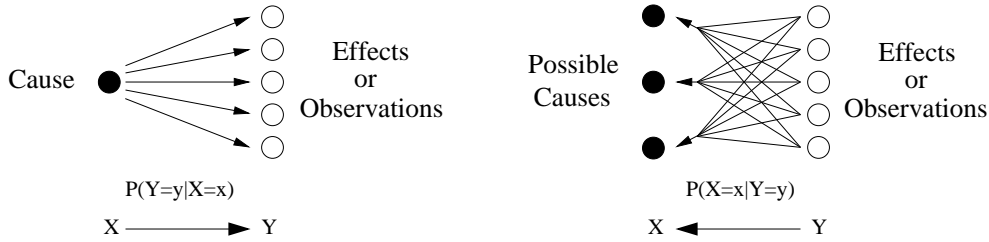


Figure 3.1: Deductive and inductive logic. Left: A cause X may have different outcomes Y depending on $P(Y = y|X = x)$. Right: The observed effects Y are due to several possible causes X each with probability $P(X = x|Y = y)$.

What can be done to determine a solution in such a case is to make the best inference, i.e. to choose the most probable cause, based on the observed data and their forward model in combination with any prior knowledge of the causes that we have or can obtain. To put it with Herodotus' words (around 500 B.C.):

"A decision was wise, even though it led to disastrous consequences, if the evidence at hand indicated it was the best one to make; and a decision was foolish, even though it led to the happiest possible consequences, if it was unreasonable to expect those consequences."

Consequently, two different models play a major role for inference: A model of the underlying cause or data and the forward model describing how the observation results from these data. The latter is usually known with sufficient but only limited accuracy, as a full description of all physical processes leading to an observation is generally impossible. The choice of the former highly influences the outcome of the inference. Therefore, it has to be chosen with great care. The forward model for our particular case of SAR images, i.e. the speckle characteristics, has been described in the preceding chapter. Possible choices of data models and their impact on the estimation will be presented in the next section and in later chapters, whereas the required theoretical tools for data analysis are described in the following.

Mathematical notation used for probability measures

In contrast to a rigorous mathematical notation, we shortly introduce a simplified mathematical description found in many publications about Markov and Gibbs random fields, which will be used throughout the rest of this thesis. We have chosen this notation for practical reasons since it allows an easier illustration of our results, although it is not always fully exact from a mathematical point of view.

Basically, the following simplifications in comparison with a fully correct notation are used in this thesis:

- We denote random variables by capital letters: X and Y are random variables.

- Sample realizations of random variables are usually denoted by small letters, e.g. x and y . However, instead of writing $X(\omega) = x$ for an assignment of the value x to a realization of X , we simplify by $X = x$.
- Instead of indexing pdfs with their corresponding random variable as in chapter two, we characterize these functions by using the random variables as arguments: We write $p(X = x)$ instead of $p_X(x)$. As a consequence, $p(X = x)$ and $p(Y = y)$ are two different functions, denoting the pdfs of the random variables X and Y , respectively. $p(X = x)$ is sometimes abbreviated with $p(x)$.
- For the sake of simplicity, we use p for both probability density and probability distribution functions. Hence, $p(X = x)$ denotes either the probability of $X(\omega) = x$ or the value of the probability density function of X at position x , according to the context.

3.1.1 Probability

Among scientists there are different conflicting interpretations of the notion of probability [Bayes, Bernoulli, Cox, Laplace]: The most popular one is surely the frequentist view, which defines the probability $P(Y = y|X = x)$ as the relative number of events Y that occur for a given cause X . In contrast to this view is Bayesian probability, which defines $P(Y = y|X = x)$ as a measure of certainty or belief that Y follows the cause X . Some people, especially frequentists, call Bayesians to be subjective, since the resulting Bayesian induction requires the choice of a prior knowledge $P(X = x)$ that encapsulates the degree of certainty of the cause X assumed by the scientist.

An example for the frequentist and the Bayesian notion of probability can be given by the probability of having a white Christmas, which can be regarded to be either frequentist, if derived from meteorological observations over the years, or Bayesian, if derived from weather conditions, say, one week earlier with the prior knowledge of the geographic location.

We do not want to continue a philosophical discussion, but state that Bayesian inference allows us to select the most probable model, thus to reject highly biased, improbable or subjective prior assumptions that have no relevance or, as it is called, evidence to the data. This property of model selection and the incorporation of prior knowledge make Bayesian estimation theory a powerful tool for inductive logic and in particular for our case of SAR image filtering and information extraction.

Logical consistency

From a mathematical point of view, the ensuing three Kolmogorov axioms must be satisfied to guarantee consistency between different conclusions drawn from probabilistic measures:

1. Positivity

$$P(X = x|I) \geq 0 \tag{3.1}$$

2. Certainty

$$P(X = x|X = x) = 1 \quad (3.2)$$

3. Sum rule

$$P(X = x_1 \cup X = x_2|I) = P(X = x_1|I) + P(X = x_2|I) \quad (3.3)$$

for $x_1 \neq x_2$

Here, the variables x , x_1 and x_2 are outcomes of a random variable X and I denotes some given background information about this process [Papoulis84]. In the following, random variables are denoted by capital letters, e.g. X and Y and the corresponding sample realizations by small letters, e.g. x and y , respectively.

Bayes' theorem

The three axioms, as pointed out above, form the basic algebra of probability theory. All following equations can be derived from them. The most important one is Bayes' [Bayes] theorem

$$P(X = x|Y = y, I) = \frac{P(Y = y|X = x, I)P(X = x|I)}{P(Y = y|I)}, \quad (3.4)$$

which describes the probability of a hypothesis X given some data Y . Since X does not depend on $P(Y = y|I)$, the probability $P(X = x|Y = y, I)$ is proportional to the probability of the observation given the hypothesis $P(Y = y|X = x, I)$ multiplied by a term encapsulating prior information, $P(X = x|I)$. The terms in Bayes' equation all have formal names: $P(Y = y|X = x, I)$ is known as the likelihood function, it describes the likelihood of the data Y given a hypothesis X . The quantity on the right of the numerator, $P(X = x|I)$, is called prior probability and, in combination with the likelihood function and the evidence $P(Y = y|I)$, yields the posterior probability $P(X = x|Y = y, I)$.

For most cases of estimation of a parameter X from an observation Y the evidence term is omitted because it acts as a normalization constant independent of X . However, as we will see later, the evidence plays a crucial role for model selection.

Probability density function

For a continuous random variable Y the relationship between its probability density function $p(Y = y|I)$ and a probability P is given by

$$P(Y < y_1|I) = \int_{-\infty}^{y_1} p(Y = y|I)dy. \quad (3.5)$$

Consequently, the probability that Y lies in the interval between y_1 and y_2 is given by

$$P(y_1 \leq Y < y_2|I) = \int_{y_1}^{y_2} p(Y = y|I)dy \quad (3.6)$$

and we find for infinitesimal small intervals

$$P(Y \in (y, y + dy]|I) = p(Y = y|I)dy. \quad (3.7)$$

As a result, to preserve uniformity between the continuous and the discrete case and since pdf may also stand for probability distribution function pertaining to a discrete set of possibilities, we will denote anything related to probabilities by p in the following.

Marginalization and normalization

Another important technique often used in probability theory is marginalization. Marginalization allows to deal with so-called nuisance parameters. These parameters, e.g. unwanted instrumental parameters, which enter the process of inference, but are of no particular interest for the estimation itself, can be integrated out:

$$p(X = x|I) = \int_{-\infty}^{+\infty} p(X = x, Y = y|I)dy \quad (3.8)$$

In the above equation, Y is interpreted as a nuisance parameter and is eliminated by integrating over its whole space.

The last important property we want to mention is that of normalization. As in the discrete case, the area under the pdf is always normalized to unity, i.e.

$$\int_{-\infty}^{+\infty} p(Y = y|X = x, I)dy = 1. \quad (3.9)$$

With this set of rules and theorems, we have enough tools at hand to address the problem of probabilistic data analysis and parameter estimation within the Bayesian framework.

3.1.2 Parameter Estimation

Given a probability distribution or probability density function both denoted by $p(X = x|Y = y)$ the task of parameter estimation is to infer a parameter value \hat{x} from an observation y . Within Bayesian inference this estimate \hat{x} is the most probable value with regard to a given distribution. Formally, it does not play a role whether induction or deduction is used here, since both X and Y can either be hypothesis or measurement. In any case, we infer a variable X conditioned by another variable Y .

To determine the most probable value of X , we have to find the value that maximizes the probability $p(X = x|Y = y)$, i.e. we search for

$$\hat{x} = \arg \max_x p(X = x|Y = y). \quad (3.10)$$

If $p(X = x|Y = y)$ is given analytically and well behaving, we can derive \hat{x} by using the following conditions that must be fulfilled at the maximum of $p(X = x|Y = y)$:

$$\left. \frac{dp(X = x|Y = y)}{dx} \right|_{x=\hat{x}} = 0 \quad (3.11)$$

and

$$\left. \frac{d^2 p(X = x|Y = y)}{dx^2} \right|_{x=\hat{x}} < 0. \quad (3.12)$$

To make analytical computations easier, the log-probability $\log p(X = x|Y = y)$ is usually used. This operation reduces the dynamics of $p(X = x|Y = y)$ while preserving its extrema, since the logarithm is a monotonic function. Furthermore, the logarithm transforms the product of two terms into a sum making it easier to deal with the posterior product $p(Y = y|X = x, I)p(X = x|I)$. Thus, the condition for the maximum of $p(X = x|Y = y)$ at $X = \hat{x}$ can be rewritten as

$$\left. \frac{d \log p(X = x|Y = y)}{dx} \right|_{x=\hat{x}} = 0, \quad (3.13)$$

where we neglect the condition for the second derivative.

Cramér-Rao bound

We remark that \hat{x} is an estimate and therefore naturally affected by an error. For practical applications, it is often necessary to evaluate the accuracy of the estimation to have a measure of trust for \hat{x} . Obviously, this measure should depend on the shape of $p(X = x|Y = y)$. To describe the estimation uncertainty, the true value x of X can be considered to lie in an interval around the estimate \hat{x} , i.e.

$$x = \hat{x} \pm \sigma. \quad (3.14)$$

For the Gaussian case, this σ , which is nothing but the standard deviation of the estimator, can be computed by

$$\sigma^2 = \left(- \left. \frac{d^2 \log p(X = x|Y = y)}{dx^2} \right|_{x=\hat{x}} \right)^{-1}. \quad (3.15)$$

The positivity of σ^2 directly follows from Eq. 3.12. The probability of the true x to lie in the interval from $\hat{x} - \sigma$ to $\hat{x} + \sigma$ can then be found to be

$$P(\hat{x} - \sigma \leq x < \hat{x} + \sigma | Y, I) = 0.67. \quad (3.16)$$

For this probability cannot be increased, the variance σ^2 of the estimation error is naturally bounded and is determined by the so-called Cramér-Rao bound [Kroschel], which is defined as

$$\sigma^2 \geq E \left\{ \left. - \frac{d^2 \log p(X = x|Y = y)}{dx^2} \right|_{x=\hat{x}} \right\}^{-1} \quad (3.17)$$

and describes the lowest possible variance of the parameter to be estimated. Note that the Cramér-Rao bound is defined for $p(X = x|Y = y)$ being the likelihood function, but similar bounds can be derived if additional prior knowledge is used [Kroschel].

To illustrate these findings, we give an instructive example introducing two basic estimators: We wish to estimate the value of an unknown parameter X that is affected by white additive Gaussian noise $\mathcal{N}(\mu, \sigma^2)$ with known mean μ and variance σ^2 . For simplicity we

assume $\mu = 0$. The resulting likelihood function for the observed random process Y is given by

$$p(Y = y|X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right). \quad (3.18)$$

Note, that in the case of Gaussian noise, as in this example, the general estimators that will be presented below correspond to the least-mean-squares (LMS) and conditional least-mean-squares (CLMS) estimators.

Maximum likelihood estimation (ML)

To infer the parameter X , we use Bayes' equation (Eq. 3.4) for inference and find for Gaussian additive noise

$$p(X = x|Y = y) \propto p(Y = y|X = x), \quad (3.19)$$

if no further knowledge about X is given, i.e. if $p(X = x)$ is taken to be a uniform distribution. By maximizing $p(X = x|Y = y)$ as a function of x we obtain

$$\hat{x}_{ML} = \arg \max_x p(Y = y|X = x) = y, \quad (3.20)$$

which is called the maximum likelihood (ML) estimate of X . For ML estimation only knowledge or assumptions about the noise, i.e. the likelihood function, is required [Kreyszig, Kroschel]. Concerning the Cramér-Rao bound of the obtained ML estimate, we find that the interval containing the true value of X with probability $P = 0.67$ is determined by the standard deviation σ of the additive noise. From $\hat{x} = y$, we see that no further knowledge about X has been gained since the best estimate of X is Y itself.

Maximum a posteriori estimation (MAP)

Let us now assume that we know, e.g. from a study of the physical causes of X , that X is Gaussian distributed, this time with $\mathcal{N}(0, \sigma_X^2)$ and that σ_X is known as well. Consequently, we can introduce this prior knowledge into our estimation problem by writing

$$p(X = x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{x^2}{2\sigma_X^2}\right). \quad (3.21)$$

The maximization of the posterior probability $p(X = x|Y = y)$ then yields

$$\hat{x}_{MAP} = \arg \max_x p(X = x|Y = y) = \frac{\sigma_X^2}{\sigma^2 + \sigma_X^2} y \neq \hat{x}_{ML} \quad (3.22)$$

and is called maximum a posteriori (MAP) estimate. The MAP estimate is not equal to the ML estimate and we notice that \hat{x}_{MAP} is drawn from the observed value y towards zero, depending on a function of the variances of the corresponding variables. This should not surprise us because the true mean value $E\{X\} = 0$ is encapsulated in the prior distribution of X .

Calculating the standard deviation, or confidence interval, of the MAP estimate \hat{x}_{MAP} from the correct value x of the process X , we find after some algebra

$$x = \hat{x}_{MAP} \pm \frac{\sigma}{\sqrt{1 + \frac{\sigma^2}{\sigma_X^2}}} = \hat{x}_{MAP} \pm \sigma_{MAP}, \quad (3.23)$$

which is lower than in the ML case. The variance of the MAP estimator decreases with smaller values of σ_X as expected, because the uncertainty about X imposed by the prior diminishes.

Comments on ML and MAP estimation

So, why should anyone use the ML estimator? In fact, the ML estimator is a MAP estimator, where, due to no available knowledge about the parameter to be estimated, a uniform prior distribution is used. We will outline later, that, in this case, a so-called non-informative prior should be employed.

The problem of MAP estimation, which is also known as the first level of Bayesian inference, and a point often subject to strong critics, is the choice of the correct prior distribution. We can deduce from the presented example that a wrong prior, say with $E\{X\} \neq 0$, will yield biased results if the true mean is indeed zero. The estimated values will be drawn from the observed values towards the mean of the employed Gaussian prior. Hence, any prior information and its related parameters must be chosen carefully in order not to pull the final estimate away from the correct solution, but, on the contrary, to bring it closer. To achieve this goal it is imperative to select the prior with regard to the data at hand. A fixed prior that is not adapted to the actually measured data $Y = y$ can always be suspected to lead to unreliable or unrealistic estimates.

3.1.3 Model Selection

The model selection process, or second level of Bayesian inference, is one of the major advantages of Bayesian estimation theory and, in our opinion, justifies the employment of prior knowledge, which is a highly disputed issue among experts. Using the same formalism as before, the probability of a certain model M given the observed process Y can be written as

$$p(M|Y = y) = \frac{p(Y = y|M)p(M)}{p(Y = y)}. \quad (3.24)$$

Usually, a prior $p(M)$ favoring different models is not considered, although this is in principle possible. Hence, we assume $p(M)$ to be uniform. By analogy to Eq. 3.19, the posterior probability is then equivalent to the likelihood function of the data Y given the model M :

$$p(M|Y = y) \propto p(Y = y|M) \quad (3.25)$$

The term $p(Y = y|M)$ already appeared in Bayes' equation as the evidence, but was neglected because it served only as a normalization constant. Now, $p(Y = y|M)$ becomes

crucial for the determination of the model quality. The evidence is obtained by marginalization, i.e. by integration over X , and results in

$$p(Y = y|M) = \int_{-\infty}^{+\infty} p(Y = y|X = x, M)p(X = x|M)dx. \quad (3.26)$$

Considering a library of different models M_1, \dots, M_N , this equation enables us to calculate the likelihood of each model and, thus, to select the most probable one in light of the measured random variable Y . Note that the likelihood $p(Y = y|X = x, M)$ can often be replaced by $p(Y = y|X = x)$ when the model is already completely encapsulated in X .

Approximation and Interpretation of the Evidence

For multi-dimensional data and already moderately complicated pdfs, an analytical integration over the posterior product $p(Y = y|X = x)p(X = x|M)$ is no more feasible. In many cases, especially with uni-modal pdfs, however, the posterior can be approximated by a Gaussian distribution centered around the MAP estimate of X , i.e. around the maximum of the posterior [MacKay]. In the one dimensional case displayed in Fig. 3.2, this approximation then yields for the evidence

$$\begin{aligned} p(Y = y|M) &\approx p(Y = y|X = \hat{x}_{MAP})p(X = \hat{x}_{MAP}|M)\sqrt{2\pi\sigma_{MAP}^2} \\ &\approx p(Y = y|X = \hat{x}_{MAP}) \cdot \Omega, \end{aligned} \quad (3.27)$$

where σ_{MAP} denotes the standard deviation of the posterior distribution and Ω is known as the Occam factor. For the case of good data with $\sigma_{MAP} \approx \sigma$, i.e. the width of the prior is much larger than the uncertainty σ of the likelihood function, this factor punishes over-complex models. As stated by William of Occam, an English Franciscan monk, in the thirteenth century:

”Frustra fit per ultra quod potest fieri per pauciora” or ”It is vain to do with more what can be done with fewer”.

Therefore, a good model must both be able to sufficiently explain the measured data and have a limited complexity, i.e. the model should not be over-parameterized. However, the Occam factor Ω by itself is not always relevant for the model selection process, e.g. when the evidence is governed by the likelihood term $p(Y = y|X = \hat{x}_{MAP})$, and its interpretation may be misleading. For poor data, e.g. data affected by speckle noise, the above conclusion may even be reversed, i.e. the Occam factor favors complex models to simpler ones [Blahut].

From Eq. 3.27 we see that the evidence is the product of Ω and the posterior fit at \hat{x}_{MAP} to the data. Thus, if the Gaussian assumption is fulfilled, the Bayesian method for model comparison by evaluation of the evidence is computationally no more demanding than finding the MAP estimate for each parameter and the corresponding width of the posterior product.

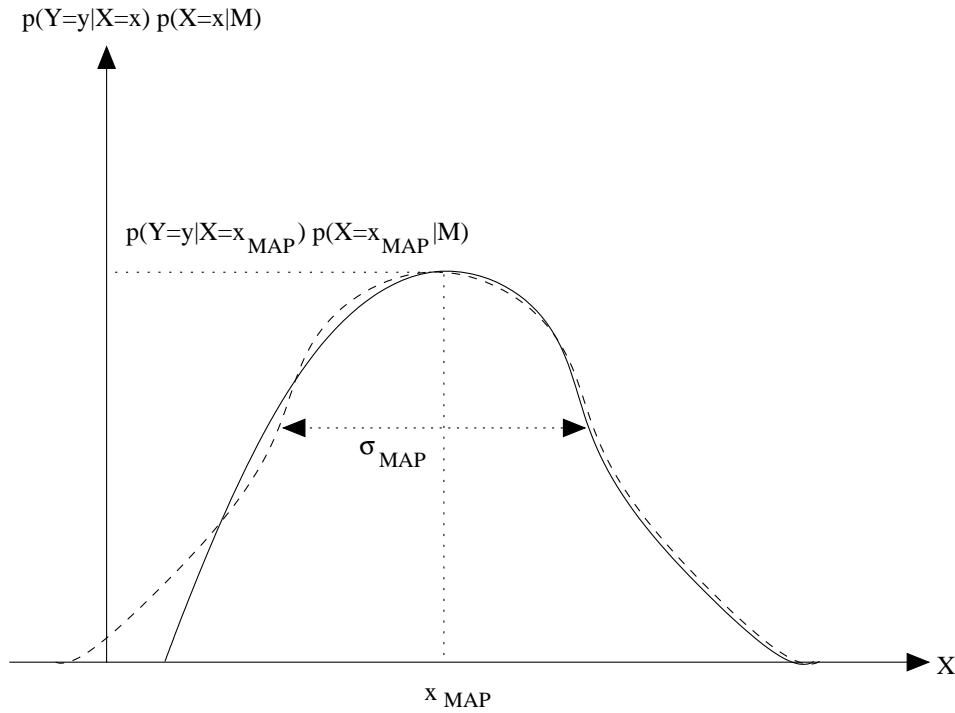


Figure 3.2: Approximation of the posterior (solid) by a Gaussian distribution (dashed). The mean of the Gaussian is equal to the MAP estimate of X , since the posterior reaches its maximum at this position.

Model selection example

To give an example, we go back to the estimation of a parameter X affected by additive Gaussian noise as presented before. The same assumptions concerning mean and variance of the noise apply, and X again is a realization of a Gaussian process with zero mean as in the MAP estimation example. Furthermore, we assume $\sigma_X = \sigma$ for computational simplicity.

We know that Y is the sum of two statistically independent random Gaussian processes of zero mean and variances σ^2 and σ_X^2 , respectively. As of this, the variance of our measurements is given by $\sigma_Y^2 = \sigma^2 + \sigma_X^2$. Assuming a uniform prior centered around zero and of width $6\sqrt{\sigma^2 + \sigma_X^2} = 6\sqrt{2}\sigma$ (a more conservative width would be $8\sqrt{2}\sigma$) to approximately cover the full range (99%) of possible values of Y , we find with Eq. 3.26 for the evidence of the ML estimation, i.e. the uniform prior

$$p(Y = y|M_{ML}) = \frac{1}{6\sqrt{\sigma^2 + \sigma_X^2}} = \frac{1}{6\sqrt{2}\sigma}. \quad (3.28)$$

In a similar way, the evidence of the Gaussian prior employed for the MAP estimate with

$\mathcal{N}(0, \sigma_X^2)$ yields

$$\begin{aligned} p(Y = y|M_{MAP}) &= \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_X^2)}} \exp\left(-\frac{y^2}{2(\sigma^2 + \sigma_X^2)}\right) \\ &= \frac{1}{\sqrt{4\pi\sigma^2}} \exp\left(-\frac{y^2}{4\sigma^2}\right). \end{aligned} \quad (3.29)$$

If σ_X is unknown, the above equation will also allow to determine its most probable value with regard to the data. Considering a measurement with $y = 0$, we find for the ratio of these two model evidences

$$\frac{p(Y = y|M_{MAP})}{p(Y = y|M_{ML})} = \frac{6\sqrt{2}\sigma}{\sqrt{4\pi\sigma^2}} \approx 2.394 \quad (3.30)$$

saying that the Gaussian prior model is more than twice as probable as a uniform prior. With increasing $|y|$ this ratio will become smaller as the uniform model fits better and the Gaussian model assumption weakens. The ratio can be found to be unity for measurements

$$|y| = 2\sigma \sqrt{-\log \frac{\sqrt{\pi}}{3\sqrt{2}}} \approx 1.868\sigma. \quad (3.31)$$

For smaller values of $|y|$, i.e. for about 81% of all observations, the Gaussian model used for MAP estimation is more evident. This result could already be guessed from the narrower confidence interval of Eq. 3.23.

3.1.4 Non-Informative Priors and the Principle of Maximum Entropy

We stated before, that in the case of full ignorance of a parameter, a uniform prior is to be employed. However, this assumption is not always fully correct and a special kind of prior should be used, as we will briefly demonstrate. Sometimes it may also happen that prior knowledge is only available in form of several constraints, but that the actual shape of the prior pdf is unknown. In order to determine this unknown prior pdf, the principle of maximum entropy can be used. The two following paragraphs are given only to complete our overview of estimation techniques and data analysis and are not relevant for the rest of this thesis:

Non-informative priors

Let us assume that we want to estimate a location parameter X , i.e. a parameter that describes the position of a distribution along the x -axis. Then, any translation of this parameter $X \rightarrow X + x_0$ should not change the assigned prior for X . To put it mathematically, the condition

$$p(X = x) = \frac{d(x + x_0)}{dx} p(X = x + x_0) \quad (3.32)$$

obtained from the relation $p(X = x) = p(X = g(x))\frac{dg(x)}{dx}$ should be fulfilled. Since x_0 is a constant, this can be solved as

$$p(X = x) = \text{constant}. \quad (3.33)$$

Consequently, the uniform distribution is a good choice for location parameters, such as the mean of the Gaussian distribution, if no additional knowledge is available.

Similar reflections apply for the determination of the prior of so-called scale parameters, that are associated with a size or a magnitude. The standard deviation of the Gaussian distribution is an example here. Since the resulting prior should be invariant towards a scaling of the random variable $X \rightarrow \alpha X$, using

$$p(X = x) = p(X = \alpha x)\frac{d(\alpha x)}{dx} \quad (3.34)$$

resulting from the above mentioned relation, we obtain

$$p(X = x) \propto \frac{1}{x}. \quad (3.35)$$

This prior was first suggested by Jeffreys [Jeffreys] in 1939 and is named after him as Jeffreys' prior. Hence, for the estimation of scale parameters the uniform assumption is not the best choice but should be replaced by $1/x$. The apparent drawback of a not-normalizable prior can generally be coped with. Interestingly, the Jeffreys' prior degenerates back into a uniform distribution for the logarithm of the random scale variable X :

$$p(\log X) = \text{constant} \quad (3.36)$$

Maximum entropy

The principle of maximum entropy [Blahut, Jaynes82, Sivia] can be used to assign prior pdfs to a random variable X , if only some constraints about X are known. It can be demonstrated [Jaynes82] that, if there is a measure to be maximized by a consistent pdf obeying the given constraints, this measure must be the entropy. The resulting pdf then possesses the highest number of possibilities to generate a random variable considering the imposed constraints. Thus, we are looking for a distribution that fulfills the given constraints and maximizes the entropy S [Shannon]. Initially, the entropy is given by

$$S = - \int_{-\infty}^{\infty} p(X = x) \log p(X = x) dx, \quad (3.37)$$

but can be extended to Kullback's information

$$K_S = - \int_{-\infty}^{\infty} p(X = x) \log \frac{p(X = x)}{m(X = x)} dx \quad (3.38)$$

for the continuous, and to

$$K_S = - \sum_{i=1}^N p(X = x_i) \log \frac{p(X = x_i)}{m(X = x_i)} \quad (3.39)$$

for the discrete case. We put here these two more general forms of Kullback's information, also known as cross-entropy, which additionally to the entropy S contain the pdf $m(X = x)$ ensuring the invariance of K_S under a change of variables. To exemplify the principle of maximum entropy, we assume that all that is known about our random variable X are its mean μ and its variance σ^2 . Under these assumptions, we find the resulting distribution by maximizing [Blahut]

$$\begin{aligned}
 C = & - \sum_{i=1}^N p(X = x_i) \log \frac{p(X = x_i)}{m(X = x_i)} \\
 & - \lambda_0 \left(1 - \sum_{i=1}^N p(X = x_i) \right) \\
 & - \lambda_1 \left(\sigma^2 - \sum_{i=1}^N p(X = x_i) (x_i - \mu)^2 \right),
 \end{aligned} \tag{3.40}$$

where λ_0 and λ_1 are Lagrange multipliers. For simplicity of mathematical analysis, we consider the discrete case in the limit of $N \rightarrow \infty$. The first term on the right-hand side is the entropic term, the center term describes the normalization condition of the distribution and the last term embodies our constraints about mean and variance.

After some algebra, we find, for the continuous case, the Gaussian distribution as the result, which best explains the given random variable X characterized solely by its mean and variance:

$$p(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x - \mu)^2}{2\sigma^2} \right) \tag{3.41}$$

In principle, analogous derivations can be performed for any kind of constraint imposed on X . We note that the positivity requirement of X in combination with a known mean value results in the negative exponential distribution. If, in addition, the variance is known, the Gamma distribution is obtained. As presented in chapter two, these distributions describe the speckle statistics.

3.1.5 Analogy to Tikhonov Regularization

The methods of Bayesian data analysis have been shown to originate from a complete theory. No ad-hoc methodologies have to be applied to obtain consistent results. This also applies for the prior information, since a formalism exists, that allows to determine the goodness of an employed model and to reject unjustified prior assumptions. Besides Bayesian methods, a variety of other theories and techniques dealing with data analysis exist. Here, we want to comment on Tikhonov regularization and its relation to the Bayesian approach.

Tikhonov Regularization

Tikhonov regularization [Tikhonov] is widely used for data inversion. The forward model directly corresponds to the Bayesian likelihood function and is not necessarily Gaussian-shaped, although this is often assumed for computational reasons. To guarantee a stable

solution in the case of under-determined systems Tikhonov constrains the solution to have a desired property, i.e. the inversion is regularized. Hence, the result is obtained by minimizing a function C consisting of a weighting of the forward fit $(y - F\{x\})^2$ and the regularization term $R\{x\}$, which is governed by a Lagrange multiplier or hyper-parameter λ . In the Bayesian theory, this can directly be interpreted as likelihood function and prior. In contrast to Bayesian methods the regularization term and λ are often heuristically chosen. However, several techniques exist to find optimal values for the Lagrange multiplier, like the L-curve technique. The analogy of Tikhonov regularization to Bayesian MAP estimation becomes visible in the following equations:

Minimizing

$$C(y, x) = (y - F\{x\})^2 + \lambda \cdot R\{x\} \quad (3.42)$$

as a function of x is equivalent to maximizing

$$\exp(-C(y, x)) = \exp(-(y - F\{x\})^2) \cdot \exp(-\lambda \cdot R\{x\}). \quad (3.43)$$

This can be regarded as a maximization of the posterior

$$p(X = x|Y = y) \propto p(Y = y|X = x) \cdot p(X = x) \quad (3.44)$$

and, thus, corresponds to the first level of Bayesian inference.

3.2 Markov Random Fields

Bayesian data analysis is well suited to image interpretation in combination with appropriate prior models for the noise-free scene. We have seen that, using a likelihood function and some assumptions about the original data, an improved estimate of this data can be obtained. The quality of this estimate, however, depends largely on the employed model and its ability to describe the image content. In this context, Markov and Gibbs random fields (MRFs and GRFs) play an important role, since they are able to statistically describe correlations, or even more generally, any kind of statistical dependence between neighboring pixels. Furthermore, they are easily applicable within the Bayesian framework.

Markov random fields have become popular in image processing since about fifteen years. Originating from statistical physics, where they have been used for the study of phase transitions, they are now widely employed to model two-dimensional lattices, such as image data. In the beginning, the use of Markov models in image processing was limited due to the constraint of causality, but after a solution to this problem had been found, they quickly became one of the standard image processing tools [Sigelle].

3.2.1 Neighborhood Systems and Cliques

The information of digital images is not only encapsulated in gray-values of individual pixels. More than that, images are usually composed of different regions and features with similar statistical properties, such as textures, lines and contours. As of this, several

independently considered pixels usually are not significant to describe all information of a certain image region, but become important by their relations and interactions with pixels in a neighborhood.

The characteristics of these local interactions between pixels, defining different regions of an image, can be modeled by a Markovian formalism, which is suitable for the envisaged framework of Bayesian data analysis. Within the related Gibbsian formalism, interactions between neighboring pixels are described by local energies, which correspond to the equivalent conditional probabilities of a Markov random field. The probability of the whole Gibbs field is encapsulated in a global energy function which can be decomposed into its local components.

Image description

As a condition for the application of Markov random fields, the modeled image is considered as a realization of a random field X that is composed of random variables X_i at different sites i representing individual pixels, i.e. $X = \{X_1, X_2, \dots\}$. We denote a concrete realization of X_i by x_i , which is a pixel gray-value lying in \mathcal{E} . Equivalently, x describes a sample realization of the whole field X , i.e. a sample image that takes one of the configurations out of $|\mathcal{E}|^{|X|}$ with the operator $|\cdot| = \text{card}(\cdot)$. As an example, this yields $256^{512 \times 512}$ possible configurations for an 8bit image of 512×512 pixels. The probability of a particular realization is given by $p(X = x)$.

Neighborhood systems

In order to model local spatial interactions between individual pixels a neighborhood system \mathcal{N} is defined in X . For a single pixel X_i at site i its neighborhood is described by [Li, Sigelle, Winkler]:

$$\mathcal{N}_i = \{j\} \quad \text{with} \quad \begin{cases} i \notin \mathcal{N}_i \\ j \in \mathcal{N}_i \Rightarrow i \in \mathcal{N}_j \end{cases} \quad (3.45)$$

where $\{j\}$ is a set of all pixel sites in the vicinity of site i , excluding i . Now, different neighborhood orders, i.e. neighborhoods of different sizes, can be defined as depicted in Fig. 3.3: In this figure, the neighborhood order of a pixel $x_{k,l}$ around a center pixel $x_{0,1}$ is denoted by the index k . The index l is a simple running variable, i.e. l, k do not denote x/y-coordinates as might be suspected at first sight, and therefore must not be mixed with the index i of x_i , which denotes the pixel site.

As an example, first and second order neighborhood configurations are shown in the center of Fig. 3.3: A first order neighborhood contains the four nearest pixels, a second order neighborhood the eight nearest pixels around its center pixel at site i , and so forth.

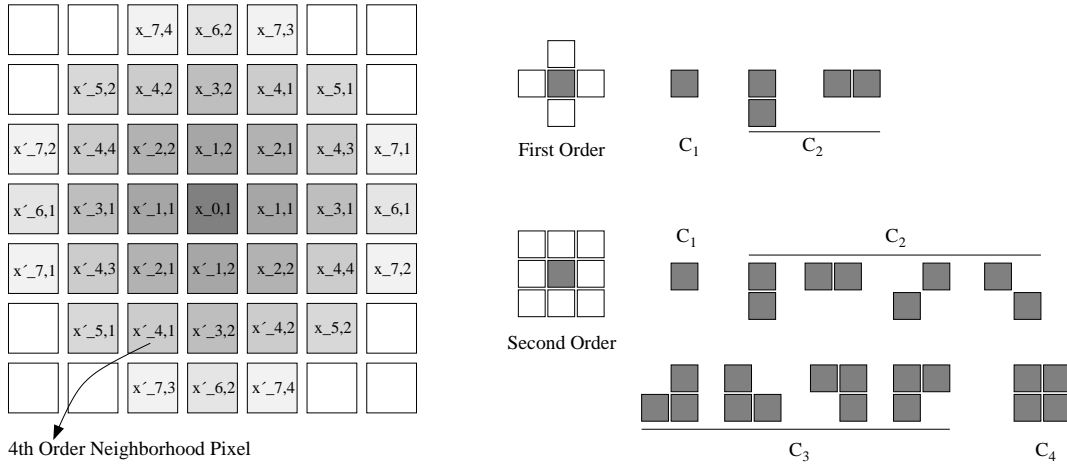


Figure 3.3: Cliques and neighborhoods. Left: Illustration of different neighborhood orders ranging from 0 (dark-gray) to 7 (light-gray). The neighborhood order of a given pixel $x_{k,l}$ is denoted by k . Center: First and second neighborhood orders for a given center pixel (dark-gray). Right: Cliques associated to the first and second neighborhood orders. In applications, the clique configurations are often limited to the cliques in C_2 , i.e. two-pixel cliques.

Cliques

Following this reasoning, a system of cliques can be set up, which depends on the size of the used neighborhood system \mathcal{N} , i.e. the neighborhood order.

A clique is defined by either a single pixel, or by a set of pixels that are neighbors to each other. Several possible clique-configurations for the first and second order neighborhoods are displayed on the right of Fig. 3.3.

We denote the ensemble of cliques relative to \mathcal{N} by \mathcal{C} and a clique consisting of k pixels by \mathcal{C}_k . For example, as shown, \mathcal{C}_2 of a first order neighborhood system contains only horizontal and vertical pixel pairs. For a second order neighborhood also diagonal pixel interactions exist. In this work, only cliques \mathcal{C}_2 , i.e. pixel pair interactions, are considered.

Energy and potential functions

Local interactions can now be described by potentials V_c for different cliques c . These potentials are a function of the gray-values of the pixels belonging to a clique. Hence, the global energy of the whole image can be written as the sum over all potentials

$$U(x) = \sum_{c \in \mathcal{C}} V_c(x). \quad (3.46)$$

In the same way, we find for the local energy at pixel X_i

$$U_i(x_i) = \sum_{c \in \mathcal{C}, i \in c} V_c(x_i), \quad (3.47)$$

which is the sum of the potentials over all cliques that include pixel X_i . As a result, the local statistical interactions of pixels within an image can be described by an appropriate set of potential functions V_c and a corresponding clique system \mathcal{C} . Note: Although writing $U_i(x_i)$ and $V_c(x_i)$, both functions may in fact depend on x_i and its complete neighborhood \mathcal{N}_i .

Example: Assuming a potential function $V_c(x_i, x_j) = (x_i - x_j)^2$ with $j \in \mathcal{N}_i$ for all cliques \mathcal{C}_2 of a first order neighborhood system, i.e. with four neighbors, we obtain the local energy

$$U_i(x_i) = (x_i - x_{left})^2 + (x_i - x_{right})^2 + (x_i - x_{up})^2 + (x_i - x_{down})^2 \quad (3.48)$$

for each center pixel X_i at site i , with x_{left} , x_{right} , x_{up} and x_{down} as the direct neighbors of X_i .

3.2.2 Markov-Gibbs Equivalence

By denoting the set of all pixels excluding X_i by \bar{X}_i with its sample realizations \bar{x}_i , i.e. $\bar{X}_i = \{X_j\}$ with $j \neq i$, a Markov random field is defined by

$$p(X_i = x_i | \bar{X}_i = \bar{x}_i) = p(X_i = x_i | x_j, j \in \mathcal{N}_i), \quad (3.49)$$

which means that a single pixel X_i can be fully statistically described by its relations to pixels within a limited neighborhood. The assumption of markovianity is justified for a large number of images, consisting of several regions with different properties, such as textures. Note that for computational reasons, we are interested only in images where the local interactions can be modeled with a relatively small neighborhood system, whereas the actual neighborhood to model complex images might be rather large.

To go from a global view to a local description of the random field and vice-versa, we introduce the Gibbs-Markov equivalence. We define a Gibbs random field by

$$p(X = x) = \frac{1}{Z} \exp(-U(x)), \quad (3.50)$$

where $U(x)$ is an energy function as defined in Eq. 3.46. The partition function Z serves as a normalization constant and is found to be

$$Z = \sum_{x \in |\mathcal{E}|^{|X|}} \exp(-U(x)). \quad (3.51)$$

It is obvious that Z cannot be computed due to the large number of possible configurations $|\mathcal{E}|^{|X|}$ of X . By rewriting Eq. 3.50 with Eq. 3.46

$$p(X = x) = \frac{1}{Z} \exp(-U(x)) = \frac{1}{Z} \exp\left(-\sum_{c \in \mathcal{C}} V_c(x)\right), \quad (3.52)$$

we see that the global energy $U(x)$ of the Gibbs random field can be decomposed into a sum of local energies. This enables a local processing based on local probabilities conditioned by neighboring pixels by means of the clique system.

Hammersley-Clifford theorem

The Hammersley-Clifford theorem [Hammersley] states the important finding that, for a limited neighborhood and discrete states of X_i , a Markov random field X characterized by a neighborhood system \mathcal{N} is equivalent to a Gibbs random field and a certain potential associated to each clique within this neighborhood. Hence, each Markov random field can be represented in the form of Eq. 3.52.

The global and the local probabilities, $p(X = x)$ and $p(X_i = x_i | \bar{X}_i = \bar{x}_i)$, can now be related. Using the local energy at pixel X_i as given in Eq. 3.47

$$U_i(x_i | x_j, j \in \mathcal{N}_i) = \sum_{c \in \mathcal{C}, i \in c} V_c(x_i, x_j, j \in \mathcal{N}_i) \quad (3.53)$$

and by decomposing the global energy term into

$$U(x) = \sum_{c \in \mathcal{C}, i \notin c} V_c(x) + U_i(x_i | x_j, j \in \mathcal{N}_i) \quad (3.54)$$

the local representation of the Gibbs random field is found to be [Sigelle]

$$\begin{aligned} p(X_i = x_i | \bar{X}_i = \bar{x}_i) &= \frac{\exp(-U_i(x_i, \bar{x}_i))}{\sum_{x_i \in \mathcal{E}} \exp(-U_i(x_i, \bar{x}_i))} \\ &= \frac{\exp(-U_i(x_i | x_j, j \in \mathcal{N}_i))}{\sum_{x_i \in \mathcal{E}} \exp(-U_i(x_i | x_j, j \in \mathcal{N}_i))} \\ &= p(X_i = x_i | X_j = x_j, j \in \mathcal{N}_i). \end{aligned} \quad (3.55)$$

This is an important result allowing to locally evaluate the conditional probability for each pixel given its neighbors. In contrast, we have seen that the evaluation of $p(X = x)$ is not possible because of the partition function Z . The local probability by itself only depends on the sum of the different potentials for each clique within the neighborhood given by the markovianity of the random field. Consequently, Eq. 3.55 represents the basis for all image processing involving Markov random fields, which are suitable to serve as regularization terms or, in the Bayesian framework, as prior distributions.

3.2.3 Gibbs Models

We have seen that Markov random fields can be described by potential functions working on a local neighborhood due to the Gibbs-Markov equivalence resulting from the Hammersley-Clifford theorem. In principle, there are no restrictions to the contents of these potential functions. The potentials attached to different cliques do not even have to be stationary but can vary throughout the image. For the problem of image restoration or information extraction, however, a certain number of "standard" potential functions exist.

Before presenting several often used Gibbs models, we comment on two relaxation methods which are used for the synthesis of sample realizations from Gibbs random fields. The images shown later in Fig. 3.4 have been generated with the Gibbs sampler. The employed model parameters for the different models are enumerated in Tab. 3.1.

Gibbs sampler

The Gibbs sampler was proposed by Geman and Geman [Geman84] and iteratively generates a realization of a Gibbs random field until convergence or, in practice, a fixed number of iterations is reached. The statistics of the synthesized image correspond to the global Gibbs distribution. Starting with a purely random initial configuration, the algorithm sequentially updates all image pixels at each iteration by sampling the new value from the local probability distribution conditioned by the neighbors of the considered pixel:

$$\begin{aligned} p(X_i = x_i | X_j = x_j, j \in \mathcal{N}_i) &= \frac{\exp\left(-U_i(x_i | x_j, j \in \mathcal{N}_i)\right)}{\sum_{x_i \in \mathcal{E}} \exp\left(-U_i(x_i | x_j, j \in \mathcal{N}_i)\right)} \\ &= \frac{1}{Z_i} \exp\left(-U_i(x_i | x_j, j \in \mathcal{N}_i)\right) \end{aligned} \quad (3.56)$$

The Gibbs sampler needs only a small number of iterations to converge but requires the computation of the local partition function Z_i , i.e. the whole distribution must be built. This may dramatically slow down the algorithm if $|\mathcal{E}|$ is large, say 256, and if the neighborhood is of higher order. Hence, the Gibbs sampler is only suitable to synthesize images with limited dynamic range.

Metropolis algorithm

The Metropolis algorithm [Metropolis], originating from statistical physics, is often used for the synthesis of Gibbs random fields since it does not require the calculation of Z_i . Like the Gibbs sampler, the Metropolis algorithm is also a stochastic relaxation algorithm, which works in a similar way.

At iteration n a new value $x_i^{(n)}$ is randomly chosen out of \mathcal{E} for each pixel. The local change of energy is then computed by

$$\Delta U = U_i(x_i^{(n)} | x_j^{(n-1)}, j \in \mathcal{N}_i) - U_i(x_i^{(n-1)} | x_j^{(n-1)}, j \in \mathcal{N}_i). \quad (3.57)$$

The new state $x_i^{(n)}$ is accepted if the energy is decreased, i.e. $\Delta U < 0$, or otherwise rejected, i.e. $x_i^{(n)} = x_i^{(n-1)}$, with probability $P = 1 - \exp(-\Delta U)$.

In this way, the global energy tends to decrease after each update, avoiding local minima by allowing a temporary increase of U with probability P . In contrast to the Gibbs sampler, it is not necessary to build the whole distribution. But on the other hand, the acceptance rate of new states, which is 100% for the Gibbs sampler, is considerably lower, requiring a much higher number of iterations to reach convergence. Other schemes for the choice of $x_i^{(n)}$, like $x_i^{(n)} = x_i^{(n-1)} + \delta x$, can also be successfully applied.

In principle, for both the Gibbs sampler and the Metropolis algorithm, a sequential updating, i.e. scanning, of all pixels is not required, but easy to implement. In general, a random updating scheme is allowed, which does not change the quality of the obtained results.

Generalized Potts model

The generalized Potts model [Wu82] can be regarded as an extension of the Ising model. The Ising model [Ising] is a binary model with two states and was developed to study ferromagnetic effects in statistical physics. Here again, the roots of the whole framework of Gibbs random fields and their synthesis become visible.

By extending the Ising model to $|\mathcal{E}|$ gray-levels, the generalized Potts model is obtained. As for all following models, only the clique system \mathcal{C}_2 with pixel pair interactions is considered. The neighborhood system can be of any size. For each clique c the potential is defined by

$$V_c(x_i, \theta_c) = \begin{cases} -\theta_c & \text{if all pixels in the clique } c \text{ are equal} \\ \theta_c & \text{otherwise} \end{cases} \quad (3.58)$$

where θ_c is a scalar model parameter of V_c describing the potential for pairs of pixels with $X_i = X_j$ and $X_i \neq X_j$, $j \in \mathcal{N}_i$. The local energy is then given as the sum over all these potentials:

$$U_i(x_i, \boldsymbol{\theta}) = \sum_{c \in \mathcal{C}, i \in c} V_c(x_i, \theta_c) \quad (3.59)$$

We remark that for the "normal" Potts model we have $\theta_c = \theta = \text{const}$ for all cliques c in \mathcal{C} . In the following, we use the vector notation $\boldsymbol{\theta}$ to denote all parameters θ_c of the local energy function.

Identical potential functions with $\theta_c = \theta$ for all cliques c result in an isotropic distribution without any directional preference imposed by a special clique configuration. Consequently, the normal Potts model is often used to model region labels in segmentation or classification applications [Kelly, Lakshmanan, Derin87, Derin90]. On the other hand, the generalized model can be used to generate textures with sharp edges [Derin87]. However, smoothly varying transitions between neighboring pixels cannot be synthesized due to the sharpness of the potential function (either $-\theta_c$ or θ_c). These properties make the generalized Potts model not suitable for the reconstruction of natural scenes. Representative examples with one fixed set of parameters, illustrating the basic visual properties of different models, are displayed in Fig. 3.4.

Example: Assuming a first order neighborhood system with its four neighbors, cliques \mathcal{C}_2 and $\theta_c = \theta$, we obtain the local energy

$$\begin{aligned} U_i(x_i) &= \sum_{c \in \mathcal{C}, i \in c} V_c(x_i) & (3.60) \\ &= V_1(x_i, x_{left}, \theta_1) + V_2(x_i, x_{right}, \theta_2) + V_3(x_i, x_{up}, \theta_3) + V_4(x_i, x_{down}, \theta_4) \\ &= -\theta_1 + \theta_2 + \theta_3 + \theta_4 \\ &= 2\theta \end{aligned}$$

for a center pixel X_i with the gray-values of its neighbors $x_i = x_{left}$, $x_i \neq x_{right}$, $x_i \neq x_{up}$ and $x_i \neq x_{down}$. For a local configuration with $x_i = x_{left} = x_{right} = x_{up} = x_{down}$, we obtain $U_i(x_i) = -4\theta$. Thus, the model favors configurations with equal gray-values.

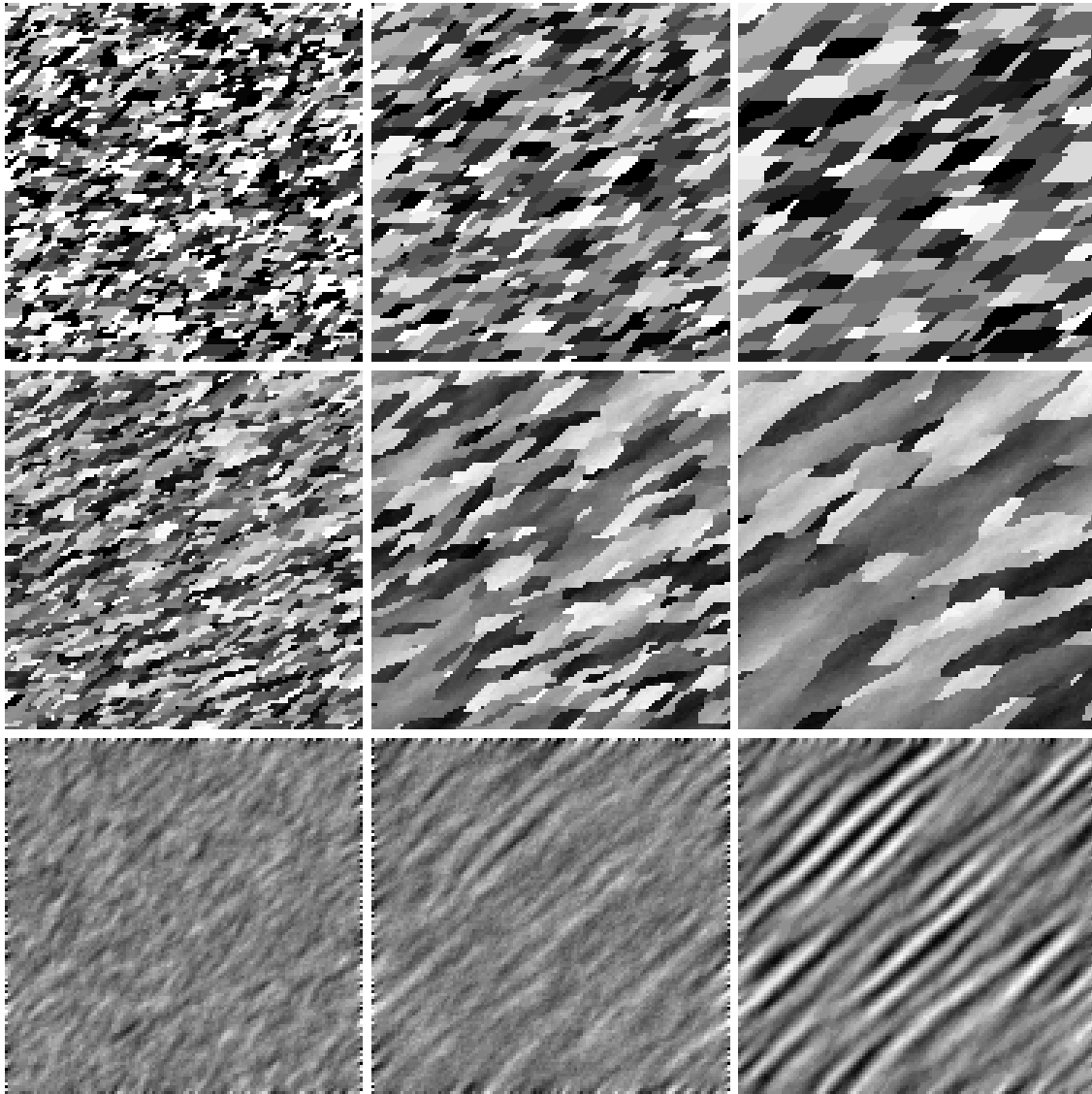


Figure 3.4: Realizations (128×128 pixels, 64 gray-levels) of second order neighborhood Gibbs random fields using different potential functions and parameters. The images synthesized with the Gibbs sampler are shown after three (left column), 25 (center column) and 100 (right column) iterations. Top row: Generalized Potts model. Center row: Realization of the discontinuity adaptive model of Eq. 3.61. Bottom row: Textures generated with the Gauss-Markov model. The corresponding parameters for each model are given in Tab. 3.1.

Discontinuity adaptive models

By choosing smoother potential functions the edge-preserving properties of the generalized Potts model can be combined with small gray-level variations in almost homogeneous areas.

Model	Δ, σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$
Potts model	-	-3	-1.5	-1.5	1.5
Discontinuity adaptive model	3	-5	-2.5	-2.5	2.5
Gauss-Markov model	2	0.225	0.2	0.2	-0.125

Table 3.1: Model parameters used for synthesis. The table enumerates the parameters of different models chosen for the generation of the textures in Fig. 3.4.

For illustration we have chosen two potential functions out of many that can be found in the literature. They are given by [Geman88, Li]

$$V_c(x_i, \theta_c, \Delta) = \frac{\theta_c}{1 + \left(\frac{\Delta x_i}{\Delta}\right)^2} \quad (3.61)$$

and

$$V_c(x_i, \theta_c, \Delta) = \theta_c \exp\left(-\left(\frac{\Delta x_i}{\Delta}\right)^2\right), \quad (3.62)$$

where $\Delta x_i = x_i - x_j$ and $j \in \mathcal{N}_i$. In general, these models show similar properties as the Potts model. The additional parameter Δ determines the smoothness of the potential function and, thus, allows more freedom to model small local variations while preserving strong edges. Hence, discontinuity adaptive models are often used for image regularization since they impose local smoothness, allow soft transitions and preserve borders.

It can be shown [Li] that the discontinuity-preserving property is determined by the first derivative of the potential function, which must converge back to zero for $\Delta x_i \rightarrow \pm\infty$ to ensure edge-preservation. We illustrate this behavior below in Fig. 3.5.

Gauss-Markov model

Compared to the class of models presented before, Gauss-Markov random fields (GMRFs) constitute a rather simple image model. They are much easier to apply and often allow (semi-)analytic algorithms, whereas general Gibbs models always result in slow, iterative and computationally heavy procedures. The Gauss-Markov model we use is defined by the local energy function

$$U_i(x_i, \boldsymbol{\theta}, \sigma) = \frac{(x_i - \mu_i)^2}{2\sigma^2}, \quad (3.63)$$

where

$$\mu_i = \sum_{k,l} \theta_{k,l} \cdot (x_{k,l} + x'_{k,l}). \quad (3.64)$$

and $\boldsymbol{\theta} = \{\theta_{k,l}\}$ is the model parameter vector.

In contrast to the previously presented clique-based potentials, the Gauss-Markov model is defined differently: The neighborhood configuration is encapsulated in a weighted sum over all neighboring pixels resulting in a prediction μ_i for the center pixel.

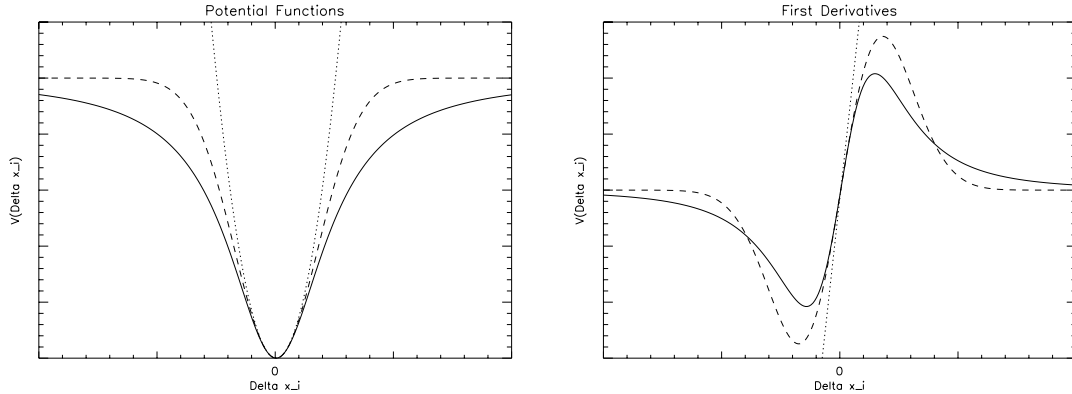


Figure 3.5: Potential functions and their first derivatives. Left: Potential functions for discontinuity adaptive models of Eq. 3.61 (solid) and Eq. 3.62 (dashed) with $\Delta = 32$ and $\theta_c = \theta = -1$. Additionally, a vertically shifted quadratic energy function (cf. Eq. 3.63) corresponding to the Gauss-Markov model is given (dotted) for $\sigma = 32$. Right: The corresponding first derivatives of these potential functions.

As of this and in order to simplify the following studies, we slightly change the notation of neighboring pixels and associated model parameters for the Gauss-Markov model: In Eq. 3.64 k, l denote the indices as given for the pixels in the left part of Fig. 3.3, i.e. k denotes the neighborhood order, l is a running number. A weighting factor $\theta_{k,l}$ is assigned to a pixel pair $x_{k,l}$ and $x'_{k,l}$ being symmetrical with regard to the center pixel $x_i = x_{0,1}$. Note that k, l are always relative to this center pixel.

Example: Assuming again a first order neighborhood system with its four closest neighbors, the prediction value μ_i of a center pixel X_i is computed by

$$\begin{aligned} \mu_i &= \theta_{1,1} \cdot (x_{1,1} + x'_{1,1}) + \theta_{1,2} \cdot (x_{1,2} + x'_{1,2}) \\ &= \theta_{1,1} \cdot (x_{right} + x_{left}) + \theta_{1,2} \cdot (x_{up} + x_{down}) \end{aligned} \quad (3.65)$$

Hence, $\theta_{1,1}$ corresponds to a horizontal interaction parameter, $\theta_{1,2}$ to a vertical one. This can easily be extended to larger neighborhoods with e.g. $\theta_{2,1}$ and $\theta_{2,2}$ denoting diagonal interactions, as used in Tab. 3.1.

Consequently, the parameters $\theta_{k,l}$ describe the correlation of X_i with its neighboring pixels. We further note the stability condition $\sum_{k,l} \theta_{k,l} = 0.5$, which must be fulfilled for a stochastic synthesis to converge. The sum always extends over all k, l of the chosen neighborhood order.

Assuming some distortion of the prediction μ_i by additive Gaussian noise N_i of zero mean and variance σ^2 the corresponding conditional pdf of Gauss-Markov random fields can be written in a closed form as

$$p(X_i = x_i | X_j = x_j, j \in \mathcal{N}_i, \boldsymbol{\theta}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma^2}\right). \quad (3.66)$$

which corresponds to a linear auto-regressive process with $x_i = \mu_i + n_i$, where n_i is a realization of N_i .

Further properties of Gauss-Markov random fields, which are, however, less important for our studies, are given by [Krishnamachari]. For example: The pdf of the joint field can be found by

$$p(X = x|\mathbf{B}, \sigma) = \frac{\sqrt{\det \mathbf{B}}}{(2\pi\sigma^2)^{|\mathbf{X}|/2}} \exp\left(-\frac{\mathbf{x}^T \mathbf{B} \mathbf{x}}{2\sigma^2}\right), \quad (3.67)$$

which is a multi-variate Gaussian distribution with the covariance matrix of X given by $\sigma^2 \mathbf{B}^{-1}$ and \mathbf{B} being a matrix as a function of the parameters $\theta_{k,l}$.

In contrast to non-linear models, the Gauss-Markov model cannot model sharp transitions. As depicted in Fig. 3.5, for small values of Δx_i the quadratic function of the Gauss-Markov model shows the same smoothing properties as the discontinuity adaptive potentials. For the edge-preserving models, however, the potential function no more increases beyond a certain threshold value of Δx_i , i.e. smoothing is no longer applied. Thus, if only the performance of the model for restoration is considered, non-linear discontinuity adaptive models are clearly to be preferred. However, the Gauss-Markov model promises to be analytically tractable. We will see that parameter estimation and the computation of a MAP estimate are much easier to be performed. This becomes even more important in the case of model parameter estimation in noisy conditions, i.e. for the estimation of hidden Markov random fields. The drawback of the model's inability to preserve sharp edges can partly be overcome by increasing the model order and some additional processing.

Other models

In addition to the presented models, there are many more potential functions adapted to various applications. To enumerate only a few of them, potential functions can be defined for linear structures [Tupin98, Tupin99], for processes that describe edges between regions of a certain homogeneity [Geman84], or other more specialized models for texture description like the auto-binomial or the exponential model [Cross, Schroeder98, Schroeder99b].

Image content representation by Markov random fields

By employing Markov and Gibbs random fields we are free to model all properties of the image content that we think are important for a particular application. Knowledge can be incorporated into a Gibbs model and then be used for information extraction [Schroeder99a] or as a prior in Bayesian image restoration. In both cases, the information regarding the model is encapsulated by the parameter vector $\boldsymbol{\theta}$ of the energy function. To give an example, $\theta_{1,1}$ and $\theta_{1,2}$ of the Gauss-Markov model describe the correlation of two neighboring pixels in the horizontal and vertical directions. Analogous relations apply for other parameters and different models, which may describe non-linear statistical dependences.

How the parameters of the potential functions are estimated for purposes of analysis and synthesis, and how the actual image restoration using both levels of Bayesian inference is

performed, will be discussed, with special emphasis on SAR imagery, in the next part.

3.3 Summary

In this chapter, the following points have been discussed:

- The Bayesian approach for data analysis was introduced and its difference in interpretation compared to the frequentist point of view was explained.
- It was shown that the incorporation of prior knowledge may help to improve the accuracy of parameter estimation. However, the correct choice of such a prior is crucial and often rises critics about the Bayesian approach.
- We demonstrated the importance of model selection to justify the use of an appropriate prior, gave an interpretation of the Occam factor and briefly outlined a technique to evaluate the evidence integral.
- The equivalence between Markov and Gibbs random fields has been shown. This allows a local modeling of spatial interactions between neighboring pixels by a stochastic description.
- We illustrated several basic potential functions and commented on their properties with regard to the application. Especially, the Gauss-Markov model rose our interest because it has the nice property of analytical tractability. However, its linear character does not allow the modeling of sharp edges.
- Markov random fields are promising to be used as prior information within a general Bayesian framework for image processing. The choice of the model or energy function to be used is application-dependent with regard to the kind of information to be extracted.

Part II

SAR Image Interpretation

4 Bayesian Image Despeckling

Synthetic Aperture Radar (SAR) systems use a coherent image formation technique to generate high-resolution images of the Earth's surface. The speckle effect, which results from coherent imaging, considerably disturbs the direct interpretation of the sensed intensity images. Consequently, methods have been developed to model the properties of speckle and to reduce its effects on the data.

In the first part of this thesis, we explained the speckle phenomenon and gave an introduction to Bayesian techniques that we want to use in this second part in order to improve the interpretability of SAR images. Before discussing Bayesian SAR image analysis using Markov random fields in the second section of this chapter, we first give an overview of already existing despeckling techniques in order to point out qualitative differences. In the second section, we illustrate the basic approach for SAR image restoration using Markov random fields by employing different potential functions and show the need for model parameter estimation. Several techniques for parameter estimation are then presented in section three and their effectiveness for the case of Gauss-Markov model parameter estimation from speckled data is investigated in section four. In this way, we develop the idea of a model-based despeckling and information extraction algorithm for SAR images, which will be the main subject of chapters five and six.

4.1 Overview of Existing Approaches

In this section, we present several of the most prominent approaches for SAR image despeckling, limiting ourselves to algorithms that are directly applied to detected images. As for the whole thesis, we do not consider any additional information that might be contained in complex data. Thus, the reviewed techniques do not only work on single look data but are also applicable to other SAR image products with reduced speckle variance. For all following investigations, we assume uncorrelated speckle noise (cf. appendix C).

The purpose of this overview is to illustrate the different degrees of modeling in existing approaches and to show their consequences on the filtering performance. Hence, we evaluate these approaches from a Bayesian point of view assuming that the quality of a filtered image increases with modeling accuracy. Results of some of the addressed filters for an artificially speckled image are illustrated below in Fig. 4.1, filter parameters are indicated and the quality of the results is measured in terms of the mean square error (MSE) relative to the

original optical image, as given in Tab. 4.1. The noisy image was synthesized by adding speckle with $L = 4$, where L denotes the equivalent number of looks (ENL), as defined in chapter two.

In the following, the noise model corresponding to Eq. 2.32 is assumed. For simplicity, we change notation to $y = x \cdot n$, where y is the observed intensity SAR image, x is the noise-free radar cross-section and n denotes the speckle noise, which is distributed according to Eq. 2.33. y , x and n are realizations of the corresponding random processes Y , X and N . The empirically measured mean and variance of y are denoted by μ_y and σ_y , respectively. Note that several of the presented filters work on square-root intensity, others on intensity images.

4.1.1 Simple Filters

The first class of filters we consider does not exploit any specific knowledge about the statistics of the noise. These filters rely on the assumption that the noise-free data are smooth and have the same mean as the noisy data. The resulting algorithms are extremely fast in computation, simple in theory, and do not use any explicit model (of course, any approach relies on implicit assumptions), neither for the noise, nor for the noise-free data. Consequently, the results are usually of poor quality.

Mean filter

The mean filter simply substitutes the gray-value of a given pixel by the mean computed in a certain neighborhood or in a scanning window around the pixel under consideration. This corresponds to the multi-looking approach in the spatial domain. Although subsampling is not applied, the spatial resolution of the averaged image is decreased in a similar way. Fine detail in the image is blurred or even disappears with increasing estimation window size. To have sufficient noise-suppression the typical window size is at least 5×5 pixels, i.e. a reduction of the speckle variance by a factor of 25. To improve the preservation of finer detail, this linear algorithm can be enhanced by using adaptive window sizes driven by the locally measured coefficient of variation.

Median filter

The median filter is also window- or neighborhood-based. A considered pixel is replaced by the median of all gray-values within the estimation window. As a result, the median filter does not introduce any new gray-values in the image, as the mean filter, and much better preserves edges [Rees]. However, fine detail, such as isolated point-scatterers, is filtered out. Compared to the mean filter, the median filter is clearly to be favored due to its edge-preserving capabilities. The median filter, which is a non-linear filter, works under the same assumptions as the linear mean filter, i.e. it does not use any explicit data or noise model.

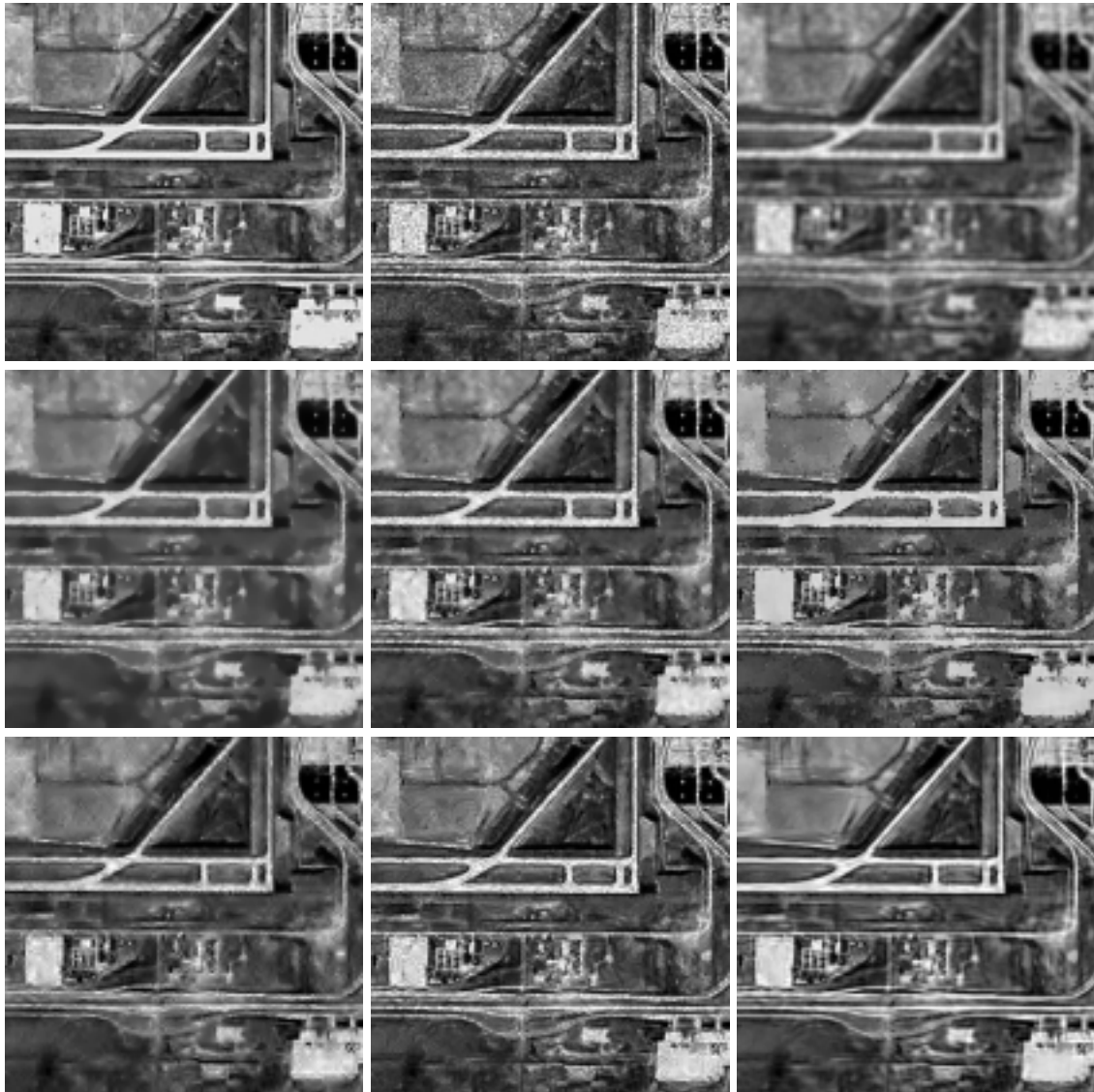


Figure 4.1: Illustration of the despeckling performance for different classes of filters applied to a synthetic image (256×256 pixels). Top row from left to right: Original optical test image without noise, speckled image ($L = 4$) to be filtered, mean filtered image (5×5 pixels). Center row: Crimmins filtered image (5 iterations), Kuan filtered image (5×5 pixels), EPOS filtered image (5×5 pixels). Bottom row: Wavelet filtered image, Wiener filtered image (8×8 pixels) and MBD filtered image (Model-Based Despeckling using Gauss-Markov random fields of seventh order). The basic versions of all filters without enhancements by additional structural information have been used for despeckling.

Geometrical filter

A completely different approach is introduced by the class of morphological filters. The original value of a center pixel is replaced by a non-linear combination of pixels from a

Speckled	Mean	Crimmins	Kuan	EPOS	Wavelet	Wiener	MBD
838	730	395	331	451	524	406	199

Table 4.1: Mean square errors of filtered images obtained from different speckle filters. The mean square error is given for the speckled and for each filtered image as presented in Fig. 4.1.

neighborhood system. The Crimmins filter [Crimmins] iteratively increments or decrements noise valleys or noise hills by applying different binary morphological masks. As for the mean and the median filters, no assumptions about the noise itself are made. However, the Crimmins filter has better chances to preserve edges and fine detail. The smoothness of the filtered image and the loss of fine structures is determined by the applied number of iterations. The Crimmins filter visually outperforms the mean and the median filter, which is easily explained by its improved ability to preserve structural information.

4.1.2 Statistical Filters

As for the mean and the median filter, statistical filters also use scanning windows ranging from typically 5×5 to 15×15 pixels. These filters are not limited to the calculation of an average gray-value, but obtain an estimate of the noise-free pixel as a weighted sum over all window pixels. The corresponding weights are driven by the locally measured values for mean and variance, i.e. by local statistics [Fjørtoft96]. This kind of filtering can be interpreted as a locally varying convolution kernel applied to the image.

Statistical filters take into account mean and variance of the noise, i.e. its basic properties. Furthermore, they consider, in a limited way, a certain roughness of the noise-free image. Reduced smoothing or no smoothing at all is applied in areas where the locally estimated coefficient of variation does not correspond to the known noise statistics. The additional roughness is recognized to be caused by image structures instead of noise. Hence, the assumption of stationary mean and variance is relaxed. We note that for most statistical filters improved versions exist that take into account additional structural information to allow a better filtering along edges.

Kuan filter

The Kuan filter [Kuan85, Kuan87] belongs to the class of minimum mean square error (MMSE) filters, i.e. $E\{(\hat{x} - x)^2\}$ is to be minimized, where \hat{x} is the estimate of x . The filter has been deduced by decomposing the observed signal y into x and an additive signal-dependent noise term, $y = x + (n - 1)x$, and can be considered optimal if both x and y are Gaussian distributed. With $E\{N\} = 1$ for intensity images, the filter equation is given by

$$\hat{x} = ky + (1 - k)\mu_y \quad (4.1)$$

with

$$k = \frac{\sigma_x^2}{\sigma_x^2 + (\mu_y^2 + \sigma_x^2)/L}. \quad (4.2)$$

The variance of the radar reflectivity is derived by $\sigma_x^2 = (L\sigma_y^2 - \mu_y^2)/(L + 1)$. L denotes the equivalent number of looks of the speckle noise.

The Kuan filter is an adaptive filter based on a test of the local coefficient of variation, i.e. mean μ_y and variance σ_y^2 of the observation. Compared to the filters presented before, this represents an important improvement, which is directly visible in the filtering results. However, the filter only computes a weighted sum of noisy and mean-filtered pixels. It does neither exploit the full knowledge of the noise distribution, nor does it include additional assumptions about the noise-free data.

Lee filter

The well-known Lee filter [Lee81, Lee83, Lee86] is a special case of the Kuan filter. It differs from the latter only in the weighting factor k because of a linear approximation made for the multiplicative noise model. For the Lee filter k can be found to be

$$k = \frac{\sigma_x^2}{\sigma_x^2 + \mu_y^2/L}. \quad (4.3)$$

Due to its more accurate modeling of the multiplicative noise behavior, the Kuan filter is to be preferred. However, the visual appearance of images filtered with Lee's filter is identical.

EPOS filter

The EPOS filter (edge-preserving optimized speckle filter) [Hagg] also relies on an analysis of mean and variance. To guarantee improved edge-preserving capabilities, the estimation window is divided into eight triangular areas. The mean of the most homogeneous area is taken as the estimate \hat{x} . In fact, borders are extremely well preserved and the filter smoothes right up to the edges, unlike the Kuan filter. However, the EPOS filter does not allow smoothly varying cross-sections. The filtering results are composed of areas of almost constant cross-section separated by sharp edges, comparable to the Potts model image of Fig. 3.4. Texture is absolutely not preserved by this kind of approach.

4.1.3 Multi-Scale Approaches for Despeckling

Multi-scale approaches rely on an analysis of the image at different resolutions. The image is represented by a pyramidal decomposition in order to profit from inter-scale dependences, to exploit correlations over several pixels that are easier to capture at lower scales, or to separately analyze lower and higher detail. This kind of approach mainly benefits from an analysis of mean and variance at different scales and from varying neighborhood sizes.

Filters based on the wavelet transform

Wavelet-based filters work on the wavelet transformed image. Since these approaches are applied under the assumption of additive Gaussian noise, the image is subject to a homomorphic transform (in the case of SAR, the logarithmically transformed image is computed) before the wavelet decomposition is applied. The wavelet coefficients are then shrunk according to various methods [Donoho] in order to reduce the noise energy in different sub-bands. The inverse-transformed image still exhibits a lot of detail, while the noise has been reduced. However, strong noise reduction is accompanied by the introduction of wavelet artefacts, which can be as disturbing as the speckle noise itself. Better wavelet shrinkage methods might solve this problem, but the main drawback remains the additive Gaussian noise approximation, which is not valid for low values of L (cf. chapter two). In addition, a model for the noise-free data is not included, and the high number of possible wavelets to apply is another free parameter. Wavelets represent a promising technique, due to their capability to capture spatial features within frequency sub-bands, but convincing results for despeckling are still difficult to find [Gagnon, Hervet].

Filters based on a multi-resolution decomposition

Multi-resolution approaches are usually extensions of existing basic filtering methods to the application at multiple scales [Aiazzi, Belhadj]. The image to be filtered is decomposed into a pyramidal representation. The different scales are filtered separately and are recombined to form the final filtering result. The noise variance is reduced with decreasing resolution, which simplifies the filtering at coarser scales. As a consequence, slightly improved results can be expected compared to the corresponding single-scale techniques. An additional degree of modeling enabling the restoration of a particular class of features is not introduced, unless multi-scale models [Bouman, Krishnamachari] which incorporate inter-scale dependences are used.

4.1.4 Wiener Filters

The Wiener filter is the optimal linear filter in the sense of minimum mean square error for stationary signals corrupted by additive noise. It requires the knowledge of the power spectra of the noise and of the noise-free scene, or equivalently, of their auto-correlation functions. Hence, the full Wiener filter is able to directly deal with correlated additive noise. In order to obtain optimal results with a Wiener filter both the noise and the noise-free signal must be Gaussian distributed, which is not the case for SAR data.

Frost filter

The Frost filter [Frost82a] is an adaptive Wiener filter resulting in a locally changing convolution kernel for the noisy image. It was derived directly for multiplicative noise under the assumption of locally stationary image data. After a good number of simplifications,

the impulse response is given by

$$h(r) = K_1 \exp\left(-K \frac{\sigma_y^2}{\mu_y^2} |r|\right). \quad (4.4)$$

Here, K is the filter parameter, which determines the strength of the filtering. K_1 is a normalization constant and $|r|$ is the radial distance from the center pixel to be filtered. The convolution kernel $h(r)$ is valid under the assumption of a scene reflectivity X obeying an auto-regressive process with an exponentially decreasing isotropic auto-correlation function.

However, the made approximations only result in a performance comparable to most other statistical filters. Unlike a full Wiener filter, the Frost filter does not rely on the computation of covariances or power-spectra to profit from a better image description.

Locally adaptive homomorphic Wiener filter

This approach is based on a homomorphic processing of the SAR image in combination with full Wiener filtering [Franceschetti, Kroschel, Papoulis77]. The homomorphic filtering is used to transform the multiplicative speckle into additive noise by taking the logarithm of the signal, thus permitting the application of a Wiener filter. The required power spectra can be calculated analytically for the noise and by an iterative procedure for the noise-free signal. However, convergence of the latter estimate to the correct power spectrum of the cross-section is not assured. In contrast to the Frost filter, the local estimation of the power spectrum allows a much better restoration of especially linear structures. On the other hand, artefacts may be introduced if the power spectrum is not correctly estimated or if the estimation window lies over highly non-stationary areas. Since rather large windows are required, typically between 8×8 and 16×16 pixels, this filter is not well suited for images with high variations in scene content.

Thanks to the information contained in the estimated power spectra, the results are of more than satisfactory quality for stationary textured areas. Note that the filtering and the estimation in the Fourier domain make this approach several magnitudes slower than statistical filters.

4.1.5 Bayesian Filters

Among the approaches presented here, the Bayesian approach is the only one that correctly models the speckle noise statistics in form of the likelihood function, i.e. knowledge about the shape of the noise distribution is exploited. As was demonstrated in the preceding chapter, the quality of the estimate is influenced by the employed prior assumptions about the noise-free data. Thus, using a good prior together with a complete description of the noise (complete in the sense of first order statistics), an improved quality of the filtered image can be expected.

Gamma-Gamma MAP filter

The Gamma-Gamma MAP (GGMAP) filter [Lopes90a, Lopes90b, Lopes93] is one of the most prominent despeckling algorithms and uses the Gamma distribution as likelihood function:

$$p(Y_i = y_i | X_i = x_i) = \frac{L^L y_i^{L-1}}{x_i^L \Gamma(L)} \exp\left(-\frac{L y_i}{x_i}\right) \quad (4.5)$$

This choice is obvious because the Gamma distribution corresponds to the intensity speckle pdf for homogeneous cross-sections known from Eq. 2.28. As mentioned in chapter two, the K-distribution is observed in non-homogeneously cluttered intensity SAR images, i.e. SAR images that contain some sort of texture, and can be shown to result from a speckled Gamma distributed cross-section. Hence, it is a reasonable choice to use the Gamma distribution for both prior and likelihood. The prior is described by

$$p(X_i = x_i | \mu_x) = \frac{\nu^\nu x_i^{\nu-1}}{\mu_x^\nu \Gamma(\nu)} \exp\left(-\frac{\nu x_i}{\mu_x}\right), \quad (4.6)$$

where μ_x is the mean value of the cross-section and $\nu = \mu_x^2 / \sigma_x^2$ is a form parameter reflecting its roughness. The variance of x is denoted by σ_x^2 . These parameters are estimated within windows centered around the pixel to be filtered. By setting the first derivative of the log-posterior to zero, it is straightforward to find the MAP estimate given by

$$\hat{x}_{i_{MAP}} = \frac{\alpha - L - 1}{2\alpha} \mu_y + \sqrt{\frac{(\alpha - L - 1)^2}{4\alpha^2} \mu_y^2 + 2L y_i \mu_y}, \quad (4.7)$$

with $\alpha = (L + 1) / (L \sigma_y^2 / \mu_y^2 - 1)$ and μ_y as the mean of the intensity observation computed in a window around y_i .

Although the modeling is more elaborate than in the approaches presented before, both visual and quantitative results are rather similar to the Kuan and Lee filters. Since the filter equation of the Gamma-Gamma MAP filter is also only based on mean and variance, this could have been expected. It is clear that outstanding results cannot be achieved by using a simple data model, which is described by only two parameters. Nevertheless, the Gamma-Gamma MAP filter is a solid and therefore widely used speckle filter. The results can be further improved by including a statistical edge detector to better cope with non-stationarities in mean backscatter.

Filtering with Gibbs random fields

Having seen that a Gamma prior seems not to be sufficient, one could think of more complex parametric models to describe spatial structures and, thus, encapsulate more information than mean and variance. For this purpose, Markov random fields are well adapted as was shown in the second section of chapter three. Following this reasoning, we will discuss the use of stochastic models in the next section and see if they really allow a better restoration of the noise-free scene. Finally, this will lead to the Model-Based Despeckling (MBD) algorithm [Walessa00] that we have developed. For comparison of the filtering performance

of its basic version without structural enhancements, as proposed in chapter five, its result for the given test image is provided in Fig. 4.1. Other approaches relying on Gibbs random fields are described in [Belhadj, Cabada, Dias, Walessa97], but show only limited success.

Several comparisons of different speckle reduction techniques and other interesting filtering approaches can be found in [Dewaele, McConnell96, Oliver91, Procello, Schwarz, Shengh, Touzi99] and [McConnell95, Nezry, White, Wu92], respectively. Additional remarks concerning the restoration of blurred data and of SAR images with correlated speckle are given in appendix C.

4.2 SAR Image Restoration with Markov Random Fields

In the preceding chapter, we have discussed the theory of Markov random fields and described several methods of how to synthesize a realization whose properties are determined by a parameter vector $\boldsymbol{\theta}$ describing pixel interactions between different pairs of pixels, i.e. cliques. Now, we will address the application of Markov random fields as low level image processing tools, i.e. for the filtering or restoration of SAR images. In this context, we put emphasis on the importance of the employed model parameters. We will demonstrate that a fixed parameter set, i.e. a fixed prior, cannot be a suitable choice. The prior and its parameters must be chosen with regard to the data.

4.2.1 Markov Random Fields in a Bayesian Framework

As explained in the first section of chapter three, the use of suitable prior knowledge increases the estimation accuracy. Consequently, the uncertainty imposed by the likelihood function on the observed data becomes smaller. The use of Markov random fields allows to introduce assumptions about spatial interactions of the pixels in the noise-free image. Depending on the used model, i.e. the chosen potential function and the likelihood, different kinds of information can be extracted from the data, such as a segmentation with the Potts model, an image restoration with a discontinuity adaptive model or a texture analysis with the Gauss-Markov model.

Considering a degraded image y , which is assumed to be a realization of a random field Y , we are looking for the undegraded image x being a realization of another random field X . The two processes X and Y are related via the likelihood function. In the case of speckle, this is the Gamma or square-root Gamma distribution. Information about X is contained in Y , but cannot be directly extracted, since it is hidden under noise. Hence, one speaks of a hidden Markov random field.

Looking for the most probable realization x given the observation y , we write Bayes' equation

$$p(X = x|Y = y, \boldsymbol{\theta}) = \frac{p(Y = y|X = x, \boldsymbol{\theta})p(X = x|\boldsymbol{\theta})}{p(Y = y|\boldsymbol{\theta})}, \quad (4.8)$$

where the term $p(X = x|\boldsymbol{\theta})$ is a Markov or Gibbs prior characterized by the parameter vector $\boldsymbol{\theta}$. Since the random process Y does not explicitly depend on $\boldsymbol{\theta}$ being already contained

in X , we simply write $p(Y = y|X = x)$ instead of $p(Y = y|X = x, \theta)$ for the likelihood function in the following.

Assuming uncorrelated noise, which strictly speaking is not true for SAR data, we can decompose the likelihood function into a product of local terms in analogy to the maximum pseudo-likelihood approach of [Besag86]

$$p(Y = y|X = x) = \prod_{i=1}^{|X|} p(Y_i = y_i|X_i = x_i) \quad (4.9)$$

and find the posterior to be a Gibbs distribution of the form of Eq. 3.52

$$p(X = x|Y = y, \theta) \propto \exp \left(\sum_{i=1}^{|X|} \log p(y_i|x_i) - \sum_{c \in \mathcal{C}} V_c(x, \theta) \right), \quad (4.10)$$

which can be written as a function of local terms. Realizations of this Gibbs random field can again be synthesized using the Gibbs sampler or the Metropolis algorithm. However, in the case of image restoration this is not our goal: We are looking for the MAP estimate \hat{x}_{MAP} , i.e. for the realization that maximizes the posterior distribution. This maximization of the posterior as a function of $|X|$ variables can be performed by a slightly modified form of the relaxation algorithms that we have used for texture synthesis and is called simulated annealing. Several commonly used optimization methods, such as simulated annealing (SA) [Aarts, Kirkpatrick, VanLaarhoven], iterated conditional modes (ICM) [Besag86, Winkler] and the Newton-Raphson algorithm [Sivia] are outlined in more detail in appendix B.

In the following, we concentrate on the influence of the model parameters on the restoration result and on their estimation from the data and assume that the optimization problem for cross-section estimation is solved by a suitable algorithm like simulated annealing or ICM.

4.2.2 SAR Image Despeckling with Parametric Gibbs Priors

In the following examples, we consider square-root intensity images, unless otherwise stated. In contrast to most other filters, which work in the intensity domain, square-root intensity images must be used here, in order to limit the dynamic range. We have already pointed out that, for computational reasons, the number of possible states should be small. The calculation of the local partition function Z_i for 2^{16} gray-values is computationally not feasible. Even for eight bit images the resulting optimizations are rather slow. Apart from that, the proposed models usually behave better for images with a moderate dynamic range, like optical data. It is clear that the use of a particular prior for square-root images implies a different model for intensity images, and vice-versa. By transforming an image from one space to another, the corresponding model will change as well. The behavior of the Gauss-Markov model in different signal spaces will be studied in the next chapter.

Beach sand	Δ, σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	MSE
Discontinuity adaptive, fixed	12	-2	-2	-2	-2	260
Discontinuity adaptive, estimated	85	-19.7	-19.4	8.1	4.8	167
Gauss-Markov, fixed	6	0.125	0.125	0.125	0.125	283
Gauss-Markov, estimated	6.5	0.41	0.40	-0.17	-0.13	176
Straw	Δ, σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	MSE
Discontinuity adaptive, fixed	12	-2	-2	-2	-2	390
Discontinuity adaptive, estimated	55	-0.8	-8	-8.7	-3.3	161
Gauss-Markov, fixed	6	0.125	0.125	0.125	0.125	549
Gauss-Markov, estimated	6.5	0.02	0.29	0.27	-0.08	165

Table 4.2: Model parameters used for despeckling and resulting mean square errors. The model parameters, fixed or estimated, and the mean square errors of the restorations are given for the images depicted in Figs. 4.2 and 4.3. The mean square errors of the speckled data are $MSE = 299$ and $MSE = 355$ for the "sand" and the "straw" texture, respectively.

To calculate the MAP estimate of the noise-free scene we use the likelihood function for square-root SAR images of Eq. 2.36 given by

$$p(Y_i = y_i | X_i = x_i) = 2 \left(\frac{y_i}{x_i} \right)^{2L-1} \frac{L^L}{x_i \Gamma(L)} \exp \left(-L \left(\frac{y_i}{x_i} \right)^2 \right), \quad (4.11)$$

where y_i is the observed gray-value at pixel position i , x_i is the radar cross-section and L denotes the equivalent number of looks, i.e. the speckle noise level.

Despeckling with discontinuity adaptive models

We use a discontinuity adaptive Gibbs prior as a regularization term to impose a certain smoothness on the data. In a first example, we do not encode different directional preferences in the parameter vector θ . The model characteristics are determined by Δ and $\theta_c = \theta = const$ for all cliques, i.e. by two free parameters whose values are fixed heuristically. In a second example, the MAP estimate is computed by using the full parameter vector which has been estimated from the original noise-free image, serving as training data. Consequently, we consider a prior, which encapsulates real knowledge about the data to be estimated.

In our synthetic examples we have chosen $L = 4$ for the noise level and a second order neighborhood system, i.e. five model parameters, to keep the computational load limited. The employed prior is Gibbsian according to Eq. 3.52 with the potential function

$$V_c(x_i, \theta_c, \Delta) = \frac{\theta_c}{1 + \left(\frac{\Delta x_i}{\Delta} \right)^2}. \quad (4.12)$$

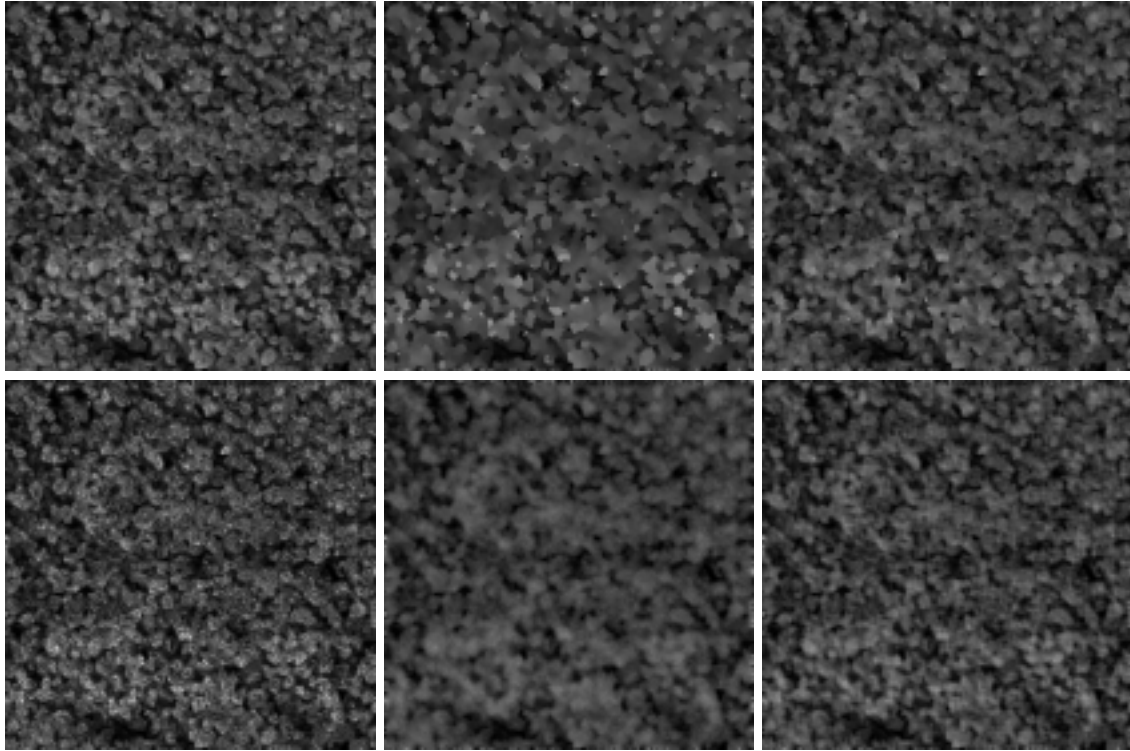


Figure 4.2: Influence of the model and its parameters on the texture restoration. Top row: Original noise-free "beach sand" image (256×256 pixels), restoration using the discontinuity adaptive model with fixed parameters, restoration using the discontinuity adaptive model with estimated parameters. Bottom row: Speckled image with $L = 4$ (MSE=299), restoration using the Gauss-Markov model with fixed parameters, restoration using the Gauss-Markov model with estimated parameters. Parameters and mean square errors are given in Tab. 4.2.

Simulated annealing (cf. appendix B) was used to maximize the resulting global posterior, which is of the form of Eq. 4.10, and to ensure convergence of the estimate to a good local minimum of the global energy, since in practice the global minimum is difficult to be found even by simulated annealing. The results for two images with different contents are presented in Figs. 4.2 and 4.3. The employed model parameters and the achieved mean square errors are provided in Tab. 4.2. The used textures have been extracted from the Brodatz album.

We can see from the examples with parameters estimated from the noise-free data that both the "straw" and the "beach sand" images are more than satisfactorily restored. This is due to the model's ability to describe different image contents by means of its parameters. A fixed parameter set for one image with heterogeneous contents, or even for different images, will never yield satisfactory results. This is reflected both in the visual representation and in the mean square error. For the "straw" texture, the heuristically chosen smoothing parameters even increase the mean square error emphasizing the influence of the prior.

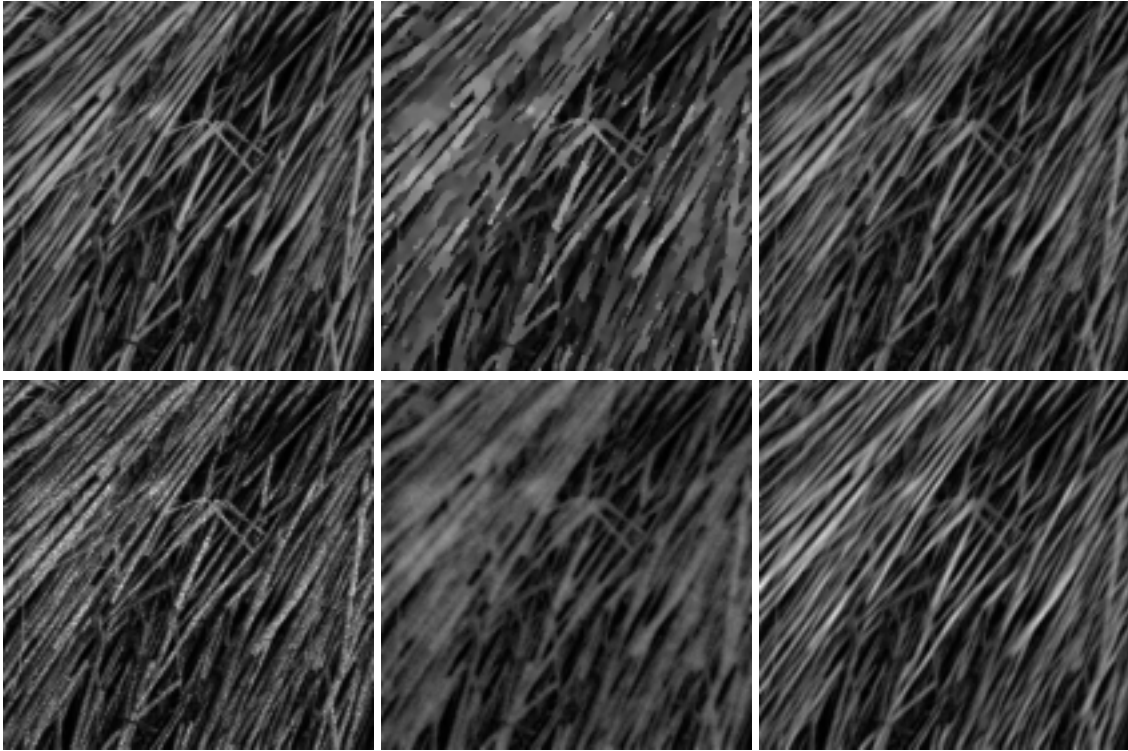


Figure 4.3: Influence of the model and its parameters on the texture restoration. Top row: Original noise-free "straw" image (256×256 pixels), restoration using the discontinuity adaptive model with fixed parameters, restoration using the discontinuity adaptive model with estimated parameters. Bottom row: Speckled image with $L = 4$ (MSE=355), restoration using the Gauss-Markov model with fixed parameters, restoration using the Gauss-Markov model with estimated parameters. Parameters and mean square errors are given in Tab. 4.2

We learn from these two examples that correct model parameters give much better results than values that have been chosen heuristically. In practice however, the parameter values are not known and must be estimated from the noisy data. Parameter estimation for discontinuity adaptive models, or non-linear models in general, from noise-free images is already difficult, as we will demonstrate in the next section. To estimate parameters of hidden Markov fields, especially in the case of multiplicative noise, is a lot more challenging.

Despeckling with the Gauss-Markov model

The same experiments have been performed with the linear Gauss-Markov model. In the first example, the model parameters have been chosen heuristically. In the second example, they have been estimated from the original noise-free data. Again, we use a second order neighborhood system to make a fair comparison with the discontinuity adaptive model. Higher model orders, up to a certain complexity, yield better results, but parameter esti-

mation for the non-linear models becomes too slow, since an iterative approach is required.

As before, the square-root intensity case was considered with the likelihood function of Eq. 4.11. The used local Gauss-Markov prior is given by

$$p(X_i = x_i | X_j = x_j, j \in \mathcal{N}_i, \boldsymbol{\theta}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma^2}\right) \quad (4.13)$$

as already defined in Eq. 3.66. This time, ICM (cf. appendix B) was used to calculate the MAP estimate. As a matter of fact, the optimization space of the posterior distribution is rather smooth and convex, since both the Gauss-Markov prior and the likelihood function represent uni-modal pdfs. For reasonable values of σ of the Gauss-Markov (GMRF) model, the likelihood function is almost constant around the significant width of the prior distribution allowing to approximate the posterior by a Gaussian. We will intensively discuss this property in chapter five.

By comparing the results illustrated in Figs. 4.2 and 4.3, the previously drawn conclusions are confirmed: To obtain consistent and meaningful results the hidden model parameters of the noise-free scene must be directly estimated from the noisy data, which represent the only source of information to guarantee optimal results. This can be considered as the only way to ensure a reliable MAP estimate, which is not misguided by a prior that has no evidence in the data. The heuristically fixed smoothing prior heavily blurs the textures similarly to a strong mean filter and results in a higher mean square error than the noisy image itself.

On the other hand, we notice the main deficiency of the Gauss-Markov model: While the results for the "straw" image are convincing, the restoration of the "beach sand" image looks a little bit blurred along the edges. This effect increases for images that show more sharp edges, which are difficult to be described by the Gauss-Markov model. At this point, the non-linear model behaves much better for both examples. However, this advantage in comparison with the Gauss-Markov model cannot be justified by computation times for parameter estimation, which are in the range of two orders of magnitude higher.

Conclusions

We have seen that both the employed prior model and its model parameters affect the quality of the restored images. While the choice of the parametric model will rather lie on practical considerations, like computing times and mathematical handling, the selection of the model parameters plays a crucial role, due to their considerable impact on the filtering result. We state again, that model parameter estimation for prior Markov models is indispensable for proper image filtering with space-variant non-linear Bayesian restoration methods. Heuristically chosen parameters, e.g. in form of a fixed regularization term, cannot yield optimal results, but may, on the contrary, even further degrade the observed data.

4.3 Model Parameter Estimation Techniques

In the preceding section we gave several examples for the restoration of speckled textures and stated that it is imperative to estimate model parameters, instead of choosing them heuristically. Of course, in the real case, this estimation can only be performed from the noisy observation: We are interested in extracting knowledge about the noise-free data that may help to determine a prior. Except for rare cases, a direct estimation of hidden parameters is impossible.

In the following, we first address the problem of parameter estimation from noise-free data. Although this seems to be of little use for our problem, we will see later that this sort of estimation is needed in a particular algorithm for the extraction of hidden model parameters. In principle, the description of the estimation algorithms is kept general enough to be applicable for all classes of Gibbs models. We put special emphasis on the estimation of hidden model parameters for Gauss-Markov random fields in the next section of this chapter.

4.3.1 Model Parameter Estimation from Complete Data

The extraction of model parameters from data assumed to be noise-free plays an important role, e.g. in current research related to large databases that allow a query by image content [Datcu99, Rehrauer, Schroeder99a, Schroeder99b], instead of a query by geographic coordinates or acquisition time. As in our approach, estimated model parameters are used as a description of actual image features and are easily extracted from optical data exhibiting no or very low additive noise.

Under the assumption of a noiseless acquisition, we deal with so-called complete data $Y = X$ and describe the observed realization of a Gibbs random field by

$$p(X = x|\boldsymbol{\theta}) = \frac{1}{Z} \exp(-U(x, \boldsymbol{\theta})), \quad (4.14)$$

where the energy function $U(x, \boldsymbol{\theta})$ is characterized by a parameter vector $\boldsymbol{\theta}$. For simplicity of notation, we denote all parameters including scale parameters like σ or Δ by the vector $\boldsymbol{\theta}$. In order to estimate the set of parameters that best explains the observation x , our task consists in maximizing the probability

$$p(\boldsymbol{\theta}|X = x) \propto p(X = x|\boldsymbol{\theta}) \quad (4.15)$$

as a function of $\boldsymbol{\theta}$. The above proportionality results from a maximum likelihood estimation of the parameters, i.e. we usually do not employ any prior assumptions about $\boldsymbol{\theta}$. Various methods to perform this maximization exist. The most prominent ones will be pointed out below.

Maximum pseudo-likelihood (MPL) estimation

We have noted in chapter three that the computation of the partition function Z is usually impossible. Hence, the global probability $p(X = x|\boldsymbol{\theta})$ cannot be evaluated. However, it can

be shown [Besag86] that

$$p(X = x|\boldsymbol{\theta}) \approx \prod_{i=1}^{|X|} p(X_i = x_i|\boldsymbol{\theta}) \quad (4.16)$$

and consequently with Eq. 4.14,

$$\log p(X = x|\boldsymbol{\theta}) \approx - \sum_{i=1}^{|X|} U_i(x_i|\boldsymbol{\theta}) - \sum_{i=1}^{|X|} \log Z_i \quad (4.17)$$

are valid approximations, if $|X|$ is large enough. The expression in Eq. 4.16 is called pseudo-likelihood and can be maximized as a function of $\boldsymbol{\theta}$ by gradient approaches, since the local partition function Z_i is computable. Convergence is guaranteed because the pseudo-likelihood is a convex function if the relationship between local energies and model parameters is linear, which is the case for the directional elements of $\boldsymbol{\theta}$, but not for the scale parameters Δ or σ . Nevertheless, the local partition function has to be evaluated, which is still computationally heavy, if it cannot be determined analytically. Hence, in these cases, the computation time is proportional to the number of image gray-levels and to the size of the neighborhood. Parameter estimation for large neighborhoods and 256 gray-levels is already no more feasible, even for small images.

Approximation of non-linear models by a combination of Gaussians

The principal reason for the slowness of the maximum pseudo-likelihood (MPL) parameter estimation for non-linear models lies in the required numerical computation of the local partition function Z_i . If an analytical expression is not available, Z_i has to be calculated by numerically integrating over the whole un-normalized distribution.

A faster approach may consist in approximating the local distribution by a mixture of Gaussians for which the normalization factor can easily be computed. As a matter of fact, a local Gibbs distribution is often the product of N Gaussian-like functions, where N is the number of considered cliques. As an example, we derived an approximation for the potential function of Eq. 3.62 which, for a single clique c , is found to be

$$\exp(-V_c) = \exp\left(-\theta_c \exp\left(-\left(\frac{\Delta x_i}{\Delta}\right)^2\right)\right) \approx 1 + \exp\left(-\frac{\Delta x_i^2}{2\sigma_{DA}^2}\right) (\exp(-\theta_c) - 1), \quad (4.18)$$

with

$$\sigma_{DA}^2 = \begin{cases} -\Delta^2/(2\theta_c) & \text{for } \theta_c < -1 \\ \Delta^2\theta_c/2 & \text{for } \theta_c > 1 \\ \Delta^2/2 & \text{otherwise,} \end{cases} \quad (4.19)$$

having been derived analytically and experimentally from the curvature at the maximum of the non-linear potential function. We also derived the normalization constant of such a single clique contribution to be described by

$$Z_c = G + (\exp(-\theta_c) - 1) \sqrt{2\pi\sigma_{DA}^2}, \quad (4.20)$$

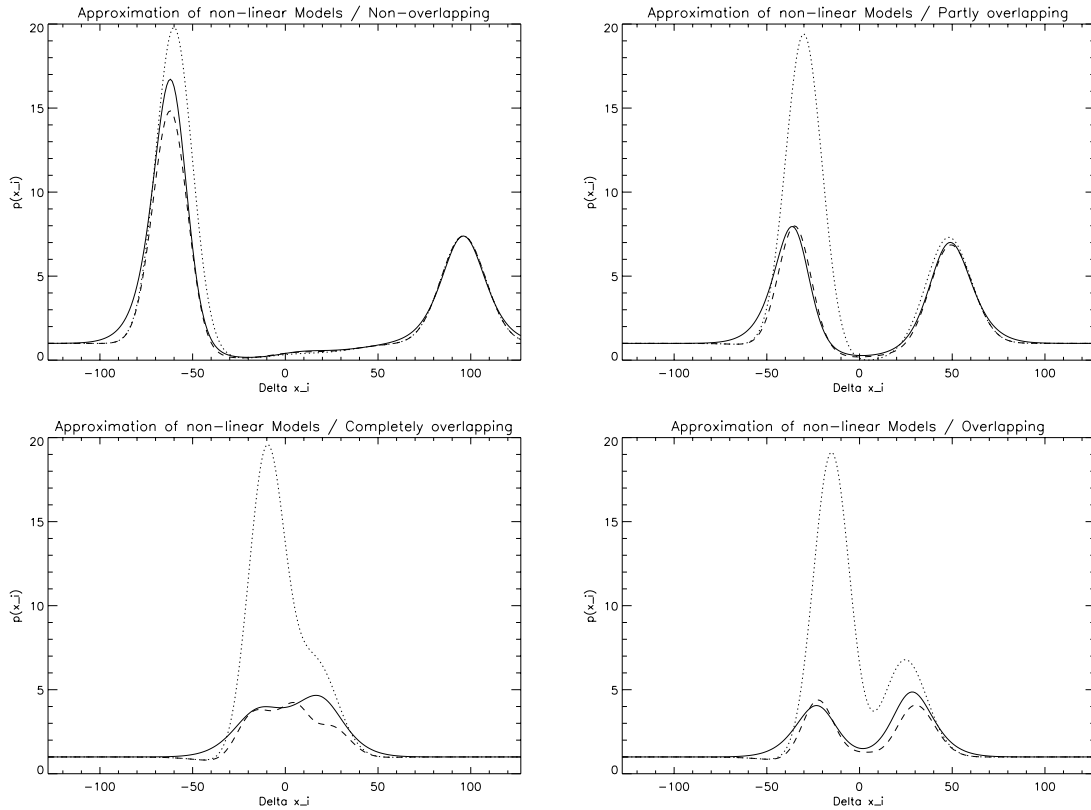


Figure 4.4: Approximation of a non-linear model by a combination of Gaussians. The pdfs for the exact non-linear potential (solid), the approximations by a summation of Gaussians (dotted) and by a multiplication of the terms of Eq. 4.18 (dashed) are given for a second order neighborhood with $\boldsymbol{\theta} = (\theta_{1,1}, \theta_{1,2}, \theta_{2,1}, \theta_{2,2}) = (-3, 2, 0.5, -2)$ and $\Delta = 24$. Clockwise: Both approximations are valid for non-overlapping contributions of the different cliques. With increasing overlap, the first approximation, which allows an easy computation of the partition function, becomes invalid. For a complete overlap, the multiplicative approximation is no more able to correctly describe the non-linear behavior of the model either.

where G is the number of used gray levels.

The complete localized Gibbs distribution can now be roughly approximated by a sum of these different Gaussian-shaped distributions for each clique corrected by an additive constant, which allows the direct computation of Z_i . However, a more accurate approximation is given by the product of the terms for each clique given in Eq. 4.18. Unfortunately, in the latter case the computation of Z_i is no more straightforward.

The resulting approximations for several local conditional pdfs are exemplified in Fig. 4.4. It can be seen that the summation of different Gaussians is only a good approximation if the individual contributions of the cliques do not overlap. In fact, this case is not interesting

for real textures, which have neighboring gray levels of similar value, hence Δx_i is rather small for all cliques. The multiplicative approximation is much more accurate. But here, the similarity also degrades for very overlapping contributions of the different cliques. Even if Z_i was easily computable, this approximation would not be always satisfactory. Experiments with such an approximated partition function Z_i did not show convergence of the model parameter estimates to their correct values.

We deduce from these examples that the presented simple approximation of non-linear models does not give convincing results for the important case of overlapping contributions from different cliques. As a matter of fact, it is in this case, where the power of the non-linear potential functions, which cannot be easily approximated by an additive or multiplicative combination of Gaussians, lies. We conclude that more research in this field is needed to find better approximations with necessarily increased complexity.

Maximum pseudo-likelihood estimation for Gauss-Markov random fields

It is of particular interest for practical applications that, for the Gauss-Markov model, the maximum pseudo-likelihood approach results in a rather fast, not to say instantaneously computed, analytical solution [Chellappa, Manjunath] of the form

$$\hat{\boldsymbol{\theta}} = \mathbf{A}^{-1} \mathbf{b} \quad \text{and} \quad \sigma^2 = E\{(x_i - \mu_i)^2\}, \quad (4.21)$$

which is determined by correlations between neighboring pixels. The model parameter vector to be estimated is defined as

$$\hat{\boldsymbol{\theta}} = (\hat{\theta}_{1,1}, \hat{\theta}_{1,2}, \dots)^T, \quad (4.22)$$

where T denotes the transposition of vectors. Concerning the computation of the standard deviation σ , x_i denotes single pixel values and $\mu_i = \sum_{k,l} \hat{\theta}_{k,l} (x_{k,l} + x'_{k,l})$ gives their value predicted by the Gauss-Markov model with parameters $\hat{\theta}_{k,l}$. The matrix \mathbf{A} contains the correlation factors of all combinations of pixel pairs, i.e. left and right neighbors $x_{k,l}$ and $x'_{k,l}$ of a center pixel, within a given neighborhood since $\theta_{k,l}$ is the interaction parameter for both neighbors

$$\mathbf{A} = \begin{pmatrix} E\{(x_{1,1} + x'_{1,1}) \cdot (x_{1,1} + x'_{1,1})\} & E\{(x_{1,1} + x'_{1,1}) \cdot (x_{1,2} + x'_{1,2})\} & \cdots \\ E\{(x_{1,2} + x'_{1,2}) \cdot (x_{1,1} + x'_{1,1})\} & E\{(x_{1,2} + x'_{1,2}) \cdot (x_{1,2} + x'_{1,2})\} & \cdots \\ E\{(x_{2,1} + x'_{2,1}) \cdot (x_{1,1} + x'_{1,1})\} & E\{(x_{2,1} + x'_{2,1}) \cdot (x_{1,2} + x'_{1,2})\} & \cdots \\ E\{(x_{2,2} + x'_{2,2}) \cdot (x_{1,1} + x'_{1,1})\} & E\{(x_{2,2} + x'_{2,2}) \cdot (x_{1,2} + x'_{1,2})\} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (4.23)$$

and the vector \mathbf{b} is composed of the correlation coefficients between a center pixel $x_{0,1}$ and its neighbors $x_{k,l}$ and $x'_{k,l}$

$$\mathbf{b} = (E\{x_{0,1} \cdot (x_{1,1} + x'_{1,1})\}, E\{x_{0,1} \cdot (x_{1,2} + x'_{1,2})\}, \dots)^T. \quad (4.24)$$

Obviously, the solution of this set of linear equations depends on the invertibility of the matrix \mathbf{A} . Parameters for uniform regions with zero variance cannot be derived in this way,

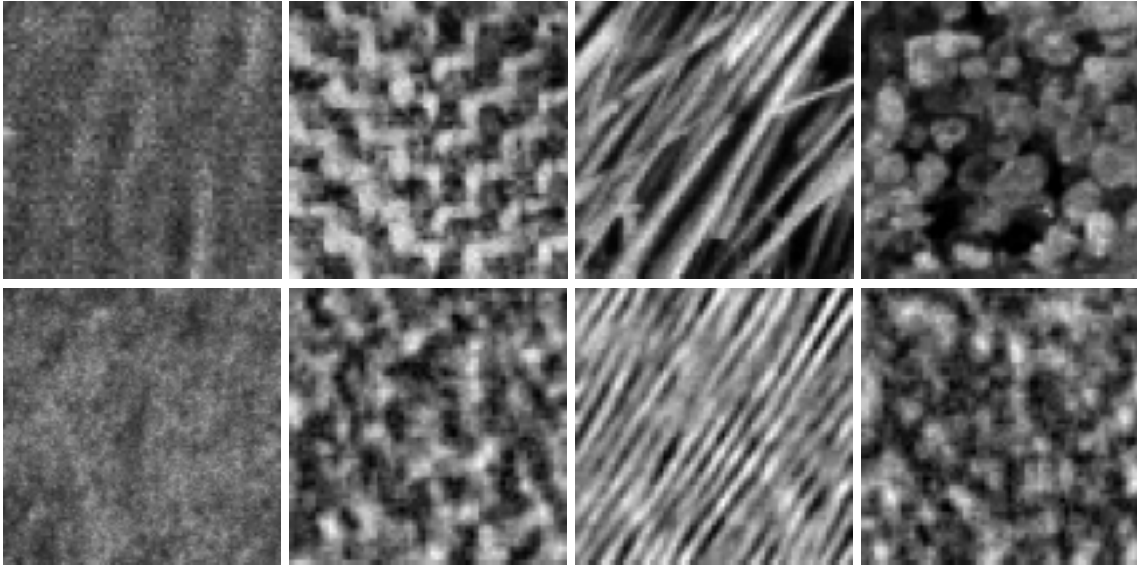


Figure 4.5: Analysis and synthesis using Gauss-Markov random fields of fifth neighborhood order. Top row: Original Brodatz textures (water, herringbone weave, straw and beach sand, 64×64 pixels). Bottom row: Synthesized textures generated with the parameters extracted from the original data. The estimated parameters are given later in Tab. 4.3 and in Fig. 4.7.

although equal values for all parameters $\theta_{k,l}$ represent a reasonable solution. Hence, stable and robust estimates cannot be always guaranteed with this method.

Thanks to the analytical solution, parameter estimation for arbitrary neighborhood orders is straightforward and fast. An example of image analysis and synthesis for Gauss-Markov random fields is given in Fig. 4.5. The model parameters of a fifth order neighborhood have been extracted from the original texture images and have been used to regenerate similar textures. The values of the six most significant parameters are enumerated later in Tab. 4.3 and the full parameter vectors are displayed further below in Fig. 4.7.

In the presented examples, parameters and realizations of the second and the fourth texture are rather similar. This is because the "weave" texture is not stationary concerning the directional information. In such cases, the model parameters tend to show isotropic smoothing characteristics, which is partly visible in the synthesized image. This is an important property, since textures that cannot be sufficiently explained by the model will be simply smoothed in the image restoration process. Such textures will not be filtered with random parameters, which would result in strong artefacts.

Furthermore, the sampled textures only approximate the properties of the original ones. However, a highly accurate description of the image contents for re-synthesis by a parametric model is not necessary if the goal is image restoration. In our case, enough additional information about the noise free image x is encapsulated in the noisy observation y by means of the likelihood term.

Coding method

The coding method was introduced in [Besag74] and is basically a maximum likelihood estimation method providing parameters that maximize the conditional probability of a subset of the random field, the so-called coding, which is statistically independent from other codings. This method requires the solution of a set of non-linear equations, which can be obtained by gradient methods. However, the coding method showed not to be very reliable in experimental results. Moreover, an established method to combine the different parameter estimates from different codings into one single parameter set does not exist.

Stochastic gradient algorithm for complete data

Another parameter estimation approach that we want to mention starts from the full global probability of the Gibbs random field, which is written as

$$p(X = x|\lambda) = \frac{1}{Z} \exp(-\lambda U(x)) \quad (4.25)$$

and results in an iterative updating scheme for the hyper-parameter λ to be estimated. An extended approach should work for a full parameter vector θ of the local energy function. The function $U(x)$ is the potential function without the free parameter λ . The estimate of λ at iteration $n + 1$ is then given by [Younes88]

$$\lambda^{(n+1)} = \lambda^{(n)} + \frac{U(x^{(n)}) - U(x)}{(n+1)V}. \quad (4.26)$$

We do not want to go too deeply into this method, but intend to illustrate its different quality: The denominator in the above equation is an approximation of the variance of $U(x)$, but what is more interesting is the term $U(x^{(n)})$, which is the potential of the sampled field $x^{(n)}$ at iteration n with parameter $\lambda^{(n)}$. Thus, starting with a random parameter $\lambda^{(0)}$, a configuration $x^{(0)}$ is sampled, λ is updated and the whole procedure is re-iterated. Note that $x^{(n)}$ is used as a starting point to generate the next sample $x^{(n+1)}$.

Although we did not investigate this method, we find it worth to be briefly pointed out here. The idea of a stochastic gradient with sampling from the distribution under investigation represents a different class of approach. However, it can be supposed that convergence is rather slow and that computation times are quite high, due to the required sampling from the Gibbs distribution.

Model selection for complete data

Up to now, we considered the case of parameter estimation from noise-free data for a given model M . Of course, it is possible not only to fit the model to the data by estimation of its parameters, but also to select the model that best explains the data by means of the evidence. Writing the full Bayesian equation

$$p(\theta|X = x, M) = \frac{p(X = x|\theta, M)p(\theta|M)}{p(X = x|M)}, \quad (4.27)$$

we see that $\boldsymbol{\theta}$ is estimated for a particular model M and that the likelihood of that model is described by the evidence term

$$p(X = x|M) = \int p(X = x|\boldsymbol{\theta}, M)p(\boldsymbol{\theta}|M)d\boldsymbol{\theta}, \quad (4.28)$$

where the integration is performed over the whole space of $\boldsymbol{\theta}$. Approximations of this possibly multi-dimensional integral can be made and, in principle, the evidences for the Gauss-Markov and a discontinuity adaptive model can be evaluated.

We do not perform this model comparison here because, on one hand, the computation of the evidence for non-linear models is rather difficult and, on the other hand, we only intend to point out the model selection for noise-free data in order to compare it to the case of noisy data as presented below.

4.3.2 Model Parameter Estimation from Incomplete Data

For SAR image restoration we do not have the favorable case of complete data, but must estimate the required model parameters from noisy, i.e. incomplete, data. We want to extract knowledge about X from the observation y . Consequently, we are unable to compute the required parameters by maximizing $p(X = x|\boldsymbol{\theta})$. To better illustrate the problem, we write Bayes' equation used for restoration

$$p(X = x|Y = y, \boldsymbol{\theta}, M) = \frac{p(Y = y|X = x, \boldsymbol{\theta}, M)p(X = x|\boldsymbol{\theta}, M)}{p(Y = y|\boldsymbol{\theta}, M)} \quad (4.29)$$

and find for the evidence term

$$p(Y = y|\boldsymbol{\theta}, M) = \int p(Y = y|X = x, \boldsymbol{\theta}, M)p(X = x|\boldsymbol{\theta}, M)dx, \quad (4.30)$$

which results from the integration of the posterior product over the whole space of X . In contrast to the noise-free case, we do not separate the model M and its parameters, but treat them together by maximizing the evidence $p(Y = y|\boldsymbol{\theta}, M)$ as a function of $\boldsymbol{\theta}$ for a fixed model M . This can be interpreted as a decomposition of M into several sub-models each characterized by a different parameter vector $\boldsymbol{\theta}$. Of course, this scheme can be performed for several models M , thus enabling the selection of the best model together with its parameters. A model selection of the type $p(Y = y|M)$, regardless of its parameters, requires another probabilistic layer and is not suited for practical application. We illustrated before that the equivalent computation of $p(X = x|M)$ for the noise-free case is possible, because this additional noise-layer is missing.

In the following, we neglect M and estimate parameters for a fixed model. Several methods exist to solve the problem of estimating $\boldsymbol{\theta}$ out of the observation y .

Markov chain Monte Carlo (MCMC)

To solve the problem of parameter estimation from incomplete data, the maximization of the evidence integral in Eq. 4.30 is necessary. The computation of this integral, however, is not trivial and in the general case not feasible.

To approximate the evidence for a given parameter set θ , a stochastic scheme was proposed [Linden, Skilling]. In fact, what is needed is the histogram $p(Y = y|X = x, \theta)p(X = x|\theta)$ in order to calculate the evidence, which is its normalization constant. This histogram can be built by sampling from the posterior product for fixed values of θ and, thus, allows an approximation of the evidence term.

It is clear that this approach is computationally very heavy and converges only slowly. Moreover, it has to be performed for a large number of values of θ to find the optimum. Besides the fact that full statistical certainty is never achieved, as only a finite number of samples is drawn, we consider this approach unsuitable for our goal of parameter estimation for SAR image restoration and information extraction, for the practical reason of computation time. However, for many cases this Monte Carlo method may be the only solution since it can always provide the sought inference about the model parameters.

Expectation maximization (EM) algorithm

A method, which is frequently used for the estimation of hidden Markov parameters, especially in speech processing, is the expectation maximization (EM) algorithm [Dempster, EM]. We define a function

$$Q(\hat{\theta}, \hat{\theta}^{(n)}) = \int \log(p(\hat{\theta}|X = x, Y = y)) p(X = x|Y = y, \hat{\theta}^{(n)}) dx, \quad (4.31)$$

which is described by the parameter vector $\hat{\theta}$ to be estimated and its current guess $\hat{\theta}^{(n)}$ obtained at iteration n . By iteratively maximizing this function, a value for $\hat{\theta}$ which represents the global or a local maximum of the evidence can be found. The maximization is done in two steps:

1. E-Step: Compute $Q(\hat{\theta}, \hat{\theta}^{(n)})$ as a function of $\hat{\theta}$ using the current guess $\hat{\theta}^{(n)}$. The integration has to be performed over the whole space of X .
2. M-Step: Maximize the Q function with respect to $\hat{\theta}$ in order to obtain a new parameter vector update $\hat{\theta}^{(n+1)}$ and re-iterate.

It was proven by [Dempster] that this full EM algorithm always converges. But it cannot be assured that the global maximum is reached. Nonetheless, in our case, the integration over the space of X is even more challenging than the computation of the evidence by itself. Hence, the EM algorithm in this form is of no particular use for our problem. However, the Q function can be simplified if the posterior $p(X = x|Y = y, \hat{\theta}^{(n)})$ is very peaked and approximates a Dirac function in the limit. For this case, the EM algorithm reduces to:

1. E-Step: Compute the MAP estimate \hat{x}_{MAP} of the data using the observation y and the current estimate $\theta^{(n)}$ of the parameter vector by maximizing the posterior $p(X = x|Y = y, \hat{\theta}^{(n)})$.

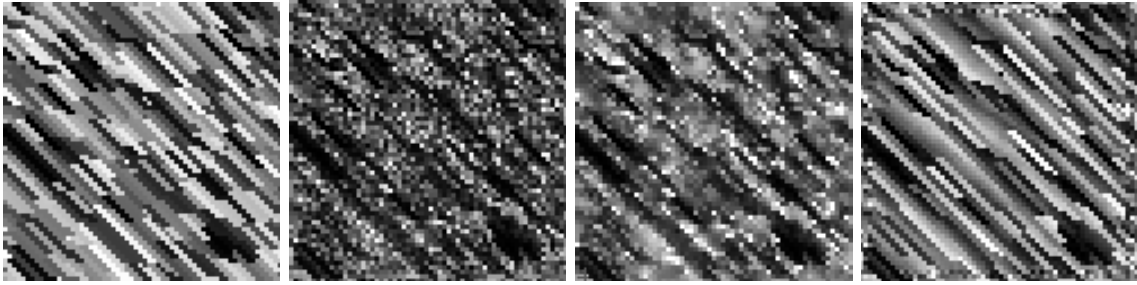


Figure 4.6: Texture restoration using Gibbs random fields. From left to right: Original data (64×64 pixels) sampled from a discontinuity adaptive Gibbs model, noisy texture with single look speckle, Kuan filtered image, restored texture using the same Gibbs model as for synthesis of the original data. The model parameters used for restoration have been iteratively estimated from the speckled data with the EM-type algorithm.

2. M-Step: Use the estimated value \hat{x}_{MAP} as the actually measured data to determine the ML estimate of θ by maximizing $p(X = \hat{x}_{MAP}|\theta)$ and re-iterate until convergence is reached.

In the following, we refer to this technique as the EM-type algorithm [Moon, Wei], since it is a reduced version of the original EM approach. In this method, to obtain θ the computation of the MAP estimate \hat{x}_{MAP} is needed for the E-step, and the parameter estimation for complete data (e.g. MPL estimation) is needed for the M-step. For both steps we have pointed out suitable algorithms.

As an illustration, we present a qualitative example for image restoration using non-linear models in Fig. 4.6. A synthetic texture was generated, degraded by speckle, and has been filtered using Markov random fields with the presented EM-type algorithm for model parameter estimation. For comparison, the same image has also been filtered with the Kuan filter. In contrast to the Kuan filtered image, which shows only reduced similarity to the original data, the restored image exhibits clear similar textural properties. As a drawback, computing times for the used non-linear model are too high to be suitable for practical application due to the computationally demanding MPL parameter estimation. We also note that the original texture is a direct realization of the employed prior model, which allows a better restoration. For real textures, we have only rarely observed convergence to a reliable reconstruction, since the simplified EM-type algorithm may not always converge to a good estimate of θ .

Stochastic gradient algorithm for incomplete data

To complete this overview, we finally note that a similar stochastic gradient algorithm exists for parameter estimation from incomplete data [Younes89, Younes91], as for the noise-free case. Since the same restrictions apply as for complete data, no tests have been done to evaluate its performance.

4.4 Estimation of Hidden Gauss-Markov Model Parameters

In the following, we limit ourselves to the case of linear Gauss-Markov random fields. The principal capabilities of non-linear models for information extraction and image restoration have been demonstrated, but the required computations are by far too time consuming for practical applications. As of this, linear models, such as the Gauss-Markov model, are preferred in spite of their limited modeling capabilities, since they are much easier to deal with in the Bayesian framework. For the case of complete data, parameter estimation for the Gauss-Markov model is straightforward and can be done analytically. Hence, a similarly reduced computational complexity of parameter estimation for speckled data can be suspected.

4.4.1 Gauss-Markov Model Parameter Estimation from Speckled Data

We address three possible approaches for parameter estimation from incomplete data and evaluate their performance concerning the goal of SAR image restoration. Using synthetic data, examples will be given for the estimated parameters extracted from assumed stationary textures and for image restoration of a non-stationary scene requiring small estimation windows. Consequently, two criteria are important for the selection of an appropriate method: Quality of parameter estimation and behavior in non-stationary conditions with few data, i.e. stability and robustness of the estimator.

Parameter estimation from speckled data by expectation maximization

The EM-type approach is easily applicable in the Gauss-Markov case. Both the computation of the MAP estimate and the maximum pseudo-likelihood parameter estimation are very fast to perform, the former because of the analytically given partition function, the latter because of the analytical solution for the parameter estimates by a simple matrix inversion. However, results might not be optimal due to the possibility of getting trapped in local extrema or because of the made simplifications in the derivation of the EM-type algorithm, which neglects information from solutions in the vicinity of the actual MAP estimate (assumption of a peaked posterior).

Analytical parameter estimation from speckled data

The derivation of the maximum pseudo-likelihood estimate for Gauss-Markov random fields showed that the model parameters are completely determined by the correlation coefficients of all pixels within the considered neighborhood. Consequently, estimates of the matrix \mathbf{A} and of the vector \mathbf{b} obtained from noisy data can be used to compute the parameter vector $\hat{\boldsymbol{\theta}}$, which describes the statistical properties of the noise-free image x in exactly the same way as for complete data using Eq. 4.21.

Under the assumption of multiplicative speckle noise n , i.e. $y = x \cdot n$, and for square-root intensity images, our estimation of the elements of \mathbf{A} and \mathbf{b} directly yields the following

relations

$$E\{x_{0,1} \cdot (x_{a,b} + x'_{a,b})\} = \frac{E\{y_{0,1} \cdot (y_{a,b} + y'_{a,b})\}}{E\{n\}^2}, \quad (4.32)$$

$$E\{(x_{a,b} + x'_{a,b}) \cdot (x_{a,b} + x'_{a,b})\} = E\{y_{a,b}^2\} + E\{y'_{a,b}{}^2\} + 2 \frac{E\{y_{a,b} \cdot y'_{a,b}\}}{E\{n\}^2} \quad (4.33)$$

and

$$E\{(x_{a,b} + x'_{a,b}) \cdot (x_{c,d} + x'_{c,d})\} = \frac{E\{(y_{a,b} + y'_{a,b}) \cdot (y_{c,d} + y'_{c,d})\}}{E\{n\}^2}, \quad (4.34)$$

where the mean of the noise process $E\{n\}$, which was given in Eq. 2.37, depends on L , i.e. on the equivalent number of looks (ENL). In this derivation, we used $E\{n_{a,b} \cdot n_{c,d}\} = E\{n\}^2$ for uncorrelated speckle and $E\{n^2\} = 1$.

As for complete data, an analytical estimate can also be derived for the variance of the model prediction σ^2 in the noisy case. We find

$$\begin{aligned} \sigma^2 &= E\{(x_i - \mu_i)^2\} \\ &= E\{x_i^2\} - 2E\{x_i \mu_i\} + E\{\mu_i^2\} \\ &= E\{x_i^2\} - 2E\left\{x_i \cdot \sum_{k,l} \theta_{k,l} \cdot (x_{k,l} + x'_{k,l})\right\} + E\left\{\left(\sum_{k,l} \theta_{k,l} \cdot (x_{k,l} + x'_{k,l})\right)^2\right\}, \end{aligned} \quad (4.35)$$

which can be expressed as a function of the expectation values of the noisy data using

$$E\{x_i^2\} = \frac{E\{y_i^2\}}{E\{n^2\}} = E\{y_i^2\} \quad (4.36)$$

and the equations Eqs. 4.32 - 4.34 derived before.

Computationally, the estimation of Gauss-Markov parameters from speckled data with this method is as fast as the parameter estimation from complete data, i.e. almost instantaneous in comparison with other methods. Moreover, the estimation accuracy can be assumed to be very good depending, of course, on the quality of the estimates for \mathbf{A} and \mathbf{b} , which is determined by the estimation window size. Concerning the solution of the resulting set of linear equations, the same restrictions in terms of stability as for the noise-free case apply.

Parameter estimation by iterative evidence maximization (IEM)

Another approach for Gauss-Markov parameter estimation from speckled data is the iterative maximization of an approximation of the evidence $p(Y = y|\boldsymbol{\theta}, M)$ [Datcu98, Walessa99a]. We mentioned that the computation of the evidence integral is usually very difficult and rather slow, if performed e.g. by MCMC. However, if the integrand is composed of two unimodal distributions, e.g. a square-root Gamma and a Gaussian distribution as in our case, the resulting evidence can be approximated as shortly hinted at in chapter three. Being able to numerically evaluate $p(Y = y|\boldsymbol{\theta}, M)$ within reasonable time, an iterative optimization technique can be employed for its maximization. Before discussing this method in detail in the following chapter, we will first compare its performance to the other two presented approaches.

Beach sand	σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{3,1}$	$\theta_{3,2}$
Original	8.489	0.4048	0.3933	-0.08244	-0.04275	-0.0688	-0.07476
Noisy	23.27	0.1711	0.1758	0.07485	0.09635	0.01614	0.01557
EM-type	3.138	0.2519	0.1739	0.1713	0.1577	-0.04427	0.1403
Analytical	8.919	0.3317	0.3844	-0.008458	0.0009847	0.005694	-0.07582
IEM	9.302	0.2166	0.2139	0.09477	0.1139	0.0075	0.002045
Water	σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{3,1}$	$\theta_{3,2}$
Original	14.28	0.1604	0.1572	0.07497	0.04989	0.004371	0.03087
Noisy	33.12	0.06279	0.07801	0.05794	0.04945	0.02556	0.04952
EM-type	4.453	0.1409	0.04386	0.08708	0.07772	-0.01477	-0.006031
Analytical	14.39	0.1482	0.1327	0.1176	0.05179	0.01128	0.02759
IEM	14.8	0.06455	0.07545	0.05455	0.06	0.02182	0.06273
Straw	σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{3,1}$	$\theta_{3,2}$
Original	7.629	0.1719	0.3886	-0.07157	0.1926	-0.02289	-0.1167
Noisy	25.04	0.1038	0.1527	0.009229	0.1301	-0.0267	0.01916
EM-type	3.454	0.08539	0.1551	-0.03883	0.3754	-0.02549	0.02944
Analytical	8.271	0.132	0.3923	0.002109	0.1684	0.07316	-0.147
IEM	5.818	0.1511	0.233	-0.008864	0.193	-0.04432	-0.02523
Weave	σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{3,1}$	$\theta_{3,2}$
Original	12.05	0.3673	0.4307	-0.1035	-0.05603	-0.03876	-0.05532
Noisy	34.76	0.1513	0.1782	0.07711	0.08753	0.02788	0.03775
EM-type	4.992	0.2176	0.2147	0.05355	0.04623	0.1036	0.07762
Analytical	8.271	0.132	0.3923	0.002109	0.1684	0.07316	-0.147
IEM	14.7	0.195	0.2086	0.09227	0.1086	0.02136	0.03227

Table 4.3: Gauss-Markov random field parameters estimated for the four Brodatz textures. The six most significant directional parameters of a fifth order neighborhood system used for estimation are displayed. Parameters have been estimated under the assumption of complete data from both original and noisy data using the MPL approach. The EM-type algorithm, the analytical estimation technique and the iterative evidence maximization (IEM) method have been used to estimate parameters from incomplete data.

4.4.2 Evaluation of Estimation Techniques for Gauss-Markov Parameters

To evaluate the quality of the estimated parameters obtained from different techniques, we use the four Brodatz textures "beach sand", "water", "straw" and "herringbone weave" (Fig. 4.8). Employing the analytical MPL method, we estimate reference model parameters for each noiseless texture using an estimation window of the size of the texture (256×256 pixels), which guarantees reliable estimates. For that purpose, we assume the textures to be stationary as far as their spatial properties are concerned. A fifth order neighborhood system is assumed to capture the complexity of the different image contents resulting in

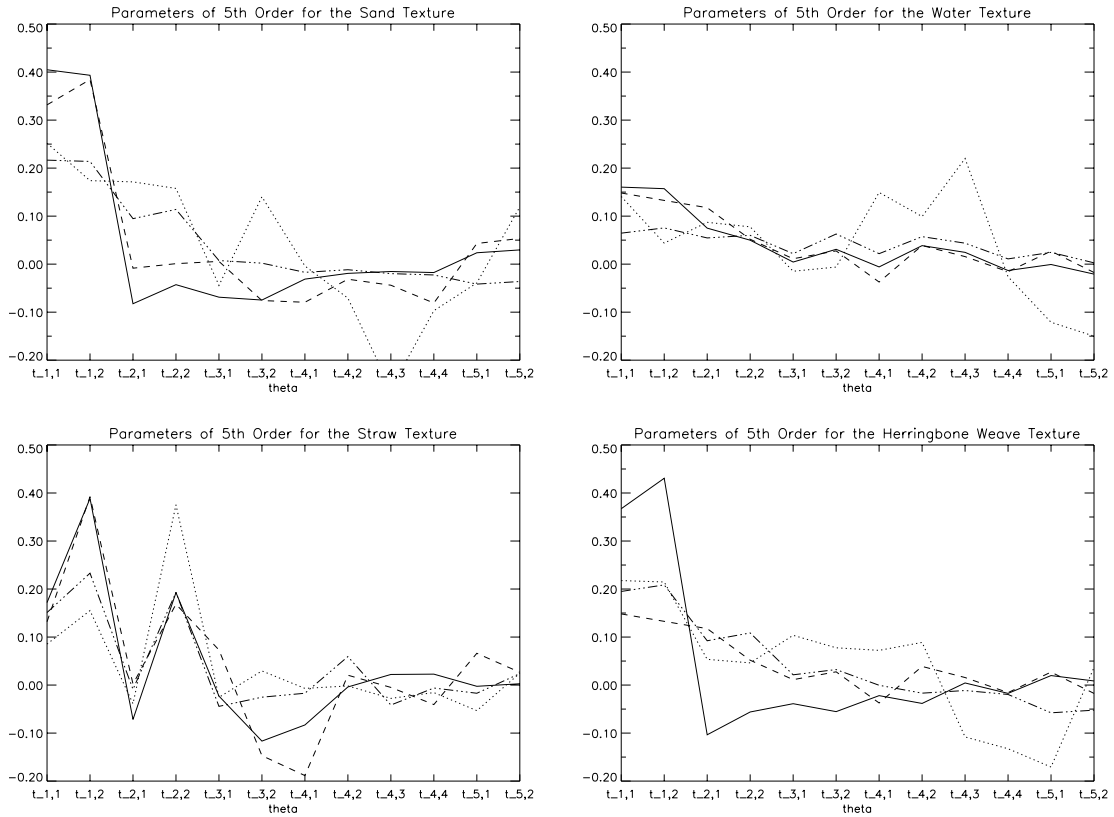


Figure 4.7: Plots of estimated directional parameters $\theta_{k,l}$ extracted with different algorithms. Results are given for the four Brodatz textures. Parameters are displayed for the original texture (solid) and for the estimates obtained by the EM-type algorithm (dotted), by the analytical method (dashed) and by IEM (dotted-dashed). To reduce complexity of the plots the MPL estimates from noisy data are omitted.

twelve directional parameters $\theta_{k,l}$ and the standard deviation σ for the Gauss-Markov model. For parameter estimation from incomplete data, we added speckle noise with $L = 3$ and applied the presented parameter estimation techniques.

Another experiment was made to verify the performance of the estimators on non-stationary images. We generated a synthetic image and restored it with locally adaptive versions of the algorithms under study (Fig. 4.9). Small sliding estimation windows (typically 16×16 pixels) are used to obtain local parameter estimates, which are required for the computation of the MAP estimates. The results of all these tests are depicted in Figs. 4.7, 4.8, 4.9 and the corresponding Tabs. 4.3 and 4.4.

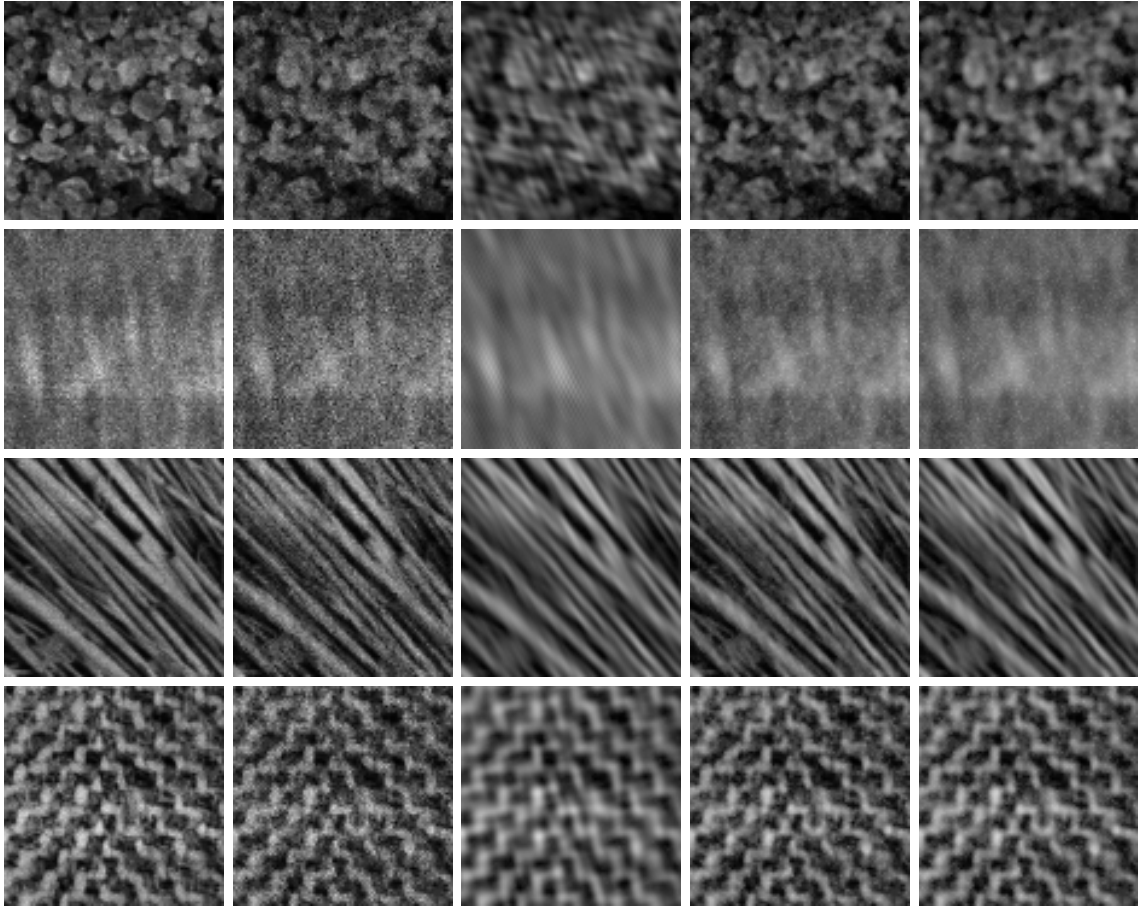


Figure 4.8: Restoration of speckled textures ($L = 3$) using Gauss-Markov random fields and different parameter estimation techniques. Columns from left to right: Original textures without noise, restoration of speckled textures with MPL-estimated parameters from noisy data, restoration obtained with EM-type parameter estimation, with analytical parameter estimation and with iterative evidence maximization (IEM).

Parameter estimation from speckled data by MPL estimation for complete data

In addition to the presented methods for parameter estimation from incomplete data, we verified the performance of the MPL approach for complete data in the case of model parameter estimation from speckled images and found that normal MPL estimation of the Gauss-Markov (GMRF) parameters without taking into account the noise properties yields very poor results. The estimates, which are given in Tab. 4.3, provide some information about the textural activity of the data, but surely cannot be satisfactorily used to discriminate different textures of the cross-section. For image restoration, these results are of no use at all. Consequently, an example of locally adaptive image restoration has not been computed.

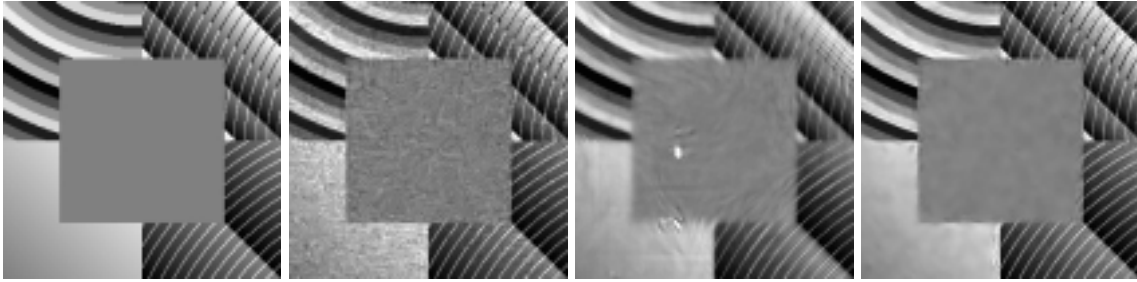


Figure 4.9: Despeckling results from a synthetically speckled image ($L = 3$, 256×256 pixels) using different techniques for local parameter estimation of a fifth order neighborhood. From left to right: Original synthetic image, restoration of the speckled image with the EM-type algorithm (window size 16×16 , $\text{MSE}=425$), result of the analytical estimation method (window size 64×64 , $\text{MSE}=412$) and result obtained with parameter estimation by IEM (window size 16×16 , $\text{MSE}=179$).

Parameter estimation from speckled data by expectation maximization

For the EM-type approach, a fixed number of 25 iterations was used. The estimated Gauss-Markov parameters capture the basic textural properties like different directional preferences, but do not convince in terms of their error to the "true" parameters. A striking characteristic is the value estimated for σ , which is very low and results in a rather smooth restoration of the image, as depicted in Fig. 4.8. Concerning the mean square error (MSE) of the restoration and the evidence of the estimated parameters shown in Tab. 4.4, results are sometimes even worse than those obtained by filtering with parameters extracted without considering speckle. This is not acceptable for SAR image despeckling. The same conclusion can be drawn from the result of the locally adaptive filtering, as illustrated in Fig. 4.9. Many artefacts, which result from the instability of the parameter estimates, are visible in the filtered image. Consequently, the EM-type approach cannot be considered to be suitable for parameter estimation from speckled data.

Analytical parameter estimation from speckled data

The analytical parameter estimation from speckled data yields very fast and very good estimates. The similarity of the estimated parameters with the reference parameters of the noise-free textures, as illustrated in Fig. 4.7, is impressive. Apparently, this makes this approach the first choice for information extraction from noisy data. The four restored textures are of high quality, with mean square errors significantly lower than for the EM-type algorithm. However, concerning the locally adaptive image filtering, inverse conclusions have to be drawn. The estimated parameters show to be extremely unstable for small window sizes. We increased the initial window size of 16×16 pixels to 64×64 pixels in order to ensure stable parameters throughout most of the image, although some artefacts remained visible in Fig. 4.9. However, large estimation windows are not acceptable for image

MSE	Sand	Water	Straw	Weave
Noisy	189	418	188	409
EM-type	198	256	198	435
Analytical	133	220	137	263
IEM	135	217	136	251
Log-Evidence	Sand	Water	Straw	Weave
Noisy	-4.569	-4.966	-4.692	-4.974
EM-type	-4.623	-5.020	-4.537	-5.034
Analytical	-4.476	-4.867	-4.473	-4.880
IEM	-4.446	-4.863	-4.495	-4.828

Table 4.4: Quality of image restoration and texture parameters. Upper table: Mean square error between original texture and restored image obtained by different parameter estimation techniques. "Noisy" denotes MPL estimation of the model parameters for complete data, i.e. from the speckled data without considering noise. Lower table: Normalized log-evidence $p(Y = y|\boldsymbol{\theta}, M)$ of the parameters estimated with the presented algorithms. Both the analytical method and IEM yield similar excellent results. For numerical computation of $p(Y = y|\boldsymbol{\theta}, M)$ approximations made in the IEM algorithm have been used.

restoration due to the inherent non-stationary behavior of SAR images. As a consequence, small features and textures are strongly smoothed, as shown in Fig. 4.9 and as reflected in the mean square error. The possible instability becomes especially apparent in areas of low textural activity.

This disappointing result can be explained by considering the linear system of the matrix \mathbf{A} and the vector \mathbf{b} . The estimation of $k = 12$ directional model parameters for a fifth order neighborhood requires the computation of $k(k+3)/2 = 90$ empirical expectation values from the speckled data. To guarantee stability, these estimates must possess a certain accuracy, which increases with the estimation window size. By sequentially increasing the assumed value of the noise mean $E\{n\}$ in the algorithm up to $E\{n\} = 1$, which is equivalent to the normal MPL estimate without considering noise, stability grows, but the parameter estimates are no more accurate. Robustness is also limited due to the required matrix inversion. In regions of low textural activity, i.e. with similar correlation in all directions, \mathbf{A} cannot be inverted any more or its inverse matrix becomes unstable, resulting in parameters with values far out of the nominal range. A comparative example is illustrated in Tab. 4.5. We see that the analytical estimation yields extremely unstable parameters with σ not defined (NaN). However, the sum $\sum_{k,l} \theta_{k,l}$ of all twelve parameters yields 0.5, i.e. the correct normalization, as expected.

We conclude that, although computationally simple and very accurate for stationary textures with a certain activity, the analytical approach is not suited for locally adaptive image restoration in combination with information extraction.

Flat Cross-Section	σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{3,1}$	$\theta_{3,2}$
Analytical	NaN	-0.686	-4.677	1.413	0.3028	-1.664	1.668
IEM	11.01	0.04545	0.04545	0.04091	0.04091	0.04091	0.04091

Table 4.5: Gauss-Markov random field parameters estimated for a uniform cross-section (256×256 pixels with mean $\mu_x = 100$) degraded by speckle with $L = 3$. The six most significant directional parameters of a fifth order neighborhood system used for estimation and the standard deviation σ are displayed. The potential instability of the analytical solution is reflected in $\sigma = NaN$.

Parameter estimation by iterative evidence maximization

The results obtained by the iterative maximization of the approximated evidence are of almost equal quality compared to those of the analytical estimator, as can be seen from Fig. 4.7. The quality of the restored textures can be judged to be equal by visual evaluation and in terms of the mean square error. However, the iterative evidence maximization approach is about two orders of magnitudes slower. Its main advantages compared to the analytical approach are its robustness and the stability imposed by the iterative optimization. Comparing the examples of Fig. 4.9, by far the best results, both visually and with regard to the mean square error, are obtained. The filtered image is almost free of artefacts and shows a good restoration of fine structures. Artefacts occur mainly at transitions between areas of different textural properties, i.e. at strong non-stationarities. This may require a special processing at borders, which could probably be partly avoided by the use of discontinuity adaptive models. Concerning the estimation for uniform cross-sections (Tab. 4.5) the twelve iteratively estimated parameters show to be very stable and closely take their expected values of $0.5/12 = 0.0416$. The value of σ reflects the remaining uncertainty imposed by the speckle noise.

Conclusions

We deduce from the above experiments that only the last of the three presented approaches for Gauss-Markov model parameter estimation from incomplete data is suitable to meet the conditions necessary for information extraction from SAR data and for speckle filtering: Accurate and robust parameter estimation from few data. In the next chapter, the iterative evidence maximization (IEM) approach will be presented in detail and its application to real SAR data will be demonstrated, an issue which was not yet addressed.

4.5 Summary

In this chapter, the following points have been discussed:

- Several commonly used despeckling filters have been presented. We discussed their qualitative differences in terms of modeling of the noise and of the noise-free cross-

section. Generally, the quality of the despeckled data increases with model complexity. Almost all approaches rely on the assumption of uncorrelated noise and require some pre-processing of the data to meet this condition.

- Complex prior assumptions require iterative relaxation methods for the maximization of the posterior. With regard to the optimization space, either slow optimal stochastic algorithms like simulated annealing or faster sub-optimal optimization schemes like the ICM algorithm are to be used.
- Regardless of the kind of parametric prior, optimal filtering results can only be achieved if the model parameters are estimated from the data. Fixed parametric models cannot be considered to be satisfactory since they introduce a strong bias and may even further degrade the noisy observation.
- Methods for information extraction from noise-free data have been presented. We showed that the computational complexity of the estimation mainly depends on the used model, either non-linear or linear. In principle, non-linear models possess higher modeling capabilities than linear ones. However, non-linear models are difficult to handle making parameter estimation too slow for practical applications.
- In the case of incomplete data, parameter estimation becomes more challenging. Several methods have been presented. In the following, we limited ourselves to the linear Gauss-Markov model for practical reasons and discussed several parameter estimation techniques. A quantitative comparison under different conditions clearly favored one of the tested approaches.
- The different steps leading to the choice of Gauss-Markov random fields for information extraction and image restoration have been illustrated. Based on these results, we develop a Bayesian algorithm for SAR image filtering and feature extraction in the following chapter. However, due to the known shortcomings of the Gauss-Markov model, an additional processing, which is discussed in chapter six, will be necessary.

5 Model-Based Despeckling with Gauss-Markov Random Fields

Especially for high-resolution SAR imagery it is important to preserve textural information while removing speckle noise, in order to simplify a subsequent image interpretation. To achieve this goal, we use a Bayesian approach for image restoration relying on texture models. In the preceding chapters, we have demonstrated that this is best done with linear models, such as Gauss-Markov random fields, due to their better tractability for computations. Non-linear models may possess additional qualities but are very difficult to handle.

We calculate the MAP estimate of the noise-free image using adaptive model parameters. These parameters must be estimated by taking into account the characteristics of the noise which affects the data, making this a difficult and sometimes not feasible task. A number of estimation techniques for incomplete data have been presented in the last chapter, among which we have chosen the IEM approach as the most suitable one. This estimation technique has not been explained yet and will be the main subject of the following sections.

In this chapter, all relevant issues concerning filtering and information extraction from SAR data will be addressed. We start with a discussion on the computation of the MAP estimate, which will lead to a justified approximation of the evidence term allowing its evaluation. After this, the basic algorithm for parameter estimation and filtering will be outlined.

A subsequent analysis will address a possible bias of the resulting estimator and contains an error analysis of the filtered image. In the last section, we discuss several additional issues like the selection of the neighborhood order, the choice of the estimation window size, an interpretation of the evidence and the properties of the Gauss-Markov model in different signal spaces, i.e. filtering of logarithmic and intensity data. Proposed solutions for identified drawbacks of the introduced approach will be discussed in chapter six.

5.1 MAP SAR Image Estimation with a Gauss-Markov Prior

The observed noise-free radar cross-section x of the random process X that we want to estimate from a given observation y of a process Y is affected by the realization n of the speckle noise process N in a multiplicative way. We have already seen in chapter two that

a multiplicative noise is characteristic for all coherent imaging systems and usually is of very high variance compared to the mean image gray-values. As in the preceding chapter, we consider the probability density function of the likelihood of the observed square-root intensity image $y = x \cdot n$, which is given at site i analogously to Eq. 2.36 by

$$p(Y_i = y_i | X_i = x_i) = 2 \left(\frac{y_i}{x_i} \right)^{2L-1} \frac{L^L}{x_i \Gamma(L)} \exp \left(-L \left(\frac{y_i}{x_i} \right)^2 \right). \quad (5.1)$$

Discussions of the behavior of the Gauss-Markov model for intensity or log-intensity data will follow in a later section.

In order to preserve texture and to extract textural information, we use Gauss-Markov random fields (GMRFs), which represent an auto-regressive process, as prior information. This choice was motivated by the Gauss-Markov model's analytical tractability making it interesting for practical application. Transforming the auto-regressive process into a corresponding pdf yields the already known distribution (cf. Eq. 3.66) for the stochastic Gauss-Markov random field model of X :

$$p(X_i = x_i | X_j = x_j, j \in \mathcal{N}_i, \boldsymbol{\theta}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x_i - \mu_i)^2}{2\sigma^2} \right) \quad (5.2)$$

We remind that μ_i is a linear combination of gray-values in a certain neighborhood around x_i , which are weighted by the elements of the model parameter vector $\boldsymbol{\theta}$:

$$\mu_i = \sum_{k,l} \theta_{k,l} \cdot (x_{k,l} + x'_{k,l}) \quad (5.3)$$

5.1.1 Fast Computation of the MAP Estimate

Knowing the local representations of the likelihood function and the prior Gibbs random field, as well as its parameters, a MAP estimate can be computed that maximizes the global energy function of the posterior field, as seen in chapter three. Usually, stochastic relaxation methods are required for this optimization task. In our case, however, it is advisory to use a deterministic algorithm, like ICM (cf. appendix B), since convergence to the final solution is much faster and we are faced with a convex optimization space, as we will demonstrate in a later section.

ICM solution for a Gaussian-shaped prior

At each iteration, the ICM algorithm requires the computation of the pixel values that maximize the local posterior distribution at each site, i.e. each pixel x_i is updated with its most probable value. Unlike for non-linear models, this computation can be performed analytically for the posterior product of the square-root Gamma and the Gaussian distribution. We have to maximize

$$p(X_i = x_i | Y_i = y_i, \boldsymbol{\theta}) \propto p(Y_i = y_i | X_i = x_i, \boldsymbol{\theta}) p(X_i = x_i | \boldsymbol{\theta}), \quad (5.4)$$

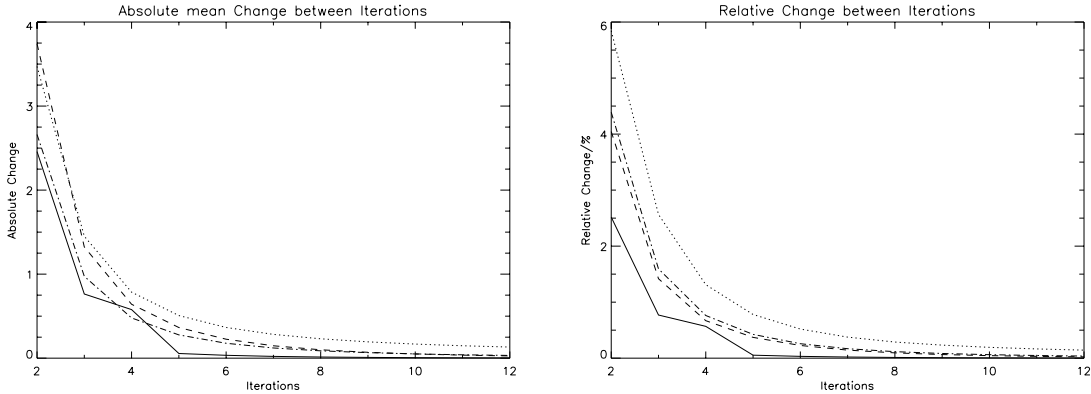


Figure 5.1: Convergence of the ICM (Iterative Conditional Modes) algorithm using the Gauss-Markov prior. Left: Mean absolute change $|x^{(n)} - x^{(n-1)}|$ of gray-values between to subsequent iterations n and $n - 1$. Right: Relative change $|x^{(n)} - x^{(n-1)}|/x^{(n-1)}$ between two iterations. For visualization we show only the range starting from iteration two. The plots are given for the filtering of the "sand" (dot-dashed), the "water" (solid), the "straw" (dotted) and the "weave" (dashed) texture. The "water" texture, which exhibits no dominant structures, shows the fastest convergence.

as a function of x_i . This can be achieved by setting the first derivative of the product to zero. The logarithm may be used here for simplification since it is a monotonic function that does not change the location of any maxima. Consequently, we find what we call the MAP equation:

$$\frac{\partial}{\partial x_i} \log p(X_i = x_i | Y_i = y_i, \boldsymbol{\theta}) = -\frac{2L}{x_i} + \frac{2Ly_i^2}{x_i^3} - \frac{x_i - \mu_i}{\sigma^2} = 0 \quad (5.5)$$

For the proposed model and the presented likelihood function we obtain as a final result a fourth order polynomial which yields four solutions for x_i depending on the model parameter vector $\boldsymbol{\theta}$ including σ , the observed data y_i and the current neighborhood configuration of x_i characterized by μ_i :

$$x_i^4 - \mu_i x_i^3 + 2L\sigma^2 x_i^2 - 2L\sigma^2 y_i^2 = 0 \quad (5.6)$$

A simple case study of the possibly complex-valued roots gives the desired solution for the maximum of the local posterior. Convergence of the ICM algorithm is usually reached after a few iterations. In practice, we limited ourselves to a maximum of ten iterations, which provides very good results. This choice is confirmed by a study of the changes from one iteration to the next, as depicted in Fig. 5.1. After five iterations, visual changes are no more perceivable. Full convergence is usually achieved after 25 iterations.

Unfortunately, the computations involved to solve the fourth order polynomial are rather heavy, preventing us from reaching low calculation times which are especially important for the following computation of the approximated evidence and the related model parameter

estimation. The fourth order polynomial can be solved analytically, but requires to deal with complex numbers.

ICM solution for a square-root Gamma-shaped prior

By presenting the Gamma-Gamma MAP filter, we have pointed out that the Gamma distribution explains the observation of K-distributed intensity SAR data for slightly textured scenes. A Gaussian-shaped prior is only approximately able to achieve this. To overcome this apparent drawback, we change the shape of the employed prior from a Gaussian to a square-root Gamma distribution, which is the equivalent of the Gamma distribution for the square-root intensity case as stated in chapter two. Hence, we rewrite the prior as

$$p(X_i = x_i | X_j = x_j, j \in \mathcal{N}_i, \boldsymbol{\theta}, \nu) = 2 \left(\frac{x_i}{\tilde{\mu}_i} \right)^{2\nu-1} \frac{\nu^\nu}{\tilde{\mu}_i \Gamma(\nu)} \exp \left(-\nu \left(\frac{x_i}{\tilde{\mu}_i} \right)^2 \right), \quad (5.7)$$

where ν is a form parameter equivalent to the scale parameter σ , and $\tilde{\mu}_i$ encapsulates information about the neighboring pixels of x_i similar to μ_i in the Gauss-Markov case.

Following the same reasoning as before, we find the new value of x_i that maximizes the local posterior of the changed prior to be a root of

$$x_i^4 + \frac{2L - 2\nu + 1}{2\nu} \tilde{\mu}_i^2 x_i^2 - \frac{L}{\nu} \tilde{\mu}_i^2 y_i^2 = 0, \quad (5.8)$$

which is very easy to solve and results in a single unique positive solution. As of this, the ICM optimization for a square-root Gamma prior is much faster. In our implementation computation times could be reduced by a factor of about ten. Unfortunately, the estimation of the model parameters, which is computationally more challenging than the calculation of the MAP estimate, as will be demonstrated, is slower by about the same factor, if performed by the IEM algorithm.

The square-root Gamma model cannot be considered to be suitable for the description of textures with non-stationary mean. As illustrated in Fig. 2.9, the width of the distribution increases with its mean value. Hence, a fixed value of ν estimated from data with non-stationary mean will result in a reduced smoothing for brighter image regions compared to darker ones. As a consequence, the application of the square-root Gamma model for texture restoration from speckled data with a non-stationary mean makes no sense.

Approximation of a Gaussian by the square-root Gamma distribution

Nevertheless, we can exploit the identified computational advantage concerning the MAP estimate by approximating the Gaussian shape of the Gauss-Markov model by the square-root Gamma distribution, i.e. knowing the model parameters for the Gauss-Markov prior, the MAP estimation is computed much faster with an approximated prior distribution. To perform this approximation, we make use of the fact that the coefficient of variation for both amplitude and square-root intensity images is almost equal (cf. Fig. 2.10), which results in

$$\nu = 0.5227^2 \frac{\tilde{\mu}_i^2}{\sigma^2} \quad (5.9)$$

for the form parameter ν , where the relation of Eq. 2.43 has been used. In order not to introduce a bias in the estimator, the maxima of the Gaussian and the square-root Gamma distributions must lie at the same position μ_i . In this case, their respective mean values μ_i and $\tilde{\mu}_i$ are not identical. Using the derived relation

$$\mu_i = \tilde{\mu}_i \sqrt{\frac{2\nu - 1}{2\nu}}, \quad (5.10)$$

we find the two parameters of the approximating square-root Gamma distribution to be given by

$$\nu = \frac{1}{2} + \left(\frac{0.5227\mu_i}{\sigma} \right)^2 \quad (5.11)$$

and

$$\tilde{\mu}_i = \sqrt{\mu_i^2 + \frac{\sigma^2}{2 \cdot 0.5227^2}}. \quad (5.12)$$

From Eq. 5.11, we see that ν is not fixed, but adapts itself to the predicted mean μ_i of the Gauss-Markov model for every single pixel.

5.1.2 Error Evaluation for the Square-Root Gamma Approximation

The evaluation of the accuracy of the square-root Gamma approximation is not a trivial task since the approximating function depends on parameters like x_i , $\tilde{\mu}_i$ and ν . If the quality of the maximizer x_i of the local probability distribution is to be evaluated, the values of L and y_i must also be considered. Because of its higher relevance, we decided to study this latter case.

In Fig. 5.2, the relative estimation error $|x_i - \hat{x}_i|/x_i$ of \hat{x}_i obtained under the approximation compared to the reference value x_i resulting from the Gaussian-shaped prior is illustrated. As shown in the three plots for different values of σ and for a fixed value of the observation $y_i = 100$, the relative error is rather small, lying below two percent. For the most important practical case of $\sigma = 8$ the error is even negligible. The two other cases with $\sigma = 2$ and $\sigma = 32$ can be considered as limiting cases for real SAR data, as shown by experimental results. However, the relative error sometimes dramatically grows for small values of μ_i , since the square-root Gamma distribution is limited to positive values not allowing to correctly approximate the negative lobes of the Gaussian. Increasing errors will also occur for decreasing values of y_i because the relative width of the likelihood with regard to the prior also influences the approximation accuracy.

The Gaussian and its approximations are displayed in Fig. 5.3 for $\sigma = 32$, i.e. for a rather high approximation error. As shown in Fig. 2.9 the approximation accuracy will increase for higher values of ν , i.e. lower values of σ , since the square-root Gamma distributions comes closer to a Gaussian shape. Similar experiments for single look speckle showed equivalent results as the considered case with $L = 3$. Hence, we consider this approximation, which introduces only a negligible error in the estimation of the cross-section while allowing much faster and easier computations, to be valid. We will use it without special notice throughout the rest of this work for the computation of the MAP estimate.

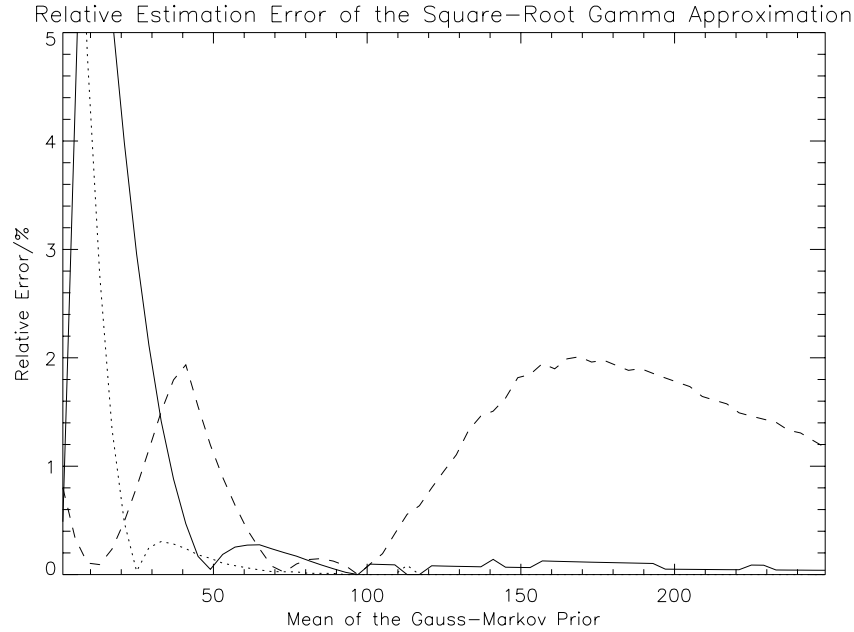


Figure 5.2: Relative estimation error resulting from the square-root Gamma approximation. The relative error of \hat{x}_i compared to the estimate obtained with the Gauss-Markov prior x_i is plotted as a function of μ_i , i.e. of the prediction of the prior, for $L = 3$, $y_i = 100$ and $\sigma = 2$ (dotted), $\sigma = 8$ (solid) and $\sigma = 32$ (dashed).

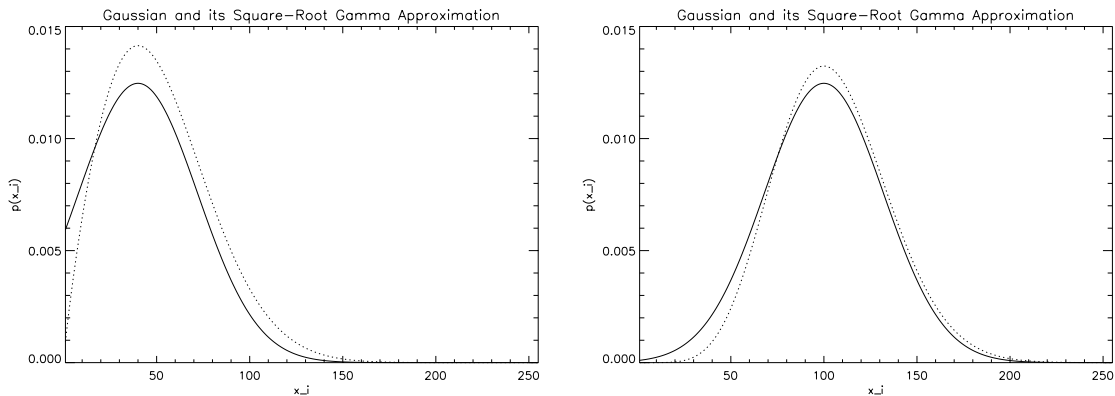


Figure 5.3: A Gaussian and its square-root Gamma approximation. Left: Approximation for $\mu_i = 40$. Right: Approximation for $\mu_i = 100$. The Gaussian (solid) and its approximation (dotted) are shown for a high (left) and a small (right) relative estimation error with $\sigma = 32$ and $L = 3$.

This conclusion is confirmed in Tab. 5.1, which compares the filtering results obtained with and without the discussed square-root Gamma approximation that reduces computing

MSE	Original - Gaussian	Original - SqrtGamma	Gaussian - SqrtGamma
Sand	145	142	3.2
Water	240	230	7.2
Straw	172	166	20.0
Weave	274	274	10.3

Table 5.1: Mean square errors for the restoration of the four Brodatz textures. The mean square errors are given for the original data compared to the results obtained with the Gaussian and the approximating square-root Gamma prior, as well as for the difference between the two filtering results. We considered artificially speckled data with $L = 3$.

times by an order of magnitude: As required, the mean square error between the filtered images is rather small. However, it is striking that the mean square error to the original data decreases for the square-root Gamma approximation of the Gauss-Markov prior. This could be due to the fact that the original data indeed are not Gaussian, but are better described by a skewed pdf. Both the "water" and the "sand" texture possess nearly Gamma-shaped histograms and rather slowly varying mean values.

5.2 Model Parameter Estimation by Iterative Evidence Maximization

Being able to calculate a MAP estimate of the noise-free signal x , a crucial problem still remains: How to choose the model parameter vector $\boldsymbol{\theta}$? Here, we take advantage of the second level of Bayesian inference to estimate the set of parameters that best explains the observed noisy image y . Assuming a uniform prior $p(\boldsymbol{\theta})$, Bayes' equation for the model parameters is reduced to the following proportionality

$$p(\boldsymbol{\theta}|Y = y) \propto p(Y = y|\boldsymbol{\theta}) = \int p(Y = y|X = x)p(X = x|\boldsymbol{\theta})dx, \quad (5.13)$$

where the integral has to be performed over the whole space of X and, again, the likelihood function for speckle does not explicitly depend on $\boldsymbol{\theta}$. This integral is called evidence and reflects the probability of the data y given an assumed model characterized by $\boldsymbol{\theta}$. The task consists in maximizing this evidence as a function of $\boldsymbol{\theta}$, i.e. in finding the most likely model parameter vector. Unfortunately, this maximization is generally a quite difficult task, especially for non-Gaussian multi-dimensional functions, as they often occur in image processing, e.g. in form of non-linear Gibbs models. We have already addressed several approaches to the problem of model parameter estimation from incomplete data in chapter four.

For the special case of estimation from a uni-modal posterior distribution, this task can be solved. We introduce the chosen approach and verify if made assumptions do really apply for SAR image restoration using Gauss-Markov random fields as a texture model.

5.2.1 Approximation of the Evidence Integral

To be able to perform the integration over the posterior product $p(Y = y|X = x)p(X = x|\boldsymbol{\theta})$ and to keep the problem tractable several approximations must be made:

- In a first step, we approximate the multi-dimensional posterior product by a multi-variate Gaussian distribution, which is centered around the MAP estimate \hat{x}_{MAP} , i.e. around the maximum of the posterior [MacKay]. This is shown later to be a good approximation for the used likelihood function in combination with the Gauss-Markov random field model.
- In a second step, we consider the integrand of Eq. 5.13 to consist of mutually independent random variables, allowing us to break the conditional probability density functions into the products of their components. Of course, this statistical independence is not given but has been shown to be a good approximation for large $|X|$, i.e. a large number of pixels, being equivalent to the maximum pseudo-likelihood approach [Besag86].

General mathematical approximation

By applying the proposed simplifications, we find for the first step, the second step and for the final result of the posterior product, i.e. the integrand,

$$\begin{aligned}
 p(Y = y|X = x)p(X = x|\boldsymbol{\theta}) &\approx p(y|\hat{x}_{MAP})p(\hat{x}_{MAP}|\boldsymbol{\theta}) \exp\left(-\frac{1}{2}\boldsymbol{\Delta x}^T \mathbf{H} \boldsymbol{\Delta x}\right) & (5.14) \\
 &\approx \prod_{i=1}^{|X|} p(y_i|x_i)p(x_i|\boldsymbol{\theta}) \\
 &\approx \prod_{i=1}^{|X|} p(y_i|\hat{x}_{iMAP})p(\hat{x}_{iMAP}|\boldsymbol{\theta}) \exp\left(-\frac{1}{2}h_{ii}(x_i - \hat{x}_{iMAP})^2\right),
 \end{aligned}$$

where \hat{x}_{iMAP} is the MAP estimate of x_i obtained with a fixed parameter vector $\boldsymbol{\theta}$. $\boldsymbol{\Delta x}$ denotes $x - \hat{x}_{MAP}$ in vector form and h_{ii} are the main diagonal elements of the Hessian matrix \mathbf{H} , which is given by (∇ denotes the Nabla operator)

$$\begin{aligned}
 \mathbf{H} &= -\nabla\nabla \log\left(\prod_{i=1}^{|X|} p(Y_i = y_i|X_i = x_i)p(X_i = x_i|\boldsymbol{\theta})\right)\Bigg|_{x_i=\hat{x}_{iMAP}} & (5.15) \\
 &= -\nabla\nabla \sum_{i=1}^{|X|} \log(p(Y_i = y_i|X_i = x_i)p(X_i = x_i|\boldsymbol{\theta}))\Bigg|_{x_i=\hat{x}_{iMAP}}.
 \end{aligned}$$

As a consequence, we are able to perform the integration over the resulting function. The approximated evidence is simply described by the properties of the multi-variate Gaussian and the posterior fit of the MAP solution to the data:

$$p(Y = y|\boldsymbol{\theta}) = \int p(Y = y|X = x, \boldsymbol{\theta})p(X = x|\boldsymbol{\theta})dx \quad (5.16)$$

$$\begin{aligned}
&\approx \int \prod_{i=1}^{|X|} p(y_i|\hat{x}_{i_{MAP}})p(\hat{x}_{i_{MAP}}|\boldsymbol{\theta}) \exp\left(-\frac{1}{2}\boldsymbol{\Delta x}^T\mathbf{H}\boldsymbol{\Delta x}\right) dx_1 \dots dx_{|X|} \\
&\approx \frac{(2\pi)^{\frac{|X|}{2}}}{\sqrt{\det \mathbf{H}}} \prod_{i=1}^{|X|} p(y_i|\hat{x}_{i_{MAP}})p(\hat{x}_{i_{MAP}}|\boldsymbol{\theta})
\end{aligned}$$

Please note that this can be rewritten using the Occam factor Ω , which was mentioned in chapter three. Thus, we obtain for easier interpretation

$$\begin{aligned}
p(Y = y|\boldsymbol{\theta}) &\approx \frac{(2\pi)^{\frac{|X|}{2}}}{\sqrt{\det \mathbf{H}}} p(y|\hat{x}_{MAP})p(\hat{x}_{MAP}|\boldsymbol{\theta}) \\
&\approx p(y|\hat{x}_{MAP}) \cdot \Omega \\
&\approx p(Y = y|X = \hat{x}_{MAP}) \cdot \Omega.
\end{aligned} \tag{5.17}$$

Finally, since we prefer to use the logarithmic form of $p(Y = y|\boldsymbol{\theta})$ for numerical reasons, the log-evidence can be expressed as [Walessa99a]

$$\begin{aligned}
\log p(Y = y|\boldsymbol{\theta}) &\approx \frac{1}{2} \log \left(\frac{2\pi^{|X|}}{\det \mathbf{H}} \right) + \sum_{i=1}^{|X|} (\log p(y_i|\hat{x}_{i_{MAP}}) + \log p(\hat{x}_{i_{MAP}}|\boldsymbol{\theta})) \\
&\approx \sum_{i=1}^{|X|} \left(\frac{1}{2} (\log 2\pi - \log h_{ii}) + \log p(y_i|\hat{x}_{i_{MAP}}) + \log p(\hat{x}_{i_{MAP}}|\boldsymbol{\theta}) \right),
\end{aligned} \tag{5.18}$$

where the assumption

$$\det \mathbf{H} \approx \prod_{i=1}^{|X|} h_{ii} \tag{5.19}$$

was employed. This approximation is valid since it implies that all covariances resulting from the sparsely set matrix \mathbf{H} are zero (off-main-diagonal values are neglected), which is in accordance with the already made assumption of statistical independence in Eq. 5.14. Moreover, this simplification is required for practical reasons, as it prevents us from the computation of determinants of dimension $|X| \times |X|$, which is not feasible for sufficiently large estimation windows.

Approximation for a Gauss-Markov prior

Using the preceding simplifications, only the components on the main diagonal of the matrix \mathbf{H} are needed. These components, which are inversely proportional to the variances of the approximating Gaussians, can be found to be

$$\begin{aligned}
h_{ii} &= - \frac{\partial^2}{\partial x_i^2} \sum_{j=1}^{|X|} \left(-2L \log x_j - L \left(\frac{y_j}{x_j} \right)^2 - \frac{(x_j - \mu_j)^2}{2\sigma^2} \right) \Bigg|_{x_i = \hat{x}_{i_{MAP}}} \\
&= \frac{6Ly_i^2}{\hat{x}_{i_{MAP}}^4} - \frac{2L}{\hat{x}_{i_{MAP}}^2} + \frac{1}{\sigma^2} \left(1 + \sum_{k,l} \theta_{k,l}^2 \right) \\
&= \frac{6Ly_i^2}{\hat{x}_{i_{MAP}}^4} - \frac{2L}{\hat{x}_{i_{MAP}}^2} + \frac{1}{\sigma^2} (1 + |\boldsymbol{\theta}|^2)
\end{aligned} \tag{5.20}$$

for the square-root Gamma likelihood function in combination with the Gauss-Markov random field model, and are mainly characterized by the variance σ^2 and the norm $|\boldsymbol{\theta}|$ of the directional model parameter vector of the Gauss-Markov prior.

Approximation for a square-root Gamma prior

By analogy to the Gauss-Markov case, the expression for h_{ii} can also be derived for a square-root Gamma-shaped prior, as given in Eq. 5.7, instead of the Gauss-Markov prior. This yields

$$h_{ii} = \frac{6Ly_i^2}{\hat{x}_{i_{MAP}}^4} + \frac{2\nu - 2L - 1}{\hat{x}_{i_{MAP}}^2} + \frac{2\nu}{\mu_i^2} + 2\nu \sum_{k,l} \frac{\theta_{k,l}^2}{\mu_{k,l}^2} \left(3 \frac{\hat{x}_{k,l_{MAP}}^2}{\mu_{k,l}^2} - 1 \right), \quad (5.21)$$

which is much more complicated than Eq. 5.20 and depends on the predicted values $\mu_{k,l}$ of all neighbors around x_i . This result is indicated only for completeness since, as discussed before, we consider this prior not to be suitable for texture restoration and use the square-root Gamma distribution only as an approximation of the Gaussian to speed up the computation of the MAP estimate.

Error evaluation for the Gaussian approximation of the posterior

In order to determine the error introduced by approximating the evidence by an integral over a Gaussian, we proceed similarly to the case of the square-root Gamma approximation: Keeping σ , L and y_i fixed, we calculate the MAP estimate $\hat{x}_{i_{MAP}}$ and the evidence for a running value of μ_i . For comparison, we plot the relative error of the numerically obtained reference evidence according to Eq. 5.13 and the value obtained by the proposed approximation of Eq. 5.16. These results and a histogram of the calculated relative errors computed for values of μ_i from 0 to 255 are displayed in Fig. 5.4. The parameter σ was fixed to 12, which is a reasonable value for practical applications. The interval was limited to a maximum value of 255 for practical reasons. Actual gray-values of SAR data may be much larger but can be rescaled to this gray-value range.

As shown, the relative error decreases with growing L and increasing y_i . For very low values of y_i the approximation of the posterior may be rather poor for reasons equivalent to those already mentioned for the square-root Gamma approximation: The approximating Gaussian is not limited to values larger than zero and the likelihood function becomes rather peaked with a width similar to the one of the prior. However, this large error quickly decreases with growing values of y_i , as illustrated in the plots, and attains values clearly below one percent. The displayed histograms, which have been computed for the relative errors resulting from all combinations of y_i and μ_i with values from 0 to 255, confirm this observation.

We conclude that the computation of the approximated evidence is of sufficient accuracy and can be used for model parameter estimation. The numerical evaluation of the evidence is still computationally demanding for it requires the computation of the MAP estimate.

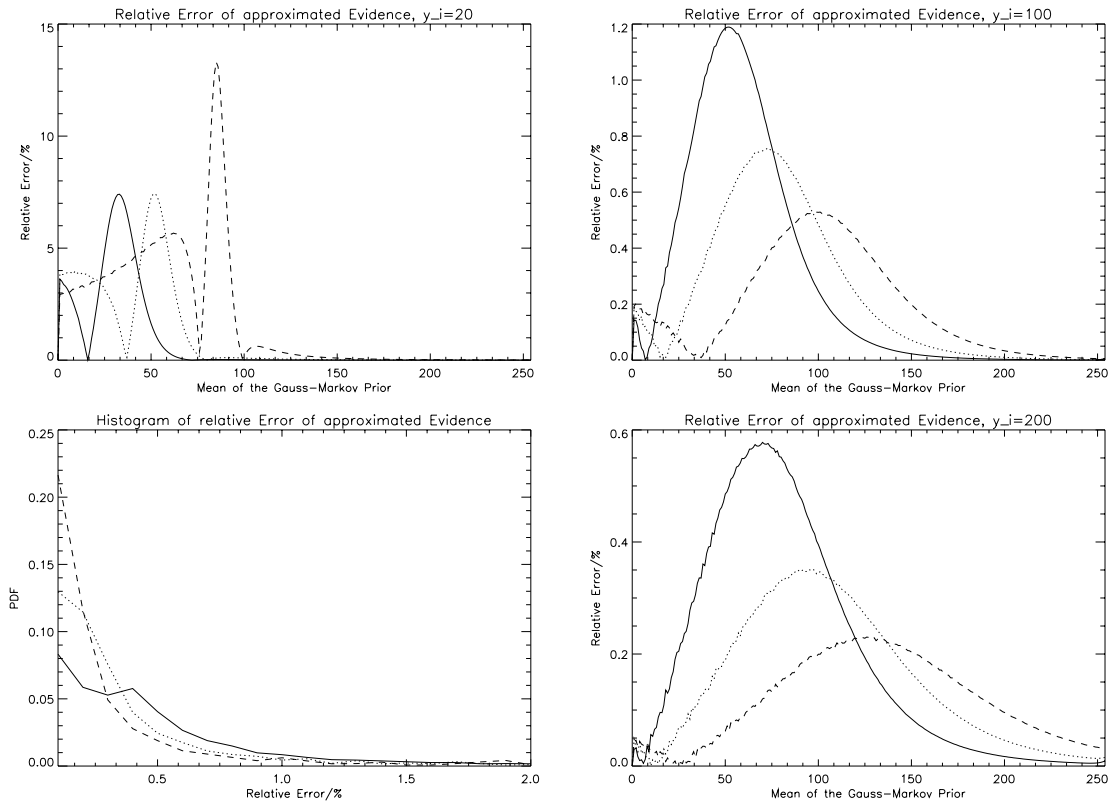


Figure 5.4: Relative errors of the approximated evidence for a single pixel. Clockwise: Relative errors for $y_i = 10$, for $y_i = 100$, for $y_i = 200$ as a function of μ_i , and histogram of relative errors for the interval of y_i from 0 to 255. Curves are displayed for $L = 1$ (solid), $L = 3$ (dotted) and $L = 8$ (dashed). In all cases $\sigma = 12$ was assumed.

However, the MAP estimate and the model parameters are obtained simultaneously, providing the restored image and a description of its textural contents.

5.2.2 Iterative Model Parameter Estimation

We have demonstrated in the last section that the evidence for a given model and a fixed parameter set can be numerically computed. However, to determine the model parameters that maximize this measure, an iterative procedure is necessary since an analytical solution is not available. This is both due to the complexity of the integral and the chosen approximation, which requires the MAP estimate of x for the evidence computation.

For the purpose of maximization, we tested two different approaches: A simple hill-climbing algorithm with a fixed step width for the parameter update, and a two step approach, which, as an interesting property, can be regarded as an extension of the EM-type algorithm.

Evidence maximization by hill-climbing

In principle, the maximization of the approximated evidence as a function of θ and σ can be performed with any optimization technique. For computational reasons only deterministic approaches are suitable. Since an ICM approach is not possible because we cannot directly find each maximizing individual parameter, a simple hill-climbing algorithm remains a reasonable choice.

The approach that we implemented works in the following way:

1. Start with an initial parameter configuration serving as a first guess and compute the MAP estimate and the evidence for this set of parameters.
2. At iteration n , update a single parameter using a fixed step-width, i.e. $\theta_{k,l}^{(n+1)} = \theta_{k,l}^{(n)} + \delta_{k,l}^{(n)}$, and compute the new MAP estimate and the resulting evidence.
3. Keep the new parameter value if the evidence is increased, or reject it and invert $\delta_{k,l}$, i.e. $\delta_{k,l}^{(n+1)} = -\delta_{k,l}^{(n)}$, otherwise.
4. Re-iterate several times from step two for all parameters $\theta_{k,l}$ and σ until convergence is reached, i.e. until the evidence no more increases.

Typically, as a first guess we use a heuristically determined value of $\sigma = 10$ and a model parameter vector describing uniform areas with $\theta_{k,l} = \theta$ for all its elements. This parameter vector is normalized to 0.5 after each update to guarantee stability. Consequently, all parameter values are changed at each update. This approach proved to work fine in practice and yields consistent parameter estimates. Typical values for $\delta_{k,l}$ are 0.001 and 0.125 for δ_σ , which was assigned a different value for faster convergence. Note that the values for $\delta_{k,l}$ must be eventually lowered for neighborhoods larger than fifth order because of a higher number of parameters.

The proposed approach gives convincing results concerning the parameter estimates but, as its main drawback, showed to be computationally very demanding. After each update, the MAP estimate has to be calculated and the full expression of the evidence has to be computed. This becomes even more challenging when the number of parameters to determine increases, i.e. when the neighborhood size grows. For practical applications, this method showed to be by far too slow.

Evidence maximization by Occam factor optimization

In order to faster estimate the Gauss-Markov random field parameters, we adopt an approach which shows to be related to the expectation maximization algorithm of [Dempster, Moon] discussed in the preceding chapter. Equivalently, the proposed method can be split into two steps:

1. Compute the MAP estimate of the data using the observation y and the current estimate of the model parameters θ and σ .

2. Keep the MAP estimate fixed and determine the model parameters $\boldsymbol{\theta}$ and σ that maximize the expression for the approximated evidence $p(Y = y|\boldsymbol{\theta})$ of Eq. 5.17. Reiterate from step one until convergence is reached.

For the optimization task of step two, the identical hill-climbing approach with starting parameters as presented before is used. However, in contrast to the hill-climbing method for evidence maximization, the main computational load now lies on the calculation of the approximated evidence and not on the MAP estimate. A new MAP estimate is only computed once the model parameters maximize $p(Y = y|\boldsymbol{\theta})$ for the current configuration of \hat{x}_{MAP} . In this way, the maximization of the evidence can be performed much faster since it reduces to the maximization of

$$\frac{(2\pi)^{\frac{|X|}{2}}}{\sqrt{\det \mathbf{H}}} p(\hat{x}_{MAP}|\boldsymbol{\theta}) = \Omega, \quad (5.22)$$

at each iteration. The likelihood term is independent of the Gauss-Markov parameters and can be omitted for the optimization of the evidence in step two. Hence, the maximization is done solely on the Occam factor Ω , which we call the Occam factor maximization step.

Interpretation of the Occam factor maximization step

The above method becomes very interesting if we consider the expectation maximization algorithm or, more exactly, the EM-type approach: The first step of the proposed optimization algorithm and the E-step are identical. Concerning the M-step, the second part of the presented technique is slightly different. Instead of maximizing

$$p(\hat{x}_{MAP}|\boldsymbol{\theta}) = \prod_{i=1}^{|X|} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\hat{x}_{i_{MAP}} - \mu_i)^2}{2\sigma^2}\right) \quad (5.23)$$

as in the maximization step of the EM-type algorithm, we consider an additional factor of

$$\frac{(2\pi)^{\frac{|X|}{2}}}{\sqrt{\det \mathbf{H}}} = \int \exp\left(-\frac{1}{2}\boldsymbol{\Delta x}^T \mathbf{H} \boldsymbol{\Delta x}\right) dx_1 \dots dx_{|X|}. \quad (5.24)$$

The product of these two terms was shown to result in the Occam factor as the measure to maximize in step two.

In the last chapter, we demonstrated that the EM-type algorithm does not provide robust and satisfactory results. Consequently, the difference between stable and unstable estimates can only result from the factor given in Eq. 5.24. As a matter of fact, by neglecting this factor for parameter estimation equivalent results to those of the EM-type algorithm are obtained.

Considering only a single pixel and neglecting the first two additive terms of Eq. 5.20 which result from the likelihood and are functions of \hat{x}_{MAP} , we simplify Eq. 5.22 for the purpose of easier interpretation to

$$\sqrt{\frac{2\pi\sigma^2}{1 + |\boldsymbol{\theta}|^2}} p(\hat{x}_{i_{MAP}}|\boldsymbol{\theta}) = \Omega_i. \quad (5.25)$$

From this simplified term the qualitative difference between the proposed method of iteratively maximizing the approximated evidence and the EM-type approach becomes directly clear:

- The under-estimation of σ , which was observed for the EM-type algorithm, is compensated by σ present in the numerator of Eq. 5.25, thus favoring larger values.
- The unstable behavior of the estimates of the directional parameters is reduced by $|\boldsymbol{\theta}|^2$ in the denominator, suppressing outliers or high values of single parameters, which would result in a higher norm. A parameter vector with a small norm, i.e. a small variance of the parameter values, is more probable.

A direct equivalence of the proposed method to the full expectation-maximization algorithm could not be detected by a comparison of analytically computed examples. Nevertheless, these approaches seem to be related in a certain way. More theoretical research is required to qualify the proposed technique with regard to other methods.

5.2.3 Evaluation of the Proposed Optimization Approaches

The performances of the two presented optimization techniques for evidence maximization have been tested on the four speckled texture images ($L = 3$) from the Brodatz album. The filtering conditions, such as step widths or the number of iterations for the MAP estimator, have been chosen to be equal, and the estimation window was taken to be of the size of the whole texture, i.e. 256×256 pixels. We measured the performance in terms of the mean square error of the filtered image to the original one and in terms of the obtained log-evidence normalized to one pixel. The corresponding results are provided in Tab. 5.2.

Visually, no differences can be detected between the filtered images. However, the hill-climbing method behaves slightly better concerning both the mean square error of the restoration and the evidence of the estimated texture parameters. Significant differences occur only for the mean square error of the "water" texture. It is interesting to observe in our examples that better evidences also yield better mean square errors, which confirms the employment of the evidence as a criterion for model parameter extraction. Note that the indicated values for the MSE may differ from previously given examples, due to changed step widths or a different number of iterations.

Although better results can be achieved with the hill-climbing method, we choose the Occam factor maximization method for model parameter estimation, since it is about one magnitude faster in computation. This advantage is mostly due to the fact that the computation of the MAP estimate is not required for every update of the parameter vector, as it is the case for the hill-climbing method. Furthermore, the value of the Occam factor can be computed much faster than the full evidence. The computational load of the hill-climbing method is beyond of what can be accepted for the processing of larger scenes.

	E_{hill}	E_{occam}	MSE_{hill}	$\text{MSE}_{\text{occam}}$
Sand	-4.377	-4.397	133	137
Water	-4.854	-4.881	200	244
Straw	-4.296	-4.344	135	146
Weave	-4.805	-4.817	250	257

Table 5.2: Log-evidences and mean square errors for the restoration of the four Brodatz textures with a Gauss-Markov model of fifth neighborhood order. Both measures are given for the optimization by hill-climbing and by Occam factor maximization.

5.3 Properties of the Gauss-Markov Texture Estimator

Having derived the estimator for both the MAP estimate of the cross-section and the ML estimate of the model parameters, a quantitative evaluation is required. For the users of SAR data the preservation of the image mean value is of particular interest for interpretation and classification. Hence, we have to check for a possible bias introduced by the Gauss-Markov filtering. The accuracy of the estimated model parameters has already been the subject of chapter four. For this reason, besides a short example of local parameter extraction, we will only verify the robustness of the iterative estimation approach by applying it to noise-free data in the limit of $L \rightarrow \infty$. At the end of this section will follow an analysis of the minimum achievable MSE of the obtained MAP estimate, which is related to the Cramér-Rao bound for ML estimation.

5.3.1 Bias of the Gauss-Markov Texture Estimator

In order to check for a possible bias of the estimated mean cross-section, we generated six test images (128×128 pixels) with different mean values μ_x being constant throughout the whole image. We added speckle of $L = 1$, $L = 3$ and $L = 8$ and measured the mean of the filtering result $\hat{\mu}_x$ for these eighteen different cases. As before, we work in the square-root intensity domain.

In Tab. 5.3 the obtained results $\hat{\mu}_x$ for the filtered images are enumerated and compared to the original mean values μ_x . We observe that a bias in form of an under-estimation of the original mean value is introduced. For a better analysis of this phenomenon, we plotted the estimated mean against its correct value in Fig. 5.5. From the left plot, we deduce that the bias is not of additive but of multiplicative form, which is confirmed by the plot on the right where the ratio of the mean values $\mu_x/\hat{\mu}_x$ is displayed. This ratio is almost constant for the whole range of tested values. Consequently, we have to deal with a constant multiplicative bias as a function of L which can be easily corrected.

A closer look reveals that the required correction factor is approximately equal to the inverse mean of the square-root intensity speckle noise (cf. Eq. 2.37), which results to be 1.128, 1.042 and 1.016 for the given values of L . When these correction factors are used, a

μ_x	25	50	100	150	200	250	$\overline{\mu_x/\hat{\mu}_x}$
$\hat{\mu}_x, L=1$	22.513	45.606	90.827	135.92	180.46	226.49	1.104
$\hat{\mu}_x, L=3$	23.917	48.147	96.962	145.23	193.31	241.39	1.036
$\hat{\mu}_x, L=8$	24.569	49.227	98.599	147.90	196.91	246.87	1.015

Table 5.3: Bias of the Gauss-Markov texture estimator. The original noise-free mean μ_x and the mean gray-value of the filtering results $\hat{\mu}_x$ for different speckle levels L are displayed. We also give the empirical mean for the ratio of these values.

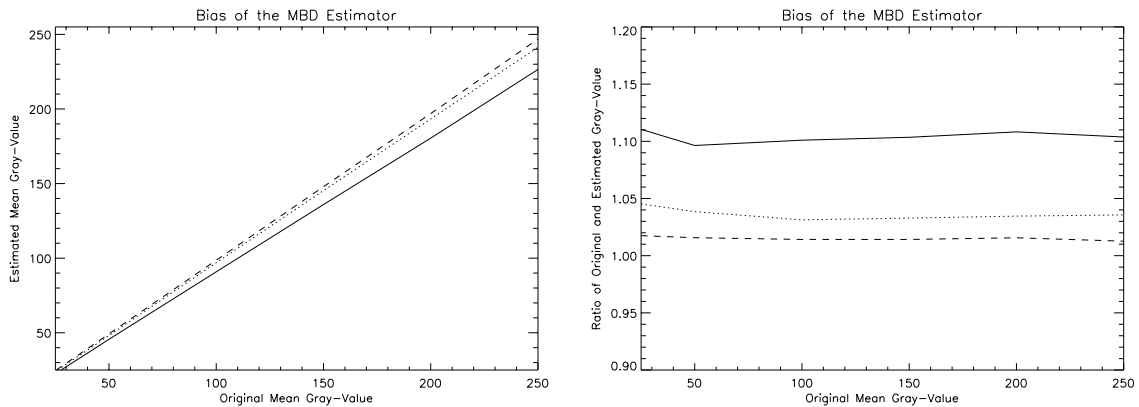


Figure 5.5: Multiplicative bias of the estimator. Left: Plot of the estimated mean $\hat{\mu}_x$ over the original value μ_x for $L = 1$ (solid), $L = 3$ (dotted) and $L = 8$ (dashed) showing the multiplicative character of the introduced bias. Right: Ratio of the original to the estimated mean $\mu_x/\hat{\mu}_x$. This ratio is approximately equal to the inverse of the square-root intensity speckle mean.

relative error of more than one percent only occurs for single look data.

On square-root intensity data the estimator introduces the same bias as a simple mean filter, which is due to the averaging that is performed by the Gauss-Markov model in order to obtain the prediction value μ_i . The unbiased ML solution, however, consists in averaging the squares of the observed gray-values and taking the square-root afterwards. Nonetheless, the proposed estimator very accurately preserves the mean of the observed data and allows for a simple adjustment to the maximum likelihood value by means of an analytically given correction factor, i.e. by dividing the obtained result by the mean of the square-root intensity speckle noise.

5.3.2 Quality of Model Parameter Estimation by IEM

The accuracy of the model parameter estimation from speckled data has already been analyzed in chapter four. It was shown that the estimation by IEM yields results similar to

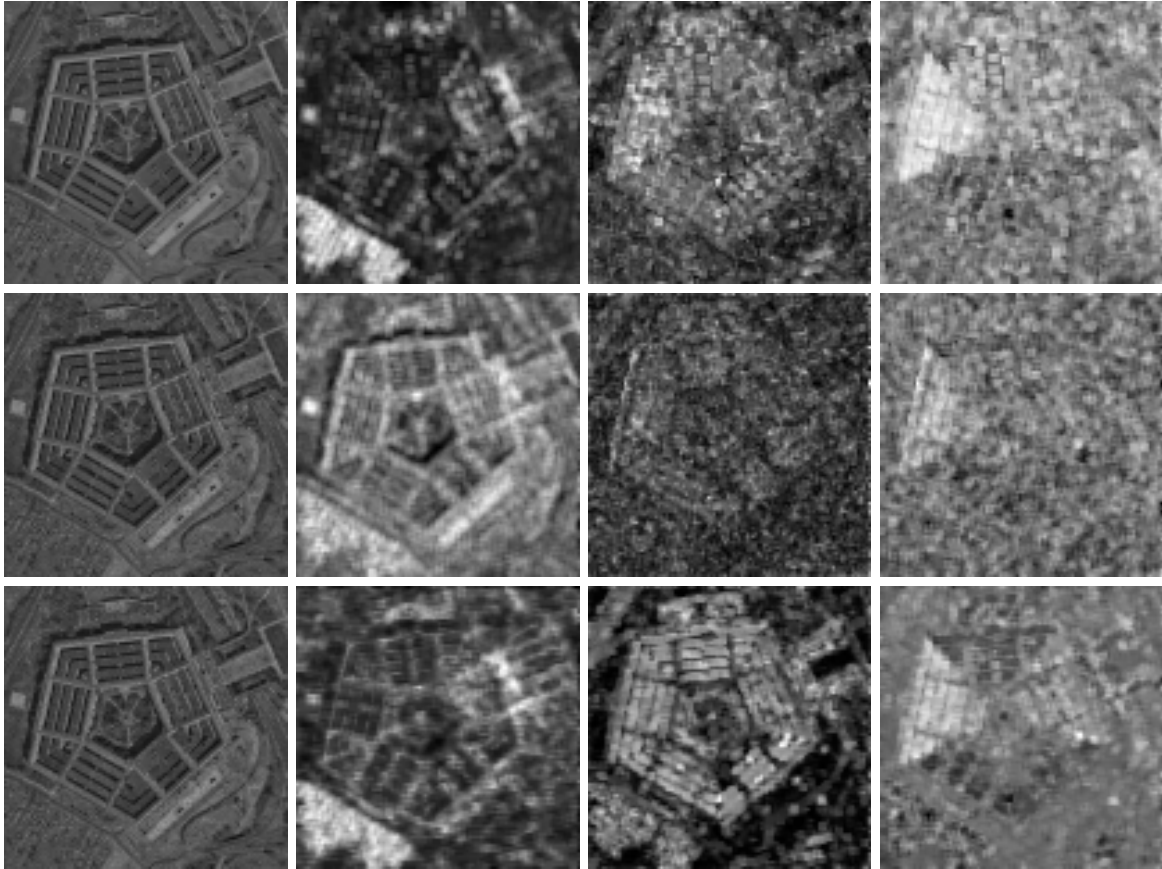


Figure 5.6: Local model parameter estimation from complete and incomplete data. Top row: Original noise-free image (512×512 pixels) and model parameters extracted by MPL estimation. Center row: Speckled image ($L = 4$) and its MPL parameters. Bottom row: Speckled image ($L = 4$) and model parameters extracted by IEM considering the speckle noise. From left to right, the parameter images show the estimated values of σ , $|\boldsymbol{\theta}|$ and $\theta_{2,1}$ for a fifth order neighborhood. For parameter estimation windows of 15×15 pixels have been used.

the ones obtained by the derived analytical estimator and are close enough to the reference values estimated from noise-free data. A quantitative quality measure is difficult to provide since the Cramér-Rao bound cannot be derived analytically in a closed form. For information extraction and SAR image restoration, the accuracy can be assumed to be sufficient.

Local model parameter estimation from complete and incomplete data

In Fig. 5.6 we depict an example of model parameter estimation from complete and incomplete data. In the first row, the parameters obtained by MPL estimation from the noise-free data are displayed. The next two rows show parameters extracted from speckled

data by MPL estimation without taking into account the noise and by IEM, where speckle is considered.

By comparing the last two rows, conclusions already made in chapter four are confirmed: If the influence of speckle is neglected (MPL estimate from noisy data), the estimated parameters are only of limited value. A clear distinction between differently textured areas by directional information cannot be made. We note however, that the information in σ is much less affected, since it is strongly correlated to the average image brightness.

Concentrating on the information contained in the directional parameters, we observe a much clearer separation between different textures for the IEM estimate (third row) than for the MPL estimate from complete data (first row). This especially applies for the IEM-estimated norm $|\boldsymbol{\theta}|$, which clearly reflects the structures of the original data.

This apparent inconsistency of more descriptive parameter estimates from noisy than from noise-free data is caused by the fit of the data to the Gauss-Markov model. Since the complete data are no direct realization of the used model, the extracted parameters exhibit a noisy appearance. In the case of estimation from incomplete data however, this limited model fit can be attributed to the speckle in the image. The underlying data are "made" to behave according to the imposed Gauss-Markov model allowing an apparently less noisy estimation.

Model parameter estimation by IEM from complete data

In order to check the robustness of the estimator, we consider the limiting case of $L \rightarrow \infty$ and estimate the texture parameters from the original noise-free Brodatz images by iterative evidence maximization (IEM) for incomplete data. The results, which are given for a fifth order neighborhood system, are shown in Tab. 5.4.

The correspondence of the original and the estimated parameters for all four textures is extremely high, confirming the validity of the made assumptions and the convergence of the introduced model parameter estimation approach. While the directional parameters are in almost perfect agreement, the scale parameter σ is slightly under-estimated. This is explained by the value of L , which we have chosen to be $L = 256$ in our experiment. For larger values of L , convergence to the correct value of σ can be observed. It is clear that the assumption of higher noise than actually given in the image results in lower values for σ since part of the model misfit, which is described by σ , is automatically attributed to the overestimated image noise.

5.3.3 Lower Bound of the Mean Square Error

For parameter estimation it is desirable to have a measure indicating the reliability of the obtained estimates. We cannot derive this measure for the model parameters as already mentioned above, but a lower bound can be determined for the mean square error of the filtered image. It can be shown, that this minimum mean square error is directly related to the Hessian matrix \mathbf{H} of the approximated posterior that we have already derived.

Beach sand	σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{3,1}$	$\theta_{3,2}$
Original	8.489	0.4048	0.3933	-0.08244	-0.04275	-0.0688	-0.07476
Estimated	7.689	0.4294	0.4212	-0.1161	-0.07242	-0.07515	-0.08606
Water	σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{3,1}$	$\theta_{3,2}$
Original	14.28	0.1604	0.1572	0.07497	0.04989	0.004371	0.03087
Estimated	13.67	0.1660	0.1578	0.07326	0.04598	0.002348	0.02962
Straw	σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{3,1}$	$\theta_{3,2}$
Original	7.629	0.1719	0.3886	-0.07157	0.1926	-0.02289	-0.1167
Estimated	6.492	0.1703	0.4239	-0.08333	0.1921	-0.001515	-0.1379
Weave	σ	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{3,1}$	$\theta_{3,2}$
Original	12.05	0.3673	0.4307	-0.1035	-0.05603	-0.03876	-0.05532
Estimated	11.22	0.3912	0.4539	-0.1352	-0.07788	-0.04788	-0.06970

Table 5.4: Gauss-Markov random field parameters estimated for the four noise-free Brodatz textures with $L \rightarrow \infty$. The six most significant directional parameters of a fifth order neighborhood system used for estimation are displayed. Parameters have been estimated from noise-free data using the iterative evidence maximization (IEM) method for incomplete data in the limiting case of $L \rightarrow \infty$ and are compared to the corresponding MPL estimates from the original data.

The minimum mean square error for the estimation of a parameter X using prior information is given by [Kroschel]

$$E \left\{ (x - \hat{x})^2 \right\} \geq E \left\{ -\frac{\partial^2}{\partial x^2} \log p(Y = y, X = x) \right\}^{-1}, \quad (5.26)$$

where the expectation is over both X and Y . This measure is closely related to the Cramér-Rao bound, which describes the estimation error for maximum likelihood estimation without prior information.

The minimum achievable variance for each gray-value at site i as in Eq. 5.26 is given by the element ii of the inverse matrix of the expectation of \mathbf{H} . We note, that this is not equal to the inverse of the diagonal elements h_{ii} themselves, if the off-diagonal elements of \mathbf{H} are non-zero: The inverse of the diagonal elements of a matrix are in general not equal to the diagonal elements of its inverse. However, we make this approximation at the risk of under-estimating the correct variance for each estimate. The actual minimum estimation error may be much higher. For the four Brodatz textures we provide approximated results in Tab. 5.5 by computing the empirical mean of $1/h_{ii}$ over all pixels.

In fact, the minimum errors are much lower than the actually measured mean square errors. In principle, two causes may explain this observation: On one hand, the given minimum errors are only valid if the used prior and its parameters really describe the noise-free data, i.e. the data are a realization of a Gauss-Markov random field. Either this is not fully true, or the accuracy of the model parameter estimates is not sufficient. On the

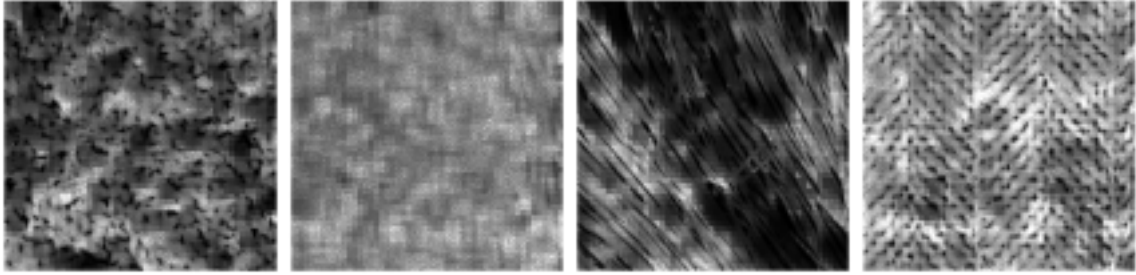


Figure 5.7: Images of the lower bound of the mean square error in each pixel for the four Brodatz textures. From left to right: "Sand", "water", "straw", and "weave" textures. The images illustrate the dependence of the lower bound on the average image brightness reflecting the multiplicative property of speckle.

MSE	Sand	Water	Straw	Weave
Lower Bound	71	134	54	154
Measured Value	133	199	115	262

Table 5.5: Measured mean square error and its theoretical lower bound. The two values are given for the four filtered Brodatz textures with speckle of $L = 3$ and a seventh order neighborhood system for restoration.

other hand, the given lower bound can only be reached if the estimator is efficient. For an efficient estimate to exist, the posterior must be a Gaussian. This is only approximately and not exactly the case as shown before.

The images of Fig. 5.7, which show the lower bounds of the mean square error for each pixel of the four Brodatz textures, reflect the speckle noise properties. As expected, the maximum estimation accuracy for bright regions is lower than the one for darker regions since the variance of the affecting speckle increases with image intensity.

5.4 Optimization of the Filtering Quality

In this section, we point out several issues that directly influence the filtering quality of the proposed algorithm: We address the problem of model order selection for the Gauss-Markov model in order to obtain an optimal restoration. The influence of the size of the windows, which are used for local parameter estimation and spatially adaptive filtering, on the robustness of the filter will also be discussed and a first example of model-based SAR image despeckling will be given for illustration. The section ends with an investigation of the most suitable signal space for SAR image restoration in combination with Gauss-Markov random fields.

5.4.1 Model Order Selection

Different neighborhood sizes of the Gauss-Markov model can be used for both information extraction from SAR data and SAR image filtering. The selection of the most probable model order is performed within the Bayesian framework. We maximize $p(Y = y|\boldsymbol{\theta}, M_{order})$ as a function of $\boldsymbol{\theta}$ for a given model order M_{order} , i.e. we perform a model parameter estimation for different neighborhood sizes and choose the one with highest evidence. Since computing time increases with the number of models to test, we usually use a fixed model order for practical application of the proposed filter. We remind that a full model order selection independent of the estimate $\hat{\boldsymbol{\theta}}$ requires an additional Bayesian layer making this a too complicated task.

Evidence for different model orders

To illustrate the model selection process in more detail we study the different terms of the approximated evidence given by

$$\begin{aligned} p(Y = y|\boldsymbol{\theta}) &\approx \frac{(2\pi)^{\frac{|X|}{2}}}{\sqrt{\det \mathbf{H}}} p(y|\hat{x}_{MAP}) p(\hat{x}_{MAP}|\boldsymbol{\theta}) \\ &\approx p(y|\hat{x}_{MAP}) \cdot \Omega. \end{aligned} \quad (5.27)$$

The terms that determine this evidence are the likelihood fit $p(y|\hat{x}_{MAP})$ and the Occam factor Ω , in which the fit to the prior model $p(\hat{x}_{MAP}|\boldsymbol{\theta})$ is of particular interest. With increasing model order, i.e. a larger neighborhood size and more texture parameters to be estimated, the following behavior can be observed for the measures plotted in Fig. 5.8:

Likelihood fit: Since we are dealing with data of poor quality, the distance in the image space between the true noise-free image and the observed speckled one is rather large. A decrease of the likelihood fit $p(y|\hat{x}_{MAP})$ can be measured with increasing model complexity as the filtering result moves further away from the observation.

Prior model fit: The fit of the prior model $p(\hat{x}_{MAP}|\boldsymbol{\theta})$ increases monotonically with the model order, since a more complex model is able to better explain the actual image content.

Gauss-Markov σ : Equivalent arguments apply for σ of the Gauss-Markov model, which decreases with growing model order, since the residual fit between predicted and estimated data improves.

Occam factor Ω : For the Gauss-Markov model, the Occam factor behaves similar to the prior model fit. Apparently, the second factor, which is the area under the Gaussian-approximated posterior, shows to have only a small influence. σ , which is a dominant factor, decreases only moderately leaving the general evolution of the Occam factor unchanged.

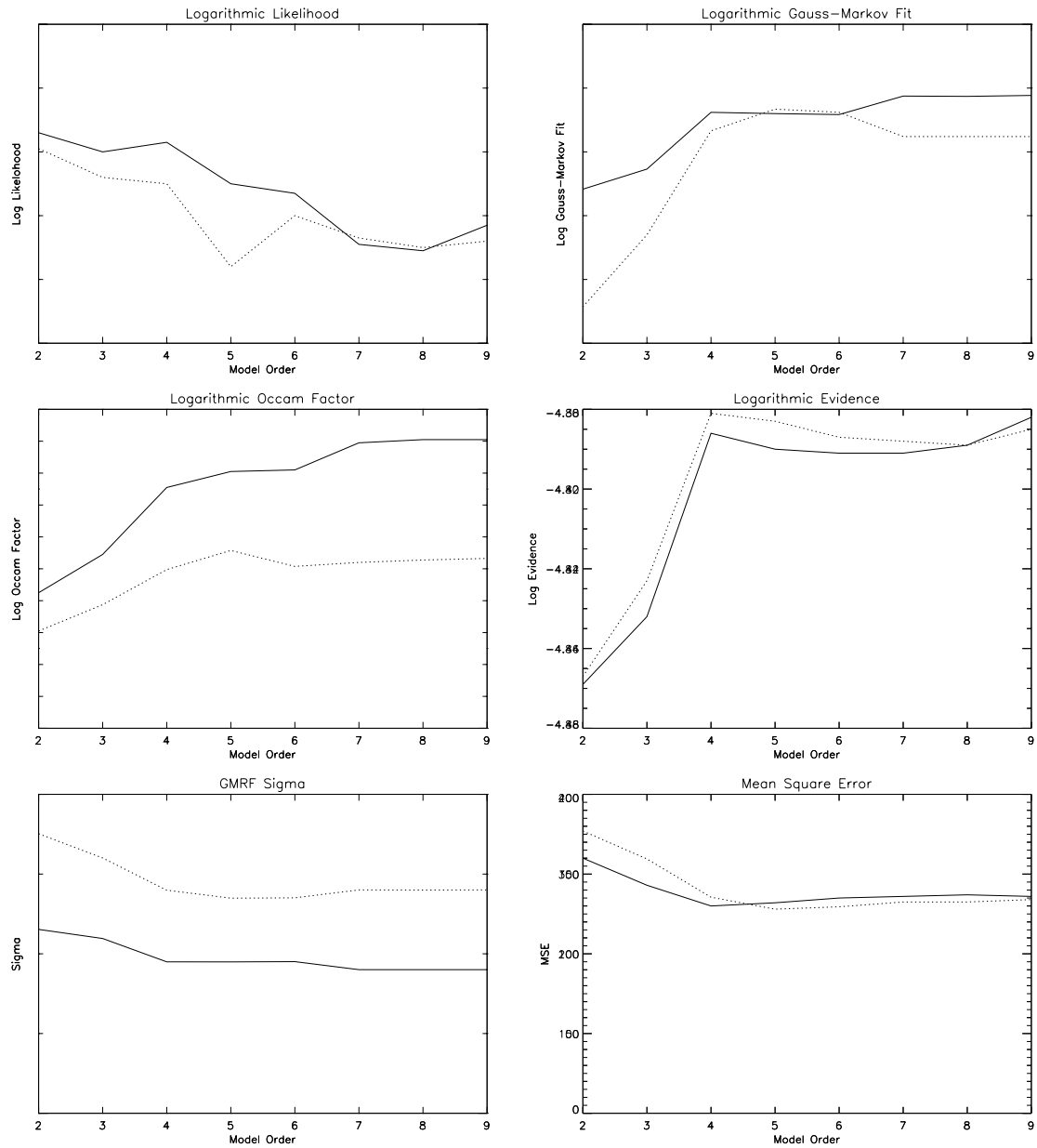


Figure 5.8: Model order selection for the Gauss-Markov model. The evolution of different measures related to the model order selection, like Occam factor or evidence, is qualitatively displayed for the "sand" (solid) and the "weave" texture (dotted). Due to different scaling, the various measures for "sand" and "weave" are not directly comparable.

Evidence: After a steep increase, the evidence remains almost constant. Additional model complexity does not result in considerable improvements, neither does it degrade

the restoration: Equally probable, but less complex models are included in more complex ones. Higher order models with an increased computational load do not bring additional benefits beyond a certain threshold of complexity.

Mean square error: An equivalent behavior can be observed for the mean square error, which seems to be directly related to the evidence. For the two given examples, fourth or fifth order neighborhood systems seem to be the best choice both in terms of evidence and mean square error.

To briefly sum up the consequences of these observations, we note that with increasing model order a saturation of the evidence is observed. All models with order beyond a certain threshold are about equally probable. A ninth order model can easily take the parameter values of an assumed optimal fifth order model with its additional model parameters set to zero resulting in no penalty for the increased complexity. From these equally probable models the one of lowest order should be selected since it requires the least computational complexity without resulting in a loss of filtering quality.

Conclusions

Model order selection is possible for SAR image restoration with Gauss-Markov random fields. However, in contrast to what is generally expected, a clear maximum of the evidence as a function of the model order is not observed. Instead, we found a saturation of the evidence for model orders beyond the most efficient one. As of this and since the whole model order selection process is rather slow, we directly use a high order model (fifth to seventh order) to obtain optimal estimates of the noise-free image. This choice results in higher computing times than the employment of a possibly equivalent lower order model, but has no disadvantageous effects on the filtered image if the parameters, whose estimation complexity increases with the neighborhood size, are correctly estimated.

Using higher order models (e.g. seventh order), a decrease of the mean square error between 5-10% in comparison with a simpler model (e.g. third order) was experimentally observed. This is a rather small improvement. However, more complex models result in sharper, less blurred and visually more attractive restorations.

5.4.2 Local Parameter Estimation and Spatially Adaptive Filtering

Due to spatially changing image content, the model parameter estimation and the filtering of real SAR images have to be performed locally. Parameters must be estimated for a single pixel or a group of pixels by using estimation windows of small size. We have shown in chapter four that parameter estimation from few data may pose problems in terms of stability and accuracy of the extracted model parameters. We illustrated that the analytically derived solution often fails under these conditions and, therefore, rely on the introduced IEM approach.

In Tab. 5.6 we compare the results of local model parameter estimation and filtering with the images obtained by a global estimation approach with a single parameter vector for the

whole image: As expected, with local parameter estimation the MSE decreases for non-stationary textures, like the "straw" texture, but remains unchanged for almost stationary textures, like the "water" texture. The constant mean square error of the "weave" texture is somewhat surprising, but may be explained by the fact that this particular texture cannot be properly described at all by the employed Gauss-Markov model, e.g. due to a wrong resolution or scale at which the analysis is performed.

Parameter estimation window size

Working locally on the image, we have to fix the window size that is used to locally estimate the model parameters. The larger this window, which we call parameter estimation window, the more reliable the values of the extracted model parameters: According to the Cramér-Rao bound, the variance of an estimate decreases proportionally with the number of considered pixels. The need for higher accuracy increases with growing model order, i.e. the number of parameters to estimate, since the stability imposed by the normalization on every parameter weakens. For models of higher order, e.g. seventh order with nineteen parameters, larger windows are necessary than for a third order model with only seven parameters.

An example of SAR image filtering using different parameter estimation window sizes for a seventh order model is illustrated in Fig. 5.9. Parameter estimation with too small windows (5×5 pixels) results in strong artefacts. Considering the amount of parameters to extract, we note that larger windows are required than generally used for statistical filters. For a window of 11×11 pixels the number of artefacts is reduced. A restoration with larger windows is free of artefacts.

It is striking that the visual appearance of images filtered with very large windows of 21×21 pixels or larger is still convincing. Similarly to the behavior of statistical filters, we would expect a strong blurring of the filtered image, which does not occur for the Gauss-Markov model. This property can be partly explained by the fact that no assumptions about stationarity of the mean backscatter are made. Of course, the scene has to be stationary with regard to the model, i.e. concerning its textural properties. However, since textural properties often vary only slowly throughout the image, the employment of larger estimation windows can be justified. We will address cases where this assumption is not given and propose a solution to this problem in chapter six.

We conclude that parameter estimation window sizes must be rather large to obtain results free of artefacts. Consequently, we use windows sizes of at least 21×21 pixels for a seventh and 11×11 pixels for a third order neighborhood system, which yield positive results.

Parameter validity window size

Since usually no abrupt changes in textures occur, it is unreasonable to estimate the model parameter vector for each individual pixel, which is also computationally prohibitive. Instead, we locally determine the textural properties for a number of pixels which are located

MSE	Sand	Water	Straw	Weave
Global Estimation	141	197	133	263
Local Estimation	133	199	115	262

Table 5.6: Comparison of local and global model parameter estimation. The mean square error of the restored Brodatz textures is given for global (estimation window size 256×256 pixels) and local (estimation window size 21×21 pixels, parameter validity window size 11×11 pixels) model parameter estimation. The restoration was performed with a seventh order neighborhood system and $L = 3$.

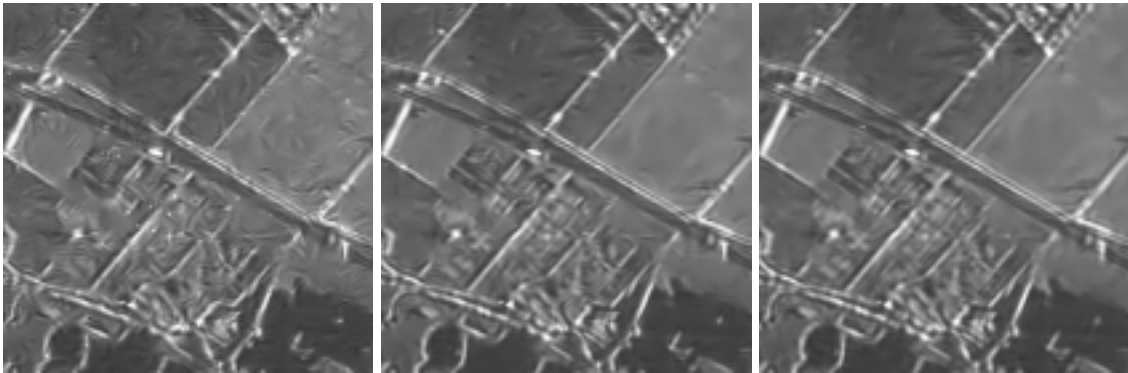


Figure 5.9: Influence of the parameter estimation window size on image restoration. From left to right: Restoration of a part (256×256 pixels) of the aerial SAR image of Fig. 1.1 for different estimation window sizes of 5×5 , 11×11 and 21×21 pixels. For all restorations a seventh order neighborhood system with nineteen model parameters has been used.

in the center of the estimation window. These pixels, for which only one single texture vector is estimated, lie within what we call the parameter validity window. Hence, the resolution of the extracted texture parameters is reduced compared to the original data. In practice, we found windows of 5×5 up to 11×11 pixels to give satisfactory results reducing the computational load for parameter estimation by a factor of 25 or 121, respectively. We note that for most images the quality is hardly influenced by the size of this window, but computing times reduce significantly.

Example of SAR image restoration

In Fig. 5.10 we depict an example of SAR image despeckling using the Model-Based Despeckling algorithm as described so far. An additional processing, which is used to preserve isolated targets, will be addressed in chapter six. A fifth order neighborhood system (13 parameters) was used and we have chosen window sizes of 21×21 and 11×11 pixels for the parameter estimation and for the parameter validity window, respectively. The compu-

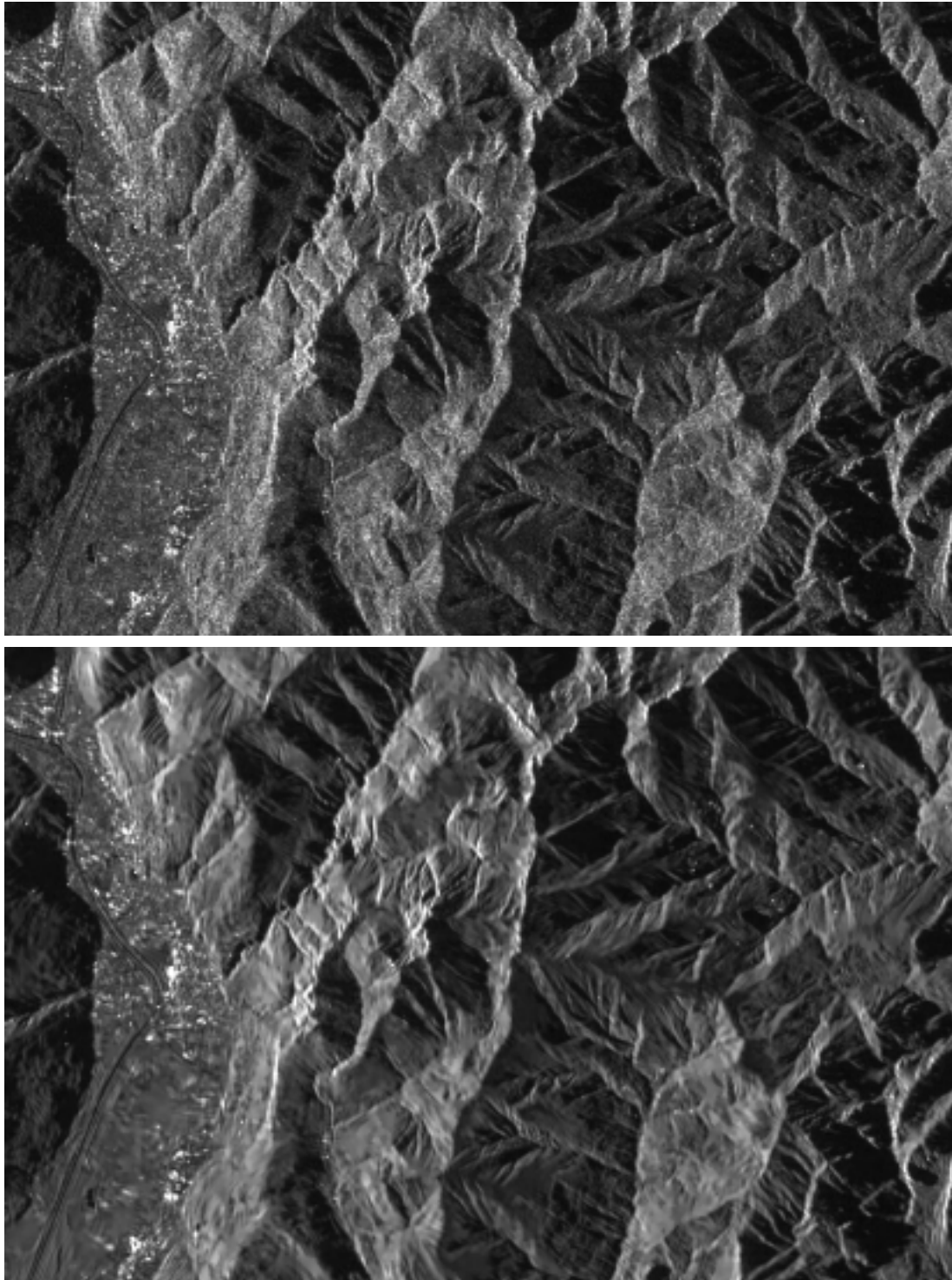


Figure 5.10: Model-based despeckling of an X-SAR image (768×512 pixels, $L = 3.5$) using the Gauss-Markov model. Top: Original X-SAR image over Switzerland with mostly mountainous regions. Bottom: Despeckled image of high visual quality for a fifth order neighborhood of the Gauss-Markov model.

tation time for the displayed result is in the range of 20 minutes on a Sun Ultra 10 with 300 MHz clock frequency. This still rather high processing time illustrates the necessity of the made simplifications, such as the approximation of the Gaussian prior by a square-root Gamma distribution for MAP estimation, the optimization by Occam factor maximization and the reduced resolution of the extracted model parameters, which only make a reasonable application of the proposed technique possible.

The restored image exhibits all important features of the original data and shows a visually attractive restoration of the mountainous areas. Speckle is completely filtered out. However, the image looks slightly blurred. This blurring is reduced with increasing model order, but is characteristic of all Gauss-Markov filtered data. Nonetheless, we consider this a minor drawback, since the effect is rather small. A good image restoration with preservation of textural information is obtained, while spatial resolution is not significantly reduced. A more elaborate comparison and evaluation of the filtering performance on synthetic and real SAR data will be given in chapter seven.

5.4.3 Gauss-Markov MAP Filtering in Different Signal Spaces

We have limited ourselves to the Gauss-Markov model because of its easier applicability compared to non-linear models. However, the space in which this model is best applied has not yet been determined. The employment of a fixed model in one space results in a changed and possibly non-linear model in other spaces. Hence, the behavior of the Gauss-Markov model in different spaces and its influence on the filtering performance is worth a study.

We do not try to analytically derive the corresponding models in different spaces but test the quality of the filtering results. For this purpose, we use the four Brodatz test images again and apply the Gauss-Markov filtering to the corresponding square-root intensity, intensity and log-intensity images. The likelihood function is modified according to the changed properties of the noise.

Square-root intensity images

This is the case we have considered so far and for which all corresponding derivations have been made. For the adaptation of the existing algorithm to the other two cases under study, we change the likelihood function, which results in a change of the MAP equation and of the approximation for the evidence integral. Here, especially the components of the Hessian matrix h_{ii} are affected. The introduced algorithm has to be only slightly modified to work in other signal spaces.

Intensity images

For intensity images, we use the Gamma distribution as in Eq. 2.28 as likelihood function. For the MAP equation we find a third order polynomial

$$x_i^3 - \mu_i x_i^2 + L\sigma^2 x_i - L\sigma^2 y_i = 0, \quad (5.28)$$

which has to be solved to obtain the MAP estimate. In a similar way, the approximation of the evidence is changed and we obtain for the components of \mathbf{H} , which is the Hessian of the posterior,

$$h_{ii} = \frac{2Ly_i}{\hat{x}_{iMAP}^3} - \frac{L}{\hat{x}_{iMAP}^2} + \frac{1}{\sigma^2} \left(1 + \sum_{k,l} \theta_{k,l}^2 \right). \quad (5.29)$$

As far as the iterative optimization is concerned, the step size δ_σ for the Gauss-Markov parameter σ is also modified due to the higher dynamic of the image. Its start value for the estimation is increased accordingly in order to have a better first guess and to speed up convergence.

Log-intensity images

For log-intensity images the likelihood function is given in Eq. 2.45, which is the Fisher-Tippett distribution for $L = 1$. However, for this pdf subsequent derivations of the MAP equation and the approximated evidence cannot be done analytically. As of this, we approximate the likelihood function by a Gaussian, which is a valid assumption for larger values of L , say $L \geq 3$. Mean and standard derivation of the additive almost Gaussian noise in the log-intensity domain can be calculated according to Eq. 2.46 and Eq. 2.47 as a function of the equivalent number of looks L .

In this way, we find a weighted mean for the value that maximizes the local posterior distribution

$$\hat{x}_{iMAP} = \frac{(y_i - \psi(L) + \log L) \sigma^2 + \mu_i \psi'(L)}{\psi'(L) + \sigma^2}, \quad (5.30)$$

and obtain for the components of the Hessian

$$h_{ii} = \frac{1}{\psi'(L)} + \frac{1}{\sigma^2} \left(1 + \sum_{k,l} \theta_{k,l}^2 \right). \quad (5.31)$$

As for the intensity case, these are the only modifications of the original algorithm to be made, apart from an adjustment of the step size and the first guess of σ . The basic iterative approach for model parameter estimation and MAP filtering remains unchanged.

Basically, this case represents a Gauss-Markov model-based filtering of images affected by additive Gaussian noise and is the basis for a digital elevation model (DEM) filter that has been successfully applied during the SRTM mission [SRTM] (cf. appendix D for additional information).

Evaluation of Gauss-Markov MAP filtering in different signal spaces

The modified versions of the algorithm working in different spaces have been used to filter the speckled ($L = 3$) Brodatz textures with a fifth order neighborhood system. The mean square errors of the obtained results are enumerated in Tab. 5.7. We observe that filtering in the square-root intensity domain yields the best results. Especially images filtered in the

MSE	MBD-S	MBD-Z	MBD-I
Sand	145	181	228
Water	240	242	286
Straw	172	217	215
Weave	274	394	647

Table 5.7: Mean square errors for the restoration of the four Brodatz textures in different signal spaces. The mean square errors are given for the original data compared to the results obtained with the MBD approach working in the square-root intensity (MBD-S), in the log-intensity (MBD-Z) and in the intensity domain (MBD-I). We considered artificially speckled data with $L = 3$.

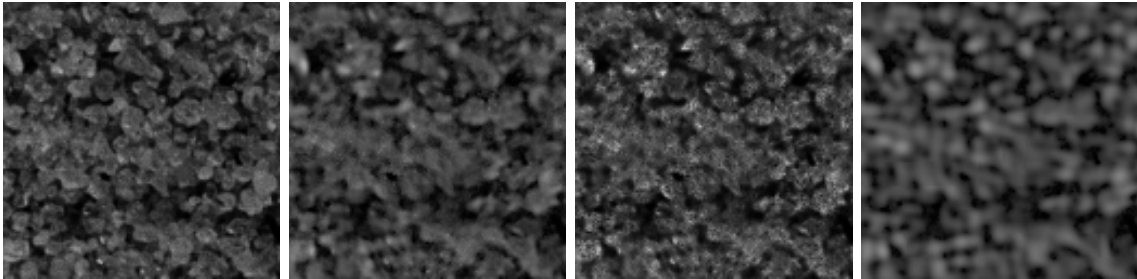


Figure 5.11: Restoration of the "sand" texture (128×128 pixels, $L = 3$) in different signal spaces. From left to right: Original texture, restoration of the speckled texture in the square-root intensity, in the log-intensity and in the intensity domain. All obtained results are displayed in the square-root intensity domain.

intensity domain exhibit a rather smooth visual appearance. Here, weak structures are no more recognized and are not restored by the Gauss-Markov model. On the contrary, results computed in the log-domain still show full detail, but also preserve a noisy appearance. The quality of the images filtered in the square-root intensity space lies between these two extrema. Enough detail is preserved while the noise is sufficiently reduced.

This observation is interesting but not too surprising. The two opposite cases of intensity and log-intensity images, i.e. expansion and reduction of the dynamic range with regard to the square-root intensity image, yield qualitatively contrary results with stronger and weaker smoothing. A representative result for the three discussed cases is displayed in Fig. 5.11.

Filtering in the intensity space does not only yield the worst results, since the dynamic range is apparently too high for the Gauss-Markov model, but the parameter estimation also converges slower. Concerning the MSE, filtering in the log-intensity or in the square-root intensity domain sometimes yields comparable results, however with qualitatively different visual properties. Since the approximation of additive Gaussian noise in the log-space is no more valid for single look data, the filtering performance decreases. For this reason and due to the measured MSE we reject this restoration approach and conclude that the Gauss-

Markov model best behaves in the square-root intensity space. All subsequent filtering will be done in this domain.

5.5 Summary

In this chapter, the following points have been discussed:

- We have derived the MAP equation for Gauss-Markov random fields as a prior and the speckle distribution as likelihood function. Since the optimization space is well-behaved, a deterministic ICM algorithm is used for maximization of the global probability. For faster computation the Gaussian prior was approximated by a square-root Gamma distribution, which was shown to introduce only negligible error.
- For model parameter estimation the computation of the evidence is required. We approximated the posterior product by a multi-variate Gaussian enabling an evaluation of the evidence for a given model parameter vector. We demonstrated the usefulness of this approach for model parameter estimation and proved its accuracy by investigating its behavior in the limiting case of $L \rightarrow \infty$.
- Iterative schemes for evidence maximization as a function of the model parameters have been studied. The introduced method showed to be related to the EM-type algorithm. However, we identified the function to be maximized in the M-step equivalence to be the Occam factor. This allowed an explanation of the qualitatively different results that are obtained with these two approaches.
- We have shown that an introduced bias of the MAP estimator can be easily corrected by a factor, which is equivalent to the noise mean. The mean of the original square-root intensity image is then preserved with sufficient accuracy.
- The relations between evidence, Occam factor, σ and mean square error have been illustrated and we commented on the selection of the optimal model order for Gauss-Markov random fields.
- Since the model order selection process is rather slow, we decided to use a model of fixed higher order, which yielded an improved filtering performance in comparison with lower orders at the cost of an increased computing time.
- SAR image despeckling requires local processing to preserve spatially varying structures. We commented on different window sizes and provided an example for spatially adaptive model-based SAR image restoration.
- We investigated the signal space in which the texture preserving despeckling using Gauss-Markov random fields is best performed. Optimal results in terms of mean square error have been observed in the square-root intensity domain. Intensity data exhibited too much smoothing, whereas rather sharp but noisier images have been obtained by filtering in the log-intensity domain.

6 Extraction of Non-Linear Features

In the last chapter we have derived the model parameter estimation from speckled data for Gauss-Markov random fields and the computation of the MAP estimate of the noise-free SAR scene. However, as noted in chapters three and four, a linear model cannot explain all relevant SAR image features. For many applications the preservation of sharp edges between regions of different uniform mean backscatter is crucial. Other applications require information about isolated targets. For both cases, a linear auto-regressive model is not well suited: The Gauss-Markov model too strongly smoothes edges and blurs isolated strong scatterers. Due to these shortcomings of the proposed approach, an additional processing of the image to capture these non-linear features is required.

We start with a discussion on edge-detection to preserve non-stationarities in mean backscatter and present a specially adapted segmentation approach. The technique is employed to limit the neighborhood system of the Gauss-Markov model at sharp transitions by breaking the bond between a center pixel and its neighbors [Geman84, Jeng, Smits], thus avoiding the smoothing of edges. However, edges also exist in textured areas, where the Gauss-Markov model is perfectly able to describe the image contents. Here, additional structural information does not improve but, on the contrary, degrade the restoration. Hence, a distinction between edges within texture and edges between regions of homogeneous backscatter must be made. This adaptation of the model to different kinds of edges is the subject of the second section.

In the third section we deal with the detection and the processing of strong scatterers. Two cases play an important role: Very strong scatterers disturb a correct model parameter estimation, thus a special pre-processing is required to limit this effect. On the other hand, isolated targets are usually blurred after Gauss-Markov filtering. These targets must be detected and restored in order to preserve information. Both tasks rely on a statistical image analysis and result in a detection scheme which is controlled by a maximum false alarm rate. We close this chapter with a description of the resulting full Model-Based Despeckling algorithm.

6.1 Edge-Detection for Model-Based Despeckling

In order to preserve non-stationarities in mean backscatter, we apply an edge-detection step in our algorithm. The extracted edge information is used to drive the Gauss-Markov

despeckling. Similar approaches also exist for many statistical filters, like the Lee or the GGMAP filter, which use this kind of information in order to limit the size of their estimation window. Hence, a special processing is often employed to fix deficiencies of the underlying model assumptions, which are necessarily kept simple for computational reasons. We commented on this point in preceding chapters referring to the high complexity of non-linear discontinuity adaptive Gibbs random fields.

6.1.1 Region-Growing Backscatter Segmentation

Since the preservation of edges between regions of homogeneous backscatter is of particular interest for applications and since a full texture segmentation is computationally too challenging, we only search for non-stationarities, i.e. edges, in mean backscatter. A lot of work has already been performed in this field and two different approaches are possible: Edge-detection and backscatter segmentation. Since edge-detection techniques usually do not yield closed borders, we prefer the latter approach. We require closed edges in order to avoid the spread of information from one stationary region to another through "holes" in the detected boundary. However, in principle any reliable edge-detector or segmentation algorithm from which edge information can be derived may be applied.

For our purpose, we have adopted the basic idea of a statistical segmentation algorithm proposed in [Cook96] and adjusted this approach to fulfill our requirements for Gauss-Markov edge-preserving filtering, such as a pixel-accurate and clean segmentation.

Maximum likelihood segmentation

Assuming that a SAR image is composed of r regions R_i with uniform backscatter values, μ_{x_0} to $\mu_{x_{r-1}}$, and under the assumption of statistical independence, the likelihood of the observed data y can be written as

$$\begin{aligned} p(Y = y | \mu_{x_0}, \dots, \mu_{x_{r-1}}) &= \prod_{i=0}^{r-1} \prod_{j \in R_i} p(Y_j = y_j | \mu_{x_i}) \\ &= \prod_{i=0}^{r-1} \prod_{j \in R_i} 2 \left(\frac{y_j}{\mu_{x_i}} \right)^{2L-1} \frac{L^L}{\mu_{x_i} \Gamma(L)} \exp \left(-L \left(\frac{y_j}{\mu_{x_i}} \right)^2 \right), \end{aligned} \quad (6.1)$$

where $p(Y_j = y_j | \mu_{x_i})$ is the likelihood function for square-root intensity data of Eq. 4.11. Equivalent results follow with Eq. 4.5 for the use of intensity data.

By maximizing Eq. 6.1 with regard to r different backscatter values and by assigning a label in form of the most probable value out of μ_{x_0} to $\mu_{x_{r-1}}$ to each observed pixel y_j , the image y is segmented into r classes. This is a classical maximum likelihood segmentation if the mean values μ_{x_0} to $\mu_{x_{r-1}}$ are known in advance. However, neither the number of classes r , nor their mean gray-values μ_{x_i} are known in practice. Additionally, a maximum likelihood segmentation does not meet our requirement of high accuracy, since the assigned labels do not form homogeneous and closed regions, but, on the contrary, give a rather noisy appearance making the extraction of reliable edge information impossible.

Introduction of prior knowledge by region-growing optimization

The results of the maximum likelihood segmentation can be improved by introducing prior knowledge, e.g. in form of the often used Potts model, corresponding to a MAP segmentation. However, the Potts model requires the employment of simulated annealing (cf. appendix B) for optimization and an additional model parameter estimation step. Consequently, we adopt a different approach by introducing prior knowledge by means of a region-growing optimization of Eq. 6.1, which favors the formation of homogeneous labels.

We perform the maximization by iteratively estimating the classes' mean gray-values μ_{x_i} and stochastically attaching region labels to all image pixels. The number of existing classes r is assumed to be known. The prior to have uniformly labeled regions avoiding a noisy maximum likelihood segmentation is encapsulated in the following updating scheme:

1. Starting with a completely random segmentation into r classes, the whole image is scanned pixel by pixel. The number of classes r is fixed by the user. Initial mean values μ_{x_i} are computed for each class.
2. For each pixel which is located at a current border, the probability according to Eq. 6.1 of that pixel to belong to the class of a randomly chosen neighbor of a different class is calculated and compared to the probability of its current class.
3. The decision for the new class of each border pixel is sampled from these two probabilities and the estimates for the class mean values μ_{x_0} to $\mu_{x_{r-1}}$ are recomputed from the new segmentation.
4. The procedure is re-iterated until convergence in form of a low number of changes in each iteration is reached.

In this way, a homogeneous segmentation is achieved. However, pixels inside closed labels are no more considered, an updating is only performed at borders between different labels. This may result in a loss of detail, since the algorithm is unable to rediscover previously lost small segments within larger ones. Furthermore, very small segments of several pixels, such as scatterers, may not be detected if the initial random configuration represents an unfavorable starting position. Nonetheless, the approach works fine for structures of small sizes, if r is not chosen too high, and is independent of region-shape or edge-orientation.

Example of region-growing segmentation

An example of the presented approach is illustrated in Fig. 6.1. We have generated a synthetic image of eight different gray-levels, as depicted in the histogram of Fig. 6.3. The original image, which is composed of squared regions of sizes 64×64 down to 8×8 pixels, was rotated by fifteen degrees in order to demonstrate the technique's independence on edge-orientation, which might affect approaches based on pre-defined window sizes and edge-patterns. From the histogram of the speckled data ($L = 3$) the number of classes in the image cannot be determined.

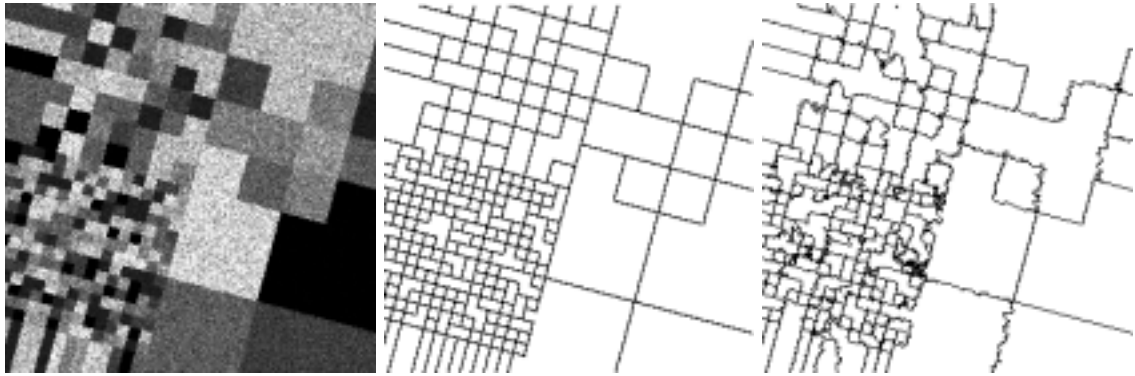


Figure 6.1: Region-growing segmentation for edge-detection. From left to right: Speckled synthetic image (256×256 pixels, $L = 3$), original edge-map, edge-map extracted from the segmentation with $r = 8$ after 200 iterations.

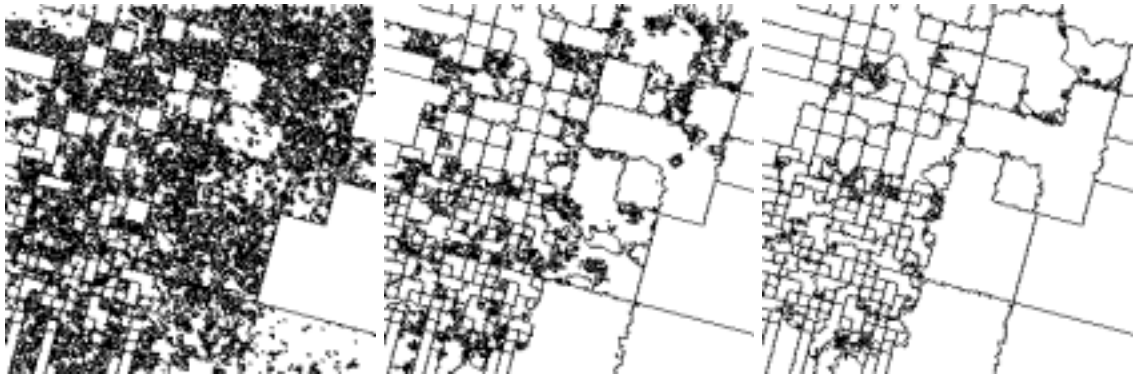


Figure 6.2: Evolution of the edge-detection process. From left to right: The edges resulting from the segmented image with $r = 8$ are given after 20, 50 and 100 iterations.

Edges extracted from the segmented image after 20, 50 and 100 iterations, i.e. full image sweeps, with $r = 8$ are displayed in Fig. 6.2. We see that dark regions are segmented first and bright regions last, which is due to the increasing speckle variance and the decreasing ratio of the mean intensity between brighter regions making segmentation more difficult.

The detected edges derived from the segmentation after 200 iterations and the original edge map are shown in Fig. 6.1. We see that the correspondence is rather high, although some edges between regions of lower contrast have not been found. Considering the estimated mean values μ_{x_i} for the eight classes, as enumerated in Tab. 6.1, we note that one class of high intensity (C6, $\mu_x = 160$) has not been identified and was mixed with class seven (C7) of $\mu_x = 192$. This merging occurred due to the small amplitude ratio of 1.2 of these two classes and the fact that class seven highly dominates class six in terms of the number of pixels. Class six is by far the smallest class in the test image. However, in spite of this shortcoming, the segmentation is unaffected by the introduced image rotation and also yields convincing results for the smallest segments of 8×8 pixels.

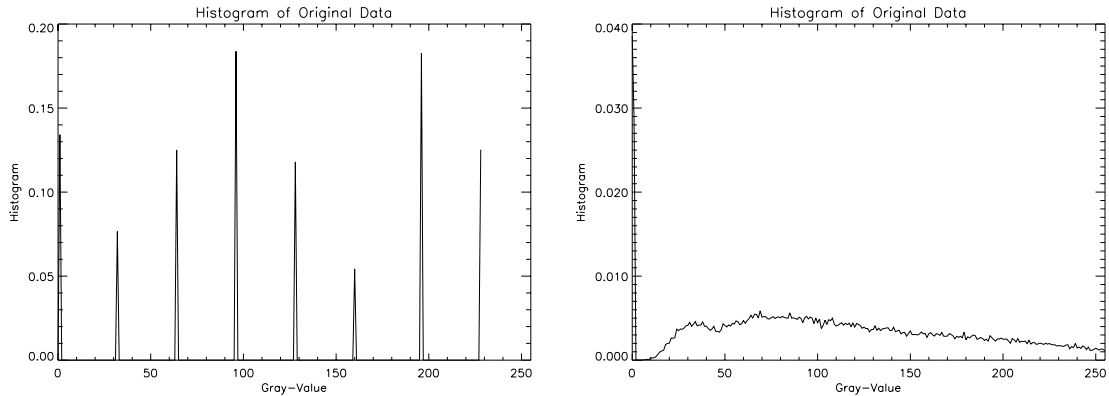


Figure 6.3: Histograms of the noise-free and the speckled test image. Left: Histogram of the noise-free image exhibiting eight classes. Right: Histogram of the speckled data not allowing the determination of the number of classes.

μ_{x_i}	C1	C2	C3	C4	C5	C6	C7	C8
Original	1	32	64	96	128	160	192	224
Estimated	1	32	64	97	136	194	196	225

Table 6.1: Original and estimated class mean values obtained by the region-growing segmentation. The mean values of the eight classes C1-C8 of the synthetic image and their estimates obtained from the segmentation algorithm are enumerated.

Number of assumed classes r	4	6	8
Percentage of inaccurate/false edges to estimated edges	31%	34%	36%
Percentage of correctly detected edges to existing edges	59%	61%	65%

Table 6.2: Edge-detection accuracy of the region-growing segmentation. The percentage of inaccurate edges with regard to the number of estimated edges and the percentage of correctly detected edges with regard to the number of existing edges are given. For correctly detected edges, pixel accuracy is required.

Concerning the edge-detection accuracy, we performed the segmentation for $r = 4$, $r = 6$ and $r = 8$ classes and measured the percentage of correctly identified edges compared to the number of edge-pixels in the original edge-map and the percentage of non-corresponding or wrong edges with reference to the number of edges obtained by the segmentation. Correspondence of edges is only given for a pixel-accurate detection, a shift of one pixel results in non-correspondence. The found values are given in Tab. 6.2: With r growing, the edge-detection rate increases. For $r = 8$ about 65% of all existing edges have been identified with pixel accuracy. However, the number of false edges increases as well, as the number of found edge pixels grows with r . The percentage of false edges (with shifts of possibly

only one pixel) from the number of found border pixels is about 36% for $r = 8$. One-pixel shifts especially occur at edges between regions of low or moderate contrast, where the segmentation tends to follow the noise. Border pixels of higher intensity are attributed to the class of higher mean, and vice-versa.

In view of the fact that we require pixel accuracy for edge-detection, the achieved results are considered satisfactory. Even if the number of existing classes is under-estimated, the percentage of correctly identified borders is sufficient, with the false alarm rate being lower than for larger values of r (cf. first row of Tab. 6.2). Since the edge-detection for Gauss-Markov filtering will be performed locally and not globally, we use a value of $r = 3$ for detection of significant edges in windows of typically 21×21 pixels.

6.1.2 Detection of False Alarms and Region-Merging

For a further verification of the existence of edges found by the region-growing segmentation approach, we rely on a statistical ratio edge-detection proposed by [Touzi88]. The conditional pdf of the so-called bounded ratio detector ρ of two regions consisting of N_1 and N_2 pixels, respectively, is given for square-root intensity data by

$$p(\rho|C) = \frac{2 \Gamma(L(N_1 + N_2))}{\rho \Gamma(N_1 L) \Gamma(N_2 L)} \left(\frac{(N_2/N_1)^{N_2 L} (\rho^2 C)^{N_1 L}}{(\rho^2 C + N_2/N_1)^{L(N_1+N_2)}} + \frac{(N_1/N_2)^{N_1 L} (\rho^2/C)^{N_2 L}}{(\rho^2/C + N_1/N_2)^{L(N_1+N_2)}} \right), \quad (6.2)$$

where $C = \mu_{y_2}/\mu_{y_1}$ is the empirically measured contrast of these regions and ρ lies between zero and one. To determine a decision threshold ρ_{max} , below which an edge is assumed to exist for a measured contrast $\max(C, 1/C)$, the probability of false alarm P_{fa} is used. The probability that a homogeneous area with $C = 1$ is assigned to an edge is given by

$$P_{fa} = \int_0^{\rho_{max}} p(\rho|1) d\rho. \quad (6.3)$$

Hence, the decision threshold can be determined by specifying a minimum value for P_{fa} . In our algorithm, we have chosen $P_{fa} = 0.0001$ for the verification of the previously extracted edges. However, we do not compute the contrast of two complete segmented areas, but only consider gray-values for the computation of C which are no more than two pixels away from the edge between the two segments under investigation. In this way, we are able to reject edges found on a monotonically and smoothly increasing cross-section being equivalent to a ramp in the one-dimensional case. Borders between all found segments are verified and neighboring segments are eventually merged if the existence of an edge is not confirmed.

6.2 Edge Information for Gauss-Markov Despeckling

Extracted edges from the region-growing segmentation can be used to limit the neighborhood size of the Gauss-Markov model in order to preserve non-stationarities in mean

backscatter. However, we do not consider edges between differently textured regions, since in real SAR data, textural differences are rather small and transitions are usually not very sharp. Moreover, a pixel accurate texture-segmentation is already a difficult problem when rather simple texture models are employed [Oliver96].

For the edge information obtained from the segmentation we can distinguish three cases:

1. A false alarm is given, i.e. no edge is present. This may especially happen, if the number of assumed classes r is higher than the actual number of classes in the image.
2. An edge is found that is part of a stationary texture which can be modeled by Gauss-Markov random fields. In this case, the neighborhood of the model should not be changed, the edge information should not be considered.
3. An edge is detected, which describes a non-stationary image behavior that cannot be modeled with Gauss-Markov random fields. Here, the edge information should be taken into account for filtering by locally adjusting the Gauss-Markov neighborhood system.

While several methods exist to determine the occurrence of case one, a discrimination between case two and three is a difficult problem. Non-stationarity in mean backscatter, which is detected by the employed segmentation approach, does not necessarily require the adaptation of the neighborhood system. Even on the contrary, the Gauss-Markov model is not based on the assumption of uniform backscatter, which represents its main advantage compared to most statistical filters.

6.2.1 Adaptation of the Gauss-Markov Neighborhood System

The smoothing of edges is caused by pixels that linearly contribute to the model prediction $\mu_i = \sum_{k,l} \theta_{k,l} (x_{k,l} + x'_{k,l})$ and which belong to regions with different statistical properties. Hence, an information exchange between regions of different statistical contents results in a blurring or other artefacts. To avoid this problem, the computation of μ_i has to be performed only on pixels within a stationary region, where stationary is meant in terms of Gauss-Markov random fields. We will see in the next section that the determination of this stationarity is not trivial.

If a boundary between stationary regions has been identified, all pixels in the original neighborhood of x_i which do not belong to the same stationary region as x_i are excluded from the computation of μ_i , i.e. their parameter $\theta_{k,l}$ is set to zero. This is illustrated for a second order neighborhood in Fig. 6.4. To guarantee stability of the Gauss-Markov random field the modified parameter vector $\tilde{\theta}$ has then to be re-normalized, i.e. $\sum_{k,l} \tilde{\theta}_{k,l} = 0.5$. The approach for the computation of the corresponding MAP estimate remains unchanged. Consequently, this processing only affects the calculation of μ_i for pixels whose neighborhood extends into regions of different statistical properties.

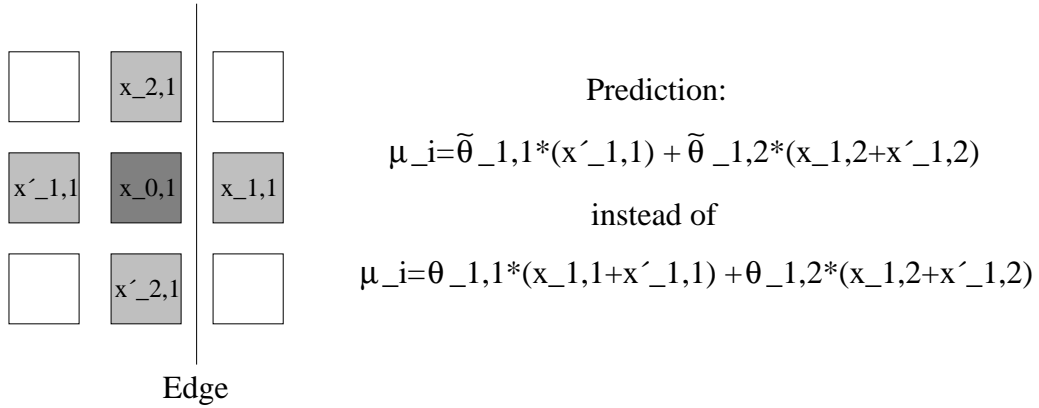


Figure 6.4: Adaptation of the Gauss-Markov neighborhood system. The computation of the new predicted gray-value μ_i is illustrated for a second order neighborhood system, where pixel $x_{1,1}$ belongs to a different region separated by a detected edge. The model parameters are re-normalized to guarantee stability.

6.2.2 Introduction of Edge Information

After the verification of detected edges, the task remains to decide on how to interpret this extracted information. Here, we take advantage of the fact that the preservation of borders is especially of interest for regions of different homogeneous backscatter, such as agricultural areas. As stated, the modeling of such edges cannot be satisfactorily performed by the employed texture model. Consequently, we intend to use edge information only to describe borders between areas of uniform backscatter. This implies the modification of the neighborhood at identified borders and an adjustment of the texture parameter vector θ , in order to describe homogeneous areas instead of textural information.

Restoration under two different assumptions

Following this reasoning, any image allows the restoration with two different models according to the following assumptions:

1. A MAP estimate \hat{x}_{MAP_1} is computed with Eq. 5.6 using the extracted texture parameters without any information about edges, i.e. we assume that the noise-free image can be explained by the texture model alone even if edges are present.
2. The noise-free image is assumed to consist of several regions of different uniform backscatter. The image is filtered using Eq. 5.6 with an adaptive neighborhood system at the region borders resulting in an estimate \hat{x}_{MAP_2} . The model parameters are set to values that cause a pure averaging of the amplitude image, i.e. $\theta_{k,l} = \theta = const$ for all k, l within a region label, where θ is determined by the model order and the number of parameters. σ is kept as estimated.

Synthetic examples for the restoration under these assumptions are shown in Figs. 6.5 and 6.6. For the "chess-board" image the blurring becomes obvious when no edge information is employed (\hat{x}_{MAP_1}). On the other hand, the use of edge information results in a much better restoration \hat{x}_{MAP_2} since the made assumption of homogeneous regions is valid. However, for the "straw" image, the restoration \hat{x}_{MAP_1} without edge information gives the best result: Additional structural information and the assumption of homogeneous regions leads to sharp edges but gives an unnatural appearance (\hat{x}_{MAP_2}). We remark that statistical filters using geometrical information about edges provide results similar to the images on the right of Figs. 6.5 and 6.6. Textural information as in Fig. 6.6 (center) cannot be preserved.

We conclude that the correct choice of the filtering assumption highly influences the image quality. Either structural information in form of edges or textural information has to be preserved. The problem consists in finding out whether edges are due to texture as in the "straw" image or if they are better explained by the assumption of uniform backscatter regions as in the "chess-board" example.

6.2.3 Fusion of Texture and Edge Information

Since a correct model selection within the Bayesian framework cannot be easily applied, we adopt a different and computationally less demanding approach to determine the assumption under which the filtering should be performed.

After having checked the validity of the found edges we apply a restoration under both assumptions, i.e. textured areas and homogeneous regions separated by edges, and fuse these two results in order to obtain the final image. This fusion is controlled by the coefficient of variation measured in each obtained segment. The whole selection process works as follows:

1. Perform a region-growing segmentation to separate areas of different homogeneous backscatter.
2. Verify the validity of the found edges by the ratio edge-detector and eventually merge certain segments.
3. Perform a normal Gauss-Markov texture parameter estimation and compute the corresponding MAP estimation \hat{x}_{MAP_1} of the noise-free image.
4. Perform a MAP filtering \hat{x}_{MAP_2} by adapting the neighborhood of the Gauss-Markov model according to the detected edges and use model parameters that perform a simple averaging. The value of the Gauss-Markov parameter σ is taken from the preceding parameter estimation.
5. Measure the coefficient of variation $CV^2 = \sigma_y^2 / \mu_y^2$ in each segment obtained by the region-growing algorithm. Since closed boundaries are needed for this step, this is the main reason to apply a segmentation instead of an edge-detection algorithm.
6. For each segment whose coefficient of variation is equal or lower than the expected value of Eq. 2.39, replace the gray-values of \hat{x}_{MAP_1} by the values of \hat{x}_{MAP_2} to generate the final result in \hat{x}_{MAP_1} .

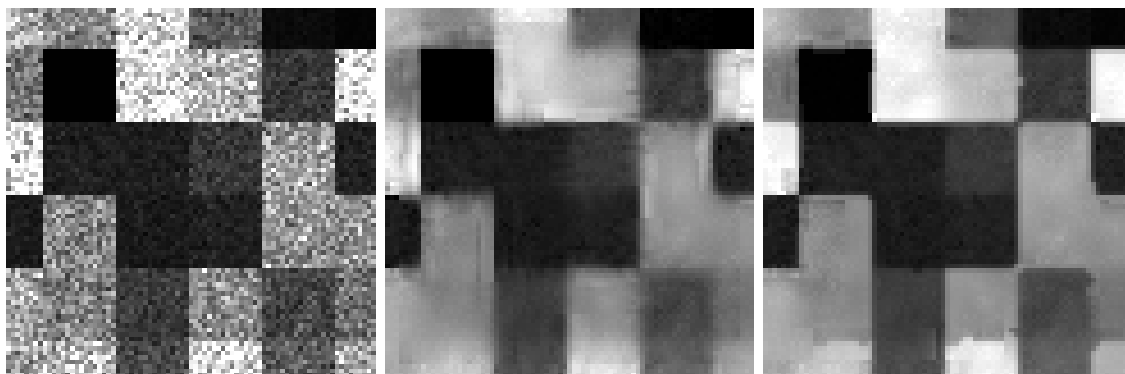


Figure 6.5: Restoration under two different assumptions. From left to right: Speckled "chess-board" image (80×80 pixels, $L = 3$), Gauss-Markov texture restoration \hat{x}_{MAP_1} without using edge information (MSE=259), restoration \hat{x}_{MAP_2} using edge information under the assumption of homogeneous backscatter (MSE=160). \hat{x}_{MAP_2} gives the best restoration.

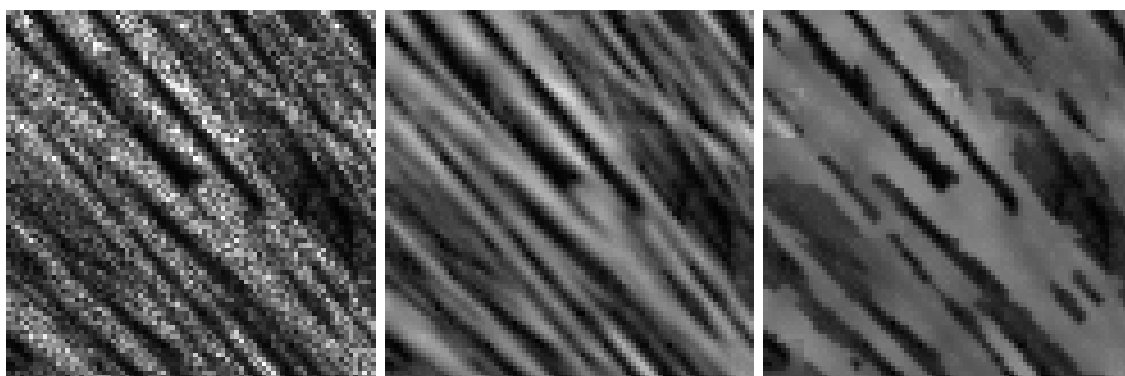


Figure 6.6: Restoration under two different assumptions. From left to right: Speckled "straw" image (80×80 pixels, $L = 3$), Gauss-Markov texture restoration \hat{x}_{MAP_1} without using edge information (MSE=141), restoration \hat{x}_{MAP_2} using edge information under the assumption of homogeneous backscatter (MSE=388). \hat{x}_{MAP_1} gives the best restoration.

We found this approach to work much better than other experimented approaches based on the likelihood of the restoration or on the cross-entropy of the ratio image y/\hat{x}_{MAP} . Even for difficult cases the presented approach selects the correct restoration as shown in Fig. 6.5 (right) and Fig. 6.6 (center). This is very important in order to obtain a visually attractive image restoration with a low mean square error.

6.3 Extraction and Preservation of Strong Scatterers

Strong scatterers or targets, which represent an important feature in SAR images, require an additional treatment for detection and preservation after filtering, since they cannot be detected by the approach described above. Isolated targets are strongly smoothed by the filter and also disturb its model parameter estimation. Thus, we envisage a two-step approach to deal with this special SAR image feature: Removal of very strong targets as a pre-processing and detection and re-insertion of blurred and previously removed targets as a post-processing step after the Gauss-Markov filtering.

6.3.1 Pre-Processing of Targets for Parameter Estimation

The problem we are faced with is depicted in Fig. 6.7. We have generated a synthetic SAR image of a homogeneous cross-section with several targets, which is filtered with the introduced approach. The result on the right of the top row shows a blurring of these targets, which ideally should be unaffected by the filtering, and a rather weak smoothing of their surrounding homogeneous areas.

This behavior is due to the Gauss-Markov model, which tries to capture the full image contents. However, a complex and highly non-stationary scene cannot be sufficiently modeled by an auto-regressive process: The smoothing around the targets is rather weak because of a high value of the Gauss-Markov σ , which reflects the model's inability to properly describe the noise-free data. Hence, neither the targets nor the surrounding homogeneous cross-section are restored in a satisfactory way.

Detection of strong scatterers

To solve this problem, we introduce a pre-processing step intended to remove strong targets before filtering with the Gauss-Markov model. This pre-processing, with a target detection based on [Lopes93], is done as follows:

1. For a given center pixel calculate the mean μ_{inner} on a given neighborhood (e.g. $N_1 = 2 \times 2 = 4$ pixels) and the mean μ_{outer} on a larger surrounding rectangular area (e.g. $N_2 = 4 \times 4 - 2 \times 2 = 12$ pixels).
2. If the ratio μ_{inner}/μ_{outer} exceeds a threshold $1/\rho_{max}$, replace the gray-values of the pixels in the inner window by their original values divided by this ratio.
3. Perform this processing for all pixels by scanning the whole image.

It is important to carefully select the threshold $1/\rho_{max}$ in order to remove only real scatterers and no speckle of high variance. We intend to only attenuate scatterers that really disturb the parameter estimation. Isolated targets, that are not very bright, i.e. whose gray-values lie within the speckle distribution of the surrounding uniform area, hardly affect the model parameter values, since they are interpreted as noise. The employed threshold $1/\rho_{max}$ varies

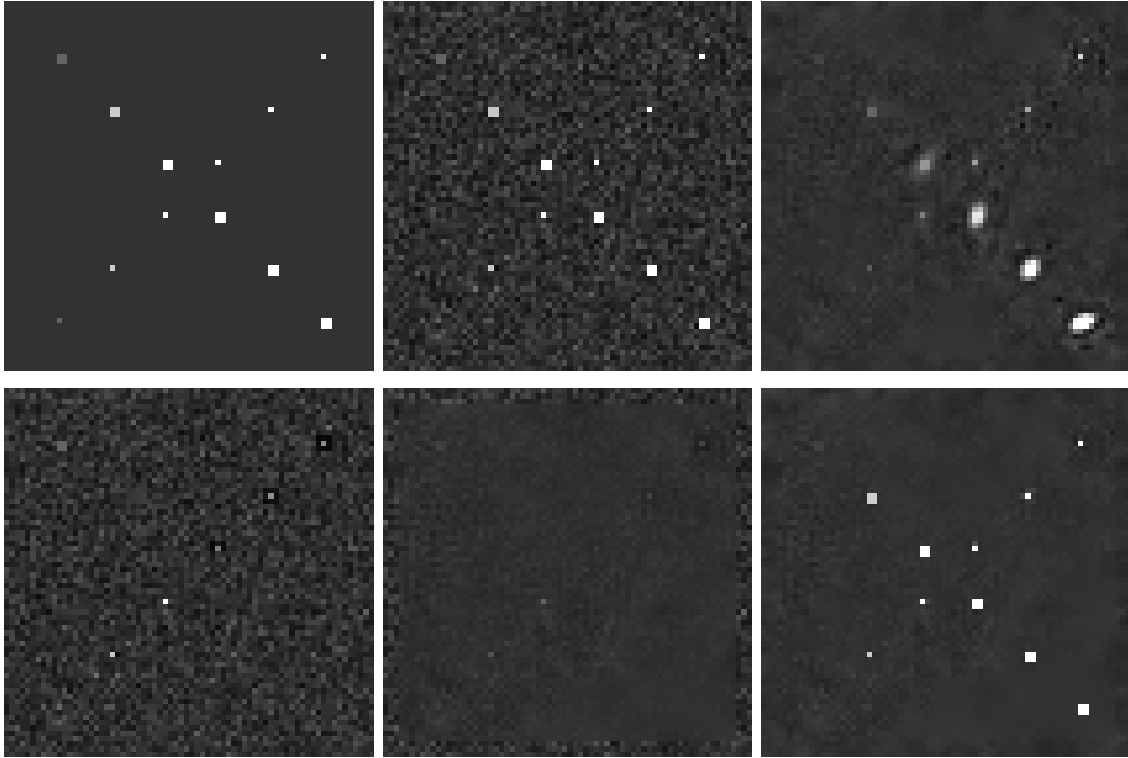


Figure 6.7: Detection and restoration of strong targets. Top: Original target map (70×70 pixels), targets on a speckled background ($L = 3$), restoration with the Gauss-Markov model of seventh order. Bottom: Image after pre-processing for target removal, Gauss-Markov filtered image, final restoration after post-processing. The gray-values of the targets increase from left ($x_i = 100$) to right ($x_i = 600$). The background gray-value is $x_i = 50$.

with the speckle strength L and is chosen by means of a constant false alarm rate according to Eqs. 6.2 and 6.3 with $C = 1$ [Lopes93].

For the example given in Fig. 6.7 with $N_1 = 4$, $N_2 = 12$ and $L = 3$, we have chosen a threshold of $1/\rho_{max} = 3$ resulting in a false alarm rate for the detection of homogeneous regions below $P_{fa} \approx 6 \times 10^{-8}$. A threshold of $1/\rho_{max} = 2$ is still a reasonable choice with $P_{fa} \approx 0.0002$. The evolution of the false alarm rates as a function of the chosen threshold $1/\rho_{max}$ and the speckle level L is displayed in the left plot of Fig. 6.8. Naturally, P_{fa} decreases with L growing.

Despeckling of pre-processed data

The effects of this pre-processing in combination with despeckling are illustrated in the bottom row of Fig. 6.7. The Gauss-Markov filtering is applied to almost scatterer-free data (left) and we observe that the intermediate despeckled image is well smoothed (cen-

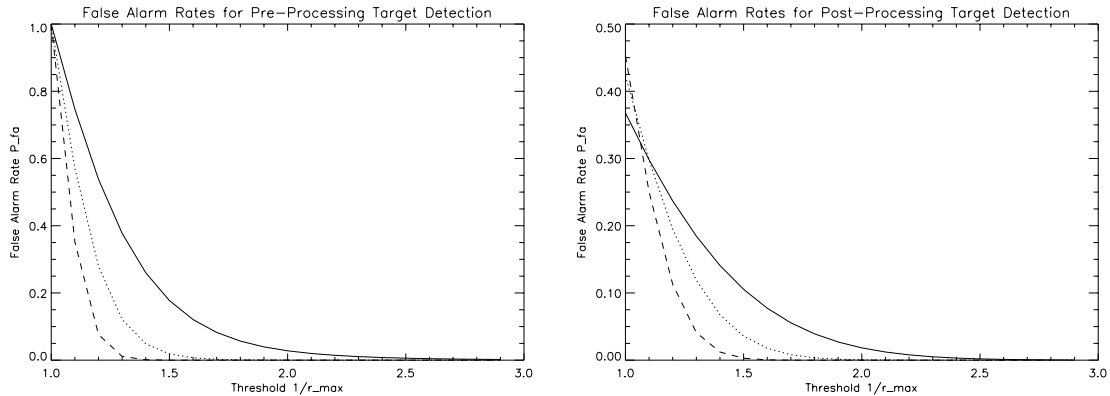


Figure 6.8: False alarm rates as a function of the thresholds for pre- and post-processing. Left: False alarm rates for the pre-processing threshold of the ratio μ_{inner}/μ_{outer} with $N_1 = 4$ and $N_2 = 12$. Right: False alarm rates for the post-processing threshold of the ratio y_i/\hat{x}_{iMAP} . Plots are given for $L = 1$ (solid), $L = 3$ (dotted) and $L = 8$ (dashed).

ter). As expected, all scatterers unaffected by the pre-processing have been smoothed and disappeared.

In a last step, the original gray-values of the eliminated scatterers are simply put back into the final result. Note that only original gray-values y_i that have higher values than their counterparts \hat{x}_{iMAP} of the Gauss-Markov filtered data are re-inserted in order to avoid dark pixels around strong scatterers. This additional step is necessary to compensate for the decreased resolution which is introduced by the employment of windows for the target detection.

6.3.2 Post-Processing for the Detection of Blurred Targets

Targets which are not detected in the pre-processing will be completely blurred after Gauss-Markov despeckling, as illustrated in the center image in the bottom row of Fig. 6.7. To compensate for this effect, we introduce a special post-processing of the filtered data, which is based on the ratio image y/\hat{x}_{MAP} . In the optimal case, i.e. $\hat{x}_{MAP} = x$, which is the true noise-free image, this ratio is identical to the speckle n affecting the observation y . Hence, a filter-induced structural degradation, such as a blurring of edges or targets, can be detected in this ratio. For the given example, the resulting ratio image is visually similar to the center image of the top row of Fig. 6.7.

Consequently, in order to detect blurred scatterers, we compare the ratio at each site i to the pdf of the speckle process. Values which are beyond a fixed threshold are assumed not to be caused by speckle but by structural degradation. The new post-processing threshold $1/\rho_{max}$ is again chosen according to a constant false alarm rate, which can be computed

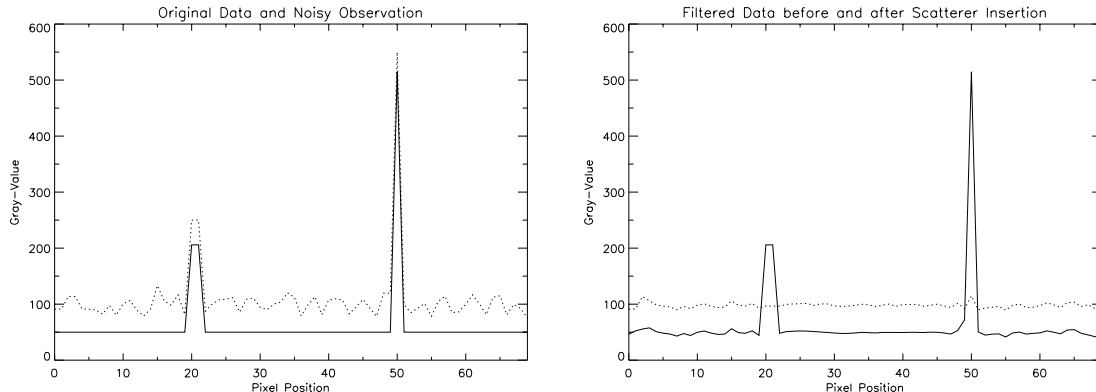


Figure 6.9: Cross-sections of the target images. Left: Original image (solid), targets on a speckled background (dotted). Right: Restoration of the pre-processed image (dotted), final restoration after post-processing (solid). For visualization, the dotted plots are vertically shifted by a value of 50.

from the speckle distribution with

$$P_{fa} = 1 - \frac{2L^L}{\Gamma(L)} \int_0^{1/\rho_{max}} n^{2L-1} \exp(-Ln^2) dn, \quad (6.4)$$

and should be chosen to be quite small. The integration is performed over the square-root intensity speckle distribution of Eq. 2.36 with $\mu_S = 1$. In our implementation we have chosen $P_{fa} = 0.0005$ corresponding to a threshold of $1/\rho_{max} = 2.215$ for $L = 3$. This guarantees the detection of real targets and avoids the preservation of pixels with speckle of higher variance. The false alarm rates for the post-processing step are depicted in the right plot of Fig. 6.8. As before, the choice of the threshold for negligible error rates depends on the equivalent number of looks L of the speckle noise.

The final restoration is given in the last image of Fig. 6.7. In the post-processing, previously undetected targets that were blurred out are identified in the ratio image and are put back into the final result together with the targets removed in the pre-processing. Note however, that the two weak targets on the far left of the original image with $x_i = 100$ are lost, since they are not distinguished from the speckle noise.

For a more quantitative evaluation of the target restoration we display several cross-sections of the same experiment in Fig. 6.9: The restored targets do not show any blurring, their amplitude is exactly preserved and the surrounding pixels are highly smoothed without artefacts around the scatterers.

6.4 Outline of the Model-Based Despeckling Algorithm

This section represents a short algorithmic summary of the basic parts of the Model-Based Despeckling algorithm. To prepare the filter evaluation discussed in the next chapter, we

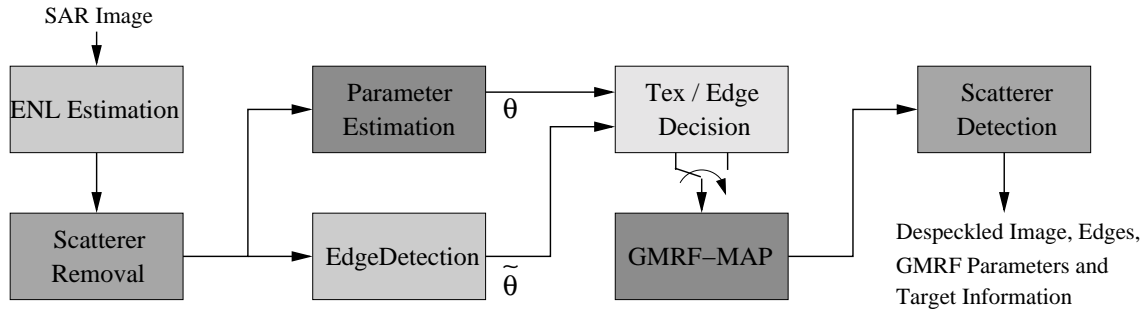


Figure 6.10: Flowchart of the full Model-Based Despeckling algorithm. The basic pre- and post-processing steps in combination with model parameter extraction and MAP estimation are illustrated.

describe the full approach and give its main parameters. Finally, we provide a SAR filtering example illustrating the incorporation of additional structural information.

6.4.1 Algorithmic Flowchart

A general flowchart of the implemented algorithm is presented in Fig. 6.10. We shortly comment on the different blocks to summarize their function. The processing is window-based and works by a pixel-wise (only for the highest spatial resolution of the parameter vector) scanning of the whole image. The filter performs a locally adaptive despeckling of the image considering its non-stationary behavior and a varying scene content.

ENL Estimation If unknown, the equivalent number of looks L is determined. L is computed in the most homogeneous detected area with a size of 35×35 pixels. This approach was found to give accurate results unless the image is heavily textured.

Scatterer Removal Prior to filtering, strong targets are detected and removed from the image to be despeckled as described above in this chapter.

Parameter Estimation The parameters of the Gauss-Markov random field are locally estimated by means of the presented iterative evidence maximization approach. The size of the parameter estimation window is typically 21×21 pixels for a seventh order neighborhood. One single model parameter vector is usually estimated for every 7×7 pixels.

Edge Detection The local area under investigation is segmented into $r = 3$ classes providing border information. The segmentation is based solely on the mean backscatter and does not consider textural properties. Eventually, several classes are merged together to reduce the edge-detection false alarm rate.

Tex/Edge Decision The homogeneity of the detected segments is verified by an analysis of the measured coefficient of variation. The decision for textured or homogeneous

areas with edges controls the following computation of the MAP estimate, i.e. the fusion of the two different estimates.

GMRF-MAP For textured areas the MAP estimate is computed with the previously estimated model parameter vector θ and the full neighborhood system. For assumed homogeneous regions the neighborhood is limited at the borders and the directional parameters are chosen to result in a simple averaging of the observed image.

Scatterer Detection As outlined in this chapter, previously removed scatterers are re-inserted. Blurred additional targets are detected in the ratio image and are also restored.

6.4.2 Model-Based Despeckling with Structural Information

In Fig. 6.11 we exemplify the employment of structural information for the proposed filtering approach with Gauss-Markov random fields. The upper original ERS SAR image is overlaid with the extracted edge information. The lower image shows the filtering result, where this information has been exploited. Despite the used Gauss-Markov model, edges between regions of homogeneous backscatter are well preserved due to the locally adaptive neighborhood system. The quality of the restoration of textured regions, such as mountains, is not affected by this additional processing. The preservation of isolated targets can be verified in the lower left part of the shown images.

6.5 Summary

In this chapter, the following points have been discussed:

- Filtering with linear Gauss-Markov random fields requires an additional processing for edge-preservation. Borders found with pixel accuracy and independent of edge-orientation are used to adapt the neighborhood system avoiding the blurring of edges between regions of different first order statistics.
- A region-growing backscatter segmentation is employed which is an extended ML segmentation favoring the formation of homogeneous, closed segments. Segmentation is preferred to edge-detection since continuous edges are required. Borders between segments are verified and different segments are eventually merged.
- The restoration can either be performed with the originally estimated model parameters or by taking into account the identified edges between regions of different mean backscatter. The latter restoration is performed with a locally adaptive neighborhood system.
- The decision for any of these two assumption is based on a local analysis of the coefficient of variation in order to verify the hypothesis of edges between areas of homogeneous backscatter or between textured regions.

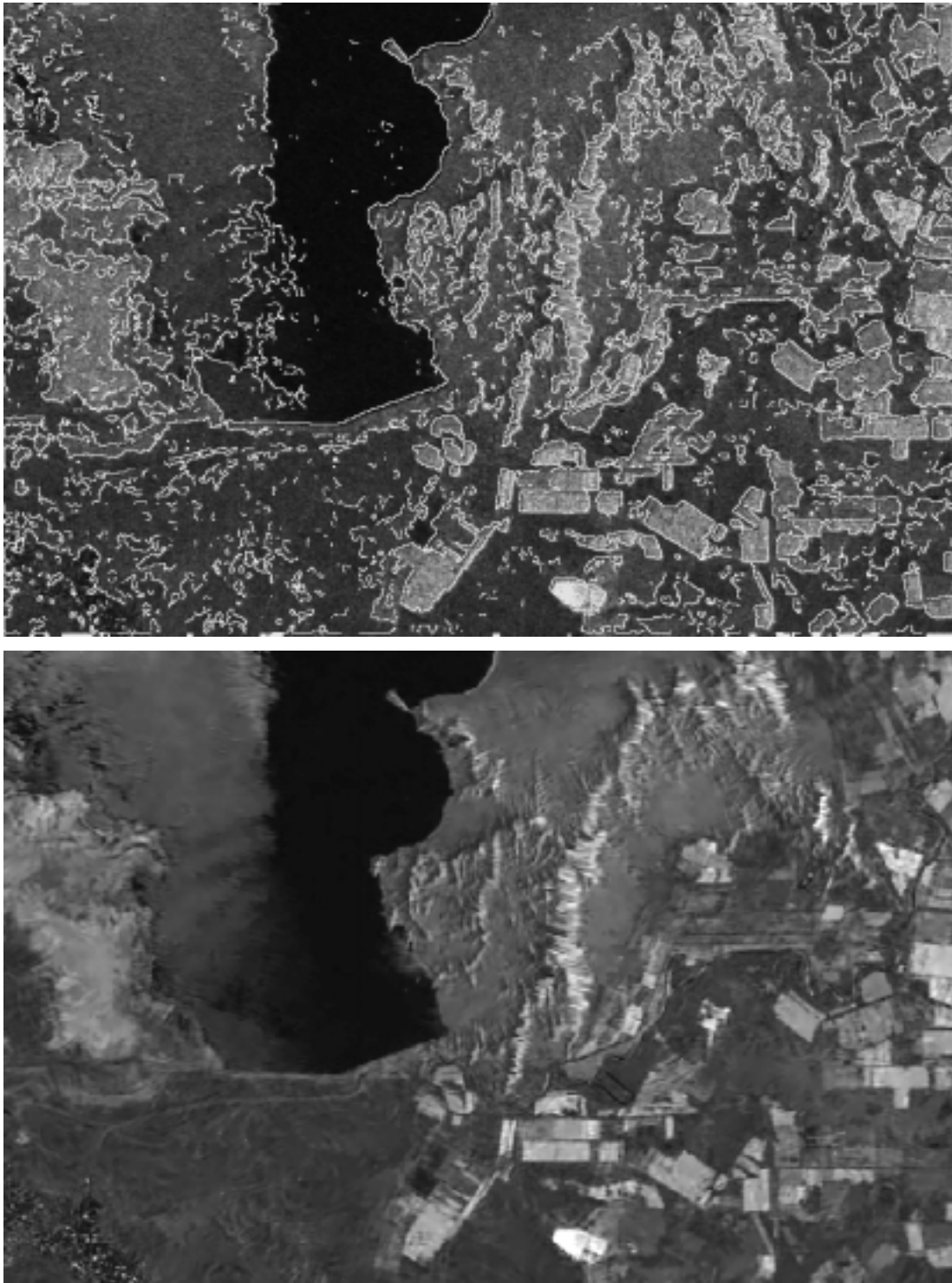


Figure 6.11: Model-Based Despeckling (MBD) with structural information. Top: Geocoded ERS SAR image (768×512 pixels, $L \approx 4$) with overlaid edge information. Bottom: MBD filtered ERS SAR image taking into account edge and target information.

- Isolated targets are blurred by conventional Gauss-Markov texture despeckling and hinder a correct parameter estimation. Hence, a special processing of strong scatterers is required.
- As a pre-processing step, isolated targets are detected by a window-based statistical analysis relying on the probability of false alarm. Found targets are removed to enable a proper parameter estimation.
- After Gauss-Markov despeckling, previously removed targets are re-inserted into the image. Targets that have not been detected in the pre-processing and that are consequently blurred are restored by a statistical analysis of the ratio image between observed and filtered data.

7 Illustration and Study of Results

In the preceding chapters we have developed the idea of the Model-Based Despeckling algorithm for filtering and information extraction from SAR data. Unlike most commonly used approaches, the introduced technique relies on the Bayesian formalism in combination with a complex data model, whose parameters are estimated from the noisy observation. Additional information is extracted to cope with several shortcomings of the chosen prior model and to adapt it to SAR specific properties, like e.g. isolated targets. This high degree of modeling causes a much higher computational load than required by other approaches, but promises to yield results of superior quality.

In this chapter, we will qualitatively and quantitatively evaluate the proposed algorithm as outlined in chapter six and compare it to other filters. This evaluation is not easy to perform, since no convincing objective quality measures for filtered data exist. This applies even more if the noise-free data or a ground truth is not available, as it is the case for real SAR data. As of this, we use synthetic data to perform an objective analysis by measuring different image quality indicators for different filters in the first section of this chapter. In the second section, we will test the developed approach on real SAR data and compare it to one of the best existing speckle filters. A quality analysis will mostly rely on visual inspection of the obtained results since we renounce to provide subjective heuristical measures. An example of the processing of a full scene is provided in section three, where we will also illustrate the additionally extracted image information in terms of edges and texture parameters. The chapter ends with a short summary of the presented results.

7.1 Quantitative Filter Evaluation based on Synthetic Data

To objectively evaluate the filtering quality of different despeckling approaches, we use synthetic data since the original noise-free image is often required for the computation of quality measures like the mean square error. Several prominent despeckling techniques are tested on four synthetically speckled images with different image contents. Since a single quality measure cannot reflect the whole truth we employ four different measures to describe the quality of the filtered images. After a short presentation of these measures, we provide the filtering results, show images of magnified details and comment on identified advantages and drawbacks of the filters under study.

7.1.1 Quality Measures

A single number can never describe the full quality of an image. Moreover, quality measures are often rather related to a particular application for which an image is to be used. However, we did not intend to generate filtered images for a specific application, but tried to obtain an optimal estimate of the noise-free scene. Thus, we use rather general, application independent quality indicators to evaluate despeckled images. We note that apart from numerical values the visual impression of an image still is one of the best, but necessarily subjective, indicators of image quality.

Mean square error

The mean square error (MSE) between original noise-free data and the filtering result is one of the most popular quality measures because it is easy to compute. However, this measure, which describes the energy of the remaining image noise, only partly reflects the visual quality of the filtered image, since it does not consider structural or spatial information like the preservation of edges. Nevertheless, the MSE is a standard measure and constitutes the basis of our filter evaluation.

Image mean

The preservation of the image mean is of particular interest for the subsequent interpretation of filtered SAR data. Not much related to visual image quality, an unbiased estimate of the mean intensity is crucial for a physical interpretation of the sensed images. In our examples we compare the mean of the original amplitude image to the one of the filtering result.

Maximum image smoothness

The smoothness of a filtered image gives information about the noise suppression capabilities of the employed filter. For window-based approaches, the maximum achievable smoothness is usually a function of the estimation window size. We measure image smoothness in terms of the equivalent number of looks of the despeckled image. This value is computed in the most homogeneous image region with a size of 35×35 pixels. Structural information in the estimation area falsifies and reduces the value of this measure. This especially occurs for images of high variation which do not contain large enough homogeneous regions.

Speckle ratio image

If the noise-free image is unknown, the only source of information about the filtering quality lies in the properties of the noise which was filtered out. For speckle, this noise image n is obtained by the ratio of the observation y to the filtered image \hat{x} , i.e. $n = y/\hat{x}$. The extracted noise n should exactly possess the statistical properties of the speckle process of the observed image. As a measure, we indicate the equivalent number of looks L of n

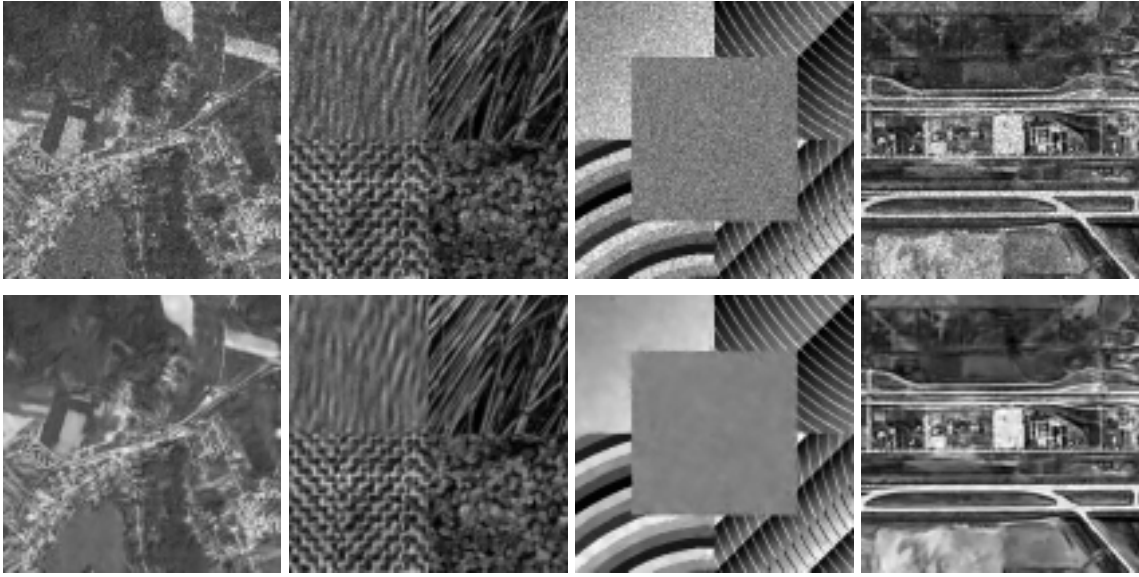


Figure 7.1: Artificially speckled test images (256×256 pixels, $L = 4$) and MBD filtered results. Top row: "Wessling" image, "Brodatz" image, "synthetic" image and "runway" image. Bottom row: Images filtered with the full MBD approach including structural enhancements.

determined over the whole image. However, this measure does not include any structural image degradations, which are also visible in the ratio image n . For that purpose, the ratio image by itself is often used to visually describe the quality of filtered images. In the ideal case, no spatial structures should be perceivable. The ratio images displayed below are scaled between values of 0.5 and 1.5 for better visualization [Oliver98].

7.1.2 Filtering of Artificially Speckled Test Images

As test data, we have chosen four images from different sources. To cover a wide range of image contents with varying textures, sharp edges and small detailed structures, we use two optical remote sensing images, a composition of four Brodatz textures and a synthetic data set with regular geometrical image contents. This selection of test data was artificially degraded with uncorrelated four-look speckle noise. The noisy and the corresponding MBD-filtered data are illustrated in Fig. 7.1.

For filter comparison, we filtered the speckled data ($L = 4$) with the developed Model-Based Despeckling (MBD), the GGMAP filter with structural enhancements, the Kuan filter, the EPOS filter and with a homomorphic wavelet-shrinkage method similar to [Donoho]. The computed image quality measures for the despeckled data are enumerated in Tab. 7.1. For all images identical filtering parameters have been used: The Kuan, the GGMAP and the EPOS filter have been applied with estimation windows of 7×7 pixels. For the MBD

Wessling	MBD	GGMAP	KUAN	EPOS	WVT
MSE	245	403	293	408	360
Mean $\mu_x = 99$	99	96	97	95	93
ENL	274	175	93	54	20
Speckle $L = 4$	5.3	3.5	6.4	4.5	6.0
Brodatz	MBD	GGMAP	KUAN	EPOS	WVT
MSE	152	374	277	386	367
Mean $\mu_x = 81$	80	79	78	79	73
ENL	47	62	27	33	19
Speckle $L = 4$	5.5	4.7	9.0	4.1	3.5
Synthetic	MBD	GGMAP	KUAN	EPOS	WVT
MSE	184	283	329	310	607
Mean $\mu_x = 110$	110	106	106	105	100
ENL	1327	292	220	74	111
Speckle $L = 4$	5.0	4.9	6.8	4.7	2.9
Runway	MBD	GGMAP	KUAN	EPOS	WVT
MSE	183	369	376	483	493
Mean $\mu_x = 101$	100	98	97	97	92
ENL	56	46	41	16	25
Speckle $L = 4$	5.5	5.2	7.3	7.0	3.7

Table 7.1: Quantitative filter evaluation for the four synthetic test images. The mean square error to the original image, the image mean, the ENL of the smoothest local image region and of the speckle ratio image are given for all tested filters.

filter, we have chosen a fifth order neighborhood system, an estimation window of 21×21 pixels and a parameter validity window of 7×7 pixels.

Small detailed cutouts of 96×96 pixels are shown in Figs. 7.2, 7.3, 7.4 and 7.5 for all four test images including the noise-free, the speckled, the MBD filtered and the GGMAP filtered data, as well as their corresponding ratio images. The GGMAP filtered images have been chosen to be displayed in combination with the MBD filtered data, since this filter possesses the highest degree of modeling of all standard filters and provides the visually best results, although the mean square errors are slightly higher than those of the Kuan filtered images. A general visual impression of the despeckling performance of the other tested filters is given in Fig. 4.1 of chapter four.

”Wessling” image

The ”Wessling” image exhibits many small structures in form of roads, but also agricultural areas with sharp edges. As can be seen in Fig. 7.2, these edges are well preserved by the MBD filter, while the center of the village looks a little bit blurred. On the contrary, the GGMAP filtered image looks rather sharp. However, this is simply because the original

noisy gray-values are not filtered throughout the whole image, as can be observed in the corresponding ratio image. The ratio image of the MBD filtered data, which give the best mean square error, shows no significant structures.

”Brodatz” image

Since the MBD filter was designed to restore textured images, we expect it to outperform all other filters on this test image. This expectation is confirmed both by the measured quality indicators and by the detailed images as they are displayed in Fig. 7.3. The MSE is almost 50% lower than the second best result obtained by the Kuan filter. The visual correspondence between original and despeckled data is almost perfect, which is also confirmed by the ratio image. The shortcomings of the Gamma model with edge-detection, as used in the GGMAP filter, are clearly visible both in the filtering result and the ratio image, where textural structures can be seen: The GGMAP filter is unable to model complex textural information.

”Synthetic” image

This test image was generated to examine the filters’ capabilities to cope with non-stationary image contents and to verify their noise reduction performance in homogeneous areas. In terms of the mean square error the MBD filter dominates all other filters by far. The same applies for the image smoothness measure, as indicated in Tab. 7.1. However, at the border between the homogeneous area and the strongly textured region with lines (cf. Fig. 7.4) the filtering performance decreases. Since the mean gray-value of these two areas is in a similar range this non-stationarity is not detected by the region-growing edge-detection step of the MBD filter. While the GGMAP filter leaves the gray-values of the lines unchanged and thus preserves some noise, these structures are well filtered by the MBD approach. However, the MBD ratio image indicates a reduced filtering for very dark image regions with a mean gray-value close to one.

”Runway” image

The ”runway” image impressively demonstrates the filtering quality of the MBD filter. Visually, original and filtered data are almost identical, as illustrated in Fig. 7.5. The mean square error is by far lower than for all other filters. No significant structures can be detected in the ratio image. Apparently, this image fits rather well to the Gauss-Markov model, explaining the outstanding performance. The GGMAP filter performs less convincing. As reflected in the ratio image, the despeckled data still exhibit a lot of noise. The visual similarity of the filtered data to the noisy image is much higher than the similarity to the original data.

7.1.3 Conclusions

In order to quantitatively determine the filtering quality of the proposed MBD filter and other filters, we have chosen artificially speckled images to test various important properties

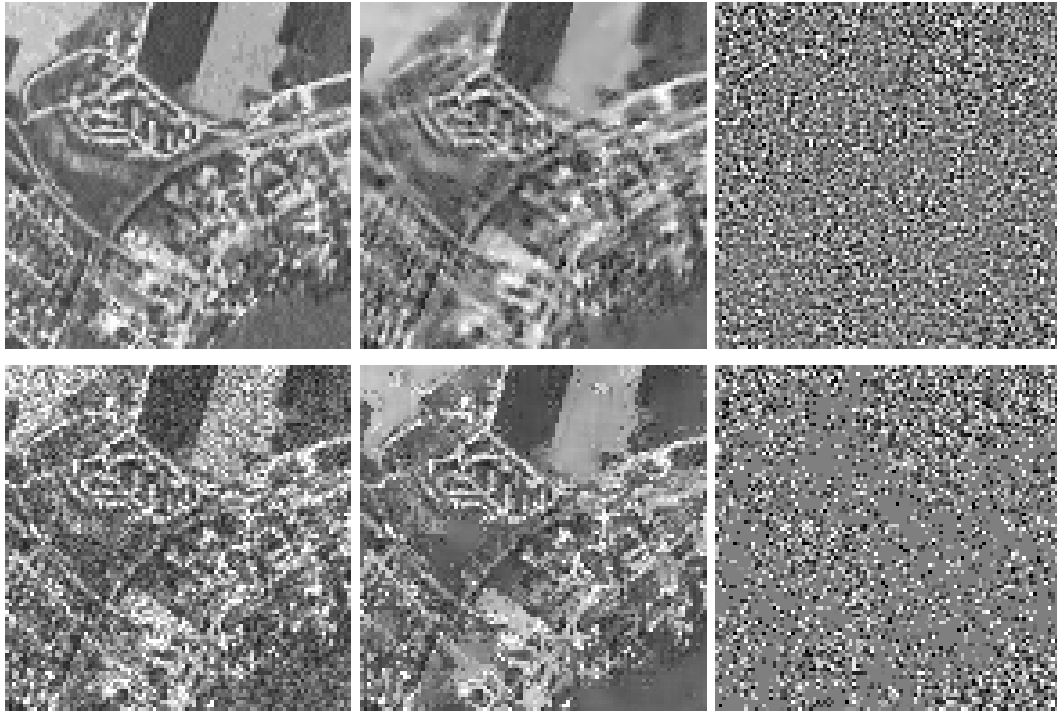


Figure 7.2: Detailed cutout of the "Wessling" test image. Top row: Original data (96×96 pixels), MBD filtered data, MBD ratio image. Bottom row: Speckled data, GGMAP filtered data, GGMAP ratio image.

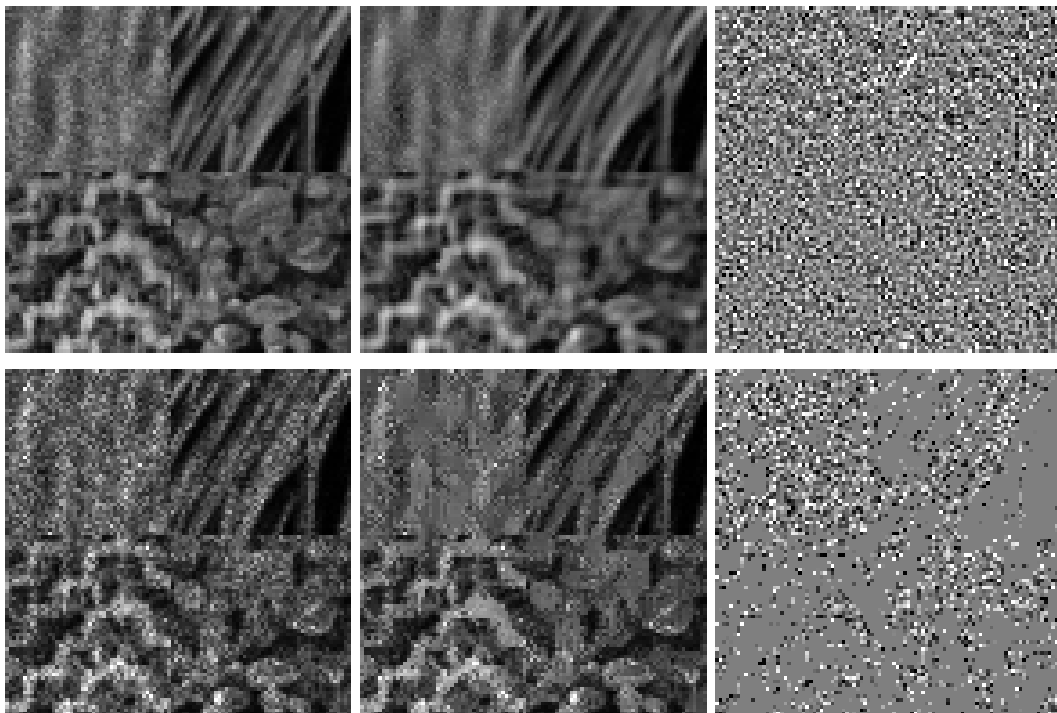


Figure 7.3: Detailed cutout of the "Brodatz" test image. Top row: Original data (96×96 pixels), MBD filtered data, MBD ratio image. Bottom row: Speckled data, GGMAP filtered data, GGMAP ratio image.

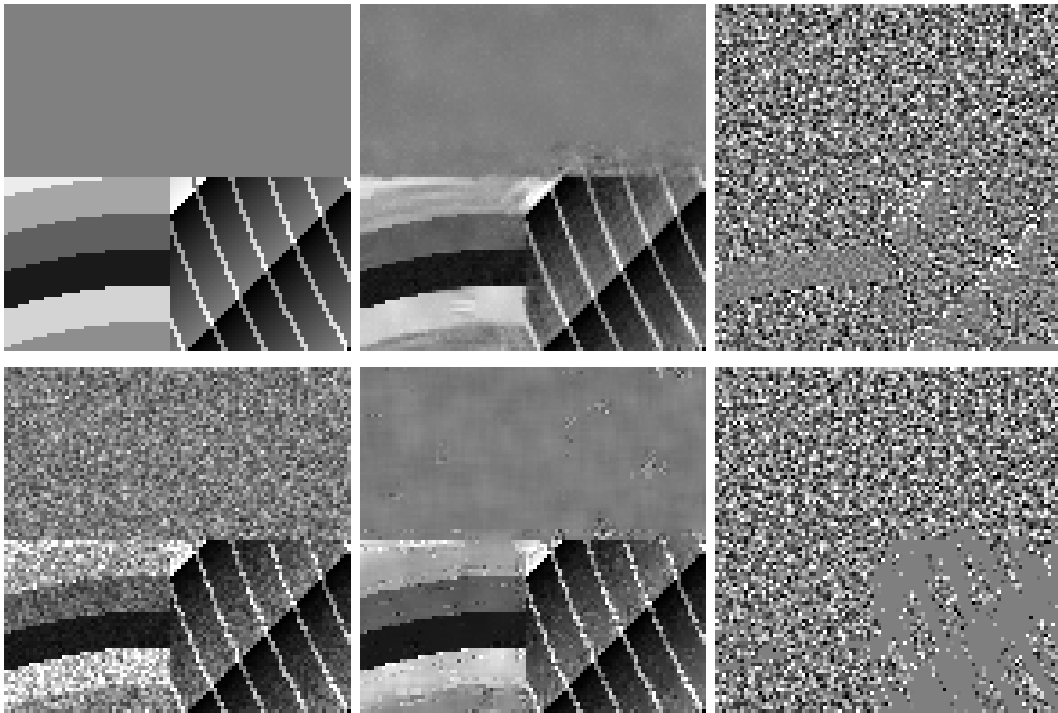


Figure 7.4: Detailed cutout of the "synthetic" test image. Top row: Original data (96×96 pixels), MBD filtered data, MBD ratio image. Bottom row: Speckled data, GGMAP filtered data, GGMAP ratio image.

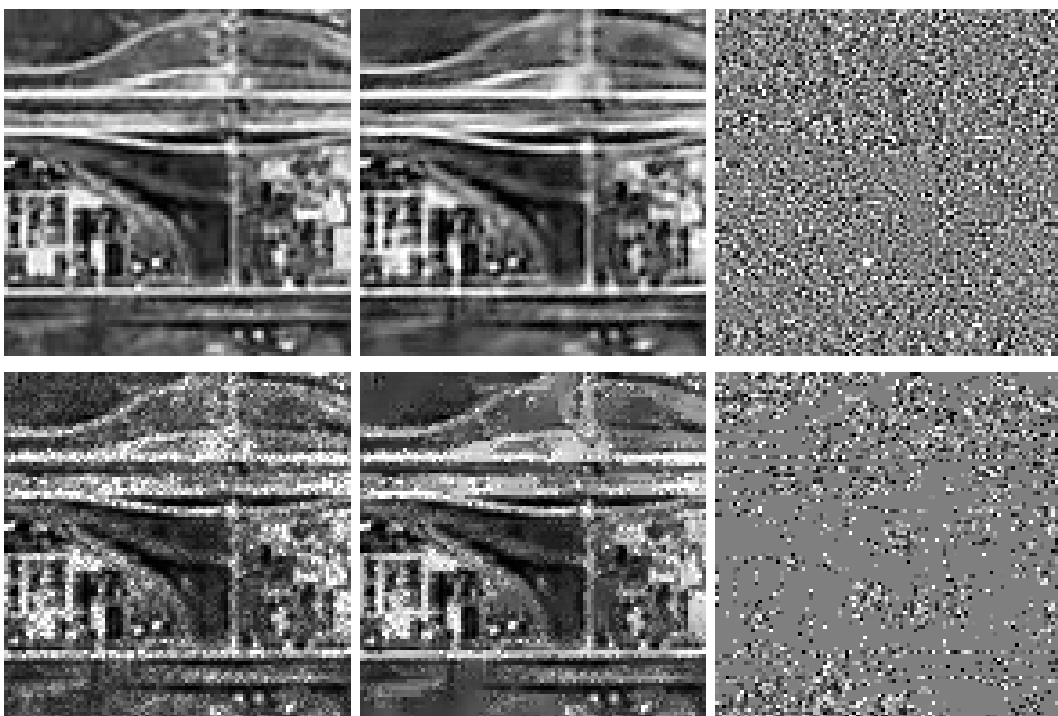


Figure 7.5: Detailed cutout of the "runway" test image. Top row: Original data (96×96 pixels), MBD filtered data, MBD ratio image. Bottom row: Speckled data, GGMAP filtered data, GGMAP ratio image.

of despeckling algorithms. The proposed MBD approach showed convincing results both for objective quality measures and subjective visual image inspection.

Especially as far as the attainable mean square error is concerned, the new filter sometimes dramatically outperforms existing methods. Equivalently good results can be found for the achievable image smoothing which does not depend on the size of the estimation window. Due to an included correction factor for the introduced multiplicative bias (cf. chapter five) the image mean is also satisfactorily preserved. We note however from the ENL of the ratio image, that the variance of the extracted speckle is too low compared to its true value ($L = 4$). Nonetheless, in comparison with the other investigated filters, the extracted value for the ratio image remains rather stable (slightly above five), as denoted in Tab. 7.1. This indicates some kind of robustness towards the introduction of artefacts, which does not apply for the other filters whose strongly varying measurements may be caused by artefacts as a function of the image contents.

Another important observation is that other filters preserve a certain degree of residual noise in the image or give too many false alarms concerning the detection of edges or other small features. This effect, which much less often occurs for MBD filtered data, may hinder a subsequent interpretation of the despeckled images. Nonetheless, this issue is strongly application dependent. On the other hand, the almost complete noise-reduction of the MBD filter is reached at the expense of a slight blurring of the filtered data. Visually, the other tested filters apparently provide much sharper images simply by less strongly filtering and staying closer to the observation. It was proven by the mean square error and a visual evaluation of the given image examples that an altogether much better estimate of the radar cross-section is provided by the MBD filter.

7.2 Qualitative Filter Evaluation based on SAR Data

We have provided a first filter evaluation for synthetic data. In this section several results obtained from real SAR data will be presented. Unlike in the last section, we do not give any quantitative measures concerning the image quality but simply show the despeckled data and their corresponding ratio images. For an evaluation of the filtering performance we further rely on a visual inspection of the filtered and the ratio images. We renounce to give the values of a developed quality measure based on the power spectra of the ratio images and of other measures, since they may be judged subjective and do not give further insight to what is already visually perceivable in the ratio images.

7.2.1 Filtering of Synthetic Aperture Radar Images

By analogy to the last section, we have selected two rather different types of SAR data in order to be able to better evaluate the quality of the MBD filter. This time, we make comparisons only to the GGMAP filter. First, the full images are displayed, then we again give small cutouts with higher detail. In addition, we now also show the cutout of the corresponding mean filtered data processed with a window size leading to the same noise

reduction in homogeneous areas as the MBD filter. For the given examples, this yielded an averaging kernel of about 11×11 pixels, i.e. a multi-looking of more than 100 times.

X-SAR image

In Fig. 7.6 the despeckling result of an X-SAR image is illustrated. This image contains mainly agricultural areas with sharp borders between different regions, but also thin linear structures. Hence, both a successful edge-detection and a reliable texture parameter estimation are required to guarantee good results. The MBD filtered data are of high visual quality. Homogeneous areas are highly smoothed while edges and small structures are well preserved.

This impression is confirmed in the detail images presented in Figs. 7.7 and 7.8. While a slight blurring is visible in the MBD filtered result shown in Fig. 7.7, a good restoration of the line structures in the upper right part of the image of Fig. 7.8 is observed. The GGMAP filter shows the same behavior as before, i.e. preservation of noise at non-stationarities and inability to restore textural information. Nonetheless, the qualitative difference of the ratio images is not as pronounced as for the presented synthetic examples. The corresponding mean filtered data show the loss of resolution of multi-look images that is necessary to obtain a noise reduction which is equivalent to the MBD filtered image.

FGAN SAR image

The FGAN (Forschungsgesellschaft für Angewandte Naturwissenschaften e.V.) image is a high-resolution airborne SAR image, which was kindly provided by FGAN for experiments. This image exhibits many isolated small structures like buildings, as well as textured forest areas and agricultural regions. This makes this image especially interesting for an evaluation of the texture preserving capabilities of the developed filter. The original and the filtered image are provided in Fig. 7.9. In the despeckled image a clear textural separation between forest and field areas is observed. The build-up area in the upper right part of the image is also more than satisfactorily restored.

The detail images in Figs. 7.10 and 7.11 allow a better visual inspection: Different textures for forest and flat areas can be easily identified in the MBD filtered data, whereas the GGMAP filtered data seem to be composed of small homogeneous patches. In our opinion the MBD despeckled data give a much more natural and realistic visual impression. However, a slight blurring effect is observed. The mean filtered images with identical noise suppression contain no remaining detail.

SRTM SAR image

A last short illustrative example of SAR image filtering, which shows the power of the employed Gauss-Markov texture model is given in Figs. 7.12-7.15. The images show an area over White Sands, New Mexico, USA, the first dataset acquired during the SRTM

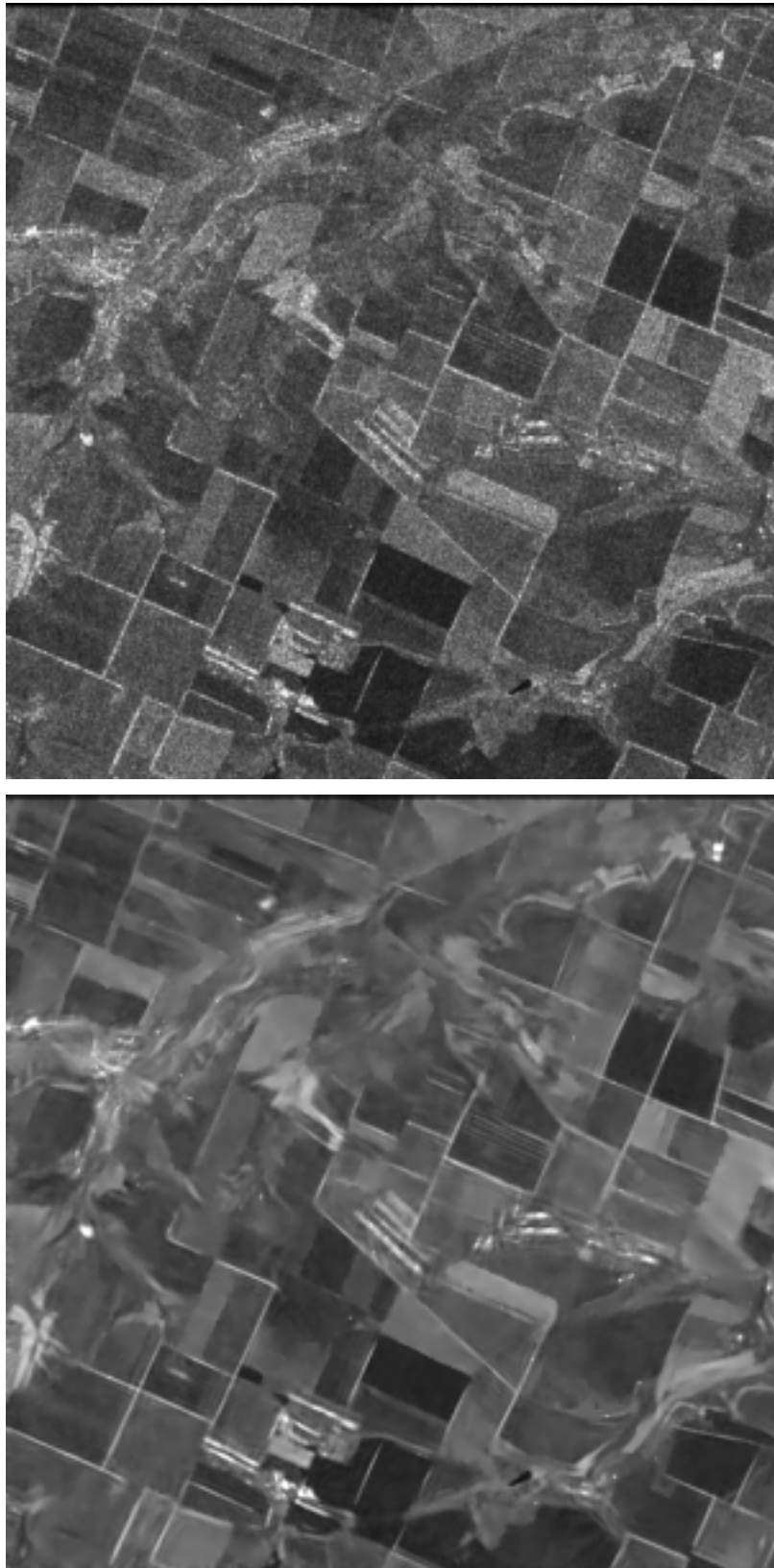


Figure 7.6: X-SAR image over Ukraine. Top: Original X-SAR image (512×512 pixels, $L \approx 3$). Bottom: MBD filtered X-SAR image.

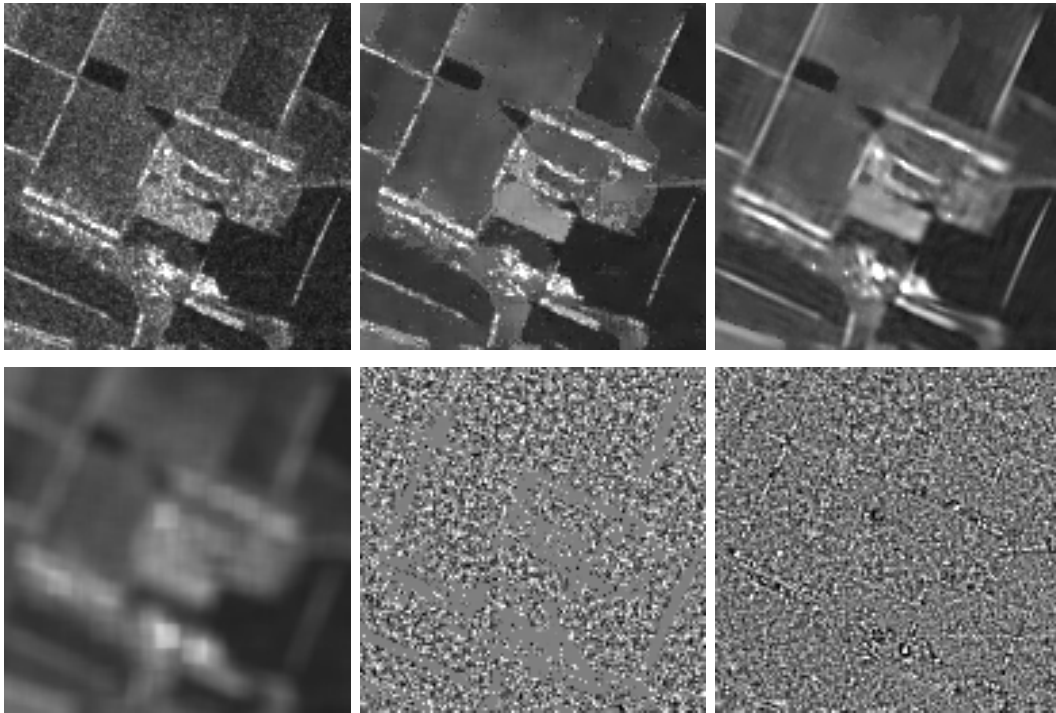


Figure 7.7: First detailed cutout of the X-SAR image. Top row: Original X-SAR image (128×128 pixels), GGMAP filtered image, MBD filtered image. Bottom row: Mean filtered image, GGMAP ratio image, MBD ratio image.

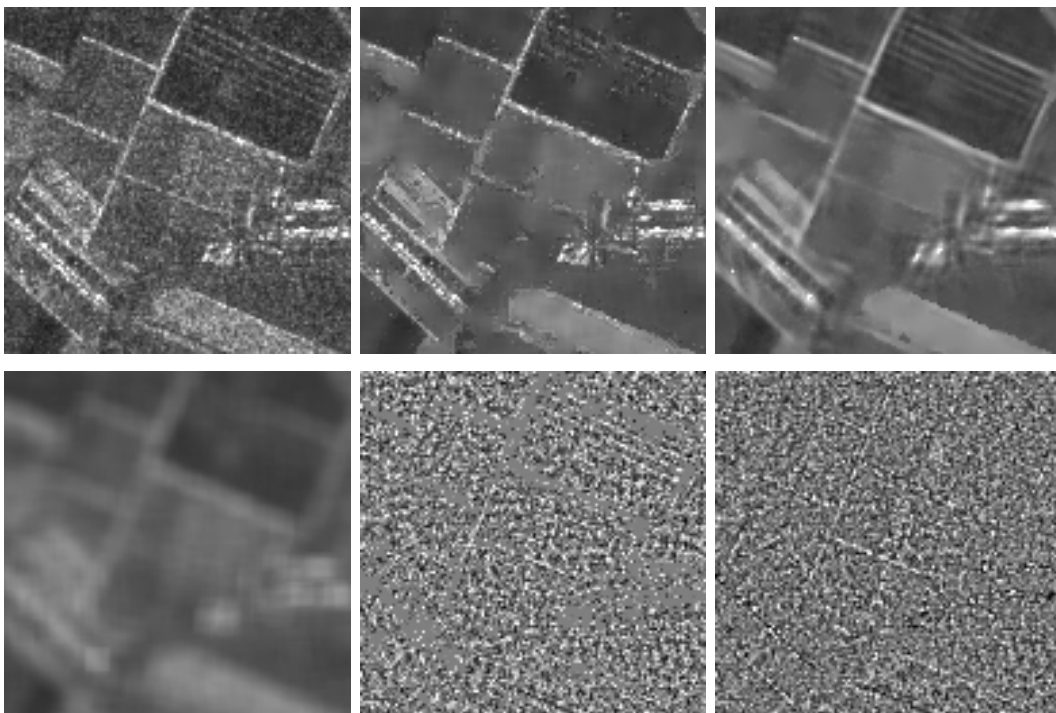


Figure 7.8: Second detailed cutout of the X-SAR image. Top row: Original X-SAR image (128×128 pixels), GGMAP filtered image, MBD filtered image. Bottom row: Mean filtered image, GGMAP ratio image, MBD ratio image.



Figure 7.9: FGAN airborne SAR image over Germany. Top: Original FGAN SAR image (512×512 pixels, $L \approx 8$). Bottom: MBD filtered FGAN SAR image.

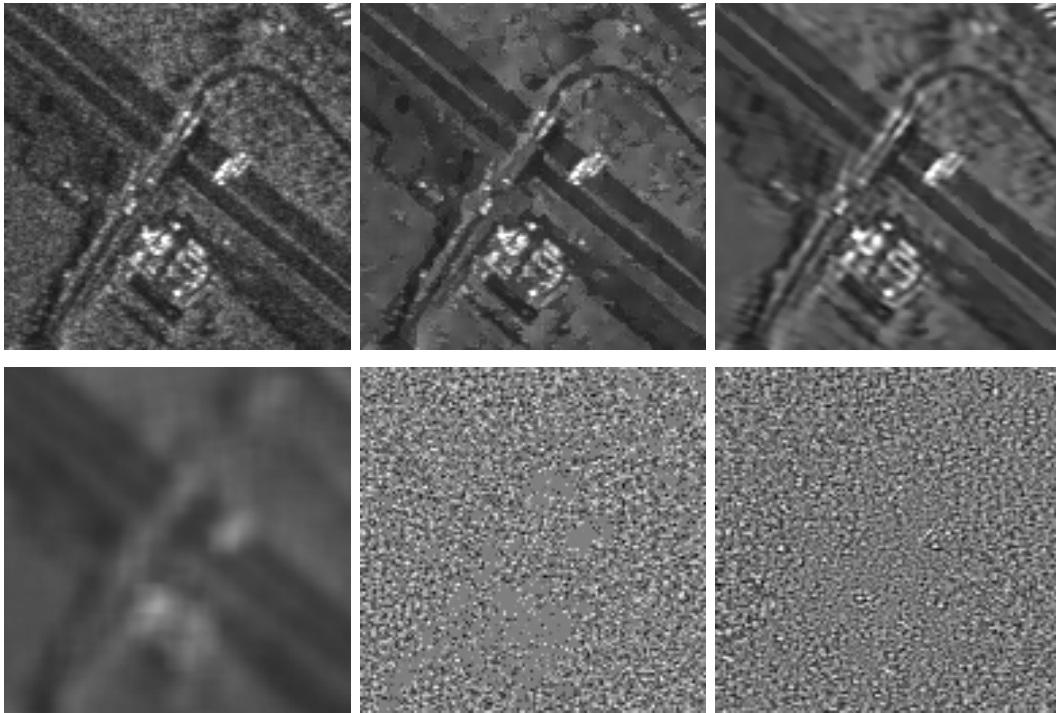


Figure 7.10: First detailed cutout of the FGAN SAR image. Top row: Original FGAN image (128×128 pixels), GGMAP filtered image, MBD filtered image. Bottom row: Mean filtered image, GGMAP ratio image, MBD ratio image.

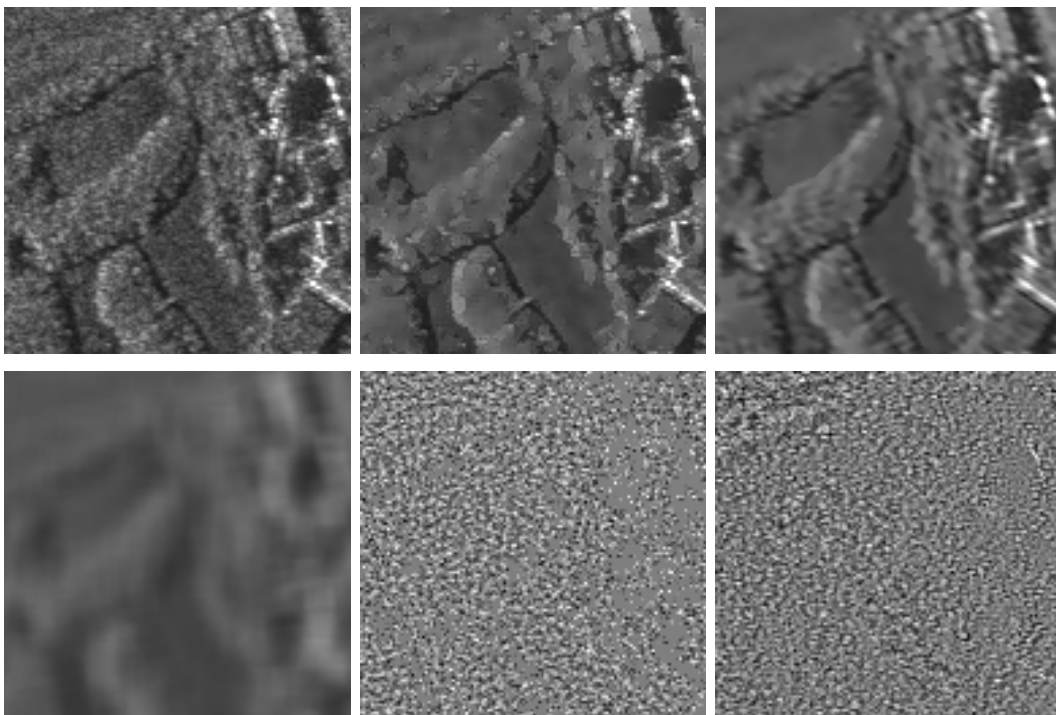


Figure 7.11: Second detailed cutout of the FGAN SAR image. Top row: Original FGAN image (128×128 pixels), GGMAP filtered image, MBD filtered image. Bottom row: Mean filtered image, GGMAP ratio image, MBD ratio image.

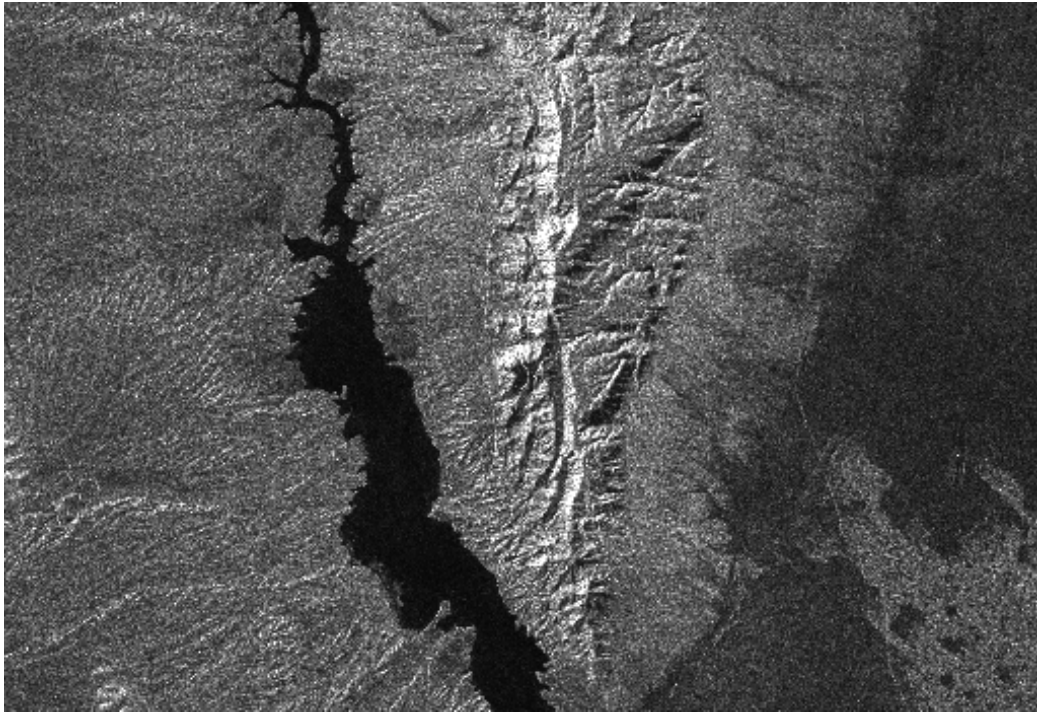


Figure 7.12: Original SRTM (Shuttle Radar Topography Mission) image. The cutout (1024×684 pixels, $L \approx 3$) of the geocoded scene over White Sands, New Mexico, USA, shows interesting drainage patterns and mountainous regions.

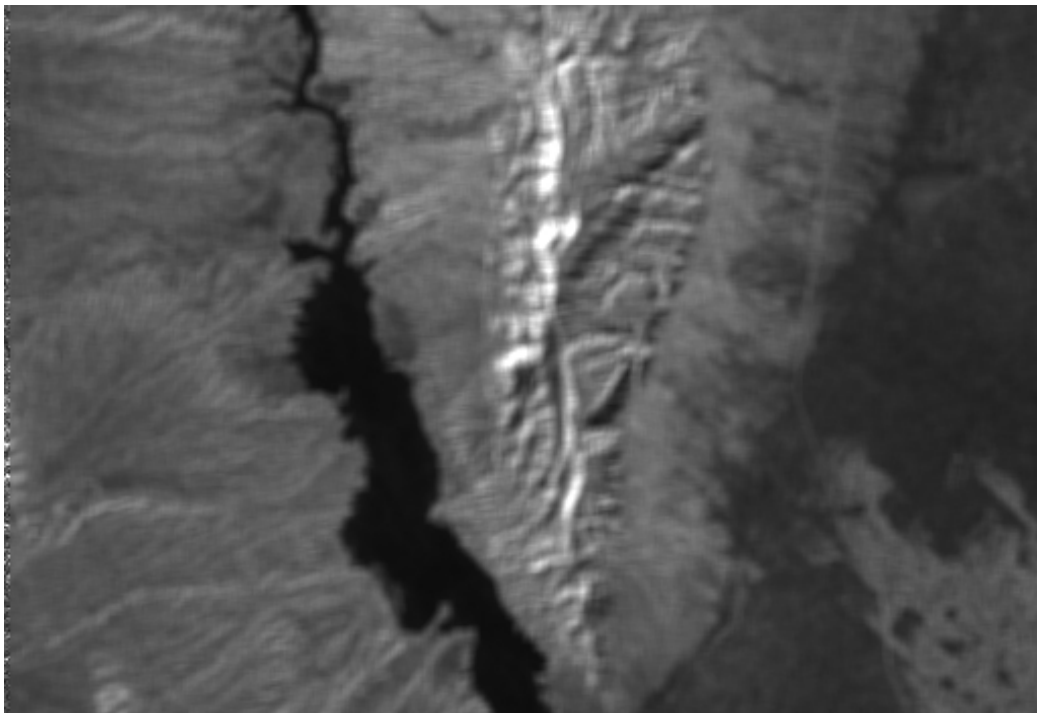


Figure 7.13: Mean filtered SRTM image. The image illustrates the noise smoothing and detail preserving capabilities of the MBD filter compared to multi-looking.

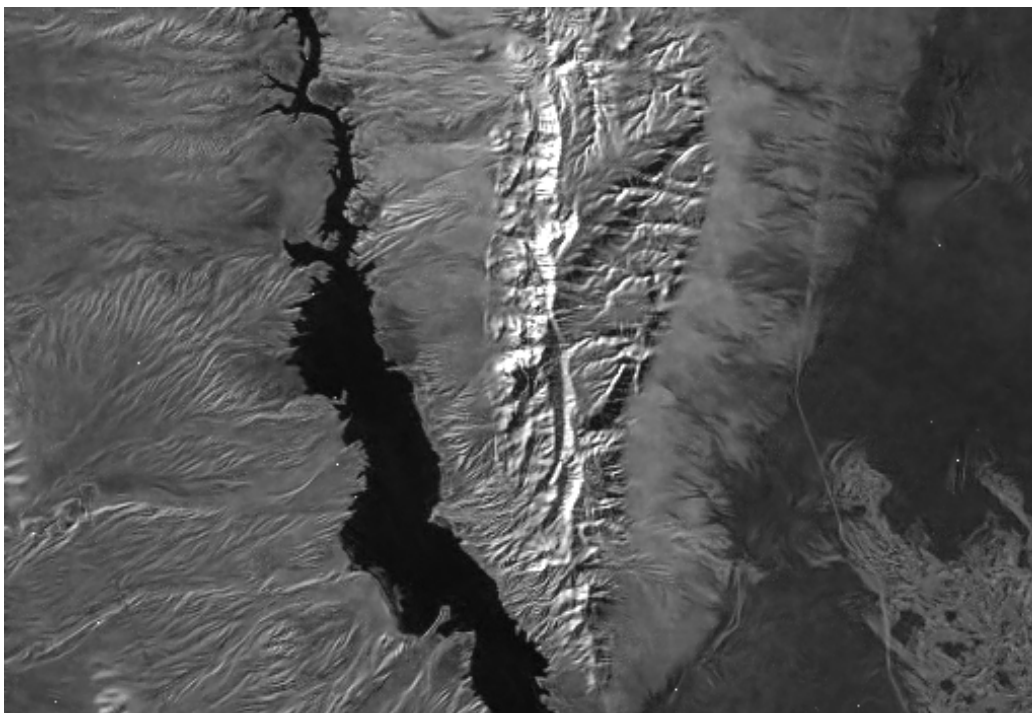


Figure 7.14: MBD filtered SRTM image. Textural information of the original SAR image is preserved and visually enhanced.

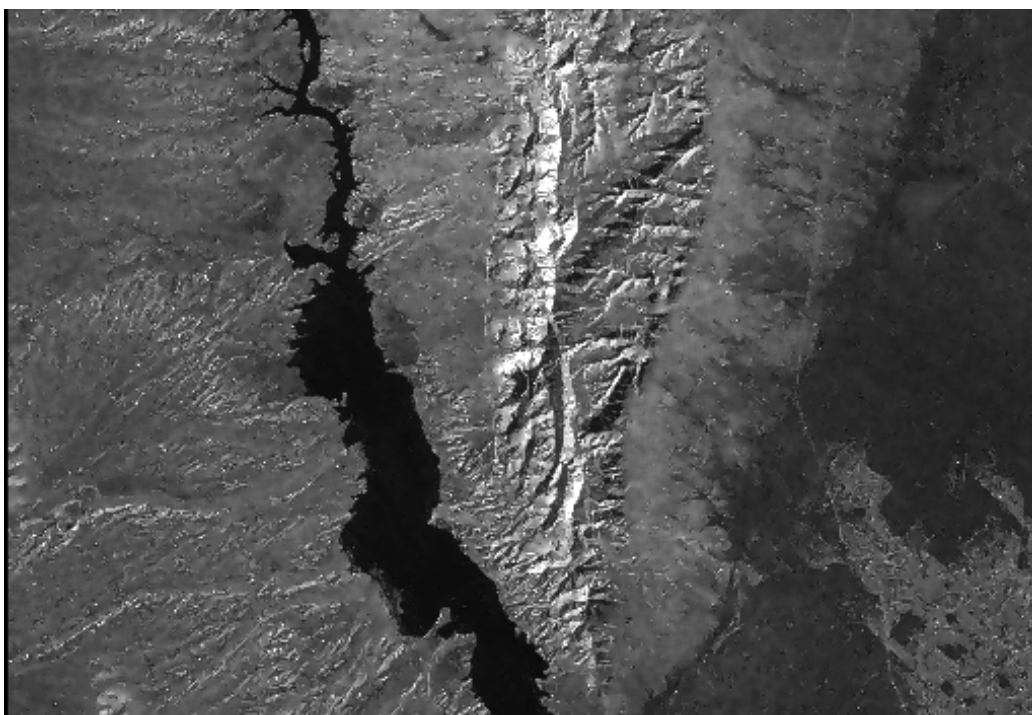


Figure 7.15: GGMAP filtered SRTM image. The over-detection of features and the noise-preservation in areas of higher textural activity are clearly visible.

mission. For comparison to the MBD filter, we provide again the GGMAP filtered and the mean filtered images.

The MBD despeckled image might be easily mistaken for optical data: The speckle noise has been completely filtered out and the drainage patterns in the mountainous areas are impressively preserved and even enhanced compared to the original data, i.e. significant information is much easier visible in the filtered image than in the original observation.

7.2.2 Conclusions

A quantitative and qualitative analysis of the filtering performance of the introduced MBD filter in comparison with standard techniques yielded favorable results for the model-based approach. While visual attractiveness is a rather subjective criterion, more objective quality measures, such as the mean square error, though not capable to fully describe the image quality, confirm the more accurate restoration of MBD filtered images.

The new approach accurately captures and restores textural information and is able to preserve edges and isolated targets. As main drawbacks, a slight blurring of the filtered image is observed, several stationarity problems between abruptly changing textures in synthetic images occurred and the required computational load is excessively high compared to standard approaches.

We especially consider this last point as the main disadvantage of the proposed method. Computation times higher by roughly two orders of magnitude may hinder a wide-spread use of our approach. However, these computational requirements are easily explained by the complexity of the filter, which requires more than the calculation of mean and variance.

In the given examples, we did not consider single look data for the following reason: Single look data show too high speckle correlation making the application of the proposed filter difficult. Of course, the correlation problem affects all other approaches as well, but less strongly since they do not require the estimation of spatial interactions to the same extent. Moreover, all despeckling filters perform rather poorly on single look data, due to the low signal to noise ratio. Hence, filtered single look data are usually not suited for a subsequent interpretation. Up to now, speckle filtering provides profitable results only for at least slightly multi-looked SAR image products.

7.3 Full Scene Processing for Information Extraction

In this section, an example of SAR image processing with more emphasis on additionally extracted information, i.e. texture parameters and edge-information, is presented. The power of the developed approach does not only lie in a more accurate estimation of the radar cross-section, but also in the extraction of information about the image contents. Common approaches designed to extract textural information are either based on analytically analyzable models [Oliver94, Oliver96] or use complex spatial texture models [Schistad92, Schistad97, Walessa98]. However, the latter usually do not allow to take into account the properties of speckle. We have demonstrated in chapter four that MPL Gauss-

Markov model parameter estimation from speckled data yields only limited information about the image.

7.3.1 Information Extraction from SAR Data

In Figs. 7.16-7.21 an example of the information extracted from SAR data by the MBD filter is given. The originally processed scene is depicted in Fig. 7.16 and was kindly provided by Aerosensing GmbH.

Edge-information

Fig. 7.17 presents the extracted edge information. Two different kinds of edges according to the distinction made in chapter six can be discriminated. Edges between homogenous regions are shown in black, edges in textured areas, which occur mainly in the forest areas, in gray. The edge image gives a rough impression of the image contents allowing to identify dominant changes in mean backscatter and areas of high textural activity. Note that the edge-detection step was especially designed to improve the cross-section estimation and not as a stand-alone edge-detection algorithm.

Backscatter information

The despeckled radar cross-section is depicted in Figs. 7.18 and 7.19 estimated by the MBD and the GGMAP filter, respectively. The MBD filtered data do not only look visually more attractive, but are much smoother and allow a better textural distinction between homogeneous and forest areas. The GGMAP filtered image exhibits a similar roughness for both textured and untextured regions.

Textural information

The norm $|\boldsymbol{\theta}|$ of the extracted Gauss-Markov model parameters is shown in Fig. 7.20. In this image, the clear distinction between forest and areas of uniform backscatter, which cannot be discriminated by intensity information alone, becomes visible. Note that due to the projection to one dimension the norm contains only reduced textural information with regard to the full parameter vector.

The result of an unsupervised k-means clustering of the norm $|\boldsymbol{\theta}|$ and the backscatter information from the filtered image \hat{x}_{MAP} into six classes is given in Fig. 7.21. Although this segmentation is derived from a single and highly noisy information channel, i.e. a single SAR image, a good general distinction between main image regions, such as build-up areas, forest and flat regions, is possible.

A new database system developed at DLR in collaboration with the ETH Zürich exploits this kind of textural information contained in the parameters of the Gauss-Markov or similar texture models in order to enable a query by image contents. The system is designed to

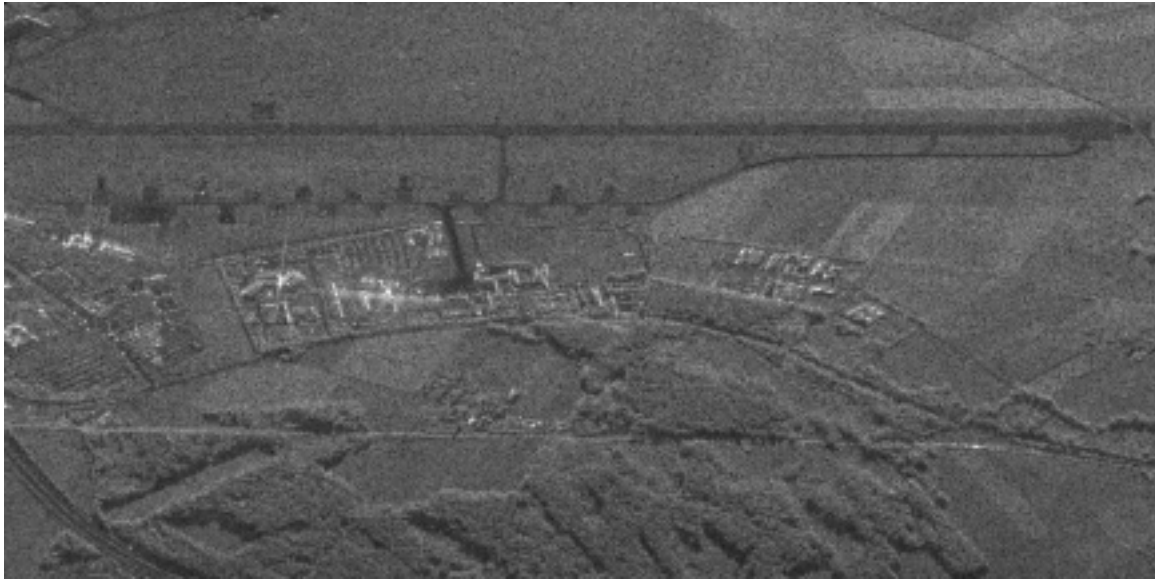


Figure 7.16: SAR image of DLR/Oberpfaffenhofen. A subsampled part (1800×900 pixels, $L \approx 3$) of the whole scene is shown. The full scene was kindly provided by Aerosensing GmbH.



Figure 7.17: Extracted edge information of the DLR/Oberpfaffenhofen SAR image. Edges between homogeneous regions are shown in black. Edges in textured areas are gray.



Figure 7.18: MBD filtered DLR/Oberpfaffenhofen SAR image. Textural information of the original SAR image is preserved and visually enhanced.



Figure 7.19: GGMAP filtered DLR/Oberpfaffenhofen SAR image. Targets and edges are well preserved. A clear textural distinction between fields and forest cannot be made.

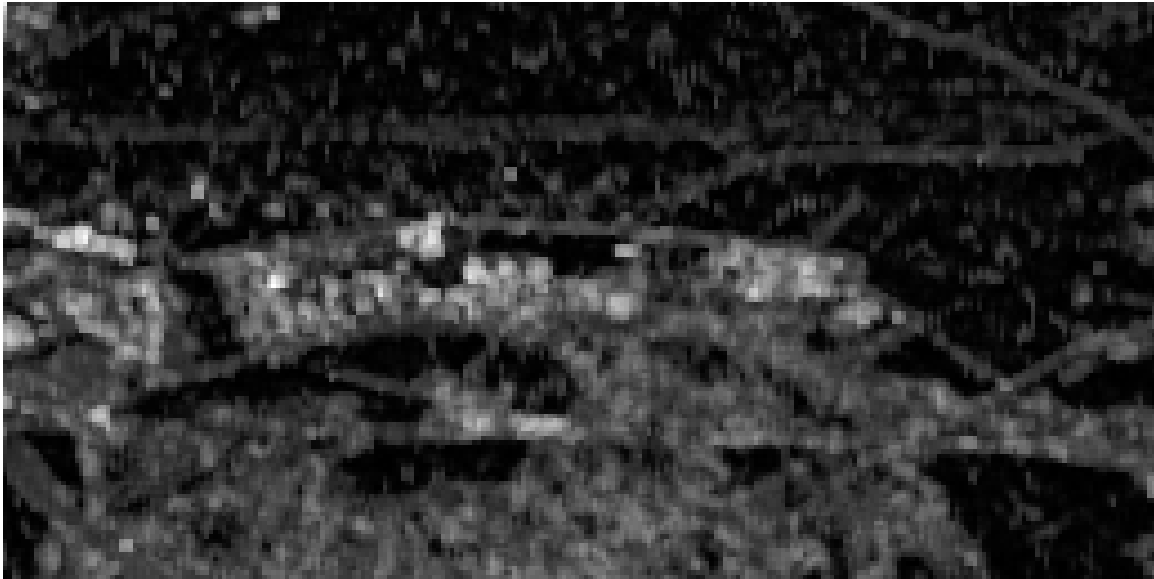


Figure 7.20: Texture information extracted from the DLR/Oberpfaffenhofen SAR image. The norm $|\boldsymbol{\theta}|$ of the directional texture parameters is displayed.

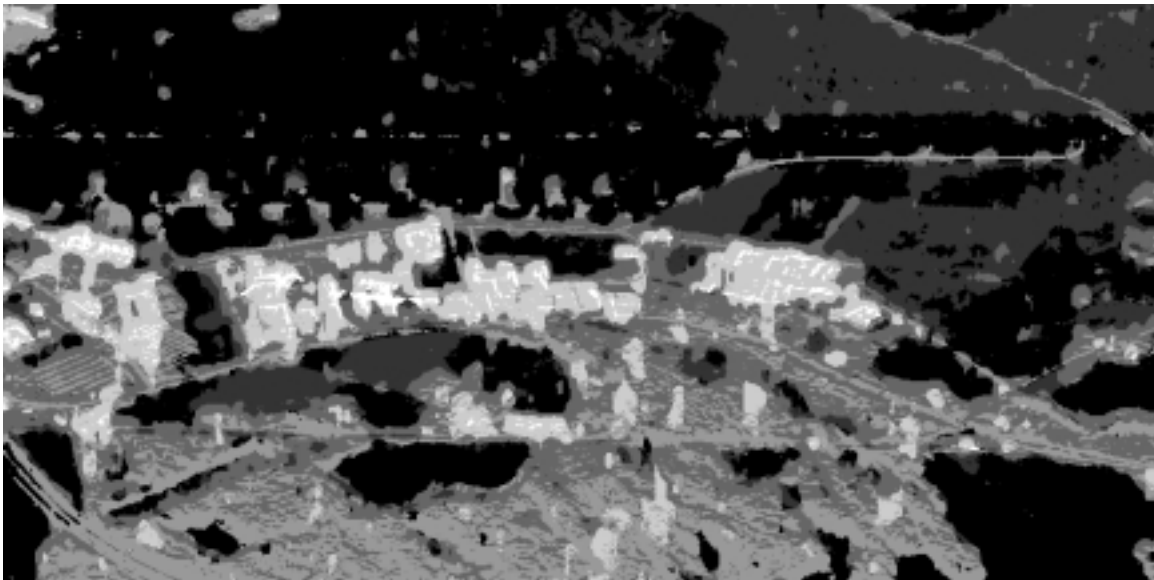


Figure 7.21: Unsupervised clustering of texture and backscatter information from the DLR/Oberpfaffenhofen SAR image. The norm $|\boldsymbol{\theta}|$ and the backscatter information from the MBD filtered image have been used for a clustering into six classes.

work for texture information extracted from complete, i.e. optical, and incomplete, i.e. radar, image data and can be experimented over the web [RSIA1, RSIA2]. Allowing the user to define his own search criteria, the query is not limited to pre-defined or pre-extracted labels.

7.3.2 Conclusions

In addition to an estimate of the radar cross-section, edge and texture information is extracted by the presented approach. Besides increasing the quality of SAR image despeckling, this information has value on its own and can be used for further image interpretation, such as image classification, or serve as a feature in database systems. Short examples have been provided.

However, the extraction of features alone is not sufficient if they are not assigned a meaning. A step towards this direction was made with the joint query by image content project of DLR and ETH Zürich [Datcu98, Rehrauer, Schroeder99b].

7.4 Summary

In this chapter, the following points have been discussed:

- A filter comparison has been presented, which was based on synthetic image examples to be able to compute various quality measures. Since image quality cannot be measured consistently by a few numbers, we also relied on a visual evaluation of the image quality.
- As objective measures, we have chosen the mean square error, the image mean, the maximum noise reduction and the first order statistical properties of the extracted speckle noise. Furthermore, speckle ratio images have been provided to visually evaluate introduced spatial distortions.
- Both the visual and the numerical analysis provided convincing results for the proposed Model-Based Despeckling approach. Its improved filtering quality is mainly due to the employment of more complex models in comparison with the other tested approaches. However, the increased complexity also results in comparatively rather high computation times.
- The MBD filtered images look visually very attractive, but exhibit a slight blurring effect, which is equivalent to a reduction of image resolution. While other filters often provide rather sharp images, this is mainly because their results stay much closer to the noisy observation and either preserve a certain degree of speckle or give many false alarms in terms of detected edges and small features.
- The extracted information in form of texture parameters, edges and point scatterers, which enables the high quality of the filtered images, can be used as a feature vector in remote sensing image databases or for further image interpretation.

8 Conclusions

This thesis presents a new approach to SAR image filtering and information extraction from SAR data. SAR images are affected by the speckle phenomenon, which hinders a straightforward interpretation of this particular kind of remote sensing images. We have developed the idea of a combined approach for feature extraction and employment of this information to generate filtered SAR data of high quality. Both the low-level feature extraction and the higher level image restoration are performed within the Bayesian framework to be able to incorporate a maximum of available information. The presented approach constitutes one of the first attempts to extract hidden model parameters of Gibbs random fields, i.e. complex textural information, from incomplete speckled data. Additional SAR specific features that cannot be described by the employed Gauss-Markov model are extracted by a non-Bayesian statistical data analysis.

Unlike other methods, we fully take into account the speckle noise properties for the estimation of relevant model parameters. Thus, the introduced algorithm requires no heuristically set parameter values for SAR image despeckling but only uses information already contained in the noisy observation. Due to a sophisticated modeling, more information about the noise-free scene can be extracted allowing a much better restoration than achievable with other currently available approaches. The employed model is not based on stationarity assumptions concerning the mean SAR backscatter, as the Gamma model, but encapsulates spatial interactions of pixels within small neighborhoods in form of model parameters of a Gibbs random field. The proposed algorithm requires a considerable amount of computing power for the low-level information extraction but, at the same time, does not only provide a despeckled SAR image but also a vector of additional features which characterize the actual image content. In principle, this additional information, which is extracted from under the speckle noise resulting in more reliable estimates, enables a further image interpretation. However, it must be clear that a single highly degraded information channel, i.e. a non-polarimetric SAR image, contains only a limited amount of usable information. Hence, the value of the extracted features for real applications must be further investigated, since many applications require multi-temporal or multi-sensor data in order to increase the available information content and often even need features completely different from those provided by our approach.

For these cases, the despeckled image is of special interest, since the sought information is always contained in the radar cross-section. Consequently, valuable information for a high variety of applications can be much more easily extracted from filtered images of high quality

than from the original noisy data. Although the direct interpretation of unfiltered SAR data promises more accurate and more reliable results, this also requires the development of highly specialized application-dependent algorithms. In most cases, this approach is too expensive and filtered data are preferred. In our opinion, the employment of filtered images can be justified in many applications, if the despeckled data do not contain significant artefacts and preserve the dominant characteristics of the underlying cross-section, like the mean backscatter value, basic textural properties and edge information. In this way, we think that the proposed Model-Based Despeckling approach constitutes an important step towards an easier interpretation of SAR data, since it provides a much better and more reliable estimate of the radar cross-section than many other available techniques.

8.1 Summary of the Thesis

The employed Bayesian approach requires prior knowledge about both the parameters to estimate and the noise process that causes the degradation of the observed data. Consequently, we started with a general outline of the SAR sensor in order to better understand its particular imaging properties. We have seen that SAR is a time-ranging system, which allows to generate images from backscattered intensity. In order to obtain an adequate resolution, signal processing techniques like pulse compression and synthetic aperture are used. We have furthermore mentioned that radiometrical and geometrical information in SAR images are influenced by the topography of the imaged scene requiring an appropriate correction of these effects. However, additionally to these distortions a phenomenon called speckle highly degrades the radiometric information of SAR images and hinders their interpretation. Speckle can usually be interpreted as a multiplicative granular noise which becomes visible only in detected SAR data. The spatial correlation of speckle is usually not considered in SAR image processing techniques and can be reduced by approaches like subsampling or multi-looking. We described the most important statistical properties of the speckle process, e.g. its likelihood function, which are necessary to deal with this highly disturbing effect within the Bayesian framework for SAR image filtering.

We introduced the basics of Bayesian estimation theory and illustrated that the incorporation of prior knowledge may help to improve the accuracy of parameter estimation, but also showed that the correct choice of a prior model is a crucial issue often rising critics about the Bayesian approach. To be able to perform a model selection and to determine the parameters of a given prior, we gave an interpretation of the Occam factor and briefly outlined a technique to evaluate the evidence integral. We illustrated the equivalence between Markov and Gibbs random fields, which allow a local modeling of spatial interactions between neighboring pixels by a stochastic description and are well suited to serve as prior models. Several potential functions have been given and we commented on their basic properties. Especially, the Gauss-Markov model rose our interest because it is partly analytically tractable. However, its linear character does not allow the modeling of sharp edges. We concluded that Markov or Gibbs random fields are promising to be used as prior information within a Bayesian framework for SAR image processing, since they are able to describe different spatial image properties.

As an overview of the state-of-the-art, several commonly used despeckling filters have been presented. We discussed their qualitative differences in terms of the modeling of the noise and of the noise-free cross-section and concluded that the quality of the despeckled data increases with model complexity. This motivated the employment of more powerful prior models, such as Gibbs random fields, which usually require iterative relaxation methods for the maximization of the resulting posterior. With regard to the optimization space, either slow optimal stochastic algorithms like simulated annealing or faster sub-optimal optimization schemes like the ICM algorithm have to be used. We illustrated that regardless of the kind of parametric prior optimal filtering results can only be achieved if the model parameters are estimated from the observed data. It was shown that fixed parametric models cannot be considered to be satisfactory since they may introduce a strong bias and sometimes even further degrade the noisy observation. For this purpose, we investigated methods for information extraction and model parameter estimation from noisy data. We noted that the computational complexity of the model parameter estimation mainly depends on the kind of the employed model and demonstrated that non-linear models principally possess higher modeling capabilities than linear ones. However, parameter estimation for non-linear models proved to be difficult and too slow for practical applications. In the following, we limited ourselves to the linear Gauss-Markov model for computational reasons and discussed several parameter estimation techniques for incomplete data. A quantitative comparison clearly favored a Bayesian iterative approach based on the maximization of the evidence.

Based on the above findings, we developed the idea of a Bayesian algorithm for SAR image filtering and feature extraction. We derived the MAP equation for Gauss-Markov random fields as a prior and the speckle distribution as likelihood function. Since the optimization space is well-behaved, we used a deterministic ICM algorithm for maximization of the posterior probability and approximated the Gaussian prior by a square-root Gamma distribution for faster computation. In order to calculate the evidence, which is required for model parameter estimation from incomplete data, we approximated the posterior product by a Gaussian enabling an evaluation of the evidence for a given model parameter vector. The accuracy of this approach was verified by investigating the resulting model parameter estimation for the limiting case of complete data. We studied two iterative schemes for evidence maximization. One of these methods showed to be related to the EM-type algorithm and we identified the function to be maximized in the M-step equivalence to be the Occam factor, which explained the qualitative difference between these two estimation techniques. We showed that an introduced bias of our MAP estimator can be easily compensated by an analytically given correction factor and that the mean of the original square-root intensity image is preserved with sufficient accuracy. Afterwards, we investigated the relations between evidence, Occam factor, σ of the Gauss-Markov model and mean square error for the selection of the optimal model order and decided to use a model of fixed high order, which yields an improved filtering performance in comparison with lower orders at the cost of a moderately increased computing time. Since SAR image despeckling requires a local processing to preserve spatially varying structures, we shortly studied the influence of the estimation window size. Finally, we investigated the signal space in which the despeckling with Gauss-Markov random fields is best performed and found optimal results in terms of the mean square error for a processing in the square-root intensity domain.

We identified the need to adapt the linear Gauss-Markov model to specific SAR image features, which were not satisfactorily restored with the used auto-regressive model. An additional processing for edge-preservation and scatterer detection was required. For this purpose, we used a region-growing backscatter segmentation which favors the formation of homogeneous, closed region labels. The restoration can now either be performed with the originally estimated model parameters or by taking into account the identified edges by means of an adaptive neighborhood system. The decision for any of these two assumptions was based on a local analysis of the coefficient of variation in order to verify the hypothesis of edges between non-textured areas of homogeneous backscatter. Since isolated targets have been found to be blurred after despeckling and to hinder a correct parameter estimation, a special processing of strong scatterers was introduced. Isolated scatterers are detected by a statistical analysis relying on a constant false alarm rate and are removed prior to filtering. In a post-processing step, removed targets and blurred scatterers that have not been detected in the pre-processing are re-inserted in the image contingent on an analysis of the ratio image between observed and filtered data.

8.2 Evaluation of Obtained Results and Outlook

For an evaluation of the proposed despeckling method, a filter comparison has been presented which was based on synthetic image examples and the computation of various quality measures. We stated that image quality cannot be consistently measured by a few numbers, and therefore also relied on a visual quality evaluation. As objective measures, we have chosen the mean square error, the preservation of the image mean, the maximum noise reduction and the first order statistical properties of the extracted speckle noise. Furthermore, speckle ratio images have been provided to visually evaluate introduced spatial distortions. However, we remind that image quality measures are application dependent. Users whose applications require smooth data will favor a different restoration than those who require the sharp preservation of edges. Consequently, the value of a certain filter always depends on the application it is intended for. On the other hand, for the development of our approach we did not have a particular application in mind but aimed at an optimal estimate of the radar cross-section. For this reason, and although this measure is not optimal, we mainly relied on the mean square error to determine the goodness of filtered synthetic data.

Both the visual and the numerical analysis provided convincing results for the proposed Model-Based Despeckling approach. Its improved filtering quality is mainly due to the employment of a much more complex model which is able to capture spatial information. Naturally, this increased complexity also results in comparatively rather high computation times, which however can be justified for the processing of high-resolution SAR data with high detail. For most SAR images of lower resolution which do not exhibit textural information, simpler and much faster filters may still be preferred.

We noted that the MBD filtered images look visually very attractive, but exhibit a slight blurring effect, which is equivalent to a reduction of image resolution and results from the employed Gauss-Markov model. While other filters often provide rather sharp images, their results stay much closer to the noisy observation preserving a certain degree of noise.

The extracted information in form of texture parameters and edges, which enables the high quality of the filtered images, has been shown in several examples to be of sufficient accuracy as well. Otherwise a reliable cross-section estimation of high quality would not have been possible.

The proposed approach allows a good restoration of SAR images and provides additional features. We have demonstrated that this information can be further used to characterize the image content. A first step in this direction was made with a query by image content system developed in a joint collaboration between DLR and the ETH Zürich where the extracted SAR features have been included. To improve the performance of both the despeckling and the query by image content system, the employment of more powerful models should be investigated. Non-linear models might make the introduced processing for edges and targets obsolete and capture additional spatial information. However, the parameter estimation from incomplete data for such models is still computationally prohibitive. A fast estimation technique, e.g. based on an approximation of non-linear models as hinted at in chapter four, might constitute a big leap forward for the application of such methods on a larger scale. We have demonstrated that the employment of sophisticated and more powerful models with parameter estimation taking into account the degrading noise indeed allows an improved data analysis and a better image restoration.

A Overview of SAR Sensors

In the following tables Tabs. A.1 and A.2 we provide an overview of several important airborne and spaceborne SAR sensors and SAR related space shuttle missions. This list cannot be complete due to the fast emerging increase of available, especially airborne, SAR sensors, which nowadays are often operated by private companies.

A.1 Spaceborne SAR Sensors

	Sensor	Nation	Altitude	Band	Polarization	Resolution
Satellite	SEASAT	USA	800km	L 1.3GHz	HH	23m
	ERS-1,2	Europe	785km	C 5.3GHz	VV	25m
	JERS-1,2	Japan	565km	L 1.2GHz	HH	30m
	RADARSAT	Canada	792km	C 5.3GHz	HH	28m
Space Shuttle	SIR-A	USA	259km	L 1.28GHz	HH	4.7m
	SIR-C	USA, Italy, Germany	215km	L 1.28GHz C 5.3GHz	HH, HV, VH, VV	6.1m
	X-SAR	USA, Italy, Germany	215km	X 9.6GHz	VV	6.1m

Table A.1: Selection of important spaceborne SAR sensors. The most important specifications like frequency band and resolution are given.

A.2 Airborne SAR Sensors

Nation	Agency	Frequency	Polarization	Resolution	Swath
USA	NASA/JPL	0.44, 1.25, 5.3GHz	HH, HV, VH, VV	8m	10-18km
Canada	INTERA	9.38GHz	HH	6m	60km
Denmark	TUD	5.3GHz	VV	2m	12-48km
France	CNES	9.38GHz	HH, VV	3m	11.6km
China	CAS	5.3GHz	HH, HV, VH, VV	10m	35 km
Germany	DLR	0.45, 1.3, 5.3, 9.6GHz	HH, VV	2.5m	4km

Table A.2: Selection of important airborne SAR sensors. The most important specifications like frequency and resolution are given.

B Optimization Methods

To find the lowest energetic state of a Gibbs distribution, i.e. the most probable realization, only small modifications to the Gibbs sampler or Metropolis algorithm for synthesis are necessary. Basically, there are two possible approaches to solve an optimization problem: The first approach consists in stochastic relaxation methods. These methods are usually rather slow, but are theoretically able to find the global minimum. The second approach consists in deterministic or semi-deterministic methods that converge rather fast, but cannot guarantee the final solution to be the global minimum and often get stuck in local minima depending on the space to be explored. These two basic approaches are illustrated in Fig. B.1. The choice for the method to apply must therefore depend on practical considerations, like computation times and the quality of the achieved solution, which is conditioned by the shape of the function to maximize.

B.1 Simulated Annealing

The task of finding the global minimum of a Gibbsian energy function $U(x)$ can be solved by introducing a new parameter T called temperature, which originates from statistical physics, where it is used to describe the temperature of an annealing process. Thus, this stochastic optimization procedure to find \hat{x}_{MAP} is known as simulated annealing (SA) [Aarts, Kirkpatrick, VanLaarhoven]. For the posterior Gibbs distribution according to Eq. 4.10 with the additional parameter T

$$p_T(X = x|Y = y) = \frac{1}{Z(T)} \exp\left(-\frac{U(x, y)}{T}\right), \quad (\text{B.1})$$

we find two limiting cases:

- $T \rightarrow \infty$ yields a uniform distribution with $p_T(X = x|Y = y) = \text{const}$.
- $T \rightarrow 0$ yields $p_T(X = x \neq \hat{x}_{MAP}|Y = y) = 0$ and $p_T(X = \hat{x}_{MAP}|Y = y) = 1$.

Both the Gibbs sampler and the Metropolis algorithm can use such a modified Gibbs distribution with a temperature T , which is sequentially lowered throughout the optimization process, in order to determine \hat{x}_{MAP} . Starting with a high value for T , all states of X are almost equally probable. By lowering T according to a pre-defined annealing schedule, the

distribution becomes more peaked and sampled configurations slowly tend to reach the most probable solution, i.e. the MAP estimate, regardless of the shape of the multi-dimensional energy landscape or the initial starting configuration.

It is crucial that the temperature is decreased slowly enough in order to allow to recover from local minima. It can be proven [Aarts] that the algorithm converges towards the global minimum if T is lowered not faster than $T^{(n)} = c/\log(n+1)$, where n denotes the iteration and c is a constant, which usually can be set to one. Since this annealing schedule is by far too slow for practical applications, many other schemes exist, like e.g. $T^{(n)} = k^n$, where k usually ranges from 0.9 to 0.99. Nonetheless, a stochastic relaxation is a computationally heavy method, which should only be employed if faster methods do not give satisfactory results. Similar, heuristically motivated methods [Creutz, Dueck] have been evaluated and applied in [Walessa95, Walessa96].

B.2 Iterated Conditional Modes

Another important optimization algorithm is the iterated conditional modes (ICM) approach [Besag86, Winkler]. This approach is fully deterministic and consequently cannot guarantee convergence to the global minimum. Its main advantages are its speed and a convergence after a small number of iterations. Starting with a good initial guess of the optimal configuration \hat{x}_{MAP} , which could be the observed data y , the algorithm works like this:

1. Each pixel X_i is assigned the value that minimizes the local energy function or, equivalently, maximizes the local conditional posterior probability $p(X_i = x_i | Y_i = y_i, X_j = x_j, j \in \mathcal{N}_i)$. This procedure decreases the global energy, thus converges towards the global or a local optimum.
2. Step one is repeated several times for the whole image until either a local minimum is reached, i.e. \hat{x} does not change any more, or until the number of pixel changes or their magnitude lies below a fixed threshold.

If, in step one, the value resulting in the steepest descent cannot be found analytically, as it is the case for most energy functions, the whole distribution has to be build, as done in the Gibbs sampler. Since this slows down convergence by a factor governed by the number of gray-values, some modifications of the ICM exist to circumvent this problem: By analogy to the Metropolis algorithm, a new candidate value is randomly chosen. If the local energy decreases, the current pixel is updated with this value, otherwise, the new value is always rejected. In contrast to the ICM technique, which is a steepest descent method, this modified approach is a slower converging and simpler optimization technique.

We emphasize again two important conditions to take maximum advantage of the ICM algorithm: The locally optimal value should be given analytically, and the energy landscape should be smooth and convex to favor convergence to the global optimum. If the latter condition is not fulfilled, it might prove helpful to perform an optimization at multiple

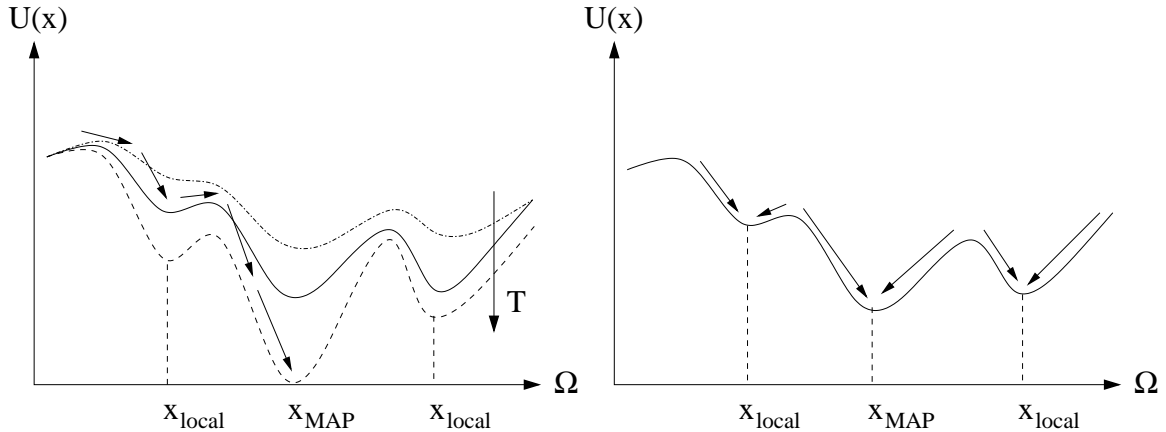


Figure B.1: Convergence of simulated annealing (SA), iterated conditional modes (ICM) and Newton-Raphson. Left: Optimization space at different temperatures (decreasing in the direction of the arrow) of the simulated annealing. If T is decreased correctly and slowly enough the estimate converges to the global minimum regardless of its initial state. Right: Optimization by ICM or Newton-Raphson. The initial configuration and the shape of the energy function determine the convergence to a local or to the global minimum.

scales [Terzopoulos, Walesa95] due to an often better behaving optimization function at coarser resolutions and the exploitation of inter-scale information.

B.3 Newton-Raphson Algorithm

If the posterior probability is well-behaved, a classical Newton-Raphson algorithm [Sivia] can also be applied for optimization. The Newton-Raphson algorithm is an iterative deterministic approach which is based on the computation of the gradient of the function to maximize. This requires the posterior distribution to exist in analytical form. Starting with a guess $x^{(n)}$ at iteration $n = 0$ the new update is computed by

$$x^{(n+1)} = x^{(n)} - (\nabla \nabla \log p(X = x | Y = y, \theta))^{-1} \nabla \log p(X = x | Y = y, \theta) \Big|_{x=x^{(n)}}, \quad (\text{B.2})$$

with ∇ as the Nabla operator and $\nabla \nabla \log p(X = x | Y = y, \theta)$ as the matrix of second derivatives.

We note that these derivatives can be calculated for the proposed posterior product of the square-root Gamma distribution and the Gauss-Markov model. For the given problem of cross-section estimation however, the required computational load compared to an ICM approach can be assumed to be much higher, since a very fast approximation has been derived for the latter. Qualitative differences between Newton-Raphson and ICM are not to be expected for the above mentioned posterior.

C Restoration of Blurred Images

The forward model used throughout this thesis only describes the noise process that degrades the data. However, this is a simplification of the reality, where the data are usually degraded by both a transformation and noise. This transformation can either be a point-spread-function modeling the image acquisition system or a non-linear transformation describing the relations between different physical parameters at the input and at the output. This more accurate imaging model [Geman84, Geman88] can be described by

$$y = F\{H(x)\} \odot n, \quad (\text{C.1})$$

where the original data x are subject to a linear transformation H (usually a blurring kernel), an additional non-linear transformation (e.g. the logarithm) denoted by F and some additive or multiplicative (denoted by the operator \odot) noise process N , which is assumed to be statistically independent of the random variable X . The respective sample realizations of X and N are x and n .

C.1 Deblurring of Noisy Images

It was shown in [Geman84] that this kind of full inverse problem can be modeled as a Gibbs distribution with a modified likelihood function, where x is replaced by the deterministic forward model $F\{H(x)\}$. In this case, the local likelihood term also comprises a neighborhood system which is determined by the width of the blurring function H . Unfortunately, this general model cannot be treated analytically any more and the computational load grows with the size of H .

As prior information for X , the full range of Gibbs models can be used without restrictions. In the case of a "simple" deblurring of images which are slightly affected by noise, an entropic prior is often used [Burch, Gull, Toma]. Conventional maximum entropy restoration is a special case of MAP estimation, in which the prior distribution for X is given by

$$p(X = x) = \frac{1}{Z} \exp \left(-\lambda \sum_{i=1}^{|X|} x_i \log x_i \right). \quad (\text{C.2})$$

The variable λ serves as a hyper-parameter which weights the relation between likelihood and prior and should be determined by estimation. The L-curve method is often used for this purpose, but Bayesian approaches can also be applied.

C.2 Filtering of SAR Images with Correlated Speckle

Although SAR data are also affected by a blurring function, the imaging process for detected data behaves differently: The blurring must be interpreted as to affect the already speckled data, i.e.

$$y = H(x \cdot n). \quad (\text{C.3})$$

Note that this is only approximately true for the pure image formation process. In fact, it is the complex SAR data that are affected by a point-spread-function and usually negligible thermal noise. Hence, the physical causes leading to both speckle and to the blurring effect in the observed image are different from what is described by the above model.

Nonetheless, we have to deal with correlated speckle noise (cf. chapter two). However, the restoration approach for blurred images, i.e. the model of Eq. C.1 with uncorrelated noise, is not applicable unless complex data are considered. Furthermore, the implementation of the more accurate model in Eq. C.3 is not straightforward and complicated. As a consequence, in order to cope with correlated speckle, methods like multi-looking or subsampling are often applied in a pre-processing step to decrease correlation [Quegan]. We already stated that complex SAR data are often slightly oversampled and that detected data may also exhibit significant correlation.

In the proposed Bayesian approach, we do not take into account speckle correlation and only consider possibly pre-processed SAR images of almost uncorrelated speckle for two main reasons:

1. The modeling of correlated speckle within the Bayesian framework is a difficult problem. Even if this was possible, resulting algorithms would probably be very complex and by far too slow for practical applications.
2. Since we aim at texture extraction from SAR data using small neighborhoods, correlated noise of a similar correlation length strongly hinders a reliable parameter estimation.

Hence, we neither consider a blurring function to increase image resolution, nor do we model the correlation properties of speckle. Until now, there are only a few approaches that incorporate the information of speckle correlation for the case of homogeneous cross-sections [Fjørtoft99b]. With the exception of the homomorphic Wiener filter, similar approaches do not yet exist for the full problem of SAR image restoration.

D Generalized Model-Based Noise Filtering

The new approach for noise filtering presented in this work was originally developed to reduce the effect of speckle in SAR images. Using the Bayesian formalism, we have demonstrated in chapter five that the estimation of hidden model parameters and the restoration of noisy images can be adapted to different cases by a change of the employed likelihood function. In this way, the proposed approach is easily modified to work for other applications of image restoration: Here, we present a generalized version of the filter for non-stationary additive Gaussian noise that was developed for the enhancement of interferometric digital elevation models (DEMs).

D.1 Model-Based Filtering of Non-Stationary Noise

Assuming a noise-free image x of the random process X , which is affected by zero mean additive white Gaussian noise with variance σ_n^2 , we find for the local likelihood function of the observed data y of process Y at site i

$$p(Y_i = y_i | X_i = x_i) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(y_i - x_i)^2}{2\sigma_n^2}\right). \quad (\text{D.1})$$

Equivalently to the log-intensity case presented in chapter five, this yields a weighted mean for the value that maximizes the local posterior distribution if a Gauss-Markov random field (cf. Eq. 3.66) is chosen as prior information:

$$\hat{x}_{i_{MAP}} = \frac{y_i\sigma^2 + \mu_i\sigma_n^2}{\sigma^2 + \sigma_n^2}, \quad (\text{D.2})$$

where σ and μ_i are the parameters of the Gauss-Markov prior. To estimate the texture parameters of the prior model, we need the diagonal components of the Hessian matrix of the posterior, which are found to be

$$h_{ii} = \frac{1}{\sigma_n^2} + \frac{1}{\sigma^2} \left(1 + \sum_{k,l} \theta_{k,l}^2\right). \quad (\text{D.3})$$



Figure D.1: Generalized Gauss-Markov noise filtering. Top row: Original Lena image (256×256 pixels), mean filtered image (7×7 pixels, $\text{MSE}=227$), Lee filtered image (7×7 pixels, $\text{MSE}=351$). Bottom row: Noisy image (additive Gaussian noise with $\sigma_n = 32$, $\text{MSE}=998$), wavelet filtered image ($\text{MSE}=170$), generalized Gauss-Markov filtered image (fifth order, $\text{MSE}=105$).

By means of these results, the presented Model-Based Despeckling algorithm can be adapted to work for additive noise. The model parameter estimation by iterative evidence maximization remains unchanged. An example of this generalized Gauss-Markov filtering for additive noise with known variance and a short comparison to other popular filters is illustrated in Fig. D.1.

The above approach can be extended to work for non-stationary noise by locally estimating the noise variance σ_n^2 . For this purpose, we do not measure the empirical variance $\text{Var}(y)$ of the observation, but locally determine $\sigma_n^2 \approx \text{Var}(y - \hat{x})$, where \hat{x} is a rough guess of the noise-free image obtained by e.g. a simple mean filter. This approach is necessary since the locally measured variance $\text{Var}(y)$ contains fluctuations induced by image structures resulting in an over-estimation of σ_n^2 and a too strong smoothing.

As \hat{x} is only a rough guess, the estimation of σ_n^2 is still affected by distortions: An under-estimation may occur if the estimate \hat{x} exhibits remaining noise. Consequently, the filtering result strongly depends on the local guess for σ_n^2 , whose quality is determined by the estimation window size for $\text{Var}(y - \hat{x})$ and the window size of the mean filter employed to compute \hat{x} .

In principle, this generalized noise adaptive filter can also be used for SAR image despeckling. However, it suffers from the typical shortcomings of the Gauss-Markov model to describe isolated targets and sharp edges. Moreover, since the noise is assumed to be locally additive and the estimate of its variance is subject to error, the filtering performance for SAR data is inferior to the presented Model-Based Despeckling algorithm without structural enhancements.

A similar approach of using Gauss-Markov random fields for the restoration of images affected by additive noise with additional structural information employed to limit the local neighborhood of the model was presented in [Jeng]. However, in this approach model parameters are not estimated from the observed noisy image but are derived from complete training data. We commented on the correct selection of model parameters in chapter four.

D.2 Model-Based DEM Filtering during the SRTM Mission

The presented generalized model-based filter for additive non-stationary noise was employed during the SRTM mission to filter digital elevation models (DEMs) derived from single-pass interferometry. These DEMs, though of very high quality due to high coherence of the image pairs, still exhibit a certain degree of noise, which requires filtering as a post-processing.

This noise, whose properties are unknown, varies throughout the DEM as a function of the local coherence and requires a locally adaptive noise estimation. Since digital elevation models do not contain sharp borders or isolated peaks, but show rather smooth variation, the employment of Gauss-Markov random fields as a prior proved to be a good choice. This became especially visible when a comparison with standard noise filters was made: Wavelet-based methods [Donoho] often introduce highly disturbing artefacts, while the Lee filter for additive noise [Lee81] does not smooth regions of high variance. Mean or median filters do not preserve image detail.

An example of DEM filtering with the proposed approach is illustrated in Fig. D.2. For visualization purposes we show wrapped DEMs, although the filtering was performed on unwrapped data. Unlike the original SRTM DEM, the filtered image is rather smooth but still contains all relevant structural information. Fig. D.3 shows the corresponding local slope $\Delta = \sqrt{\Delta x^2 + \Delta y^2}$ as a function of the gray-level gradients in range and azimuth Δx and Δy . While all details are hidden under noise for the unfiltered DEM, valuable slope information can be directly derived from the filtered data.

A perspective view of a part of the first DEM of the SRTM mission (White Sands, New Mexico, USA) derived by DLR is given in Fig. D.4. The noise, which affects the original interferometric DEM, is easily visible in this illustration. The filtered digital elevation model is shown below, where the high noise area of low coherence in the lower part of the original DEM was suppressed by incorporation of the image intensity, which can be used as an additional source of information for DEM regularization [Nico]. A full panoramic view of original and denoised height information with the overlaid unfiltered and despeckled image amplitudes, respectively, is finally presented in Fig. D.5.

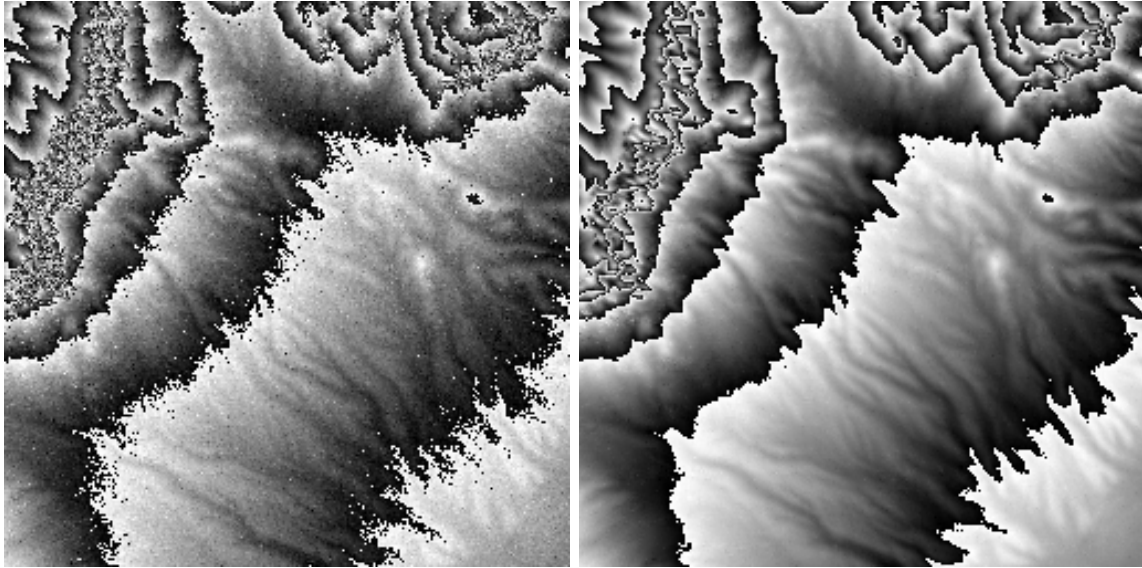


Figure D.2: Model-based DEM (Digital Elevation Model) filtering. Left: Original DEM (256×256 pixels) derived from single-pass interferometry during the SRTM mission. Right: Filtered DEM with the proposed generalized model-based noise filter. Wrapped DEMs are displayed for better visualization. The filtering was performed on unwrapped data.

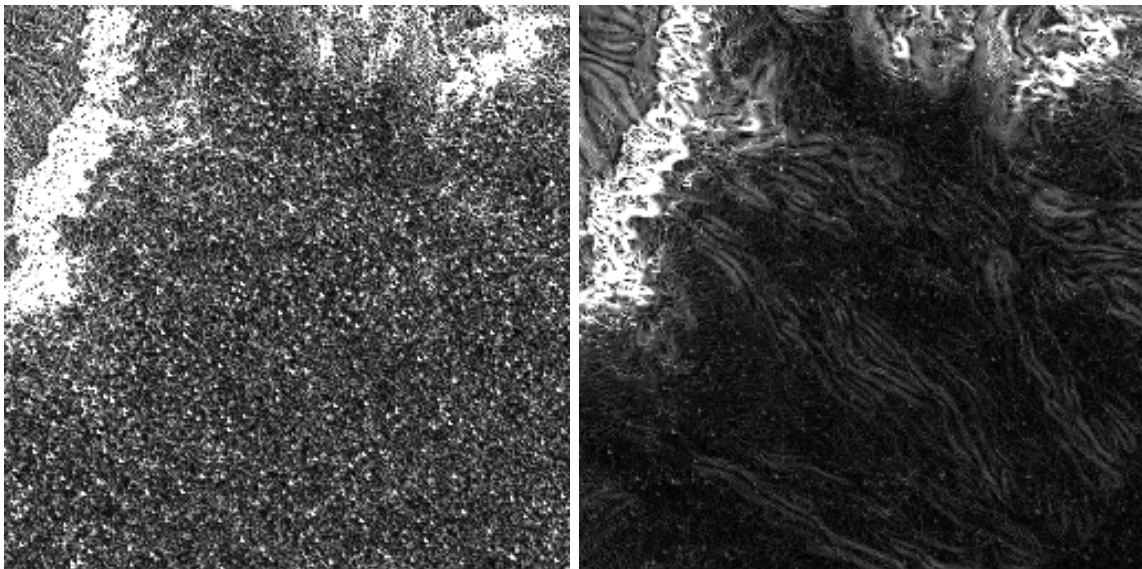


Figure D.3: Gradient images of the original and filtered unwrapped DEM. Left: Measured gradient in each pixel extracted from the original SRTM DEM as shown in Fig. D.2. Right: Measured gradient in each pixel extracted from the corresponding filtered DEM.

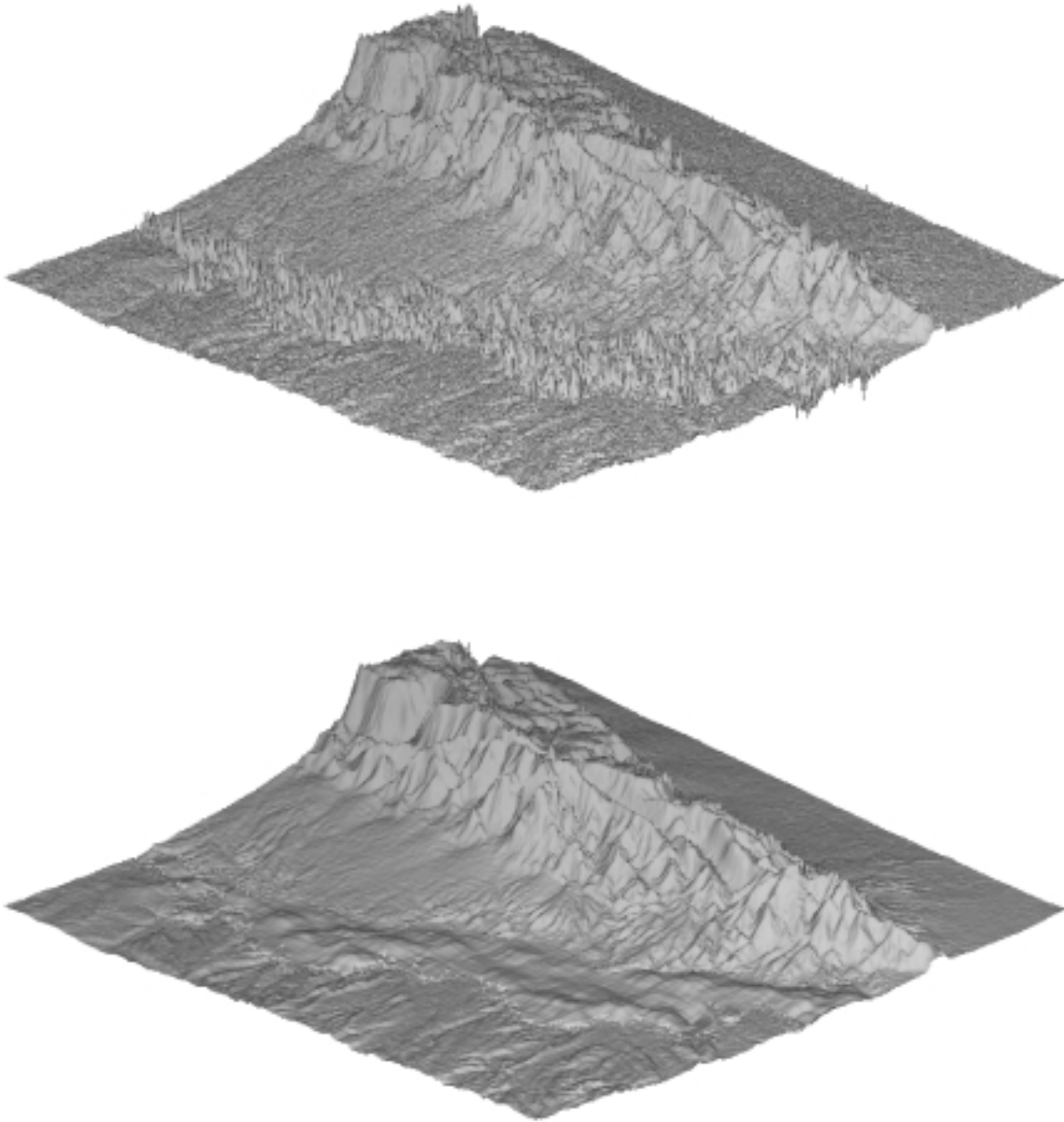


Figure D.4: Perspective view of original and filtered DEM. Top: Original DEM (512×512 pixels, first SRTM DEM over White Sands, New Mexico, USA) derived from single-pass interferometry during the SRTM mission. Bottom: Corresponding view of the filtered DEM.

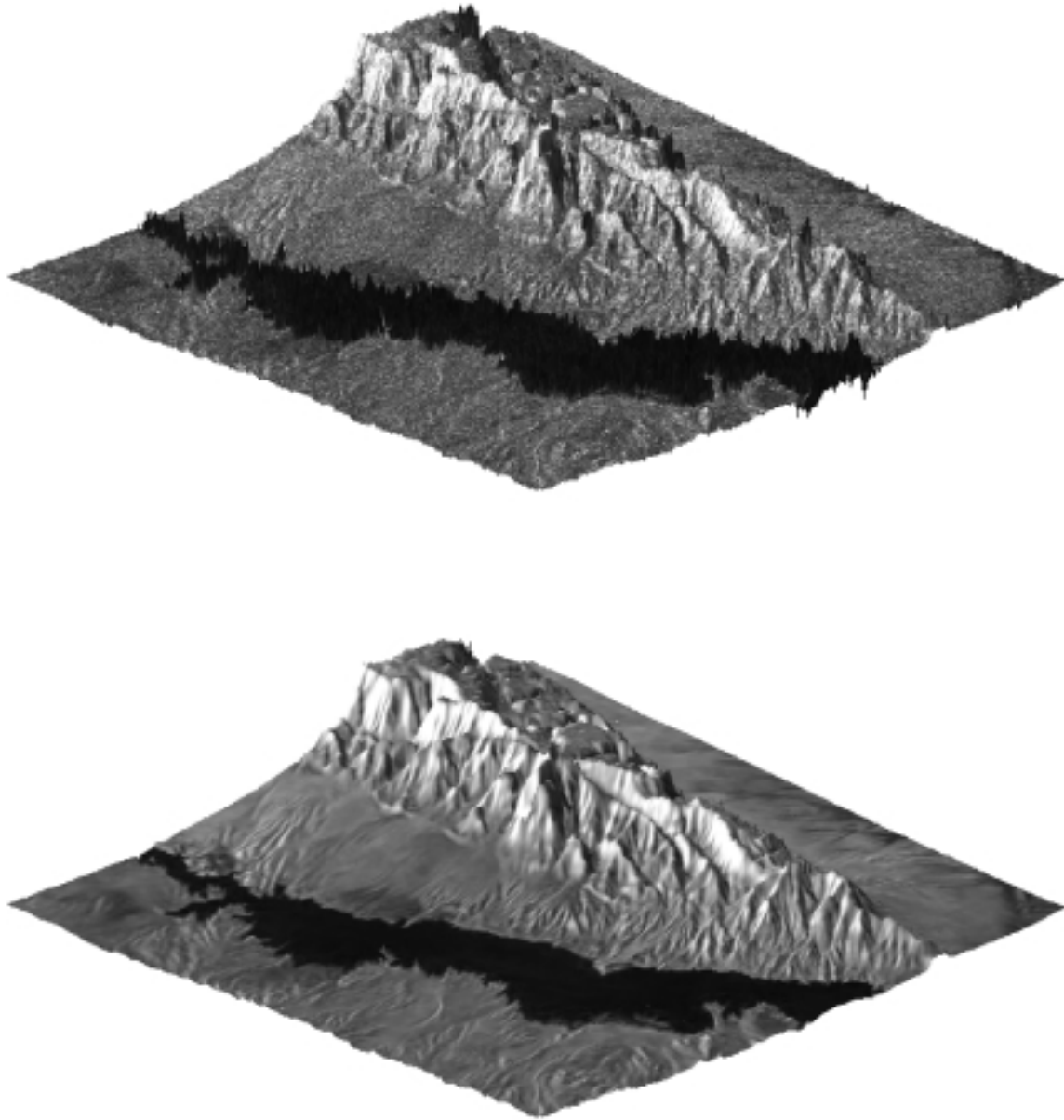


Figure D.5: Perspective view of original and filtered landscape. Top: Original DEM (512×512 pixels) with overlaid measured SAR image amplitude. Bottom: Corresponding view of the filtered DEM overlaid with the despeckled SAR image amplitude.

E Symbols and Acronyms

E.1 List of Symbols

We list the most important functions and symbols that can be found in this thesis. Due to the diversity of covered subjects, some symbols may have several meanings. The correct interpretation, however, should become clear from the context.

The following symbols have been used in this work:

A	amplitude
\mathbf{A}	matrix of correlation factors
B	signal bandwidth
$\mathbf{B}\sigma^2$	covariance matrix of N
\mathbf{b}	vector of correlation factors
C	cost function, contrast between two regions
C_{n_I}	auto-covariance function of intensity speckle
CV_A	coefficient of variation for amplitude
CV_I	coefficient of variation for intensity
CV_P	coefficient of variation for power
CV_S	coefficient of variation for square-root intensity
CV_Z	coefficient of variation for log intensity
\mathcal{C}	clique system
\mathcal{C}_k	clique system of k elements
c	speed of light, clique index, constant
D	antenna width
E	complex backscattered field, log-evidence
E_0	backscattered field variable
E_i	complex backscattered field of one scatterer
E_{i0}	magnitude of backscattered field
$E\{\}$	expectation value
\mathcal{E}	space of gray-values
$F\{\}$	non-linear transformation
f_c	carrier frequency
f_T	pulse repetition frequency
G	number of gray-levels
H	sensor altitude

H	Hessian matrix
$H()$	linear blurring kernel
h_{ii}	main diagonal element of the Hessian
I	intensity, additional information
I_0	intensity variable
i	index, pixel site
j	complex number, pixel site
K	proportionality constant, filter parameter
K_S	Kullback's information
$K_{\nu-L}$	Bessel function of order $\nu - L$
K_1	normalization constant
k	wave number, number of expectation values to calculate, weighting constant, number of pixels within a clique
k, l	indices of neighboring pixels
L	equivalent number of looks, form parameter of the Gamma distribution
M	data model
M_1, \dots, M_N	different data models
M_{MAP}	maximum a posteriori model
M_{ML}	maximum likelihood model
M_{order}	model with a given order
m	probability density or distribution function
N	number of individual scatterers or elements, noise process
N_1, N_2	number of pixels of two adjacent segments
N_i	noise process at site i
\mathcal{N}	neighborhood system
\mathcal{N}_i	neighborhood of x_i
$\mathcal{N}(a, b)$	Gaussian noise process of mean a and variance b
n	number of current iteration, realization of noise, ratio image
n_i	realization of noise at site i
n_I	intensity speckle
n_S	square-root intensity speckle
n_Z	log intensity speckle
n_i	realization of additive noise at site i
P	received power or intensity, probability
P_0	receiver power or intensity variable
P_{fa}	probability of false alarm
$p()$	probability density or distribution function
$p_\xi()$	probability density or distribution function of a random variable ξ
$p_T()$	Gibbs distribution with temperature T
Q	Q function of the EM algorithm
R	radial distance from the sensor
$R\{\}$	regularization term
$R(t)$	radial distance from sensor as a function of time
R_0	minimum radial distance from sensor
R_i	radial distance to a scatterer, region of homogeneous backscatter

R_{n_I}	auto-correlation function of intensity speckle
R_I	auto-correlation function of intensity
r	radial distance, number of regions for segmentation
r_x	range resolution
r_y	azimuth resolution
S	square-root intensity, entropy
S_0	square-root intensity variable
s	standard deviation
T	pulse repetition time, temperature
t	time
U	global energy
U_i	local energy
V	weighting factor, global potential function
V_c	local potential function
v	sensor velocity
$Var()$	variance of a random variable
X	random variable
X_i	random variable at site i
x	range coordinate, realization of X
x_{down}	pixel below a given center pixel
x_{left}	pixel left to a given center pixel
x_{right}	pixel right to a given center pixel
x_{up}	pixel above a given center pixel
\mathbf{x}	cross-section in vector notation
\hat{x}	estimate of x
\hat{x}_{MAP}	maximum a posteriori estimate of x
$\hat{x}_{MAP_1}, \hat{x}_{MAP_2}$	MAP estimates obtained under different model assumptions
\hat{x}_{ML}	maximum likelihood estimate of x
x_i	realization of X_i
\hat{x}_i	estimate of x at site i
\hat{x}_{iMAP}	maximum a posteriori estimate of x at site i
$x_i^{(n)}$	realization of X_i at iteration n
$x_{k,l}$	neighboring pixel to x_i
$x'_{k,l}$	symmetric neighbor to x_i with regard to $x_{k,l}$
$x^{(n)}$	realization of X at iteration n
x_0	shift along the x-axis
Y	random variable
Y_i	random variable at site i
y	azimuth coordinate, realization of Y
y_i	realization of Y_i
y_1	lower bound
y_2	upper bound
Z	log intensity, global partition function
Z_0	log intensity variable
Z_S	log square-root intensity

Z_c	partition function for the approximated model
Z_{dB}	intensity in decibels
Z_i	local partition function
α	scaling factor, filtering constant
β_y	antenna azimuth aperture
$\Gamma()$	Gamma function
Δ	spread of the potential function, smoothing parameter, slope
ΔR	general sensor resolution
ΔU	energy difference
Δt	time interval
Δx	difference between center and neighboring pixels, shift and slope in range direction
$\Delta \mathbf{x}$	difference vector
Δy	azimuth distance, shift and slope in azimuth direction
$\delta_{k,l}$	step width for directional texture parameters
$\delta_{k,l}^{(n)}$	step width for directional texture parameters at iteration n
δ_σ	step width for σ
δx	step width
δy	infinitesimal offset
ζ_i	phase shift of one scatterer
θ	model parameter
θ_c	model parameter for clique c
$\theta_{k,l}$	model parameter corresponding to $x_{k,l}$ and $x'_{k,l}$
$\tilde{\theta}_{k,l}$	renormalized model parameter
$\theta_{k,l}^{(n)}$	model parameter at iteration n
θ_ν	incidence angle
$\boldsymbol{\theta}$	model parameter vector
$\tilde{\boldsymbol{\theta}}$	renormalized model parameter vector
$\hat{\boldsymbol{\theta}}$	estimate of the model parameter vector
$\hat{\boldsymbol{\theta}}^{(n)}$	model parameter vector at iteration n
λ	wavelength of pulse, hyper-parameter
$\lambda^{(n)}$	estimate of λ at iteration n
λ_0	Lagrange multiplier
λ_1	Lagrange multiplier
μ_A	mean of amplitude
μ_I	mean of intensity
μ_S	mean of square-root intensity
μ_Z	mean of log intensity
μ_{ZS}	mean of log square-root intensity
μ_i	mean of the Gauss-Markov model
μ_{inner}, μ_{outer}	empirical mean computed in an inner, outer window
$\tilde{\mu}_i$	mean of the square-root Gamma approximation
μ_x	mean of the radar reflectivity, mean of the Gamma distribution
$\hat{\mu}_x$	estimated mean of the radar reflectivity

μ_{x_i}	mean backscatter value of segment i
μ_y	mean of the observation y
μ_σ	mean of the cross-section
ν	form parameter of the Gamma distribution
ρ	bounded ratio detector
ρ_{max}	decision threshold for the bounded ratio detector
σ	standard deviation, cross-section, standard deviation of the Gauss-Markov model
σ_A	standard deviation of the amplitude
σ_{DA}	width of the discontinuity adaptive potential
$\sigma_{ E }$	standard deviation of the Rayleigh distribution
σ_{MAP}	standard deviation of the posterior distribution
σ_P	standard deviation of the negative exponential distribution
σ_S	standard deviation of the square-root intensity
σ_X	standard deviation of the random variable X
σ_n	standard deviation of additive Gaussian noise
σ_x	standard deviation of the radar reflectivity x
σ_y	standard deviation of the observation y
τ	pulse length
τ'	equivalent pulse length
$\Phi_x(t)$	phase shift of signal in range
$\Phi_y(t)$	phase shift of signal in azimuth
ϕ	phase
ϕ_0	phase variable
ϕ_i	phase of one scatterer
$\psi()$	Digamma function
$\psi'()$	first derivative of the Digamma function
Ω	Occam factor
ω	angular frequency
∇	Nabla operator

E.2 List of Acronyms

The following acronyms have been used in this work:

CNES	Centre National d'Études Spatiales
CV	coefficient of variation
DEM	digital elevation model
DLR	Deutsches Zentrum für Luft- und Raumfahrt e.V.
dB	decibels
ENL	equivalent number of looks

ENVISAT	European Environmental Polar Orbit Earth Observation Mission
EM	expectation-maximization
EPOS	edge-preserving optimized speckle filter
ERS	European Remote-Sensing Satellite
FGAN	Forschungsgesellschaft für Angewandte Naturwissenschaften e.V.
GGMAP	Gamma-Gamma maximum a posteriori
GMRF	Gauss-Markov random field
GRF	Gibbs random field
ICM	iterated conditional modes
IEM	iterative evidence maximization
JERS	Japanese Earth Resources Satellite
JPL	Jet Propulsion Laboratory
MAP	maximum a posteriori
MBD	Model-Based Despeckling
MBD-I	Model-Based Despeckling for intensity images
MBD-S	Model-Based Despeckling for square-root intensity images
MBD-Z	Model-Based Despeckling for log intensity images
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
MMSE	minimum mean square error
MPL	maximum pseudo-likelihood
MRF	Markov random field
MSE	mean square error
NASA	National Aeronautics and Space Administration
pdf	probability density or distribution function
RADARSAT	Radar Satellite
SA	simulated annealing
SAR	synthetic aperture radar
SEASAT	Sea Satellite
SIR	Shuttle Imaging Radar
SRTM	Shuttle Radar Topography Mission
X-SAR	X-band synthetic aperture radar

Bibliography

- [Aarts] E.H.L. Aarts, J. Korst, "Simulated Annealing and Boltzman Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing," John Wiley & Sons Ltd., Tiptree, Essex, 1989.
- [Aiazzi] B. Aiazzi, L. Alparone, S. Baronti, G. Borri, C. Susini, "Multiresolution De-Speckle Based on Laplacian Pyramids," Proc. IGARSS'96, Vol. 1, pp. 411-413, Lincoln, Nebraska, USA, May 1996.
- [Arsenault] H.H. Arsenault, G. April, "Properties of Speckle Integrated with a Finite Aperture and Logarithmically Transformed," Journal of the Optical Society of America, Vol. 66, No. 11, pp. 1160-1163, Nov. 1976.
- [Bamler] R. Bamler, B. Schättler, "SAR Data Acquisition and Image Formation," in G. Schreier (Ed.), SAR Geocoding: Data and Systems, Wichmann, Karlsruhe, pp. 53-102, 1993.
- [Barber] B.C. Barber, "Theory of Digital Imaging from Orbital Synthetic-Aperture Radar," International Journal of Remote Sensing, Vol. 6, No. 7, pp. 1009-1057, 1985.
- [Bayes] T. Bayes, "An Essay Towards Solving a Problem in the Doctrine of Chances," Phil. Trans. of the Royal Society, Vol. 53, pp. 370-418, 1763.
- [Belhadj] Z. Belhadj, S. Ben Jebara, H. Maatar, "Adaptive Speckle Filtering by Multiresolution Analysis," Proc. IGARSS'99, Vol. 2, pp. 1235-1237, Hamburg, Germany, Jun. 1999.
- [Bernoulli] J. Bernoulli, "Ars Conjectandi," Thurnisiorum, Basel, 1713.
- [Besag74] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," Journal of the Royal Statistical Society, series B, Vol. 36, pp. 192-236, 1974.
- [Besag86] J. Besag, "On the Statistical Analysis of Dirty Pictures," Journal of the Royal Statistical Society, series B, Vol. 48, No. 3, pp. 259-302, 1986.
- [Blahut] R.E. Blahut, "Principles and Practice of Information Theory," Addison-Wesley.
- [Bouman] C.A. Bouman, M. Shapiro, "A Multiscale Random Field Model for Bayesian Image Segmentation," IEEE Trans. Image Processing, Vol. 3, No. 2, pp. 162-177, Dec. 1994.

- [Bretthorst] G.L. Bretthorst, "Bayesian Spectrum Analysis and Parameter Estimation," Lecture Notes in Statistics, Vol. 48, Springer-Verlag, Berlin, 1988.
- [Bronstein] I.N. Bronstein, K.A. Semendjajew, "Taschenbuch der Mathematik," Verlag H. Deutsch, Frankfurt/Main, 1981.
- [Burch] S.F. Burch, S.F. Gull, J. Skilling, "Image Restoration by a Powerful Maximum Entropy Method," Computer Vision, Graphics and Image Processing, Vol. 23, pp. 113-128, 1983.
- [Cabada] J.C. Cabada, A. Lopès, R. Fjørtoft, "On the Use of Ratio Likelihood and GNC for the Restoration of SAR Images," Proc. IGARSS'98, Vol. 1, pp. 16-18, Seattle, Washington, USA, Jul. 1998.
- [Caves] R.G. Caves, "Automatic Matching of Features in Synthetic Aperture Radar Data to Digital Map Data," PhD thesis, University of Sheffield, Great Britain, Jun. 1993.
- [Chellappa] R. Chellappa, S. Chatterjee, R. Bagdazian, "Texture Synthesis and Compression Using Gaussian-Markov Random Field Models," IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-15, No. 2, pp. 298-303, Mar./Apr. 1985.
- [Cook94] R. Cook, I. McConnell, "MUM (Merging Using Moments) Segmentation for SAR Images," Proc. European Symposium on Remote Sensing, SAR Data Processing for Remote Sensing, SPIE Vol. 2316, pp. 92-103, Rome, Italy, 1994.
- [Cook96] R. Cook, I. McConnell, D. Stewart, C.J. Oliver, "Segmentation and Simulated Annealing," Proc. European Symposium on Remote Sensing, SAR Image Analysis, Modelling, and Techniques III, SPIE Vol. 2955, pp. 30-37, Taormina, Italy, Sep. 1996.
- [Cox] R.T. Cox, "Probability, Frequency and Reasonable Expectation," American Journal of Physics, Vol. 14, pp. 1-13, 1946.
- [Creutz] M. Creutz, "Microcanonical Monte Carlo Simulation," Physical Review Letters, Vol. 50, pp. 1411-1414, 1983.
- [Crimmins] T.R. Crimmins, "Geometric Filter for Reducing Speckle," Applied Optics, Vol. 24, pp. 1438-1443, 1985.
- [Cross] G.R. Cross, A.K. Jain, "Markov Random Field Texture Models," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 5, No. 1, pp. 25-39, Jan. 1983.
- [Datcu98] M. Datcu, K. Seidel, M. Walessa, "Spatial Information Retrieval From Remote Sensing Images - Part I: Information Theoretical Perspective," IEEE Trans. Geoscience and Remote Sensing, Vol. 36, No. 5, pp. 1431-1445, Sep. 1998.
- [Datcu99] M. Datcu, K. Seidel, G. Schwarz, "Information Mining in Remote Sensing Image Archives," I. Kanellopoulos, G. Wilkinson, T. Moons (Eds.), Machine Vision and Advanced Image Processing in Remote Sensing (MAVIRIC), Springer, pp. 199-212, 1999.

- [Dempster] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, Vol. 39, No. 1, pp. 1-38, 1977.
- [Derin87] H. Derin, H. Elliott, "Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 9, No. 1, pp. 39-55, Jan. 1987.
- [Derin90] H. Derin, P.A. Kelly, G. Vezina, S.G. Labitt, "Modeling and Segmentation of Speckled Images Using Complex Data," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 28, No. 1, pp. 76-87, 1990.
- [Desnos] Y.-L. Desnos, V. Matteini, "Review on Structure Detection and Speckle Filtering on ERS-1 Images," *EARSeL Advances in Remote Sensing*, Vol. 2, No. 2 - VI, pp. 52-65, 1993.
- [Dewaele] P. Dewaele, P. Wambacq, A. Oosterlinck, J.L. Marchand, "Comparison of some Speckle Reduction Techniques for SAR Images," *Proc. IGARSS'90*, Vol. 3, pp. 2417-2422, College Park, Maryland, USA, May 1990.
- [Dias] J.M.B. Dias, T.A.M. Silva, J.M.N. Leitão, "Adaptive Restoration of Speckled SAR Images," *Proc. IGARSS'98*, Vol. 1, pp. 19-23, Seattle, Washington, Jul. 1998.
- [Donoho] D.L. Donoho, "De-Noising by Soft-Thresholding," *IEEE Trans. on Information Theory*, Vol. 41, No. 3, pp. 613-627, 1995.
- [Dueck] G. Dueck, "New Optimisation Heuristics: The Great Deluge Algorithm and the Record-to-Record Travel," *Journal of Computational Physics*, Vol. 104, pp. 86-92, 1993.
- [EM] For an extensive list of references to papers describing applications of the EM algorithm, see <http://www.engineering.usu.edu/Departements/ece/Publications/Moon>.
- [Fjørtoft96] R. Fjørtoft, F. Lebon, F. Séry, A. Lopès, P. Marthon, E. Cubero-Castan, "A Region-Based Approach to the Estimation of Local Statistics in Adaptive Speckle Filters," *Proc. IGARSS'96*, Vol. 1, pp. 457-459, Lincoln, Nebraska, USA, May 1996.
- [Fjørtoft99a] R. Fjørtoft, "Segmentation d'images radar par détection de contours," PhD thesis, Institut National Polytechnique de Toulouse, France, Mar. 1999.
- [Fjørtoft99b] R. Fjørtoft, A. Lopès, "Comparison of Estimators of the Mean Radar Reflectivity from a Finite Number of Correlated Samples," *Proc. IGARSS'99*, Vol. 3, pp. 1549-1551, Hamburg, Germany, Jun. 1999.
- [Franceschetti] G. Franceschetti, V. Pascazio, G. Schirinzi, "Iterative Homomorphic Technique for Speckle Reduction in Synthetic-Aperture Radar Imaging," *Journal of the Optical Society of America*, Vol. 12, No. 4, pp. 686-694, Apr. 1995.
- [Frost82a] V.S. Frost, J.A. Stiles, K.S. Shanmugan, J.C. Holtzman, "A Model for Radar Images and its Application to Adaptive Filtering of Multiplicative Noise," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 4, pp. 157-166, 1982.

- [Frost82b] V.S. Frost, K.S. Shanmugan, J.C. Holtzman, "Edge Detection for Synthetic Aperture Radar and Other Noisy Images," Proc. IGARSS'82, Vol. FA2, pp. 4.1-4.9, Munich, Germany, Jun. 1982.
- [Gagnon] L. Gagnon, A. Jouan, "Speckle Filtering of SAR Images - A Comparative Study Between Complex-Wavelet-Based and Standard Filters," Proc. Conf. Wavelet Applications in Signal and Image Processing V, SPIE Vol. 3169, San Diego, 1997.
- [Gelautz] M. Gelautz, "Integration of Layover Information into SAR Image Analysis," PhD thesis, Technische Universität Graz, Austria, May 1997.
- [Geman84] S. Geman, D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 6, No. 6, pp. 721-741, Nov. 1984.
- [Geman88] D. Geman, "Random Fields and Inverse Problems in Imaging," Lecture Notes in Mathematics, Springer-Verlag, pp. 117-193, 1988.
- [Goodman] J.W. Goodman, "Statistical Properties of Laser Speckle Patterns," in Laser Speckle and Related Phenomena, J.C. Dainty (Ed.), Springer, 1975.
- [Gull] S.F. Gull, G.J. Daniell, "Image Reconstruction from Incomplete and Noisy Data," Nature, Vol. 272, pp. 686-690, 1978.
- [Hagg] W. Hagg, M. Sties, "Efficient Speckle Filtering of SAR Images," in Proc. IGARSS'94, Vol. 4, pp. 2140-2142, Pasadena, California, Aug. 1994.
- [Hammersley] J.M. Hammersley, P. Clifford, "Markov Field on Finite Graphs and Lattices," 1971. Unpublished.
- [Hervet] E. Hervet, R. Fjørtoft, P. Marthon, A. Lopès, "Comparison of Wavelet-based and Statistical Speckle Filters," Proc. SAR Image Analysis, Modelling, and Techniques III, SPIE Vol. 3497, Barcelona, Spain, Sep. 1998.
- [Hoekman] D.H. Hoekman, "Speckle Ensemble Statistics of Logarithmically Scaled Data," IEEE Trans. Geoscience and Remote Sensing, Vol. 29, No. 1, pp. 180-182, 1991.
- [Ising] E. Ising, "Beitrag zur Theorie des Ferromagnetismus," Zeitschrift für Physik, Vol. 31, pp. 253-258, 1925.
- [Jaynes82] E.T. Jaynes, "On the Rationale of Maximum-Entropy Methods," Proc. of the IEEE, Vol. 70, No. 9, pp. 939-952, Sep. 1982.
- [Jaynes] E.T. Jaynes, "Probability Theory: The Logic of Science." In preparation.
- [Jeffreys] H. Jeffreys, "Theory of Probability," Clarendon Press, Oxford, 1939.
- [Jeng] F.-C. Jeng, J.W. Woods, "Compound Gauss-Markov Random Fields for Image Estimation," IEEE Trans. Signal Processing, Vol. 39, No. 3, pp. 683-697, Mar. 1991.

- [Kelly] P.A. Kelly, H. Derin, K.D. Hartt, "Adaptive Segmentation of Speckled Images using a Hierarchical Random Field Model," *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. 36, No. 10, pp. 1628-1641, Oct. 1988.
- [Kirkpatrick] S. Kirkpatrick, C.D. Gellatt, M.P. Vecchi, "Optimization by Simulated Annealing," IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA, 1982.
- [Kreyszig] E. Kreyszig, "Statistische Methoden und ihre Anwendungen," Vandenhoeck & Ruprecht, Göttingen, 1979.
- [Krishnamachari] S. Krishnamachari, R. Chellappa, "Multiresolution Gauss-Markov Random Field Models for Texture Segmentation," *IEEE Trans. Image Processing*, Vol. 6, No. 2, pp. 251-267, Feb. 1997.
- [Kroschel] K. Kroschel, "Statistische Nachrichtentheorie: Signal- und Mustererkennung, Parameter- und Signalschätzung," Springer, Berlin, 1996.
- [Kuan85] D.T. Kuan, A.A. Sawchuk, T.C. Strand, P.C. Chavel, "Adaptive Noise Smoothing Filter for Images with Signal-Dependant Noise," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 7, No. 2, pp. 165-177, 1985.
- [Kuan87] D.T. Kuan, A.A. Sawchuk, T.C. Strand, P.C. Chavel, "Adaptive Restoration of Images with Speckle," *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. 4, pp. 157-166, 1987.
- [Lakshmanan] S. Lakshmanan, H. Derin, "Simultaneous Parameter Estimation and Segmentation of Gibbs Random Fields Using Simulated Annealing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 11, No. 8, pp. 799-813, Aug. 1989.
- [Laplace] P.S. de Laplace, "Théorie analytique des probabilités," Courcier Imprimeur, Paris, 1812.
- [Lee81] J.-S. Lee, "Refined Filtering of Image Noise Using Local Statistics," *Computer Graphics and Image Processing*, Vol. 17, pp. 24-32, 1981.
- [Lee83] J.-S. Lee, "A Simple Speckle Smoothing Algorithm for Synthetic Aperture Radar Images," *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 13, pp. 85-89, 1983.
- [Lee86] J.-S. Lee, "Speckle Suppression and Analysis for Synthetic Aperture Radar Images," *Optical Engineering*, Vol. 25, No. 5, pp. 636-643, May 1986.
- [Li] S.Z. Li, "Markov Random Field Modeling in Computer Vision," T.L. Kunii (Ed.), Springer-Verlag, Tokyo, 1995.
- [Linden] W. von der Linden, R. Preuss, V. Dose, "The Prior-Predictive Value: A Paradigm of Nasty Multidimensional Integrals," W. von der Linden et al. (Eds.), *Maximum Entropy and Bayesian Methods*, Kluwer Academic Publishers, pp. 319-326, 1999.
- [Lopes90a] A. Lopès, R. Touzi, E. Nezry, "Adaptive Speckle Filters and Scene Heterogeneity," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 28, No. 6, pp. 992-1000, Nov. 1990.

- [Lopes90b] A. Lopès, R. Touzi, E. Nezry, H. Laur, "Maximum A Posteriori Speckle Filtering and First Order Textural Models in SAR Images," Proc. IGARSS'90, Vol. 3, pp. 2409-2412, College Park, Maryland, USA, May 1990.
- [Lopes93] A. Lopès, E. Nezry, R. Touzi, H. Laur, "Structure Detection and Statistical Adaptive Speckle Filtering in SAR Images," International Journal of Remote Sensing, Vol. 14, No. 9, pp. 1735-1758, 1993.
- [Lüke90] H.D. Lüke, "Signalübertragung. Grundlagen der digitalen und analogen Nachrichtenübertragungssysteme," Springer Verlag, Berlin, 1990.
- [Lüke92] H.D. Lüke, "Korrelationssignale. Korrelationsfolgen und Korrelationsarrays in Nachrichten- und Informationstechnik, Meßtechnik und Optik," Springer Verlag, Berlin, 1992.
- [McConnell95] I. McConnell, R.G. White, C.J. Oliver, R. Cook, "Radar Cross-Section Estimation of SAR Images," Proc. Europto Conf. on SAR Image Analysis, Simulation and Modelling, SPIE Vol. 2584, pp. 164-175, Paris, France, 1995.
- [McConnell96] I. McConnell, C.J. Oliver, "Comparison of Annealing and Iterated Filters for Speckle Reduction in SAR," Proc. Europto Conf. on SAR Image Analysis, Simulation and Modelling, SPIE Vol. 2958, pp. 74-85, Taormina, Italy, 1996.
- [MacKay] D.J.C. MacKay, "Bayesian Interpolation," Neural Computation, Vol. 43, 1992.
- [Manjunath] B.S. Manjunath, R. Chellappa, "Unsupervised Texture Segmentation using Markov Random Field Models," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 13, No. 5, pp. 478-482, May 1991.
- [Metropolis] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, "Equations of State Calculations by Fast Computing Machines," Journal of Chemical Physics, Vol. 21, pp. 1087-1091, 1953.
- [Moon] T.K. Moon, "The Expectation-Maximization Algorithm," IEEE Signal Processing Magazine, pp 47-60, Nov. 1996.
- [Nezry] E. Nezry, A. Lopès, F. Yakam-Simen, "Prior Scene Knowledge for the Bayesian Restoration of Mono- and Multi-Channel SAR Images," Proc. IGARSS'97, Vol. 2, pp. 758-760, Singapore, Aug. 1997.
- [Nico] G. Nico, M. Walessa, M. Datcu, "Topographic Mapping: A New Approach," Proc. IGARSS'98, Vol. 1, pp. 91-93, Seattle, Washington, Jul. 1998.
- [Oliver91] C.J. Oliver, "Information from SAR Images," J. Phys. D: Appl. Phys., Vol. 24, pp. 1493-1514, 1991.
- [Oliver93] C.J. Oliver, "Optimum Texture Estimators for SAR Clutter," J. Phys. D: Appl. Phys., Vol. 26, pp. 1824-1835, 1993.
- [Oliver94] C.J. Oliver, A. Blake, R.G. White, "Optimum Texture Analysis of Synthetic Aperture Radar Images," SPIE Vol. 2230, pp. 389-398, 1994.

- [Oliver96] C.J. Oliver, I. McConnell, D. Stewart, "Optimum Texture Segmentation of SAR Clutter," Proc. European Conf. on Synthetic Aperture Radar (EUSAR'96), pp. 81-84, Königswinter, Germany, Mar. 1996.
- [Oliver98] C.J. Oliver, S. Quegan, "Understanding Synthetic Aperture Radar Images," Artech House, Boston and London, 1998.
- [Papoulis77] A. Papoulis, "Signal Analysis," McGraw-Hill, New York, 1977.
- [Papoulis84] A. Papoulis, "Probability, Random Variables and Stochastic Processes," McGraw-Hill, Singapore, 1984.
- [Pairman] D. Pairman, S.E. Belliss, S.J. McNeill, "Terrain Influences on SAR Backscatter around Mt. Taranaki, New Zealand," IEEE Trans. Geoscience and Remote Sensing, Vol. 35, No. 4, pp. 924-932, Jul. 1997.
- [Procello] L.J. Procello, N.G. Massey, R.B. Innes, J.M. Marks, "Speckle Reduction in Synthetic-Aperture Radars," Journal of the Optical Society of America, Vol. 66, No. 11, pp. 1305-1311, Nov. 1976.
- [Quegan] S. Quegan, "Interpolation and Sampling in SAR Images," IEEE Trans. Geoscience and Remote Sensing, Vol. 28, pp. 641-646, 1990.
- [Raney] R.K. Raney, G.J. Wessels, "Spatial Considerations in SAR Speckle Simulation," IEEE Trans. Geoscience and Remote Sensing, Vol. 26, No. 5, pp. 666-672, Sep. 1988.
- [Rees] W.G. Rees, M.J.F. Satchell, "The Effect of Median Filtering on Synthetic Aperture Radar Images," International Journal of Remote Sensing, Vol. 18, No. 13, pp. 2887-2893, 1997.
- [Rehrauer] H. Rehrauer, "The Role of Scale for Image Description and Image Query: An application to Remote Sensing Images," PhD thesis, ETH Zürich, Switzerland, March 2000.
- [Schistad92] A.H. Schistad, A.K. Jain, "Texture Analysis in the Presence of Speckle Noise," Proc. IGARSS'92, Vol. 2, pp. 884-886, Houston, Texas, May 1992.
- [Schistad97] A.H. Schistad Solberg, A.K. Jain, "Texture Fusion and Feature Selection applied to SAR Imagery," IEEE Trans. Geoscience and Remote Sensing, Vol. 35, No. 2, pp. 475-479, March 1997.
- [Schreier] G. Schreier, "Geometrical Properties of SAR Images," in G.Schreier (Ed.), SAR Geocoding: Data and Systems, Wichmann, Karlsruhe, pp. 103-134, 1993.
- [Schroeder98] M. Schröder, H. Rehrauer, K. Seidel, M. Datcu, "Spatial Information Retrieval from Remote Sensing Images - Part II: Gibbs Markov Random Fields," IEEE Trans. Geoscience and Remote Sensing, Vol. 36, No. 5, pp. 1446-1455, Sep. 1998.
- [Schroeder99a] M. Schröder, M. Walessa, H. Rehrauer, K. Seidel, M. Datcu, "Gibbs Random Field Models: A Toolbox for Spatial Information Extraction," Computer and Geosciences, 1999. Accepted for publication.

- [Schroeder99b] M. Schröder-Brzosniowsky, "Stochastic Modeling of Image Content in Remote Sensing Image Archives," PhD thesis, ETH Zürich, Switzerland, Nov. 1999.
- [RSIA1] M. Schröder, H. Rehrauer, "WWW online demonstrator: Interactive learning and probabilistic retrieval," <http://www.vision.ee.ethz.ch/~rsia/ClickBayes>, 1999. At least Netscape Communicator 4.06 or Internet Explorer 4.0 required.
- [RSIA2] M. Schröder, H. Rehrauer, "Multi-Mission Demonstrator (MMDEMO) for Content-Based Queries in Remote Sensing Data," Internet Online Query at: <http://www.vision.ee.ethz.ch/~rsia/mmdemo/cgi-bin/newsession1.cgi>.
- [Schwarz] G. Schwarz, M. Walessa, M. Datcu, "Speckle Reduction in SAR Images - Techniques and Prospects," Proc. IGARSS'97, Vol. 4, pp. 2031-2034, Singapore, 1997.
- [Shannon] C.E. Shannon, "A Mathematical Theory of Communication," Bell Sys. Tech. J., Vol. 27, pp 379-423 and 623-656, 1948.
- [Shengh] Y. Shengh, Z.-G. Xia, "A Comprehensive Evaluation of Filters for Radar Speckle Suppression," Proc. IGARSS'96, Vol. 3, pp. 1559-1561, Lincoln, Nebraska, USA, May 1996.
- [Sigelle] M. Sigelle, F. Tupin, "Champs de Markov en Traitement d'Image," Support de cours du module C3M, ENST Paris, 1999.
- [Sivia] D.S. Sivia, "Data Analysis: A Bayesian Tutorial," Clarendon Press, Oxford, 1996.
- [Skilling] J. Skilling, "Probabilistic Data Analysis: An Introductory Guide," Journal of Microscopy, Vol. 190, Pts 1/2, pp. 28-36, Apr./May 1998.
- [Small] D. Small, F. Holecz, E. Meier, D. Nüesch, A. Barmettler, "Geometric and Radiometric Calibration of RADARSAT Images," Proc. of Geomatics in the Era of RADARSAT, Ottawa, Canada, May 1997.
- [Smits] P.C. Smits, S.G. Dellepiane, "Synthetic Aperture Radar Image Segmentation by a Detail Preserving Markov Random Field Approach," IEEE Trans. Geoscience and Remote Sensing, Vol. 35, No. 4, pp. 844-857, Jul. 1997.
- [SRTM] Shuttle Radar Topography Mission (SRTM), Internet Homepages: <http://www.dlr.de/srtm> and <http://www.jpl.nasa.gov/srtm>, 2000.
- [Terzopoulos] D. Terzopoulos, "Image Analysis Using Multigrid Relaxation Methods," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 8, pp. 129-139, 1986.
- [Tikhonov] A.N. Tikhonov, V.A. Arsenin, "Solutions of Ill-Posed Problems," Halsted Press, New York, 1977.
- [Toma] C.E. Toma, M. Datcu, "Maximum Entropy and Minimum Cross-Entropy Methods in Image Processing," Proc. SPIE Vol. 1827, pp. 133-144, 1992.

- [Touzi88] R. Touzi, A. Lopès, P. Bousquet, "A Statistical and Geometrical Edge Detector for SAR Images," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 26, No. 6, pp. 764-773, Nov. 1988.
- [Touzi99] R. Touzi, "Speckle Filtering of Stationary and Nonstationary Scene Signals in SAR Imagery," *Proc. IGARSS'99*, Vol. 3, pp. 1807-1809, Hamburg, Germany, Jun. 1999.
- [Tupin98] F. Tupin, H. Maitre, J.-F. Mangin, J.-M. Nicolas, E. Pechersky, "Detection of Linear Features in SAR Images: Application to Road Network Extraction," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 36, No. 2, pp. 434-453, Mar. 1998.
- [Tupin99] F. Tupin, I. Bloch, H. Maitre, "A First Step Toward Automatic Interpretation of SAR Images Using Evidential Fusion of Several Structure Detectors," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 37, No. 3, pp. 1327-1343, May 1999.
- [Tur] M. Tur, K.C. Chin, J.W. Goodman, "When is Speckle Noise Multiplicative?," *Applied Optics*, Vol. 21, No. 7, pp. 1157-1159, Apr. 1982.
- [Ulander] L.M.H. Ulander, "Radiometric Slope Correction of Synthetic-Aperture Radar Images," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 34, No. 5, pp. 1115-1122, Sep. 1996.
- [VanLaarhoven] P.J.M. Van Laarhoven, E.H.L. Aarts, "Simulated Annealing: Theory and Applications," D. Reidel Publishing, Dordrecht, Holland, 1987.
- [Walessa95] M. Walessa, "Correlation Matching over Scale and Stochastic Optimization," Studienarbeit, Département Images, ENST Paris & IENT, RWTH Aachen, 1995.
- [Walessa96] M. Walessa, "Synthese von Sequenzen und Arrays mit guten inversen Filtern mittels stochastischer Optimierungsverfahren," Diplomarbeit, Institut für elektrische Nachrichtentechnik, RWTH Aachen, Sep. 1996.
- [Walessa97] M. Walessa, M. Datcu, "Bayesian Approach to SAR Image Reconstruction," *Proc. IGARSS'97*, Vol. 2, pp. 767-769, Singapore, 1997.
- [Walessa98] M. Walessa, M. Datcu, "Bayesian Reconstruction and Texture Segmentation of SAR Images," *Proc. IGARSS'98*, Vol. 1, pp. 13-15, Seattle, Washington, Jul. 1998.
- [Walessa99a] M. Walessa, M. Datcu, "Texture Reconstruction in Noisy Images," W. von der Linden et al. (Eds.), *Maximum Entropy and Bayesian Methods*, Kluwer Academic Publishers, pp. 65-72, 1999.
- [Walessa99b] M. Walessa, "Texture Preserving Despeckling of SAR Images using GMRFs," *Proc. IGARSS'99*, Vol. 3, pp. 1552-1554, Hamburg, Germany, Jun. 1999.
- [Walessa00] M. Walessa, M. Datcu, "Model-Based Despeckling and Information Extraction from SAR Images," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 38, No. 5, pp. 2258-2269, Sep. 2000.

- [Wei] G.C.G. Wei, M.A. Tanner, "A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms," *Journal of the American Statistical Association, Theory and Methods*, Vol. 85, No. 411, pp. 699-704, Sep. 1990.
- [White] R.G. White, "A Simulated Annealing Algorithm for Radar Cross-Section Estimation and Segmentation," *Proc. SPIE* Vol. 2234, pp. 231-239, 1994.
- [Winkler] G. Winkler, "Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical Introduction," Springer, Berlin, 1995.
- [Wu82] F.Y. Wu, "The Potts Model," *Review of Modern Physics*, Vol. 54, No. 1, pp. 235-267, 1982.
- [Wu92] Y. Wu, H. Maitre, "Smoothing Speckled Synthetic Aperture Radar Images by Using Maximum Homogeneous Regions Filters," *Optical Engineering*, Vol. 31, No. 8, pp. 1785-1792, Aug. 1992.
- [XSAR] X-SAR Data, Internet Homepage: <http://isis.dlr.de/xsar>.
- [Younes88] L. Younes, "Estimation and Annealing for Gibbsian Fields," *A. Inst. Henri Poincaré*, Vol. 24, No. 2, pp. 269-294, 1988.
- [Younes89] L. Younes, "Parametric Inference for Imperfectly Observed Gibbsian Fields," *Probability Theory and Related Fields*, Vol. 82, pp. 625-645, 1989.
- [Younes91] L. Younes, "Parameter Estimation for Imperfectly Observed Gibbs Fields and some Comments on Chaldmond's EM Gibbsian Algorithm," P. Barone and A. Frigessi (Eds.), *Stochastic Models, Statistical Methods and Algorithms in Image Analysis, Lecture Notes in Statistics*, Springer, 1991.

