

Geostatistics

WILEY SERIES IN PROBABILITY AND STATISTICS
APPLIED PROBABILITY AND STATISTICS SECTION

Established by **WALTER A. SHEWHART** and **SAMUEL S. WILKS**

Editors: *Vic Barnett, Noel A. C. Cressie, Nicholas I. Fisher,
Iain M. Johnstone, J. B. Kadane, David G. Kendall, David W. Scott,
Bernard W. Silverman, Adrian F. M. Smith, Jozef L. Teugels;
Ralph A. Bradley, Emeritus, J. Stuart Hunter, Emeritus*

A complete list of the titles in this series appears at the end of this volume.

Geostatistics

Modeling Spatial Uncertainty

JEAN-PAUL CHILÈS

Bureau de Recherches Géologiques et Minières

PIERRE DELFINER

TOTAL Exploration Production



A Wiley-Interscience Publication

JOHN WILEY & SONS, INC.

New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

This book is printed on acid-free paper. ☺

Copyright © 1999 by John Wiley & Sons, Inc. All rights reserved.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

Library of Congress Cataloging-in-Publication Data:

Chilès, Jean-Paul.

Geostatistics: modeling spatial uncertainty/Jean-Paul Chilès,
Pierre Delfiner.

p. cm. – (Wiley series in probability and statistics.
Applied probability and statistics section)

“A Wiley-Interscience publication.”

Includes bibliographical references and index.

ISBN 0-471-08315-1 (alk. paper)

I. Earth sciences—Statistical methods. I. Delfiner, Pierre.

II. Title. III. Series: Wiley series in probability and statistics.
Applied probability and statistics.

QE33.2.S82C45 1999

550'.72—dc21

98-3599

Printed in the United States of America

10 9 8 7 6

Contents

Preface	ix
Abbreviations	xi
Introduction	1
Types of Problems Considered, 2	
Description or Interpretation?, 7	
1. Preliminaries	11
1.1. Random Functions, 11	
1.2. On the Objectivity of Probabilistic Statements, 22	
1.3. Transitive Theory, 24	
2. Structural Analysis	29
2.1. General Principles, 29	
2.2. Variogram Cloud and Sample Variogram, 34	
2.3. Mathematical Properties of the Variogram, 57	
2.4. Regularization and Nugget Effect, 74	
2.5. Variogram Models, 80	
2.6. Fitting a Variogram Model, 104	
2.7. Variography in Presence of a Drift, 115	
2.8. Simple Applications of the Variogram, 128	
2.9. Complements: Theory of Variogram Estimation and Fluctuation, 137	
3. Kriging	150
3.1. Introduction, 150	
3.2. Notations and Assumptions, 152	

3.3.	Kriging with a Known Mean, 154	
3.4.	Kriging with an Unknown Mean, 164	
3.5.	Estimation of a Spatial Average, 193	
3.6.	Selection of a Kriging Neighborhood, 201	
3.7.	Measurement Errors and Outliers, 210	
3.8.	Case Study: The Channel Tunnel, 215	
3.9.	Kriging under Inequality Constraints, 224	
4.	Intrinsic Model of Order k	231
4.1.	IRF-0 and IRF- k , 231	
4.2.	A Second Look at the Model of Universal Kriging, 233	
4.3.	Allowable Linear Combinations of Order k , 236	
4.4.	Intrinsic Random Functions of Order k , 243	
4.5.	Generalized Covariance Functions, 252	
4.6.	Estimation in the IRF Model, 265	
4.7.	Generalized Variogram, 276	
4.8.	Automatic Structure Identification in the General Case, 281	
5.	Multivariate Methods	292
5.1.	Introduction, 292	
5.2.	Notations and Assumptions, 293	
5.3.	Simple Cokriging, 296	
5.4.	Universal Cokriging, 298	
5.5.	Case of Gradient Information, 313	
5.6.	Multivariate Random Functions, 321	
5.7.	Shortcuts, 351	
5.8.	Space-Time Models, 362	
6.	Nonlinear Methods	375
6.1.	Introduction, 375	
6.2.	Simple Methods for Estimating a Point Distribution, 376	
6.3.	Local Estimation of a Point Distribution by Disjunctive Kriging, 388	
6.4.	Simple Methods for Estimating a Block Distribution, 419	
6.5.	Local Estimation of a Block Distribution by Disjunctive Kriging, 437	

7. Conditional Simulations	449
7.1. Introduction and Definitions, 449	
7.2. Direct Conditional Simulation of a Continuous Variable, 462	
7.3. Conditioning by Kriging, 465	
7.4. Turning Bands, 472	
7.5. Nonconditional Simulation of a Continuous Variable, 478	
7.6. Nonconditional Simulation of an IRF- k , 506	
7.7. Simulation of a Categorical Variable, 520	
7.8. Object-Based Simulations: Boolean Models, 545	
7.9. Constrained Simulations, 561	
7.10. Practical Considerations, 571	
7.11. Case Studies, 577	
8. Scale Effects and Inverse Problems	593
8.1. Introduction, 593	
8.2. Upscaling Permeability, 594	
8.3. Stochastic Differential Equations, 602	
8.4. Inverse Problem in Hydrogeology, 611	
Appendix	636
References	650
Index	687

Preface

This book covers a relatively specialized subject matter, geostatistics, as it was defined by Georges Matheron in 1962, when he coined this term to designate his own methodology of ore reserve evaluation. Yet it addresses a larger audience, for the applications of geostatistics now extend to many fields in the earth sciences, including not only the subsurface but also the land, the atmosphere, and the oceans.

The reader may wonder why such a narrow subject should occupy so many pages. Our intent was to write a short book. But this would have required us to sacrifice either the theory or the applications. We felt that neither of these options was satisfactory—there is no need for yet another introductory book, and geostatistics is definitely an applied subject. We have attempted to reconcile theory and practice by including application examples, which are discussed with due care, and about 160 figures. This results in a somewhat weighty volume, although hopefully more readable.

This book gathers in a single place a number of results that were either scattered, not easily accessible, or unpublished. Our ambition is to provide the reader with a unified view of geostatistics, with an emphasis on *methodology*. To this end we detail simple proofs when their understanding is deemed essential for geostatisticians, and omit complex proofs that are too technical. Although some theoretical arguments may fall beyond the mathematical and statistical background of practitioners, they have been included for the sake of a complete and consistent development that the more theoretically inclined reader will appreciate. These sections, as well as ancillary or advanced topics, are set in smaller type.

Many references in this book point to the works of Matheron and the Center for Geostatistics in Fontainebleau, which he founded at the Paris School of Mines in 1967 and headed until his retirement in 1996. Without overlooking the contribution of Gandin, Matérn, Yaglom, Krige, de Wijs, and many others, it is from Matheron that geostatistics emerged as a discipline in its own right—a body of concepts and methods, a theory and a practice—for the study of spatial phenomena. Of course this initial group spawned others, notably in

Europe and North America, under the impetus of Michel David and André Journel, followed by numerous researchers trained in Fontainebleau first, and then elsewhere. This book pays tribute to all those who participated in the development of geostatistics, and our large list of references attempts to give credit to the various contributions in a complete and fair manner.

This book is the outcome of a long maturing process nourished by experience. We hope that it will communicate to the reader our enthusiasm for this discipline at the intersection between probability theory, physics, and earth sciences.

ACKNOWLEDGMENTS

This book owes more than we can say to Georges Matheron. Much of the theory presented here is his work, and we had the privilege of seeing it in the making during the years that we spent at the Center for Geostatistics. In later years he always generously opened his door to us when we asked for advice on fine points. It was a great comfort to have access to him for insight and support. We are also indebted to the late Geoffrey S. Watson who showed an early interest in geostatistics and introduced it to the statistical community. He was kind enough to invite one of the authors to Princeton University, and as an advisory editor of the Wiley Interscience Series, made this book possible. We wish he had been with us to see the finished product.

The manuscript of this book greatly benefited from the meticulous reading and quest for perfection of Christian Lantuéjoul, who suggested many valuable improvements. We also owe much to discussions with Paul Switzer, whose views are always enlightening and helped us relate our presentation to mainstream statistics. We have borrowed some original ideas from Jean-Pierre Delhomme, who shared the beginnings of this adventure with us. Bernard Bourguin contributed to the illustrations. This book could not have been completed without the research funds of the Bureau de Recherches Géologiques et Minières, whose support is gratefully acknowledged.

We would like to express our thanks to John Wiley & Sons for their encouragement and exceptional patience during a project which has spanned many years, and especially to Bea Shube, the Wiley-Interscience Editor when we started, and her successors Kate Roach and Steve Quigley.

Finally, we owe our families, and especially our dear wives Chantal and Edith, apologies for all the time we stole from them and thanks for their understanding and forbearance.

La Villetterre
July 12, 1998

Abbreviations

c.d.f.	cumulative density function
i.i.d.	independent identically distributed
IRF	intrinsic random function
IRF- k	intrinsic random function of order k
m.s.	mean square
OK	ordinary kriging
p.d.f.	probability density function
RF	random function
SK	simple kriging
SRF	stationary random function
UK	universal kriging

Introduction

Geostatistics aims at providing quantitative descriptions of natural variables distributed in space or in time and space. Examples of such variables are

- ore grades in a mineral deposit,
- depth and thickness of a geological layer,
- porosity and permeability in a porous medium,
- density of trees of a certain species in a forest,
- soil properties in a region,
- rainfall over a catchment area,
- pressure, temperature, and wind velocity in the atmosphere,
- concentrations of pollutants in a contaminated site.

These variables exhibit an immense complexity of detail that precludes a description by simplistic models such as constant values within polygons, or even by standard well-behaved mathematical functions. Furthermore for economic reasons these variables are often sampled very sparsely. In the petroleum industry, for example, the volume of rock sampled typically represents a minute fraction of the total volume of a hydrocarbon reservoir. The following figures, from the Brent field in the North Sea, illustrate the orders of magnitude of the volume fractions investigated by each type of data (“cuttings” are drilling debris, and “logging” data are geophysical measurements in a wellbore):

Cores	0.000 000 001
Cuttings	0.000 000 007
Logging	0.000 001

By comparison, if we used the same proportions for an opinion poll of the 50 million US households, we would interview only between 0.05 and 50 households, while 1500 is standard. Yet the economic implications of sampling for natural resources development projects can be significant. The cost of an offshore development platform like that of the Brent field is about 2 billion

dollars. Similarly in the mining industry “the decision to invest up to 1–2 billion dollars to bring a major new mineral deposit on line is ultimately based on a very judicious assessment of a set of assays from a hopefully very carefully chosen and prepared group of samples which can weigh in aggregate less than 5 to 10 kilograms” (Parker, 1984).

Naturally these examples are extreme. Such investment decisions are based on studies involving many disciplines besides geostatistics, but they illustrate the notion of *spatial uncertainty* and how it affects development decisions. The fact that our descriptions of spatial phenomena are subject to uncertainty is now generally accepted, but for a time it met with much resistance, especially from engineers who are trained to work deterministically. In the oil industry there are anecdotes of managers who did not want to see uncertainty attached to reserves estimates because it did not look good—it meant incompetence. For job protection it was better to systematically underestimate reserves. (Ordered by his boss to get rid of uncertainty, an engineer once gave an estimate of proven oil reserves equal to the volume of oil contained in the borehole!) Such conservative attitude led to the abandonment of valuable prospects. In oil exploration profit comes with risk.

Geostatistics provides the practitioner with a methodology to quantify spatial uncertainty. Statistics come into play because probability distributions are the meaningful way to represent the range of possible values of a parameter of interest. In addition, a statistical model is well suited to the apparent randomness of spatial variations. The prefix “geo” emphasizes the spatial aspect of the problem. Spatial variables are not completely random but usually exhibit some form of structure, in an average sense, reflecting the fact that points close in space tend to assume close values. G. Matheron (1965) coined the term *regionalized variable* to designate a numerical function $z(x)$ depending on a continuous space index x , and combining high irregularity of detail with spatial correlation. Geostatistics can then be defined as “the application of probabilistic methods to regionalized variables.” This is different from the vague usage of the word in the sense “statistics in the geosciences.” In this book geostatistics refers to a specific set of models and techniques, largely developed by G. Matheron, in the lineage of the works of L. S. Gandin in meteorology, B. Matérn in forestry, D. G. Krige and H. J. de Wijs in mining, and A. Y. Khinchin, A. N. Kolmogorov, P. Lévy, N. Wiener, A. M. Yaglom, among others, in the theory of stochastic processes and random fields.

We will now give an overview of the various geostatistical methods and the types of problems they address and conclude by elaborating on the important difference between description and interpretation.

TYPES OF PROBLEMS CONSIDERED

The presentation follows the order of the chapters. For specificity the problems presented refer to the authors’ own background in earth sciences applica-

tions, but newcomers with different backgrounds and interests are encouraged to find equivalent formulations of the problems in their own disciplines. Geostatistical terms will be introduced and highlighted with italics.

Epistemology

The quantification of spatial uncertainty requires a model specifying the mechanism by which spatial randomness is generated. The simplest approach is to treat the regionalized variable as deterministic and the positions of the samples as random, assuming for example that they are selected uniformly and independently over a reference area, in which case standard statistical rules for independent random variables apply, such as that for the variance of the mean. If the samples are collected on a systematic grid, they are not independent and things become more complicated, but a theory is possible by randomizing the grid origin.

Geostatistics takes the bold step of associating randomness with the regionalized variable itself, by using a *stochastic model* in which the regionalized variable is regarded as one among many possible *realizations* of a *random function*. Does this make any sense? The objects we deal with—a mineral deposit, a petroleum reservoir—are perfectly deterministic. Probabilities and their experimental foundation in the famous “law of large numbers” require the possibility of repetitions, which are impossible with objects that are essentially unique. The objective meaning and relevance of a stochastic model under such circumstances is a fundamental question of epistemology that needs to be resolved. The clue is to carefully distinguish the model from the reality it attempts to capture. Probabilities do not exist in Nature but only in our models. We do not choose to use a stochastic model because we believe Nature to be random (whatever that may mean), but simply because it is analytically useful. We should also keep in mind that models have their limits and represent reality only up to a certain point. And finally, no matter what we do and how carefully we work, there is always a possibility that our predictions and our assessments of uncertainty turn out to be completely wrong, because for no foreseeable reason the phenomenon at unknown places is radically different than anything observed (what Matheron calls the risk of a “radical error”).

Structural Analysis

Having observed that spatial variability is a source of spatial uncertainty, we have to quantify and model spatial variability. What does an observation at a point tell us about the values at neighboring points? Can we expect continuity in a mathematical sense, or in a statistical sense, or no continuity at all? What is the signal-to-noise ratio? Are variations similar in all directions or is there anisotropy? Do the data exhibit any spatial trend? Are there characteristic scales and what do they represent? Is the histogram symmetric or skewed?

Answering these questions, among others, is known in geostatistics as *structural analysis*. One key tool is a structure function, the *variogram*, which de-

scribes statistically how the values at two points become different as the separation between these points increases. The variogram is the simplest way to relate uncertainty with distance from an observation. Other two-point structure functions can be defined that, when considered together, provide further clues for modeling. If the phenomenon is spatially homogeneous and densely sampled, it is even possible to go beyond structure functions and determine the complete *bivariate distributions* of measurements at pairs of points. In applications there is rarely enough data to allow empirical determination of multiple-point statistics beyond two points, a notable exception being when the data originate, or are borrowed, from images.

Unexpected difficulties arise when the data exhibit a systematic spatial effect, or trend, which in geostatistical theory is modeled as a space-varying mean called *drift*. The determination of the variogram in the presence of a drift is often problematic due to the unclear separation between global and local scales. A special theory with structural tools insensitive to drifts has been developed to deal with such cases (*intrinsic random functions*).

Survey Optimization

In resources estimation problems the question arises as to which sampling pattern ensures the best precision. The variogram alone permits a comparison between random, random stratified, and systematic sampling patterns. But in practice the design is often constrained by operational and economic considerations, and the real question is how to optimize the parameters of the survey. Which grid mesh should be used to achieve a required precision? What is the optimal spacing between survey lines? What is the best placement for an additional appraisal well? Does the information expected from acquiring or processing more data justify the extra cost and delay? What makes life interesting is that these questions must be answered, of course, prior to acquiring the data.

Interpolation

We often need to estimate the values of a regionalized variable at places where it has not been measured. Typically these places are the nodes of a regular grid laid out on the studied domain, the interpolation process being then sometimes known as “gridding.” Once grids are established, they are often used as “the” representation of reality, without reference to the original data. They are the basis for new grids obtained by algebraic or Boolean operations, contour maps, volumetric calculations, and the like. Thus the computation of grids deserves much care and cannot rely on simplistic interpolation methods.

The estimated quantity is not necessarily the value at a point; in many cases a grid node is meant to represent the grid cell surrounding it. This is typical for inventory estimation or for numerical modeling. Then we estimate the mean value over a cell, or a block, and more generally some weighted average.

In all cases we wish our estimates to be “accurate.” This means, first, that on the average our estimates are correct; they are not systematically too high

or too low. This property is captured statistically by the notion of *unbiasedness*. It is especially critical for inventory estimation and was the original motivation for the invention of *kriging*. The other objective is precision, and it is quantified by the notion of *error variance*, or its square root the *standard error*, which is expressed in the same unit as the data.

The geostatistical interpolation technique of kriging comes in different flavors qualified by an adjective: *simple kriging*, *ordinary kriging*, *universal kriging*, *intrinsic kriging*, and so on, depending on the underlying model. The general approach is to consider a class of unbiased estimators, usually linear in the observations, and to find the one with minimum uncertainty, as measured by the error variance. This optimization involves the statistical model established during the structural analysis phase, and there lies the fundamental difference with standard interpolation methods: these focus on modeling the interpolating surface whereas geostatistics focuses on modeling the phenomenon itself.

Integration of Multiparameter Information

In applications the greatest challenge is often to “integrate” (i.e., combine) information from various sources. To take a specific example, a petroleum geologist must integrate into a coherent geological model information from cores, cuttings, open hole well logs, dip and azimuth computations, electrical and acoustic images, surface and borehole seismic, and well tests. The rule of the game is: “Don’t offend anything that is already known.” Geostatistics and multivariate techniques provide the framework and the tools to build a consistent model.

The technique of *cokriging* generalizes kriging to multivariate interpolation. It exploits the relationships between the different variables as well as the spatial structure of the data. An important particular case is the use of slope information in conjunction with the variable itself. When the implementation of cokriging requires a statistical inference beyond reach, shortcuts can be used. The most popular ones are the *external drift* method and *collocated cokriging*, which use a densely sampled auxiliary field to compensate for the scarcity of observations of the variable of interest.

Indicator Estimation

We are interested in the event: “at a given point x the value $z(x)$ exceeds the level z_0 .” We can think of z_0 as a pollution alert threshold, or a cutoff grade in mining. The event can be represented by a binary function, the *indicator function*, valued 1 if the event is true, and zero if it is false, whose expected value is the probability of the event “ $z(x)$ exceeds z_0 .” Note that the indicator is a nonlinear function of the observation $z(x)$. The mean value of the indicator over a domain V represents the fraction of V where the threshold is exceeded. When we vary the threshold, it appears that indicator estimation amounts to the determination of the histogram or the cumulative distribution function of the values of $z(x)$ within V . The interesting application is to estimate this

locally over a domain v to obtain a *local distribution function* reflecting the values observed in the vicinity of v . *Disjunctive kriging*, a nonlinear technique based on a careful modeling of bivariate distributions, provides a solution to this difficult problem.

Selection and Change of Support Problems

The *support* of a regionalized variable is the averaging volume over which the data are measured or defined. Typically there are *point values* and *block values*, or high-resolution and low-resolution measurements. As the size of the support changes, the histogram of the variable is deformed, but there is no straightforward relationship between the distributions of values measured over two different supports, except under very stringent Gaussian assumptions. For example, ore sample grades and blocks grades cannot both be exactly lognormally distributed—although they might as approximations. Predicting the change of distribution when passing from one size of support to another, generally point to block, is the *change of support* problem. Specific *isofactorial models* are proposed to solve this problem.

Change of support is central in inventory estimation problems in which the resource is subject to *selection*. Historically the most important application has been in mining, where the decision to process the ore or send it to waste, depending on its mineral content, is made at the level of a block, say a cube of 10-m side, rather than, say, a teaspoon. The recoverable reserves then depend on the local distributions of block values. Modeling the effect of selection may be a useful concept in other applications, such as the delineation of producing beds in a petroleum reservoir, the remediation of contaminated areas, or the definition of pollution alert thresholds.

Simulation

Kriging, as any reasonable interpolation method, has a *smoothing effect*. It does not reproduce spatial heterogeneity. In the world of images we would say that it is not true to the “texture” of the image. This can cause significant biases when nonlinear effects are involved. To take a simple example, compare the length of an interpolated curve with the length of the true curve: it is much shorter—the true curve may not even have a finite length! Similarly, for the same average permeability, a porous medium has a very different flow behavior if it is homogeneous or heterogeneous.

This is where the stochastic nature of the model really comes into play. The formalism of *random functions* involves a family of alternative realizations similar in their spatial variability to the reality observed but different otherwise. By *simulation techniques* it is possible to generate some of these “virtual realities” and produce pictures that are true to the fluctuations of the phenomenon. A further step toward realism is to constrain the realizations to pass through the observed data, thus producing *conditional simulations*. By generating several of these digital models we are able to materialize spatial uncertainty. Then if we are interested in some quantity that depends on the

spatial field in a complex manner, we can compute a result for each simulation and study the statistical distribution of the results.

Scaling and Inverse Problems

Fluid flow, contaminant transport, heat transfer, wave propagation are typical examples of physical phenomena governed by partial differential equations. These equations depend on spatially distributed parameters, such permeability or acoustic impedance that can be considered regionalized variables. Unfortunately, it is usually impossible to measure the parameters directly at the scale at which we need them for a description of the flow. Laboratory measurements of permeability, for example, investigate very small volumes compared with the size of the modeling grid mesh. A shift to a larger scale, or *upscaling*, is needed, but it is not a simple matter of linear averaging over a larger support because permeability is not additive. The general upscaling problem is this: Find the physical characteristics of a fictitious homogeneous medium which, at a higher scale, behaves like the real heterogeneous medium. This involves the equations of the problem as well as the spatial distribution of the parameter.

Inverse problems refer to the reconstruction of the regionalized parameters of the equations from the observable response of the system at a limited number of points and a few values of the parameters themselves. The solution obtained matches both types of data and verifies the physical equation. Upscaling and inverse problems are still topics of active research.

Problems Omitted

A wide class of spatial problems concern the processing and analysis of images. This is a world by itself, and we will not enter it, even though there will be occasional points of contact. An image analysis approach very much in line with geostatistics, and developed in fact by the same group of researchers, is Mathematical Morphology (see Serra, 1982). Variables regionalized in time will also be left out. Even though geostatistical methods apply, the types of problems considered are often of an electrical engineering nature and are better handled by digital signal processing techniques. However, geostatistics is useful for phenomena regionalized in both space and time.

Finally the study of point patterns (e.g., the distribution of trees in a forest) and the modeling of data on a lattice are intentionally omitted from this book. The reader is referred to Cressie (1991) for a comprehensive overview of these two approaches and to Guyon (1993) for a presentation of Markov fields on a lattice.

DESCRIPTION OR INTERPRETATION?

Geostatistical methods are goal-oriented. Their purpose is not to build an explanatory model of the world but to solve specific problems using the minimal

prerequisites required, following the principle of parsimony. They are descriptive rather than interpretive models. We illustrate this important point with an example borrowed from contour mapping.

Mathematically inclined people—including the present authors—have long thought that computer mapping was the definitive, clean, and objective replacement of hand contouring. Hand-drawn maps are subjective; they can be biased consciously or unconsciously. Even when drafted honestly they seem suspect: If two competent and experienced interpreters can produce different maps from the same data, why should one believe any of them? And of course there is always the possibility of a gross clerical error such as overlooking or misreading some data points. By contrast, computer maps have all the attributes of respectability: they don't make clerical mistakes, they are "objective," reproducible, and fast. Yet this comparison misses an important point: it neglects the semantic content of a map. For a geologist, or a meteorologist, a map is far more than a set of contours: it represents *the state of an interpretation*. It reflects the attempt of its author to build a coherent picture of the geological object, or the meteorological situation, of interest.

This is demonstrated in a striking manner by a synthetic sedimentological example constructed by O. Serra, a pioneer in the geological interpretation of well logs. He considered a regular array of wells (the favorable case) and assigned them sand thickness values, without any special design, in fact using only the round numbers 0, 10, 20, 30. From this data set he derived four very different isopach maps. Figure 0.1a pictures the sand body as a meandering channel; Figure 0.1b as an infill channel with an abrupt bank to the east; Figure 0.1c as a transgressive sand filling paleo-valleys; and Figure 0.1d as a barrier bar eroded by a tidal channel. Each of these maps reflects a different depositional environment model, which was on the interpreter's mind at the time and guided his hand.

Geostatistical models have no such explanatory goals. They model mathematical objects, a two-dimensional isopach surface, for example, not geological objects. The complex mental process by which a geologist draws one of the above maps can better be described as *pattern recognition* than interpolation. Compared with this complexity, interpolation algorithms look pathetically crude, and this is why geological maps are still drawn by hand. To the geostatistician's comfort, the fact that widely different interpretations are consistent with the same data makes them questionable. For one brilliant interpretation (the correct one) how many "geofantasies" are produced?

Another way to qualify description versus interpretation is to oppose *data-driven* and *model-driven* techniques. Traditionally geostatistics has been data driven rather than model driven: it captures the main structural features from the data, and knowledge of the subject matter does not have much impact beyond the selection of a variogram model. Therefore it cannot discriminate between several plausible interpretations. We can, however, be less demanding and simply require geostatistics to take external knowledge into account, and in particular an interpretation proposed by a physicist or a geologist. The current

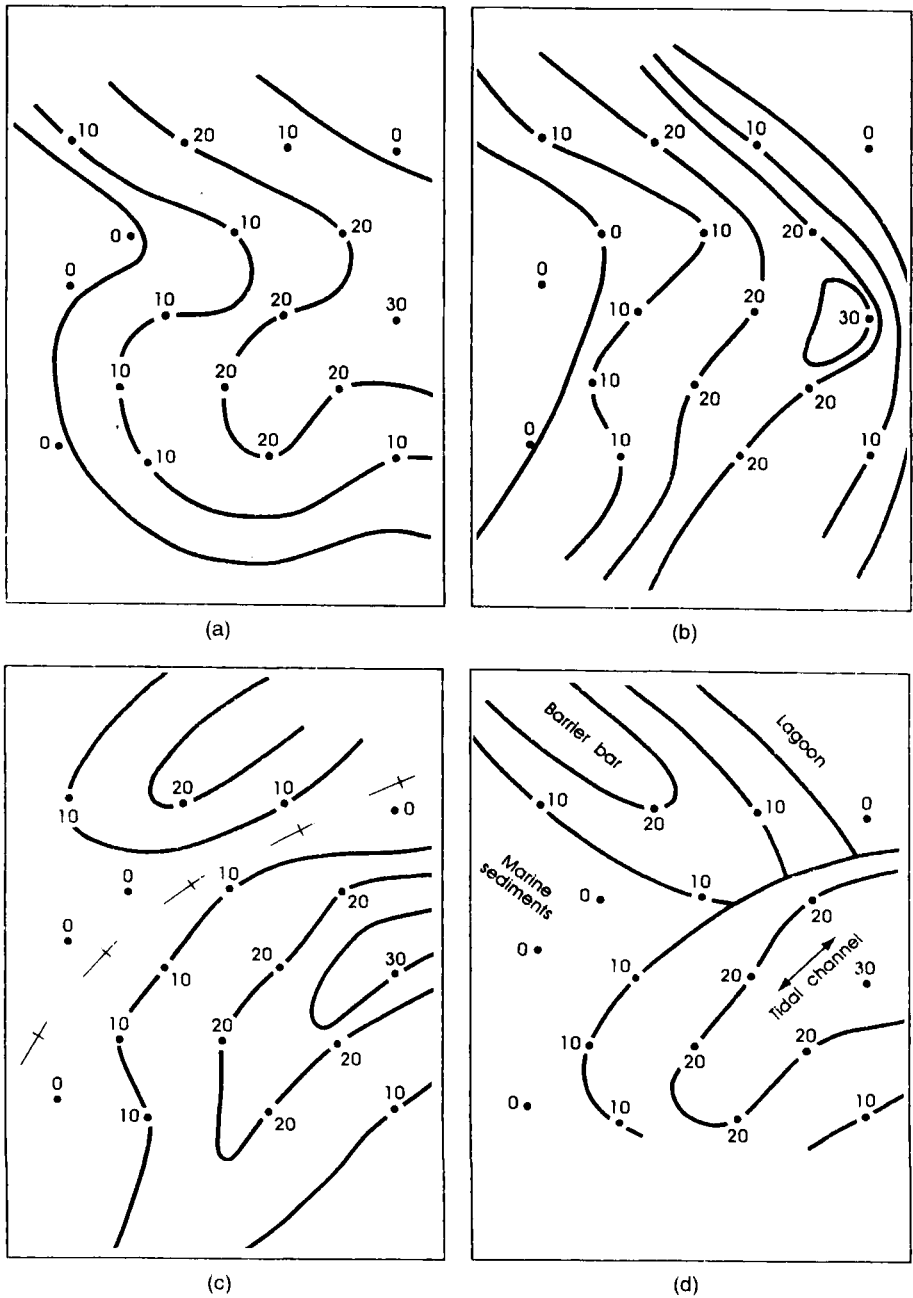


FIGURE 0.1. Four interpretations of the same synthetic data set (hand-drawn isopach maps): (a) meandering channel; (b) infill-channel; (c) transgressive sand filling paleo-valleys; (d) barrier bar eroded by a tidal channel. From O. Serra, personal communication.

trend in geostatistics is precisely an attempt to include physical equations and model-specific constraints.

Hydrogeologists, who sought ways of introducing spatial randomness in aquifer models, have pioneered the research to incorporate physical equations in geostatistical models. Petroleum applications where data are scarce initially have motivated the development of *object-based* models. For example, channel sands are simulated directly as sinusoidal strips with an elliptic or rectangular cross-section. This is still crude, but the goal is clear: import geological concepts to the mapping or simulation processes. We can dream of a system that would find “the best” meandering channel consistent with a set of observations. The difficulty is that the parent genetic models usually remain vastly underdetermined from the observations at hand.

To summarize, the essence of the geostatistical approach is to recognize the inherent variability of natural spatial phenomena and the fragmentary character of our data, and incorporate these notions in a model of a stochastic nature. It identifies the structural relationships in the data and uses them to solve specific problems. It does not attempt any physical or genetic interpretation but uses them as much as possible when they are available.

CHAPTER 1

Preliminaries

1.1. RANDOM FUNCTIONS

G. de Marsily started the defense of his hydrogeology thesis by showing the audience a jar filled with fine sand and announced “here is a porous medium.” Then he shook the jar and announced “and here is another,” shook it again “and yet another.” Indeed at the microscopic scale the geometry is defined by the arrangement of thousands of individual grains with different shapes and dimensions, and it changes as the grains settle differently each time. Yet at the macroscopic scale we tend to regard it as the same porous medium because its physical properties do not change. This is an ingenious illustration of the notion of a random function in three-dimensional space.

Random functions are useful models for regionalized variables.

1.1.1. Definitions

Notations. Throughout this book the condensed notation x is used to denote a point in the n -dimensional space considered. For example, in 3D x stands for the coordinates (x_1, x_2, x_3) (usually called x, y, z). The notation $f(x)$ represents a function of x as well as its value at x . The notation f is used for short, and sometimes the notation $f(\cdot)$ to emphasize that we consider the function taken as a whole and not its value at a single point. Since x is a point in \mathbb{R}^n , dx stands for an element of length ($n = 1$), of surface ($n = 2$), or volume ($n = 3$) and $\int_V f(x)dx$ represents the integral of $f(x)$ over a domain $V \subset \mathbb{R}^n$. For example, if $n = 2$ and V is the rectangle $[a_1, b_1] \times [a_2, b_2]$,

$$\int_V f(x)dx = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} f(x_1, x_2)dx_2$$

We will seldom need an explicit notation for the coordinates of a point so that from now on, except when stated otherwise, x_1, x_2, \dots , will represent distinct points in \mathbb{R}^n rather than the coordinates of a single point.

Coming back to the sand jar, we can describe the porous medium by the indicator function of the grains, namely the function $I(x) = 1$ if the point x (in 3D space) is in a grain and $I(x) = 0$ if x is in a void (the pores). Each experiment (shaking the jar) determines at once a whole function $\{I(x) : x \in V\}$ as opposed to, say, throwing a die that only determines a single value (random variable). In probability theory it is customary to denote the outcome of an experiment by the letter ω and the set of all elementary outcomes, or events, by Ω . To make the dependence on the experiment explicit, a random variable is denoted by $X(\omega)$, and likewise our random indicator function is $I(x, \omega)$. For a fixed $\omega = \omega_0$, $I(x, \omega_0)$ is an ordinary function of x , called a *realization* (or *sample function*); any particular outcome of the jar shaking experiment is a realization of the random function $I(x, \omega)$. On the other hand, for a fixed point $x = x_0$ the function $I(x_0, \omega)$ is an ordinary random variable. Thus mathematically a random function can be regarded as an infinite family of random variables indexed by x .

We can now give a formal definition of a random function (from Neveu, 1970, some details omitted; see also Appendix A.1):

Random Function. Given a domain $D \subset \mathbb{R}^n$, with a positive volume, and a probability space (Ω, \mathcal{A}, P) , a random function (abbreviation: RF) is a function of two variables $Z(x, \omega)$ such that for each $x \in D$ the section $Z(x, \cdot)$ is a random variable on (Ω, \mathcal{A}, P) . Each of the functions $Z(\cdot, \omega)$ defined on D as the section of the RF at $\omega \in \Omega$ is a realization of the RF. For short the RF is simply denoted by $Z(x)$, and a realization by the lowercase $z(x)$.

In the literature a random function is also called a *stochastic process* when x varies in a 1D space, and can be interpreted as time, and a *random field* when x varies in a space of more than one dimension.

In geostatistics we act *as though* the regionalized variable under study $z(x)$ is a realization of a parent random function $Z(x)$. Most of the time we will not be able to maintain the notational distinction between $Z(x)$ and $z(x)$, and will get away with it by saying that the context should tell what is meant. The same is true for the distinction between an estimator (random) and an estimate (fixed).

Spatial Distribution. A random function is described by its *finite-dimensional distributions*, namely the set of all multidimensional distributions of k -tuples $(Z(x_1), Z(x_2), \dots, Z(x_k))$ for all *finite* values of k and all configurations of the points x_1, x_2, \dots, x_k . For short we will call this the *spatial distribution*.

In theory, the spatial distribution is not sufficient to calculate the probability of events involving an infinite noncountable number of points, such as the following important probabilities:

- $\Pr\{\sup\{Z(x) : x \in V\} < z_0\}$ no point within the domain V exceeds z_0
- $\Pr\{\exists x \in V : Z(x) = 0\}$ zero crossing in domain V
- $\Pr\{\text{every realization of } Z(\cdot) \text{ is continuous over } V\}$

This difficulty is overcome by adding the assumption of separability of the random function. A random function is separable if all probabilities involving a noncountable number of points can be uniquely determined from probabilities on countable sets of points (e.g., all points in \mathbb{R}^n with rational coordinates), and hence from the spatial distribution. A fundamental result established by Doob (1953, sec. 2.2) states that for any random function there always exists a separable random function with the same spatial distribution. In other words, among random functions that are indistinguishable from the point of view of their spatial distribution, we pick and work with the smoothest possible version (see Note 3 in Chapter 2). For completeness let us also mention that tools more powerful than the spatial distribution are required to represent random sets (e.g., Matheron, 1975a) but will not be needed in this book.

Moments. The mean of the RF is the expected value $m(x) = EZ(x)$ of the random variable $Z(x)$ at the point x . It is also called the *drift* of Z , especially when $m(x)$ varies with location. The (centered) covariance $\sigma(x, y)$ is the covariance of the random variables $Z(x)$ and $Z(y)$:

$$\sigma(x, y) = E[Z(x) - m(x)][Z(y) - m(y)]$$

In general, this function depends on both x and y . When $x = y$, $\sigma(x, x) = \text{Var}Z(x)$ is the variance of $Z(x)$. Higher-order moments can be defined similarly.

Naturally, in theory, these moments may not exist. As usual in probability theory the mean is defined only if $E|Z(x)| < \infty$. If $E[Z(x)]^2$ is finite at every point, $Z(x)$ is said to be a second-order random function: it has a finite variance, and the covariance exists everywhere.

Convergence in the Mean Square. A sequence of random variables X_n is said to converge in the mean square (m.s.) sense to a random variable X if

$$\lim_{n \rightarrow \infty} E|X_n - X|^2 = 0$$

Taking $X_n = Z(x_n)$ and $X = Z(x)$, we say that an RF $Z(x)$ on \mathbb{R}^n is m.s. continuous if $x_n \rightarrow x$ in \mathbb{R}^n implies that $Z(x_n) \rightarrow Z(x)$ in the mean square. This definition generalizes the continuity of ordinary functions.

1.1.2. Hilbert Space of Random Variables

It is interesting to cast the study of random functions in the geometric framework of Hilbert spaces. To this end consider for maximum generality a family of complex-valued random variables X defined on a probability space (Ω, \mathcal{A}, P) and having finite second-order moments

$$E|X|^2 = \int |X(\omega)|^2 P(d\omega) < \infty$$

These random variables constitute a vector space denoted $L^2(\Omega, \mathcal{A}, P)$ which can be equipped with the scalar product $\langle X, Y \rangle = E X \bar{Y}$ defining a norm¹ (or distance) $\|X\| = \sqrt{E|\bar{X}|^2}$ (the upper bar denotes complex conjugation). In this sense we can say that two random variables are *orthogonal* when they are *uncorrelated*. Then $L^2(\Omega, \mathcal{A}, P)$ is a Hilbert space (every Cauchy sequence converges for the norm). An example is the infinite-dimensional Hilbert space of random variables $\{Z(x) : x \in D\}$ defined by the RF Z .

A fundamental property of a Hilbert space is the possibility of defining the orthogonal projection of X onto a closed linear subspace K as the unique point X_0 in the subspace nearest to X . This is expressed by the so-called *projection theorem* (e.g., Halmos, 1951):

$$X_0 \in K \text{ achieves } \inf_{Y \in K} \|X - Y\| \Leftrightarrow \langle X - X_0, Y \rangle = 0 \quad \text{for all } Y \in K \quad (1.1)$$

Since $X_0 \in K$, it satisfies $\langle X - X_0, X_0 \rangle = 0$ so that

$$\|X - X_0\|^2 = \|X\|^2 - \|X_0\|^2 \quad (1.2)$$

This approximation property is the mathematical basis of kriging theory.

1.1.3. Conditional Expectation

Consider a pair of random variables (X, Y) , and let $f(y | x)$ be the density of the conditional distribution of Y given that $X = x$. The *conditional expectation of Y given $X = x$* is the mean of that conditional distribution

$$E(Y | X = x) = \int_{-\infty}^{+\infty} y f(y | x) dy$$

$E[Y | X = x] = \phi(x)$ is a function of x only, even though Y appears in the expression. It is also known as the *regression function* of Y on X . When (X, Y) are jointly Gaussian,² this function is a straight line. If the argument of ϕ is the random variable X , ϕ is itself a random variable denoted by $E(Y | X)$. This definition carries over to the case where there are several conditioning variables X_1, \dots, X_N .

It is possible to develop a theory of conditional expectations without reference to conditional distributions, and this is mathematically better and provides more insight. The idea is to find the best approximation of Y by a function of X . Specifically we assume X and Y to have finite means and variances and pose the following problem: Find a function $\phi(X)$ such that $E[Y - \phi(X)]^2$ is a minimum. The solution is the conditional expectation $E(Y | X)$.

This solution is unique (up to an equivalence between random variables) and is *characterized* by the following property:

$$E\{[Y - E(Y | X)]h(X)\} = 0 \quad \text{for all measurable } h(\cdot) \quad (1.3)$$

In words, the error $Y - E(Y | X)$ is uncorrelated³ with any finite-variance random variable of the form $h(X)$. Notice that this is a particular application of the projection formula (1.1).

In particular, when $h(X) \equiv 1$, we get

$$E[E(Y | X)] = E(Y) \quad (1.4)$$

The conditional variance is defined by

$$\text{Var}(Y | X) = E(Y^2 | X) - [E(Y | X)]^2$$

from which we deduce the well-known formula

$$\text{Var}(Y) = \text{Var}[E(Y | X)] + E[\text{Var}(Y | X)] \quad (1.5)$$

The variance about the mean equals the variance due to regression plus the mean variance about regression.

For $h(X) = E(Y | X)$ we have

$$E\{[Y - E(Y | X)]E(Y | X)\} = 0$$

so that

$$\text{Cov}(Y, E(Y | X)) = \text{Var}(E(Y | X))$$

which shows that Y and $E(Y | X)$ are always positively correlated with

$$\rho^2 = \frac{\text{Var}(E(Y | X))}{\text{Var}(Y)} \quad (1.6)$$

From (1.5) the residual variance takes on the familiar form

$$E[\text{Var}(Y | X)] = (1 - \rho^2) \text{Var}(Y) \quad (1.7)$$

(note that here ρ is not the correlation between Y and X but between Y and its regression.)

Further to the unbiasedness property let us mention the property of *conditional unbiasedness*, which we will often invoke in this book in relation to kriging:

$$\phi(X) = E(Y | X) \quad \Rightarrow \quad E(Y | \phi(X)) = \phi(X)$$

The proof follows immediately from the characteristic property (1.3), since

$$E\{[Y - \phi(X)]h(X)\} = 0 \quad \text{for all } h(\cdot)$$

entails that $\phi(X)$ also satisfies

$$E\{[Y - \phi(X)]h(\phi(X))\} = 0 \quad \text{for all } h(\cdot)$$

Some Properties of Conditional Expectation

The following results can be derived directly from the characteristic formula and are valid almost surely (a.s.):

Linearity	$E(aY_1 + bY_2 X) = aE(Y_1 X) + bE(Y_2 X)$
Positivity	$Y \geq 0 \text{ a.s.} \Rightarrow E(Y X) \geq 0 \text{ a.s.}$
Independence	$X \text{ and } Y \text{ are independent} \Rightarrow E(Y X) = E(Y)$
Invariance	$E(Yf(X) X) = f(X)E(Y X)$
Successive projections	$E(Y X_1) = E[E(Y X_1, X_2) X_1]$

1.1.4. Stationary Random Functions

Strict Stationarity. A particular case of great practical importance is when the finite-dimensional distributions are invariant under an arbitrary translation of the points by a vector h :

$$\Pr\{Z(x_1) < z_1, \dots, Z(x_k) < z_k\} = \Pr\{Z(x_1 + h) < z_1, \dots, Z(x_k + h) < z_k\}$$

Such RF is called *stationary*. Physically this means that the phenomenon is homogeneous in space and, so to speak, repeats itself in the whole space. The sand in the jar is a good image of a stationary random function in three dimensions, at least if the sand is well sorted (otherwise, if the jar vibrates, the finer grains will eventually seep to the bottom, creating nonstationarity in the vertical dimension).

Second-Order Stationarity. When the random function is stationary, its moments, if they exist, are obviously invariant under translations. If we consider the first two moments only, we have for points x and $x + h$ of \mathbb{R}^n ,

$$\begin{cases} EZ(x) = m \\ E[Z(x) - m][Z(x + h) - m] = C(h) \end{cases}$$

The mean is constant and the covariance function only depends on the *separation* h . We will see in Section 2.3.2 that a covariance must be a *positive definite* function.

By definition, a random function satisfying the above conditions is *second-order stationary* (or weakly stationary, or wide-sense stationary). In this book, unless specified otherwise, stationarity will always be considered at order 2, and the abbreviation SRF will designate a second-order stationary random function.

An SRF is *isotropic* if its covariance function only depends on the length $|h|$ of the vector h and not on its orientation.

Intrinsic Hypothesis. A milder hypothesis is to assume that for every vector h the increment $Y_h(x) = Z(x+h) - Z(x)$ is an SRF in x . Then $Z(x)$ is called an *intrinsic random function* (abbreviation: IRF) and is characterized by the following relationships:

$$\begin{cases} E[Z(x+h) - Z(x)] = \langle a, h \rangle \\ \text{Var}[Z(x+h) - Z(x)] = 2\gamma(h) \end{cases}$$

$\langle a, h \rangle$ is the *linear drift* of the IRF (drift of the increment) and $\gamma(h)$ is its *variogram* function, studied at length in Chapter 2.

If the linear drift is zero, that is, if the mean is constant, we have the usual form of the intrinsic model

$$\begin{cases} E[Z(x+h) - Z(x)] = 0 \\ E[Z(x+h) - Z(x)]^2 = 2\gamma(h) \end{cases}$$

Gaussian Random Functions. A random function is Gaussian if all its finite-dimensional distributions are multivariate Gaussian. Since a Gaussian distribution is completely defined by its first two moments, knowledge of the mean and the covariance function suffices to determine the spatial distribution of a Gaussian RF. In particular, second-order stationarity is equivalent to full stationarity.

A Gaussian IRF is an IRF whose increments are multivariate Gaussian.

A weaker form of Gaussian behavior is when all *bivariate* distributions of the RF are Gaussian; the RF is then sometimes called *bi-gaussian*. A yet weaker form is when only the marginal distribution of $Z(x)$ is Gaussian. This by no way implies that $Z(x)$ is a Gaussian RF but this leap of faith is sometimes made.

1.1.5. Spectral Representation

The spectral representation of SRFs plays a key role in the analysis of time signals. It states that a stationary signal is a mixture of statistically independent sinusoidal components at different frequencies. These basic harmonic constituents can be identified physically by means of filters

that pass oscillations in a given frequency interval and stop others. This can also be done digitally using the discrete Fourier transform.

In the case of spatial processes the physical meaning of frequency components is generally less clear, but the spectral representation remains a useful theoretical tool, especially for simulations. For generality and in view of future reference we will state the main results in \mathbb{R}^n , which entails some unavoidable mathematical complication.

Theorem. A real, continuous, zero-mean RF defined on \mathbb{R}^n is stationary (of order 2) if and only if it has the spectral representation

$$Z(x) = \int e^{2\pi i \langle u, x \rangle} Y(du) \quad (1.8)$$

for some unique *orthogonal random spectral measure* $Y(du)$ (see Appendix A.1). Here i is the unit pure imaginary number, $u = (u_1, \dots, u_n)$ denotes an n -dimensional frequency vector, du an element of volume in \mathbb{R}^n , $x = (x_1, \dots, x_n)$ a point of \mathbb{R}^n and $\langle u, x \rangle = u_1 x_1 + \dots + u_n x_n$ the scalar product of x and u .

For any Borel sets B and B' of \mathbb{R}^n the measure Y satisfies

$$\begin{cases} E[Y(B)] = 0 \\ E[Y(B)\overline{Y(B')}] = 0 & \text{if } B \cap B' = \emptyset \\ Y(B \cup B') = Y(B) + Y(B') & \text{if } B \cap B' = \emptyset \end{cases}$$

$Z(x)$ being real, we have in addition the symmetry relation $Y(-B) = \overline{Y(B)}$, where $-B$ denotes the symmetric of B with respect to the origin. Note that the random variables associated with disjoint sets B and B' are uncorrelated, hence the name *orthogonal measure*.

Now define $F(B) = E|Y(B)|^2$. F is a positive bounded symmetric measure called the *spectral distribution function*. We have in particular

$$\begin{cases} F(B \cup B') = F(B) + F(B') & \text{if } B \cap B' = \emptyset \\ E[Y(B)\overline{Y(B')}] = F(B \cap B') \end{cases}$$

It follows readily from (1.8) and the symmetry of F that the covariance of $Z(x)$ has the spectral representation

$$C(h) = E[Z(x)\overline{Z(x+h)}] = \int e^{2\pi i \langle u, h \rangle} F(du)$$

For time signals, the power of the RF $Z(x)$, which is the energy dissipated per unit time, is generally proportional to $Z(x)^2$. If the SRF has zero mean, $C(0)$ is equal to $EZ(x)^2$ and plays the role of an average power, and the measure F represents the decomposition of this power into the different frequencies. Note that the integral $\int F(du)$ of the spectral distribution is equal to the total power $C(0)$.

Real Spectral Representation. It is interesting to separate the real and imaginary parts of the spectral measure Y in the form

$$Y(B) = U(B) - iV(B)$$

where U and V are two *real* random measures (notice the $-i$ in the definition of V). From the properties of Y , we can deduce the following properties that will be useful for simulations

$$\left\{ \begin{array}{ll} U(-B) = U(B) & V(-B) = -V(B) \\ EU(B)U(B') = EV(B)V(B') = 0 & \text{if } B \cap B' = B \cap (-B') = \emptyset \\ EU(B)V(B') = 0 & \forall B, B' \\ EU(B)^2 = EV(B)^2 = F(B)/2 & \text{if } \{0\} \notin B \\ E|U(\{0\})|^2 = F(\{0\}) & V(\{0\}) = 0 \end{array} \right. \quad (1.9)$$

Also $Z(x)$ has the representation

$$Z(x) = \int \cos(2\pi\langle u, x \rangle) U(du) + \int \sin(2\pi\langle u, x \rangle) V(du)$$

1.1.6. Ergodicity

Ergodicity is an intimidating concept. The practitioner has heard that the RF should be ergodic, since “this is what makes statistical inference possible,” but is not sure how to check this fact and proceeds anyway, feeling vaguely guilty of having perhaps overlooked something very important. We will attempt here to clarify the issues. In practice, ergodicity is never a problem. When no replication is possible, as with purely spatial phenomena, we can safely choose an ergodic model. If the phenomenon is repeatable, typically time-dependent fields or simulations, averages are computed over the different realizations, and the only issue (more a physical than a mathematical one) is to make sure that we are not mixing essentially different functions.

A detailed discussion of ergodicity can be found in Yaglom (1987, vol. 1, ch. 3), and an analysis of its meaning in the context of unique phenomena in Matheron (1978). We have summarized the most important results so that practitioners can pay their respects to ergodicity once for all, and move on.

Ergodic Property

In order to carefully distinguish a random function from its realizations, we will revert, in this section only, to the full notation $Z(x, \omega)$ where ω is the random event indexing the realization. By definition, a stationary random function $Z(x, \omega)$ is ergodic (in the mean) if the spatial average of $Z(x, \omega)$ over a domain $V \subset \mathbb{R}^n$ converges to the expected value $m = EZ(x, \omega)$ when V tends to infinity:

$$\lim_{V \rightarrow \infty} \frac{1}{|V|} \int_V Z(x, \omega) dx = m \quad (1.10)$$

In this expression the norming factor $|V|$ denotes the volume of the domain V , and the limit is understood, as we will always do, in the mean square sense. In \mathbb{R}^n it is important to specify how V tends to infinity, since we may imagine

that V becomes infinitely long in some directions only, but we exclude this and assume that V grows in *all* directions. For example, V may be the cube $[0, t]^n$ where $t \rightarrow \infty$. Of course the limit does not depend on the particular shape of V .

To gain insight into the meaning of this property, it is interesting to revisit the sand jar a last time and do a little thought experiment. We consider a point x at a fixed location relative to the jar and shake the jar repeatedly, recording each time a 1 if x falls in a grain and a 0 otherwise. From this we can evaluate the mean of $I(x, \omega)$, namely the probability that x is in a grain, which should not depend on x . It is intuitively obvious that we will get the same result if we keep the jar fixed and select the point x at random within the jar, the probability of landing in a grain being equal to the proportion of the space occupied by the grains.

The ergodic property can be extremely important for applications, since it allows the determination of the mean from a *single* realization of the stationary random function, and precisely most of the time we only have one realization to work with. Not all stationary random functions are ergodic. The classic counter example is the RF $Z(x, \omega) \equiv A(\omega)$ whose realizations are constants drawn from the random variable A . Clearly for each realization the space integral (1.10) is equal to the constant level $A(\omega)$ but not to the mean of A . Another more realistic example of nonergodic RF is to consider a family of different stationary and ergodic RFs and select one of them according to the outcome of some random variable A , thus defining the composite RF $Z(x, \omega; A)$. On each realization the space integral converges to the mean $m(a) = E(Z(x, \omega; A) | A = a)$ of the particular RF $Z(x, \omega; a)$, but this is different from the overall mean $E m(A)$. Here we have the most common source of stationary but nonergodic random functions arising in practice. As has been pointed out in the literature, nonergodicity usually means that the random function comprises an artificial union of a number of distinct ergodic stationary functions.

Ergodic Theorem

This theorem states that if $Z(x, \omega)$ is a stationary random function (of order 2) the space integral (1.10) *always* converges to some value $m(\omega)$, but this value in general depends on the realization ω : it is a random variable, not a constant:

$$\lim_{V \rightarrow \infty} \frac{1}{|V|} \int_V Z(x, \omega) dx = m(\omega) \quad (1.11)$$

This result is a direct consequence of the stationarity of $Z(x, \omega)$ and again requires V to grow in all directions. The random variable $m(\omega)$ has mean m and a fluctuation equal to the atom at the origin of the random spectral measure associated with the RF $Z(x, \omega)$:

$$m(\omega) = m + Y(\{0\}, \omega)$$

Since $E|Y(\{0\}, \omega)|^2 = F(\{0\})$, it appears that $Z(x, \omega)$ possesses the ergodic property if and only if its spectral measure F has no atom at the origin.

An equivalent condition, known as Slutsky's ergodic theorem is

$$\lim_{V \rightarrow \infty} \frac{1}{|V|} \int_V C(h) dh = 0 \quad (1.12)$$

This is always satisfied if $C(h) \rightarrow 0$ as $h \rightarrow \infty$, as is usually the case, but the condition is not necessary. The left-hand side of (1.12) represents the variance of the space integral (1.11), and we will revisit this in Section 2.3.5 with the notion of *integral range*.

Ergodicity in the Covariance

In the above we have only considered first-order ergodicity, or ergodicity in the mean. It is also important to be able to determine the covariance from a single realization. This implies second-order ergodicity, or ergodicity in the covariance. To establish the ergodicity of the covariance, the same theory can be applied to the product variable $Q_h(x) = Z(x)Z(x+h)$ considered as a second-order stationary random function of x with h fixed. This involves the stationarity of fourth-order moments. For Gaussian RFs the fourth-order moments depend on the second-order moments, and simple results can be obtained. The analogue of Slutsky's condition for the convergence of covariance estimates is then

$$\lim_{V \rightarrow \infty} \frac{1}{|V|} \int_V [C(h)]^2 dh = 0 \quad (1.13)$$

This condition is more restrictive than (1.12). Its equivalent spectral formulation is that the spectral measure F has no atom *anywhere*. In other words, the covariance has no sinusoidal component. Again the convergence $C(h) \rightarrow 0$ as $h \rightarrow \infty$ suffices to fulfill (1.13), but the proof is only valid for Gaussian RFs.

Now what?

In the case of a unique phenomenon there is no way of knowing if the space integral would have converged to a different value on another realization, since there is, and can be, only one. As will be seen in a moment ergodicity is not an objective property in the sense that it cannot be falsified. Therefore we *choose* to model $Z(x, \omega)$ as an ergodic random function whose mean is the limit of the space integral (1.10). Likewise we take the limit of the sample covariance (a space integral) as the *definition* of the covariance of the parent RF. Any other choice would have no relevance to the situation considered.

Strictly speaking there still is a problem. Recall that in practice, we work in a bounded domain and cannot let it tend to infinity. This is a matter of scale. If the domain is large enough for the integral (1.12) to be small, the mean can

be estimated reliably. But if the variance is still large, due to a slow fall off of the covariance, the estimation of the mean is difficult, and it is preferable to avoid using it at all and only consider increments. This is the justification for using the variogram instead of the covariance. The possibility of statistical inference of the variogram is discussed in Section 2.9.

When dealing with space-time phenomena observed at a fixed set of monitoring stations, we typically consider spatially nonstationary models and compute time-averaged estimates of spatial means and covariances. We are thus treating the data as a collection of (correlated) stationary and ergodic random functions of time (multiple time series). These assumptions have to be checked carefully.

Micro-Ergodicity

As we have noted earlier, it is impossible to extend the domain to infinity. Matheron (1978) introduced the notion of *micro-ergodicity*, concerned with the convergence of space integrals when the domain D remains fixed but the sampling density becomes infinite. This concept is distinct from standard ergodicity. For example, neither the mean nor the variance of a stationary and ergodic RF is micro-ergodic, but the slope of the variogram at the origin is micro-ergodic if the variable is not too smooth (Section 2.9.2). Micro-ergodic parameters represent physically meaningful properties.

1.2. ON THE OBJECTIVITY OF PROBABILISTIC STATEMENTS

What sense does it make to speak of the probability of a unique event? When we are told that “there is a 60% chance of rain tomorrow,” we know the next day if it rains or not, but how can we check that the probability of rain was indeed 60% on that day at a specific place? We can’t. The only probabilistic statement that can be disproved is “there is a zero chance of rain tomorrow”: if it does rain the next day, then clearly the forecast was wrong. The same problem essentially arises for spatial “prediction.” What is the physical meaning of a statement such as “there is a 0.95 probability that the average porosity of this block is between 20% and 25%”? Potentially we could measure the porosity and check if it lies in the interval, but we will never know if the 0.95 was correct. Yet despite their unclear meaning we tend to find probabilistic statements useful in giving us an appreciation of uncertainty.

In reality we establish the credibility of weather forecasts not from a single prediction but over time. Someone with enough motivation could check if out of 100 days associated with a forecast of a 60% chance of rain about 60 days were indeed rainy, and do this for all % chance classes. A successful track record, without proving the correctness of the forecast on any given day, proves at least that it is correct *on the average*. It validates the forecasting methodology.

One may object that since we introduced repetitions we are no longer really dealing with a unique phenomenon. But the distinction between unique and repeatable situations is not as clear-cut as it seems. Strictly speaking, it is impossible to repeat the “same” experiment: they always differ in some aspects; simply we judge those unimportant. On the other hand, even though every petroleum reservoir, every mine, every forest is unique, they all belong to classes of situations, shaly sand reservoirs, copper deposits, or tropical woods that are similar enough to give rise to specific methodologies which over time can be validated objectively. This is “external” objectivity.

The practitioner who is interested in the evaluation of this specific deposit or that specific forest would rather have criteria for “internal” objectivity that are based on those unique situations. If we cannot pin down the meaning of a probabilistic statement on a singular event, then the question becomes which concepts, statements, and parameters have an objective, observable, measurable counterpart in reality? Matheron (1978) devotes a fascinating essay entitled “Estimating and Choosing” to this quest for objectivity. The central idea is this: *the only objective quantities are those that may be calculated from the values of a single realization over a bounded domain D* . Indeed, in the absence of repetitions, the maximum information we can ever get is the complete set of values $\{z(x) : x \in D\}$. Objective quantities are essentially space integrals of functions of $z(x)$, referred to as *regionals*: all the values of $z(x)$ itself, block averages, mean values above thresholds, and so on, along with regional mean, variogram, or histogram. On the contrary, the expected value m , the (true) variogram γ , or the marginal distribution of $Z(x)$ are *conventional* parameters. To emphasize the difference, Matheron says that we *estimate* a regional whose exact value is unknown but nevertheless exists independently of us, namely is potentially observable, but we *choose* the value of a conventional parameter.⁴

These considerations lead to a striking reversal of point of view where regionals cease to be mere estimates of “true” parameters to become the physical reality itself, while their theoretical counterparts turn into conventional parameters. For example, the regional variogram γ_R should not be regarded as the regional version of γ but rather γ as being the theoretical version of γ_R . Likewise the fluctuation variance (in the probabilistic model) of a regional is not indicative of the difficulty of the statistical inference of its expected value but rather of the lack of objective meaning of this parameter.

The objectivity of statements can be defined by two criteria. The stronger one is to regard a statement as objective if it is *decidable*, which means that it can be declared true or false once we know $z(x)$ for all $x \in D$. The weaker form of objectivity is K. Popper’s demarcation criterion for scientific hypotheses: it must be possible to design experiments whose outcomes are liable to *falsify* predictions derived from these hypotheses, that is, events with probabilities (nearly) equal to 0 or 1 (in the model). If such attempts are successful, the hypothesis is falsified. If it withstands testing, we may not conclude that it is true but only that it is corroborated (not refuted).

The statement “ $z(x)$ is a realization of a random function $Z(x)$ ” or even “of a stationary random function” has no objective meaning. Indeed, since D is bounded, it is always possible by periodic repetitions and randomization of the origin to construct a stationary random function having a realization that coincides over D with the observed $z(x)$. Therefore no statistical test can disprove stationarity in general. We *choose* to consider $z(x)$ as a realization of $Z(x)$ over D . It does not mean that this decision is arbitrary—in practice, it is suggested by the spatial homogeneity of the data—but simply that it cannot be refuted. As stated earlier, ergodicity is also not an objective property.

If repetitions are the objective foundation of probabilities and if only regionals are physically meaningful, then in the case of a unique phenomenon the objectivity of our measures of uncertainty must be based on spatial repetitions. These are obtained by moving the configuration involved, for example, a block and its estimating data points, throughout a domain D_0 . Denoting by Z_v the block value and by Z^* its estimator, the estimation variance of such block is interpreted as the spatial average of the squared error $(Z^* - Z_v)^2$ over D_0 . By construction, this variance is *not localized* (i.e., is constant) within D_0 , but neither is the kriging variance calculated in the stationary model (since it only depends on the geometry and on the variogram). Of course the domain D_0 can itself be local and correspond to a homogeneous subzone of the total domain D . However, D_0 should not be too small; otherwise, we will lack repetitions. We are tempted to say that there is a trade-off between objectivity and spatial resolution.⁵

1.3. TRANSITIVE THEORY

To avoid the epistemological problems associated with the uniqueness of phenomena, Matheron (1965) first developed an estimation theory in purely spatial terms which he named *transitive theory*. In this approach the regionalized variable $z(x)$ is deterministic and only assumed to be identically zero outside a bounded domain D ; it represents a so-called transition phenomenon, a spatial equivalent of a transient phenomenon in time. We will focus here on the global estimation problem, namely the evaluation of the integral of $z(x)$ which typically represents the total amount of some resource. The transitive theory can also be developed for local estimation, but it has no advantage over the more elegant random function approach.

In this model randomness is introduced through sampling. The easiest would be to use the classic Monte Carlo method and select N samples randomly and independently, leading to an unbiased estimator with a variance equal to σ^2/N where σ^2 is the spatial variance of $z(x)$. However, systematic sampling is usually more efficient. We will present the theory in this case, mainly for background, but also to justify a neat formula for surface estimation. Transitive theory will not be used elsewhere in the book.

1.3.1. Global Estimation by Systematic Sampling

Consider the estimation of the integral

$$Q = \int z(x) dx$$

which is finite since $z(x)$ is zero outside the domain D . If $z(x)$ is a mineral grade (in g/ton) and if the ore density d is a constant, Qd is the quantity of metal in the deposit; if $z(x)$ is an indicator function, Q is the volume of D . We assume that the domain D is sampled on a rectangular grid that extends as far as needed beyond the boundaries of D .

As usual, we reason in \mathbb{R}^n and denote by a the elementary grid spacing (a_1, a_2, \dots, a_n) and by $|a|$ its volume (i.e., the product $a_1 a_2 \dots a_n$). The origin of the grid, which is one of its points, is denoted by x_0 , and k denotes the set of positive or negative integers $k = (k_1, k_2, \dots, k_n)$. The simplest estimate of Q is

$$Q^*(x_0) = |a| \sum_{k \in \mathbb{Z}^n} z(x_0 + ka)$$

where \mathbb{Z} is the set of relative integers. If we select the origin x_0 at random and uniformly within the parallelepiped $\Pi = [0, a_1] \times [0, a_2] \times \dots \times [0, a_n]$, then Q^* becomes a random variable whose expected value is

$$EQ^* = \frac{1}{|a|} \int_{\Pi} Q^*(x_0) dx_0 = \frac{1}{|a|} \int_{\Pi} dx_0 |a| \sum_k z(x_0 + ka) = \int z(x) dx = Q$$

It is unbiased. Defining the *transitive covariogram* $g(h)$ by

$$g(h) = \int z(x) z(x+h) dx$$

similar calculations show that

$$E(Q^*)^2 = |a| \sum_k g(ka) \quad \text{and} \quad Q^2 = \int g(h) dh$$

so that the variance of the error $Q^* - Q$, or *estimation variance*, is given by the formula

$$E(Q^* - Q)^2 = |a| \sum_k g(ka) - \int g(h) dh \quad (1.14)$$

This estimation variance, denoted $\sigma^2(a)$, appears as the error incurred by approximating the integral $\int g(h) dh$ by a discrete sum over the grid. It decreases as the grid becomes finer and also the function $z(\cdot)$ smoother.

This variance is always nonnegative provided that the covariogram $g(h)$ is modeled as a positive definite function ($g(h)$ is the convolution of $z(x)$ by $z(-x)$). The transitive covariogram plays the role of the covariance in an RF model, and in fact they are related, since the regional

noncentered covariance over D is given by

$$C_R(h) = \frac{1}{K(h)} \int_{D \cap D_{-h}} z(x)z(x+h)dx \quad \text{where} \quad K(h) = |D \cap D_{-h}|$$

so that $g(h) = K(h)C_R(h)$.

An expansion of formula (1.14) as an Euler Mac-Laurin series leads, for small a , to a decomposition of the variance $\sigma^2(a)$ into two terms:

$$\sigma^2(a) = T_1(a) + T_2(a)$$

The first term $T_1(a)$ is related to the behavior of $g(h)$ near the origin; the second one $T_2(a)$ depends on its behavior near the range (the distance b beyond which $g(h)$ becomes identically zero). $T_2(a)$ is the *fluctuating term*, also called *Zitterbewegung* (the German for jittery motion). It is a periodic function of the remainder ε of the integer division b/a and cannot be evaluated from the grid data; since it has a zero mean, it is simply ignored.

The regular term $T_1(a)$ can be approximated from the expansion of $g(h)$. For example, in 2D and for an isotropic covariogram with a linear behavior near the origin, the explicit result is

$$\sigma^2(a) = \sigma^2(a_1, a_2) \simeq -g'(+0) \left[\frac{1}{6} a_1^2 a_2 + 0.0609 a_2^3 \right] \quad a_1 \leq a_2 \quad (1.15)$$

where $g'(+0)$ is the slope of $g(h)$ at the origin and is < 0 .

1.3.2. Estimation of a Surface Area

If $z(x)$ is the indicator function of a geometric object, its covariogram is necessarily linear at the origin. More precisely, for a vector h in a direction α , we have $g(h) = g(0) - D_\alpha |h| + \dots$, where D_α is the “total diameter” in the direction α . If the object is convex, D_α is simply the so-called tangent diameter (or caliper diameter); otherwise, $2D_\alpha$ is the total length of the contour of the object projected orthogonally along the direction α (see Section 2.3.4). Here we consider an object with surface area A and a total diameter D that is approximately the same in all directions. Replacing A by its estimate $A^* = N a_1 a_2$, where N is the number of positive samples, we can express the variance (1.15) in the dimensionless form σ_A^2/A^2 :

$$\frac{\sigma_A^2}{A^2} \simeq \frac{D}{\sqrt{A}} \frac{1}{N^{3/2}} \left[\frac{1}{6} \sqrt{\lambda} + 0.0609 \lambda^{-3/2} \right] \quad \lambda = \frac{a_1}{a_2} \leq 1 \quad (1.16)$$

Note that the variance decreases like $1/N^{3/2}$ rather than $1/N$. We can evaluate D from the contour of the object by counting the number of boundary segments $2N_1$ and $2N_2$ respectively, parallel to a_1 and a_2 , including possible holes in the contour (the total perimeter comprises $2(N_1 + N_2)$ segments)

$$D = N_1 a_1 = N_2 a_2$$

and upon replacement in (1.15) and division by A^2 , we get

$$\frac{\sigma_A^2}{A^2} \simeq \frac{1}{N^2} \left[\frac{1}{6} N_2 + 0.0609 \frac{N_2^2}{N_2} \right] \quad N_2 \leq N_1 \quad (1.17)$$

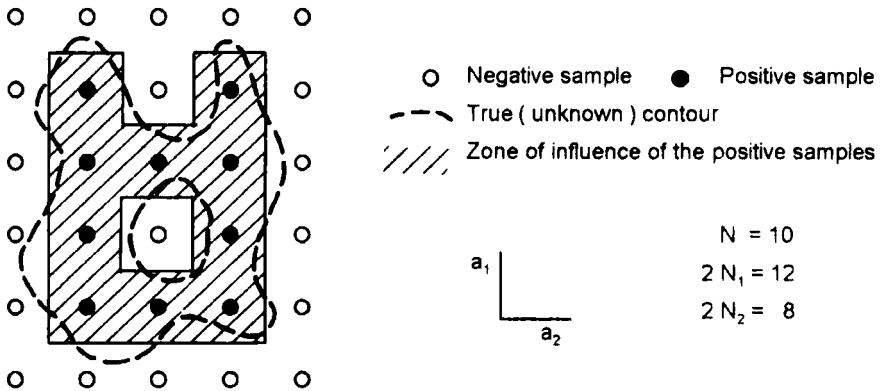


FIGURE 1.1. Estimation of a surface area by systematic sampling.

This formula remains valid if the object is not isotropic but has a main direction of elongation parallel to one of the grid axes—which is the natural orientation for the grid. Indeed we can then restore isotropy, at least approximately, by an affine transformation parallel to one of the grid axes; this changes a_1 or a_2 as well as A and D_1 or D_2 but not N , N_1 , N_2 , or σ_A^2/A^2 which are dimensionless. Thus (1.17) is a simple, self-contained (no calculation and modeling of $g(h)$ needed), and yet theoretically founded formula for evaluating the error in the estimation of a surface area.

To illustrate its use, consider the example shown in Figure 1.1. We read from the figure

$$N = 10 \quad \begin{array}{l} 2N_1 = 12 \\ 2N_2 = 8 \end{array} \quad \text{so that} \quad \frac{\sigma_A^2}{A^2} = \frac{1}{100} \left[\frac{4}{6} + 0.0609 \frac{36}{4} \right] = \frac{1.21}{100}$$

The relative error standard deviation on the surface area is therefore 11%.

An interesting indication can be derived concerning the optimal grid mesh. In case of an isotropic object the variance in (1.16) is minimized for $\lambda = 1$, i.e. a square grid. If the object is not isotropic formula (1.16) can be applied to the transformed figure, where the new grid spacings are for example $a'_1 = a_1$ and $a'_2 = (D_1/D_2)a_2$. In this isotropic case the optimal grid satisfies $a'_1 = a'_2$ so that $a_1/a_2 = D_1/D_2$ or equivalently $N_1 = N_2$. The optimal sampling grid is one that is adapted to the anisotropy of the object. (1.17) also shows that for the optimal grid ($N_1 = N_2$) the estimation variance increases with the perimeter of the object, the minimum variance being achieved for a circle, which for a given surface area has the minimum perimeter.

NOTES

1. Strictly speaking, $\|X\| = 0$ implies that $X = 0$ only up to a set of probability zero, but as usual, equivalence classes of random variables are considered.
2. "Gaussian" and "normal" will be used as synonyms.
3. This does not imply independence between the error and X ; if $X = Y^2$, Y symmetric about 0, $E(Y | X) = 0$, but Y is not independent of X .
4. In the statistical literature, *to predict* means *to estimate* in Matheron's sense and *to estimate* means *to choose*. In this book we will estimate observables in Matheron's sense but fit model parameters rather than choose them.

5. The image of “dithering” comes to mind. This is a binarization technique to transform a half-tone image into a black-and-white image. A gray level is obtained by judiciously distributing black and white dots in the cells of a matrix: a 4×4 matrix allows 16 gray levels, and a 16×16 matrix is required to render 256 gray levels. Thus there is a trade-off between the representation of gray level amplitude and the spatial resolution.

CHAPTER 2

Structural Analysis

Au-delà de l'outil, et à travers lui, c'est la vieille nature que nous retrouvons, celle du jardinier, du navigateur, ou du poète.

Beyond our tools, and through them, it is old mother nature that we reach, an experience that we share with gardeners, sailors, or poets.

Saint-Exupéry

2.1. GENERAL PRINCIPLES

2.1.1. Introduction

The theory of stochastic processes and random functions has been in use for a relatively long time to solve problems of interpolation or filtering. The methods proposed are based on the first two moments of the random functions. In the real world, however, these are never known a priori and must be determined first. One of the strengths of geostatistics is to propose a methodology for the identification of the features of the stochastic model as well as for the interpolation itself. The present chapter is devoted to the analysis of the structural characteristics of spatial data. The spatial phenomena studied here are generally

- unique, nonreproducible,
- defined in a two- or three-dimensional domain,
- too complex for a precise deterministic description,
- known from unevenly distributed sample points.

These phenomena are regionalized variables $\{z(x) : x \in D \subset \mathbb{R}^n\}$. We decide to regard them as realizations of random functions.

A random function (RF) $\{Z(x) : x \in \mathbb{R}^n\}$ is characterized by its finite-dimensional distributions (also called here *spatial distribution* for short), namely the set of all multidimensional distributions of k -tuples $(Z(x_1), Z(x_2), \dots, Z(x_k))$ for all values of k and all configurations of the points x_1, x_2, \dots, x_k . Even if

a very large number of realizations of a random function were available, the combinatorial possibilities are such that, in practice, one could calculate sample multidimensional distributions only for the simplest k -tuples. When a single realization is available, which is the common case, these distributions cannot be determined, except under an assumption of stationarity which introduces repetition in space: two point configurations that are identical up to a translation are considered as statistically equivalent. Since the sample points are unevenly distributed, the only (nearly) identical configurations that can be found are pairs of sample points. A large part of the book is therefore dedicated to methods involving only the knowledge of two-point statistics. The complete knowledge of all these statistics, namely the bivariate distributions, is required for the nonlinear techniques presented in Chapter 6. For linear methods, which are the most widely used, it suffices to know the second-order moments. These are the focus of the present chapter.

The main tool will be the variogram. We will distinguish three main definitions: (1) the variogram of the random function, or theoretical variogram, whose knowledge is required to solve the interpolation problems presented in the next chapter; (2) the variogram of the regionalized variable, or regional variogram, which could be calculated if we knew the value of the regionalized variable at every point of its domain of study; and (3) the sample variogram, which can be calculated from the data. Our task can therefore be split into two phases: (1) compute a sample variogram that best approximates the regional variogram and (2) fit a theoretical model to this sample variogram.

2.1.2. Covariance versus Variogram

We will consider two classes of random functions: stationary random functions and intrinsic random functions. Throughout the book, unless stated otherwise, we consider second-order stationarity.

Covariance of a Stationary Random Function

As we have seen in Section 1.1.4, a stationary random function (SRF) $Z(x)$ is characterized by its mean

$$m = EZ(x)$$

and its covariance function (or covariance for short)

$$C(h) = E[Z(x) - m][Z(x + h) - m] \quad (2.1)$$

A related function is the *correlogram* $\rho(h) = C(h)/C(0)$, which is interpreted as the correlation coefficient between $Z(x)$ and $Z(x + h)$. The covariance and the correlogram show how this correlation evolves with the *separation*, or *lag*,

h . Note that h is a vector. These functions therefore depend both on its length, which is the distance between x and $x + h$, and on its direction. When the covariance depends only on distance, it is said to be *isotropic*. A covariance is an even function, and by the Schwarz inequality it is bounded by its value at the origin (i.e., the variance of the SRF):

$$C(h) = C(-h) \quad |C(h)| \leq C(0)$$

More precisely, we will show that a covariance is a positive definite function. Since the random function has a finite variance, it fluctuates around the mean. Some phenomena do not show this behavior: if we compute the sample mean and variance over increasingly large domains, the sample mean does not stabilize, and the sample variance always increases. This motivates the next model.

Variogram of an Intrinsic Random Function

An intrinsic random function (IRF) is a random function whose *increments* are second-order stationary. It is characterized by its linear drift

$$m(h) = E[Z(x + h) - Z(x)] = \langle a, h \rangle$$

and its variogram

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(x + h) - Z(x)] \quad (2.2)$$

To prove that the drift is linear, start from the obvious relation

$$Z(x + h + h') - Z(x) = [Z(x + h) - Z(x)] + [Z(x + h + h') - Z(x + h)]$$

Passing on to the mathematical expectation gives $m(h + h') = m(h) + m(h')$, which implies that m is a linear function of the vector $h = (h_1, \dots, h_n)'$, namely $m(h) = \langle a, h \rangle = a_1 h_1 + \dots + a_n h_n$ for some gradient vector $a = (a_1, \dots, a_n)'$. From now on, unless explicitly stated otherwise, we will consider that the IRF has no drift, or $m(h) \equiv 0$. The opposite case will be included in the model of universal kriging (see the introduction of Section 3.4).

The variogram shows how the dissimilarity between $Z(x)$ and $Z(x + h)$ evolves with separation h . Like the covariance, it is in general anisotropic. Obviously the variogram is an even, nonnegative function valued 0 at $h = 0$:

$$\gamma(h) = \gamma(-h) \quad \gamma(h) \geq 0 \quad \gamma(0) = 0$$

A variogram also cannot be an arbitrary function. We will show that $-\gamma(h)$ must be a conditionally positive definite function.

The variogram $\gamma(h)$ is sometimes called “theoretical” to remind us that it is a theoretical construct involving neither a particular realization nor a particular

region. Let us mention that $\gamma(h)$ is also called “semivariogram.” However, the term “variogram” tends to become established for its simplicity and can be supported by theoretical arguments (see the definition of the generalized variogram given in Section 4.7.1); it was already used by Jowett (1955a, c), where it represented, however, the graph of the sample or experimental version of $\gamma(h)$.

Bounded Variograms and Stationarity

An SRF is obviously also an IRF and therefore has a variogram. In that case the variogram is linked to the covariance by the relation

$$\gamma(h) = C(0) - C(h) \quad (2.3)$$

Thus the variogram of an SRF is bounded by $2 C(0)$. Equation (2.3) shows that if the covariance is known, the variogram is also known. Conversely, if the variogram of an IRF is bounded by a finite value, $\gamma(h)$ is of the form (2.3) for a stationary covariance $C(h)$ and the IRF only differs from an SRF by a random constant (Matheron, 1973a, pp. 454–457). It is then equivalent to know $\gamma(h)$ or $C(h)$. The variogram therefore has a larger degree of generality than the covariance.

Variogram and Sample Variance

The theoretical variance of $Z(x)$ is equal to $C(0)$ if Z is an SRF, or does not exist (i.e., is infinite) if Z is a nonstationary IRF. But the sample variance of a finite number of values always has a finite expectation. Let us consider a particular realization $z(x)$ of the IRF $Z(x)$. If the N values $\{z(x_\alpha) : \alpha = 1, \dots, N\}$ are known, we can define the sample variance, which characterizes the dispersion of the $z(x_\alpha)$ around their mean, as

$$s^2(0 | N) = \frac{1}{N} \sum_{\alpha=1}^N [z(x_\alpha) - \bar{z}]^2 \quad \text{with} \quad \bar{z} = \frac{1}{N} \sum_{\alpha=1}^N z(x_\alpha)$$

The notation $s^2(0 | N)$ will be generalized in Section 2.8.2 to other dispersion variances. The 0 means that the unit samples are considered as punctual (zero volume), and N recalls that the variance is related to a sampling pattern $\{x_\alpha : \alpha = 1, \dots, N\}$. This variance can be expressed in the form

$$s^2(0 | N) = \frac{1}{2N^2} \sum_{\alpha=1}^N \sum_{\beta=1}^N [z(x_\beta) - z(x_\alpha)]^2 \quad (2.4)$$

Now, if the values $z(x_\alpha)$ are not available, $s^2(0 | N)$ cannot be calculated, but we can obtain its expected value $\sigma^2(0 | N)$ by randomizing (2.4) with respect

to the realization. It is equal to

$$\sigma^2(0 | N) = \frac{1}{N^2} \sum_{\alpha=1}^N \sum_{\beta=1}^N \gamma(x_\beta - x_\alpha) \quad (2.5)$$

In the case of an SRF, this expression is equivalent to

$$\sigma^2(0 | N) = C(0) - \frac{1}{N^2} \sum_{\alpha=1}^N \sum_{\beta=1}^N C(x_\beta - x_\alpha) = C(0) - \text{Var} \left[\frac{1}{N} \sum_{\alpha=1}^N Z(x_\alpha) \right]$$

The expectation of the sample variance is always smaller than the theoretical variance. If the x_α are so distant that the $Z(x_\alpha)$ are mutually uncorrelated, this amounts to

$$\sigma^2(0 | N) = \left(1 - \frac{1}{N} \right) C(0)$$

The discrepancy between the expectation of the sample variance and the theoretical variance becomes negligible as $N \rightarrow \infty$.

Variogram as a Structural Tool

Unless we are considering an SRF with known mean, an exceptional situation, the above gives us two reasons to favor the variogram over the covariance. The first is theoretical: since the class of IRFs include the SRFs, the variogram is a more general tool than the covariance. This is why it was introduced in the 1940s for the study of turbulent flow (e.g., see Kolmogorov, 1941a, b; Obukhov, 1949a, b; and Gandin, 1963, for applications to meteorology).

The second reason is practical: the variogram, unlike the covariance, does not require the knowledge of the mean. In practice, this mean is not known and has to be estimated from the data, which introduces a bias. For example, the covariance at $h = 0$ will be approximated by the sample variance, which is biased downward according to (2.5). This bias cannot be corrected unless the covariance function, or at least the correlation function, is already known, which is not the case (unless the data can be considered uncorrelated, but this is a very special case). A similar bias can also corrupt the behavior of the covariance near the origin: Matheron (1970, ch. 2, exer. 18) shows that in 1D, for a covariance that is linear over the data domain, the slope at the origin of the sample covariance is equal to 4/3 the true slope. Use of the sample covariance when the mean is not known can thus result in erroneous interpretations. The variogram is not affected by these problems, since it automatically filters the mean. It was precisely for this reason that Jowett (1955a) used the sample variogram rather than the covariance. For completeness we should mention that as early as 1926 A. Langsaeter had already used the variogram to characterize the variability of data derived from forest surveys (reported in Matérn, 1960, p. 51).

To summarize, we can say that even if the objective is the covariance, the structural tool is the variogram. This can justify why, following Obukhov (1949a, b), Yaglom (1987) calls the variogram the *structure function*, a phrase to which we give a more general meaning.

2.2. VARIOGRAM CLOUD AND SAMPLE VARIOGRAM

We now change our point of view and turn to the empirical aspects of variogram analysis.

2.2.1. Preliminary: Exploratory Data Analysis

Geostatistical tools do not replace, but complement, the usual statistical tools. Before computing sample variograms and other spatial statistics, it is wise to perform an exploratory data analysis (Tukey, 1977), namely to compute usual univariate and bivariate statistics such as posted maps, histograms, scatterplots, stem-and-leaf plots, and box plots. Multivariate techniques such as cluster analysis or principal component analysis can also be used to reduce multivariate problems to univariate ones. Since this book is devoted to geostatistical methods, we will not develop this point, but the relevance of a geostatistical study depends largely on the quality of this preliminary phase. At this stage most inconsistencies in the data can be detected (gross errors on data coordinates for example), as well as mixing of different populations that should be studied separately (bimodal histogram), or oversampling of particular areas such as the richest parts of an orebody or the most contaminated zones of a plume, which should require the use of declustering techniques. It is not uncommon in a real study that the task of obtaining clean data and clear objectives is as long as the geostatistical study itself. Examples of exploratory analysis abound in the literature; for example, see Webster and Oliver (1990) with applications to soil and land resources, Rossi et al. (1992) with applications to ecology, and Cressie (1991). The geostatistical tools we present now offer the opportunity to go further in the analysis by exploring the spatial relationships between data pairs (e.g., see the applications presented by Bradley and Haslett, 1992).

2.2.2. Variogram Cloud

Let us step for a moment outside the probabilistic context and consider a regionalized variable $\{z(x) : x \in D \subset \mathbb{R}^n\}$, with known values $z_\alpha = z(x_\alpha)$ at N sample points $\{x_\alpha : \alpha = 1, \dots, N\}$. Physical intuition suggests that two points that are close take on close values because these values were generated under similar physical conditions. A geologist would say that the two points have the same “geological environment.” On the contrary, at long distances the genetic conditions are different and greater variations are to be expected. This intuition

of variability with distance can be quantified with the variogram cloud. This tool, first used by Gandin (1963, p. 47) for the study of meteorological fields, and reintroduced and systematically exploited by Chauvet (1982), is a plot of all sample points pairs (α, β) showing

- along the x-axis, distance $r_{\alpha\beta} = |x_\beta - x_\alpha|$;
- along the y-axis, halved squared increment $\frac{1}{2}(z_\beta - z_\alpha)^2$.

Figure 2.1 shows a nice example of this, from Bastin et al. (1984). It concerns rainfall data (Fig. 2.1a), a time-repeated phenomenon. Each point in Figure 2.1b is in fact the mean of 1425 values obtained from 1425 six-hourly observations. When computed from a single realization, the cloud is of course much more dispersed.

The variogram cloud can show different behaviors along the different directions of the separation $h_{\alpha\beta} = x_\beta - x_\alpha$, namely display an anisotropy. This is frequent in 2D, and especially in 3D where vertical variability is rarely of the same nature as horizontal variability (layered media). The cloud is then calculated by classes of direction. The main anisotropy directions are often suspected from geological knowledge, and the variogram cloud computed along these directions. If this is not the case, it is necessary to compute the cloud in several directions to detect a possible anisotropy. In 2D, at least four equally spaced directions are usually considered (the coordinate axes and the diagonals). In 3D, equally spaced directions are obtained by considering regular polyhedra. The icosahedron, with its 20 vertices and 20 equilateral triangle faces, defines the finest regular discretization of the sphere: the lines joining midpoints on opposite edges determine 15 regularly spaced directions.¹

Note that the choice to represent halved squared increments is made by reference to the definition of the variogram of an IRF. It is of course possible, and even recommended, to similarly display other characteristics of the increment $z_\beta - z_\alpha$, for example the increment itself, which makes it possible to check the absence of drift, especially at the border of the domain of study.

2.2.3. Sample Variogram

From the variogram cloud it is possible to extract the following information:

- *The sample variogram.* This is the curve giving the mean of the halved squared increment as a function of distance; in practice, it is calculated by classes of distance, taking in each class the center of gravity of the sample points of the variogram cloud (Fig. 2.1d).
- *Any other characteristic of the cloud.* This is calculated by classes of distance (median, quartiles).
- *Box plots.* These present several characteristics of the variogram cloud in a single figure, usually the mean, the median, and some other quantiles of the halved squared increments for each class of distance (Fig. 2.1c).

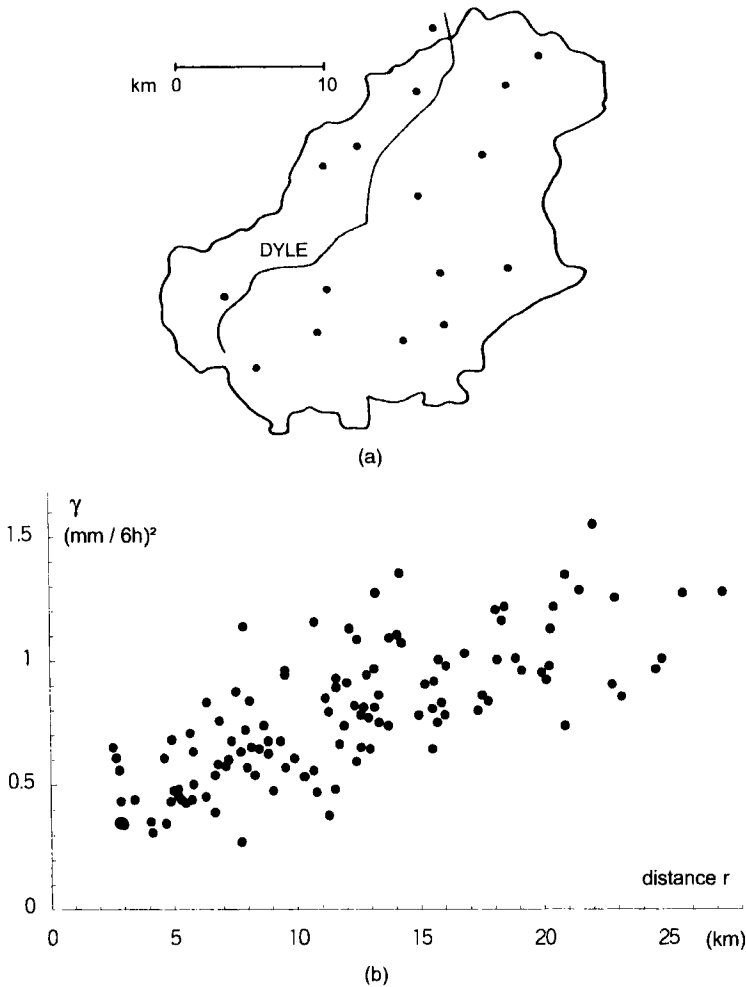


FIGURE 2.1. Variography of rainfall data (Dyle River Basin, Belgium): (a) locations of the rain gauges; (b) variogram cloud (each point represents the mean of 1425 values); (c) box plot showing mean (\times), median (—), quartiles, extreme values, and number of pairs; and (d) sample variogram (circle area proportional to number of pairs). From Bastin et al. (1984), © American Geophysical Union.

Like the variogram cloud, the sample variogram can be anisotropic and is therefore calculated and displayed by classes of direction. The sample variogram can also be defined directly: denoting by N_h the count of pairs of points separated (approximately) by the lag h , it is defined by

$$\hat{\gamma}(h) = \frac{1}{2N_h} \sum_{x_j - x_a \simeq h} [z(x_j) - z(x_a)]^2 \quad (2.6)$$

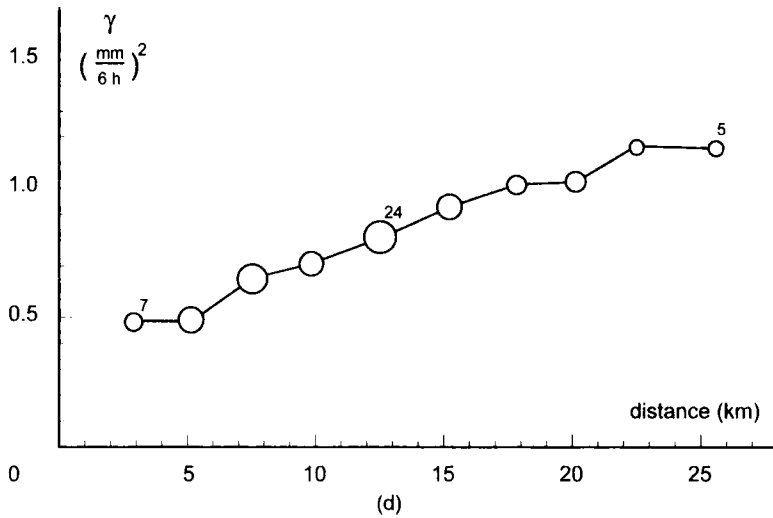
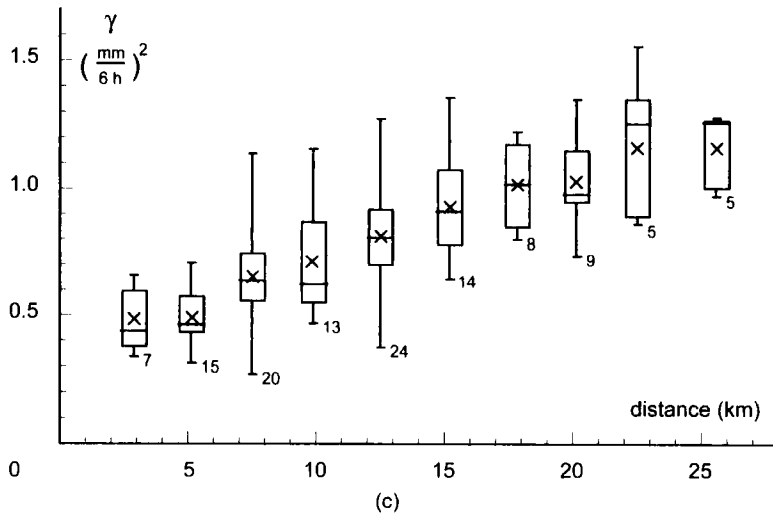


FIGURE 2.1. (Continued).

It is of course calculated for discrete values of h . For example, if the N points x_α are equally spaced in 1D and at a distance Δx , the sample variogram at lag $h = k \Delta x$ is

$$\hat{\gamma}(k \Delta x) = \frac{1}{2(N-k)} \sum_{\alpha=1}^{N-k} [z_{\alpha+k} - z_\alpha]^2$$

This definition can be generalized to data on a 2D or 3D grid, thus producing a sample variogram that is a 2D or 3D array, symmetric with respect to the origin (sometimes named a *variogram map*).

The sample variogram achieves a decomposition of the sample variance into distances. Indeed the average of all terms $\hat{\gamma}(h)$ for all possible lags including the lag $h = 0$, weighted by N_h , is the mean of $\frac{1}{2}(z_\beta - z_\alpha)^2$ for all pairs (α, β) , including those for which $\alpha = \beta$, and it coincides with the sample variance of the data. Spectral analysis achieves a similar decomposition of the variance into different frequencies. There is, however, an important difference: the variogram decomposition, unlike the spectral decomposition, does not produce independent components.

2.2.4. Regional Variogram

When studying a domain D , the ideal sample variogram is that which can be calculated when the domain D is perfectly known. This variogram, named the *regional variogram*, is defined as an areal average of $\frac{1}{2}[z(x+h) - z(x)]^2$ by the formula

$$\gamma_R(h) = \frac{1}{2|D \cap D_{-h}|} \int_{D \cap D_{-h}} [z(x+h) - z(x)]^2 dx \quad (2.7)$$

The actual domain of integration is not D but the intersection of D with its translate by the vector $-h$, denoted $D \cap D_{-h}$. Indeed x and $x+h$ both belong to D if and only if x belongs to $D \cap D_{-h}$. This intersection usually shrinks as the modulus of h increases. $|D \cap D_{-h}|$ is the measure of $D \cap D_{-h}$, namely its length, area, or volume according to the dimensionality of the space. Since in practice only a limited number of sample points are available, one may wonder to what extent the sample variogram is representative of the regional variogram. It is usually considered that a sample variogram value is not reliable if it has been calculated from less than 50 pairs; this is, however, only a broad indication which, to be refined, needs to take into consideration the lag, the locations of the pairs, the shape of the variogram, the histogram of the data, the higher-order moments, and so on. Section 2.9.1 gives further insight into this question.

At this point, the usual question is: Which of the regional variogram and the theoretical variogram matters? From a statistical point of view, the theoretical variogram is more meaningful, since it is a property of the permanent process underlying the observations rather than a feature of incidental sample fluctuations. It is the essential rather than the anecdote. On the other hand, from an engineering point of view, the parent process is an abstraction, and what really matters is the behavior of a specific variable over a specific domain. We will abandon here this quasi-philosophical discussion to simply note that geostatistical practice reconciles these points of view. The sample variograms that we calculate from the data are in fact discrete approximations of regional variograms and thus inherit their physical significance. But the interpretation of these empirical variograms is performed in reference to properties of theoretical variograms which alone can provide a common background to analyze diverse particular situations.

2.2.5. Robust Variograms

As is shown in Section 2.9.1, the sample variogram can give a poor estimate of the regional variogram if the histogram has a long tail. This motivates the search for other variogram estimators.

Robustness and Resistance

Let us consider a set of data all equal to 0, except one that has a value of 1 (Fig. 2.2a). The corresponding sample variogram is of the form

$$\hat{\gamma}(h) = \frac{1}{2} \frac{n_h}{N_h}$$

where N_h is the total number of pairs of points separated by a lag h and n_h is the number of pairs involving the value 1.

This is equal, up to the $\frac{1}{2}$ factor, to the proportion of pairs of points that include the value 1. The variogram is therefore only representative of the geometric configuration of the sample points and of the position of the point valued 1, and it has a rather erratic look (Fig. 2.2b). This may occur, in reality, in two types of situation:

1. When a data point is corrupted by a gross clerical error (relative to the dynamic range of the true data), such as an absent value that has been conventionally set to a high value (typically -9999).
2. When the data do include exceptional values, such as by the presence of a gold nugget in a sample while all the other samples are waste.

In the “ideal” case of a Gaussian RF, the sample variogram is the mean of squared increments that follow a chi-square distribution on one degree of freedom (a gamma distribution with shape parameter $\alpha = \frac{1}{2}$). This distribution has a fairly long tail, and without a large number of squared increments, their sample mean is not a stable estimator of the theoretical mean. Hence the idea of seeking variogram estimators that are more stable than the classic estimator (2.6).

This type of problem is common in statistics and motivated the development of robust and resistant methods (Tukey, 1960; Huber, 1964, 1981; Mosteller and Tukey, 1977). The concept of resistance appeared in the context of exploratory data analysis and does not involve distributional assumptions: a procedure is said to be *resistant* if its result is insensitive to changes, even large ones, in a few of the data (erroneous or anomalous data). For example, the sample median is resistant, whereas the sample mean is not.

Just like an optimal procedure, a *robust* procedure is defined in relation to a statistical model. The optimal procedure is the one that achieves maximum efficiency when the data conform to the model, but it can lead to erroneous results in the opposite case. A robust procedure is slightly less efficient in the

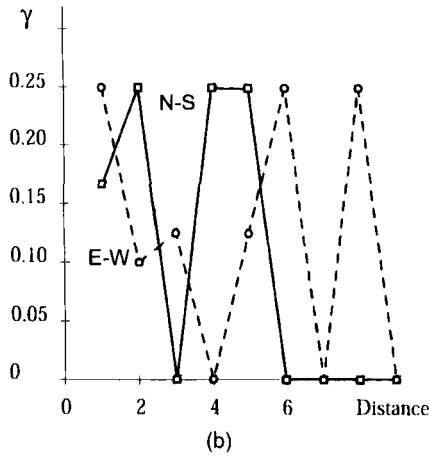
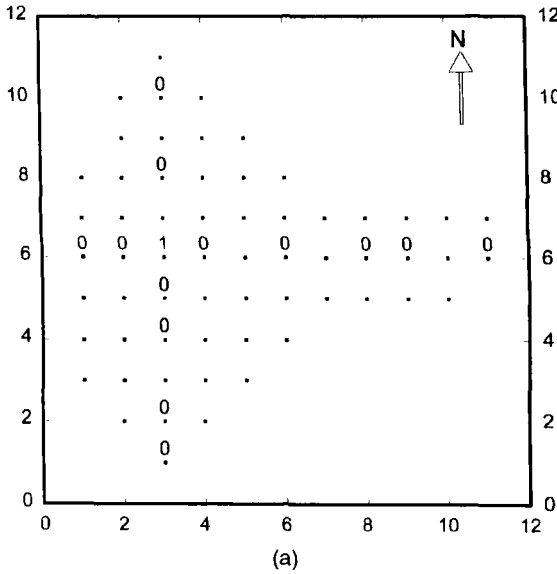


FIGURE 2.2. Synthetic example of an anomaly: (a) data; (b) variogram.

case where the data conform to the ideal model, but it still gives good efficiency when the data depart slightly from this model, and does not give absurd results in case of larger deviations. A classic example is the contamination of a normally distributed batch of “good” observations by a small proportion of “bad” observations, also normally distributed, with the same mean but a much larger variance. The sample median is then a robust estimate of the theoretical mean whereas the sample mean is not. In practice, as pointed out by Huber (1981), the distinction between resistance and robustness is largely arbitrary, and from now on we will only refer to robustness.

Robust techniques are not a substitute for conscientiousness and will not permit a blind processing of any data set. It is always preferable to spot suspect data and deal with them appropriately. Tools as simple as posted maps, histograms, correlation plots, or variogram clouds are excellent for locating suspect data, especially with the availability of software for exploratory analysis of spatial data (Haslett, 1992; Bradley and Haslett, 1992). Data proved to be grossly erroneous must then be corrected or eliminated. However, one must beware of automatically eliminating all extreme data: in many cases (geochemistry, pollution, bathymetry) it is precisely the anomalous data that are important, and if they are inconsistent with the model (stationary, Gaussian, etc.), then it is the model that must be re-examined and not the reality.

Robust Geostatistical Procedures

It is not possible here to give a complete account of all robust variograms proposals. We will only discuss the main principles and a few examples. But first we must point out an ambiguity in the expression “robust variogram” which can denote:

1. an alternative structure function to the variogram, also based on increments but lending itself to a more stable estimation;
2. an estimator of the ordinary variogram $\gamma(h)$, but more stable than the standard estimator $\hat{\gamma}(h)$ defined by (2.6).

Most robust variograms rest on one of, or variations of, the following principles:

- Replacing squared increment by lower-order powers, which are less dispersed and less sensitive to anomalous data.
- Replacing the average of the available pairs (for a given distance class) by their median, which is particularly insensitive to extreme values.
- Eliminating or clipping large increments.

From this point of view a typical robust variogram proposed by Dowd (1984) is the median of the magnitude of increments (for a given distance class). Variants proposed include the quantile variograms (Armstrong and Delfiner, 1980), the “Huberized” variogram (ibid.), and the variogram of order $\frac{1}{2}$ (Cressie and Hawkins, 1980). Other methods have also been studied, in particular, jack-knifing (Chung, 1984). We can also consider the variogram of a transformed variable and the indicator variograms (see Section 2.5.3).

These variograms are better at revealing the possible structure of the phenomenon. But when we are dealing with a skewed distribution, such as that of squared increments, the sample mean and the sample median, say, do not estimate the same parameter. We nevertheless calculate a median variogram, or any other robust variogram, for its good properties, but the variogram thus

obtained is no longer an estimate of the ordinary variogram $\gamma(h)$ required for kriging. Similarly the variogram of an indicator of the variable under study has usually a different behavior than the variogram of the variable itself. Two solutions can be considered:

1. To develop a robust geostatistics involving only the robust variogram: one then leaves the formalism of the L^2 norm (minimization of the quadratic error). This does not lead to tractable ways of constructing estimates and assessing their uncertainties, though solutions exist. For example, a large number of mathematical results are known in the scope of the L^1 norm (minimization of the absolute value of the error), in relation with methods of linear programming (Dantzig, 1963), and particularly in the context of an exponential distribution, which has a larger spread than a Gaussian distribution. But change of support models are no longer available, although necessary for many applications. Hawkins and Cressie (1984) and Dowd (1984) discuss attempts along these lines (cf. Section 3.7.2).
2. To find a means of passing from the robust variogram to the ordinary variogram $\gamma(h)$. This is possible provided that the distribution of the data pairs is specified. Most robust methods include this step under the assumption of a contaminated bi-gaussian RF (one for which all bivariate distributions are Gaussian). These methods estimate the variogram of the non-contaminated variable rather than that of the whole (which in general should show an additional nugget term that needs to be evaluated separately, e.g., from the global variance of the data). Robustness against extreme data is achieved only within fairly restrictive hypotheses outside which bias problems persist.

Median Variogram and Quantile Variograms

The classic variogram estimator (2.6) is

$$\hat{\gamma}(h) = \text{Mean}_{x_j - x_i \simeq h} \left\{ \frac{1}{2} [z(x_j) - z(x_i)]^2 \right\}$$

where Mean represents the mean calculated from N_h pairs of points separated by the lag h . Since the median is more robust than the mean, it is worthwhile considering the *median variogram* taken, for each value of h , as the median of the sample halved squared increments. Armstrong and Delfiner (1980) define the *quantile variogram* $\hat{\gamma}_p(h)$ more generally as

$$\hat{\gamma}_p(h) = Q_p_{x_j - x_i \simeq h} \left\{ \frac{1}{2} [z(x_j) - z(x_i)]^2 \right\}$$

where Q_p denotes the quantile associated with the proportion p ($0 < p < 1$). Calculation of $\hat{\gamma}_p(h)$ simply requires us to sort for each value of h , the available halved squared increments by increasing values so as to construct the cumulative frequency curve.

Assuming that the bivariate distributions of the RF $Z(x)$ associated with the regionalized variable $z(x)$ are Gaussian, $Z(x+h) - Z(x)$ follows a Gaussian distribution with mean 0 and variance $2\gamma(h)$. Therefore $\frac{1}{2}[Z(x+h) - Z(x)]^2$ is distributed as $\gamma(h)\chi_1^2$, where χ_1^2 has a chi-square

distribution on one degree of freedom:

$$\frac{1}{2} [Z(x+h) - Z(x)]^2 \stackrel{D}{=} \gamma(h) \chi_1^2$$

where D represents the equality in distribution. It follows that quantile variograms are all proportional to the variogram $\gamma(h)$, the proportionality factors being the quantiles of the distribution of χ_1^2 ; in particular, we have

$$Q_{0.25}(\chi_1^2) = 0.101 \quad Q_{0.50}(\chi_1^2) = 0.455 \quad Q_{0.75}(\chi_1^2) = 1.324$$

Similar properties are obtained with quantiles defined on halved absolute increments. Still assuming Gaussian bivariate distributions, this gives

$$\frac{1}{2} |Z(x+h) - Z(x)| \stackrel{D}{=} \sqrt{\frac{1}{2} \gamma(h)} |U|$$

where U follows a standard normal distribution. The quantile variograms are thus proportional to $\sqrt{\frac{1}{2} \gamma(h)}$, the proportionality factors being this time the quantiles of the distribution of $|U|$. In particular,

$$Q_{0.25}(|U|) = 0.318 \quad Q_{0.50}(|U|) = 0.674 \quad Q_{0.75}(|U|) = 1.150$$

Improved variogram estimates may be obtained by linear combinations of quantile variograms.

“Huberized” Variogram

Armstrong and Delfiner (1980) have also adapted to the variogram a robust estimator defined by Huber (1964). When calculating the variogram at lag h the halved squared increments are clipped. The threshold is taken proportional to their expected value, namely $\gamma(h)$. More precisely, let (α, β) be a pair such that $x_\beta - x_\alpha = h$: if $\frac{1}{2} [z(x_\beta) - z(x_\alpha)]^2$ exceeds a threshold $c^2 \gamma(h)$, then the threshold value is used for calculating the variogram. Since this results in a downward bias of the variogram, a correction is introduced that reduces the bias in the case of a bi-gaussian RF: the term $\frac{1}{2} [Z(x_\beta) - Z(x_\alpha)]^2 / \gamma(h)$ is then the square of a standard Gaussian, and therefore

$$E[\min(\frac{1}{2} [Z(x_\beta) - Z(x_\alpha)]^2, c^2 \gamma(h))] = f(c) \gamma(h)$$

with

$$f(c) = E[\min(U^2, c^2)]$$

where $U \sim \mathcal{N}(0, 1)$ (standard normal random variable).

For clipping it is necessary to already know $\gamma(h)$. The “Huberized” variogram $\hat{\gamma}_H(h)$ is defined as the value that achieves the equality

$$\text{Mean}_{x_\beta - x_\alpha = h} \{ \min(\frac{1}{2} [z(x_\beta) - z(x_\alpha)]^2, c^2 \hat{\gamma}_H(h)) \} = f(c) \hat{\gamma}_H(h)$$

and is calculated by an iterative method. Convergence is very fast. In practice, one selects a relatively high threshold ($c \geq 2$) that only clips the high values, and one can use, for example,

$$f(2.0) = 0.921 \quad f(2.5) = 0.978 \quad f(3.0) = 0.995$$

Variogram of Order $\frac{1}{2}$

Cressie and Hawkins (1980) have observed that if the RF $Z(x)$ is bi-gaussian, the distribution of the increment of order $\frac{1}{2}$,

$$\frac{1}{2}|Z(x+h) - Z(x)|^{1/2}$$

is close to a Gaussian distribution, and that approximately

$$E[\frac{1}{2}|Z(x+h) - Z(x)|^{1/2}] = A\gamma(h)^{1/4}$$

with

$$A = \frac{\frac{1}{2}\Gamma(\frac{3}{4})}{\pi^{1/2}} = 0.346$$

where Γ is the Euler gamma function. If G_h denotes the mean of the N_h terms $\frac{1}{2}|Z(x_\beta) - Z(x_\alpha)|^{1/2}$ such that $x_\beta - x_\alpha \simeq h$, Cressie and Hawkins find that

$$E[G_h^4] = \frac{1}{8} \left(0.457 + \frac{0.494}{N_h} + \frac{0.045}{N_h^2} \right) \gamma(h)$$

which allows the calculation of an unbiased estimator of $\gamma(h)$. This result is exact if the N_h terms are independent, which they are not. Cressie and Hawkins show that the interdependence among these terms seems to have a negligible impact for problems of practical interest. To help us in such an analysis, the variogram cloud and box plot can be adapted to display increments of order $\frac{1}{2}$ instead of square increments.

Variogram of Fuzzy Data

Let us mention a different case. In certain situations we are faced with qualitative or vague data that cannot be reduced to a single value like *crisp* data (often called *hard data* in the geostatistical literature) nor represented by a (subjective) probability distribution like *uncertain* data. We will call them *imprecise* data. Such data often result from experts' opinion ("this rock has a low porosity") and can be described using the fuzzy set formalism (Zadeh, 1965; Zimmermann, 1985).

Let us consider a space E . An ordinary set A of the space E can be defined by its indicator function $I_A(z)$ equal to 1 when z belongs to A and to 0 in the opposite case: the membership of z to A is in 0 or 1. The fuzzy formalism extends this to partial membership: a fuzzy set A is defined by its membership function $\mu_A(z)$ which assigns to every $z \in E$ a value belonging to the interval $[0, 1]$. A fuzzy number is a fuzzy set of the real line \mathbb{R} with some restrictions (the function μ_A is convex over its support, piecewise continuous, and there exists exactly one $z_0 \in \mathbb{R}$ such that $\mu_A(z_0) = 1$). A typical example is a triangle-shaped function. The experts will translate their opinion into: "Values outside the interval $[0.01, 0.10]$ cannot represent the porosity of that rock, whereas 0.03 corresponds to the maximum membership."

The calculation of sample covariance functions and variograms, as well as kriging, can be adapted to fuzzy numbers by using fuzzy arithmetics. The interested reader is referred to Bardossy et al. (1988), and to Kacewitz (1994) who summarizes fuzzy approaches in geostatistics. Fuzzy sample variograms may be helpful when the number of crisp data is too small to permit the calculation of a workable sample variogram from these data.

2.2.6. Analysis of Heterogeneous Data

Whichever estimator is used, it must be calculated from a homogeneous population. Therefore it is important to treat cases separately:

1. Areas that exhibit different variabilities or are separated by discontinuities.
2. Data measured on different supports (which give neither the same histogram nor the same variogram).
3. Measurements of different qualities (corresponding to measurement errors with different magnitudes).
4. Pairs of points for which the distance is well known and those for which an uncertainty exists.

The variograms obtained in these different cases are of course different, but related and must be analyzed in a consistent manner (see Section 2.4). Similarly one must pay a particular attention to cases of oversampling because they are often in areas of high variability (e.g., one tends to oversample rich areas, which are often more variable). A variogram computed without care will be unusable. Two examples illustrate this point.

Examples

1. A deposit was investigated by E–W profiles. In the northern part the profiles were sampled on a 40-m grid, whereas in the southern part, which is narrower but richer, the grid was tightened to 20 m (Fig. 2.3a). The E–W variograms calculated for each area have the same shape but differ in vertical scaling, which is indicative of a proportional effect (see the next section). If one had merely calculated a global E–W variogram of the deposit from all the profiles, the resultant curve would have been very poorly structured (Fig. 2.3b).

2. A site contaminated by chemicals was investigated using a nearly regular sampling pattern. Additional samples were then taken in the anomalous areas, thus forming clusters of points (Fig. 2.4a). If an ordinary variogram were calculated from all the data, these areas would be over represented, especially at short distances, and since they are areas of high variability, the variogram would be useless (Fig. 2.4b).

Finally in certain cases the Euclidean distance is inappropriate and should be replaced, for example, by a “curvilinear” distance for measurements at sea along a very ragged coastline, or by geological distances “down dip,” “along strike,” and “across strike” for measurements in folded beds (e.g., see Fig. 2.5, and Dagbert et al., 1984).

2.2.7. Physical Interpretation of the Variogram

The graph of the sample variogram $\hat{\gamma}(h)$ against $|h|$, plotted for a given direction of h or for all directions taken together, generally shows the following

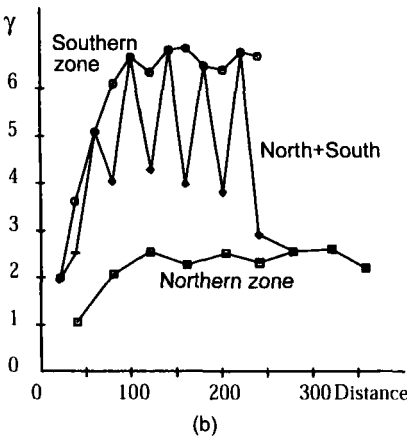
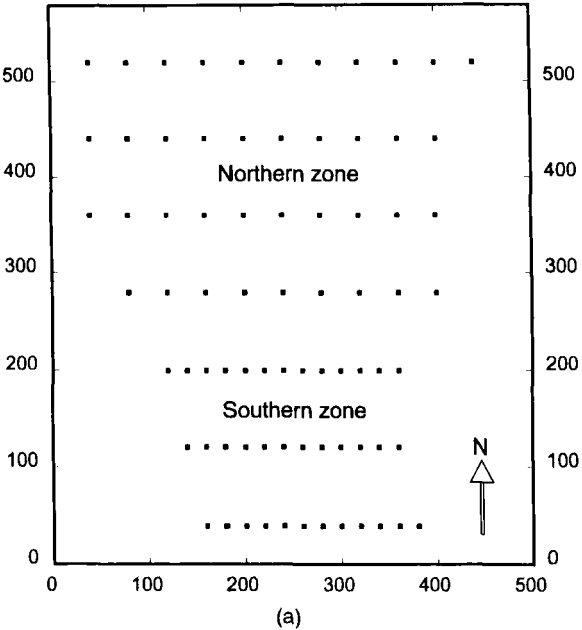


FIGURE 2.3. Effect of refining the sampling grid in the rich zone of a deposit (northern area): (a) data; (b) variograms of the northern and southern areas and variogram of the whole deposit calculated directly, all in the E-W direction.

behavior:

1. It starts at zero (for $h = 0$, $z(x + h) - z(x) = 0$).
2. It increases with $|h|$.
3. It continues to increase, or else stabilizes at a certain level.

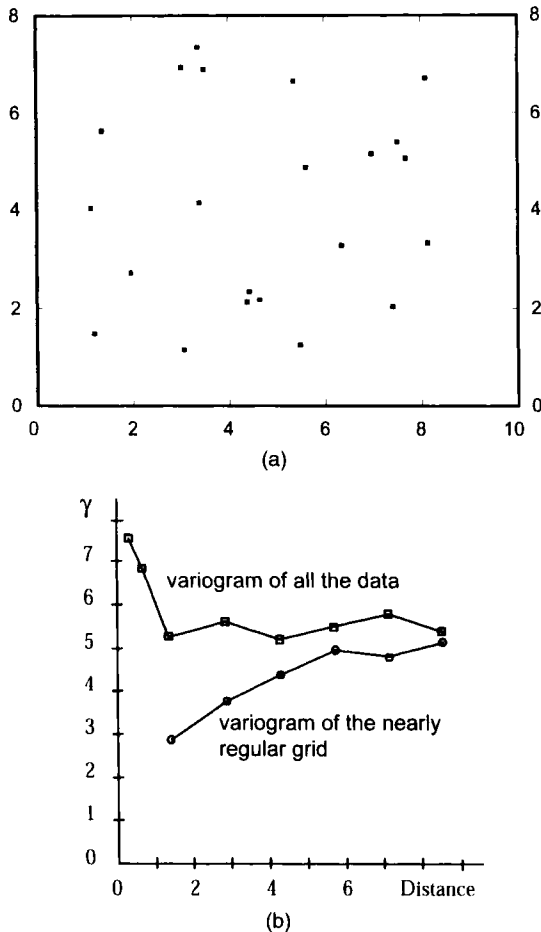


FIGURE 2.4. Effect of oversampling anomalous zones: (a) data; (b) variogram of the regular-grid data and variogram of all the data.

We review below its main features. In practice, the interpretation of a variogram must be done in relation with the contextual knowledge about the variable of interest, for example with geology (e.g., see Rendu, 1984, and the enlightening examples presented by Rendu and Readdy, 1980).

Range and Sill

The rate of the variogram increase reflects the degree of dissimilarity of ever more distant samples. The variogram can increase indefinitely if the variability of the phenomenon has no limit at large distances (no “recall force”; see Fig. 2.6 taken from Krige, 1978, where the variogram is calculated from 3 to 1000 m). If, conversely, the variogram reaches a limiting value, called the *sill*, it means that there is a distance beyond which $Z(x)$ and $Z(x + h)$ are uncorrelated.

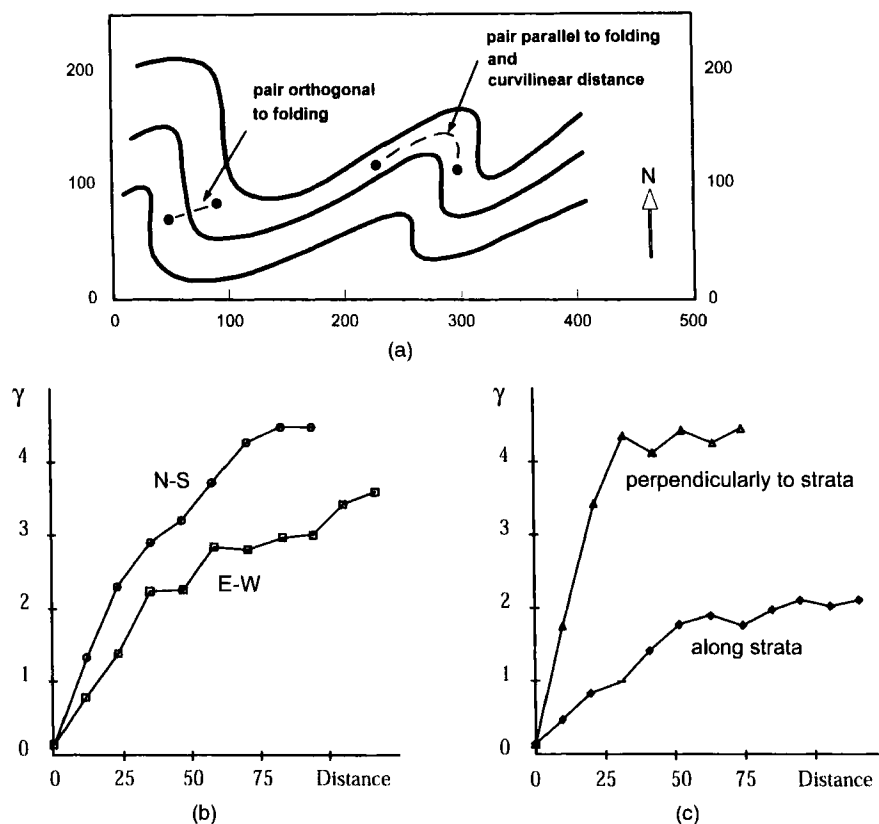


FIGURE 2.5. Calculation of a sample variogram in folded beds: (a) stratification pattern; (b) variogram calculated in the Euclidean system; (c) variogram calculated according to the stratification.

This distance is called the *range* (see Fig. 2.7). Such behavior characterizes what is called a *transition phenomenon* because it is often observed in the case of phenomena where discontinuities, such as bedding joints, lenses, and beds, delimit compartments at the boundaries of which the variability is high. The range then has the same order of magnitude as the compartments. More generally, the range gives an exact sense to the conventional concept of *area of influence* of a sample.

The variogram can reveal *nested structures*, that is to say, hierarchical structures, each characterized by its own range. Serra (1968), in his study of the Lorraine iron basin, has exhibited up to seven nested structures ranging from the petrographic scale, due to oolites, up to the megastructure scale with a range of 10–20 km, passing by decimeter- to meter-size concretions and 100-m-size lenses. Fig. 2.8 shows a simpler example, from Goovaerts et al. (1993), which concerns environmental data (springwater solute contents): the variogram of alkanility which is displayed here reveals

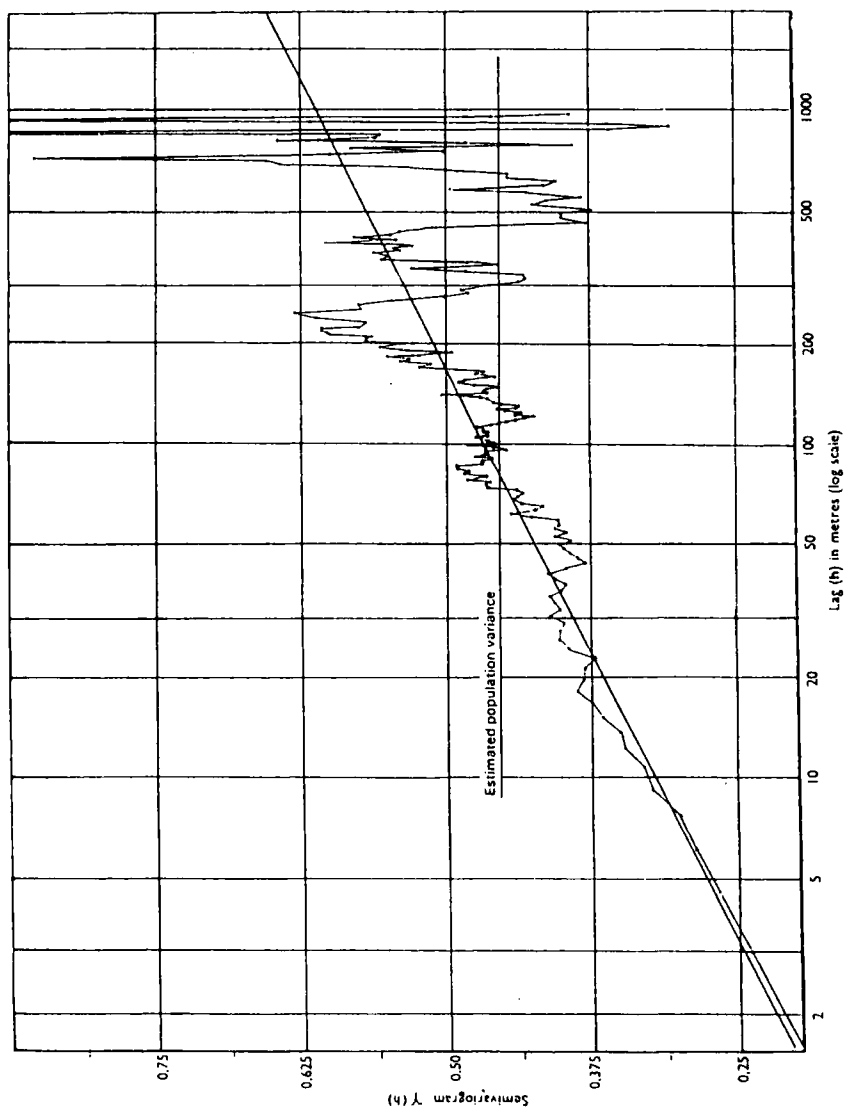


FIGURE 2.6. Variogram with no sill: gold accumulation at the President Steyn mine. Notice the logarithmic scale of the distance axis. From Krige (1978).

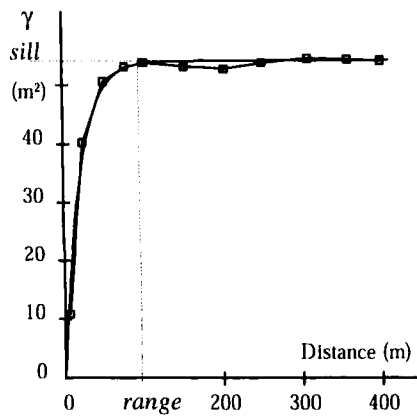


FIGURE 2.7. Range and sill on the variogram of the thickness of nickel-bearing garnieritic ore.

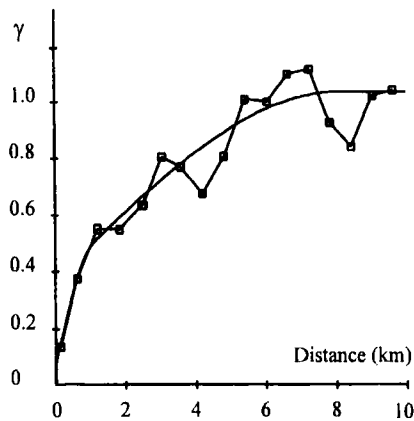


FIGURE 2.8. Nested structure on a variogram of environmental data (alkalinity of springwater, Dyle River Basin, Belgium). From Goovaerts et al. (1993), © American Geophysical Union.

two scales of variation: (1) a short-range component (1 km) corresponding to local sources of contaminants due to human activities, and (2) a long-range component (9 km) interpreted as regional changes in the geologic characteristics of the aquifer. These interpretations make it possible to estimate each component through a factorial kriging analysis (see Section 5.6.6).

Behavior near the Origin and Nugget Effect

Having just examined the variogram behavior at large distances, it is equally interesting to examine its behavior near the origin because this is linked to the

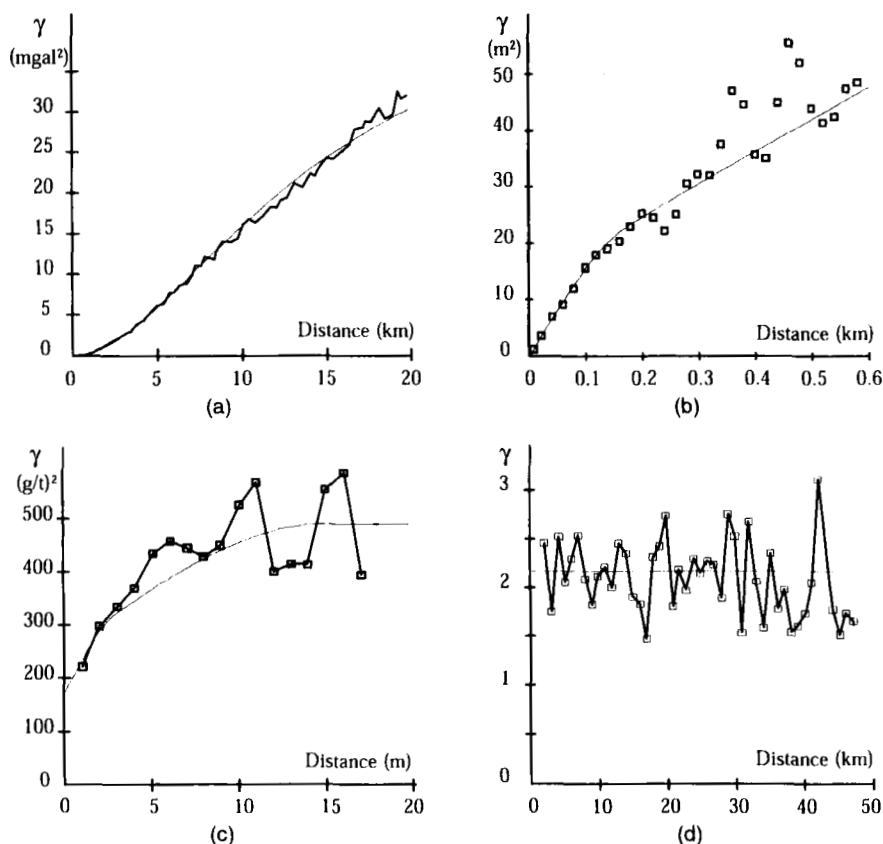


FIGURE 2.9. Examples of variogram behaviors near the origin: (a) parabolic (gravity); (b) linear (depth of a geological formation, Paris Basin); (c) with nugget effect (gold grade, Salsigne); (d) apparent pure nugget effect (logarithm of permeability, Paris Basin). From Chilès and Guillen (1984) (a); B. Bourguine, personal communication (b); Chilès and Liao (1993), with kind permission from Kluwer Academic Publishers (c); J. C. Martin, personal communication (d).

continuity and to the spatial regularity of the regionalized variable. Figure 2.9 shows four typical behaviors:

1. *A parabolic behavior.* This characterizes a highly regular regionalized variable that is usually differentiable at least piecewise. Such behavior, if it persists over large distances, can also be associated with the presence of a strong drift.

2. *A linear behavior.* The regionalized variable is continuous, at least piecewise, but being no longer differentiable, it is less regular than in the previous case.

3. *A discontinuity at 0—nugget effect.* $\hat{\gamma}(h)$ does not seem to tend to zero when $h \rightarrow 0$. This means that the regionalized variable is generally not con-

tinuous and is thus very irregular. The origin of this denomination is as follows: in gold deposits, gold commonly occurs as nuggets of pure metal that are much smaller than the size of a sample. This results in strong grade variability in the samples, even when physically very close and therefore in a discontinuity of the variogram at the origin. By extension, the term “nugget effect” (in the wide sense) is applied to all discontinuities at the origin, even if their cause is different. In general, the nugget effect is due to:

- a microstructure or “geological noise,” namely a component of the phenomenon with a range shorter than the sampling support (true nugget effect);
- a structure with a range shorter than the smallest interpoint distance;
- measurement or positioning errors.

In the absence of close sampling points, it is impossible to tell from the variogram itself which cause is applicable; moreover they can be mixed. Knowledge about the physics of the problem is essential for modeling the variogram. If the model selected is discontinuous at the origin, the corresponding random function is not m.s. continuous.

4. *A flat curve—pure nugget effect or white noise.* There is no correlation between the two points, however close they may be. This is the extreme case of total absence of structure.

From a theoretical point of view, one must qualify the preceding conclusion, for absence of correlation does not necessarily imply independence and absence of structure. For example, if X and Y are two independent random variables valued -1 or $+1$ with equal probability, $X - Y$ and $X + Y$ have the same distribution and can take on the values $-2, 0, +2$. They are uncorrelated but not independent because the possible pairs $(X - Y, X + Y)$ are those with a zero value and a value equal to ± 2 . Figure 2.10 displays simulations with a flat variogram obtained by different methods (Matheron, personal communication). One is pure noise but the other four exhibit patterns. The patchwork image, for example, is obtained by simulating 0–1 noise separately on two orthogonal discrete axes i and j , thus giving two one-dimensional simulations $Y_1(i)$ and $Y_2(j)$, and forming $Y(i, j) \equiv Y_1(i) + Y_2(j) \bmod 2$. The patterning reflects the lack of independence of the $Y(i, j)$ (N^2 values are constructed from $2N$ values). These are, however, curiosities hardly ever found in the usual applications of geostatistics.

Anisotropy

When the variogram does not vary with direction, it is said to be isotropic. It is then a function of the modulus of the vector h , namely of the distance between the points. In the opposite case, the variogram is anisotropic. Two

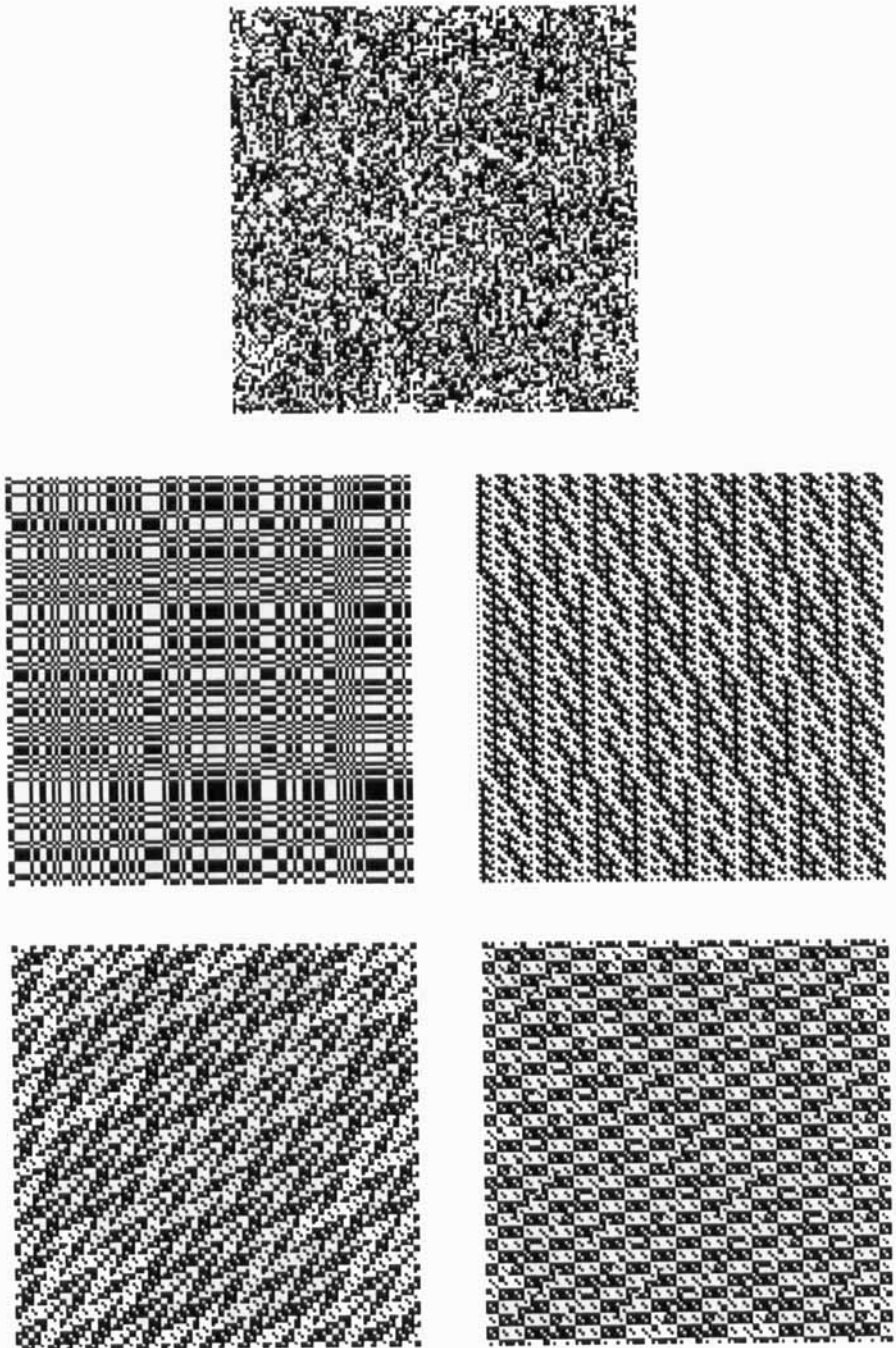


FIGURE 2.10. Realizations of binary random functions with a purely flat variogram. The one shown in the upper image is random noise, but the others exhibit patterning (grid of 128×128 pixels).

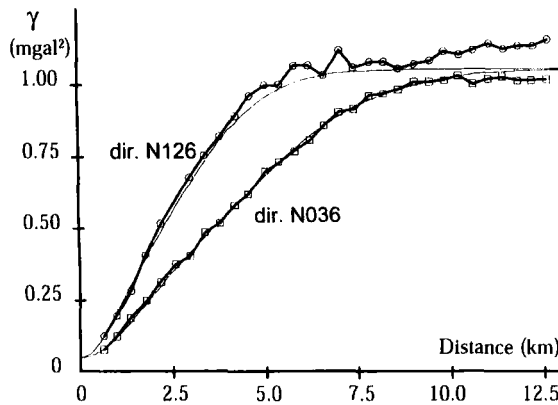


FIGURE 2.11. Geometric anisotropy (gravity residual). From Simard (1980).

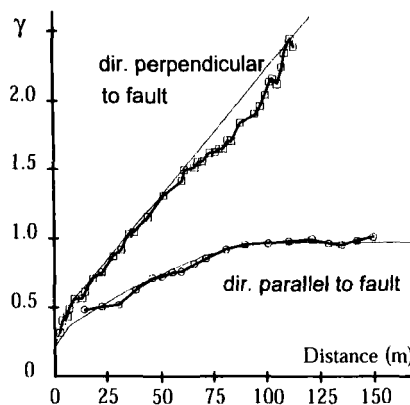


FIGURE 2.12. Zonal anisotropy (variogram of logarithm of gold grade calculated parallel and perpendicular to a fault). From Champigny and Armstrong (1989), with kind permission from Kluwer Academic Publishers.

typical cases of anisotropy are the following:

1. The sill is constant but the range varies with the direction: the variogram of a lenticular formation can show a larger range in the direction of lens elongation than in the other directions. Figure 2.11 shows an example obtained with gravity data (Simard, 1980).
2. The variogram displays a lower sill in a specific direction. Figure 2.12 shows a 2D example where this is due to a fault: the variability is stronger in the direction normal to the fault (Champigny and Armstrong, 1989). In 3D the vertical generally plays a particular role: variations are greater between strata than within a single stratum.

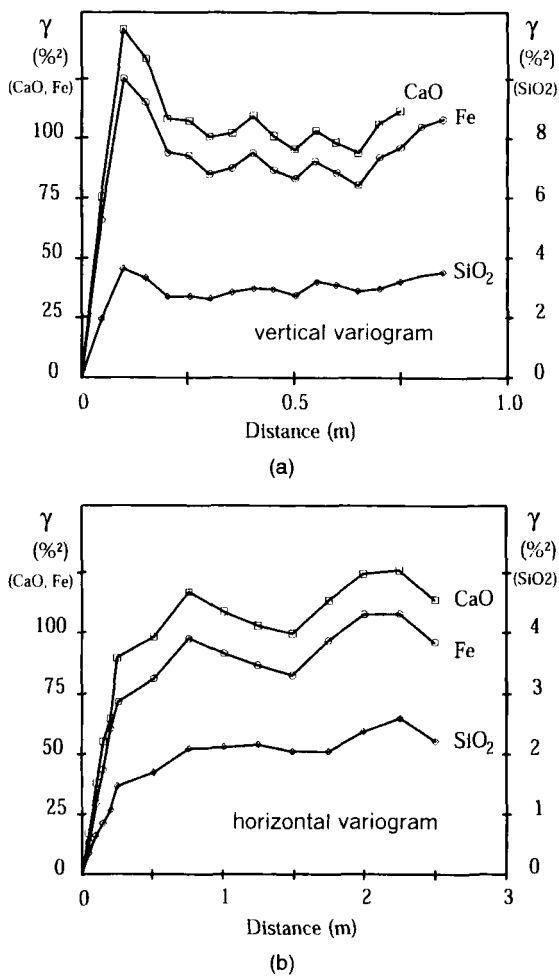


FIGURE 2.13. Variograms of Fe, CaO, and SiO₂ concentrations: (a) vertical variograms; (b) horizontal variograms. The vertical variograms show a hole effect resulting from a migration phenomenon. From Serra (1967).

Hole Effect

The hole effect is characterized by the presence of one or more bumps on the variogram which correspond to an equivalent number of holes on the covariance. It reflects a tendency for high values to be systematically surrounded by low values, and vice versa. One must nevertheless take care not to interpret simple fluctuation of the variogram as a hole effect, and it is advisable to consider the presence of a hole effect only if there is a reasonable physical explanation. Figure 2.13 shows, as an example, variograms of the iron, CaO, and SiO₂ grades obtained by Serra (1967, pp. 85–91) in a study of the

Lorraine oolitic iron ore; the data were reconstructed from 170 cubic samples, with 4-cm sides, collected at 5-cm intervals in vertical channels. The vertical variograms show a very clear hole effect over a distance of 10 cm: it would seem that the migration of carbonates in this moderately reduced ore gave rise to calcite nodules about 10-cm thick (and 50-cm long). The hole effect can also arise from competition between plants although, as noted by Matérn (1960, p. 62), the effect is commonly masked by strong correlations among the soil properties or by too large a sample surface.

Periodicities

A particular case of the hole effect warrants a separate treatment: it is when the variogram shows a periodic, or at least a pseudoperiodic, behavior. Here again one must be certain that this behavior is really significant, especially when studying spatial variables. Time phenomena are frequently periodic due to the influence of the basic daily and yearly cycles on natural phenomena and human activities. Such clear periodicity scarcely exists in space, except in cases where time is involved indirectly. Two examples are given by Matérn (1960): the impact of the annual cycle in sedimentary rocks and soils, and the effect of waves on a beach. One can also cite measurements of magnetism that contain a temporal component that must be separated from the spatial component (Séguret, 1989). Serra (1982, pp. 288–290) gives a good example of pseudoperiodicity at microscopic scale, observed on the binary image of a polypod microfossil (Fig. 2.14). Note also that problems that are by nature clearly oscillatory seem to be better approached by Fourier methods (in the frequency domain) provided that the data are sampled on a regular grid.

Proportional Effect

Variability is sometimes higher in areas with high average values than in areas with low average values. This may corrupt the sample variogram. Typically the sample variogram may reflect more the variations in the average value of the data used for each lag

$$\hat{m}(h) = \frac{1}{2N_h} \sum_{x_j - x_i \approx h} [z(x_j) + z(x_i)]$$

than the spatial variability itself (e.g., see the ecological data studied by Rossi et al., 1992). Then one should calculate local variograms. If all computed variograms look the same and differ only by a multiplicative factor, this is a case of proportional effect; the most common situation is when the sill is proportional to the square of the local mean (the local sample variance is then also proportional to the square of the local mean). This occurs when the regionalized variable has a lognormal histogram.

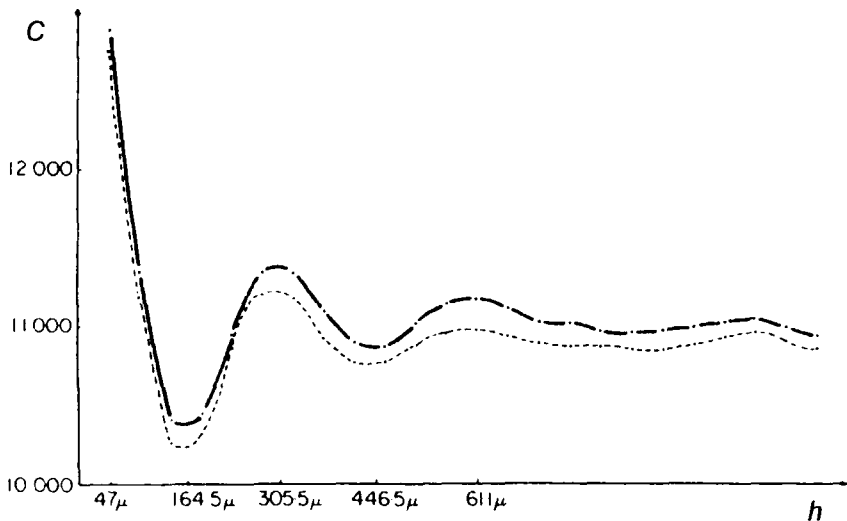


FIGURE 2.14. Pseudoperiodicity on the covariance of the image of a fossil polypod. From Serra (1982), with permission of Academic Press Limited, London.

Presence of a Drift

Theoretically the variogram at large distances should increase more slowly than a parabola (see Section 2.3.1). In practice, however, one can see sample variograms that increase as fast as $|h|^2$, if not faster. The sample variogram may also exhibit very strong and very complex variations with direction. In both cases this indicates the presence of a space-varying mean, or “drift,” representing a trend in the data; for example, the depth of the seafloor increases with distance from the coastline. A variogram of the raw data is therefore of little use: what is needed is the underlying variogram, that of the phenomenon without its drift. Models other than IRFs (universal kriging, IRF- k) and other structural tools (variogram of residuals, generalized variogram) have to be used (see Sections 2.7 and 4.7).

2.3. MATHEMATICAL PROPERTIES OF THE VARIOGRAM

To proceed further in the applications, it is necessary to know the variogram $\gamma(h)$ for any value of h . We cannot use the regional variogram, and even less the sample variogram, instead of the theoretical variogram. Indeed we will show that $-\gamma(h)$ must be a conditionally positive definite function and that the regional variogram as well as the sample variogram do not possess this property in general, even in the case of regularly spaced data. Therefore, to pass from the sample variogram to the theoretical variogram, we turn to theoretical models that are known to be valid variogram functions. We will see in Section 2.6 how to fit a model to a sample variogram. But let us first review

the main mathematical properties of variograms and characterize the class of functions that can represent variograms. Additional results can be found in Yaglom (1987).

2.3.1. Continuity and Differentiability

The degree of regularity of an SRF or IRF is directly related to the behavior of its variogram at the origin.

Continuity

The continuity most suited to second-order models is continuity in the mean square (m.s.): by definition, an RF $Z(\cdot)$ is m.s. continuous at point x if

$$\lim_{h \rightarrow 0} E[Z(x+h) - Z(x)]^2 = 0$$

It follows immediately that an IRF is m.s. continuous everywhere if and only if its variogram is continuous at $h = 0$. And it can easily be shown that if the variogram is continuous at 0, then it is continuous at all $h \in \mathbb{R}^n$.

On the other hand, if the variogram has a discontinuity at the origin, it may also have one elsewhere.² In this section we only consider continuous variograms, this for two reasons:

1. We deliberately ignore the case of RFs whose covariance or variogram has discontinuities elsewhere than at the origin. These RFs have no application in geostatistics.
2. If the variogram of an IRF is continuous everywhere except at the origin, it can be decomposed into a sum of two uncorrelated terms: a pure nugget effect corresponding to random noise, and a continuous variogram corresponding to a m.s. continuous IRF (Matérn, 1960, p. 12).

Since m.s. convergence does not imply almost-sure convergence, the m.s. continuity of an RF does not imply continuity of its realizations. These realizations can, for example, show discontinuities such as faults, and be continuous only in the compartments delimited by these faults. We will see numerous examples in the construction of simulations (e.g., the Poisson process). For separable Gaussian IRFs, however, m.s. continuity implies continuity of almost every realization.³ For more information on this subject, see, for example, Yaglom (1987, vol. 1, p. 65, and vol. 2, ch. 1, n. 12) or Sobczyk (1991, sec. 15).

Differentiability

The notion of differentiability associated with second-order models is the mean square differentiability. Let us first consider the case of an IRF defined on the line. By definition, the random variable $Z'(x)$ is the m.s. derivative of the

random process $Z(\cdot)$ at point x if the finite difference $[Z(x+h) - Z(x)]/h$ converges to $Z'(x)$ in the m.s. sense when $h \rightarrow 0$:

$$\lim_{h \rightarrow 0} E \left[\frac{Z(x+h) - Z(x)}{h} - Z'(x) \right]^2 = 0$$

It is shown in the theory of random processes that $Z'(x)$ exists if and only if the covariance of $[Z(x+h) - Z(x)]/h$ and $[Z(x+h') - Z(x)]/h'$ has a limit when h and h' tend to zero independently.⁴

An immediate consequence is that an IRF $Z(\cdot)$ in \mathbb{R} is m.s. differentiable everywhere if and only if $\gamma(h)$ has a second derivative at 0. Further, for $Z(\cdot)$ having stationary increments, its derivative $Z'(\cdot)$ is a stationary random process and has for covariance the second derivative of the variogram of $Z(\cdot)$:

$$\text{Cov}(Z'(x), Z'(x+h)) = \gamma''(h)$$

Thus, if $\gamma(h)$ has a second derivative at 0, it has a second derivative at all $h \in \mathbb{R}$.

These results are easily extended to the partial m.s. derivatives of an IRF $Z(\cdot)$ in \mathbb{R}^n . If all the partial derivatives exist, $Z(\cdot)$ is said to be m.s. differentiable. As in the continuity case, the m.s. differentiability of an RF does not imply differentiability of its realizations.

Behavior of $\gamma(h)/|h|^2$

The variogram of an SRF is finite. That of a nonstationary IRF can increase to infinity, but not in an uncontrolled fashion, as is expressed by the following two properties (e.g., Yaglom, 1987, vol. 1, pp. 397–400):

1. If $\gamma(h)$ is the variogram of a m.s. continuous IRF, $\gamma(h)/|h|^2 \rightarrow 0$ as $|h| \rightarrow \infty$.
2. If $\gamma(h)$ is the variogram of a m.s. differentiable IRF, $\gamma(h)$ has a majorization of the form

$$\gamma(h) \leq A|h|^2 \quad \forall h \in \mathbb{R}^n$$

2.3.2. Conditional Positive Definiteness

The properties described above are not sufficient to characterize covariance or variogram functions. The only truly necessary condition for a function to be a covariance or a variogram is that all variance calculations lead to a nonnegative result.

Covariance and Positive Definiteness

In the stationary case, let us consider any linear combination $\sum_{i=1}^N \lambda_i Z(x_i)$ of N terms. Its variance, necessarily positive or zero, can be expressed with the

covariance $C(h)$ as

$$\text{Var} \left[\sum_{i=1}^N \lambda_i Z(x_i) \right] = \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j C(x_j - x_i) \quad (2.8)$$

By definition, a function $C(h)$ in \mathbb{R}^n for which the right-hand side of relation (2.8) is always positive (or zero), for any choice of N , x_i , and λ_i , is a *positive definite*⁵ function in \mathbb{R}^n . A covariance is thus necessarily a positive definite function. Conversely, if $C(h)$ is a positive definite function, one can construct a Gaussian SRF with $C(h)$ as its covariance (the knowledge of the covariance function is sufficient to define a consistent set of Gaussian finite-dimensional distributions; see, for example, Doob, 1953, sec. II.3 and XI.3; see also the spectral representation of an SRF in Section 2.3.3). Thus the covariance functions in \mathbb{R}^n and the positive definite functions in \mathbb{R}^n are an identical class.

The family of covariance functions in \mathbb{R}^n satisfies the following *stability properties*:

1. If $C_k(h)$, $k \in \mathbb{N}$, are covariances in \mathbb{R}^n , then $C(h) = \lim_{k \rightarrow \infty} C_k(h)$ is a covariance in \mathbb{R}^n , provided that this limit exists for all h .
2. If $C(h; t)$ is a covariance in \mathbb{R}^n for all values $t \in A \subset \mathbb{R}$ of the parameter t , and if $\mu(dt)$ is a positive measure on A , then $\int C(h; t) \mu(dt)$ is a covariance in \mathbb{R}^n provided that the integral exists for all h .
3. If $C_1(h)$ and $C_2(h)$ are two covariances in \mathbb{R}^n , $C_1(h) \times C_2(h)$ is a covariance in \mathbb{R}^n .

Properties 1 and 2 result from the definition of positive definiteness. Property 3 is obtained by establishing the covariance of the product of two independent SRFs Z_1 and Z_2 with covariances C_1 and C_2 , respectively.

It is useful to know if the covariance $C(h)$ ensures that the variance (2.8) is always strictly positive (except of course when all the λ_i are zero), since the simple kriging system then always has a unique solution. Such a covariance function will be said *strictly positive definite*. An example of a covariance that is not strictly positive definite is the cosine model $C(h) = \cos(\langle \omega, h \rangle)$, where ω is a given vector of \mathbb{R}^n , and more generally any linear combination (with positive coefficients) of a finite number of covariances of this form.

Allowable Linear Combinations

In the stationary case any finite linear combination has a finite variance, given by (2.8). This is no longer true for random functions that have a variogram but no covariance. In the intrinsic case the only linear combinations for which one can calculate the variance are linear combinations of increments. These are called allowable linear combinations and are characterized by the

condition

$$\sum_{i=1}^N \lambda_i = 0 \quad (2.9)$$

It is clear that all linear combinations of increments satisfy equation (2.9), since each increment satisfies it. And conversely, any linear combination $\sum_{i=1}^N \lambda_i Z(x_i)$ satisfying (2.9) is equal to $\sum_{i=1}^N \lambda_i [Z(x_i) - Z(x_0)]$, for any choice of the origin x_0 , and is a linear combination of increments.

Conditional Positive Definiteness

The variance of the allowable linear combination $\sum_{i=1}^N \lambda_i Z(x_i)$ can be expressed in terms of the variogram by

$$\text{Var} \left[\sum_{i=1}^N \lambda_i Z(x_i) \right] = - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(x_j - x_i) \quad (2.10)$$

This result is a particular case of equation (2.11) which is proved below. A function $G(h)$ in \mathbb{R}^n , for which an expression of the form

$$\sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j G(x_j - x_i)$$

is always positive or zero provided that $\sum_{i=1}^N \lambda_i = 0$, is said to be a *conditionally positive definite* function in \mathbb{R}^n . Thus, if $\gamma(h)$ is a variogram, $-\gamma(h)$ is a conditionally positive definite function. This necessary condition is also sufficient provided that $\gamma(0) = 0$ (see Section 2.3.3). Moreover, if $-\gamma(h)$ is a conditionally positive definite function, one can construct a Gaussian IRF with $\gamma(h)$ as its variogram (the reason is similar to that establishing the existence of Gaussian SRFs with a given covariance function).

Stability properties 1 and 2 of covariances carry over to variograms.

Positive Definiteness of Isotropic Functions

An isotropic covariance or variogram is expressed as a function of $r = |h|$. It is important, however, to keep in mind the dimension n of the space, since a positive definite isotropic function (conditionally or not) in \mathbb{R}^n of course satisfies the same property in \mathbb{R}^m for $m < n$ but not necessarily for $m > n$. Thus the triangle function

$$C(r) = \begin{cases} 1 - \frac{r}{a} & \text{if } r \leq a \\ 0 & \text{if } r \geq a \end{cases} \quad (a > 0)$$

is positive definite in \mathbb{R}^1 (see Section 2.3.3) but not in \mathbb{R}^2 , and therefore not in \mathbb{R}^n , $n \geq 2$. Indeed Armstrong and Jabin (1981) exhibit a linear combination in \mathbb{R}^2 for which the application of (2.8) leads to a negative variance.

Covariance of Two Allowable Linear Combinations

The variogram also allows the calculation of the covariance of two allowable linear combinations. In the case of an SRF, the covariance of any two linear combinations $\sum_{i=1}^N \lambda_i Z(x_i)$ and $\sum_{j=1}^{N'} \mu_j Z(x_j)$, where the x_i do not necessarily represent the same points as the x_j , is expressed in terms of the covariance function by

$$\text{Cov} \left(\sum_{i=1}^N \lambda_i Z(x_i), \sum_{j=1}^{N'} \mu_j Z(x_j) \right) = \sum_{i=1}^N \sum_{j=1}^{N'} \lambda_i \mu_j C(x_j - x_i)$$

Similarly in the case of an IRF, but considering this time only allowable linear combinations, this covariance is written in terms of the variogram as

$$\text{Cov} \left(\sum_{i=1}^N \lambda_i Z(x_i), \sum_{j=1}^{N'} \mu_j Z(x_j) \right) = - \sum_{i=1}^N \sum_{j=1}^{N'} \lambda_i \mu_j \gamma(x_j - x_i) \quad (2.11)$$

To see this, it suffices to introduce an arbitrary origin x_0 . Since $\sum_{i=1}^N \lambda_i$ and $\sum_{j=1}^{N'} \mu_j$ are zero, the quantity sought is equal to

$$\sum_{i=1}^N \sum_{j=1}^{N'} \lambda_i \mu_j \text{Cov}(Z(x_i) - Z(x_0), Z(x_j) - Z(x_0))$$

The identity

$$\begin{aligned} (Z_j - Z_i)^2 &= [(Z_j - Z_0) - (Z_i - Z_0)]^2 \\ &= (Z_i - Z_0)^2 - 2(Z_i - Z_0)(Z_j - Z_0) + (Z_j - Z_0)^2 \end{aligned}$$

where Z_i , Z_j , and Z_0 stand for $Z(x_i)$, $Z(x_j)$, and $Z(x_0)$, makes it possible to calculate the product $(Z_i - Z_0)(Z_j - Z_0)$ from squared increments, and thus the covariance from the variogram

$$\text{Cov}(Z(x_i) - Z(x_0), Z(x_j) - Z(x_0)) = \gamma(x_i - x_0) + \gamma(x_j - x_0) - \gamma(x_j - x_i)$$

Inserting this into the double sum yields (2.11).

The restriction to allowable linear combinations is not a problem from the point of view of usual geostatistical applications. For example, if the value

$Z(x_0)$ at an unmeasured point x_0 is estimated by the mean \bar{Z} of the N sample points, \bar{Z} is not an allowable linear combination but the estimation error $\bar{Z} - Z(x_0)$ is, and thus its variance can be calculated.

The computational gymnastics to express the variance of a linear combination or the covariance of two linear combinations is generally easier with the covariance function $C(h)$. If this does not exist, and if one is dealing with an IRF, it suffices to replace $C(h)$ by $-\gamma(h)$ in the result provided, of course, that only allowable linear combinations are considered. In the framework of the IRF- k theory, an ordinary IRF is an IRF-0 and $-\gamma(h)$ is a generalized covariance of order 0 (Section 4.5.1).

Stochastic Integrals

The preceding results can be extended to stochastic integrals. The stochastic integral of the SRF $Z(x)$ associated with a numerical weighting function $w(x)$ in \mathbb{R}^n is defined by

$$Z_w = \int w(x)Z(x)dx$$

The conditions of its existence are given by the following theorem (e.g., Yaglom, 1987, vol. 1, pp. 67–69): The stochastic integral Z_w exists, in the mean square sense, if and only if the integral $\iint w(x)C(x' - x)w(x')dx dx'$ is finite, and this integral is then the (finite) variance⁶ of Z_w :

$$\text{Var}(Z_w) = \iint w(x)C(x' - x)w(x')dx dx' \quad (2.12)$$

If the functions $w_1(x)$ and $w_2(x)$ define two stochastic integrals Z_{w_1} and Z_{w_2} , they have the covariance

$$\text{Cov}(Z_{w_1}, Z_{w_2}) = \iint w_1(x)C(x' - x)w_2(x')dx dx' \quad (2.13)$$

The definition of stochastic integrals can be extended to the case where $Z(x)$ is an IRF provided that we only consider allowable weighting functions, namely satisfying $\int w(x)dx = 0$. The stochastic integral Z_w then exists if and only if the integral $-\iint w(x)\gamma(x' - x)w(x')dx dx'$ is finite, and this integral is then the variance of Z_w . When the weighting functions are allowable and the corresponding stochastic integrals exist, it suffices to replace C by $-\gamma$ in (2.12) and (2.13).

From a practical point of view the double integral (2.12) can be expressed as a simple integral using the change of variable $h = x' - x$ (*Cauchy algorithm*):

$$\iint w(x)C(x' - x)w(x')dx dx' = \int g(h)C(h)dh$$

where $g(h) = \int w(x)w(x+h)dx$ is the covariogram of w , whose properties are examined in Section 2.3.4.

2.3.3. Spectral Representation

Spectral or harmonic analysis is concerned with the decomposition of functions into Fourier series or integrals. It is widely used in the physical sciences. Fourier series apply to periodic functions, whereas Fourier integrals apply to functions that decay to zero rapidly enough at infinity. Many functions, however, belong to neither of these two categories. From this point of view SRFs and IRFs have the advantage over ordinary functions that they always have a spectral representation, which in addition has a clear physical significance. The main results have been established by Kolmogorov (1940a) and Cramér (1942) in \mathbb{R}^1 and by Yaglom (1957) in \mathbb{R}^n . Here we will just state the results on the representation of continuous covariances and variograms, as well as that of the corresponding random functions, and draw a few conclusions. We refer the interested reader to the standard references in this field, notably Doob (1953) in \mathbb{R}^1 , Monin and Yaglom (1965, ch. 6) in \mathbb{R}^3 , and the very complete book of Yaglom (1987).

Spectral Representation of a Covariance and an SRF

We have seen that there is an identity between the class of continuous covariance functions in \mathbb{R}^n and the class of positive definite functions in \mathbb{R}^n . Bochner's theorem (e.g., Feller, 1971, ch. XIX) identifies the characteristic functions $\Phi(u)$ of probability distributions in \mathbb{R}^n with the positive definite continuous functions satisfying $\Phi(0) = 1$. Hence we have the following corollary, which has also been proved directly and independently by Khinchin (1934):

A continuous real function $C(h)$ defined in \mathbb{R}^n is a covariance if and only if it is the (inverse) Fourier transform of a positive bounded symmetric measure $F(du)$

$$C(h) = \int e^{2\pi i \langle u, h \rangle} F(du) = \int \cos(2\pi \langle u, h \rangle) F(du) \quad (2.14)$$

with

$$\int F(du) < \infty \quad (2.15)$$

where u represents a frequency⁷ (i is here the unit pure imaginary number). The integral $\int F(du)$ of the spectral measure is equal to the total power $C(0)$. Note that the covariance is not necessarily a *strictly* positive definite function. A sufficient condition for this is that the support of the spectral measure F is the whole space \mathbb{R}^n (this condition ensures that any stochastic integral, and not only any finite linear combination, has a strictly positive variance).⁸

As we saw in Section 1.1.5 an SRF is also characterized by its spectral representation: a m.s. continuous RF is an SRF (of order 2) if and only if it is

of the form

$$Z(x) = \int e^{2\pi i \langle u, x \rangle} Y(du) \quad (2.16)$$

where Y is an orthogonal complex random measure ($Y(du)$ and $\overline{Y(dv)}$ have zero correlation when du and dv are nonoverlapping) such that

$$E|Y(du)|^2 = F(du) \quad (2.17)$$

The above results are valid for complex random functions. In the case where $Z(x)$ is real-valued, the complex random measure Y satisfies $Y(-du) = \overline{Y(du)}$.

Example 1. If

$$Z(x) = \sum_{j=1}^k [a_j \cos(2\pi \langle u_j, x \rangle) + b_j \sin(2\pi \langle u_j, x \rangle)]$$

where $a_1, \dots, a_k, b_1, \dots, b_k$ are zero-mean uncorrelated random variables such that $E(a_j^2) = E(b_j^2) = \sigma_j^2 > 0$ and u_1, \dots, u_k are distinct vectors, $Z(x)$ is an SRF with covariance $C(h) = \sum_{j=1}^k \sigma_j^2 \cos(2\pi \langle u_j, h \rangle)$. \square

Example 2. If $Z(x) = a\sqrt{2}\cos(2\pi \langle U, x \rangle + \Phi)$, where a , U , and Φ are three independent random variables, a with mean zero and variance $\sigma^2 = C(0)$, U with distribution $F(du)/\sigma^2$, and Φ with a uniform probability density on $[0, 2\pi[$, $Z(x)$ is an SRF whose covariance is given by (2.14). The realizations of this RF with random amplitude, random frequency, and random phase are sinusoids in \mathbb{R}^1 , cylinders with a sinusoidal base in \mathbb{R}^2 , and so on. \square

When $C(h)$ falls off sufficiently rapidly to ensure that $C(h)$ is absolutely integrable in \mathbb{R}^n , namely $\int |C(h)|dh < \infty$, which is the case for usual covariances, the measure F is the integral of a bounded continuous function $f(u)$ called the *spectral density* (or *power spectrum* abbreviated as *spectrum*)

$$F(du) = f(u)du$$

This spectral density is then the Fourier transform (in the sense of ordinary functions) of the covariance

$$\begin{aligned} C(h) &= \int e^{2\pi i \langle u, h \rangle} f(u)du = \int \cos(2\pi \langle u, h \rangle) f(u)du \\ f(u) &= \int e^{-2\pi i \langle u, h \rangle} C(h)dh = \int \cos(2\pi \langle u, h \rangle) C(h)dh \end{aligned} \quad (2.18)$$

To check that a given absolutely integrable function $C(h)$ is a covariance, it suffices to calculate its Fourier transform and to verify that it is always nonnegative. Thus in \mathbb{R}^1 the triangle function

$$C(h) = \begin{cases} 1 - \frac{|h|}{a} & \text{if } |h| \leq a \\ 0 & \text{if } |h| \geq a \end{cases} \quad (a > 0) \quad (2.19)$$

is a covariance because its Fourier transform is

$$f(u) = a \left(\frac{\sin(\pi au)}{\pi au} \right)^2$$

Conversely, as we have seen earlier this is not a positive definite function in \mathbb{R}^2 (nor a fortiori in \mathbb{R}^n , $n > 1$).

The function (2.19) is, up to a multiplicative factor, the autoconvolution $C(h) = (w * \tilde{w})(h)$ of the function $w(x) = 1_{|x| \leq a/2}$. The autoconvolution of a square integrable function is always a positive definite function, called a *covariogram* (see the next section).

As a consequence of the stability property 2 applied to the family (2.19), any function of the form

$$C(h) = \int_{|h|}^{\infty} (1 - |h|/t) \mu(dt) \quad (2.20)$$

where μ is a bounded positive measure on \mathbb{R}_+ , is a covariance in \mathbb{R}^1 . This formula generates the class of symmetric functions that are convex over $[0, \infty[$, namely satisfy

$$C(\lambda h_1 + (1 - \lambda)h_2) \leq \lambda C(h_1) + (1 - \lambda)C(h_2) \quad h_1, h_2 > 0, \quad 0 < \lambda < 1$$

and tend to zero as $h \rightarrow +\infty$. Therefore any function of this family is a covariance in \mathbb{R}^1 . This result is often referred to as *Pólya's theorem* (Pólya, 1949). For example, $\exp(-|h|/a)$, $a > 0$, is a covariance in \mathbb{R}^1 . We can also obtain this result directly by computing the Fourier transform of this function, which is $f(u) = 2a/(1 + 4\pi^2 a^2 u^2)$. We will see that $\exp(-|h|/a)$ is also a covariance in \mathbb{R}^n for all n .

Spectral Representation of a Variogram and an IRF

The characterization of variograms was established by Schoenberg (1938b) and von Neumann and Schoenberg (1941):

If $\gamma(h)$ is a continuous function in \mathbb{R}^n , satisfying $\gamma(0) = 0$, the following three properties are equivalent:

1. $\gamma(h)$ is a variogram.

2. $e^{-t\gamma(h)}$ is a covariance for all $t > 0$.
3. $\gamma(h)$ is of the form

$$\gamma(h) = \int \frac{1 - \cos(2\pi\langle u, h \rangle)}{4\pi^2|u|^2} \chi(du) + Q(h) \quad (2.21)$$

where $Q(h)$ is a positive quadratic form and χ a positive symmetric measure with no atom at the origin and satisfying

$$\int \frac{\chi(du)}{1 + 4\pi^2|u|^2} < \infty \quad (2.22)$$

This expression is valid for the general case of an IRF with a random linear drift $m(h) = \langle a, h \rangle$ such that $E(a) = 0$ and for a noncentered definition of the variogram $\gamma(h) = \frac{1}{2}E[Z(x+h) - Z(x)]^2$. Consequently it is possible to have a quadratic term $Q(h)$ that does not appear with the usual centered definition of $\gamma(h)$.

Proof. (1) \Rightarrow (2). Let $Z(x)$ be a Gaussian IRF without drift, or with a zero-mean random linear drift, and with the (noncentered) variogram $\gamma(h)$ (such IRFs do exist; see below). The complex RF $X(x) = \exp\{i(Z(x) - Z(0))\sqrt{t}\}$ has for noncentered covariance

$$E[X(x)\overline{X(x+h)}] = E[\exp(-i(Z(x+h) - Z(x))\sqrt{t})] = \exp(-t\gamma(h))$$

$e^{-t\gamma(h)}$ is thus a covariance function for all $t > 0$.

(2) \Rightarrow (3). If one considers $e^{-t\gamma(u)}$ to be a characteristic function, it is the characteristic function of an infinitely divisible distribution. The general form of such characteristic functions is given by P. Lévy's theorem (e.g., Feller, 1971, ch. XVII). Considering only real-valued functions γ , this theorem states (3).

(3) \Rightarrow (1). When $\gamma(h)$ is of the form (2.21), $-\gamma(h)$ is a conditionally positive-definite function (straightforward proof). \square

For example, since $\exp(-t|h|)$ is a covariance in \mathbb{R}^1 (and also in \mathbb{R}^n) for all $t > 0$, the function $\gamma(h) = |h|$ is a variogram.

Comparing equations (2.14) and (2.21), we can define the spectral measure of a variogram as $F(du) \equiv \chi(du)/(4\pi^2|u|^2)$. It is defined in $\mathbb{R}^n - \{0\}$. Condition (2.22) on χ is equivalent to the following two conditions on the spectral measure:

$$\int_{|u| < \varepsilon} |u|^2 F(du) < \infty \quad \int_{|u| > \varepsilon} F(du) < \infty$$

where ε is an arbitrary positive value. One also finds this type of condition for IRF- k , and the reader is referred to Section 4.5.3 for further discussion. Let us simply note that the low frequencies can contain an infinite energy since the condition $\int_{|u| < \varepsilon} F(du) < \infty$ is not required. We are thus dealing with the phenomenon known as an *infrared catastrophe*. When, on the contrary, the integral of the spectral measure is finite, $\gamma(h)$ is always of the form

$C(0) - C(h)$, with $C(h) = \int \cos(2\pi\langle u, h \rangle) F(du)$, so that $F(du)$ is the spectral measure of a covariance. And when the integral of $|u|^2 F(du)$ over the whole space is finite, then $\gamma(h)$ is twice differentiable and $Z(x)$ is a m.s. differentiable IRF.

Pólya's theorem can be extended easily to variograms (see Matheron, 1988). Any positive symmetric function $\{\gamma(h) : h \in \mathbb{R}^1\}$ that satisfies $\gamma(0) = 0$ and is concave over $[0, \infty[$, namely is such that

$$\gamma(\lambda h_1 + (1 - \lambda)h_2) \geq \lambda \gamma(h_1) + (1 - \lambda)\gamma(h_2) \quad h_1, h_2 > 0, \quad 0 < \lambda < 1,$$

is a variogram in \mathbb{R}^1 .

The spectral characterization of SRFs extends to the IRFs in \mathbb{R}^1 and \mathbb{R}^n (Kolmogorov, 1940a; Yaglom, 1957): an RF is an IRF if and only if it has the form

$$Z(x) = Z_0 + \int_{\mathbb{R}^n - \{0\}} e^{2\pi i \langle u, x \rangle} Y(du)$$

where Z_0 is a random variable (equal to $Z(0)$) and where Y is an orthogonal complex random measure such that

$$E|Y(du)|^2 = F(du) = \frac{\chi(du)}{4\pi^2|u|^2}$$

Spectral Representation of an Isotropic Covariance or Variogram

Let $C(h)$ be a covariance in \mathbb{R}^n , which for simplicity we presume to be absolutely integrable. The covariance and its spectral density $f(u)$ are then Fourier transforms of each other. The Fourier transform of an isotropic function being itself isotropic, if we assume the covariance is isotropic, namely of the form $C(h) = C_n(r)$ with $r = |h|$, the spectral density is of the form $f(u) = f_n(\rho)$ with $\rho = |u|$. The index n recalls that these functions represent isotropic functions in \mathbb{R}^n , which is important since the Fourier transform \mathcal{F}_n that relates C and f depends on n . Considered as relating two functions of a single variable, C_n and f_n , \mathcal{F}_n represents the Hankel transform of order n , and (2.18) takes on the form

$$\begin{aligned} f_n(\rho) &= 2\pi \rho^{1-n/2} \int_0^\infty r^{n/2} J_{n/2-1}(2\pi \rho r) C_n(r) dr \\ C_n(r) &= 2\pi r^{1-n/2} \int_0^\infty \rho^{n/2} J_{n/2-1}(2\pi \rho r) f_n(\rho) d\rho \end{aligned} \quad (2.23)$$

where J_ν represents the Bessel function of the first kind of order ν (A.2). Despite this apparent symmetry, the two functions are usually not interchangeable: in particular, f_n is a nonnegative function, whereas C_n can take on negative values.

If we concentrate $f(u)$ on the surface of the hypersphere with radius $1/(2\pi)$ and replace $f_n(\rho) d\rho$ in (2.23) by $A \delta_{1/(2\pi)}(d\rho)$ with $A = (4\pi)^{n/2-1} \Gamma(n/2)$, we obtain the *J-Bessel model*

$$\kappa_n(r) = 2^{n/2-1} \Gamma\left(\frac{n}{2}\right) r^{1-n/2} J_{n/2-1}(r)$$

The multiplicative factor A is chosen so that $\kappa_n(0) = 1$. The covariance $\kappa_n(r)$ exhibits a hole effect, this effect being less pronounced as n increases (for $n = \infty$, we obtain the Gaussian model).

The second equation (2.23) expresses the following characterization: a necessary and sufficient condition for a function $C_n(r)$ to be an isotropic covariance in \mathbb{R}^n is to be of the form

$$C_n(r) = \int_0^\infty \kappa_n\left(\frac{r}{t}\right) \mu(dt) \quad (2.24)$$

for a bounded positive measure μ on \mathbb{R}_+ . Therefore no isotropic covariance in \mathbb{R}^n can have a relative hole effect more pronounced than κ_n .

More explicitly, for $n = 1, 2, 3$, the relations between C_n and f_n are

$$\begin{aligned} n = 1: \quad C_1(r) &= 2 \int_0^\infty \cos(2\pi\rho r) f_1(\rho) d\rho \\ n = 2: \quad C_2(r) &= 2\pi \int_0^\infty \rho J_0(2\pi\rho r) f_2(\rho) d\rho \end{aligned} \quad (2.25)$$

$$n = 3: \quad C_3(r) = \frac{2}{r} \int_0^\infty \rho \sin(2\pi\rho r) f_3(\rho) d\rho \quad (2.26)$$

Generally speaking, the formula involves trigonometric functions when n is odd and the somewhat intractable Bessel function of integer order when n is even.

As far as the variogram is concerned, if the measure χ is of the form $\chi(du) = \varphi_n(|u|)du$, relation (2.21) becomes in the isotropic case

$$\gamma_n(r) = \frac{\pi^{n/2-2}}{2\Gamma(n/2)} \int_0^\infty \left[1 - \Gamma\left(\frac{n}{2}\right) (\pi\rho r)^{1-n/2} J_{n/2-1}(2\pi\rho r) \right] \rho^{n-3} \varphi_n(\rho) d\rho$$

Recall that a positive definite function (conditionally or not) in \mathbb{R}^n is not necessarily so in \mathbb{R}^m , $m > n$.

It is convenient to have isotropic models that are covariances in \mathbb{R}^n for any n . An example is the *Gaussian model*

$$C(r) = \exp\left(-\frac{r^2}{a^2}\right) \quad (a > 0)$$

Indeed the Hankel transform of order n exchanges the functions $\exp(-\pi r^2)$ and $\exp(-\pi \rho^2)$. This is a classic result. These functions are, respectively, the density and the characteristic function of the isotropic Gaussian multidimensional distribution with variance $\sigma^2 = 1/(2\pi)$. The Gaussian model is extremely regular, having derivatives of any order.

The following characterization holds in the general case: A necessary and sufficient condition for a function $C(r)$ to be an isotropic covariance in \mathbb{R}^n for all n is that $C(r)$ is of the form

$$C(r) = \int_0^\infty \exp\left(-\frac{r^2}{t^2}\right) \mu(dt) \quad (2.27)$$

where μ is an arbitrary bounded positive measure on \mathbb{R}_+ (randomized scale parameter).

This condition is indeed sufficient, as results from the stability condition (2). It is also necessary, as shown by Schoenberg (1938a).⁹ Note that (2.27) generates *all* types of behaviors near the origin because the base model is infinitely differentiable: randomization produces models that can be *less* regular than the base model but never more regular. For example, randomizing the

scale parameter t in (2.27) by a one-sided Gaussian distribution leads to the *exponential model*¹⁰

$$C(r) = \exp\left(-\frac{r}{a}\right) \quad (a > 0)$$

which is thus a covariance in \mathbb{R}^n for any n . This model is a completely monotone function.¹¹ It constitutes the base model of isotropic covariances with this property, since any completely monotone function that is an isotropic covariance in \mathbb{R}^n for all n is of the form

$$C(r) = \int_0^\infty \exp\left(-\frac{r}{t}\right) \mu(dt) \quad (2.28)$$

for a bounded positive measure μ on \mathbb{R}_+ (e.g., see Schoenberg, 1938a; Feller, 1971, p. 439). The family of completely monotone isotropic covariances includes, for example, the *iterated exponential model* $\exp(\exp(-r/a)) - 1$ ($a > 0$) and the *gamma model* $(1 + r/a)^{-\alpha}$ ($a, \alpha > 0$; for $\alpha = 1$ this model is named the *hyperbolic model*).

Schoenberg also proves that $C(r)$ is a completely monotone isotropic covariance if and only if $C(r^2)$ is an isotropic covariance in \mathbb{R}^n for all n . A consequence is for example that the *generalized Cauchy model* $(1 + r^2/a^2)^{-\alpha}$ ($a, \alpha > 0$) is a covariance whatever the space dimensionality.¹²

Let us end with a useful result whose proof can be found in Yaglom (1987, vol. 1, pp. 358–360): since an isotropic covariance in \mathbb{R}^n , $n > 1$, is also a covariance in \mathbb{R}^1 , it has an n -dimensional spectral density f_n and a one-dimensional spectral density f_1 as well. These spectral densities are related by

$$f_1(\rho_1) = \frac{2\pi^{(n-1)/2}}{\Gamma((n-1)/2)} \int_{\rho_1}^\infty f_n(\rho) (\rho^2 - \rho_1^2)^{(n-3)/2} \rho d\rho \quad (2.29)$$

This relation can be easily inverted when $n = 2$ or 3 :

$$f_2(\rho) = -\frac{1}{\pi} \int_\rho^\infty \frac{df_1(\rho_1)}{d\rho_1} (\rho_1^2 - \rho^2)^{-1/2} d\rho_1 \quad f_3(\rho) = -\frac{1}{2\pi\rho} \frac{df_1(\rho)}{d\rho}$$

To calculate the spectral density of an isotropic covariance in \mathbb{R}^2 or \mathbb{R}^3 , it is often simpler to calculate the spectral density $f_1(\rho)$ in \mathbb{R}^1 and apply these relations rather than directly use relations (2.25) and (2.26). Note that the 1D spectral density associated with an isotropic 3D density is necessarily decreasing. Inverting (2.29) is difficult when $n > 3$, especially when n is even (see Yaglom, 1987, vol. 2, ch. 4, n. 45).

To conclude, it may also be interesting to exploit the *turning bands* operator of Section 7.4 which establishes a one-to-one mapping between isotropic covariances or variograms in \mathbb{R}^n and covariances or variograms in \mathbb{R}^1 .

2.3.4. Covariograms

Definition and Properties

Let $w(x)$ be a function in \mathbb{R}^n , both integrable and square integrable, and let $\varphi(u)$ be its Fourier transform. We can define the covariogram $g(h)$, $h \in \mathbb{R}^n$, as the convolution of $w(x)$ by the function $\check{w}(x) = w(-x)$:

$$g(h) = (w * \check{w})(h) = \int w(x)w(x+h)dx \quad (2.30)$$

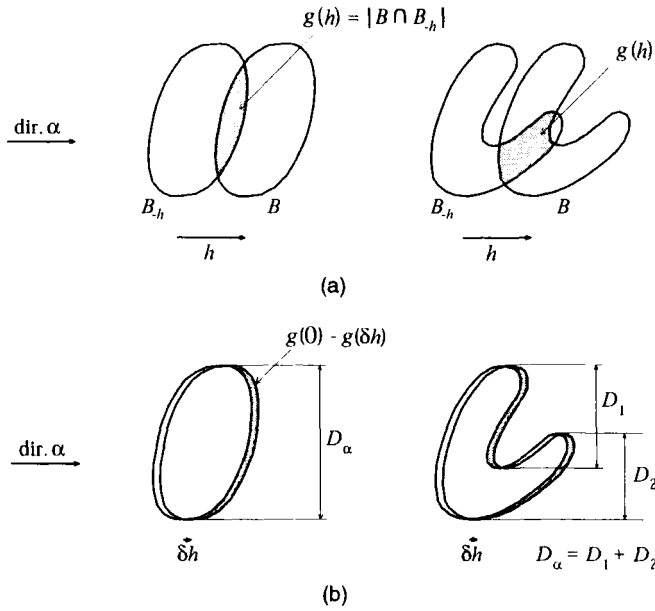


FIGURE 2.15. Geometric covariogram of a bounded domain B : (a) geometric interpretation of $g(h)$; (b) caliper diameter and total diameter in the direction α .

Since the Fourier transformation exchanges convolution and multiplication, the Fourier transform of $g(h)$ is $|\varphi(u)|^2$. It is thus always positive and $g(h)$ is a covariance in \mathbb{R}^n .

In addition to the properties of covariances, the covariogram satisfies the equation

$$\int g(h)dh = \left[\int w(x)dx \right]^2 \quad (2.31)$$

If the support of w is bounded, covariograms have a finite range, in the sense of the integral range defined in Section 2.3.5 and in the strict sense (zero value when $|h|$ exceeds a finite limit, which can vary according to the direction).

The covariogram has a particularly simple interpretation when $w(x)$ is the indicator function of a bounded domain B (Fig. 2.15a)

$$w(x) = 1_{x \in B} = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$$

The product $w(x)w(x+h)$ is then equal to 1 if x and $x+h$ belong to B (i.e., if x belongs both to B and to the translate B_{-h} of B in the translation $-h$) and equal to 0 if not. Thus $g(h)$ is the measure of $B \cap B_{-h}$ (its volume in \mathbb{R}^3 , its area in \mathbb{R}^2 , its length in \mathbb{R}^1). This measure will be denoted by $|B \cap B_{-h}|$. In this case $g(h)$ is called the *geometric covariogram* of B .

Again let us consider the case of an indicator function. As shown in Figure 2.15b, for a small displacement δh in the direction α , the difference $g(0) - g(\delta h)$ is the surface area swept by a vector δh included in B when its end point scans the boundary of B . For a convex domain B this area is $D_\alpha |\delta h|$ where D_α is the tangent diameter (caliper diameter) in the direction α , and we can write in \mathbb{R}^2

$$g(0) - g(\delta h) = D_\alpha |\delta h|$$

This formula remains valid for a nonconvex B , but D_α now represents the *total diameter* in the direction α , namely the total length of the orthogonal projection along α of the portion of contour scanned by δh , or equivalently, one-half the total projected length of the complete contour of B . In any case the partial derivative of $g(h)$ at 0 in the direction α is $g'_\alpha(0) = -D_\alpha$. The left and right partial derivatives at $h = 0$ are of opposite signs, and $g(h)$ is therefore not differentiable at $h = 0$. Differentiability of $g(h)$ at 0 is only possible for functions $w(x)$ that display less abrupt transitions than an indicator function. These properties are easily transposed to \mathbb{R}^n , where the measure of the total projection of B in \mathbb{R}^{n-1} replaces the total diameter.

The covariogram is the basic tool of *transitive methods*, where $w(x)$ is the studied regionalized variable: in the case of a mineral deposit, it is the indicator function or the concentration of a constituent present only in the deposit. A brief presentation of this approach is given in Section 1.3. It will suffice here to consider the covariogram as a means of creating covariance and variogram models, and we will take ordinary analytical function for $w(x)$, the simplest example being the indicator of the sphere of \mathbb{R}^n which gives the spherical model of \mathbb{R}^n (see Section 2.5.1). By reference to the transitive theory, the obtained models are known as *transition models*.

Radon Transform

Let us explicitly write $w(x)$ as a function $w_n(x_1, x_2, \dots, x_{n-1}, x_n)$ of the n coordinates of point $x \in \mathbb{R}^n$. By integration parallel to the x_n axis, we define a function $w_{n,1}$ in \mathbb{R}^{n-1} :

$$w_{n,1}(x_1, \dots, x_{n-1}) = \int_{\mathbb{R}} w_n(x_1, \dots, x_{n-1}, x_n) dx_n$$

The operation allowing one to pass from w_n to $w_{n,1}$ is a Radon transform (cf. Gel'fand et al., 1962, sec. 1.1; Santaló, 1976, sec. IV.19.8), also known in geostatistics as a transitive "*montée*" along the x_n axis.¹³ It represents an accumulation along the direction of integration. By repeating this, we can define the Radon transform or transitive *montée* of order m ($m < n$) in the hyperplane defined by the axes of x_{n-m+1}, \dots, x_n .

Let also $g_n(h_1, h_2, \dots, h_{n-1}, h_n)$ denote the covariogram of w_n . It is readily seen from the definition of $w_{n,1}$ that its covariogram $g_{n,1}$ takes on the form

$$g_{n,1}(h_1, \dots, h_{n-1}) = \int_{\mathbb{R}} g_n(h_1, \dots, h_{n-1}, h_n) dh_n$$

Thus the covariogram of the Radon transform of w_n is the Radon transform of g_n . Through iteration this property is generalized to the m -Radon transform.

If the function w_n is isotropic, so are the Radon-transformed variables and all their covariograms. Let us denote by $g_n(r)$, $g_{n,1}(r)$, and $g_{n,2}(r)$ the initial covariogram and the Radon-transformed covariograms of order 1 and 2 as functions of the modulus r of a vector of the \mathbb{R}^n , \mathbb{R}^{n-1} , and \mathbb{R}^{n-2} space, respectively. Elementary calculations show that

$$\begin{aligned} g_{n,1}(r) &= 2 \int_0^\infty g_n \left(\sqrt{r^2 + \rho^2} \right) d\rho \\ g_{n,2}(r) &= 2\pi \int_r^\infty u g_n(u) du \end{aligned} \quad (2.32)$$

Therefore a transform of order 2, and more generally all transforms of even order obtained by iteration of (2.32), reduce to a very simple integral, whereas a transform of odd order is generally cumbersome. The same relations also obviously apply between w_n and the transformed variables $w_{n,1}$ and $w_{n,2}$.

Still in the isotropic case, a Radon-transformed function is differentiable once more than the original function. Therefore, if we start from a covariogram g_n that is not twice differentiable (e.g., a geometric covariogram), $g_{n,2}$ is then twice differentiable and is a valid model for a differentiable SRF. More generally the covariogram $g_{n,2q}$ (provided that $2q < n$) is $2q$ times differentiable and constitutes a covariance model of q times differentiable SRFs. In applications one must not forget that the m -Radon transform of a covariogram in \mathbb{R}^n produces a model that is only valid in \mathbb{R}^{n-m} . Matheron (1965, ch. II) presents several families of Radon-transformed covariograms.

2.3.5. Integral Range

Let $Z(x)$ be an SRF with covariance $C(h)$, and let us consider the spatial average $Z_V = (1/|V|) \int_V Z(x) dx$. According to (2.12), its variance is

$$\text{Var}(Z_V) = \frac{1}{|V|^2} \int_V \int_V C(x' - x) dx dx' \quad (2.33)$$

If $C(h)$ has a finite range and if the support V is large with respect to the range, the integral $\int_V C(x' - x) dx$ is equal to $\int C(h) dh$, except when x' is close to the boundary of V (at a distance less than the range). To a first approximation the variance of Z_V is then of the form

$$\text{Var}(Z_V) \simeq \frac{A}{|V|} \sigma^2 \quad (2.34)$$

where $\sigma^2 = C(0)$ is the variance of $Z(x)$, and where

$$A = \frac{1}{\sigma^2} \int C(h) dh \quad (2.35)$$

Expression (2.34) remains valid for a covariance that reaches zero asymptotically when $|h| \rightarrow \infty$ provided that (2.35) is finite. We call A the *integral range*. In 1D a useful mnemonic is to remember that the integral range is the range

of the triangle covariance having the same value at the origin and the same area under the curve as $C(h)$. Equivalently A is the integral of the correlogram $\rho(h) = C(h)/\sigma^2$. Alternative names given by Yaglom (1987) are “correlation time” (in a spatial context we would say “correlation length”) or “integral time scale” in 1D, “correlation area” or “integral area scale” in 2D space.¹⁴ Its dimension is that of a volume of the space \mathbb{R}^n . If, for example, $C(h)$ is the covariogram of a function $w(x)$ in \mathbb{R}^n , in view of (2.30) and (2.31) its integral range is

$$A = \frac{[\int w(x)dx]^2}{\int [w(x)]^2 dx}$$

Two special cases deserve mention: (1) if $w(x)$ is the indicator function of a bounded domain of \mathbb{R}^n , A is the measure of this domain; (2) if $w(x)$ is a function integrating to zero, we obtain $A = 0$, which means that the approximate formula (2.34) is no longer valid and that the variance of Z_V decreases faster than $1/|V|$.

If we write $N = |V|/A$, (2.34) takes on the form

$$\text{Var}(Z_V) \simeq \frac{\sigma^2}{N}$$

This is the conventional formula for the variance of the arithmetic mean of N independent points (i.e., mutually located at distances greater than the range). In other words, from the point of view of the mean value the domain V is equivalent to N independent samples.

The integral range is related to the ergodicity of the SRF: if it is finite, the SRF is ergodic, since the variance of Z_V tends to zero when V tends to infinity. However, the SRF may be ergodic in cases when the integral range does not exist. For example, if the integral $\int_V C(h)dh$ does not converge when V tends to infinity but remains bounded, Slutsky's condition (1.12) is satisfied and $Z(x)$ is ergodic. This case is exceptional in practice, but it occurs, for example, in \mathbb{R}^1 with the periodic covariance $C(h) = \cos(h/a)$. Also note that when $C(h)$ is a covariance in \mathbb{R}^n , $n > 1$, its integral range in a subspace of \mathbb{R}^n does not coincide with its integral range in \mathbb{R}^n . In particular, $Z(x)$ may be ergodic in \mathbb{R}^n but not in a subspace of \mathbb{R}^n .

From the above considerations it appears that the integral range represents a yardstick by which we can judge how large a domain of study V is with respect to the scale of the phenomenon. It is an indicator of practical ergodicity. Lantuéjoul (1991) illustrates this with very demonstrative examples.

2.4. REGULARIZATION AND NUGGET EFFECT

The data usually relate to a sampling support that is not punctual. They are affected by microstructures and various error sources that are pooled together

into the nugget effect. A delineation of the various components of the nugget effect permits an improved data collection and sampling control (e.g., Sinclair and Vallée, 1994).

2.4.1. Change of Support—Regularization

Our data may be measured over different sampling supports (cores, channel samples, mine blocks, plots of different sizes, etc.) or represent a weighted average over a volume of investigation (e.g., geophysical tools, remote sensing data). This averaging process can be modeled as a convolution of a (perhaps hypothetical) point-support variable. From this it is easy to establish the relationship between covariances and variograms for different supports.

Let $Z(x)$ be a point-support RF, and define the regularized RF $Z_p(x)$ by the stochastic convolution

$$Z_p = Z * \check{p}$$

that is, $Z_p(x) = \int p(u)Z(x+u)du$, where $p(u)$ is a sampling function and $\check{p}(u) = p(-u)$. This function could be the indicator function of a sample v , normalized by its volume $|v|$, in which case $Z_p(x)$, generally denoted by $Z_v(x)$, is simply the mean value over the sample of support v centered at x . It could also be a more general weighting function that we assume integrable and square integrable so that $Z_p(x)$ exists.

If $Z(x)$ is an SRF with covariance $C(h)$, it follows from (2.13) and the Cauchy algorithm that $Z_p(x)$ is an SRF whose covariance $C_p(h)$ is given by

$$C_p = C * P \quad (2.36)$$

that is, $C_p(h) = \int P(u)C(h+u)du$, where $P = p * \check{p}$ is the covariogram of p : C_p derives from C by regularization by the covariogram P . In particular, the variance of the SRF Z_p is $C_p(0) = \int P(h)C(h)dh$. When $Z_p(x)$ is a weighted moving average, namely when p sums to one, P also sums to one as shown by (2.31), and $C_p(0)$ is then a weighted average of $C(\cdot)$. The variance of Z_p is thus smaller than $C(0)$.

If $Z(x)$ is not an SRF but an IRF with variogram $\gamma(h)$, then similarly $Z_p(x)$ is an IRF whose variogram is given by

$$\gamma_p(h) = (\gamma * P)(h) - (\gamma * P)(0) \quad (2.37)$$

These transformations generally have little effect on the shape of the variogram at large distances but impart a more regular behavior at the scale of the support of the function $p(u)$. This behavior is not isotropic if the function $p(u)$ is not isotropic. Thus for a linear variogram $\gamma(h) = b|h|$ and a regularization along segments with the same orientation and same length l (regularization along

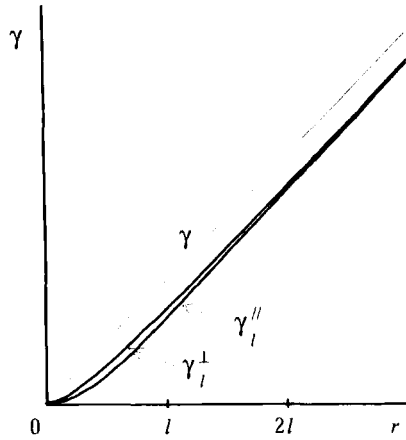


FIGURE 2.16. Linear variogram regularized by a segment of length l along the direction of segment (γ_l^{\parallel}) and perpendicular to segment (γ_l^{\perp}) .

cores), we obtain a regularized variogram $\gamma_l(h)$ such that (with $r = |h|$):

- For a vector h with the same direction as the segments:

$$\gamma_l^{\parallel}(h) = \begin{cases} b \frac{r^2}{l^2} \left(l - \frac{r}{3} \right) & \text{if } r \leq l \\ b \left(r - \frac{l}{3} \right) & \text{if } r \geq l \end{cases} \quad (2.38)$$

- For a vector h orthogonal to the cores:

$$\gamma_l^{\perp}(h) = b \left[\frac{1}{3} \sqrt{l^2 + r^2} + \frac{r^2}{l} \log \frac{l + \sqrt{l^2 + r^2}}{r} - \frac{2}{3} \frac{r^2}{l^2} \left(\sqrt{l^2 + r^2} - r \right) - \frac{l}{3} \right] \quad (2.39)$$

- And more complex expressions for the other directions.

We can see from Figure 2.16 that $\gamma_l(h)$ has a more regular behavior than the point variogram $\gamma(h)$. This behavior is parabolic along the direction of the cores; its great regularity is due to a partial overlap of the cores for $|h| < l$; in practice, the smallest distance for which we calculate a sample variogram is l , and for $|h| \geq l$ the regularized variogram differs from the point variogram only by a constant value; if we extrapolate it to the origin, we obtain an apparent negative nugget effect. Note that (2.39) can be used as an isotropic variogram model for $h \in \mathbb{R}^n$ for all n because it can be considered as the result of a regularization in \mathbb{R}^{n+1} .

Relation (2.37) is usually applied with a known sampling function. It is also possible to invert it to determine P and, to some extent, p (thus the volume investigated), knowing the point and regularized variograms (Royer, 1988).

2.4.2. Microstructure

The formalism of regularization provides a representation of the scaling up mechanism by which a microstructure turns up as a nugget effect at the macroscopic level. Consider a phenomenon made up of microscopic heterogeneities with some characteristic dimension that is very small in comparison with the size of the sampling unit. A good example is a piece of sedimentary rock, such as a core, typically made up of a large number of grains. At the grain level the phenomenon is characterized by a microstructure with a covariance $C_\mu(h)$ whose range is of the order of the diameter of the grains. When observations are average values over a support v with geometric covariogram K , the microstructure is regularized into a component with covariance

$$C_0 = \frac{1}{|v|^2} C_\mu * K$$

Since the range of C_μ is small in relation to that of K , we have

$$C_0(h) = \frac{1}{|v|^2} \int C_\mu(u) K(h+u) du \simeq \frac{1}{|v|^2} \left[\int C_\mu(u) du \right] K(h) = \frac{f_0 K(h)}{|v|^2}$$

where $f_0 = \int C_\mu(h) dh$ (the value of the spectral density at zero frequency). This covariance is equal to $f_0/|v|$ at the origin ($K(0) = |v|$) and vanishes as soon as h exceeds the dimensions of sampling unit v . Since distances smaller than the sampling unit are not considered in practice (except of course at $h = 0$), the effect of the microstructure is to offset the sample variogram at the origin by the amount

$$C_0 = \frac{f_0}{|v|} \quad (2.40)$$

Thus the nugget effect is a reminiscence of a microstructure that is no longer perceptible at the scale of study, and disappears as the sampling unit becomes larger. The macrostructure, made of medium- to long-range components, evolves very slowly at the scale of the sampling unit and remains practically unaffected by the regularization. Increasing the sample support enhances the contrast between micro and macro structures. Note that this theory assumes spatially homogeneous heterogeneities (else a location-dependent nugget effect should be considered).

From the point of view of geostatistical calculations where one considers supports and distances that are large compared with the range of the microstructure, this component appears to be purely random, namely without spatial correlation. In 1D this corresponds to the so-called white noise process, defined as a zero-mean process with a constant spectral density over all frequencies (the term “white” is by analogy with the flat spectrum of white light, as opposed to a preponderance of low frequencies for red light and of high frequencies for blue light).

Strictly speaking, such process cannot exist in the ordinary sense for its total power, the variance, is infinite. However, it can be modeled within the scope of *generalized random processes* (e.g., Yaglom, 1987, vol. 1, pp. 117–120 and sec. 24.1) which are processes defined not by point values but by the values obtained by convolution with regular enough weighting functions. In this model the microstructure can be represented by the macroscopic covariance $f_0 \delta(h)$ where $\delta(h)$ is the Dirac delta-function (Appendix A.1). It is as though the integral $f_0 = \int C_\mu(h) dh$ were concentrated at the origin. The benefit of this formalism is to be able to standardize the treatment of nugget effect under change of support by adding a term $f_0 \delta$ in all formulas involving a covariance.

For example, consider the regularization formula (2.36) with the sampling function $p(x) = I_v(x)/|v|$, where $I_v(x)$ is the indicator function of v . Substituting $f_0 \delta$ for C and using the fact that $\delta * f = f$ for any function f , we obtain

$$C_p = f_0 \delta * P = f_0 P = \frac{f_0}{|v|^2} K$$

which coincides with the covariance $C_0(h)$. In formulas involving the variogram the nugget effect is modeled as $f_0[1 - \delta(h)]$.

The proportionality of the covariance $C_0(h)$ with the geometric covariogram of the support shows the possibility of modeling the true nugget effect (i.e., due to microstructures) as the number of points of a Poisson point process with intensity $\lambda = f_0/|v|^2$ (see Section 7.5.2). Such model fits well with the idea of actual nuggets (stones) scattered randomly in the rock. It also highlights the isotropy of the true nugget effect. Indeed, the Poisson point process is isotropic and an affine transformation of the space does not introduce any anisotropy but just changes the intensity of the process.¹⁵

2.4.3. Measurement Errors

Experimental data often contain measurement errors. Instead of giving the value $Z(x_\alpha)$ at location x_α , the measurement gives $Z_\varepsilon(x_\alpha) = Z(x_\alpha) + \varepsilon_\alpha$, where ε_α is a measurement error. The numerical value of this error being unknown, ε_α is considered as a random variable. If measurement errors at the sample points are independent of Z , nonsystematic, mutually independent and with

the same variance, the variogram $\gamma_\varepsilon(h)$ of Z_ε is

$$\gamma_\varepsilon(h) = C_\varepsilon + \gamma(h) \quad (2.41)$$

where $C_\varepsilon = \sigma_\varepsilon^2$ is the variance of the measurement errors and $\gamma(h)$ the variogram of Z . Formula (2.41) remains valid at $h = 0$ if one is considering pairs of independent measurements made at the same location. Otherwise, $\gamma_\varepsilon(0) = 0$. The variogram of Z_ε then shows an offset at the origin similar to that produced by a microstructure.

The accuracy of measurements is often known. For example, in dividing up a core or a pile of ore to obtain a sample for analysis, the variance of the sampling error, or Gy's fundamental error, can be calculated (Gy, 1975, 1979; for the relationship between sampling theory and geostatistics, see Deverly, 1984a, b; François-Bongarçon, 1993). Similarly a chemical analysis technique or an instrument can be calibrated.

If the measurement errors are not with the same variance, the apparent nugget effect will be equal to the average variance. If the errors are spatially correlated, C_ε must be replaced by a structured component with sill C_ε . More complex cases (e.g., correlation with Z) are handled with multivariate models.

2.4.4. Positioning Errors

In case of a positioning error, the value believed to be at point x_α has in fact been measured at some other point $x_\alpha + U_\alpha$. Instead of studying the data $\{Z(x_\alpha) : \alpha = 1, \dots, N\}$ we are in fact studying $Z_1(x_\alpha) = Z(x_\alpha + U_\alpha)$, where the U_α are random vectors. The problem often occurs with data originating from a dense sampling of profiles, where the errors are correlated. But let us first consider the case where the U_α are uncorrelated and have the same p.d.f. $p(u)$. If $Z(x)$ is an SRF or an IRF, the variogram of the observations is given by

$$\begin{aligned} \gamma_1(h) &= \frac{1}{2} \iint E[Z(x+h+u') - Z(x+u)]^2 p(u)p(u') du du' \\ &= \int \gamma(h+u'-u) p(u)p(u') du du' \end{aligned} \quad (2.42)$$

where γ is the variogram of Z , i.e.

$$\gamma_1 = \gamma * P \quad \text{with} \quad P = p * \check{p}$$

This formula is valid even for $h = 0$ if we consider pairs of distinct measurement points supposed to have the same location but subject to two different positioning errors. Of course, if we are considering twice the same value, $\gamma_1(0) = 0$. It is thus essential in the analysis of such data to compute the variogram at $h \simeq 0$ from pairs of distinct points only.

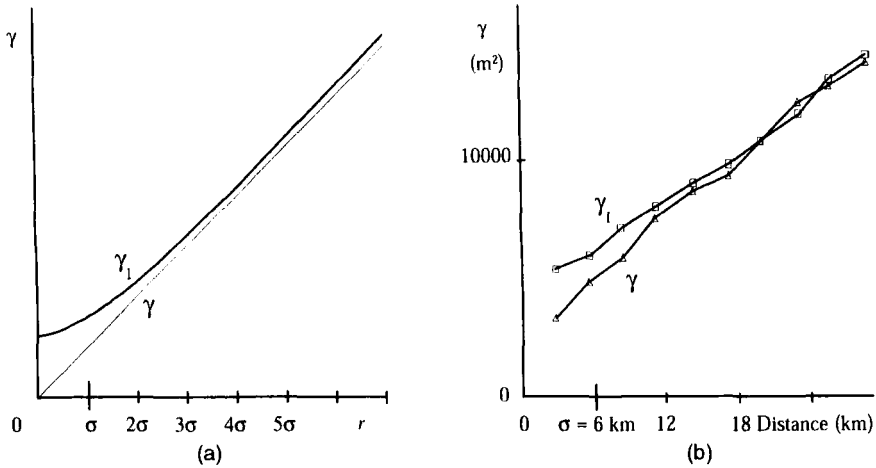


FIGURE 2.17. Linear variogram corrupted by an isotropic Gaussian positioning uncertainty U (2D) with mean square $E(|U|^2) = \sigma^2$: (a) model; (b) example in bathymetry (including measurement error and positioning uncertainty). From Chilès (1976), with kind permission from Kluwer Academic Publishers.

A comparison with formula (2.37) shows that positioning errors act as a regularization plus the addition of a nugget effect

$$C_0 = \iint \gamma(u' - u) p(u) p(u') du du' = \int \gamma(h) P(h) dh$$

adding to the possible nugget effect of Z (Fig. 2.17a). This discontinuity reflects the fact that two apparently close data points can in fact be significantly apart. Formula (2.42) can be easily generalized to correlated errors provided that the joint p.d.f. of U and U' as a function of $h = x' - x$ is known.

These features are exploited in the analysis of data along marine survey lines (Chilès, 1976, 1977; Fig. 2.17b): the sample variogram computed from pairs of points belonging to the same profile is very little affected by the positioning errors because they are highly correlated at short distances, whereas the sample variogram computed from pairs of points on different profiles conforms to the model of uncorrelated positioning errors and can be used to assess the distribution of this error.

2.5. VARIOGRAM MODELS

2.5.1. Isotropic Covariance or Variogram Models

Anisotropic models derive from isotropic models. Let us therefore begin with the latter. We will present the covariances and variograms as functions of $r = |h|$, and the corresponding spectra, also isotropic, as functions of $\rho = |u|$.

For variograms associated with a covariance, we will give the analytical form of the covariance, the variogram being deduced from this by the equation $\gamma(h) = C(0) - C(h)$; conversely, all figures will display the graph of the variogram function, which is the structural tool. Models will be given in normalized form, namely $C(0) = 1$ for covariances (they are thus correlograms) and multiplicative coefficient equal to 1 for variograms.

Spherical Models and Derived Models

By autoconvolution of the indicator function of the sphere of \mathbb{R}^n with diameter a , namely in terms of the modulus $\xi = |x|$, of the function

$$w_n(\xi) = \begin{cases} 1 & \text{if } \xi \leq \frac{a}{2} \\ 0 & \text{if } \xi > \frac{a}{2} \end{cases} \quad (2.43)$$

we obtain the spherical covariogram of \mathbb{R}^n , which we can consider as a function of $r = |h|$:

$$g_n(r) = \begin{cases} a^n v_{n-1} \int_{r/a}^1 (1-u^2)^{(n-1)/2} du & \text{if } r \leq a \\ 0 & \text{if } r \geq a \end{cases} \quad (2.44)$$

In these formulas, v_n is the volume of the unit *diameter* ball of \mathbb{R}^n , which can be deduced from the volume V_n of the unit *radius* ball of \mathbb{R}^n , whose expression is given by (A.5):

$$v_n = \frac{V_n}{2^n} = \frac{\pi^{n/2}}{2^{n-1} n \Gamma(n/2)}$$

For odd $n = 2p + 1$ the integrand of (2.44) is a polynomial, which gives after integration

$$g_{2p+1}(r) = a^{2p+1} \left[v_{2p+1} - v_{2p} \sum_{l=0}^p \frac{(-1)^l}{2l+1} \binom{p}{l} \left(\frac{r}{a} \right)^{2l+1} \right] \quad (r \leq a)$$

The models used in practice, given hereafter as covariances in their normalized form, correspond to $n = 1, 2, 3$:

- *Triangle model*, also called *tent covariance*, valid in \mathbb{R}^1 :

$$C_1(r) = \begin{cases} 1 - \frac{r}{a} & \text{if } r \leq a \\ 0 & \text{if } r \geq a \end{cases} \quad (2.45)$$

- *Circular model*, valid in \mathbb{R}^2 :

$$C_2(r) = \begin{cases} \frac{2}{\pi} \left[\arccos\left(\frac{r}{a}\right) - \frac{r}{a} \sqrt{1 - \frac{r^2}{a^2}} \right] & \text{if } r \leq a \\ 0 & \text{if } r \geq a \end{cases} \quad (2.46)$$

- *Spherical model*, valid in \mathbb{R}^3 :

$$C_3(r) = \begin{cases} 1 - \frac{3}{2} \frac{r}{a} + \frac{1}{2} \frac{r^3}{a^3} & \text{if } r \leq a \\ 0 & \text{if } r \geq a \end{cases} \quad (2.47)$$

The corresponding variograms exhibit a linear behavior near the origin and reach their sill at $r = a$: the scale parameter a of the spherical models coincide with the range. When we speak of a spherical model, without specifying n , we are of course referring to the model $C_3(r)$ (Fig. 2.18a). Because of its validity in \mathbb{R}^n for $n = 1, 2$, or 3 , its well-marked range, and its ease of calculation, it is very widely used.

As was shown in Section 2.3.4, more regular models, corresponding to q -times m.s. differentiable SRFs, are obtained by Radon transform of even order $m = 2q$ ($2q < n$). The Radon transform of order $2q$ of the indicator function (2.43) of \mathbb{R}^n is an isotropic function $w_{n,2q}$ in \mathbb{R}^{n-2q} . Considered as function of $\xi = |x|$ ($x \in \mathbb{R}^{n-2q}$), $w_{n,2q}(\xi)$ represents the volume of the sphere of \mathbb{R}^{2q} with radius $\sqrt{\frac{1}{4}a^2 - \xi^2}$. Thus it depends only on q and is given by

$$w_{n,2q}(\xi) = \begin{cases} v_{2q}(a^2 - 4\xi^2)^q & \text{if } \xi \leq \frac{a}{2} \\ 0 & \text{if } \xi \geq \frac{a}{2} \end{cases}$$

Concerning now the covariograms, equation (2.32) applied to the expression (2.44) of g_n shows that $g_{n,2}$ can be expressed in terms of g_n and g_{n+2} :

$$g_{n,2}(r) = \pi \left[(a^2 - r^2)g_n(r) - \frac{v_{n-1}}{v_{n+1}}g_{n+2}(r) \right] \quad (r \leq a)$$

By recursive reasoning, $g_{n,2q}$ ($2q < n$) can be expressed in terms of $g_n, g_{n+2}, \dots, g_{n+2q}$. The result is

$$g_{n,2q}(r) = \frac{\pi^q}{q!} \sum_{l=0}^q (-1)^l \binom{q}{l} \frac{v_{n-1}}{v_{n+2l-1}} (a^2 - r^2)^{q-l} g_{n+2l}(r) \quad (r \leq a) \quad (2.48)$$

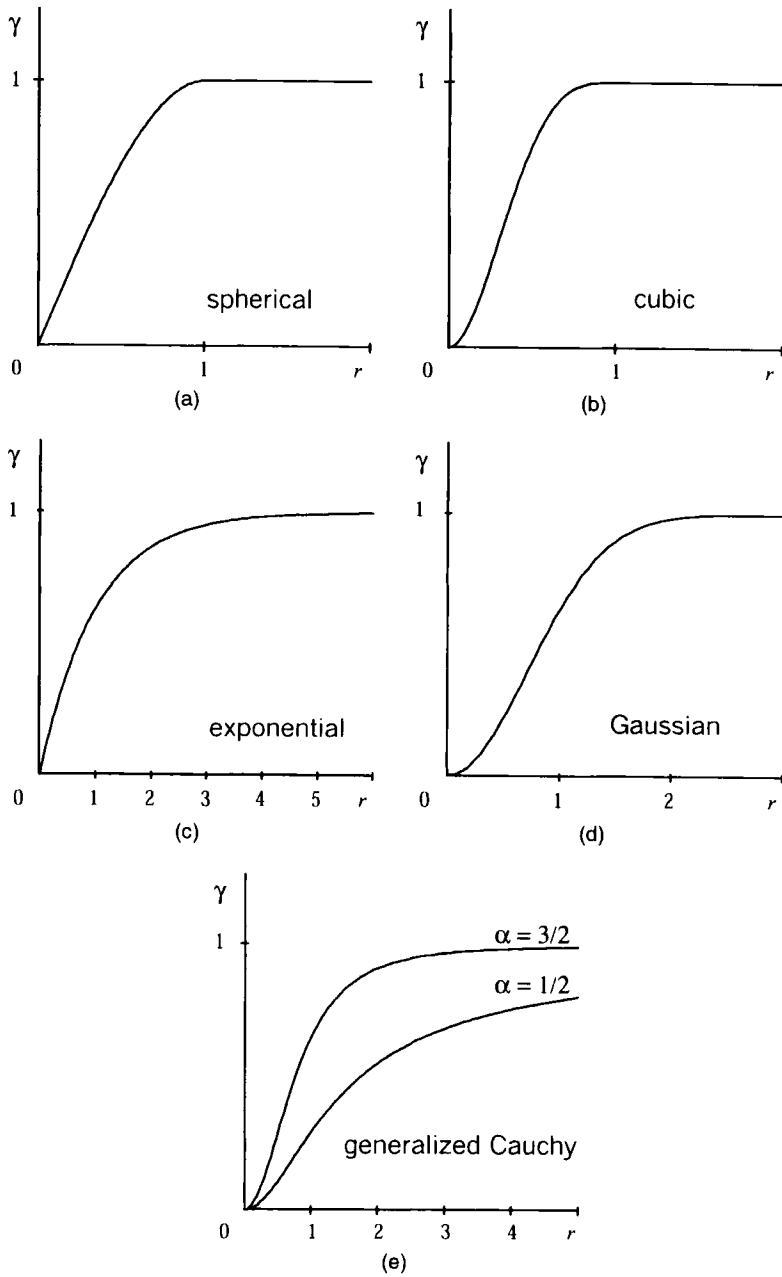


FIGURE 2.18. Variogram models with unit sill and scale parameter—1: (a) spherical; (b) cubic; (c) exponential; (d) Gaussian; (e) generalized Cauchy.

and of course $g_{n,2q}(r) = 0$ for $r \geq a$. These models are valid in \mathbb{R}^{n-2q} . Considering only the lowest-degree models valid in \mathbb{R}^3 , we obtain from formula (2.48) the following two models for $q = 1, n = 5$, and $q = 2, n = 7$, respectively (given here in their normalized form):

1. *The model known as "cubic" because its lowest irregular term is r^3 (the even terms, being infinitely differentiable, are not used in determining the degree of irregularity of a covariance). It corresponds to the Radon transform of order 2 of the spherical covariogram of \mathbb{R}^5 (Fig. 2.18b):*

$$C(r) = \begin{cases} 1 - 7 \frac{r^2}{a^2} + \frac{35}{4} \frac{r^3}{a^3} - \frac{7}{2} \frac{r^5}{a^5} + \frac{3}{4} \frac{r^7}{a^7} & \text{if } r \leq a \\ 0 & \text{if } r \geq a \end{cases} \quad (2.49)$$

This model is used for differentiable variables, such as pressure fields or geopotential in meteorology (Royer, 1975).

2. *A model for a twice m.s. differentiable SRF, which could be called the "pentamodel" because its lowest irregular term is r^5 . It corresponds to the Radon transform of order 4 of the spherical covariogram of \mathbb{R}^7 :*

$$C(r) = \begin{cases} 1 - \frac{22}{3} \frac{r^2}{a^2} + 33 \frac{r^4}{a^4} - \frac{77}{2} \frac{r^5}{a^5} + \frac{33}{2} \frac{r^7}{a^7} - \frac{11}{2} \frac{r^9}{a^9} + \frac{5}{6} \frac{r^{11}}{a^{11}} & \text{if } r \leq a \\ 0 & \text{if } r \geq a \end{cases} \quad (2.50)$$

Its graph differs very little from that of the cubic model, except in the immediate vicinity of the origin. This model is to be used only if the physical conditions ensure that the studied variable is twice differentiable, an exceptional case in the usual applications of geostatistics.

Exponential Model and Derived Models

The exponential model with scale parameter $a > 0$ is defined by

$$C(r) = \exp\left(-\frac{r}{a}\right) \quad (2.51)$$

This model is a covariance in \mathbb{R}^n for any n , since it corresponds to the positive spectral density

$$f_n(\rho) = \frac{2^n \pi^{(n-1)/2} \Gamma((n+1)/2) a^n}{(1 + 4\pi^2 a^2 \rho^2)^{(n+1)/2}}$$

(e.g., see Yaglom, 1987, vol. 1, pp. 362–363). As shown by Figure 2.18c, the variogram reaches its sill only asymptotically when $r \rightarrow \infty$, and its practical range (95% of the sill, or equivalently a correlation of only 5%) is about $3a$.

In \mathbb{R}^1 the exponential model is the covariance of continuous-time Markov processes that possess the property of conditional independence between the past and the future when the present is known. More generally, the distribution of $Z(x)$ conditional on $Z(x_1), \dots, Z(x_i), Z(x_{i+1}), \dots, Z(x_N)$, for $x_1 < \dots < x_i < x < x_{i+1} < \dots < x_N$, only depends on the two neighbors $Z(x_i)$ and $Z(x_{i+1})$.

Since the exponential model is valid in \mathbb{R}^n for any n , Radon transforms of $C(r)$ provide differentiable covariances that are also valid in \mathbb{R}^n for all n , and in particular:

- for the Radon transform of order 2:

$$C_2(r) = \left(1 + \frac{r}{a}\right) \exp\left(-\frac{r}{a}\right) \quad (2.52)$$

- for the Radon transform of order 4:

$$C_4(r) = \left(1 + \frac{r}{a} + \frac{1}{3} \frac{r^2}{a^2}\right) \exp\left(-\frac{r}{a}\right) \quad (2.53)$$

Gaussian Model

The Gaussian model with scale parameter $a > 0$ defined by

$$C(r) = \exp\left(-\frac{r^2}{a^2}\right) \quad (2.54)$$

is a covariance in \mathbb{R}^n for any n . Its practical range is about $1.73 a$, as shown in Figure 2.18d. This model is associated with an infinitely differentiable SRF and thus is extremely regular. After a Radon transform of order m , the covariance remains Gaussian, as can be seen by applying (2.32): it only differs from the initial model by the multiplicative factor $(\pi a^2)^{m/2}$. This regularity gives the SRF a deterministic character, in that knowing the value of the SRF at 0, as well as the values of its partial derivatives of all orders, determines the value of the SRF at any location x . Such regularity is hardly ever encountered in the earth sciences. This model has been used in combination with a nugget effect representing a microstructure, in meteorology for geopotential fields (Delfiner, 1973; Schlatter, 1975; Chauvet et al., 1976), and in bathymetry in areas where the seafloor surface is smooth due to water flow, erosion, and sedimentation (Herzfeld, 1989b).

Generalized Cauchy Model

The gravity and magnetic fields are governed by the laws of physics. If the geometry, density, and magnetism of the sources are known, the corresponding fields can be determined. However, short of knowing these characteristics exactly, a statistical model of the main parameters allows the determination,

if not of the fields, at least of their spectra. This has been studied by several authors, in particular, Spector and Bhattacharyya (1966) and Spector and Grant (1970). The expression of the spectrum is complex, but simpler expressions can be derived by making specific assumptions about the geometry of the sources.

A particularly interesting case is that of sources formed of parallelepipeds of random height, length, width, and orientation located at a random depth but close to an average depth $d = a/2$. The 2D spectra f_G and f_M of the gravity and magnetic fields created at the surface are then approximately of the following form:

$$f_G(\rho) = \alpha \frac{e^{-2\pi a \rho}}{\rho} \quad f_M(\rho) = \beta e^{-2\pi a \rho}$$

Through application of (2.25), we can deduce the corresponding covariances:

$$C_G(r) = \alpha' \left(1 + \frac{r^2}{a^2}\right)^{-1/2} \quad C_M(r) = \beta' \left(1 + \frac{r^2}{a^2}\right)^{-3/2}$$

These two models are in fact particular cases of the *generalized Cauchy model*

$$C(r) = \left(1 + \frac{r^2}{a^2}\right)^{-\alpha} \quad (a > 0, \alpha > 0) \quad (2.55)$$

which is a model in \mathbb{R}^n for all n because it can be expressed in the form (2.27) (e.g., see Yaglom, 1987, vol. 1, p. 365) (the Cauchy model corresponds to $\alpha = 1$). This type of model is very regular near the origin, since its Taylor expansion only contains even terms and it reaches its sill slowly (Fig. 2.18e).

In the case of gravity or magnetic data, and provided that the geometry of the sources is of the type outlined above, we note the remarkable physical significance of the scale parameter a : it is equal to twice the average depth of the sources. These variogram models enable us to take into account both the physics and the geology of the problem. They also enable optimal gravity data transformations (Marcotte and Chouteau, 1993) and an optimal decomposition of the total field into several components, either by spectral methods or by cokriging (Chilès and Guillen, 1984; see Section 5.6.6).

K-Bessel Model

The isotropic function (2.55) is positive and integrable in \mathbb{R}^n for all n . Its n -dimensional Fourier transform is therefore a covariance in \mathbb{R}^n . Considering the case $\alpha = \nu + n/2$, $\nu \geq 0$, leads to the *K-Bessel model*

$$C(r) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{r}{a}\right)^{\nu} K_{\nu}\left(\frac{r}{a}\right) \quad (a > 0, \nu \geq 0) \quad (2.56)$$

where K_ν is the modified Bessel function of the second kind of order ν (A.4) (cf., e.g., Yaglom, 1987, vol. 1, p. 363). This model is valid in \mathbb{R}^n for any n . It can have any type of behavior near the origin, since it behaves like $r^{2\nu}$ if ν is not integer and like $r^{2\nu} \log r$ if ν is integer (Fig. 2.19a).

The case $\nu = \frac{1}{2}$ corresponds to the exponential model. The case $\nu = 1$ also deserves attention: the covariance is $(r/a)K_1(r/a)$ and has been used, for example, to describe rainfall variability in hydrology (Rodríguez-Iturbe and Mejía, 1974).

$|h|^\alpha$ Model

The *power-law model*

$$\gamma(r) = r^\alpha \quad (2.57)$$

is a variogram provided that $0 < \alpha < 2$. This has been shown by Schoenberg (1938b) and Kolmogorov (1940b) in \mathbb{R}^1 .

Proof. Starting from the formula

$$A_{-\alpha}|h|^\alpha = -A_{\alpha+n} \int_{\mathbb{R}^n} [1 - \cos(2\pi \langle u, h \rangle)] |u|^{-\alpha-n} du \quad \text{with} \quad A_t = \pi^{-t/2} \Gamma\left(\frac{t}{2}\right)$$

where $\Gamma(\cdot)$ is the Euler gamma function (A.1), the function $-\Gamma(-\alpha/2)|h|^\alpha$ is obtained through application of (2.21) with

$$\chi(du) = 4\pi^{2-\alpha-n/2} \Gamma\left(\frac{\alpha+n}{2}\right) |u|^{2-\alpha-n} du$$

A change to polar coordinates in which $\rho = |u|$ shows that χ satisfies (2.22) if and only if

$$\int_0^\infty \frac{\rho^{1-\alpha}}{1 + 4\pi^2 \rho^2} d\rho < \infty$$

namely when $0 < \alpha < 2$. In that case χ is a positive measure and $-\Gamma(-\alpha/2)$ is also positive (see Appendix A.2). $|h|^\alpha$, $0 < \alpha < 2$, is therefore a variogram. \square

This model does not have a sill (Fig. 2.19b). The extreme cases of $\alpha = 0$ and $\alpha = 2$ correspond, respectively, to a pure nugget effect and to a linear RF with a random slope. For $\alpha = 1$ we obtain the much used linear variogram $\gamma(h) = |h|$.

The $|h|^\alpha$ model satisfies a property of similarity: it is invariant under a change of the scale of observation. If we go from r to sr , where s is a scale factor, we have $\gamma(sr) = s^\alpha \gamma(r)$. Therefore we cannot associate a characteristic scale to the phenomenon. This model is the only one with such property.¹⁶

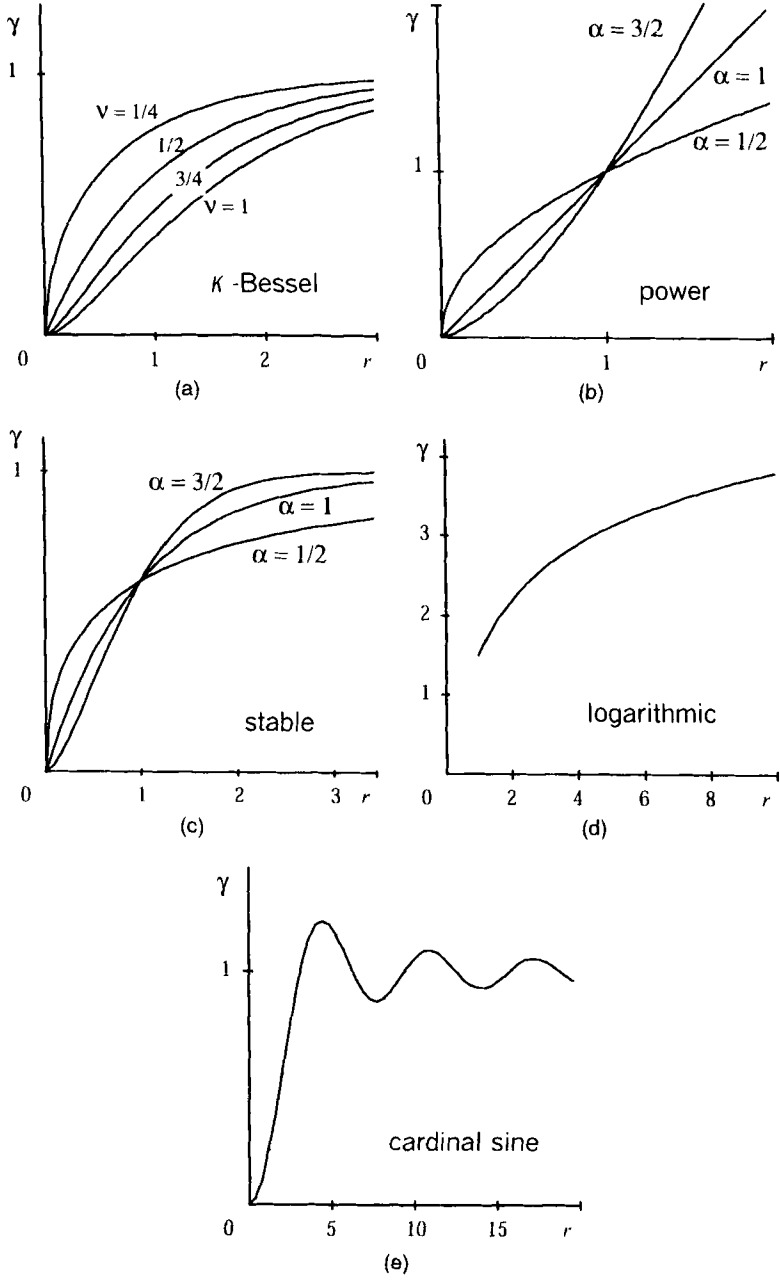


FIGURE 2.19. Variogram models with unit sill and scale parameter—2: (a) *K*-Bessel; (b) $|h|^\alpha$; (c) stable; (d) regularized logarithmic; (e) cardinal sine.

The $|h|^\alpha$ variogram has close ties with fractals. We have all seen spectacular computer-generated images of fractal landscapes (Mandelbrot, 1975b, 1977, 1982). They are extremely rugged surfaces with a noninteger Hausdorff dimension comprised between 2 and 2.5 while their topological dimension is 2. Fractal landscapes are simulations of Gaussian IRFs with variograms of type $|h|^\alpha$, which Mandelbrot named “fractional Brownian” random functions, because they generalize the Brownian motion, or Wiener-Lévy process, obtained in 1D for $\alpha = 1$ (see Mandelbrot and Van Ness, 1968). They are defined for $0 < \alpha < 2$, but $\alpha > 1$ is used for simulating relief (α is related to the standard Hurst exponent H of fractal theory by $\alpha = 2H$). A fractional Brownian random function of \mathbb{R}^n in \mathbb{R}^{n+1} (the space \mathbb{R}^n plus the coordinate z) is a surface with fractal (or Hausdorff) dimension $D = n + 1 - \alpha/2$. The situation may be very different for non-Gaussian random functions. For example, a Poisson process (Section 7.6.1) has a linear variogram but its realizations are not fractal.¹⁷

While on the subject of fractals it may be useful to point out a common confusion between two different properties, an extreme irregularity of detail, on the one hand, and the property to repeat itself at all scales, on the other (self-similarity). Because these properties are both present with the fractional Brownian random function and with many nonrandom fractal constructions, they are sometimes assumed to go together and lead to the erroneous conclusion, for example, that since a phenomenon is very irregular it ought to be self-similar. In fact the fractal character is mathematically defined by a Hausdorff dimension exceeding the topological dimension, that is, by the irregular character of the surface. If we add any other component to a fractal $Z(x)$, be it a smooth SRF or another fractal with a different α , the result is still fractal but no longer self-similar; on the other hand, a plane is a self-similar surface that is obviously nonfractal. In the Gaussian case it is the $|h|^\alpha$ behavior *near the origin*, $0 < \alpha < 2$, which is related to the fractal character so that, for example, any realization of a Gaussian SRF with a spherical or exponential covariance is fractal without being self-similar. Again the situation is different for non-Gaussian random functions.¹⁸

In 1D a Brownian motion ($\alpha = 1$) is a process with independent and stationary Gaussian increments. For $\alpha \neq 1$ the successive increments $Z(x) - Z(x - h)$ and $Z(x + h) - Z(x)$ have a correlation coefficient $\rho = 2^{\alpha-1} - 1$ regardless of the value of h , a remarkable property characterizing this model. Thus a fractional Brownian motion has long-term memory. For $\alpha > 1$ the successive increments are positively correlated, which corresponds to a phenomenon of persistence: as Mandelbrot phrases it, the curve tends “to persist in any direction upon which it has embarked.” For $\alpha < 1$, on the contrary, we have antipersistence (limited to $\rho > -0.5$): the curve tends to turn back constantly toward the point it came from.

The $|h|^\alpha$ model exhibits a large variety of behaviors near the origin, but it has no range. The application of the theorem characterizing variograms

(Section 2.3.3) provides us with a model that has the same behavior near the origin and reaches a sill (Fig. 2.19c). Indeed

$$C(r) = \exp\left(-\left(\frac{r}{a}\right)^\alpha\right) \quad (a > 0, 0 < \alpha \leq 2) \quad (2.58)$$

is a covariance in \mathbb{R}^n for all n . Since in \mathbb{R}^1 this function is the characteristic function of a stable random variable, this covariance is named the *stable model* (see Schoenberg, 1938b, and Yaglom, 1987, vol. 2, ch. 4, n. 50). The exponential and Gaussian models belong to this family.

The $|h|^\alpha$ model and its variant play an important role in the theory of turbulence and its application to meteorology: Kolmogorov (1941a, b) and Obukhov (1941, 1949b) have shown from theoretical considerations that in a fully developed turbulence the velocity components have variograms of type $|h|^\alpha$ with $\alpha = 2/3$ at short and medium distances, which has been confirmed experimentally (see also Monin and Yaglom, 1965, ch. 8). At larger distances, however, the variogram reaches a sill and the stable model or the K -Bessel model with $\nu = \frac{1}{3}$ have been used (Gandin, 1963, p. 51).¹⁹ The situation is similar for the geopotential height of isobaric surfaces, as reported by Gandin (1963, p. 44), this time with $\alpha = \frac{5}{3}$ and $\nu = \frac{5}{6}$, but twice-differentiable models are also used. Blanc-Lapierre and Fortet (1953, pp. 453–454) use the K -Bessel model for electrical noises.

Logarithmic Model

The logarithmic model

$$\gamma(r) = \log r \quad (2.59)$$

is known under the name of *de Wijs model*. The function does not vanish at zero but has a value $-\infty$, which implies that the variance of the point-support variable is infinite. In reality this model is used for describing not point variables but variables regularized by a sampling support. Since any measurement is always based on a support which, no matter how small, is not strictly a point, this restriction is not constraining. By application of (2.37), the variogram $\gamma_p(h)$ associated with such sampling function $p(x)$ is itself an ordinary variogram satisfying $\gamma_p(0) = 0$ and $\gamma_p(h) > 0$ when $|h| > 0$. Figure 2.19d shows the variogram regularized by a segment of length l , calculated along a direction orthogonal to the segment. As soon as r is sufficiently large (e.g., greater than $2l$) this variogram is approximately

$$\gamma_p(r) \simeq \log\left(\frac{r}{l}\right) + \frac{3}{2}$$

From a mathematical point of view, $\log r$ is a variogram model for a random distribution (in the sense of pseudofunctions), but as soon as one passes to reg-

ularized variables, this random distribution becomes a standard RF (Matheron, 1965, pp. 173, 242).

In \mathbb{R}^2 the logarithmic model has a Markov property similar to that of the linear model in \mathbb{R}^1 : if the IRF is Gaussian and known over a closed contour \mathcal{C} , there is independence between the inside and outside of \mathcal{C} (Matheron, 1970, ch. 3, exer. 9).

Because of its good analytical properties, the logarithmic model was widely used by the pioneers of geostatistics (Matheron, 1962; Carlier, 1964; Formery, 1964) in applications to deposits of gold and uranium and also of bauxite and base metals. This model has one remarkable property: if the variogram of the RF defined in \mathbb{R}^n is

$$\gamma(r) = n\alpha \log r$$

and if v and V are two geometrically similar sets of \mathbb{R}^n , the dispersion variance within V of the grades of samples of size v (see Section 2.8.2) is given by

$$\sigma^2(v | V) = \alpha \log \left(\frac{|V|}{|v|} \right) \quad (2.60)$$

This expresses a property of very strong similarity: the variance of the small blocks v within the large block V depends only on the ratio $|V|/|v|$ of volumes, regardless of the scale (*de Wijs formula*). The parameter α is called the absolute dispersion, since it characterizes the dispersion of the phenomenon independently of the geometry of the domain V and the support v .

This type of behavior was revealed by Krige (1952) when studying the large Orange Free State gold deposit which is subdivided into ten mines: the variance of borehole grades (or more precisely the variance of the logarithms of the gold accumulations²⁰) within the groups of boreholes, then within the mines, and finally within the whole deposit (300 km²), increases as the logarithm of the surface area of the zone in which the variance is calculated (Fig. 2.20). Most applications of the logarithmic model have been related to the gold deposits of South Africa (Krige, 1978).

The logarithmic model is called the *de Wijsian model* in honor of the work of H. J. de Wijs (1951, 1953) on the distribution of grades in mine deposits. De Wijs derived formula (2.60) by postulating a principle of similarity by which, when a block V with grade z_V is divided in two equal parts, the two half-blocks have grades $(1 - d)z_V$ and $(1 + d)z_V$, and this coefficient d does not depend on the size of V . When splitting the blocks further up to elementary blocks of size v , it is shown that the grades follow a logbinomial distribution and that their logarithms (rather than the grades themselves) have a variance of the form (2.60). Therefore for a point support this model has close ties with the lognormal distribution (Matheron, 1955; 1962, pp. 308–311, or 1987b). Its principle of similarity allows it to be considered as a fractal model (Mandelbrot, 1982, pp. 376–377), or more precisely a multifractal model (Evertsz and Mandelbrot, 1989).

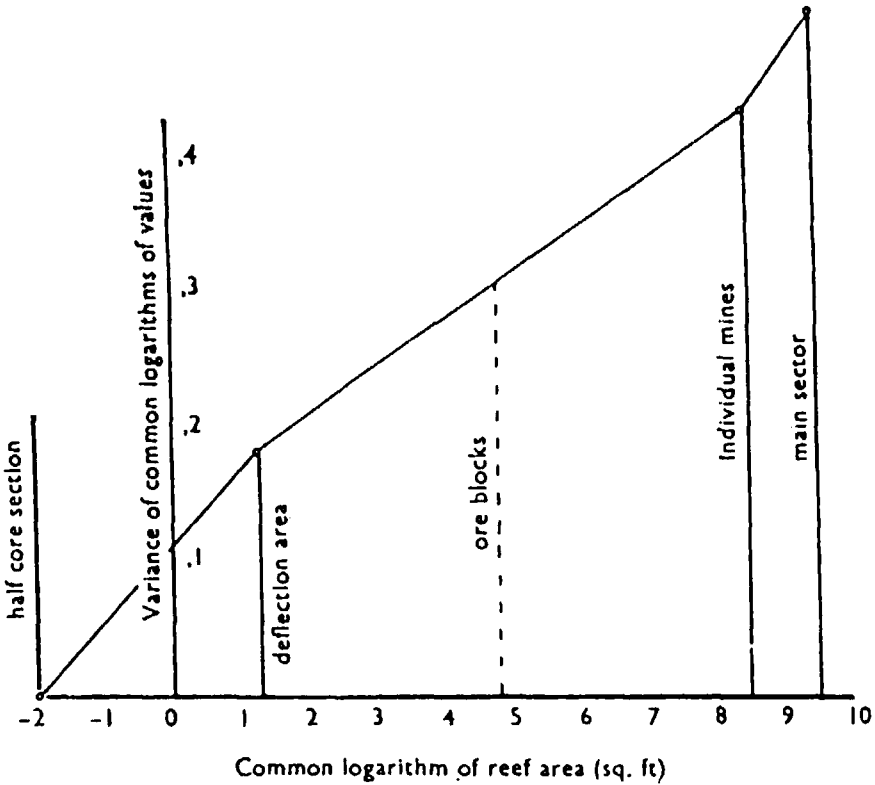


FIGURE 2.20. Dispersion variance versus size of domain: example of gold data from the Orange Free State. The horizontal axis represents the area of the domain in logarithmic scale, from 10^{-2} to 10^{10} ft². From Krige (1952).

Hole Effect Models

The magnitude of a covariance $C(h)$ being bounded by its value at the origin $C(0)$, the maximum hole effect is reached when the covariance reaches $-C(0)$. The effect is obtained in \mathbb{R}^1 with the periodic covariance

$$C(h) = \cos\left(\frac{h}{a}\right) \quad (h \in \mathbb{R}, a > 0) \quad (2.61)$$

This model is not strictly positive definite, since the variance of $Z(x + 2\pi a) - Z(x)$, equal to $2(C(0) - C(2\pi a))$, is here equal to zero. Since this model has no damping factor it is often combined with the exponential model, which gives

$$C(h) = \exp\left(-\frac{|h|}{a_1}\right) \cos\left(\frac{h}{a_2}\right) \quad (h \in \mathbb{R}, a_1, a_2 > 0) \quad (2.62)$$

In higher dimensions an isotropic covariance cannot give such clear-cut hole effect as (2.61). As any isotropic covariance in \mathbb{R}^n has the form (2.24), the maximum hole effect in \mathbb{R}^n is obtained with the *J-Bessel model*

$$C_n(r) = \kappa_n \left(\frac{r}{a} \right) = 2^{n/2-1} \Gamma \left(\frac{n}{2} \right) \left(\frac{r}{a} \right)^{1-n/2} J_{n/2-1} \left(\frac{r}{a} \right) \quad (a > 0) \quad (2.63)$$

As a consequence the ratio $C(r)/C(0)$ for an isotropic covariance cannot be less than -0.403 in \mathbb{R}^2 , -0.218 in \mathbb{R}^3 , and -0.133 in \mathbb{R}^4 (Matérn, 1960, p. 16).

These covariances are very regular: their Taylor expansions include only even terms. Their behavior at infinity is a sine wave multiplied by $1/r^{(n-1)/2}$. Letting $n = 1$ in (2.63) gives back (2.61). For $n = 3$ we obtain the *cardinal-sine model*:

$$C(r) = \left(\frac{a}{r} \right) \sin \left(\frac{r}{a} \right) \quad (h \in \mathbb{R}^3) \quad (2.64)$$

The minimum value (-0.218) is reached for $h \simeq 4.50a$ (Fig. 2.19e).

Since the model (2.62) is the product of an exponential model, which can be extended to \mathbb{R}^n , and a cosine model, which is valid only in \mathbb{R}^1 (as a function of $r = |h|$), we could imagine that it is valid only in \mathbb{R}^1 . This model can nevertheless be extended to an isotropic model in \mathbb{R}^n under a condition on the parameters. For example,

$$C(r) = \exp \left(-\frac{r}{a_1} \right) \cos \left(\frac{r}{a_2} \right) \quad (2.65)$$

is a covariance in \mathbb{R}^2 if and only if $a_2 \geq a_1$, and a covariance in \mathbb{R}^3 if and only if $a_2 \geq a_1 \sqrt{3}$ (Yaglom, 1987, vol. 1, p. 366).

2.5.2. Anisotropic Models

Geometric Anisotropy

By definition, a variogram in \mathbb{R}^n displays a geometric anisotropy if it is of the form

$$\gamma(h) = \gamma_0 \left(\sqrt{\mathbf{h}' \mathbf{Q} \mathbf{h}} \right) \quad (2.66)$$

where γ_0 is an isotropic model and \mathbf{Q} a $n \times n$ positive definite matrix. For clarity the $n \times 1$ matrix \mathbf{h} represents the components $\{h_i : i = 1, \dots, n\}$ of vector h . The eigenvalues $\{b_i^2 : i = 1, \dots, n\}$ of matrix \mathbf{Q} and the eigenvectors define

a new orthogonal coordinate system in which the quadratic form $\mathbf{h}'\mathbf{Q}\mathbf{h}$ can be expressed as a sum of squares $\sum_{i=1}^n b_i^2 \tilde{h}_i^2$ of the new components $\{\tilde{h}_i : i = 1, \dots, n\}$ of vector \mathbf{h} . Combining a change of coordinate system with a scaling of \tilde{h}_i into $\hat{h}_i = b_i \tilde{h}_i$ restores isotropy (b_i is chosen positive). The variogram can therefore be written as

$$\gamma(\mathbf{h}) = \gamma_0(|\mathbf{A}\mathbf{h}|)$$

where the matrix \mathbf{A} defines the transformation from the initial space to the isotropic space. The simplest case is when the anisotropy axes (i.e., the eigenvectors) coincide with the coordinate axes. No rotation is necessary, and (2.66) amounts then to

$$\gamma(\mathbf{h}) = \gamma_0 \left(\sqrt{\sum_{i=1}^n b_i^2 h_i^2} \right)$$

or equivalently the matrix \mathbf{A} is simply the diagonal matrix of the b_i .

The expression of \mathbf{A} is more complex when it includes a rotation. In \mathbb{R}^2 , denoting by θ_1 and $\theta_1 + \pi/2$ the main directions of anisotropy, a rotation θ_1 defines a new coordinate system with axes parallel to the anisotropy directions. The matrix \mathbf{A} is then

$$\mathbf{A} = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} \begin{bmatrix} \cos \theta_1 & \sin \theta_1 \\ -\sin \theta_1 & \cos \theta_1 \end{bmatrix}$$

If $\gamma_0(r)$ is a linear variogram with unit slope, b_1 and b_2 are the slopes of $\gamma(\mathbf{h})$ along directions θ_1 and $\theta_1 + \pi/2$, and the graph of the reciprocal $1/b(\theta)$ of the slope as a function of θ describes an ellipse. Similarly, when $\gamma_0(r)$ is a transition model with unit range, $\gamma(\mathbf{h})$ has range $a_1 = 1/b_1$ in the direction θ_1 and range $a_2 = 1/b_2$ in the direction $\theta_1 + \pi/2$, and the graph of the range $a(\theta)$ describes an ellipse. The isovariogram curves are also concentric ellipses, so that this type of anisotropy is named elliptic anisotropy. Usually the anisotropy is described by the ranges a_1 and a_2 rather than by the parameters b_1 and b_2 . θ_1 is often taken as the direction of maximum range and a_1/a_2 named the *anisotropy ratio*.

The generalization to \mathbb{R}^3 is straightforward when one of the main anisotropy axes is the vertical, which is often the case. In the general case, however, denoting the coordinate axes by Ox , Oy , and Oz , the definition of a coordinate system involves three successive rotations: (1) a rotation θ_1 around the Oz axis, leading to new Ox' and Oy' axes; (2) a rotation θ_2 around the Ox' axis so that the Oz axis comes to its final position Oz'' , and (3) a rotation θ_3 around this Oz'' axis to place the other two axes in their final position. In this final system three anisotropy parameters b_1 , b_2 , and b_3 are associated with the main

directions. The matrix \mathbf{A} is therefore

$$\mathbf{A} = \begin{bmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{bmatrix} \begin{bmatrix} \cos\theta_3 & \sin\theta_3 & 0 \\ -\sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_2 & \sin\theta_2 \\ 0 & -\sin\theta_2 & \cos\theta_2 \end{bmatrix} \begin{bmatrix} \cos\theta_1 & \sin\theta_1 & 0 \\ -\sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Isovariogram surfaces are now ellipsoids, as well as the surface representing the range or the inverse of the slope as a function of direction.

Generalizations of these models can be defined but there are consistency requirements to ensure that these are valid. For example, the slope of a model with a linear behavior near the origin cannot vary arbitrarily with direction, even if continuously. Matheron (1975a, pp. 96–98) gives a geometric characterization of the tangential cone defined by all the tangents at $h = 0$ to the graph $z = C(h)$ of a continuous covariance in \mathbb{R}^{n+1} : its intersection with the horizontal plane $z = 0$ is necessarily a symmetric, closed, and convex neighborhood of the origin (included as an interior point). Equivalently we have the same property for the intersection of the tangential cone to the variogram graph $z = \gamma(h)$ with the horizontal plane $z = 1$. When $n = 2$ and letting $b(\theta)$ denote the variogram slope²¹ in the direction of polar angle θ , the basis of the cone in the plane $z = 1$ has the polar equation $\rho(\theta) = 1/b(\theta)$ and describes a convex set. We can verify this for a geometric anisotropy ($a(\theta) = 1/b(\theta)$ defines an ellipse) and also for the zonal anisotropy presented below. A nice result is that the converse is also true: if $\rho(\theta)$ describes the boundary of a symmetric, closed, and convex set, then $\gamma(h) = b(\theta)r$ is a valid linear variogram, and therefore $C(h) = \exp(-b(\theta)r)$ is a valid covariance. Note, however, that there is no guarantee that an arbitrary function with the same tangential cone is a valid variogram or covariance model. For $n > 2$ the convexity property is no longer sufficient; it is required that the slope $b(u)$, as a function of the unit directional vector u , is the supporting function of a Steiner compact set (for $n > 2$ Steiner compact sets are a strict subset of the class of symmetric, compact, convex sets; refer to Matheron, 1975a, or Serra, 1982, for a precise definition of the Steiner class).

Zonal Anisotropy

In this model, also called *stratified anisotropy*, the variogram depends only on some components of the vector h (possibly after an appropriate change of coordinate system). In 3D the simplest case is when the variogram only depends on the vertical component h_z of vector h , namely is of the form $\gamma(h) \equiv \gamma_0(h_z)$. The variogram in a direction θ with respect to the Oz axis is then $\gamma_\theta(r) = \gamma_0(r \cos \theta)$: if γ_0 has sill C and range a , γ_θ has sill C and range $a/\cos \theta$, except in any direction orthogonal to Oz where the variogram is identically

zero. This is the variogram of a variable that remains constant in any horizontal plane and has thus a layered aspect. Another simple case is a variogram that only depends on the horizontal components h_x and h_y of h : the variable is constant along any vertical.

In practice a real phenomenon can very seldom be represented by a pure zonal model. But we often have several components $\gamma_p(h)$, at least one of which is zonal. In \mathbb{R}^n , when the elementary variogram $\gamma_p(h)$ is zonal and depends on only m components of the vector h , it must be a model admissible in \mathbb{R}^m , and not necessarily in \mathbb{R}^n . This model can of course have its own geometric anisotropy in \mathbb{R}^m . If the various components have sills C_p , $\gamma(h)$ has a sill ΣC_p , except in directions orthogonal to those of the components showing the zonal anisotropy. For example, if we have in \mathbb{R}^3 ,

$$\gamma(h) = \gamma_1 \left(\sqrt{h_x^2 + h_y^2 + h_z^2} \right) + \gamma_2 \left(\sqrt{h_x^2 + h_y^2} \right) + \gamma_3(h_z)$$

the sill is $C_1 + C_2$ in the horizontal directions, and $C_1 + C_3$ in the vertical direction, whereas it is equal to $C_1 + C_2 + C_3$ in all other directions. This makes it easy to test the zonal character of an anisotropy.

One must beware of models that partition the coordinates, such as

$$\gamma(h) = \gamma_1(h_x) + \gamma_2(h_y) \quad \text{in } \mathbb{R}^2$$

or

$$\gamma(h) = \gamma_1 \left(\sqrt{h_x^2 + h_y^2} \right) + \gamma_2(h_z) \quad \text{in } \mathbb{R}^3$$

The first model is obtained for an RF of the type $Z(x, y) = Z_1(x) + Z_2(y)$, where x and y are the coordinates of point x . For such an RF certain linear combinations can have a zero variance. Such is the case with

$$Z(x, y) - Z(x, y + v) - Z(x + u, y) + Z(x + u, y + v)$$

because for such an RF the linear combination is itself zero. Thus $-\gamma(h)$ is not a strictly positive definite function. Such a model is used only if it is imposed by the physics of the problem (here the sum of two 1D structures).

Other Anisotropies

Instead of taking a few components $\gamma_p(h)$ with different models, we could also take a large number, if not an infinity, of one-dimensional zonal anisotropic components and construct:

$$\gamma(h) = \int \gamma_u(\langle h, u \rangle) \mu(du) \quad (2.67)$$

where $\gamma_u(h)$ is a family of models valid in \mathbb{R}^1 , μ is a bounded positive measure on the unit half-sphere of \mathbb{R}^n , and u represents both a point of this sphere and the corresponding unit vector (thus $\langle h, u \rangle$ is the component of the vector h in the direction materialized by u). This model is a particular application of stability property 2 of Section 2.3.2. In \mathbb{R}^2 , by restricting ourselves to four components with the same one-dimensional model $\gamma_0(h)$, we obtain

$$\gamma(h) = \mu_1 \gamma_0(|h_x|) + \mu_2 \gamma_0\left(\frac{|h_x + h_y|}{\sqrt{2}}\right) + \mu_3 \gamma_0(|h_r|) + \mu_4 \gamma_0\left(\frac{|h_x - h_y|}{\sqrt{2}}\right)$$

For a linear elementary model $\gamma_0(h) = |h|$, the variogram $\gamma(h)$ associated with $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$ gives a practically isotropic model with slope 2.4μ . This anisotropic model (with $\mu_1, \mu_2, \mu_3, \mu_4$ different but positive) can thus compete with a model exhibiting geometric anisotropy. Its parameters are linear in the expression of the variogram and so lend themselves more easily to automatic identification (Chilès, 1978).

Another variant is the *factorized covariance*, also called *separable covariance*, whose general form in \mathbb{R}^n is

$$C(h) = \prod_{i=1}^n C_i(h_i) \quad (2.68)$$

where h_i are the n components of vector h and $C_i(h_i)$ are covariances in \mathbb{R}^1 . This is the model obtained with the product of n one-dimensional independent RFs along the various coordinate axes. It has nice conditional independence properties (cf. Sections 3.6.1 and 5.8.1).

No model allows the sill to vary continuously with the direction u of the vector h . If $\gamma_0(h)$ is a transition model, $\sigma^2(u)\gamma_0(h)$ is a variogram if and only if the symmetric function $\sigma^2(u)$ is constant almost everywhere (Matheron, personal communication). It can happen that a sample variogram does not satisfy this property. If the phenomenon exceeds what can be explained by mere fluctuations, this does not mean that our anisotropy models are not rich enough but rather that the studied regionalized variable cannot derive from a stationary model.

2.5.3. Internal Consistency of Models

As soon as one departs from the Gaussian model, it is no longer legitimate to use just any covariance or variogram function. The proof that a positive definite function is a covariance is based on the construction of a Gaussian SRF with this covariance, but it does not establish that an SRF with any other type of spatial distribution can have this covariance. The same holds true for IRFs. Although a model is never the reality, it should at least be internally consistent. This question was studied by Matheron (1987a). We will simply present some examples often found in practical applications: indicators and

lognormal SRFs. But let us first introduce a new concept that generalizes the variogram and is useful to characterize internal consistency: the variogram of order α .

Variogram of Order α

If instead of considering the squares of increments, we consider their magnitudes, or more generally their magnitudes raised to the power $\alpha > 0$, we obtain the variogram of order α (Matheron, 1987a) defined by

$$\gamma_\alpha(h) = \frac{1}{2} E|Z(x+h) - Z(x)|^\alpha \quad (2.69)$$

For $\alpha = 2$ we have the ordinary variogram, and for $\alpha = 1$ the expected value of the magnitude of increments. Matheron (1987a) shows that any variogram of order 1, and more generally any variogram of order α , $0 < \alpha < 2$, is an ordinary variogram (i.e., $-\gamma_\alpha$ is a conditionally positive definite function). But this cannot be any model. For example, the inequality

$$|Z(x+h+h') - Z(x)| \leq |Z(x+h+h') - Z(x+h)| + |Z(x+h) - Z(x)|$$

entails that the variogram of order 1 satisfies the triangular inequality

$$\gamma_1(h+h') \leq \gamma_1(h) + \gamma_1(h') \quad (2.70)$$

More generally, a variogram of order α , $\alpha > 0$, should satisfy the necessary (but not sufficient) condition

$$[\gamma_\alpha(h+h')]^{1/\alpha} \leq [\gamma_\alpha(h)]^{1/\alpha} + [\gamma_\alpha(h')]^{1/\alpha}$$

A sufficient condition is the following: if $\gamma(h)$ is a variogram of order 2, then $\gamma(h)^{\alpha/2}$, $\alpha > 0$, is a variogram of order α . In particular, $\sqrt{\gamma(h)}$ is a variogram of order 1.

In practice, however, the variogram of order α is rarely studied for itself and is modeled with reference to the associated variogram of order 2, either for a robust estimation of the latter (see Section 2.2.5) or for modeling the bivariate distributions (see Section 6.3.4). For example, for a diffusive SRF (e.g., a Gaussian random function) the variogram of order 1 is, up to a multiplicative factor, the square root of the variogram of order 2, whereas for a mosaic SRF (e.g., i.i.d. random values assigned to the cells of a random partition) these two variograms (and all the variograms of order α provided that they exist) are proportional.

Let us also mention the use of the variogram of order α with $\alpha < 2$ to study SRFs that do not have finite second-order moments and, in particular, SRFs with a stable marginal distribution (Boulanger, 1990). Standard stable random variables depend on two parameters: an index $\alpha \in]0, 2]$, and a dissymmetry coefficient $\beta \in [-1, 1]$. In the case of a symmetric distribution ($\beta = 0$), the

characteristic function is $\Phi(u) = \exp(-|u|^\alpha)$. It corresponds to a Gaussian distribution if $\alpha = 2$. Otherwise, it is a distribution with an infinite variance that has moments of order $\alpha' < \alpha$ only. Hence stable SRFs look very different from usual SRFs. Since the linear combination of independent stable variables with the same α is a stable variable with this α , stable SRFs can have variograms of order $\alpha' < \alpha$ only. These are the basis for specific structural tools.

Variogram of a Random Set

Let us consider a stationary random set, or equivalently its indicator which will be denoted by $I(x)$, to recall that it is a binary SRF. Obviously it satisfies

$$[I(x+h) - I(x)]^2 = |I(x+h) - I(x)|$$

so that its variogram is identical to its variogram of order 1 and must therefore satisfy the triangular inequality (2.70):

$$\gamma(h+h') \leq \gamma(h) + \gamma(h') \quad (2.71)$$

If this variogram behaves like $|h|^\alpha$ at short distances, then necessarily $0 < \alpha \leq 1$. In particular, an indicator cannot be m.s. differentiable. The case $\alpha = 1$ corresponds to a set whose boundary has a specific surface area of finite expectation, whereas for $\alpha < 1$ the boundary is of fractal type, namely its Hausdorff dimension in \mathbb{R}^n is greater than $n-1$.

The condition (2.71) is necessary but not sufficient. A stricter set of conditions is proved by Matheron (1993) in a more general setting than the stationary case, the variogram being considered as a separate function of x and $x+h$ and not simply of h : a necessary condition for a function $\gamma(x, x')$ to be the variogram of an indicator is that for all $m \geq 2$ and for any configuration x_1, \dots, x_m , the values $\gamma_{ij} = \gamma(x_i, x_j)$ satisfy

$$\sum_{i=1}^m \sum_{j=1}^m \varepsilon_i \varepsilon_j \gamma_{ij} \leq 0 \quad \forall \varepsilon_i \in \{-1; 0; 1\} \text{ such that } \sum_{i=1}^m \varepsilon_i = 1$$

Matheron formulates the conjecture that this condition is also sufficient. These conditions are nevertheless not easy to use. In practice, the only covariances that are known to be admissible for indicator functions are those of known random sets.

In 1D a well-known example is obtained by assigning i.i.d. values 0 or 1 to the segments delimited by a Poisson point process, which yields an exponential covariance whose scale parameter is the inverse of the intensity of the Poisson point process. It is extended to the n -dimensional space by assigning i.i.d. independent values to the Poisson polyhedra delimited by a Poisson hyperplane process, which yields an isotropic exponential covariance (cf. Section 7.7.5).

A more subtle example in 1D is the covariance

$$C(h) = \sigma^2 \exp\left(-\frac{|h|}{a_1}\right) \cos\left(\frac{h}{a_2}\right) \quad (2.72)$$

It cannot be that of a random set if $a_1 > a_2$, for in this case the corresponding variogram does not satisfy the triangular inequality (2.70). On the other hand, if $a_1 \leq a_2$, it is possible to construct a random set with this covariance. Indeed, if $a_1 = a_2 = a$ and $\sigma^2 = \frac{1}{4}$, (2.72) is the covariance of an *alternating process* based on a gamma renewal process with shape parameter $\alpha = 2$ (renewal

processes are described in Section 7.7.5).²² Now, if we intersect this random set with an independent random set having an exponential covariance, the resulting covariance is the product of covariances and is of the form (2.72) with $a_1 < a_2$. Note that the graph of this model is a damped sine wave.

Another example of an unusual model is the covariance of a process used to model the concentration of minerals in the vicinity of germs and therefore called a *migration process* (Haas et al., 1967).²³ denoting by p the proportion of 1's and by $q = 1 - p$ the proportion of 0's, its covariance is

$$C(h) = \frac{pq}{q-p} \left[q \exp\left(-\frac{|h|}{pa}\right) - p \exp\left(-\frac{|h|}{qa}\right) \right] \quad (2.73)$$

namely the difference of two exponential functions. In the case $p = q = \frac{1}{2}$, this expression is not valid, and a direct calculation (Lantuéjoul, 1994) gives

$$C(h) = \frac{1}{4} \left(1 - \frac{2|h|}{a} \right) \exp\left(-\frac{2|h|}{a}\right) \quad (2.74)$$

These models display a hole effect. A more general family of models, this time in \mathbb{R}^n , is provided by Boolean random sets (Section 7.8.1): their covariances are of the form

$$C(h) = e^{-2\lambda K(0)} (e^{\lambda K(h)} - 1) \quad (2.75)$$

where λ is a positive intensity and $K(h)$ a geometric covariogram or a mixture of geometric covariograms. Other models will be presented in Sections 7.7 and 7.8 devoted to the simulation of random sets and mosaic random functions.

Variogram of an Indicator Variable

Other random set models are obtained from an SRF with a continuous marginal distribution: for any variable $Z(x)$ we can define the indicator function associated with the threshold, or cutoff, z by

$$I(x; z) = 1_{Z(x) < z} = \begin{cases} 1 & \text{if } Z(x) < z \\ 0 & \text{otherwise} \end{cases}$$

Indicator variograms have two types of application:

1. *Calculation of robust variograms.* Except in the case of a high threshold, the indicator is not sensitive to high values. Therefore the variogram of an indicator, for example that associated with the median threshold, is sometimes used as robust variogram.
2. *Indicator kriging.* Following Journé (1982), numerous authors use kriging estimates of the indicator $I(x; z)$ to approximate the probability that $Z(x_0) < z$ at some point x_0 conditionally on the neighboring data.

However, there are caveats:

1. In the case of a long-tailed distribution, the variogram of the median indicator may be effective at revealing the underlying structure. But for a distribution with little spread the gain in robustness often comes with a loss of structure, as will be seen next in the Gaussian case.
2. The variogram of the indicator usually becomes destructured at extreme thresholds (Matheron, 1982b). The use of indicator kriging can therefore lack performance. We will return to this point in Section 6.2.4.

From the point of view of variographic analysis the variogram of an indicator is not the same as the variogram of Z . For example, as is proved in Section 6.3.3, the covariance $C(h; z)$ of the indicator $I(x; z)$ of a standard bi-gaussian SRF $Z(x)$ is related to the correlogram $\rho(h)$ of $Z(x)$ by

$$C(h; z) = \frac{1}{2\pi} \int_0^{\rho(h)} \exp\left(-\frac{z^2}{1+u}\right) \frac{du}{\sqrt{1-u^2}} \quad (2.76)$$

In particular, for the median threshold $z = 0$ this gives the well-known formula

$$C(h; 0) = \frac{1}{2\pi} \arcsin \rho(h) \quad (2.77)$$

Note that the same *centered* covariances are obtained for the indicator $1 - I(x; z)$ of the event $Z(x) \geq z$.

The models defined by (2.76) are different from those commonly used. Figure 2.21 shows graphs of the correlation of indicators as functions of ρ for various thresholds (note the destructuring effect as z increases). For a given threshold z the indicator covariance $C(h; z)$ is an increasing function of ρ and can therefore be inverted. However, the function $\rho(h)$ obtained is not necessarily a covariance, the reason being that not all random sets can be obtained by truncation of a Gaussian SRF. C. Lantuéjoul (personal communication) has proved that the function

$$\rho(h) = \sin\left(\frac{\pi}{2} \exp\left(-\frac{|h|}{a}\right)\right) \quad (2.78)$$

is a valid correlogram in \mathbb{R}^n for all n . Thus the exponential covariance $C(h) = \frac{1}{4} \exp(-|h|/a)$ can be regarded as the indicator covariance of the median-level set of a Gaussian SRF.

Coming back to (2.76), we note that for ρ close to 1 (i.e., small h and no nugget effect) we have approximately

$$\gamma(h; z) = C(0; z) - C(h; z) \simeq \frac{1}{\pi\sqrt{2}} \exp(-\tfrac{1}{2}z^2) \sqrt{1-\rho(h)}$$

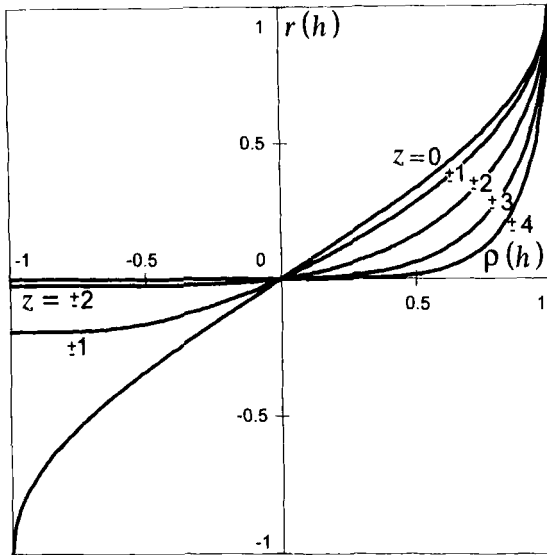


FIGURE 2.21. Graph of correlogram values $r(h) = C(h; z)/C(0; z)$ of the indicator $1_{Z(x) < z}$ (or equivalently of $1_{Z(x) > z}$) of a standard bi-gaussian SRF as a function of correlogram values $\rho(h)$ of the bi-Gaussian SRF, for $z = 0, \pm 1, \pm 2, \pm 3, \pm 4$.

Near the origin the indicator variograms of a Gaussian SRF behave like $\sqrt{1 - \rho(h)}$ and are thus far less regular than the variogram of the Gaussian. Typically, if the variogram of Z is linear near the origin, indicator variograms behave like $\sqrt{|h|}$ for all thresholds. This explains the apparent nugget effect often observed with indicator variograms. It would not be consistent, for example, to use the same variogram model for the Gaussian and for an indicator. For structural analysis and kriging, it is more advantageous to consider the Gaussian RF itself than an indicator.

By contrast, for a mosaic RF all indicator variograms are proportional to the variogram of the RF itself (see Section 6.2.4).

Matheron (1982b) shows in the general case that the indicator variograms $\gamma(h; z)$ and the variogram of order 1 of $Z(x)$, $\gamma_1(h)$, are related by

$$\int_{-\infty}^{+\infty} \gamma(h; z) dz = \gamma_1(h)$$

Furthermore knowledge of the direct and cross-covariances or variograms of the indicators at all thresholds is equivalent to the knowledge of the bivariate distributions of the SRF. Indeed

$$E[I(x; z_1)] = F(z_1)$$

$$E[I(x; z_1)I(x + h; z_2)] = F_h(z_1, z_2)$$

where F denotes the marginal distribution of $Z(x)$ and F_h the bivariate distribution of $Z(x)$ and $Z(x+h)$. It follows that the centered direct and cross-covariances for all possible thresholds are related to the centered covariance of Z by

$$\iint C(h; z_1, z_2) dz_1 dz_2 = C(h)$$

Thus in theory the study of indicators allows a determination of the bivariate distributions of an SRF. But, in practice, this is not so simple, and we will see that to get a consistent model it is preferable to model the bivariate distributions directly (see Section 6.3).

Variogram of a Lognormal Variable

When the marginal distribution is clearly non-Gaussian, we often try to make it Gaussian by a prior transformation of the variable. If, for example, $Z(x)$ has a lognormal marginal distribution, we take $Y(x) = \log Z(x)$, which has a Gaussian marginal distribution. If the bivariate distributions of the SRF $Y(\cdot)$ are Gaussian, the means m_Y and m_Z , the variances σ_Y^2 and σ_Z^2 , and the covariances $C_Y(h)$ and $C_Z(h)$ are related by the equations

$$\begin{aligned} m_Z &= \exp(m_Y + \tfrac{1}{2}\sigma_Y^2) \\ \sigma_Z^2 &= m_Z^2[\exp(\sigma_Y^2) - 1] \\ C_Z(h) &= m_Z^2[\exp(C_Y(h)) - 1] \end{aligned} \tag{2.79}$$

as shown in Appendix A.8. The variogram of a bilognormal SRF being difficult to estimate when the coefficient of variation of the lognormal distribution is not small, these equations make it possible to model the variogram of Z from that of Y . However, they are only valid if Z is bilognormal (i.e., its bivariate distributions and not only its marginal distribution are lognormal).

Conversely, if we decide to directly model the covariance of Z , it must be of the specific form (2.79), where the covariance C_Y can be any positive definite function since $Y(x) = \log Z(x)$ is a bi-gaussian SRF. In other words, not just $C_Z(h)$ but also $\log(1 + C_Z(h)/m_Z^2)$ must be positive definite, which excludes certain models. Since C_Y satisfies $-C_Y(h) \leq C_Y(0)$ according to the Schwarz inequality, it follows that C_Z must satisfy

$$-C_Z(h) \leq \frac{m_Z^2 C_Z(0)}{m_Z^2 + C_Z(0)}$$

Therefore in \mathbb{R}^1 the model $C(h) = \cos(h/a)$ cannot be the covariance of a bilognormal RF. Matheron (1987a) proves that the geostatistician's best friend, the spherical model is *not* compatible with *bilognormality* if the relative vari-

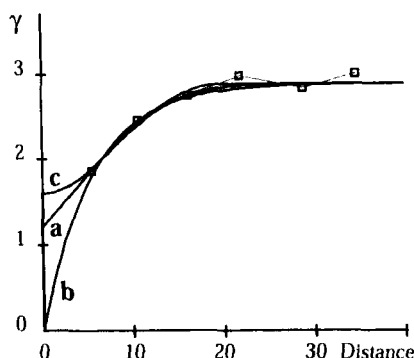


FIGURE 2.22. Sample variogram and three types of fit.

ance exceeds a finite (but unknown) threshold, and conjectures that it is not compatible with lognormality in any case (i.e., *multivariate* lognormality, not just a lognormal marginal distribution). More general transformations are examined in Section 6.3.

2.6. FITTING A VARIOGRAM MODEL

Outside a fairly limited class of models (e.g., linear model with nugget effect) variograms are nonlinear in some of their parameters, typically ranges and anisotropies. Sample variograms also exhibit fluctuations. For these reasons fitting a variogram model is generally done by the geostatistician rather than entirely automatically. Obviously the task is greatly facilitated by computer graphics which can, for example, provide a first fit, but it is up to the geostatistician to choose the type of model (type of variogram, type of anisotropy) and its parameters. Here we will only consider the key aspects of fitting. Numerous examples can be found in the geostatistical literature; for example, see Matheron (1962, 1968a), David (1977), and Journel and Huijbregts (1978).

2.6.1. Manual Fitting

Fitting the Behavior near the Origin

Fitting is often tricky, and the geostatistician introduces, consciously or not, auxiliary hypotheses derived from knowledge of the physics or geology of the problem. Let us consider the variogram of Figure 2.22. In current practice, we extrapolate the linear behavior of the first few points of the variogram and obtain the fit *a* with a nugget effect equal to 1.2. But nothing prevents us from considering other fits near the origin, the two extremes being:

- a structure with a range smaller than the data interdistance and with no nugget effect (fit *b*),

- a very regular behavior near the origin combined with a strong nugget effect (fit c).

The choice of a behavior near the origin is a vulnerable anticipation in the sense that it can always be confirmed or invalidated by additional data on a tighter sampling grid. As long as such data are not available, it is only by formulating such a hypothesis that we are able to proceed further and, for example, perform a kriging estimation. But this choice carries with it a risk of gross error: if it is not well-founded, kriging loses its optimality and the order of magnitude of the calculated kriging variances can be completely erroneous. This point is illustrated by Armstrong and Wackernagel (1988) on a set of 52 topographic data. The consequences are even more serious for a conditional simulation, which actually claims to reproduce variability in detail.

Knowledge about the phenomenon can nevertheless guide the choice of a fit. For example, in the case of Figure 2.22 we may consider the following choices:

1. If we are dealing with a geophysical potential field (gravity or magnetism) which is known to have a very regular spatial structure, fit c is called for, and the nugget effect reflects measurement errors.
2. If we are studying the top of a fairly continuous formation, and if the data come from boreholes and are free of errors, the apparent nugget effect is a short-range structure that can be modeled, as in b , for example, by using a spherical model (the choice of range is then subjective, or is taken from other data).
3. The fit a is used for variables, such as grades or porosities, that show strong variations and are measured on small supports, thus revealing the presence of microstructures.

Measurement errors, microstructure, and short-range components are often mixed. Sometimes the variance of the measurement errors is known from other means, which facilitates the fitting; for example, the stated precision of an equipment gives the standard deviation (or more commonly twice the standard deviation) of the measurement error. However, in order to accurately determine the short-range components, there is hardly any other solution than to take additional samples on a tighter grid (e.g., see Chilès et al., 1996).

Modeling the Continuous Component

Once the type of behavior near the origin has been chosen (for distances less than the first lag), there remains to model the variogram at medium and large distances. Depending on the case, we can use a single basis model or combine several elementary components corresponding to different ranges (nested structures). Figures 2.23 and 2.24 show three examples of such fits. The points

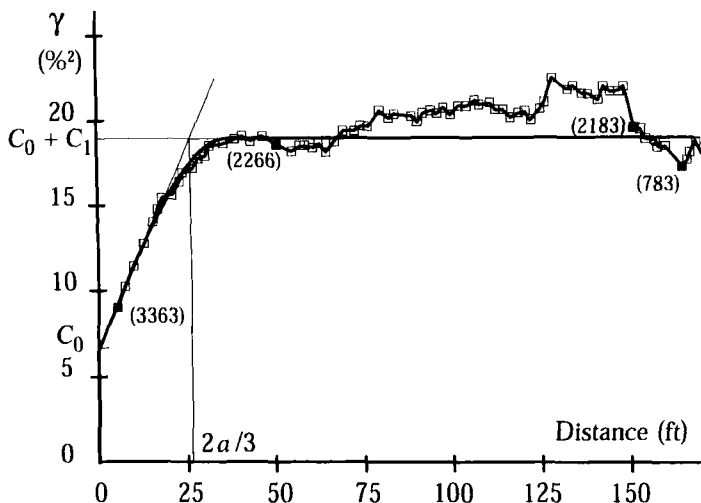


FIGURE 2.23. Example of fit by a single elementary model (% oil). In parentheses: number of pairs; a is the range. From Dowd and Royle (1977), with permission of The Australasian Institute of Mining and Metallurgy; Journel and Huijbregts (1978), with permission of Academic Press Limited, London.

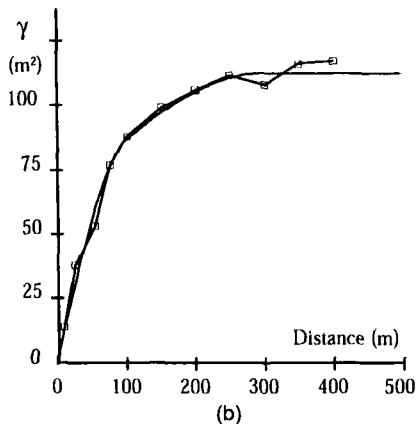
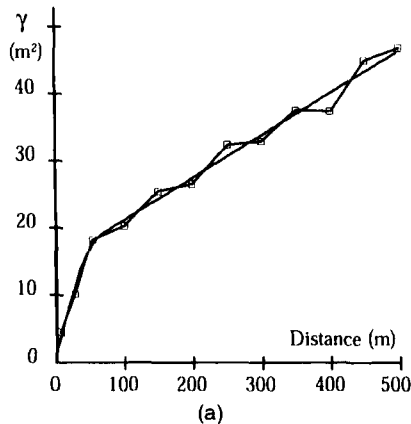


FIGURE 2.24. Examples of fit of a nested structure: (a) thickness of nickel-bearing lateritic ore; (b) thickness of cover. Tiébaghi, New Caledonia. From Chilès (1984), with kind permission from Kluwer Academic Publishers.

that need to be well reproduced are the slope at the origin, the range and the sill (at least for a variogram showing this type of behavior).

The slope at the origin is evaluated from the first points of the variogram. The sill is placed at the level at which the variogram stabilizes (as we have seen in Section 2.1.2, the sample variance tends to underestimate the sill). The range is sometimes found less easily because the point at which it meets the

sill is not always obvious. If a spherical model is considered (very common case in practice), the range can be derived from the fact that the tangent at the origin intersects the sill at two-thirds of the range.

In the absence of physical guidelines (as available for gravity or magnetism) or of internal consistency constraints, the geostatistician has a certain latitude in the choice of the basis models. Once the behavior near the origin has been set, the different possible fits will be graphically very similar at the scale of the study domain, as shown for example by Chilès (1974) with the topographic data from Noiretable (Section 2.7.4). Other comparisons have also been published; for example, see Diamond and Armstrong (1984). In other words, given the behavior near the origin, the fit is robust with regard to the analytical form of the selected model. For example, Journel and Huijbregts (1978) show that the graph of an exponential variogram can be modeled with excellent precision by the sum of two spherical models (p. 234) and that similarly the graph of a regularized logarithmic variogram can be modeled by the sum of two spherical models with a nugget effect (p. 168).

In the case of anisotropy, the variogram has to be fitted along the different directions by a global model, obviously including the anisotropies, so as to obtain a consistent model. The fits are more complex, but also richer, when a more global consistency is required, for example:

- consistency of variograms with regard to data differing in their measurement supports, measurement precision, and positioning precision;
- consistency of variograms with regard to variables linked by functional relationships, such as a variable and one of its indicators or a lognormal variable and its Gaussian transform.

These cases require a globally consistent fit including the theoretical relationships between the different models. This involves the following procedure:

1. Choice of a reference variogram model (point variogram, Gaussian variogram, etc.).
2. Calculation of the theoretical form of the different sample variograms. This calculation takes into account the theoretical relationships between the reference variogram and the variograms considered (obtained after regularization, positioning uncertainty, transformation, etc.).
3. Comparison of the theoretical variograms with the sample variograms, and iteration of steps 1 to 3 until satisfactory fits are obtained.

Modeling a Proportional Effect

A proportional effect is usually considered as a departure from stationarity: the regionalized variable derives from a locally but not globally stationary RF model. Locally, namely in the neighborhood V_0 of point x_0 , the variance of $Z(x + h) - Z(x)$ depends on h and not on x , but it varies from one neighborhood

to the other. More precisely the variogram in V_0 is modeled as a function of h and V_0 of the form

$$\gamma_{V_0}(h) = b(m_{V_0})\gamma(h) \quad (2.80)$$

where b is some function of the local mean m_{V_0} and $\gamma(h)$ a global model. However, Matheron (1974a) shows that a proportional effect is not always incompatible with global stationarity. One can consider the local variogram as the variogram of the random function conditioned on the data that belong to V_0 . By studying the form of SRF variograms conditioned on the local mean m_{V_0} , Matheron shows that if the spatial distribution is lognormal, the local variogram γ_{V_0} takes on the form (2.80) with $b(m_{V_0}) = m_{V_0}^2$, $\gamma(h)$ being the variogram of the nonconditioned SRF. In this case the value of modeling the proportional effect is to be able to provide local estimation variances conditioned on the local mean. These theoretical considerations are confirmed by the example presented by Clark (1979), which also shows that an improper modeling of a proportional effect can give worse results than a globally stationary model. An example of successful modeling of a clear proportional effect is given by Sans and Blaise (1987): the regression of the local variance on the local mean, in a bilogarithmic scale, is a perfect straight line with a slope 2.17 (uranium grades).

2.6.2. Automatic Fitting

Least Squares

Least squares techniques can provide an automatic fit of the sample variogram. In the course of a detailed geostatistical analysis, an automatic fit rarely provides definitive results. It can be the first step of a manual fit.

Generally, we will look for the variogram within a family $\gamma(h; \mathbf{b})$, where \mathbf{b} represents a vector of k parameters b_1, \dots, b_k , belonging to a subset B of \mathbb{R}^n (the parameters are for example the nugget effect, and the range and sill of an isotropic spherical component). Let us denote by $\{\hat{\gamma}(h_j) : j = 1, \dots, J\}$ the values taken on by the sample variogram for J values of the vector h , each calculated from the $N(h_j)$ available pairs. The vector \mathbf{b} can then be chosen:

- by *ordinary least squares* (David, 1977, sec. 6.3.2): \mathbf{b} is selected so as to minimize

$$Q(\mathbf{b}) = \sum_{j=1}^J [\hat{\gamma}(h_j) - \gamma(h_j; \mathbf{b})]^2$$

- by *generalized least squares* (Cressie, 1985): to take into account the correlations between the different values of the sample variogram, \mathbf{b} is chosen

so as to minimize

$$Q(\mathbf{b}) = [\hat{\gamma} - \gamma(\mathbf{b})]' \mathbf{V}^{-1} [\hat{\gamma} - \gamma(\mathbf{b})]$$

where $\hat{\gamma}$ is the vector of $\hat{\gamma}(h_j)$, $\gamma(\mathbf{b})$ the vector of $\gamma(h_j; \mathbf{b})$, and \mathbf{V} should be the variance-covariance matrix of the $\hat{\gamma}(h_j)$ (the variances can be calculated by using the covariance functions G_h defined in Section 2.9.1; the covariances can be calculated in a similar manner).

In this last case it is necessary, for the calculation of \mathbf{V} , to know $\gamma(h)$ as well as the fourth-order moments of the random function. In practice, we work within the framework of Gaussian RFs and proceed by iteration from an initial solution (e.g., obtained by ordinary least squares): the vector \mathbf{b} of the parameters being fixed, we calculate \mathbf{V} , determine the new value of \mathbf{b} minimizing $Q(\mathbf{b})$, and start over again.

A compromise between efficiency (generalized least squares) and simplicity (ordinary least squares) is *weighted least squares*, namely minimization of

$$Q(\mathbf{b}) = \sum_{j=1}^J w_j^2 [\hat{\gamma}(h_j) - \gamma(h_j; \mathbf{b})]^2$$

The weights w_j^2 should be equal to the reciprocals of $\text{Var}[\hat{\gamma}(h_j)]$. Like for generalized least squares we proceed by iteration. In the Gaussian case $\frac{1}{2}[Z(x+h) - Z(x)]^2$ is distributed as $\gamma(h)\chi_1^2$, where χ_1^2 is a chi-square variable on one degree of freedom. If the correlations between the $N(h_j)$ increments taken into account by the calculation of $\hat{\gamma}(h_j)$ are negligible, we then have

$$\text{Var}[\hat{\gamma}(h_j)] \simeq \frac{2\gamma(h_j; \mathbf{b})^2}{N(h_j)}$$

and the problem reduces to minimizing

$$Q(\mathbf{b}) = \frac{1}{2} \sum_{j=1}^J N(h_j) \left[\frac{\hat{\gamma}(h_j)}{\gamma(h_j; \mathbf{b})} - 1 \right]^2$$

Even when based on a Gaussian assumption, weighted or generalized least squares are fairly robust: if this assumption is not really valid, the quadratic form to be minimized will not be optimal, but the result will not be biased.

Maximum Likelihood and Bayesian Method

Some methods work directly from sample data, without requiring the calculation of the sample variogram. Since these methods are blind, they tend to be used only when the presence of a strong drift causes the sample variogram to be hopelessly biased. Section 4.8 discusses the use of optimal

quadratic estimators and regression. Let us examine here the maximum likelihood method and the Bayesian approach which have been applied to stationary and nonstationary fields.

The maximum likelihood method is widely used in statistics for parameter estimation and has been proposed as a means of estimating variogram parameters (Kitanidis and Lane, 1985)—some authors would say an *objective* means. Its principle is the following: assume that the N observed data are from a multivariate Gaussian distribution with mean vector $\mathbf{m} = (E[Z(x)], \dots, E[Z(x_N)])'$ and covariance matrix Σ . The joint probability density of the sample $\mathbf{z} = (z(x_1), \dots, z(x_N))'$ is

$$f(\mathbf{z}) = (2\pi)^{-N/2} |\Sigma|^{-1/2} \exp(-\frac{1}{2}(\mathbf{z} - \mathbf{m})' \Sigma^{-1} (\mathbf{z} - \mathbf{m}))$$

If the mean \mathbf{m} and the covariance Σ are unknown and depend on parameter vectors β and \mathbf{b} , respectively, one can regard \mathbf{z} as fixed and $f(\mathbf{z})$ as a function of β and \mathbf{b} , called the likelihood function $L(\beta, \mathbf{b})$. Maximizing the likelihood, or equivalently minimizing the negative log-likelihood $-\log L(\beta, \mathbf{b})$ yields estimates of the parameters.

In the case of an SRF with known mean, the vector β is absent. But this is not the usual case. When the mean of the SRF is not known, and more generally when the RF has a drift, the method has the advantage of allowing a joint determination of drift and covariance parameters, but it is then prone to biases that may be severe (e.g., see Matheron, 1970, ch. 4, exer. 16). A variant of the method, known as *restricted maximum likelihood*, is to write the likelihood of the data in terms of the $N - 1$ increments $Z(x_{\alpha+1}) - Z(x_\alpha)$ or of more general increments so as to eliminate the mean or the drift from the set of unknowns, which has the effect of reducing the bias on the variogram parameter estimates (Kitanidis and Vomvoris, 1983; Kitanidis and Lane, 1985). In practice, this approach relies of course on a strict Gaussian assumption, which is questionable, to the least. Also it lets the selection of parameters depend on sampling fluctuations.

Handcock and Wallis (1994) present an interesting study of 88 data of a meteorological field (average temperature over one winter) using a maximum likelihood method. The fitting was not done blindly but after a careful exploratory spatial data analysis. The variogram is assumed to be an isotropic K -Bessel model, and therefore \mathbf{b} has three parameters: the shape parameter ν which controls the behavior $|h|^{2\nu}$ near the origin, the scale parameter, and the sill. They obtain $\nu = 0.55$, a value close to $\frac{1}{2}$ which would correspond to an exponential covariance. Indeed a fit of the sample variogram to an exponential model would seem natural.

The authors also use the Bayesian framework to include uncertainty in the variogram estimates and derive the joint posterior distribution of the parameters. The shape parameter ν is found to lie between 0.1 and 1.5 with a sharp mode at about 0.2. Accordingly the variogram behaves like $|h|^{0.4}$ near the origin, namely with an infinite slope characterizing an extremely irregular field, even though meteorological arguments indicate that this average temperature field is continuous and may even be differentiable. This result is due to the choice of prior distribution: a probability density proportional to $1/(1 + \nu)^2$ has been used for $\nu \in]0, \infty[$ because it is supposed to be “noninformative” in the region where the likelihood has mass and rules out very large values of ν . In reality this model gives maximum prior probability to a pure nugget effect, which contradicts the physics of the problem. To produce sensible results, the method should rather use an *informative* prior distribution, but of course the results would then depend on the information introduced by the physicist or the geostatistician (see the example provided by Mostad et al., 1997).

Another approach to the determination of the “belt” of the plausible models associated with a given sample variogram is proposed by Pilz et al. (1997). It assumes Gaussianity of the random field and is based on the spectral representation of the variogram. The same problem is addressed by Solow (1985) with *bootstrap* methods. The bootstrap (Efron, 1979) is a procedure of resampling among the data. It provides approximate confidence intervals of some statistics (in particular, covariances and estimation variances) in the case of independent variables. Solow extends its applicability to correlated observations by introducing a preliminary transformation of the observations to uncorrelated quantities. The method requires the observations to be representative of the underlying population, which limits its scope.

Fuzzy Variogram Fitting

A different approach of the imprecision in variogram fitting consists in a fuzzy fit of the variogram (Bardossy et al., 1990): the variogram parameters are considered as fuzzy numbers, which makes it possible to explicitly introduce the experts' opinion. This opinion is of course subjective, but the advantage of introducing it explicitly is that it will be motivated and can be debated, which is not always the case with methods that have the appearance of objectivity.

2.6.3. Validation

Statistical Tests

In order to validate the consistency of the data with an assumed model (a "hypothesis"), the standard statistical approach is to consider some function of the observations, a *test statistic*, and derive its probability distribution under the assumed model. If the observed test statistic is "too large," namely falls in the tails of the distribution, the model is rejected. There is a probability α that a correct hypothesis is rejected, depending on the definition of the tails (typically $\alpha = 0.05$). When the assumed model depends on some parameter, the set of values of this parameter for which the model is not rejected constitutes a confidence interval at level $1 - \alpha$.

It is very difficult to construct statistical tests in geostatistics because of the spatial dependence between observations. The successive lags of the sample variogram are correlated, as already noticed by Jowett (1955b). Moreover the sample variogram, like the regional variogram, can exhibit large fluctuations at large distances compared to the theoretical model, as will be shown in Section 2.9.2. This limits the applicability of statistical tests to the main features of the variogram, mainly to its behavior near the origin. Switzer (1984) proposes a few tests for variogram scale and shape parameters and the nugget effect, and their inversion into confidence regions. For example, he considers the family $\sigma^2\gamma(h;a)$, where γ is a fixed variogram shape, σ^2 the sill, and a the range. The idea is to linearly transform the data to uncorrelated quantities of constant variance and then consider certain rank orderings. A simple application of these ideas is a test for the range: select a subset of data points whose interpoint distances all exceed the range; then test for randomness based on the rank correlation between $|Z(x_\beta) - Z(x_\alpha)|$ and $|x_\beta - x_\alpha|$, which is then a standard statistical problem (e.g., Spearman's coefficient; cf. Kendall, 1970).

Cross-Validation

A powerful model validation technique is to check the performance of the model for kriging. Here we have to anticipate on results presented in Chapter 3. Consider N data $Z(x_\alpha)$ and a variogram model fitted from the sample variogram calculated from these data. The principle of cross-validation, also called "leave-one-out method," is to estimate $Z(x)$ at each sample point x_α from neighboring data $Z(x_\beta)$, $\beta \neq \alpha$, as if $Z(x_\alpha)$ were unknown. Thus at every sample point x_α we get a kriging estimate $Z_{(\alpha)}^*$ and the associated kriging

variance $\sigma_{K\alpha}^2$. The true value $Z_\alpha = Z(x_\alpha)$ being known, we can compute the kriging error $E_\alpha = Z_{(\alpha)}^* - Z_\alpha$ and the standardized error $e_\alpha = E_\alpha / \sigma_{K\alpha}$. If $\gamma(h)$ is the theoretical variogram, E_α is a random variable with mean zero and variance $\sigma_{K\alpha}^2$, and e_α is a zero-mean unit-variance random variable. Note that kriging errors are not independent. The following results are inspected (for example, see Fig. 4.12):

- the posted standardized errors e_α ;
- the histogram of standardized errors e_α ;
- the $(Z_{(\alpha)}^*, Z_\alpha)$ scatterplot;
- the $(Z_{(\alpha)}^*, e_\alpha)$ scatterplot.

These plots should be examined in the context of the properties of the kriging estimator. In the case of simple kriging, we can check the smoothing relationship, the orthogonality of the estimate and the error, and the conditional unbiasedness in the Gaussian case. In the case of ordinary kriging, these properties do not necessarily hold. Nevertheless, scatterplots allow us to gauge how far we are from the ideal case of simple kriging. The histogram of standardized errors also shows if the kriging error can be considered Gaussian. Moreover these plots are the basis for an interactive data analysis. They highlight data that are poorly “explained” by their neighbors, that can be meaningful anomalies, as well as erroneous data that ought to be corrected or set aside. Clustered anomalies can indicate a fault or a discontinuity in their neighborhood, which must be taken into account. A contour map of standardized errors shows if the magnitude of the error is homogeneous in space or on the contrary points to a lack of stationarity, for example, a proportional effect. Once these problems are fixed adequately, one ought to recompute the sample variogram and redo the cross-validation until acceptable results are obtained. Bradley and Haslett (1992) show an illustrative example of exploratory data analysis.

Comparing the results of two cross-validations performed under different conditions can help one decide between two candidate models, such as two variogram fits, or between a local model and a global model. However, it is not a good idea to generalize this validation method to an automatic variogram-fitting procedure, for example, a blind fit by minimization of $S = (1/N) \sum_{\alpha=1}^N E_\alpha^2$ under the constraint that $s = (1/N) \sum_{\alpha=1}^N e_\alpha^2$ is close to 1. Indeed, if a variogram $\gamma(h)$ leads to $s \neq 1$, replacing it by $\gamma'(h) = \gamma(h)/s$ leaves S unchanged ($S' = S$) and leads to $s' = 1$ even if the variogram shape has nothing to do with the sample variogram. Obviously the fact that s is close to 1 is an indicator of the quality of the fit only if this constraint was not included in the fitting procedure. Besides, the value of S is often essentially due to the contribution of a few points (anomalies or erroneous data). A blind fit by minimization of S would give an overwhelming weight to a few globally or locally extreme points, which is not desirable. Several robust techniques can attenuate this effect: considering the magnitude of errors rather than their

squares, thresholding the errors (e.g., at 2.5 standard deviations), and comparing two options by their scores (counting how many times option 1 performs better than option 2 versus how many times option 2 is better than 1). Such techniques allow a refined comparison between two options but not fitting a model.

In practice, the estimation of $Z(x_\alpha)$ from its neighbors is usually accomplished with a moving neighborhood. However, too close neighbors are eliminated if the data form a cluster or are located along profiles, in order to mimic the subsequent kriging conditions. If the number of points is not too large a global neighborhood may also be used. This means solving N linear systems of size $N \times N$, which may be computationally expensive. A special technique has been developed to replace these calculations by the inversion of a single matrix of size $(N + 1) \times (N + 1)$ (see Section 3.6.3).

2.6.4. Spectral Modeling Approach

So far we have ignored spectral analysis methods, although they are widely used, especially for the study of time series. A fairly complete account is given by Yaglom (1987, vol. 1, ch. 3 in \mathbb{R}^1 , sec. 22.2 of ch. 4 in \mathbb{R}^n), along with many references (in 2D, also see Guyon, 1993, sec. 4.5). These methods require data sampled on a regular grid forming a parallelepiped in \mathbb{R}^n , which is rarely the case in the applications that we consider. Moreover, if the knowledge of the spectral measure and of the covariance are theoretically equivalent, passing from one to the other is not always easy. Signal-processing applications (e.g., the design of filters) essentially involve the spectrum, whereas geostatistical applications (kriging, change of support) involve the covariance. These are reasons to favor variographic analysis. Still it is interesting to see how our signal-processing colleagues, who have given much thought to spectral estimation and modeling, proceed. For simplicity we will restrict the presentation to the one-dimensional case. The transposition to higher dimensions is rather straightforward.

The data being sampled on a one-dimensional regular grid, we will index the data points by n instead of α . Consider a sequence of N sample values $Z_n = Z(n\Delta x)$, $n = 1, \dots, N$, and let us assume to simplify notations that $\Delta x = 1$. Furthermore $Z(x)$ is an SRF with mean zero (in practice, the mean is subtracted initially), covariance $C(h)$, and spectral density $f(u)$. From the N data Z_n it is possible to calculate $2N - 1$ values of the sample covariance

$$\hat{C}_m = \frac{1}{N - |m|} \sum_{n=1}^{N-|m|} Z_n Z_{n+|m|} \quad m = 0, \pm 1, \dots, \pm(N - 1)$$

\hat{C}_m is an unbiased estimator of the theoretical covariance $C_m = C(m\Delta x)$, provided that $Z(x)$ has indeed a zero mean. Under certain conditions (e.g., met for a Gaussian RF) \hat{C}_m is a consistent estimator of C_m in the sense that the quadratic mean of $\hat{C}_m - C_m$ tends to zero when $N \rightarrow \infty$. But for fixed N the quadratic mean of $\hat{C}_m - C_m$ is large for large values of m because \hat{C}_m is then computed with few pairs of points. Moreover, contrary to the theoretical covariances C_m , the $2N - 1$ sample covariances \hat{C}_m do not necessarily constitute a discrete positive definite function. Their inverse Fourier transforms can take on negative values. To avoid this, another estimator is considered, namely

$$\tilde{C}_m = \frac{1}{N} \sum_{n=1}^{N-|m|} Z_n Z_{n+|m|} \quad m = 0, \pm 1, \dots, \pm(N - 1)$$

For fixed N , \tilde{C}_m is a biased estimator of C_m (except if $m = 0$), but it is asymptotically unbiased when $N \rightarrow \infty$. Despite this bias, for fixed N , the quadratic mean of $\tilde{C}_m - C_m$ is generally less than or equal to that of $\hat{C}_m - C_m$ (this is obviously the case when $C_m = 0$, i.e., beyond the range). Furthermore the $2N - 1$ terms \tilde{C}_m form a discrete positive definite function. Hence the spectral density estimator

$$\tilde{f}(u) = \sum_{m=-N+1}^{N-1} e^{-2\pi i u m} \tilde{C}_m$$

This (random) function is called the *periodogram*. It can be computed directly from the data as

$$\tilde{f}(u) = \frac{1}{N} \left| \sum_{n=1}^N e^{-2\pi i u n} Z_n \right|^2$$

Since the data and the covariance are known only at discrete points, $\tilde{f}(u)$ is a (random) periodic function with period 1, and it will be considered only in the interval $u \in]-\frac{1}{2}, \frac{1}{2}]$. For fixed N , $\tilde{f}(u)$ is a biased estimator of the spectral density $f_1(u) = \sum_{m=-\infty}^{+\infty} e^{-2\pi i u m} C_m$ of the infinite random sequence Z_n , but it is asymptotically unbiased when $N \rightarrow \infty$. However, even when $N \rightarrow \infty$, the quadratic mean of $\tilde{f}(u) - f_1(u)$ does not tend to zero: $\tilde{f}(u)$ is not a consistent estimator of $f_1(u)$. Since $\tilde{f}(u)$ and $\tilde{f}(u')$ are uncorrelated (unless $u' = \pm u$), the periodogram exhibits erratic fluctuations. To attenuate this effect, the periodogram is convolved with a weighting function $A(u)$, also periodic with period 1. This leads to the *smoothed periodogram*

$$\varphi(u) = \int_{-1/2}^{1/2} A(u - u') \tilde{f}(u') du'$$

The weighting function $A(u)$ has for Fourier transform the sequence

$$a_m = \int_{-1/2}^{1/2} e^{2\pi i u m} A(u) du \quad m = 0, \pm 1, \dots, \pm(N-1)$$

The function $A(u)$ is called the *spectral window*, whereas the sequence a_m defines the *lag window*. As the Fourier transform exchanges convolution and multiplication one has

$$\varphi(u) = \sum_{m=-N+1}^{N-1} e^{-2\pi i u m} a_m \tilde{C}_m$$

To define $A(u)$ over $]-\frac{1}{2}, \frac{1}{2}]$, one chooses a unit-sum function with a maximum at zero and a rapid fall off with increasing $|u|$ (of course there is a great diversity of such functions). The sequence a_m is then equal to 1 at zero and decays progressively as $|m|$ increases, which has the effect of reducing the values of the sample covariance at large lags (those which are in general poorly estimated).

The calculations can be accomplished by means of three successive discrete Fourier transforms (DFTs): (1) a DFT of the Z_n series, which by squaring gives the periodogram for $u = k/N$, (2) an inverse DFT of $\tilde{f}_k = \tilde{f}(k/N)$ to get the covariances \tilde{C}_m , and lastly (3) a DFT of the tapered covariances $a_m \tilde{C}_m$ to get the smoothed spectral density φ .

In principle, it is not necessary to model φ , since any nonnegative integrable symmetric function can be a spectral density. However, parametric methods exist. The most common one is to start from the first $p + 1$ values of the covariance C_m ($m = 0, \dots, p$) and select the spectral density $f(u)$, $u \in [-\frac{1}{2}, \frac{1}{2}]$, maximizing the entropy H defined by²⁴

$$H = \int_{-1/2}^{1/2} \log f(u) du$$

while still matching the covariance C_m , namely satisfying the conditions

$$\int_{-1/2}^{1/2} e^{2\pi i u m} f(u) du = C_m \quad m = 0, 1, \dots, p \quad (2.81)$$

It can be shown that the solution is of the form

$$f(u) = \frac{\sigma^2}{|1 - \alpha_1 e^{-2\pi i u} - \dots - \alpha_p e^{-2\pi i u p}|^2}$$

This is the general form for the spectral density of an autoregressive process of order p (these will be presented in Section 7.5.1; the α_i are the coefficients of the autoregression and σ^2 is the variance of the innovations). In practice, the theoretical covariances C_m are not known and are replaced by the estimates \hat{C}_m . Among all spectral densities satisfying the conditions (2.81), the maximum entropy spectral density has the special property of corresponding to the linearly most unpredictable random sequence. Indeed the best linear predictor Z_{n+1}^* of Z_{n+1} from the present value Z_n and all past data Z_{n-1}, Z_{n-2}, \dots (the best estimator altogether in the Gaussian case) has the variance

$$\text{Var}(Z_{n+1}^* - Z_{n+1}) = \exp \left(\int_{-1/2}^{1/2} \log f(u) du \right)$$

as shown, for example, by Doob (1953, sec. XII.4). Maximizing entropy thus amounts to maximizing this variance. An illuminating derivation of this result is to note that the knowledge of the $(p + 1)$ covariance values C_m allows the determination of the optimal linear predictor based on the last p values. The autoregressive sequence of order p is precisely that for which the p -value predictor coincides with the predictor based on the infinite past and therefore maximizes the innovation brought by the new value Z_{n+1} (Yaglom, 1987, vol. 2, p. 104).

The above theory is concerned with random sequences and requires data on a grid. These can be obtained by sampling a continuous signal. However, the spectrum f_i of the sequence and the spectrum f of the continuous signal coincide only if frequencies higher than the Nyquist frequency $1/(2\Delta x)$ either do not exist in the signal, or have been filtered out before sampling. This is a problem for most geostatistical applications where sampling is very fragmentary and applying an anti-aliasing filter before sampling is simply impossible. Covariance estimation does not require the signal to be band-limited. The variogram has the additional advantage of not requiring a preliminary estimation of the mean, nor the stationarity of $Z(x)$, but only of its increments.

2.7. VARIOGRAPHY IN PRESENCE OF A DRIFT

When a drift is present one generally turns to the universal kriging (UK) model of Section 3.4. In this model the RF $Z(x)$ is considered as the sum of

a deterministic drift $m(x)$ (usually a polynomial with unknown coefficients) and a zero-mean stationary or intrinsic random residual $Y(x)$. We will see, first using a very simple example, and then more generally, that the presence of a drift poses difficult inference problems. Their definitive solution requires abandoning the UK model for the broader IRF- k model studied in Chapter 4, except in some simplified cases that will be presented here.

2.7.1. Introductory Example

The following values are observations of a regionalized variable $z(x)$ at 11 points $x = 0, 1, \dots, 10$:

x	0	1	2	3	4	5	6	7	8	9	10
z	0	0	0	2	4	6	8	8	10	10	12

We regard this regionalized variable as a realization of an RF $Z(x)$ and wish to find its variogram. This is obviously a hypothetical case because, in practice, one cannot hope to determine a variogram with so few data: only one to ten pairs of points are available to calculate the sample variogram. The data have been chosen to obtain demonstrative results without the need for lengthy calculations.

Raw Sample Variogram

Calculating the sample variogram of the raw data is straightforward. We obtain a curve with a clear parabolic shape, even at large distances (Fig. 2.25a). This behavior reflects the presence of a drift, which is also evident on the data themselves. Obviously, if a regionalized variable is of the form $z(x) = ax + b$, its regional variogram (2.7) is the parabola $\gamma_R(h) = \frac{1}{2}a^2h^2$.

This variogram is a mean square, whereas the variogram that we want is defined as a variance. As long as we are dealing with phenomena that have no drift, increments have zero expectation, and their mean square coincides with their variance. This is not the case here. We must therefore find the underlying variogram obtained after subtracting the drift.

Variogram of Estimated Residuals

As the “true” drift $m(x)$ is unknown, we estimate it assuming a linear shape. The exact choice of the line is more or less arbitrary and we will consider two simple cases:

1. The least squares line: $\hat{m}_1(x) = (73x - 65)/55$.
2. The line joining the end points: $\hat{m}_2(x) = 1.2x$.

From this we compute the residuals $r_1(x) = z(x) - \hat{m}_1(x)$ and $r_2(x) = z(x) - \hat{m}_2(x)$ at the sample points, and their variograms $\hat{\gamma}_1$ and $\hat{\gamma}_2$ (Fig. 2.25b, c).

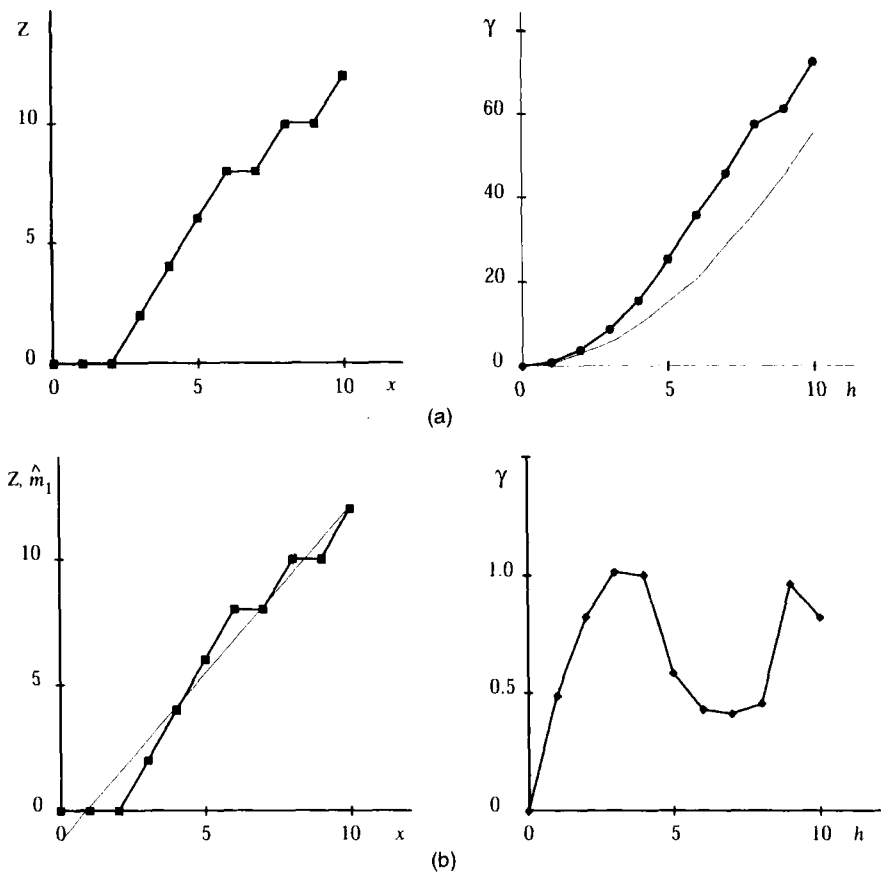


FIGURE 2.25. Example of random walk with linear drift: data (left), sample variograms (right: symbols connected by a thick line) and theoretical variograms (right: thin line): (a) raw data and raw variogram; (b) least squares residuals and variogram of residuals; (c) zero residuals at the end points and variogram of residuals; (d) "true" residuals and underlying variogram; (e) generalized variogram of order 1. (Continued on next page.)

This gives us two variograms of residuals—one too many. This is normal, since the variogram of residuals depends on the evaluation of the drift. The two curves are nevertheless fairly similar, at least for the first points, and they both reach a sill.

The variogram $\hat{\gamma}_1$ has a short range of about $h = 2$. Leaving out the value at $h = 1$, we might even be tempted to say that the variogram is flat and conclude that the residuals are uncorrelated, especially since in this case the least squares line is precisely the optimal estimator of the drift. We could also suspect a bias problem because $\hat{\gamma}_2(10)$ is zero, and this result is inherent to the method and not due to the coincidence of the values of z . In fact the residual $r_2(x)$ is by construction zero at the end points.

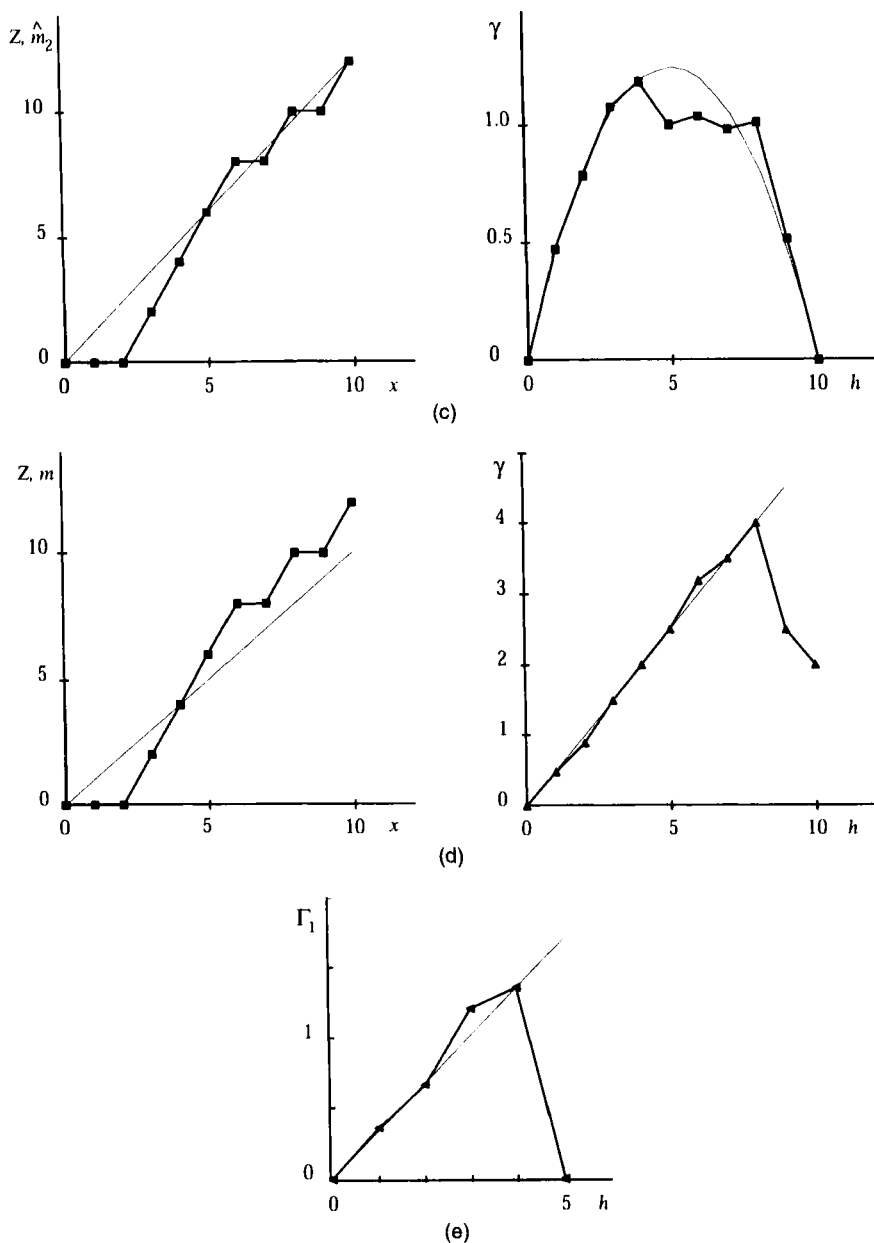


FIGURE 2.25. (Continued).

Underlying Variogram of the "True" Residuals

It is now revealed that the data were generated by flipping a coin: at each trial x is incremented by 1 and z by 0 or 2, depending on whether the outcome is

TABLE 2.1. Precision of the reconstruction of a point from its six neighbors

.	x	z	Hyp. 1 (nugget)		Hyp. 2 (linear)	
			z^*	$z^* - z$	z^*	$z^* - z$
	3	2	3	1	2	0
	4	4	4	0	4	0
	5	6	5.33	-0.67	6	0
	6	8	6.67	-1.33	7	-1
	7	8	8.33	0.33	9	1
Mean square error			0.67		0.40	

Note: Hyp. 1: kriging estimator associated with a pure nugget effect.

Hyp. 2: kriging estimator associated with a linear variogram.

z : true value; z^* : reconstructed value; $z^* - z$: estimation error.

heads or tails. The “true” drift is thus $m(x) = x$. The “true” sample residuals are therefore $y(x) = z(x) - x$. Their variogram is obviously linear with slope 0.5 (Fig. 2.25d).

Validation

We are faced with two diverging conclusions: (1) an absence of correlation, i.e. a pure nugget effect, and (2) a linear variogram. For guidance we carry out a cross-validation for each model: sample points are estimated from six neighbors (three on either side) under the two variogram hypotheses and the results are compared with the actual values.

As will be shown in Section 3.4.2, the kriging estimator of $Z(x)$ from its six neighbors is:

- the arithmetic mean of the six neighboring data in the case of a pure nugget effect;
- the mean of the two nearest neighbors $Z(x - 1)$ and $Z(x + 1)$ if the variogram is linear (the other neighbors have zero weight due to the Markov property of this model).

If we attempt to reconstruct each of the data in turn from its six neighbors (which we can do for $x = 3$ to 7), we obtain the results of Table 2.1. Notice that with one exception, the estimation error is larger for the pure nugget effect than for the linear variogram. The mean square error is 0.67 in the first case against 0.40 in the second. The linear variogram therefore appears as the better model. In the present case this can be confirmed by theoretical calculations.

Theoretical Expressions of the Variograms

Based on the properties of the coin-tossing game, we can calculate the theoretical expressions of the different variograms. The RF $Z(x)$ associated with

$z(x)$ is of the form

$$Z(x) = \sum_{p=1}^x X_p \quad (x \in \mathbb{N})$$

where the X_p are i.i.d. Bernoulli random variables with

$$\Pr\{X_p = 0\} = \Pr\{X_p = 2\} = \frac{1}{2} \quad E(X_p) = 1 \quad \text{Var}(X_p) = 1$$

From this we can easily deduce:

- the drift:

$$m(x) = E[Z(x)] = x$$

- the underlying variogram:

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(x+h) - Z(x)] = \frac{1}{2}|h|$$

- the variogram of the raw values:

$$\gamma_R(h) = \frac{1}{2} E[Z(x+h) - Z(x)]^2 = \frac{1}{2}[|h| + h^2]$$

The theoretical expression for the variograms of the estimated residuals is less straightforward, though it is easy for the variogram γ_2 of the residuals associated with the line joining the end points. It should be pointed out that if the “true” residuals are stationary, or at least intrinsic, this is not necessarily true for the estimated residuals. We must therefore calculate the variogram of the residuals γ_2 as a function of x and $x+h$. Writing $[0, L]$ for the studied interval (here $L = 10$), we obtain for $0 \leq x \leq x+h \leq L$:

$$\gamma_2(x, x+h) = \frac{1}{2} E \left[Z(x+h) - Z(x) - \frac{h}{L}(Z(L) - Z(0)) \right]^2$$

The expression between square brackets has an expected value of zero and can be written as

$$-\frac{h}{L}[Z(x) - Z(0)] + \frac{L-h}{L}[Z(x+h) - Z(x)] - \frac{h}{L}[Z(L) - Z(x+h)]$$

This is a linear combination of independent increments from which we derive

$$\gamma_2(x, x+h) = \frac{1}{2} \left[|h| - \frac{h^2}{L} \right]$$

In this case the residuals are stationary and γ_2 depends only on h , which for $L = 10$ gives

$$\gamma_2(h) = \frac{1}{2} \left[|h| - \frac{h^2}{10} \right]$$

As for the interpolation variances, by applying (2.11), we obtain:

- $\sigma_E^2 = 7/9$ for the six-neighbor pattern using the average of the six data;
- $\sigma_K^2 = 1/2$ for the two-nearest-neighbor pattern.

Figures 2.25a to d display the theoretical variograms superimposed on the experimental graphs. All the experimentally observed results are shown:

1. Positive bias and parabolic behavior of the variogram of raw values, due to the drift.
2. Negative bias of the variogram of residuals leading one to the false conclusion that residuals are uncorrelated, at least at some distance.
3. Linear behavior of the underlying variogram.

Notice that the drift estimator used for calculating $\gamma_2(h)$ is optimal for a linear variogram $\gamma(h)$, which is the case here (cf. Sections 3.4.5 and 3.4.6). The variogram of the residuals is thus corrupted by a bias, even though the estimation of the residuals is optimal. This bias is negligible at small distances but soon explodes as h increases.

Generalized Variogram

Instead of subtracting the drift from the raw data, it is possible to proceed another way. Recall the advantage of the variogram over the covariance. By working on the increment $Z(x+h) - Z(x)$ rather than on the raw variable, the variogram eliminates the problem of determining the mean around which $Z(x)$ will fluctuate. The increment filters any constant component. By iterating this, we obtain a second-order increment $Z(x+2h) - 2Z(x+h) + Z(x)$ that filters any linear component. We can define the generalized variogram (here of order 1), discussed in detail in Section 4.7.1, by

$$\Gamma_1(h) = \frac{1}{6} \text{Var}[Z(x+2h) - 2Z(x+h) + Z(x)]$$

It can be determined experimentally, directly and without bias, even when there is a linear drift. The scaling factor $\frac{1}{6}$ is chosen so that, for a purely flat variogram, $\Gamma_1(h)$ is equal to the nugget effect (for $h \neq 0$). It is clear that if the variogram $\gamma(h)$ is linear, the same is true for $\Gamma_1(h)$. Only the slope changes: it is multiplied by $\frac{2}{3}$.

Since the definition of $\Gamma_1(h)$ includes points at distances h and $2h$, here we can calculate $\hat{\Gamma}_1(h)$ only up to $h = 5$. Apart from $\hat{\Gamma}_1(5)$, calculated from

a single triplet of points, the curve obtained is just about linear with a slope $\frac{1}{3}$, which corresponds to an underlying linear variogram with slope $\frac{1}{2}$, and we have exactly the theoretical variogram $\gamma(h)$ (Fig. 2.25e).

Conclusions

In this example we observed the following facts:

1. The variogram of raw data is unusable because the drift masks the underlying variogram.
2. The variogram of residuals is heavily biased, except at very small distances.
3. With the generalized variogram it is possible to filter the drift and determine the underlying variogram.

We can conclude that the generalized variogram gets us out of the problem. But to calculate this variogram, the data need to be aligned and equally spaced (regular sampling grid, evenly sampled profiles). We have to find other methods for unevenly distributed data. It was precisely this problem that led to the theory of IRF- k , which provides an appropriate solution. But for now let us reconsider in a more general way the various points raised by our example.

2.7.2. Impact of a Drift on the Raw Variogram

As we have just seen, a drift of the form $m(x) = ax + b$ in one dimension introduces a parabolic bias term $\frac{1}{2}a^2h^2$ at the level of the regional variogram. More generally, let us consider in \mathbb{R}^n an RF of the form

$$Z(x) = m(x) + Y(x)$$

where the drift $m(x)$ is a deterministic drift and the ("true") residual $Y(x)$ an SRF or an IRF with zero mean and variogram $\gamma(h)$. The stochastic version $\gamma_R(h)$ of the regional variogram (2.7) associated with the domain D ,

$$\Gamma_R(h) = \frac{1}{2|D \cap D_{-h}|} \int_{D \cap D_{-h}} [Z(x+h) - Z(x)]^2 dx$$

has the expected value

$$E[\Gamma_R(h)] = \gamma(h) + \frac{1}{2|D \cap D_{-h}|} \int_{D \cap D_{-h}} [m(x+h) - m(x)]^2 dx \quad (2.82)$$

Should the drift have some amplitude, the additional term can mask $\gamma(h)$. It is obviously the same for the sample variogram (2.6) calculated from the raw data.

This phenomenon is often reflected in an apparent anisotropy, as can be seen on the raw variogram of Figure 2.26 which was constructed from piezometric data of the Crau aquifer (Delhomme, 1976, 1978). The anisotropic behavior is very marked at large distances and shows a very rapid increase in the NE–SW direction. The large amplitude of the deviations in this direction corresponds to the general orientation of the aquifer in which the groundwater flows from NE to SW with an average hydraulic gradient of 3 to 4 m/km. The anisotropy is related to the variation of the average gradient of the aquifer with the direction. Different effects can be seen with other types of drift; for example, a dome-shaped drift gives a dome-shaped variogram.

In geostatistical calculations, however, it is the variogram $\gamma(h)$ of $Y(x)$, known as the underlying variogram, that is used. When dealing with a repetitive phenomenon, such as meteorology, we may have a large number of similar situations and thus be able to deduce the drift at a monitoring point by averaging the observations at this point. We can then subtract this drift from the data and obtain the “true” residuals directly. This is the approach developed by Gandin (1963, p. 27) in meteorology. But when the phenomenon considered is unique, as is generally the case here, we cannot directly separate the drift from the residual. We still have to find a way of determining the underlying variogram by neutralizing the auxiliary term responsible for the bias in the raw variogram.

2.7.3. Variogram of Residuals

Let us take the approach followed in the introductory example: since the drift introduces a bias in the raw variogram, we subtract the drift and work on the residuals. Or rather, since the drift $m(x)$ is not known exactly, we take an unbiased estimate $\hat{m}(x)$ and subtract it from $Z(x)$ to obtain the (estimated) residuals $R(x_\alpha) = Z(x_\alpha) - \hat{m}(x_\alpha)$ at the N sample points. Let us compute the sample variogram of these residuals, more precisely its stochastic version, defined similarly to (2.6) by

$$\hat{\Gamma}_{\text{Res}}(h) = \frac{1}{2N_h} \sum_{x_j - x_\alpha \simeq h} [R(x_\beta) - R(x_\alpha)]^2$$

Using the notation $Z_\alpha, \hat{m}_\alpha, R_\alpha$ instead of $Z(x_\alpha), \hat{m}(x_\alpha), R(x_\alpha)$ to simplify, the contribution of the sample points x_α and x_β to the sample variogram of residuals has for expected value

$$\begin{aligned} \gamma_{\text{Res}}(x_\alpha, x_\beta) &= \frac{1}{2} E(R_\beta - R_\alpha)^2 \\ &= \gamma(x_\beta - x_\alpha) - \text{Cov}(Z_\beta - Z_\alpha, \hat{m}_\beta - \hat{m}_\alpha) + \frac{1}{2} \text{Var}(\hat{m}_\beta - \hat{m}_\alpha) \end{aligned} \quad (2.83)$$

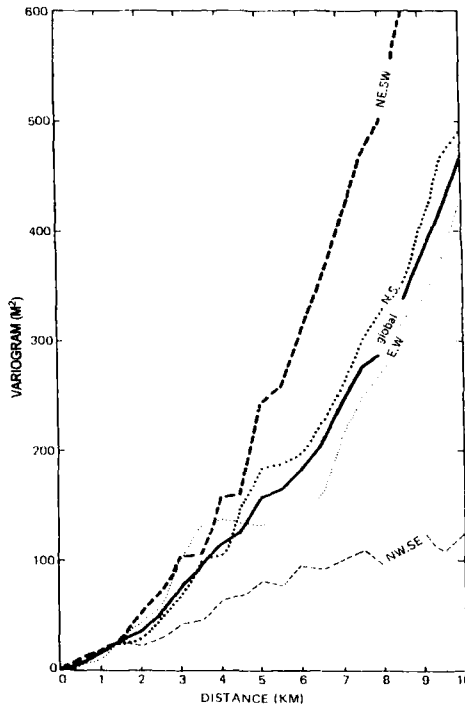


FIGURE 2.26. Raw variogram of the piezometry of the Crau aquifer, characteristic of a NE-SW linear drift. Reprinted from Delhomme (1976, 1978), with permission from BRGM and Elsevier Science.

When the estimators \hat{m}_α and \hat{m}_β are linear, (2.83) is expressed from the variogram $\gamma(h)$. It depends on x_α and x_β separately and not only on $x_\beta - x_\alpha$. Therefore we define the variogram of residuals by $\gamma_{\text{Res}}(h) = E[\hat{\Gamma}_{\text{Res}}(h)]$, namely an average of terms $\gamma_{\text{Res}}(x_\alpha, x_\beta)$. There are obviously as many variograms of residuals as there are ways of estimating the drift. The expression of $\gamma_{\text{Res}}(x_\alpha, x_\beta)$ is only simplified when the *optimal* linear drift estimator $m^*(x)$ is used, as shown in Section 3.4.5 in the framework of the universal kriging model. The central term of the right-hand side of (2.83) is then twice the third term, so that (2.83) becomes

$$\gamma_{\text{Res}}(x_\alpha, x_\beta) = \gamma(x_\beta - x_\alpha) - \frac{1}{2} \text{Var}(m_\beta^* - m_\alpha^*)$$

Thus, even if the estimator of the drift is the optimal one, which means that we already know the underlying variogram $\gamma(h)$ we are looking for, the variogram of the residuals is systematically biased downward. This bias is small at short distances but can be large at medium and large distances. It can lead to the erroneous conclusion that the residuals are uncorrelated.

To identify the underlying variogram $\gamma(h)$, we can think of using relation (2.83) to calculate the theoretical variogram of residuals $\gamma_{\text{Res}}(x_\alpha, x_\beta)$ corre-

sponding to a tentative $\gamma(h)$, and compare this with the sample variogram of the residuals $\hat{\Gamma}_{\text{Res}}(h)$ to identify $\gamma(h)$. To this end we can use for example the ordinary least squares estimator of the drift, which can be calculated easily (it is the optimal one if the underlying variogram is a pure nugget effect). This approach has been tried in the case of regularly spaced data: variogram of residuals “type curves” were constructed in moving windows, typically made of 11 successive points as in the introductory example (Huijbregts and Matheron, 1971; Sabourin, 1976). But the solution is not unique (Matheron, 1970, sec. 4.6), and very different models $\gamma(h)$ can produce very similar variogram of residuals $\gamma_{\text{Res}}(h)$. This difficulty disappears with the generalized variogram (Section 4.7).

2.7.4. A Few Favorable Cases

Despite the presence of a drift it is often possible to find a way of returning to the standard structural analysis of the stationary case.

Very Mild Drift

In the very mild drift case the term that is added to $\gamma(h)$ in (2.82) is negligible at short distances. This allows a good determination of the variogram at this scale, which is often sufficient for the requirements of kriging.

An example of this is found in a study of the Noirétable area topography (France) (Chilès and Delfiner, 1975). The aim of the study was to determine the precision with which one can reconstruct by kriging the topography of a 5 km² area from a pattern of 573 points. Figure 2.27a shows the sample variogram calculated in four directions. Beyond 500 m, the various curves diverge and exhibit a behavior characteristic of a complex-shaped drift becoming appreciable at this scale. But kriging involves neighborhoods that do not exceed 500 m in diameter. We can thus consider the phenomenon as being drift free and isotropic at this scale. There remains to fit a model to the first 500 meters of the average variogram. Several fits were proposed, all of which are suitable and give very close results, in particular, a regularized linear model given by (2.39) with $l = 500$ m (the equality of l and the maximum modeling distance of the variogram has no particular meaning) (op. cit.; Fig. 2.27b) or an $|h|^\alpha$ model with $\alpha = 1.4$ (Journel and Huijbregts, 1978; Fig. 2.27c).

Unidirectional Drift

Often the drift is not felt in all directions. If so and if the hypothesis of an isotropic variogram appears reasonable, we can use an isotropic model fitted to the sample curve obtained for the direction with no drift. We have already seen an example of this with the piezometry of the Crau aquifer (Fig. 2.26).

Another example is provided by Delfiner (1973). It concerns a study of the 500 millibars geopotential, roughly the altitude of the 500 mb atmospheric pressure surface. Because of the rotation of the earth, the geopotential shows no E–W drift. There is, however, a very clear decrease from the equator to

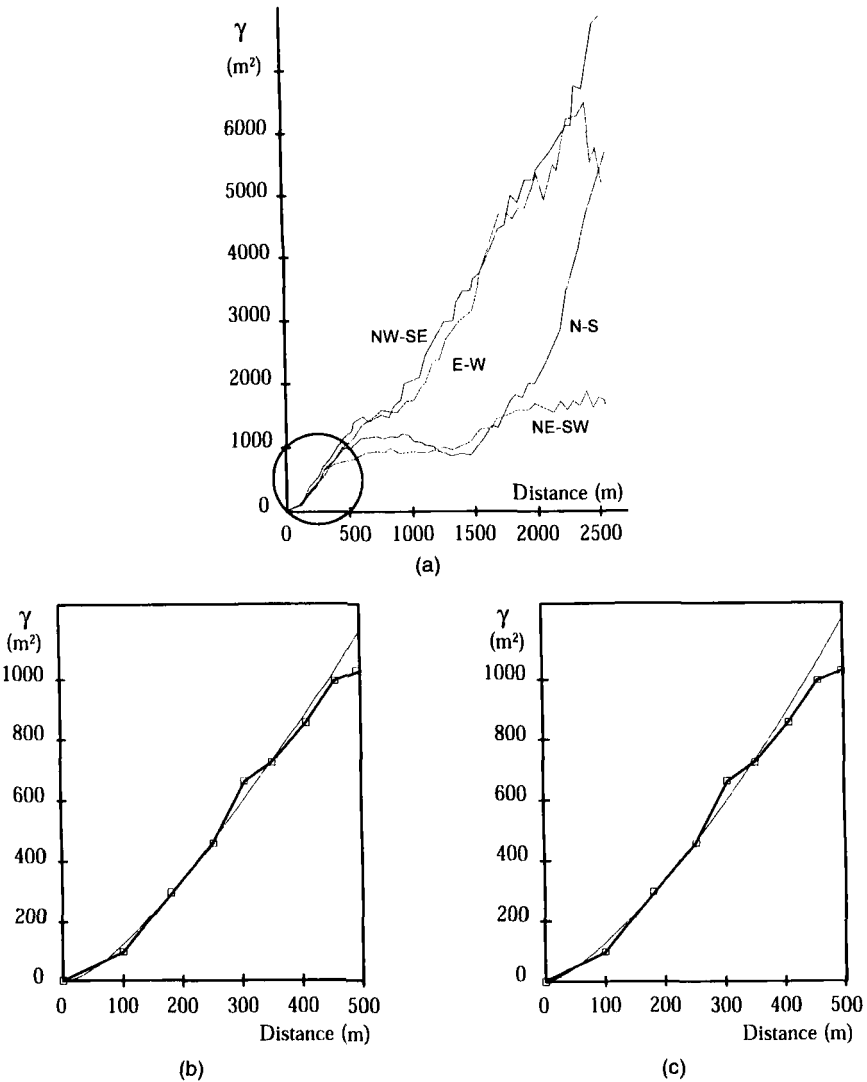


FIGURE 2.27. Topography of the Noiretable area (Massif Central): (a) directional variograms at large distances; (b) average variogram up to 500 m, and fit by a regularized linear model; (c) average variogram up to 500 m, and fit by a model of type $|h|^\alpha$. From Chilès and Delfiner (1975), with permission of SFPT (a, b); Journel and Huijbregts (1978), with permission of Academic Press Limited, London (c).

the Pole; this is a well-known phenomenon reflecting the existence of a zonal wind (the wind is to a first approximation orthogonal to the gradient of the geopotential). The phenomenon is pictured in Figure 2.28a which shows the average value of the data in sections of 5 and 10 degrees of latitude and longitude. It is no surprise that the sample variogram reaches a sill in the

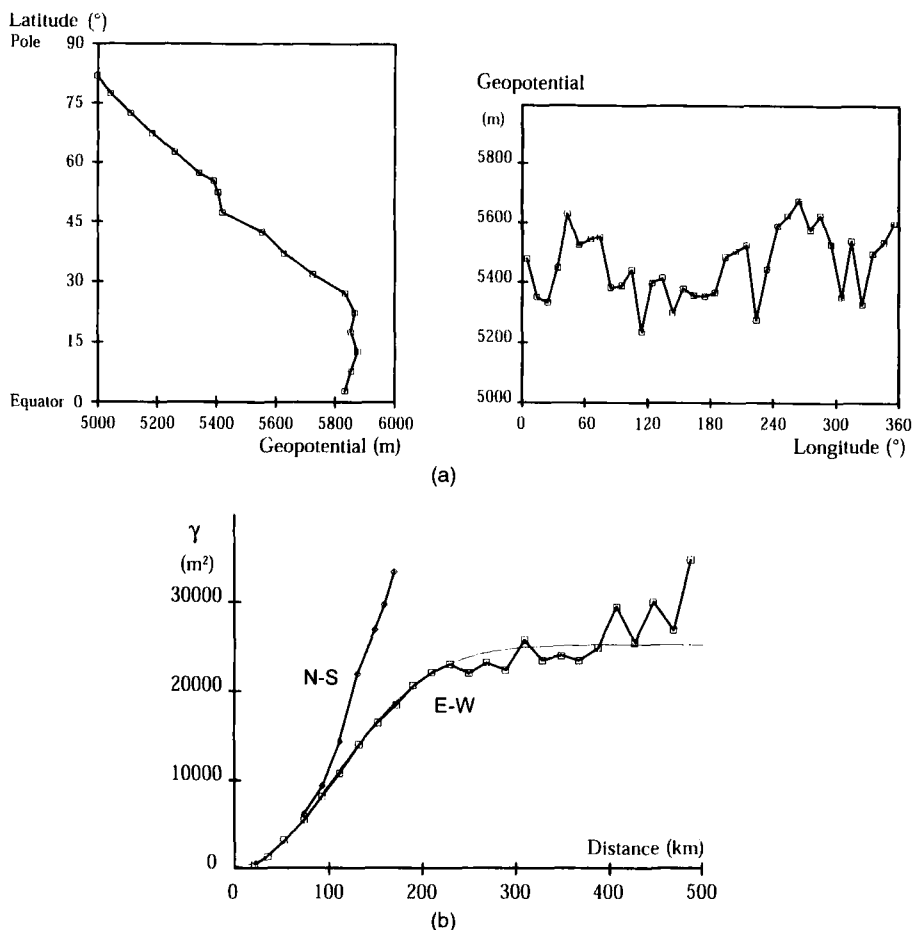


FIGURE 2.28. 500-mb geopotential: (a) profiles of the mean versus longitude and latitude; (b) E-W and N-S variograms. From Delfiner (1973), © SMF.

E-W direction and shows a parabolic behavior in the N-S direction (Fig. 2.28b). Because of the physics of the problem the basic drift functions are necessarily trigonometric functions depending only on the latitude θ . Here one can consider that, at least locally, it is of the form $a + b \cos \theta$. A model fitted to the E-W curve is taken as the variogram, namely a Gaussian model with a scale parameter $a = 1470$ km (the practical range being of the order of 2500 km), a sill $C = 25,000$ m², and a nugget effect $C_0 = 300$ m² (variance of the uncertainty of the radiosonde measurements, difficult to calibrate on the variogram but known by another way).

Notice that in this particular case, the form of the drift is not invariant by translation, unlike usual drifts (cf. Sections 3.4.5 and 3.4.6). But the spherical form of the earth is a good reason to favor spherical coordinates.

Global Drift with a Simple Form

Sometimes the drift has a very simple form at the scale of the entire region of interest, such as linear or polynomial. In such cases we can subtract a global estimate of the drift from the data and work on the residuals. Generally, the least squares estimator is chosen because, regardless of the number of data, it only requires solving a small linear system (as many equations as basis functions). The variogram of residuals is certainly biased, but since this bias is negligible at short distances, we can correctly determine the variogram at the scale of the neighborhoods to be used for subsequent kriging.

To illustrate this, Figure 2.29 shows the raw variogram and variograms of residuals for civil engineering microgravimetric data (Bouguer anomaly sampled on a 15 m grid, locally refined (Chilès, 1979b)). The study zone is a 380-m \times 350-m rectangle. The average raw variogram is clearly parabolic and reflects an essentially linear NW–SE global drift which is evident from the data. The directional raw variograms (not reproduced here) show, however, that no direction is entirely free of drift. As the drift has a good overall shape the sample variograms of residuals were calculated for polynomial drifts of degree 1, 2, and 3 determined by ordinary least squares. They show that in order to eliminate the effect of the drift, we must consider a global drift of degree 2, if not 3. The variogram has a range of 60 m, or about one-sixth of the length of the domain. In this case the variogram of residuals is practically unbiased up to half the length of the domain (this is apparent when constructing simulations without drift using the techniques presented in Chapter 7, and from comparing the simulation variogram with the simulated residuals variogram). We obtain a good fit of γ_3 (variogram of residuals associated with a drift of degree 3) by a spherical model with range $a = 60$ m and sill $C = 430 \mu\text{gal}^2$. To this we must add a measurement error variance of $4 \mu\text{gal}^2$, not perceptible on the sample variogram; geophysicists have told us, however, that the measurements are made with an error of $\pm 4 \mu\text{gal}$ (hence a standard deviation of $2 \mu\text{gal}$ and a variance of $4 \mu\text{gal}^2$). To cross-validate the fit, 146 sample points were estimated from their neighbors and gave a mean standardized square error of 1.06, which is excellent.

2.8. SIMPLE APPLICATIONS OF THE VARIOGRAM

We conclude this chapter with simple applications of the variogram to variance calculations and sampling design. Here the form of the estimator is fixed, as when practically dictated by symmetries. We wish to compute (1) elementary estimation variances, such as the variance of estimation of a square or rectangular block by a central sample or by four samples located at the vertices, and (2) a global estimation variance when the data are uniformly located within the domain of study so that the sample mean can be taken as a sensible estimator of the regional mean. The variogram also enables the calculation of

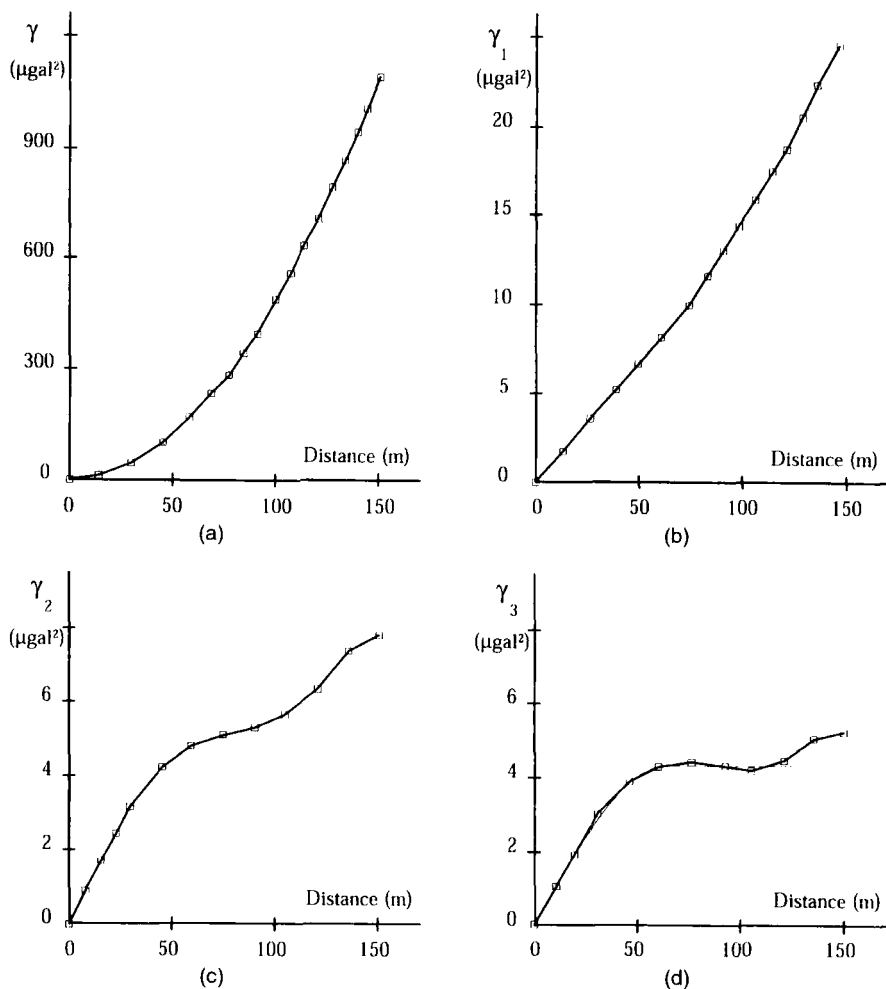


FIGURE 2.29. Microgravimetry in a quarry: (a) raw variogram; and variogram of residuals obtained after subtracting a global drift of degree k estimated by least squares: (b) $k = 1$; (c) $k = 2$; (d) $k = 3$. Notice the change in vertical scale between the raw variogram and the variograms of residuals. From Chilès (1979b).

dispersion variances of small units partitioning larger domains. These results are very useful to design a sampling scheme. In mining, for example, one may wish to achieve a given accuracy for a global estimation or to optimize a pre-exploitation grid and control the local dispersion of the average grade of elementary mining units. It is assumed here that the phenomenon has no drift. The results remain valid in the case of a linear drift if symmetries in the sampling pattern ensure that a linear component is estimated without error.

2.8.1. Estimation Variance

Let us begin with a very simple case, considering only two locations, x and $x + h$, and assuming that we know $Z(x)$ but not $Z(x + h)$; in the absence of any other information, we assign location $x + h$ the known value of location x . Since $2\gamma(h)$ is equal to the variance of $Z(x + h) - Z(x)$, $2\gamma(h)$ quantifies the error incurred when estimating $Z(x + h)$ by $Z(x)$. More generally, if we estimate the average value in a domain V

$$Z_V = \frac{1}{|V|} \int_V Z(x) dx$$

by the average of N sample points x_α ,

$$\hat{Z} = \frac{1}{N} \sum_{\alpha=1}^N Z(x_\alpha)$$

we obtain, by extending (2.10) to the case of stochastic integrals,

$$\begin{aligned} \text{Var}(\hat{Z} - Z_V) &= -\frac{1}{N^2} \sum_{\alpha=1}^N \sum_{\beta=1}^N \gamma(x_\beta - x_\alpha) + \frac{2}{N|V|} \sum_{\alpha=1}^N \int_V \gamma(x_\alpha - x) dx \\ &\quad - \frac{1}{|V|^2} \int_V \int_V \gamma(x' - x) dx dx' \end{aligned} \quad (2.84)$$

This result alternates exact and approximated expressions of the same integral, with the result that the estimation variance becomes smaller as:

- the network of sample points is tighter and more regular;
- the variogram is more regular, namely the variable itself is smoother in its spatial variation.

Such a variance is called an *estimation variance*. It is also referred to as an *extension variance* when one extends the value measured at one point, or on a small volume v , to the volume V . Denoting by

$$Z_v = \frac{1}{|v|} \int_v Z(x) dx$$

the average value of $Z(x)$ in v , (2.84) takes on the form

$$\begin{aligned} \text{Var}(Z_v - Z_V) &= -\frac{1}{|v|^2} \int_v \int_v \gamma(x' - x) dx dx' + \frac{2}{|v||V|} \int_v \int_V \gamma(x' - x) dx dx' \\ &\quad - \frac{1}{|V|^2} \int_V \int_V \gamma(x' - x) dx dx' \end{aligned} \quad (2.85)$$

The estimator \hat{Z} is seldom used for local estimation. It is preferable to consider samples in V and its immediate vicinity and to apply some weighting (kriging). For global estimation, however, if the data are evenly distributed so that they have approximately similar zones of influence, kriging will give approximately equal weight to all data points, and the estimator \hat{Z} can be used. The estimation variance can then be calculated by formula (2.84). But if N is large and/or the domain V has a complex shape, which is often the case in real applications, this formula is quite cumbersome. A variety of approximations have been proposed to handle the most common cases encountered in the global estimation of an orebody, a forest, or an agricultural field. They allow the calculation of the estimation variance of the resource in any 2D or 3D volume from systematic sampling, not necessarily on a square grid. For example, it is possible to deal with line sampling (along boreholes or transects) by cutting V into slices centered on the lines and combining errors incurred by estimating the lines from the samples, and the slices from the lines. The interested reader is referred to Matheron (1965, 1970), David (1977), and Journel and Huijbregts (1978) for a thorough description of the methods.

2.8.2. Dispersion Variance

Let us consider a domain V that can be partitioned into N cells v_i which are identical to a cell v up to a translation. Let us also consider a particular realization $z(x)$ of the IRF $Z(x)$. If the mean values $z(v_i)$ of $z(x)$ in the cells v_i are all known, the mean of the N values $z(v_i)$ is simply the mean value $z(V)$ in the domain V . We therefore define the sample variance of v_i in V by

$$s^2(v | V) = \frac{1}{N} \sum_{i=1}^N [z(v_i) - z(V)]^2 \quad (2.86)$$

This variance measures the dispersion of the values in the cells v_i partitioning V . It can also be viewed as the variance of the estimation of $z(V)$ by $z(v_i)$, the cell v_i being chosen at random among the N cells partitioning V .

Now, if the values $z(v_i)$ are not available, $s^2(v | V)$ cannot be calculated, but we can obtain its expected value $\sigma^2(v | V)$ by randomizing (2.86) with respect to the realization. This is equal to the expected value of the extension variance of $Z(v_i)$ to $Z(V)$, the cell v_i being chosen at random, or equivalently, to the mean of the N extension variances similar to (2.85) (with v_i instead of v). Because v_i is a partition of V , the cross term in (2.85) is, on the average, equal to the third term of the right-hand side, so that the variance of $Z(v_i)$ in V becomes

$$\sigma^2(v | V) = \frac{1}{|V|^2} \int_V \int_V \gamma(x' - x) dx dx' - \frac{1}{|v|^2} \int_v \int_v \gamma(x' - x) dx dx' \quad (2.87)$$

where $\sigma^2(v | V)$ is called the dispersion variance of v within V . For usual variogram models the second integral in (2.87) increases with the volume v of the support, hence the commonly observed decreasing variance-with-volume relationship. In the case where v is limited to a single point, the preceding expression is reduced to

$$\sigma^2(0 | V) = \frac{1}{|V|^2} \int_V \int_V \gamma(x' - x) dx dx' \quad (2.88)$$

which generalizes the sample variance (2.5). If the variogram $\gamma(h)$ reaches a sill and if the domain V is very large with respect to the range, the variance of $Z(x)$ in V approaches the global point variance $\gamma(\infty) = C(0)$. This does not hold true for unbounded variograms.

When v is not a point and V cannot be partitioned into identical cells v_i , formula (2.87) is taken as a definition of the dispersion variance of v within V , but it no longer has a precise significance. It can even be negative, for example one has $\sigma^2(V | v) = -\sigma^2(v | V)$.

Dispersion variances satisfy the following *additivity property*: if $v \subset V \subset \mathcal{V}$, then

$$\sigma^2(v | \mathcal{V}) = \sigma^2(v | V) + \sigma^2(V | \mathcal{V}) \quad (2.89)$$

(2.89) is sometimes called “Krigé’s relationship” and is similar to an analysis of variance formula (note that this property is also valid among experimental dispersion variances).

From a practical point of view, it should be remembered that the double integrals over the same domain, which we find in expressions (2.84), (2.85), (2.87), and (2.88), can be expressed as simple integrals using the Cauchy algorithm. Analytical expressions for the various simple or double integrals can be obtained in 2D for rectangular cells for simple models such as the spherical one. In 3D a symbolic calculation of these integrals has been considered by Marbeau and Marbeau (1989). The value of these integrals can also be read from graphs (e.g., see Journel and Huijbregts, 1978, pp. 125–147, for a spherical or exponential variogram in 2D and 3D).

2.8.3. Sampling Design

The preceding results enable us to design sampling patterns, namely to choose the locations of N points x_i that will be sampled for estimating a given domain V . This is possible because the estimation variance does not depend on the unknown values $Z(x_i)$ but only on their location and the variogram. Of course some data are necessary to identify the variogram. But once this is known, it is possible to predict the estimation variance for any proposed sampling pattern and thus optimize it to achieve a desired precision. We begin with an elementary example and then compare three main patterns: random sampling, stratified random sampling, and regular grid. Naturally in real applications the

TABLE 2.2. Estimation variance of an interval of length L from one central sample or two extreme samples for a variogram $\gamma(h) = |h|^\alpha$

α	σ_1^2	σ_2^2
0	1	0.5
0.5	$0.409 L^{1/2}$	$0.300 L^{1/2}$
1	$0.167 L$	$0.167 L$
1.5	$0.054 L^{3/2}$	$0.071 L^{3/2}$
2	0	0

design of a sampling strategy is largely constrained by practical considerations. These can be dealt with, to some extent, by considering kriging estimators.

A Simple Example

Suppose that one is interested in the mean value of some parameter, say temperature, over one-hour period; which of these two estimates is better: the temperature at half the hour or the average of two consecutive measurements on the hour?

The problem amounts to the estimation of the average of $Z(x)$ over the segment $[0, L]$ either by the estimator $\hat{Z}_1 = Z(L/2)$ or by $\hat{Z}_2 = \frac{1}{2}[Z(0) + Z(L)]$. The corresponding estimation variances are given by formula (2.84). To express them in a simple manner, let us introduce two auxiliary functions of the variogram $\gamma(h)$:

$$\begin{aligned}\chi(h) &= \frac{1}{h} \int_0^h \gamma(u) du \\ F(h) &= \frac{1}{h^2} \int_0^h \int_0^h \gamma(x' - x) dx dx' = \frac{2}{h} \int_0^h u \gamma(u) du\end{aligned}\quad (h > 0)$$

The two variances of estimation, σ_1^2 from a single central sample, σ_2^2 from two extreme samples, are then

$$\sigma_1^2 = 2\chi\left(\frac{L}{2}\right) - F(L) \quad \sigma_2^2 = 2\chi(L) - F(L) - \frac{\gamma(L)}{2}$$

Let us assume a variogram of type $\gamma(h) = b|h|^\alpha$. The auxiliary functions take on the form

$$\chi(h) = \frac{bh^\alpha}{\alpha + 1} \quad F(h) = \frac{2bh^\alpha}{(\alpha + 1)(\alpha + 2)}$$

which allows the calculation of σ_1^2 and σ_2^2 . The results are given in Table 2.2 for typical values of α and $b = 1$.

We observe the following:

- For $\alpha = 0$, the limiting case of a pure nugget effect, only the number of samples matters so that σ_2^2 is half σ_1^2 .
- This advantage decreases when α increases, and for $\alpha = 1$ the two patterns are equivalent.
- As α increases beyond 1, the variogram becomes more regular, and a single centrally located sample becomes better than two ill-placed ones.
- For $\alpha = 2$, which would correspond to a linear random function, the variances are zero.

Coming back to the initial problem, we conclude that in a turbulent environment ($\alpha = 2/3$) it is better to use the pattern with two measurements, whereas a centrally located sample is better in the case of smooth temperature variations ($\alpha > 1$).

Pure Random Sampling

Consider now the more general problem of the design of a pattern for estimating the average value Z_V of $Z(x)$ over a domain V . The survey will include N samples $Z(x_i)$, $i = 1, \dots, N$, and Z_V will be estimated by $\hat{Z} = (1/N) \sum_{i=1}^N Z(x_i)$. How should this pattern be selected?

A first possibility is a random pattern where the samples are randomly scattered within V . More specifically, the x_i are independently located in V with uniform density $1/|V|$. An elegant way to derive the estimation variance is the following: For a given realization $z(x)$, the values $z(X_i)$ are random through X_i and are independent. Since X_i is uniformly distributed within V , $z(X_i)$, considered as an estimator of z_V , satisfies

$$\begin{aligned} E[z(X_i) - z_V] &= 0 \\ E[z(X_i) - z_V]^2 &= s^2(0 | V) \end{aligned}$$

where $s^2(0 | V)$ is the variance of $z(x)$ within V . Since the estimator \hat{z} is the mean of the $z(X_i)$ and the X_i are independent, \hat{z} is unbiased and its estimation variance is $(1/N)s^2(0 | V)$. Randomizing the realization, we find that

$$\sigma_{\text{Rand}}^2 = E(\hat{Z} - Z_V)^2 = \frac{1}{N} \sigma^2(0 | V)$$

Note that this variance is that of the random pattern *before* the locations of the samples have been chosen. Once the pattern is fixed, the correct estimation variance is that of formula (2.84). The preceding result can then be used as an approximation, since it is equal to the average of the estimation variances associated with all possible realizations of the set of sampling points.

Stratified Random Sampling

This time V is divided into N similar disjoint zones of influence v_i . Within each v_i a sample is placed at random with uniform density and independently of other samples. For a given realization $z(x)$, the error is

$$\hat{z} - z_V = \frac{1}{N} \sum_{i=1}^N [z(X_i) - z_{v_i}]$$

The partial errors $z(X_i) - z_{v_i}$ are independent. By a similar argument to that just seen, we find after randomization of the realization

$$\sigma_{\text{Stra}}^2 = E(\hat{Z} - Z_V)^2 = \frac{1}{N} \sigma^2(0 | v)$$

The variance of estimation has the same form as in the purely random case, except that $\sigma^2(0 | V)$ is replaced by $\sigma^2(0 | v)$, namely the variance of a point sample within its zone of influence v . Now by the additivity relationship (2.89) we have

$$\sigma^2(0 | V) - \sigma^2(0 | v) = \sigma^2(v | V) \geq 0$$

which proves that the stratified random pattern is always more efficient than the purely random pattern.

Square Grid

In this pattern the sampling is performed at the nodes of a regular grid with square or cubic cells if the variogram is isotropic, or on a grid with a direction and a ratio of elongation adapted to the geometric anisotropy of the variogram. Since in the latter case isotropy can be restored by means of a linear transformation of the coordinates, we will only consider the isotropic case. The domain V can therefore be partitioned into square or cubic cells v_i with a sample at the center of each cell. The estimation variance should be computed by application of formula (2.84), but this is usually too cumbersome. Since the estimation error is of the form

$$\hat{Z} - Z_V = \frac{1}{N} \sum_{i=1}^N [Z(x_i) - Z_{v_i}]$$

Matheron (1965, sec. XII-3) proposes an approximation principle that consists in assuming that these partial errors are uncorrelated. This approximation has been shown to be quite good for usual isotropic variogram models provided that we use a square grid. Then

$$\sigma_{\text{Grid}}^2 = E(\hat{Z} - Z_V)^2 \simeq \frac{1}{N} \sigma_E^2(0, v)$$

where $\sigma_E^2(0, v)$ is the extension variance of a central sample to a cell v : the estimation variance is simply computed by dividing the elementary extension variance of a sample to its zone of influence by the number of samples N .

For the usual variogram models $\sigma_E^2(0, v)$ is smaller than $\sigma^2(0 | v)$, so that we finally have

$$\sigma_{\text{Grid}}^2 \leq \sigma_{\text{Stra}}^2 \leq \sigma_{\text{Rand}}^2$$

Comparison of the Three Patterns

To get an idea of orders of magnitude, let us consider the 2D case where v is a square of side l . For the unit square $S_1 = [0, 1] \times [0, 1]$ we have

$$\int_{S_1} |x| dx = \frac{\sqrt{2} + \log(1 + \sqrt{2})}{3} = 0.765$$

$$\int_{S_1} \int_{S_1} |x' - x| dx dx' = \frac{2 + \sqrt{2} + 5 \log(1 + \sqrt{2})}{15} = 0.521$$

so that by application of (2.88) and (2.84) we obtain for the square of side l and the linear variogram $\gamma(h) = b|h|$,

$$\sigma^2(0 | v) = 0.521bl \quad \sigma_E^2(0, v) = 0.244bl$$

The ratio $\sigma^2(0 | v) / \sigma_E^2(0, v) = 2.139$ shows that for a linear variogram the regular grid pattern is twice more efficient than the stratified random pattern, itself more efficient than the purely random scheme.

In order to have a complete comparison between the sampling patterns, let us further assume that the domain V is itself a square with side $l\sqrt{N}$. Then the variance of a sample within V is

$$\sigma^2(0 | V) = 0.521bl\sqrt{N}$$

We can express the three estimation variances in terms of $\sigma^2 = \sigma^2(0 | V)$ and the number of samples N which are two independent parameters:

$$\sigma_{\text{Rand}}^2 = \frac{\sigma^2}{N} \quad \sigma_{\text{Stra}}^2 = \frac{\sigma^2}{N^{3/2}} \quad \sigma_{\text{Grid}}^2 \simeq \frac{\sigma^2}{2.139N^{3/2}}$$

We can appreciate the benefit of a sampling pattern exploiting spatial correlations over crude random sampling: the variance is reduced by a factor of $N^{3/2}$ instead of N .

The same type of comparison can be made in the case of a bounded variogram. But the results now depend on an additional parameter: the range a of the variogram. Let us focus on the comparison between the stratified

random sampling and the regular grid for a spherical isotropic variogram. The ratio $\sigma^2(0|v)/\sigma_E^2(0,v)$ is always greater than 1, which proves that the regular grid performs always better than stratified random sampling. However, this advantage fades away as the size l of the grid cell becomes large with respect to the range a of the variogram: the ratio decreases from 2.14 for $l/a = 0$ to 1 for $l/a = \infty$, with an intermediate value 1.61 for $l/a = 1$. The reason is that for a large l/a ratio, the influence of a sample is purely local anyway, and the center of the square loses its strategic superiority.

The above considerations address an idealized sampling problem that came up at the early stages of the development of mining geostatistics, when statisticians, uninformed of Matérn's work in forestry (1960), were trying to force random sampling on mining engineers who preferred systematic grids. In reality the optimum data collection plan is very application-specific. It depends on the objective pursued, which may not just be the reduction of uncertainty (e.g., in petroleum exploration the placement of a new well can be optimized for production purposes rather than to improve the estimation of reserves). Practical constraints are also important, such as cost or access (what if there is a large rock where a soil sample is supposed to be collected?). The optimal sampling further depends on the estimator selected. In the above comparison we have just used the plain mean, but we could also consider more sophisticated estimators, such as kriging estimators, and calculate estimation variances for different sample placements. The interested reader may consult Rodríguez-Iturbe and Mejía (1974) for the design of rainfall networks, Bras and Rodríguez-Iturbe (1985) for sampling in hydrogeology, Gilbert (1987) for pollution monitoring, and Cressie (1991, sec. 5.6) for an overview of approaches.

2.9. COMPLEMENTS: THEORY OF VARIOGRAM ESTIMATION AND FLUCTUATION

Estimating and modeling the variogram pose problems that pertain more to epistemology than to mathematical statistics. These problems have attracted the attention of geostatisticians since the beginning (Jowett, 1955b; Matheron, 1965, ch. 13) and have fostered extensive studies (see Alfaro, 1979, 1984, and the important methodological work of Matheron, 1978). Without going into details that are beyond the scope of this book, we will outline the main points. We consider here random functions without drift. In the case of a drift, similar results are obtained with the generalized variogram.

2.9.1. Estimation of the Regional Variogram

Regional Variogram and Sample Variogram

In practice, we are faced with a regionalized variable $z(x)$ that we are studying in a bounded domain D . By interpreting $z(x)$ as a realization of an RF $Z(x)$, we have provided a theoretical definition of the variogram $\gamma(h)$. Generally, however, we study a unique phenomenon (e.g., an orebody), and it is primarily the variogram of the regionalized variable in D that we are interested

in. This variogram is the regional variogram defined by (2.7):

$$\gamma_R(h) = \frac{1}{2|D \cap D_{-h}|} \int_{D \cap D_{-h}} [z(x+h) - z(x)]^2 dx$$

where D_{-h} represents the translate of set D by vector $-h$, $D \cap D_{-h}$ the set of points such that x and $x+h$ belong to D , and $|D \cap D_{-h}|$ the measure of this set (i.e., the geometric covariogram of D for the distance h). The regional variogram is a purely deterministic and empirical quantity. If we know $z(x)$ at every point of D , γ_R is completely determined. It constitutes a summary of the structural characteristics of the regionalized variable and, in this sense, conveys a physical significance independently of the probabilistic interpretation that we can construct.

In practice, $z(x)$ is only known at a certain number of sample points x_α , $\alpha = 1, \dots, N$. As the regional variogram cannot be determined directly, we calculate the sample variogram (2.6),

$$\hat{\gamma}(h) = \frac{1}{2N_h} \sum_{x_j - x_\alpha \approx h} [z(x_j) - z(x_\alpha)]^2$$

where the sum is extended to the N_h pairs (x_α, x_j) of sample points separated (approximately) by the vector h .

Is this sample variogram a good approximation of the regional variogram? In rare cases it is possible to provide an experimental answer to this question. One such case is Narboni's (1979) exhaustive survey of the Ngolo tropical forest. The forest was subdivided into 50-m \times 50-m plots and the number of trees was counted in each plot. As a result the regional variogram of the variable "number of trees per plot" can be calculated exactly.

The study area contains about 100 transects each with 200 plots so that, by considering only one transect in 10, one obtains a 10% sampling. By varying the position of the first transect, it is possible to simulate ten different sampling choices. Figure 2.30a shows the ten corresponding sample variograms, whose average gives the regional variogram (the variograms are along the direction of the transects). These ten variograms are all similar, even though their sills deviate by as much as 10% or 15% from the average sill. So with a 10% sampling one can get a good estimate of the regional variogram.

Figure 2.30b shows the sample variograms for different sampling rates: 100% (the regional variogram, average of the curves from Fig. 2.30a), 10%, 5%, and 1%. The deterioration is obvious at the 1% sampling rate (a single transect sampled). This is also seen from the variance of the samples: the variance obtained on sampling a single transect can vary, depending on which transect is selected, between 4.4 and 11.4. With a 10% rate, however, the variance varies only between 7.2 and 9.0 for the ten possible sampling choices. Notice that the overall aspect of the sample variograms is correct, which is the most important: it is only the value of the sill that is poorly estimated. As we will see in Section 3.4.1, this inaccuracy only affects the kriging variance and not the estimate itself.

Estimation Variance of the Sample Variogram

Exhaustive sampling situations are exceptional, and we must consider the general case where the regional variogram γ_R cannot be determined experimentally. If we fix h and let

$$q_h(x) = \frac{1}{2}[z(x+h) - z(x)]^2$$

we see that $\gamma_R(h)$ is simply the average value of $q_h(x)$ over $D \cap D_{-h}$, and that $\hat{\gamma}(h)$ is the average value of the N_h data $q_h(x_i)$, where the x_i constitute a subset of the x_α . It is reasonable to expect that if sufficient data are available and fairly well distributed, $\hat{\gamma}(h)$ is close to $\gamma_R(h)$. If, for example,

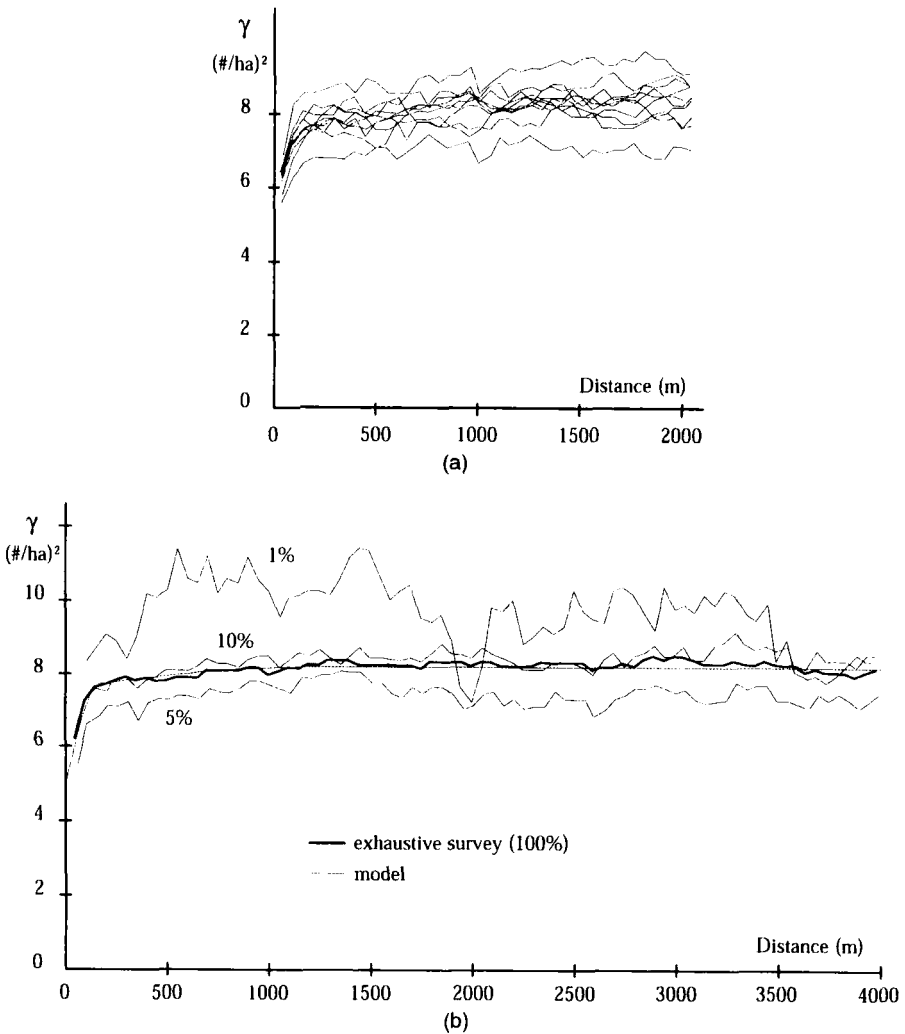


FIGURE 2.30. Survey of the Ngolo forest: (a) variograms of 10 sampling choices at a rate of 10%; (b) examples of variogram for sampling rates of 100% (average of the ten preceding variograms), 10%, 5%, and 1%. From Narboni (1979), © Bois et Forêts des Tropiques.

the data are on a regular grid, and if h is a multiple of the grid spacing, $\hat{\gamma}(h)$ is simply the discrete approximation of the integral defining $\gamma_R(h)$.

To go further, we have to determine the behavior of the regionalized variables $z(x)$ and $q_h(x)$. In our models we interpret $z(x)$ as the realization of an RF $Z(x)$, and then $q_h(x)$ is a realization of the RF

$$Q_h(x) = \frac{1}{2} [Z(x+h) - Z(x)]^2$$

We assume here that $Z(x)$ is an IRF with variogram $\gamma(h)$ and that the RF $Q_h(x)$ has second-order moments and is stationary. Let $G_h(x' - x)$ denote the covariance of $Q_h(x)$ and $Q_h(x')$. To avoid

confusion, Γ_R and $\hat{\Gamma}$ will now denote the random versions of the regional variogram and the sample variogram, or explicitly

$$\begin{aligned}\Gamma_R(h) &= \frac{1}{2|D \cap D_{-h}|} \int_{D \cap D_{-h}} [Z(x+h) - Z(x)]^2 dx \\ &= \frac{1}{|D \cap D_{-h}|} \int_{D \cap D_{-h}} Q_h(x) dx\end{aligned}\quad (2.90)$$

$$\hat{\Gamma}(h) = \frac{1}{2N_h} \sum_{x_j - x_i \simeq h} [Z(x_j) - Z(x_i)]^2 = \frac{1}{N_h} \sum_{x_j - x_i \simeq h} Q_h(x_i) \quad (2.91)$$

Naturally

$$E[\Gamma_R(h)] = E[\hat{\Gamma}(h)] = \gamma(h)$$

since $Q_h(x)$ has for expectation $\gamma(h)$. We can thus characterize the error incurred by taking the sample variogram for the regional variogram by the variance of $\hat{\Gamma}(h) - \Gamma_R(h)$. Considering their respective definitions (average of N_h values $Q_h(x_i)$ for the one, average of $Q_h(x)$ over $D \cap D_{-h}$ for the other), brings us back to a standard calculation of estimation variance to be carried out with the covariance $G_h(x' - x)$ of the RF $Q_h(x)$.

$G_h(x' - x)$, however, is a fourth-order moment of the RF $Z(x)$. Determining the precision of the calculation of the second-order moment thus requires prior knowledge of the fourth-order moment. But the latter can generally only be evaluated with mediocre precision, related to the eighth-order moment, and so on. This way the problem can be displaced endlessly. We can nevertheless determine orders of magnitude by considering classic cases of spatial distribution. The Gaussian case will be examined first, and then indications will be given for random functions with skewed marginal distributions.

Gaussian Case

We know that if U_1 and U_2 are joint Gaussian random variables with zero mean and covariance σ_{12} , the centered covariance of U_1^2 and U_2^2 is

$$\text{Cov}(U_1^2, U_2^2) = 2\sigma_{12}^2$$

From this we can deduce that for an IRF $Z(x)$ with bi-gaussian increments, the covariance $G_h(x' - x)$ is given by

$$G_h(x' - x) = \frac{1}{2} [\gamma(x' - x + h) - 2\gamma(x' - x) + \gamma(x' - x - h)]^2 \quad (2.92)$$

This expression allows us to calculate the estimation variance of the variogram. Matheron (1965, pp. 229–230) shows that when the data are on a regular grid, we have to a first approximation

$$\text{Var}[\hat{\Gamma}(h) - \Gamma_R(h)] = 4\gamma(h)\sigma_h^2$$

where σ_h^2 denotes the variance of estimation of the mean value of $Z(x)$ in $D \cap D_{-h}$ from the N_h data $Z(x_i)$ involved in the expression (2.91) of $\hat{\Gamma}(h)$. Thus, provided that there are enough pairs of points to calculate $\hat{\gamma}(h)$, we are sure of being able to evaluate $\gamma_R(h)$ with good precision: $\hat{\gamma}$ is a consistent estimator of γ_R . Figure 2.31 shows an example of the relative standard devi-

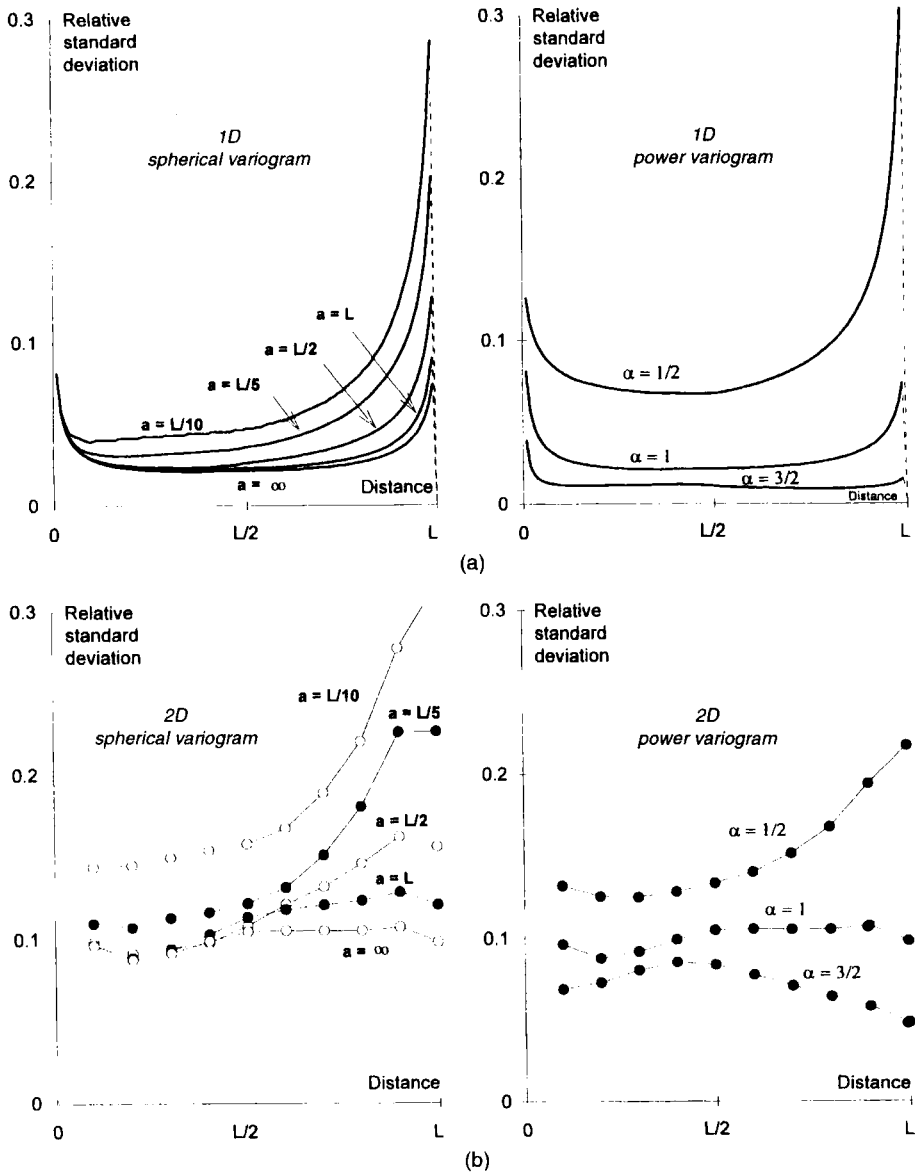


FIGURE 2.31. Estimation of the regional variogram; relative standard deviation in the Gaussian case for a spherical variogram with range a (on the left) and for a variogram of type $|h|^\alpha$ (on the right): (a) along a segment of length L sampled on a regular grid by 101 points; (b) on a square $L \times L$ sampled by 11×11 points.

ation

$$\frac{\sqrt{\text{Var}[\hat{\Gamma}(h) - \Gamma_R(h)]}}{\gamma(h)}$$

in the case of some 100 data on a regular grid in \mathbb{R}^1 or in \mathbb{R}^2 , and for a spherical variogram or a variogram of type $|h|^\alpha$. It is seen that:

- the precision improves with the regularity of the variogram (large range or high α);
- the precision in \mathbb{R}^2 is practically of the same order of magnitude for all values of h , whereas in \mathbb{R}^1 the first points are relatively poorly estimated.

General Case

To give a glimpse of the precision that one can expect with non-Gaussian RFs, let us consider three different RFs derived from independent standard Gaussian RFs $U(x)$ and $V(x)$ with the same correlogram $\rho(h)$:

1. $Z_1(x) = U(x)V(x)$. Its marginal distribution has the probability density function $f(z) = (1/\pi)K_0(|z|)$, where K_0 is the order-0 Bessel function of the second kind; its covariance is $C(h) = \rho(h)^2$.
2. $Z_2(x) = U(x)^2$. An RF with a gamma marginal distribution with parameters $\frac{1}{2}$ and $\frac{1}{2}$ (chi-square distribution on one degree of freedom); its covariance is $C(h) = 2\rho(h)^2$ and its coefficient of variation is $\sqrt{2} = 1.414$.
3. $Z_3(x) = e^{U(x)}$. A lognormal RF with covariance $C(h) = e[e^{\rho(h)} - 1]$, thus with a fairly high coefficient of variation ($\sqrt{e-1} = 1.311$).

The three RFs were chosen because it is still relatively easy to calculate the covariance $G_h(x' - x)$ from the results relative to the Gaussian RFs (see Alfaro, 1979). It is very easy to obtain an exponential covariance for the first two cases ($\rho(h)$ must also be exponential with twice the scale parameter). As for the lognormal RF, we have given it a covariance as close as possible to the exponential model. Figure 2.32 shows the results for this type of model. Three conclusions can be drawn:

1. The variogram is less precise than in the Gaussian case.
2. It is acceptable for distributions such as those of Z_1 and Z_2 which are not too long tailed.
3. The sample variogram of a variable with a long-tailed distribution, like Z_3 , bears only a distant relationship to the regional variogram.

This last result is linked to a well-known phenomenon: the presence of proportional effect. When the data have a long-tailed histogram, these findings lead one to consider either more robust variogram estimators or transformations, with the problems inherent to these techniques.

2.9.2. Modeling the Regional Variogram

Fluctuation Variance of the Regional Variogram

Even if we knew the value of the regionalized variable $z(x)$ at every point of the studied domain D , and were capable of calculating the regional variogram $\gamma_R(h)$ for any vector h , this would exhibit so many variations of detail that we would have to simplify it to be able to express it in a usable form. This amounts to considering that two very similar regional variograms have the

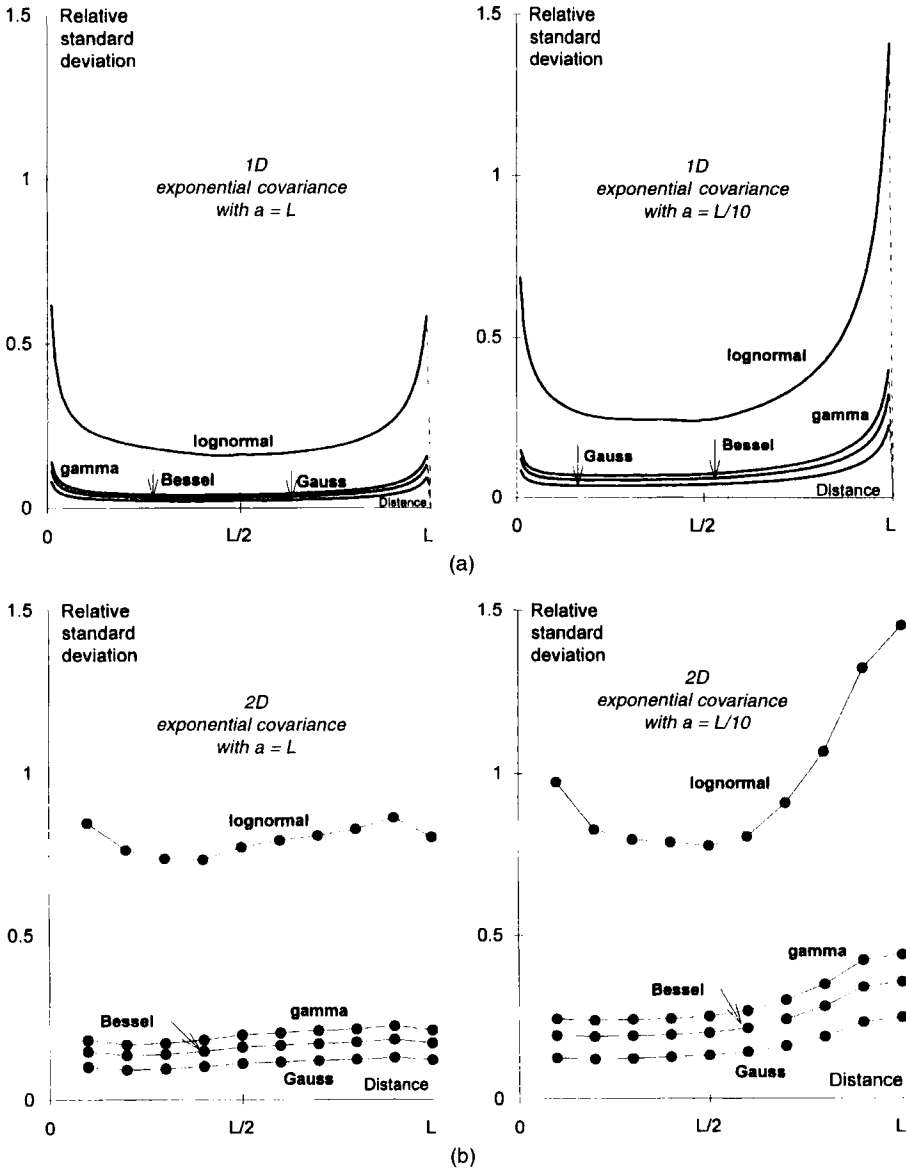


FIGURE 2.32. Estimation of the regional variogram: relative standard deviation for different spatial distributions, for an exponential variogram with scale parameter L (on the left) and $L/10$ (on the right): (a) along a segment of length L sampled on a regular grid by 101 points; (b) on a square $L \times L$ sampled by 11×11 points.

same parent variogram $\gamma(h)$. This simplification represents exactly a passage to the mathematical expectation: if one considers the studied regionalized variable as the realization of an IRF, namely as one realization among a set of similar realizations, the regional variogram of the regionalized variable is one among a family of regional variograms whose mean, or in probabilistic terms mathematical expectation, is a theoretical variogram $\gamma(h)$.

The passage to the IRF model enables us to define criteria for the precision required during modeling. In the framework of this model, the deviation $\Gamma_R(h) - \gamma(h)$ is a random variable. Its expectation is zero, and we can quantify the possible deviations by the fluctuation variance $\text{Var}[\Gamma_R(h) - \gamma(h)]$. Let us therefore consider a given value of h and use the notations of the previous section. In view of definition (2.90) of $\Gamma_R(h)$, the variance of the fluctuation of $\Gamma_R(h)$ is simply the variance of the fluctuation of the mean of $Q_h(x)$ in $D \cap D_{-h}$. It can therefore be expressed in terms of the covariance $G_h(x' - x)$ of the RF $Q_h(x)$,

$$\text{Var}[\Gamma_R(h) - \gamma(h)] = \frac{1}{|D \cap D_{-h}|^2} \int_{D \cap D_{-h}} \int_{D \cap D_{-h}} G_h(x' - x) dx dx' \quad (2.93)$$

Calculating this variance again involves the fourth-order moments of the IRF $Z(x)$. Given the same remarks as in the preceding section, let us first examine the Gaussian case.

Gaussian Case

In view of the expression (2.92) for $G_h(x' - x)$, (2.93) is expressed as a function of $\gamma(h)$. The explicit calculation is complex, so for simplification we will consider the case where $\gamma(h)$ near the origin is equivalent to $b|h|^\alpha$. The relative fluctuation variance for small $|h|$ is, to a first approximation, of the form

$$\frac{\text{Var}[\Gamma_R(h) - \gamma(h)]}{\gamma(h)^2} \simeq A|h|^{4-2\alpha} + B|h|^n$$

where n is the dimension of the space (1, 2, or 3) (see Matheron, 1978, p. 113, and the proof in \mathbb{R}^1 in Matheron, 1970, sec. 2.10.3). This relative variance tends to zero provided that $\alpha < 2$.

Micro-Ergodicity

The convergence of $\Gamma_R(h)/|h|^\alpha$ to a constant b when $h \rightarrow 0$ is ensured, provided that the RF is not m.s. differentiable. The parameter b then has an objective meaning: if we increase the number of sample points by refining the sampling grid, we can estimate it with precision. This is the concept of micro-ergodicity. It differs from conventional ergodicity, where one extends the data domain to infinity, which is of little interest to us because we always work in a bounded domain D . Micro-ergodicity refers to the case of a finite domain where we let the number of sample points tend to infinity by filling in the available space (Cressie, 1991, calls this "infill asymptotics").

The micro-ergodicity of the variogram in the neighborhood of the origin is therefore established provided that it is not too regular (we find ourselves in an inverse situation to that of the estimation of the variogram). This is easily explained: if α is very close to 2, the realization $z(x)$ is a very regular function. In choosing a sufficiently small domain D , $z(x)$ can be assimilated to a line segment. But, depending on the location of D , the line segments can have extremely variable slopes and thus almost any coefficient of the term $|h|^\alpha$ of the regional variogram: the parameter b no longer has an objective meaning. In this case the probabilistic model is poorly adapted, at least at the scale of a small domain D .

Although it is generally possible to determine $\gamma(h)$ in the neighborhood of the origin, the fluctuation variance increases very rapidly with h (except if $\gamma(h)$ has a very small range with respect to the domain D). Thus in \mathbb{R}^1 , for example, for an RF with a linear variogram known on

[0, 1] (straightforward application of (2.93); see Matheron, 1970, ch. 2, exer. 16), we have

$$\frac{\text{Var}[\Gamma_{\mathbf{R}}(h) - \gamma(h)]}{\gamma(h)^2} = \begin{cases} \frac{4}{3} \frac{h}{L-h} - \frac{1}{3} \frac{h^2}{(L-h)^2} & \text{if } 0 \leq h \leq \frac{L}{2} \\ 2 - \frac{4}{3} \frac{L-h}{h} + \frac{1}{3} \frac{(L-h)^2}{h^2} & \text{if } \frac{L}{2} \leq h \leq L \end{cases}$$

At $h = L/2$, the relative fluctuation variance already equals one, which is prohibitive.

Figure 2.33 shows the fluctuation standard deviation curves for the same cases as in Figure 2.31. Notice that the possible fluctuations in \mathbb{R}^2 are not so strong as in \mathbb{R}^1 . Nevertheless, apart from the case of a small range, the theoretical variogram and the regional variogram may have only a distant relationship if $|h|$ is greater than half the length of the domain (if not sometimes less). This is not serious insofar as geostatistical estimations depend much more on the behavior of the variogram at small distances than at large distances. From a practical point of view, we can accept that the variogram generally has no objective meaning at large distances and that it is pointless to try to refine the associated fit.

General Case

The conclusions are less encouraging for RFs that are clearly non-Gaussian, as has been shown by Alfaro (1979). We can see this from Figure 2.34, which shows the same RFs as Figure 2.32. Two conclusions can be drawn:

1. The fluctuation variance can be much larger than in the Gaussian case.
2. The relative variance no longer necessarily tends to zero when the lag h tends to zero: in other words, micro-ergodicity is no longer ensured, and the regional variogram does not even reproduce the behavior near the origin of the theoretical variogram.

In practice, this is frequently reflected in a proportional effect. If we treat it as such, which amounts to working on a variable conditioned on its local mean, we end up with a more satisfactory model. It is generally better to use methods based on prior transformation of the data into Gaussian variables (lognormal kriging, disjunctive kriging).

Micro-ergodicity is also no longer ensured for an indicator, or more generally for a mosaic RF (even if the marginal distribution is Gaussian). If the domain of study D is not large, a realization can very well be constant and so give a regional variogram that is identically zero.

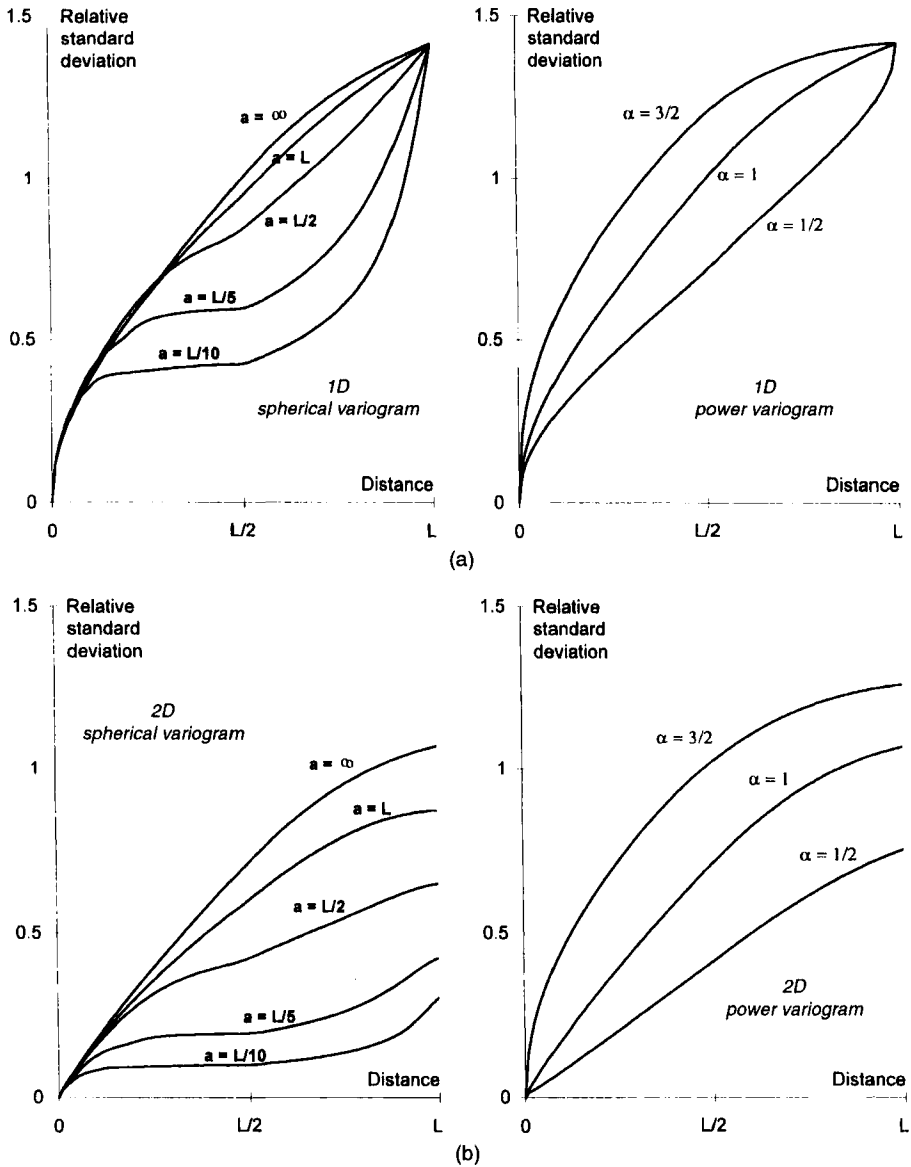


FIGURE 2.33. Fluctuation of the regional variogram: relative standard deviation in the Gaussian case for a spherical variogram with range a (on the left) and for a variogram of type $|h|^\alpha$ (on the right): (a) along a segment of length L ; (b) on a square $L \times L$.

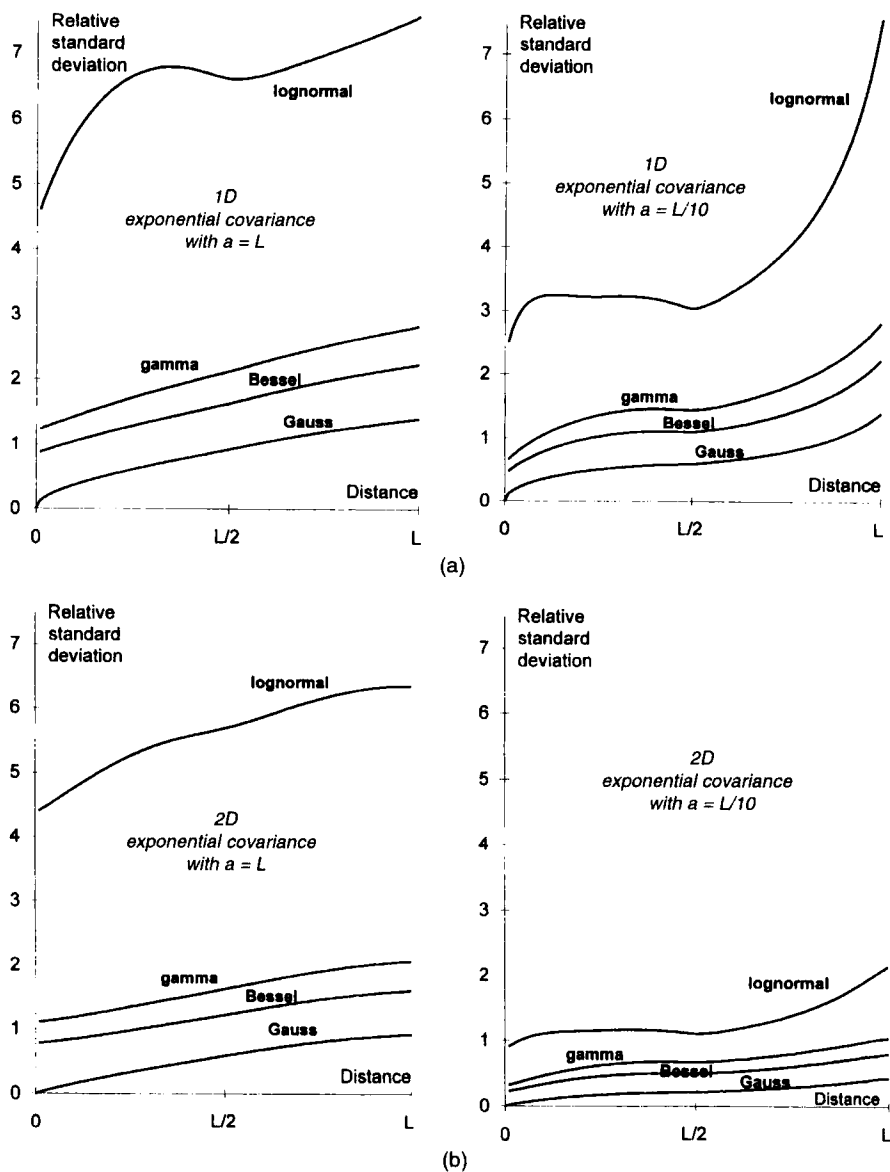


FIGURE 2.34. Fluctuation of the regional variogram: relative standard deviation for different spatial distributions, for an exponential variogram with scale parameter L (on the left) and $L/10$ (on the right): (a) along a segment of length L ; (b) on a square $L \times L$.

NOTES

1. These 15 directions form five rectangular trihedra which can be deduced from the first one (e.g., the trihedron associated with the coordinate axes) by four successive rotations defined by the same matrix

$$\mathbf{R} = \frac{1}{2} \begin{bmatrix} 1 & -(k+1) & k \\ k+1 & k & 1 \\ k & 1 & k+1 \end{bmatrix} \quad \text{with } k = \frac{\sqrt{5}-1}{2}$$

2. Consider an SRF $Z(x)$ in \mathbb{R}^n whose covariance is continuous except at the origin. The covariance of the SRF $Z_1(x) = Z(x) + Z(x+a)$, where a is a given vector of \mathbb{R}^n , is discontinuous at $h=0$, a , and $-a$. Note that if $n > 1$, this covariance is not isotropic. According to a conjecture of Schoenberg (1938a), an isotropic covariance in \mathbb{R}^n , $n > 1$, cannot have a discontinuity except at the origin.
3. This is true at least for any separable Gaussian IRF whose variogram satisfies $\gamma(h) \leq c|h|^\delta$ when $|h| \leq r_0$ for some strictly positive constants c, δ, r_0 , which is the case for all continuous variograms of practical use. In the general case, the following theorem due to Kolmogorov applies: if the separable random function Z is such that $E|Z(x+h) - Z(x)|^\alpha \leq c|h|^{1+\beta}$ for small h for some strictly positive constants α, β, c , then almost every realization of Z is continuous. In all cases the assumption of separability is essential—and fully justified in practical applications, since a nonseparable version of a random function is really a curiosity. As an example consider a Gaussian SRF $Z(x)$ with continuous realizations. Now select a random location X from any continuous distribution. The RF $Z_1(x)$ defined as equal to $Z(x)$ if $x \neq X$ and $Z(x) + 1$ if $x = X$ has the same finite-dimensional distributions as Z , but it is not separable and none of its realizations is continuous.
4. A more rigorous presentation can be found, for example, in Yaglom (1987, vol. 1, pp. 66–67).
5. Positive definite is taken synonymously to *nonnegative definite*. To exclude the value zero, we refer to *strict* positive definiteness.
6. Positive definiteness of the covariance ensures that this integral is positive or zero.
7. Fourier expansions into harmonics $\exp(i\langle\omega, x\rangle)$ instead of $\exp(2\pi i\langle u, x\rangle)$ are also used, where ω is the angular frequency.
8. We conjecture that a necessary and sufficient condition for $C(h)$ to be strictly positive definite is that the spectral measure is not concentrated on a set whose Lebesgue measure is zero. A sufficient criterion for this is the summability of $|C(h)|$ over \mathbb{R}^n , since then the spectral density is absolutely continuous with respect to the Lebesgue measure.
9. The proof is based on the fact that $\kappa_n(\sqrt{2n}r)$ tends to $\exp(-r^2)$ as $n \rightarrow \infty$.
10. This results from an application of formula (7.4.3) in Gautschi (1972, p. 302):

$$\frac{1}{a\sqrt{\pi}} \int_0^\infty \exp\left(-\frac{t^2}{4a^2} - \frac{r^2}{t^2}\right) dt = \exp\left(-\frac{r^2}{a^2}\right) \quad r \geq 0, \quad a > 0$$

11. A function f of $t \in \mathbb{R}_+$ is said to be completely monotone if and only if it is continuous at 0 and satisfies $(-1)^n f^{(n)}(t) \geq 0 \quad \forall t > 0 \quad (n = 0, 1, 2, \dots)$, namely its derivatives have a constant sign.
12. The general result derives from the following relations where C denotes a completely monotone isotropic covariance and C_1 an isotropic covariance in \mathbb{R}^n for all n :

$$\begin{aligned}
 C(r^2) &= \int_0^\infty \exp\left(-\frac{r^2}{t}\right) \mu(dt) = \int_0^\infty \exp\left(-\left(\frac{r}{\sqrt{t}}\right)^2\right) \mu(d\sqrt{t}) \\
 &= \int_0^\infty \exp\left(-\frac{r^2}{t^2}\right) \nu(dt) = C_1(r)
 \end{aligned}$$

13. In stereology, and particularly in tomographic imaging, the main concern is the inverse problem, namely the derivation of w_n knowing the value of its integral along any line or any plane.
14. Yaglom's terminology relates to $2^{-n}A$, where n is space dimension. In the literature the term "correlation length" sometimes also means what we call range, and sometimes refers to what we call scale parameter.
15. Incidentally, the persistence of the Poisson point process under a large variety of transformations rules out the possibility of detecting strain in a rock by analysis of the position of objects after deformation if the original object positions are "random" and mutually independent (Fry, 1979). But such detection is possible if the objects themselves are deformed; for example, initially spherical markers become ellipsoids.
16. Note that this property is in fact *self-affinity* rather than self-similarity (Mandelbrot, 1985): if we consider the Gaussian case (fractional Brownian random function) the invariance is ensured by two different scale changes: a factor s in the horizontal plane, and a factor $s^{\alpha/2}$ in the vertical plane. Isoline curves, which are horizontal planar sections, are self-similar. Thus the yardstick method can be applied to determine the fractal dimension of a fractional Brown surface from its contours (cf. the well-known example of the length of the coast of Britain) but not from a vertical cross section.
17. It seems that the variogram of order 1, defined in Section 2.5.3, is the right tool to diagnose the fractal character (then the realizations define fractal surfaces with fractal dimension $D = n + 1 - H$ if $\gamma_1(h) = b|h|^H$, $0 < H < 1$).
18. The criterion should be based on the behavior of the variogram of order 1 at the origin.
19. At very short distances the variogram is no longer of type $|h|^{2/3}$ but is parabolic. This behavior can be modeled by regularization of an $|h|^{2/3}$, stable, or K -Bessel model.
20. The accumulation along a vertical borehole is the product of the ore thickness by the average ore grade in the borehole; it represents a quantity of metal per unit surface.
21. More precisely $b(\theta) = \lim_{r \downarrow 0} \gamma(ru)/r$ is the directional derivative in the direction θ of the unit vector u . The symmetry of γ entails that $b(\theta) = b(\theta + \pi)$.
22. As a gamma random variable with shape parameter $\alpha = 2$ can be considered as the sum of two independent random variables with the same exponential distribution, this alternating process can be defined as follows: (1) start from a Poisson point process with intensity $\lambda = 1/a$, (2) merge the segments delimited by these points two by two (choose randomly whether the segment containing the origin is merged with the preceding or the following segment), and (3) assign the values 0 and 1 alternately to each of these new segments (choose randomly whether the new segment containing the origin is valued 0 or 1). The covariance of this random set can be obtained by considering that the bivariate distribution of $Z(x)$ and $Z(x+h)$ depends on the number of points of the initial Poisson point process falling between x and $x+h$, which is a Poisson random variable with mean $\theta = \lambda|h|$.
23. This migration process is defined as follows: (1) consider a Poisson point process with intensity $\lambda = 1/a$, (2) subdivide each interval of this point process into two unequal parts, corresponding to the proportions p (left) and $q = 1 - p$ (right) of the initial interval, and (3) assign the value 1 to the left part and the value 0 to the right part.
24. Note that this definition is slightly different from that of the entropy of a probability density function $f(z)$, which is defined by $H = - \int f(z) \log f(z) dz$.

CHAPTER 3

Kriging

Once a map is drawn people tend to accept it as reality.

Bert Friesen

3.1. INTRODUCTION

A central problem in geostatistics is the reconstruction of a phenomenon over a domain on the basis of values observed at a limited number of points. Typically one may want to build a grid model of a variable of interest in view for example of drawing a contour map or of running some numerical simulation. Or the objective may be to make an inventory and compute amounts (of ore, trees, contaminants, etc.) within given areal or volumetric units. More generally, the desired quantity may be some function of the observed variable, but attention will be restricted here to linear functions.

Mathematically this problem can be regarded as an interpolation problem. In the classic approach the unknown function is approximated by a parametric function whose form is postulated in advance, either explicitly (e.g., polynomials) or implicitly (e.g., minimum curvature condition). The parameters are selected so as to optimize some criterion of best fit at the data points, which can be statistical (least squares) or deterministic (exact fit at the points). Once the approximating function is determined, it is a simple matter to evaluate it wherever needed.

We will expand here on a different approach known as *kriging*, a term coined by G. Matheron in 1963 after the name of D. G. Krige. The main difference is that kriging starts from a statistical model of Nature rather than a model of the interpolating function. A formal analogy is provided by the prediction problem in time series: given values of the past, usually at regular time intervals, predict the value of the signal at some time in the future. First the signal is analyzed, typically by computing and modeling the spectrum, and then a filter (= a predictor) is designed. Kriging follows a similar approach but in a spatial setting where there is no general concept of past and future.

TABLE 3.1. Main forms of linear kriging

Kriging Type	Mean	Minimal Prerequisite	Model Name
Simple kriging (SK)	Constant, known	Covariance	Stationary
Ordinary kriging (OK)	Constant, unknown	Variogram	Intrinsic
Universal kriging (UK)	Varying, unknown	Variogram	UK model

A remark on terminology is in order here. In time series analysis several words are used to designate interpolation and extrapolation, each with its own nuance: *smoothing*, *filtering*, *prediction*, *forecasting*. In the spatial context the word *prediction* tends to become established in a technical sense, referring to the determination of the value of a *random* quantity, whereas *estimation* refers to the inference of some *fixed* but unknown parameter of a model. This distinction originates from regression where we have the choice to either estimate the mean value of Y given X , or the value of an individual observation given X . The two estimates are the same but their variances are different and therefore it is necessary to distinguish them. As spatial interpolation is closely related to regression the terminology has followed, although the distinction is no longer critical. Besides, the word *prediction* has a time connotation that is improper for the spatial problems considered in geostatistics—we predict an earthquake but estimate mineral reserves. Therefore in this book we use estimation, rather than prediction, terminology throughout, except in the context of regression or time series.

Two other differences between the spatial processes considered in this book and time processes must be emphasized. First, the data are usually *unevenly sampled*, which rules out most of the methods commonly used in digital signal processing of either time signals or images. Second, there are *support effects* that require, at least in concept, a continuous rather than a discrete location indexing space.

The present chapter concentrates on solving the estimation problem by use of linear estimators. The theory is developed within the scope of a second-order statistical model involving only the mean $m(x)$ and the covariance function $\sigma(x, y)$ or the variogram $\gamma(x, y)$, assumed *known*. Note that the stationarity of σ or γ is not required to derive the estimators. In some cases such as space-time phenomena, it is possible to consider a nonstationary model. In most applications, however, a stationary covariance $C(h)$ or variogram $\gamma(h)$ is used. Table 3.1 summarizes the three main forms of kriging that we will encounter and their underlying models.

The method of kriging has been extended in several directions. One extension deals with broader forms of nonstationarity than the UK model and shows how the minimal prerequisites for kriging can be reduced: it is the subject of Chapter 4. Another extension covers the multivariate case, na-

mely when $Z(x)$ is vector-valued. The method, referred to as *cokriging* is developed in Chapter 5, except for the particular problem of filtering a random error in the data which is treated here. Finally nonlinear estimators developed to evaluate nonlinear functions of $Z(x)$ are reviewed in Chapter 6.

3.2. NOTATIONS AND ASSUMPTIONS

- We denote by

$$\{Z(x) : x \in D \subset \mathbb{R}^n\}$$

a random function representing the regionalized variable of interest $z(x)$ —i.e., $z(x)$ is a realization of $Z(x)$. For simplicity we make no notational distinction between the (uppercase) parent random function $Z(x)$ and its particular (lowercase) realization $z(x)$, it being clear that all probabilistic calculations involve the random function and all numerical estimates involve the realization. The lowercase notation is used occasionally, when we want to emphasize the deterministic character of the expression (e.g., kriging as an interpolator).

- S denotes the set of points where $Z(x)$ has been sampled. In most cases S is finite and consists of N data points denoted with Greek subscripts:

$$S = \{x_\alpha : \alpha = 1, \dots, N\}$$

Occasionally S can be infinite such as when data are continuously recorded along a profile.

- Values of functions at sample points are referenced by the subscripts of these points, such as

$Z_\alpha = Z(x_\alpha)$	The data
$m_\alpha = m(x_\alpha)$	Mean value of $Z(x_\alpha)$
$\sigma_{\alpha\beta} = \sigma(x_\alpha, x_\beta)$	Covariance between $Z(x_\alpha)$ and $Z(x_\beta)$

- The estimated quantity (the “objective”) is of the general form

$$Z_0 = \int Z(x)p_0(dx)$$

for some integrable measure p_0 . The case of point estimation $Z_0 = Z(x_0)$ corresponds to a Dirac measure $p_0(dx) = \delta(x - x_0)$ at the target point x_0 ,

whereas the estimation of a spatial average over a block v

$$Z_0 = \frac{1}{|v|} \int_v Z(x) dx$$

corresponds to $p_0(dx) = (1/|v|)1_v(x)dx$, where $1_v(\cdot)$ stands for the indicator function of the block v centered at the point x_0 .

- Following established usage, kriging estimators are marked with an asterisk (*) superscript. In full explicit notations the kriging estimator of $Z(x_0)$ is of the form

$$Z^*(x_0) = \sum_{\alpha=1}^N \lambda_{\alpha}(x_0)Z(x_{\alpha}) + \lambda_0(x_0)$$

where $\lambda_{\alpha}(x_0)$ is a weight placed on $Z(x_{\alpha})$ and $\lambda_0(x_0)$ is a constant that depends on x_0 . For brevity this expression will be condensed to¹

$$Z^* = \sum_{\alpha} \lambda_{\alpha} Z_{\alpha} + \lambda_0$$

with the understanding that the summation is extended over all α indexes in S . It must be kept in mind that *the weights λ_{α} depend on the location x_0 where the function is being estimated.*

Support

We have said that the data were measured at points x_{α} . In reality the data are never collected at a single point but always involve a support of finite dimensions. This does not create any particular difficulty as long as all data have the same support and as long as this support is “compatible” with that of the objective. We speak of “point kriging” when the objective has exactly the same support as the samples (Sections 3.3 and 3.4). When the objective has a larger support, we speak of “block kriging” (Section 3.5). For compatibility the block must either be very large compared with the samples, which are then treated as points, or be a finite union of sampling units.

Neighborhoods

The theory is always derived as if all N data points were used in the estimation; this is the so-called global neighborhood case. In practice, N may be too large to allow computation and a “moving neighborhood” or “local neighborhood” has to be used, including only a subset of the data for the estimation of each grid node. Formally this does not change anything for a grid node taken in isolation: the content of S is just different. But it may alter the relationships between estimates at different grid nodes and introduce spurious discontinu-

ities. Such problems can be largely avoided by providing sufficient overlap between local neighborhoods.

3.3. KRIGING WITH A KNOWN MEAN

In this section we consider what could be called “the wonderful case of a known mean” which underlies the theory of *simple kriging* (SK). Indeed, knowing the mean of a stationary RF is already knowing very much. For some practical purposes the mean alone may provide enough information. Knowing the mean also makes the theory very simple and endows the kriging estimator with all the nice properties. In case of a Gaussian RF, it coincides with the conditional expectation $E(Z_0 | Z_1, \dots, Z_N)$ which is the ideal estimator of Z_0 in the mean square sense. In all circumstances the error $Z^* - Z_0$ is uncorrelated with every Z_α and with Z^* itself.

In the real world the mean can be known only if there are repetitions of the phenomenon, as with space-time processes, or when the number of data becomes so large as to estimate the mean almost to perfection. The properties established under this condition may thus be regarded as *limiting properties*.

3.3.1. Derivation of the Equations

For simplicity we consider here the case of point estimation. We want to estimate $Z_0 = Z(x_0)$ from N observations Z_1, \dots, Z_N , using the affine estimator

$$Z^* = \sum_{\alpha} \lambda_{\alpha} Z_{\alpha} + \lambda_0$$

The constant λ_0 and the weights λ_{α} are selected so as to minimize the error $Z^* - Z_0$, characterized by its expected mean square $E(Z^* - Z_0)^2$.

First let us concentrate on λ_0 . The mean square error (m.s.e.) can be written as

$$E(Z^* - Z_0)^2 = \text{Var}(Z^* - Z_0) + [E(Z^* - Z_0)]^2$$

Since variances are insensitive to shifts only the bias term on the right-hand side involves λ_0 . To minimize the m.s.e., it is necessary to choose λ_0 so as to cancel the bias $E(Z^* - Z_0)$:

$$\lambda_0 = m_0 - \sum_{\alpha} \lambda_{\alpha} m_{\alpha}$$

The estimator Z^* becomes

$$Z^* = m_0 + \sum_{\alpha} \lambda_{\alpha} (Z_{\alpha} - m_{\alpha}) \quad (3.1)$$

This amounts to estimating the zero-mean variable $Y(x) = Z(x) - m(x)$ by the linear estimator

$$Y^* = \sum_{\alpha} \lambda_{\alpha} Y_{\alpha}$$

and adding the mean afterward. Thus we have established that the case of a known mean is equivalent to the case of a zero mean with $\lambda_0 = 0$, and from now on in this section we will consider that $Z(x)$ has a zero mean.

The m.s.e.—which now coincides with the variance—can be expanded in terms of the centered covariance $\sigma(x, y)$ of $Z(x)$,

$$E(Z^* - Z_0)^2 = \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \sigma_{\alpha\beta} - 2 \sum_{\alpha} \lambda_{\alpha} \sigma_{\alpha 0} + \sigma_{00}$$

The minimum of this quadratic function is obtained by canceling its partial derivatives with respect to the weights λ_{α} ,

$$\frac{\partial}{\partial \lambda_{\alpha}} E(Z^* - Z_0)^2 = 2 \sum_{\beta} \lambda_{\beta} \sigma_{\alpha\beta} - 2 \sigma_{\alpha 0} = 0$$

(That this is indeed a minimum is ensured by the positivity property of the covariance function—the m.s.e. is a convex function.) The λ_{α} are solutions of the linear system of N equations

Simple Kriging System

$$\sum_{\beta} \lambda_{\beta} \sigma_{\alpha\beta} = \sigma_{\alpha 0} \quad \alpha = 1, 2, \dots, N \quad (3.2)$$

In matrix notations

$$\Sigma \lambda = \sigma_0$$

where $\Sigma = [\sigma_{\alpha\beta}]$ is the $N \times N$ matrix of data-to-data covariances, $\sigma_0 = [\sigma_{\alpha 0}]$ the N -vector of covariances between the data and the target, and $\lambda = [\lambda_{\alpha}]$ the N -vector of solutions.

These equations are the “best linear prediction” equations famous since the work of A. N. Kolmogorov (1941c) and N. Wiener (1942). They also appear as the “Yule-Walker” equations in time series (e.g., Box and Jenkins, 1976, p. 55) and as the “normal equations” in linear regression (e.g., Rao, 1973, p. 266).

The system (3.2), usually called the simple kriging system (SK), has a unique solution provided that the matrix Σ is nonsingular. This is always the case if the covariance function $\sigma(x, y)$ is strictly positive definite and if all

sample points are distinct, which we will always assume (nonunique solutions may occur in case of a pure zonal anisotropy). Solving (3.2) is a routine computational problem.

The estimation variance σ_{SK}^2 , called the *kriging variance*, associated with Z^* is obtained by substituting the solution of (3.2) in the m.s.e. Premultiplying (3.2) by λ_α and summing over all α gives

Simple Kriging Variance

$$\sigma_{SK}^2 = E(Z^* - Z_0)^2 = \sigma_{00} - \sum_{\alpha} \lambda_{\alpha} \sigma_{\alpha 0} \quad (3.3)$$

In matrix form

$$\sigma_{SK}^2 = \sigma_{00} - \lambda' \sigma_0$$

The kriging variance, or rather its square root, the kriging standard deviation, provides a measure of the error associated with the kriging estimator. It can be called, for short, the *standard error*. Its use is discussed in Section 3.4.4. Notice that it does not depend on the values of the data but only on their locations. Notice also that if the covariance is multiplied by an arbitrary (positive) constant, the kriging weights do not change and the kriging variance is multiplied by that constant. Therefore in case of a stationary covariance function, one can scale the covariance by the common variance and write the system (3.2) in terms of correlation coefficients $\rho_{\alpha\beta}$ and $\rho_{\alpha 0}$, but the kriging variance is then also scaled. Finally, and still in the stationary case, it is clear that the kriging weights and variance are *shift invariant*: they do not change if the whole kriging configuration is shifted by an arbitrary vector h .

Examples

1. Consider the simplest case of only one sample point. Then (3.2) reduces to a single equation and the kriging estimator is

$$Z^*(x_0) = \left(\frac{\sigma_{10}}{\sigma_{11}} \right) Z_1 = \rho_{10} \left(\frac{\sigma_{00}}{\sigma_{11}} \right)^{1/2} Z_1$$

where ρ_{10} is the correlation coefficient between Z_1 and Z_0 . One recognizes the standard linear regression of Z_0 on Z_1 . As a function of x_0 , $Z^*(x_0)$ is proportional to the covariance function σ_{10} and assumes the value Z_1 when $x_0 = x_1$. As x_0 moves away from x_1 , the correlation ρ_{10} usually falls off and so does, logically, the influence of Z_1 on the estimation of Z_0 . At large distances from x_1 , $\rho_{10} = 0$ so that $Z^* = 0$; we can do no better than estimate $Z(x)$ by the mean.

From (3.3) the kriging variance is

$$\sigma_{SK}^2 = \sigma_{00}(1 - \rho_{10}^2)$$

2. We now complicate the problem and consider two points x_1 and x_2 . Solving the 2×2 kriging system (3.2) for λ_1 and λ_2 and rearranging the terms lead to the estimator

$$Z^*(x) = \frac{1}{\Delta}[Z_1\sigma_{22} - Z_2\sigma_{12}]\sigma_{10} + \frac{1}{\Delta}[Z_2\sigma_{11} - Z_1\sigma_{21}]\sigma_{20}$$

with $\Delta = \sigma_{11}\sigma_{22} - \sigma_{12}^2$. This solution is represented graphically in Figure 3.1 (left), along the line joining x_1 and x_2 and for the stationary covariance function defined by equation (2.49), namely the “cubic” model with range a .

The estimator passes through Z_1 and Z_2 and tends to the mean zero as x_0 moves away from the data points, reaching zero exactly when the distance exceeds the range a . The three curves in Figure 3.1 illustrate the influence of the range a on the solution: as a decreases, the covariance function falls off more rapidly, and the information carried by the data Z_1 and Z_2 is considered more and more local.

The kriging variance is readily obtained from (3.3)

$$\sigma_{SK}^2 = \sigma_{00} - \frac{1}{\Delta}[\sigma_{22}\sigma_{10}^2 + \sigma_{11}\sigma_{20}^2 - 2\sigma_{12}\sigma_{10}\sigma_{20}]$$

This function is plotted in Figure 3.1 (right) along the line (x_1, x_2) for various values of a . It is seen that σ_{SK}^2 is zero at the sample points and reaches the maximum value of 1 under extrapolation. As could be expected, the variance reaches a local maximum at the midpoint between x_1 and x_2 . When a is small, this maximum is equal to 1 because all sites are considered as extrapolated except in the immediate vicinity of the data.

3.3.2. Interpolation Properties of the Kriging Estimator

We can now generalize some of the properties observed in the above examples.

Consistency with Data Points

The kriging estimator is an exact interpolator. If x_0 coincides with a sample point, say x_1 , then Z^* is equal to $Z(x_1)$. This can be verified by checking that the set of weights $\lambda_1 = 1$, $\lambda_\alpha = 0$, if $\alpha \neq 1$ satisfies the equations (3.2), and since the solution is unique, this is it. But it is simpler to note that $Z^* = Z(x_1)$ is certainly the best estimator of $Z(x_1)$ in the m.s.e. sense as it makes the error exactly zero. The kriging variance σ_{SK}^2 is naturally also zero.

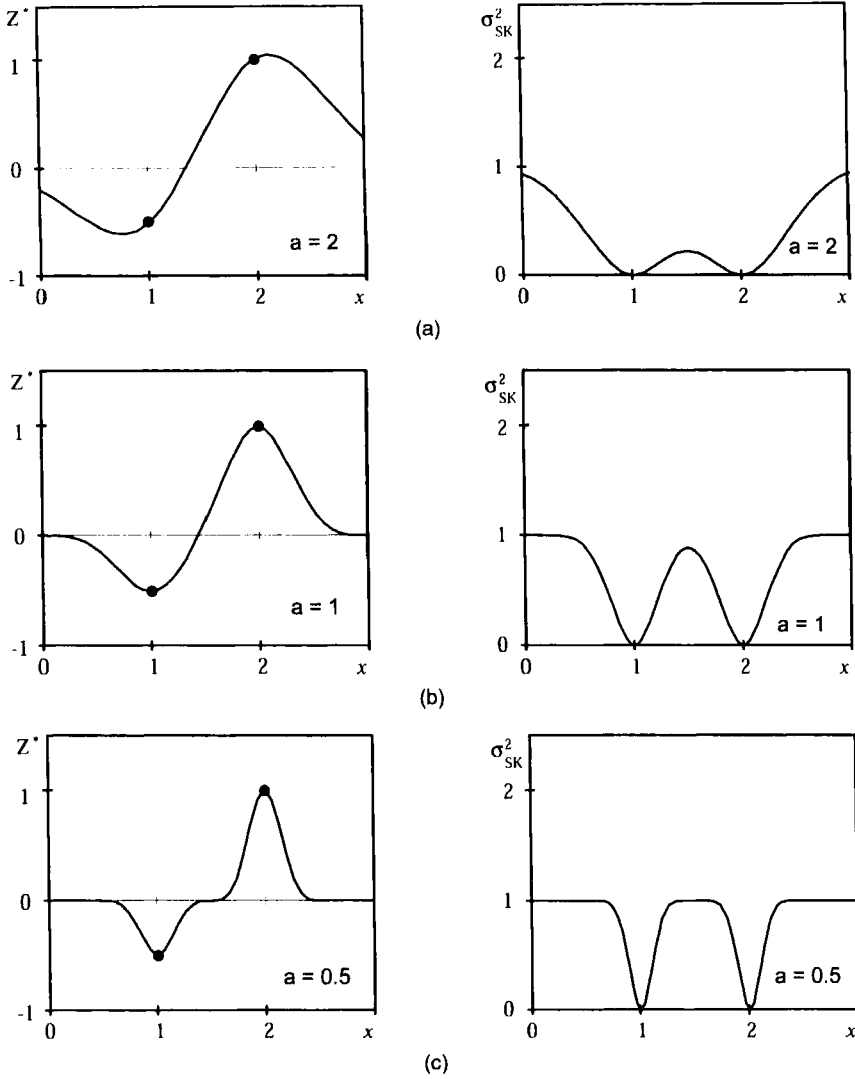


FIGURE 3.1. Simple kriging estimates and variances for the case of two data points and a cubic covariance model. From (a) to (c): the estimator becomes “wigglier” as the range decreases.

Smoothing Relationship

Since kriging performs a linear averaging, we expect kriging estimates to be less dispersed than the data. This can be proved easily by considering the variance of $Z^*(x_0)$. From (3.2),

$$\text{Var} Z^* = \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \sigma_{\alpha\beta} = \sum_{\alpha} \lambda_{\alpha} \sigma_{\alpha 0}$$

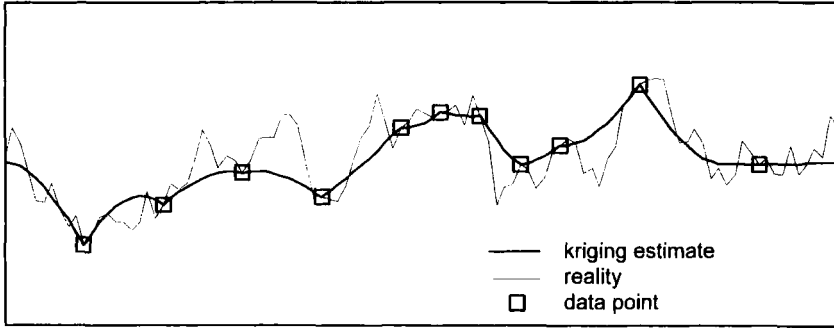


FIGURE 3.2. Illustration of the smoothing effect of kriging.

and using (3.3) together with $\text{Var}Z_0 = \sigma_{00}$, we get the so-called *smoothing relationship*

$$\text{Var}Z^* = \text{Var}Z_0 - \sigma_{\text{SK}}^2 \quad (3.4)$$

$\text{Var}Z^*$ differs from $\text{Var}Z_0$ by an amount exactly equal to the kriging variance σ_{SK}^2 (which depends on x_0). This effect is illustrated in Figure 3.2: the kriging estimate wiggles near the data points, where σ_{SK}^2 is small, and merges gradually with the mean as data become sparser.

Note that despite its name the smoothing relationship does not, by itself, prove the existence of a spatial smoothing effect in the sense of a preferential attenuation of high frequencies. (In spectral analysis terms, we have only proved that the total power is reduced.) However, this spatial smoothing does usually take place when a global kriging neighborhood is used.

Incidentally we have also shown that $\sigma_{\text{SK}}^2 \leq \text{Var}Z_0$. In a stationary model the estimation variance can never exceed the global variance, even if x_0 is very far from all data points (else the mean would become the estimator).

Kriging as an Interpolator

Though derived in a stochastic model, the function $Z^*(x) = z^*(x)$, once the data $Z_\alpha = z_\alpha$ are fixed, is a deterministic interpolator of some kind. To emphasize this fact, in this section we switch to lowercase notations for z , and to x for the interpolated point. We now wish to determine the explicit form of the kriging interpolator. To this end we solve (3.2) for λ , and letting \mathbf{z} be the N -vector of data, we get

$$\mathbf{z}^* = \mathbf{z}' \Sigma^{-1} \sigma_0 \quad (3.5)$$

In this expression only σ_0 depends on the location x . Defining $\mathbf{b} = \Sigma^{-1} \mathbf{z}$, we get

$$z^*(x) = \sum_{\alpha} b_{\alpha} \sigma(x_{\alpha}, x) \quad (3.6)$$

This is a linear combination of N covariance functions centered at the sample points x_α . The weights b_α do not depend on x , but they depend on the z_α (cf. Examples 1 and 2).

To take a physical analogy, suppose that the covariance function is stationary and that $\sigma(x_\alpha, x) = C(x - x_\alpha)$ represents the potential generated at x by a unit charge located at x_α . Then $z^*(x)$ is the potential resulting from the superposition of N charges b_1, \dots, b_N located at the sample points x_1, \dots, x_N . Note that each b_α is, in general, a linear function of *all* the data and not just of z_α . Even when the covariance has a finite range, and therefore $\sigma(x_\alpha, x) = 0$ in (3.6), the kriging estimates involve sample points beyond the range through the coefficients b_α .

Mathematically the kriging estimator can be *characterized* as the only interpolator of the form (3.6) that passes through all the data points. Indeed, if $z^*(x_\beta) = z_\beta$, at each x_β the b_α satisfy

$$\sum_{\alpha} b_{\alpha} \sigma_{\alpha\beta} = z_{\beta} \quad \text{in matrix terms} \quad \Sigma \mathbf{b} = \mathbf{z}$$

so that

$$\mathbf{z}^* = \mathbf{z}' \Sigma^{-1} \sigma_0$$

which coincides with (3.5).

The interpolation formula (3.6) shows how the covariance function $\sigma(\cdot, \cdot)$ determines the continuity and regularity properties of $z^*(x)$. If the covariance function is parabolic near the origin, $z^*(x)$ is differentiable; if it is linear near the origin, $z^*(x)$ is continuous but with cusps at the data points. If the covariance has a discontinuity at zero, there will be isolated jumps at the data points. These behaviors are illustrated in Figure 3.3. Notice in Figure 3.3b that kriging interpolation does not necessarily “look nice.” Kriging is not designed to optimize the looks of the interpolation, but its accuracy (e.g., as opposed to spline interpolation). The correct way to appreciate the quality of the results is to regard them as an array of numbers, not as a curve or a set of contours.

3.3.3. Kriging as a Projection

A simple geometric interpretation of kriging can be derived within the framework of Hilbert spaces of random variables (Section 1.1.2). In addition to providing intuitive insight, this approach permits the only rigorous proof for the case of continuously sampled data.

We consider here the finite-dimensional vector space H_{N+1} generated by all linear combinations of the $N + 1$ zero mean, finite variance, random variables $Z_0 = Z(x_0)$ and $\{Z_\alpha : \alpha = 1, \dots, N\}$, and all their limits in the mean square sense. A scalar product $\langle X, Y \rangle = E(XY)$ is defined by the *noncentered* covariance and the associated norm is $\|X\| = \sqrt{\langle X, X \rangle}$. The kriging problem can be reformulated as follows: find X in the subspace H_N generated by the data

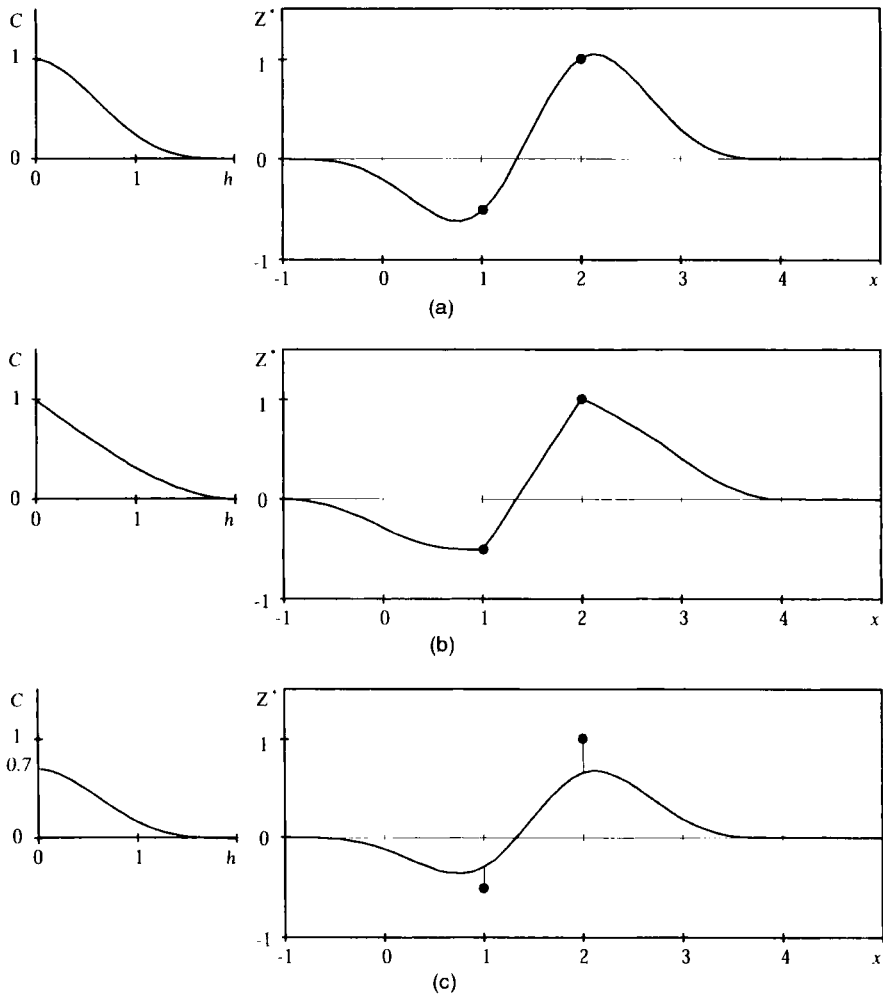


FIGURE 3.3. The dependence of the simple kriging estimate on the regularity near the origin of the covariance function: (a) parabolic behavior; (b) linear behavior; (c) discontinuity.

$\{Z_\alpha : \alpha = 1, \dots, N\}$ minimizing $\|Z_0 - X\|$. From Hilbert space theory we know that the minimum is achieved by a single element Z^* , called the “projection of Z_0 onto H_N ,” which is the foot of the perpendicular dropped from Z_0 on H_N . It is characterized by the orthogonality property (1.1),

$$\langle Z^* - Z_0, X \rangle = 0 \quad \forall X \in H_N \quad (3.7)$$

As $\{Z_\alpha : \alpha = 1, \dots, N\}$ forms a basis of H_N , the above is equivalent to the N conditions

$$\langle Z^*, Z_\alpha \rangle = \langle Z_0, Z_\alpha \rangle \quad \alpha = 1, \dots, N$$

which by expansion give the system (3.2). This concise form reveals another characteristic property of the SK estimator: Z^* has the same covariance with each Z_α as Z_0 itself.

A particular case of (3.7) is the orthogonality of the kriging error and the kriging estimator

$$\langle Z^* - Z_0, Z^* \rangle = 0 \quad (3.8)$$

Furthermore this orthogonality holds between the kriging error at x_0 and the kriging estimator $Z^*(x)$ at any other point provided that $Z^*(x)$ is obtained from the *same* data as $Z^*(x_0)$. This property is the basis of the conditioning algorithm by kriging (Section 7.3.1).

The kriging variance is easily obtained from

$$\sigma_{SK}^2 = \|Z^* - Z_0\|^2 = \langle Z^* - Z_0, Z^* \rangle - \langle Z^* - Z_0, Z_0 \rangle$$

and the application of (3.8). We get

$$\sigma_{SK}^2 = \langle Z_0, Z_0 \rangle - \langle Z^*, Z_0 \rangle$$

which by expansion is seen to coincide with (3.3).

The smoothing relationship (3.4) has a simple geometric interpretation. From (1.2), or directly from the orthogonal decomposition $Z_0 = Z^* + (Z_0 - Z^*)$, we have

$$\|Z_0\|^2 = \|Z^*\|^2 + \|Z_0 - Z^*\|^2$$

In the case of zero-mean random variables, this is equivalent to

$$\text{Var} Z_0 = \text{Var} Z^* + \sigma_{SK}^2$$

The smoothing relationship is but the Pythagorean theorem!

Notice that (3.8) entails $\text{Cov}(Z_0, Z^*) = \text{Var}(Z^*)$ so that

$$\rho^2 = \text{Corr}^2(Z_0, Z^*) = \frac{\text{Var}(Z^*)}{\text{Var}(Z_0)}$$

and the simple kriging variance takes the familiar form of a residual variance about a regression

$$\sigma_{SK}^2 = (1 - \rho^2) \text{Var} Z_0$$

These results are similar to (1.6) and (1.7) obtained in the case of the conditional expectation estimator for a good reason explained in the next section. The case of a nonzero mean can be dealt with as above by adding an extra weight λ_0 and augmenting the space H_N by the constant random variable 1.

We now turn to the case of continuous sampling (Matheron, 1969a). The set S of data points is now infinite, and we look for an estimator of the form

$$Z^* = \int_S \lambda(dx) Z(x) \quad (3.9)$$

where $\lambda(dx)$ is a weighting function on S (mathematically a measure on S). The kriging system obtained is analogous to (3.2) but with a continuous index x instead of a discrete α ,

$$\int_S \lambda(dy) \sigma(x, y) = \sigma(x, x_0) \quad \forall x \in S \quad (3.10)$$

The kriging variance is

$$\sigma_{SK}^2 = \|Z^* - Z_0\|^2 = \sigma_{00} - \int_S \lambda(dx) \sigma(x, x_0)$$

However, here some mathematical complications arise due to the infinite-dimensional nature of the Hilbert spaces considered. The subspace H generated by the data $\{Z(x) : x \in S\}$ contains all finite linear combinations of elements of S as well as their limits in the mean square sense, but these limits are not necessarily measures (they can be “generalized functions” in L. Schwartz’s sense; for example, the space may contain the derivatives of $Z(x)$). The projection Z^* of Z onto H still exists and is unique, but it is not necessarily of the form (3.9). The only valid statement is to say that if $\lambda(dx)$ satisfies (3.10), then (3.9) defines the optimal estimator. This difficulty only occurs when the covariance function is very regular, the classic example being $\sigma(x, y) = \exp(-(y - x)^2/a^2)$. In this case knowledge of the process over an arbitrarily short interval allows *perfect* extrapolation anywhere into the future (Yaglom, 1962, p. 190). But this solution is based on a Taylor’s series and not on the system (3.10). One could be tempted to ignore these mathematical difficulties and just solve the equations numerically by discretizing S , but the problem turns up as numerical instability.

3.3.4. Gaussian Regression Theory

When $Z(x)$ is a Gaussian RF, the simple kriging estimator Z^* in fact coincides with the conditional expectation $E(Z_0 | Z_1, \dots, Z_N)$. This follows immediately from the linearity of the regression function of the multivariate Gaussian distribution and the characteristic property (1.3) of conditional expectation, since we have in particular

$$E[(\lambda_1 Z_1 + \dots + \lambda_N Z_N - Z_0) Z_\alpha] = 0 \quad \alpha = 1, \dots, N$$

which are exactly the simple kriging equations (3.2).

The conditional distribution of Z_0 given $Z^* = z^*$ is also Gaussian and has for mean z^* and variance σ_{SK}^2 (which does not depend on z^*). Indeed $Z_0 - Z^*$ and Z^* are jointly Gaussian (as linear combinations of Z values) and uncorrelated. Therefore they are *independent*, and the conditional expectation of $Z_0 - Z^*$ and of $(Z_0 - Z^*)^2$ given Z^* are equal to their unconditional expectations

$$E[Z_0 - Z^* \mid Z^*] = E(Z_0 - Z^*) = 0$$

$$E[(Z_0 - Z^*)^2 \mid Z^*] = E(Z_0 - Z^*)^2 = \sigma_{SK}^2$$

It follows that

$$\begin{cases} E[Z_0 \mid Z^*] = Z^* \\ \text{Var}[Z_0 \mid Z^*] = \sigma_{SK}^2 \end{cases} \quad (3.11)$$

The property that $E(Z_0 \mid Z^*) = Z^*$, called “conditional unbiasedness,” is of great practical significance in the context of resource assessment problems.² For example, in selective mining the decision to process a block as ore or send it to waste is based on an estimate Z^* of the average grade of this block, but the actual ore recovery depends on Z_0 . Conditional unbiasedness ensures that, on the average, we get what we expect. In mining, this property is considered more essential than minimum variance (Krige himself has always insisted on this point; e.g., Krige, 1951, 1997; David, 1977; David et al., 1984; Journel and Huijbregts, 1978).

The kriging procedure thus appears especially suited to random functions which, like Gaussian ones, have regression functions that are linear in the data. In other cases kriging still provides the best linear estimator, but linear estimators may not be well adapted.

Another strong property of the Gaussian model is *homoscedasticity* of the conditional kriging error: its variance does not depend on the conditioning data values. As a consequence confidence intervals based on the kriging standard deviation σ_{SK} also constitute *conditional confidence intervals*. This is very nice but unfortunately specific to the Gaussian model.

3.4. KRIGING WITH AN UNKNOWN MEAN

In most practical situations the mean $m(x)$ is not known. An obvious approach would be to approximate it by some estimate $\hat{m}(x)$ and subtract it from the data, thereby reducing the problem to the zero-mean case. This approach is commonly used for processing time series, where it is known as “detrending.” Although such an approach is reasonable, as Duda (1982) phrases it, “it is very difficult to analyze the theoretical properties of such an ad-hoc combination of two estimation procedures. One is always left wondering how much the ‘high-frequency’ variations are corrupting the ‘low-frequency’ estimate of the

mean, and thus causing the subsequent zero-mean assumption to be violated.” Under a parametric assumption on the mean function, kriging provides an optimal solution that involves only one estimation step.

The simplest case is when the mean is a constant $m(x) = m$ and leads to *ordinary kriging* (OK). It was developed by Matheron in the early sixties and plays a special role because it is compatible with a stationary model, only involves the variogram, and is in fact the form of kriging used most.

The general model, which Matheron (1969a) named the *universal kriging* model for reasons explained hereafter, assumes that the mean can be written as a finite expansion

$$m(x) = \sum_{l=0}^L a_l f^l(x) \quad (3.12)$$

where the $f^l(x)$ are known basis functions and a_l are fixed but unknown coefficients. Usually the first basis function (case $l = 0$) is the constant function identically equal to 1, which guarantees that the constant-mean case is included in the model. The other functions are typically monomials of low degree in the coordinates of x (in practice, the degree does not exceed two). In the case of monomials, the superscript l , which is an *index*, has the meaning of a power (in 1D, $f^l(x) = x^l$). Note that (3.12) may be regarded as a local approximation to $m(x)$; that is, the coefficients a_l may vary in space but sufficiently slowly to be considered constant within estimation neighborhoods.

From a physical point of view the universal kriging model is the decomposition of the variable $Z(x)$ into the sum

$$Z(x) = m(x) + Y(x)$$

of a smooth deterministic function $m(x)$, describing the systematic aspect of the phenomenon, and called the *drift*, and a zero-mean random function $Y(x)$, called the *residual* and capturing its erratic fluctuations.³ Note that the drift refers to a technically precise notion (the mean of the RF Z), whereas *trend* is a generic term designating a general tendency, a systematic effect (besides, “trend” may imply an underlying driving force). Naturally the decomposition into drift and residual pertains to a certain *scale* of description. Seen from the road a mountain appears as a drift and local accidents of the relief as residuals, but seen from an airplane the mountain itself is a fluctuation in the mountain range.

Let us now see how not knowing the mean affects the estimation problem in the simple case of a constant mean $m(x) = a_0$. Consider again the affine estimator $Z^* = \sum_{\alpha} \lambda_{\alpha} Z_{\alpha} + \lambda_0$. Its m.s.e. can be written as

$$E(Z^* - Z_0)^2 = \text{Var}(Z^* - Z_0) + \left[\lambda_0 + \left(\sum_{\alpha} \lambda_{\alpha} - 1 \right) a_0 \right]^2$$

Only the bias term on the right-hand side involves λ_0 , but this time we cannot minimize it without knowledge of a_0 . One intuitive solution would be to replace a_0 by an estimate \hat{a}_0 and solve for λ_0 , but this estimate would necessarily depend on the data so that λ_0 would no longer be a constant. The only real solution is to set $\lambda_0 = 0$ and impose the condition $\sum \lambda_\alpha - 1 = 0$ on the weights λ_α . The bias $E(Z^* - Z_0)$ is then zero whatever the unknown constant a_0 .

The necessity of setting λ_0 to zero and imposing conditions on the weights λ_α to annihilate the bias is general when the mean $m(x)$ is not known. Thus from now on we only consider estimators of the form $Z^* = \sum_\alpha \lambda_\alpha Z_\alpha$.

3.4.1. Derivation and Discussion of the Equations

We want to estimate $Z_0 = Z(x_0)$ using a linear estimator $Z^* = \sum_\alpha \lambda_\alpha Z_\alpha$ and seek to minimize the m.s.e. which, as usual, can be decomposed as

$$E(Z^* - Z_0)^2 = \text{Var}(Z^* - Z_0) + [E(Z^* - Z_0)]^2$$

Given the linear model (3.12) for the mean, the bias can be expanded as

$$E(Z^* - Z_0) = \sum_\alpha \lambda_\alpha \sum_l a_l f_\alpha^l - \sum_l a_l f_0^l$$

using the notations

$$f_\alpha^l = f^l(x_\alpha) \quad f_0^l = f^l(x_0)$$

and the convention that summation on l extends over all possible values $l = 0, 1, \dots, L$. By interchanging the order of summations on l and α , we get

$$E(Z^* - Z_0) = \sum_l a_l \left(\sum_\alpha \lambda_\alpha f_\alpha^l - f_0^l \right)$$

In order to minimize $E(Z^* - Z_0)^2$, we have to make $[E(Z^* - Z_0)]^2$ zero whatever the unknown a_l , which implies annihilating their coefficient in the above. This leads to the set of $L + 1$ conditions

$$\sum_\alpha \lambda_\alpha f_\alpha^l = f_0^l \quad l = 0, 1, \dots, L \quad (3.13)$$

that Matheron (1969a) called *universality conditions*, hence the name universal kriging (UK). They express that the estimator Z^* is unbiased for *all* values of a_l .

Subject to these conditions the m.s.e. is equal to the variance of the error $Z^* - Z_0$ and depends only on covariances

$$\text{Var}(Z^* - Z_0) = \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \sigma_{\alpha\beta} - 2 \sum_{\alpha} \lambda_{\alpha} \sigma_{\alpha 0} + \sigma_{00}$$

The problem of kriging can be reformulated as follows: find N weights λ_{α} minimizing $\text{Var}(Z^* - Z_0)$ subject to the linear constraints (3.13). This is classically solved by the method of Lagrange multipliers. We consider the function

$$Q = \text{Var}(Z^* - Z_0) + 2 \sum_{l=0}^L \mu_l \left[\sum_{\alpha} \lambda_{\alpha} f_{\alpha}^l - f_0^l \right]$$

where $2\mu_l$, $l = 0, \dots, L$, are $L + 1$ additional unknowns, the Lagrange multipliers, and determine the unconstrained minimum of Q by equating the partial derivatives of Q to zero

$$\frac{\partial Q}{\partial \lambda_{\alpha}} = 2 \sum_{\beta} \lambda_{\beta} \sigma_{\alpha\beta} - 2 \sigma_{\alpha 0} + 2 \sum_l \mu_l f_{\alpha}^l = 0 \quad \alpha = 1, \dots, N$$

$$\frac{\partial Q}{\partial \mu_l} = 2 \left[\sum_{\alpha} \lambda_{\alpha} f_{\alpha}^l - f_0^l \right] = 0 \quad l = 0, 1, \dots, L$$

(That the extremum is indeed a minimum is again guaranteed by the convexity of $\text{Var}(Z^* - Z_0)$ as a function of the λ_{α} .) This leads to the following set of $N + L + 1$ linear equations with $N + L + 1$ unknowns

Universal Kriging System

$$\begin{cases} \sum_{\beta} \lambda_{\beta} \sigma_{\alpha\beta} + \sum_l \mu_l f_{\alpha}^l = \sigma_{\alpha 0} & \alpha = 1, \dots, N \\ \sum_{\alpha} \lambda_{\alpha} f_{\alpha}^l = f_0^l & l = 0, \dots, L \end{cases} \quad (3.14)$$

In matrix notations the system (3.14) is of the form $\mathbf{AX} = \mathbf{B}$ with the following structure:

$$\underbrace{\begin{bmatrix} \Sigma & \mathbf{F} \\ \mathbf{F}' & \mathbf{0} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \lambda \\ \mu \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} \sigma_0 \\ \mathbf{f}_0 \end{bmatrix}}_{\mathbf{B}}$$

where Σ , λ , and σ_0 are defined as for simple kriging and where

$$\mathbf{F} = \begin{bmatrix} 1 & f_1^1 & \cdot & f_1^L \\ 1 & f_2^1 & \cdot & f_2^L \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & f_N^1 & \cdot & f_N^L \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_0 \\ \mu_1 \\ \cdot \\ \mu_L \end{bmatrix} \quad \mathbf{f}_0 = \begin{bmatrix} 1 \\ f_0^1 \\ \cdot \\ f_0^L \end{bmatrix}$$

The kriging variance is obtained by premultiplying the first N equations of (3.14) by λ_α , summing over α , and then using the last $(L + 1)$ equations. The result is the UK variance:

UK Variance

$$\sigma_{\text{UK}}^2 = E(Z^* - Z_0)^2 = \sigma_{00} - \sum_{\alpha} \lambda_{\alpha} \sigma_{\alpha 0} - \sum_l \mu_l f_0^l \quad (3.15)$$

or in matrix form

$$\sigma_{\text{UK}}^2 = \sigma_{00} - \boldsymbol{\lambda}' \boldsymbol{\sigma}_0 - \boldsymbol{\mu}' \mathbf{f}_0 = \sigma_{00} - \mathbf{X}' \mathbf{B}$$

These equations were established independently by several authors, including Zadeh and Ragazzini (1950) as an extension of Wiener's prediction theory, Goldberger (1962) in the scope of a generalized linear regression model, and Matheron (1969a) within the framework of infinite-dimensional Hilbert spaces (continuous sampling).

Conditions for Nonsingularity

The linear system (3.14) has a unique solution if and only if its matrix \mathbf{A} is nonsingular. This holds under the following set of sufficient conditions: (1) that the submatrix Σ is strictly positive definite, (2) that the submatrix \mathbf{F} is of full rank $L + 1$ (equal to the number of columns). The proof follows from straightforward matrix algebra.

Strict positive definiteness of Σ is ensured by the use of a strictly positive definite covariance function and the elimination of duplicate data points. The condition on \mathbf{F} expresses that the $L + 1$ basis functions $f^l(x)$ are linearly

independent on S :

$$\sum_l c_l f^l(x) = 0 \quad \forall x \in S \quad \Rightarrow \quad c_l = 0: \quad l = 0, \dots, L$$

This is a standard condition of “sampling design,” encountered, for example, in the theory of least squares ($\mathbf{F}'\mathbf{F}$ must be nonsingular). For one thing there must be at least as many data points as there are basis functions (thus $N \geq L + 1$). Moreover the arrangement of the points must provide enough constraints to allow the determination of the coefficients a_l in the linear model (3.12). A counterexample in 2D is when $m(x)$ is a plane and all sample points are aligned: obviously the plane is not constrained by a single line. Likewise, when $m(x)$ is a quadratic function, the system is singular if all data points lie along two lines, a circle, an ellipse, a parabola, or a hyperbola. In view of these remarks one must be careful, particularly when using moving neighborhoods, not to create singular systems by a bad selection of the data points.

Solving the Kriging Equations

When selecting a computer subroutine to solve the equations (3.14), it must be noted that the matrix \mathbf{A} of the system is no longer positive definite as for the simple kriging system (3.2). This rules out direct use of the symmetric Cholesky decomposition “which has all the virtues” (Wilkinson, 1965, p. 244). However, the problem can be reduced to the solution of two subsystems with positive definite matrices. The following procedure results directly from the matrix form of (3.14):

1. Solve $\Sigma[\mathbf{X}_1 \mid \lambda_K] = [\mathbf{F} \mid \sigma_0]$ for \mathbf{X}_1 and λ_K .
2. Compute $\mathbf{Q} = \mathbf{F}'\mathbf{X}_1$ and $\mathbf{R} = \mathbf{F}'\lambda_K - \mathbf{f}_0$.
3. Solve $\mathbf{Q}\boldsymbol{\mu} = \mathbf{R}$ (\mathbf{Q} is positive definite).
4. Compute $\boldsymbol{\lambda} = \lambda_K - \mathbf{X}_1\boldsymbol{\mu}$.

Note that λ_K is the solution of the simple kriging equations (3.2). The kriging estimator in the case of an unknown mean is thus equal to the simple kriging estimator Z_K^* computed as if the mean were zero plus a term that will be interpreted as a correction for the mean (cf. Section 3.4.6).

Equations in Variogram Terms

We can redo the derivation of kriging equations starting from the expression of the error variance written in variogram terms. As shown in Section 2.3.2, this requires only that the error $Z^* - Z_0$ is an allowable linear combination, namely that $\sum_\alpha \lambda_\alpha = 1$. This condition coincides with the first unbiasedness condition provided that the constant function 1 is included in the set of basis

drift functions $f^l(x)$. If so, we obtain

$$\text{Var}(Z^* - Z_0) = - \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \gamma_{\alpha\beta} + 2 \sum_{\alpha} \lambda_{\alpha} \gamma_{\alpha 0}$$

whose constrained minimization leads to a system similar to (3.14) with $-\gamma$ in lieu of σ

Universal Kriging System

$$\begin{cases} \sum_{\beta} \lambda_{\beta} \gamma_{\alpha\beta} + \sum_l \mu_l f_{\alpha}^l = \gamma_{\alpha 0} & \alpha = 1, \dots, N \\ \sum_{\alpha} \lambda_{\alpha} f_{\alpha}^l = f_0^l & l = 0, \dots, L \end{cases} \quad (3.16)$$

and to the kriging variance

UK Variance

$$\sigma_{UK}^2 = E(Z^* - Z_0)^2 = \sum_{\alpha} \lambda_{\alpha} \gamma_{\alpha 0} + \sum_l \mu_l f_0^l$$

with the usual notations

$$\gamma_{\alpha\beta} = \gamma(x_{\alpha}, x_{\beta}) \quad \gamma_{\alpha 0} = \gamma(x_{\alpha}, x_0) \quad f_{\alpha}^l = f^l(x_{\alpha}) \quad f_0^l = f^l(x_0)$$

Equation (3.16) is theoretically equivalent to (3.14) when a covariance function exists⁴ but is also applicable when the variogram is unbounded.

In the particular case of a constant mean, the system comprises the single drift basis function 1 and is called the *ordinary kriging* system (OK). There is no variogram analog to the simple kriging system (3.2) because then $Z^* - Z_0$ is not an allowable linear combination.

In matrix notations (3.16) takes a form similar to (3.14)

$$\begin{bmatrix} \Gamma & \mathbf{F} \\ \mathbf{F}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_0 \\ \mathbf{f}_0 \end{bmatrix}$$

This system has a unique solution if $\Gamma = [\gamma_{\alpha\beta}]$ is strictly conditionally negative definite and $\mathbf{F} = [f_{\alpha}^l]$ is of full rank. The condition on Γ is usually met just by using a valid variogram model (but beware of zonal anisotropies). However,

unlike in the covariance formulation, no obvious simplification occurs in the resolution of the system (e.g., diagonal terms of Γ are 0's).

The UK kriging weights remain invariant under scaling of the covariance or variogram, while the μ_l and the kriging variance are multiplied by that scale factor. When the variogram or covariance are stationary, the UK weights and variance are also *shift invariant* provided that the set of basis drift functions remains closed under translations (see Section 3.4.6).

3.4.2. Interpretation of the Kriging Equations

The kriging equations (3.14) or (3.16) capture four aspects of the interpolation problem:

- The geometry of the sample points, through the $\sigma_{\alpha\beta}$ or $\gamma_{\alpha\beta}$ terms. These are functions of the interpoint distances and correct for the redundancy in the information.
- The position of the estimated point x_0 with respect to the data, through $\sigma_{\alpha 0}$ or $\gamma_{\alpha 0}$.
- The lateral continuity of the phenomenon, through the covariance or variogram model.
- The presence of a systematic location-dependent effect (trend), through the drift model.

The influence of the drift model depends on its complexity in relation with the data. In the vocabulary of regression analysis $N - (L + 1)$ would represent the number of degrees of freedom left in the residuals. This interpretation does not apply here because the data are correlated, but it helps to think in these terms. At one extreme, when $N = L + 1$, the solution is completely constrained by the unbiasedness conditions, and the UK estimator reduces to a purely deterministic fit. As L decreases, there are more and more degrees of freedom left, and the probabilistic nature of UK increases.

The UK estimator always coincides with the data at the sample points, and the kriging variance there is zero. This is true even in the presence of a nugget effect, but the estimate then has a discontinuity at each data point. If one is interested, rather, in the continuous component of the phenomenon, then slightly different equations should be used (see Section 3.7.1).

Examples

3. Consider again the special case of only one sample point but assume now an unknown constant mean ($N = 1$, $L = 0$). The solution of (3.16) is simply $\lambda_1 = 1$ and $\mu = \gamma_{10}$ so that $Z^*(x_0) = Z_1$ and $\sigma_{OK}^2(x_0) = 2\gamma_{10}$. This is very different from the SK solution found in Example 1.

4. In the case of two sample points at x_1 and x_2 ($N = 2$, $L = 0$) the kriging equations yield

$$\lambda_1 = \frac{1}{2} \left[1 + \frac{\gamma_{20} - \gamma_{10}}{\gamma_{12}} \right] \quad \lambda_2 = \frac{1}{2} \left[1 + \frac{\gamma_{10} - \gamma_{20}}{\gamma_{12}} \right] \quad \mu = \frac{1}{2} [\gamma_{10} + \gamma_{20} - \gamma_{12}]$$

so that

$$\begin{cases} Z^*(x_0) = \frac{Z_1 + Z_2}{2} + \frac{(\gamma_{10} - \gamma_{20})(Z_2 - Z_1)}{2\gamma_{12}} \\ \sigma_{OK}^2(x_0) = \gamma_{10} + \gamma_{20} - \frac{(\gamma_{10} - \gamma_{20})^2}{2\gamma_{12}} - \frac{\gamma_{12}}{2} \end{cases}$$

These results are particularized in Figure 3.4 for three variogram models of type $\gamma(h) = |h|^\alpha$. The Z^* curve always goes through Z_1 and Z_2 and for large $|x_0|$ behaves like $|x_0|^{\alpha-1}$.

For $\alpha = 1$ kriging simply interpolates linearly between Z_1 and Z_2 in the interval $[x_1, x_2]$ (with $\mu = 0$) and outside assumes the value of the nearest end point. The optimal estimator of $Z(x)$ at $x_0 > x_2$ is just the last value observed. This property derives fundamentally from the Markov character of processes with independent increments with which the linear variogram is closely associated. To show this, we can use a standard (invariance principle) argument of probability theory which goes as follows: *If the solution of a problem only depends on certain characteristics (e.g., the first two moments) and if we can find the solution in an easy special case (e.g., Gaussian RF), then it is the general solution.* Here we consider the special case of a Wiener process (Brownian motion) $X(t)$ without drift. It has independent and stationary increments, with a Gaussian distribution, and enjoys the Markov property of conditional independence: once we know the value $X(t)$ reached by the process at time t , its future does not depend on the path it took to get there. To predict $X(t + \tau)$ on the basis of past values $\{X(t') : t' \leq t\}$, the only relevant information is $X(t)$. The derivation of the predictor is straightforward starting from the decomposition

$$X(t + \tau) = X(t) + [X(t + \tau) - X(t)]$$

Since the increment $X(t + \tau) - X(t)$ is independent of $X(t)$, and of any earlier value, we have immediately

$$E[X(t + \tau) | X(t'), t' \leq t] = X(t) \quad (\tau \geq 0)$$

This estimator is clearly unbiased and optimal, and it happens to be linear. This is the kriging solution associated with a linear variogram in 1D. Stock prices, for example, have been modeled by a process with independent increments (e.g., Box and Jenkins, 1976, p. 150) with the disappointing consequence that the best forecast of stock price at any time in the future is just the current value of the stock.⁵

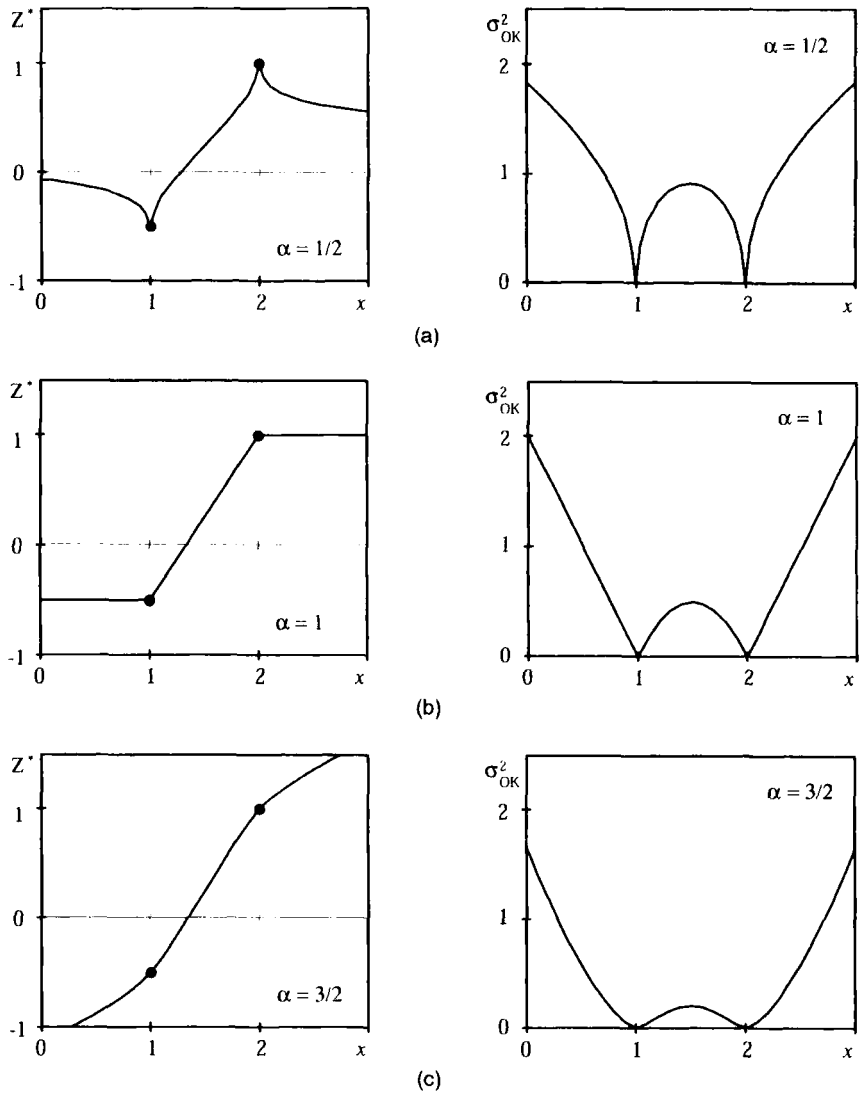


FIGURE 3.4. Ordinary kriging estimates and variance for the case of two data points and a variogram model $\gamma(h) = |h|^\alpha$: (a) $\alpha = 1/2$; (b) $\alpha = 1$; (c) $\alpha = 3/2$.

Observe that for $\alpha = 1.5$ the estimator is not confined to the data range $[Z_1, Z_2]$. When $x_0 > x_2$, for example, $\lambda_2 > 1$ and $\lambda_1 < 0$. *Kriging weights can be negative or greater than 1*, even when the mean is constant. This effect is associated with high variogram regularity (power $\alpha > 1$).

Kriging variances are zero at $x_0 = x_1$ and $x_0 = x_2$ and increase rapidly without limits as x_0 departs from the $[x_1, x_2]$ interval. Extrapolation is hazardous!

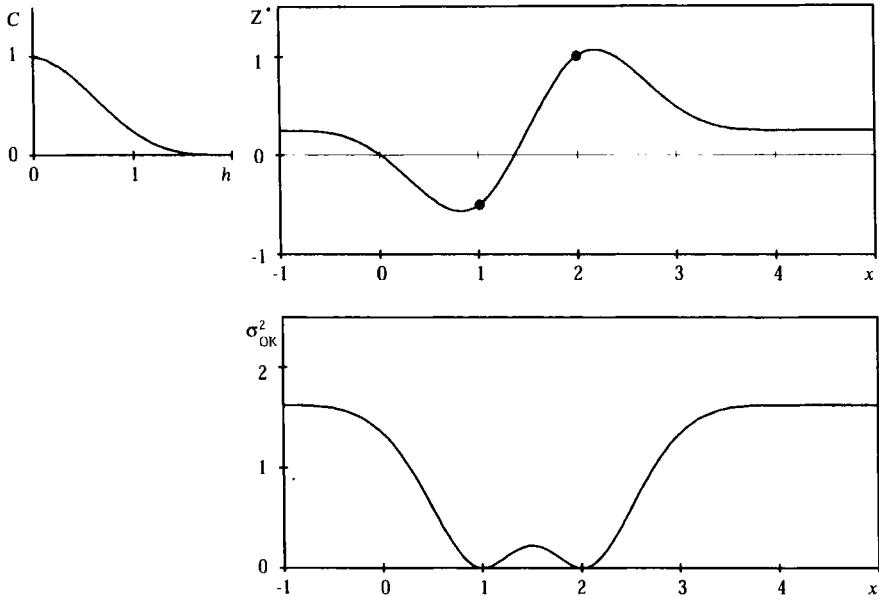


FIGURE 3.5. Ordinary kriging estimate (top) and variance (bottom) for the case of two data points and a variogram model with a sill σ^2 . Compare with simple kriging (Figs. 3.1a and 3.3a): at a distance from the data points the estimator is equal to the mean of the data (rather than the mean of the random function) and the kriging variance now exceeds the variance $C(0)$ of the RF.

Similar results are sketched in Figure 3.5 for a variogram with a finite sill σ^2 . In extrapolation the kriging estimator approaches $(Z_1 + Z_2)/2$, which is an unbiased estimator of the mean, while the kriging variance tends to $1.62\sigma^2$ and is thus larger than the global variance of $Z(x)$ itself. This is the penalty for not knowing the mean.

5. To generalize a bit the preceding example, we now consider, still in 1D, a random function with a linear variogram $\gamma(h) = b|h|$ and a linear drift $m(x) = a_0 + a_1x$. This time there are N sample points at arbitrary locations $x_1 < x_2 < \dots < x_N$. The solutions of the kriging equations (3.16) are the following, with appropriate (x_0 -dependent) values of μ_0 and μ_1 :

$$\left\{ \begin{array}{ll} x_0 \leq x_1 & Z^*(x_0) = Z_1 - \left(\frac{Z_N - Z_1}{x_N - x_1} \right) (x_1 - x_0) \\ x_i \leq x_0 \leq x_{i+1} & Z^*(x_0) = \left(\frac{x_{i+1} - x_0}{x_{i+1} - x_i} \right) Z_i + \left(\frac{x_0 - x_i}{x_{i+1} - x_i} \right) Z_{i+1} \\ x_0 \geq x_N & Z^*(x_0) = Z_N + \left(\frac{Z_N - Z_1}{x_N - x_1} \right) (x_0 - x_N) \end{array} \right.$$

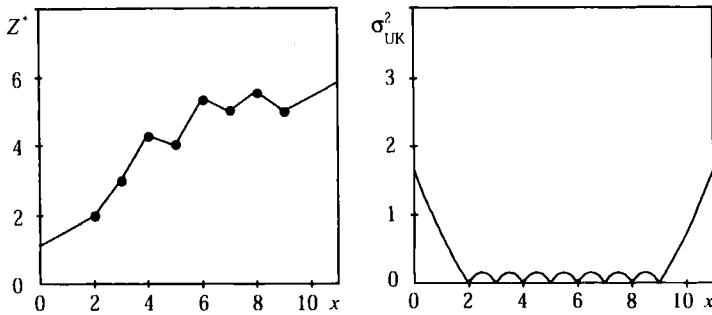


FIGURE 3.6. Universal kriging estimate and variance for the case of a linear variogram and a linear drift.

When $x_1 \leq x_0 \leq x_N$, the solution is simply linear interpolation in each subinterval. The same would hold true for a constant mean (OK). Notice again that the two adjacent sample points screen off the influence of all other data. In extrapolation the estimator is the straight line joining the first and the last data points (Figure 3.6). Considering $x_0 > x_N$, for example, the solution may be regarded as the sum of the OK estimator Z_N and a correction $\hat{a}_1 (x_0 - x_N)$ for the linear drift, in which $\hat{a}_1 = (Z_N - Z_1)/(x_N - x_1)$ is the estimator of the slope. Since $(Z_N - Z_1)$ and $(Z_0 - Z_N)$ are uncorrelated increments, the OK variance $2\gamma_{0N}$ is simply augmented by the variance of the drift correction, leading to a parabolic growth

$$\sigma_{UK}^2 = E(Z^* - Z_0)^2 = 2b(x_0 - x_N) \left(1 + \frac{x_0 - x_N}{x_N - x_1} \right) \quad (x_0 > x_N)$$

Extrapolation is even more hazardous when there is a drift!

3.4.3. Sensitivity of Kriging to Misspecification of the Variogram

So far we have assumed that the variogram was known exactly. In reality it is *estimated*, and since this estimation is not perfect, two things happen: (1) the computed kriging estimates are different from the optimal kriging estimates, and (2) the computed kriging variances are different from the true error variances associated with the suboptimal estimates. How large can these differences be?

An answer to point (1) is that the quality of the kriging estimates obtained differs from the true optimum by an amount that is *second order* in the precision to which the optimal solution is determined. In other words, even a fairly crudely determined set of kriging weights can give excellent results when it is applied to data. A similar observation is made by Press et al. (1992, p. 548) regarding the determination of the optimal Wiener filter (analogous to kriging with random errors). The important factor is to make sure that the variogram behavior near the origin is correctly represented. For example, it matters to know that the behavior is linear, but the slope itself has no influence on the kriging weights. It takes a gross misspecification of the variogram model to have a dramatic impact on kriging estimates, such as using a continuous model when significant noise is present in the data. It is also known that extremely regular models such as the Gaussian covariance should be avoided because of their inherent numerical instability—but they

are acceptable when used in conjunction with a nugget effect, even if small. On the contrary, for point (2), the computed kriging variance is directly affected by the variogram fit.

Since the computed kriging variance is obtained by “plugging-in” an estimated variogram model assumed known without error, some authors point out that the reported kriging variance does not reflect the *total* uncertainty. This effect is indeed ignored in standard geostatistical practice because its evaluation is complicated and requires strong model assumptions. For example, the Bayesian approach to this problem is to put a prior distribution on the parameters θ of a covariance model and compute the posterior probability distribution $f(Z_0 | \mathbf{Z})$ of Z_0 given the vector of observations $\mathbf{Z} = (Z_1, \dots, Z_N)'$ using the formula (Cressie, 1991, p. 171)

$$f(Z_0 | \mathbf{Z}) = \int f(Z_0 | \mathbf{Z}, \theta) f(\theta | \mathbf{Z}) d\theta$$

The optimal estimator $E(Z_0 | \mathbf{Z})$ and its variance $\text{Var}(Z_0 | \mathbf{Z})$ are moments of this distribution.

But this approach raises some objections. First, it requires us to postulate, among other things, the analytical form of the conditional distribution of Z_0 given \mathbf{Z} and θ , while we are striving simply to estimate one or two covariance parameters! Second, after considerable numerical integration (e.g., Handcock and Wallis, 1994) the results obtained

- depend largely on the chosen prior distribution;
- are mixing two very different sources of uncertainty, the spatial variability and the uncertainty on the magnitude of this variability.

What is the user supposed to do with such information? In the authors' opinion, a simpler and more useful way to account for model uncertainty is to compare scenarios based on several variants of the fit. These variants also depend on the geostatistician's choice, but *explicitly*. The sensitivity of the kriging results to the choice of the variant shows to what extent the uncertainty on the variogram is a problem and guides the design of additional sampling to improve kriging and/or variogram determination. Scenarios are not the solution to the quest for the “correct” kriging variance but achieve the operational purpose of accounting for variogram uncertainty.

Let us also mention an alternative approach based on kriging with “fuzzy” variograms (Bardossy et al., 1990). This adaptation of kriging produces fuzzy estimates and fuzzy kriging variances. It separates the spatial uncertainty from the imprecision in the model parameters: the level of the fuzzy kriging variance reflects the probabilistic uncertainty of the interpolation, assuming exact variogram parameters, whereas the interval width of the fuzzy kriged values can be used to measure the effect of the imprecision in the variogram parameters.

A number of theoretical studies have been devoted to the sensitivity of kriging to misspecification of the variogram model (see Cressie, 1991, pp. 289–299). In essence the studies consider a tightly specified family of alternative models and determine the worst that can happen to the kriging solutions if the true model were one of this family rather than the one selected. For example, Diamond and Armstrong (1984) analyze the effect of a perturbation of the variogram, and Stein (1989) derives a general purpose upper bound for the loss of efficiency incurred for misspecifying the coefficients of a linear combination of two fixed covariance models. Pilz et al. (1997) propose to model a class of plausible variogram functions and a new kriging method to minimize the maximum possible kriging variance in that class (minimax kriging).

In case of a nonstationary mean $m(x)$, there is the additional question of the type and degree of the drift. This question is better addressed by the IRF- k theory (next chapter).

3.4.4. Use of the Kriging Variance

If the kriging error has a Gaussian distribution, this distribution is completely specified by its mean (zero) and its variance σ_K^2 . Under the assumption that

the variogram is known, the kriging variance is determined without error (i.e., is nonrandom), and it is possible to make a probabilistic statement such as

$$\Pr(|Z^* - Z_0| > 2\sigma_K) \simeq 0.05$$

which leads to the traditional 95% confidence interval for Z_0

$$[Z^* - 2\sigma_K, Z^* + 2\sigma_K]$$

When the error is not Gaussian, this interval loses its rigorous significance but still makes sense as a *nominal* (or conventional) confidence interval. The question becomes: What is the real significance level of this interval for plausible distributions of the error, not necessarily Gaussian? or equivalently: How wide should the interval be to cover a 95% probability for all distributions in that class?

The answer to the last question is astonishingly simple: $\pm 3\sigma_K$. This result is a direct consequence of an inequality established by Vysochanskii-Petunin in 1980 and discussed by Pukelsheim (1994). The only assumption, a very mild one, is that the error distribution is *continuous and unimodal*. Because of its general interest we give the complete result here and particularize it to our problem.

The inequality states that if X is a random variable with a probability density f that is nondecreasing up to a mode ν and nonincreasing thereafter and if $d^2 = E(X - \alpha)^2$ is the expected squared deviation from an *arbitrary* point α , then

$$\begin{aligned} \Pr(|X - \alpha| \geq td) &\leq \frac{4}{9t^2} \quad \forall t \geq \sqrt{\frac{8}{3}} \\ &\leq \frac{4}{3t^2} - \frac{1}{3} \quad \forall t \leq \sqrt{\frac{8}{3}} \end{aligned}$$

When X is the kriging error and $\alpha = 0$, then $d^2 = \sigma_K^2$ and

$$\Pr(|Z^* - Z_0| \geq 2\sigma_K) \leq \frac{1}{9}$$

So under the stated assumptions, the nominal confidence interval has a significance level of about 90%. In order to get a 95% interval, it suffices to let $t = 3$, since $4/81 = 0.049$. The penalty for not knowing the distribution is an interval of width $6\sigma_K$ instead of $4\sigma_K$.

For $t > 1.63$ the Vysochanskii-Petunin inequality coincides with the Gauss inequality for deviations from the mode (case $\alpha = \nu$) dating back to 1821 (see Cramér, 1945, p. 183). Note that the bound is less than half the Bienaymé-Tchebycheff bound $1/t^2$. The usefulness of the Gauss inequality in the context of kriging was pointed out by Alfaro (1993).

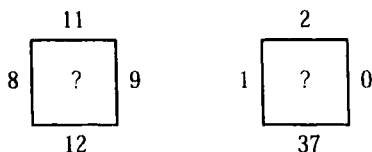


FIGURE 3.7. Which scenario is a safer bet? From Armstrong (1994), with kind permission from Kluwer Academic Publishers.

Another aspect of the kriging variance needs to be discussed. We have seen that the kriging variance does not directly depend on the data values used for the estimation: it is an *unconditional* variance. What this means is illustrated in Figure 3.7 from Ravenscroft (in Armstrong, 1994). Two blocks are estimated from four samples, they have the same kriging variance, since the data layout is the same, and also the same kriging estimate of 10. Yet clearly the right-hand case carries a higher uncertainty. The kriging variance is an *average* of such cases. If the left-hand scenario is the most frequent one, the uncertainty will occasionally be underestimated; if the right-hand scenario is the rule, the heterogeneity will translate into the variogram, and occasionally the uncertainty will be exaggerated. In this sense the kriging variance has the meaning of a spatial average as introduced in the discussion on the objectivity of probabilistic parameters for unique phenomena (Section 1.2).

Except in the case of a Gaussian RF with known mean, the kriging variance is not sufficient to model the conditional distribution of Z_0 given Z_1, \dots, Z_N . This fact sometimes comes as a disappointment, but one should not ask too much from a simple linear approach. Modeling conditional distributions is a very ambitious goal, and in fact unrealistically ambitious without a specified theoretical model (see Section 6.2.3)

The kriging variance still has its merits as a precision indicator. Referring to the terminology used by Switzer (1993), the kriging error reflects two scales of variability. The first is a “local modulation” related to the spatial configuration of the data in the kriging neighborhood: errors are smaller close to the data points, clustered samples carry less information than isolated ones, and so on. The other component is a “regional modulation” reflecting variability in the region of interest. When a global variogram is used over the whole domain, the kriging variance mainly reflects the local modulation. When data permit, however, the variogram parameters are adjusted regionally (proportional effects or other techniques), and the kriging variance, without being conditional, becomes an interesting indicator of uncertainty. Interpolated surfaces are too easily accepted as reality, especially when displayed and manipulated on powerful 3D modeling workstations. The kriging variance allows an intelligent use of these surfaces with consideration of possible errors.

3.4.5. Drift Estimation

We have seen that universal kriging provides an estimator of $Z(x_0)$ without having to estimate the mean. In fact, as intuition suggests, the mean is estimated anyhow, but *implicitly*. To see how this falls out, let us first consider the problem of drift estimation by itself.

Suppose that the objective is to estimate $m(x_0)$ at some point x_0 . Following the approach of kriging, a linear estimator is formed,

$$m^*(x_0) = \sum_{\alpha} \lambda_{\alpha} Z_{\alpha}$$

with weights selected so that

$$\begin{cases} E[m^*(x_0) - m(x_0)] = 0 & \forall a_l \\ \text{Var}[m^*(x_0) - m(x_0)] & \text{minimum} \end{cases}$$

In matrix notations where \mathbf{Z} is the N -vector of Z_{α} data and $\boldsymbol{\lambda}$ the vector of weights, this leads to the following system:

Drift Estimation System

$$\begin{cases} \boldsymbol{\Sigma} \boldsymbol{\lambda} + \mathbf{F} \boldsymbol{\mu} = \mathbf{0} \\ \mathbf{F}' \boldsymbol{\lambda} = \mathbf{f}_0 \end{cases} \quad (3.17)$$

and thus

$$m^*(x_0) = \boldsymbol{\lambda}' \mathbf{Z} = \mathbf{f}_0' (\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F})^{-1} \mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{Z}$$

with

Drift Estimation Variance

$$E[(m^*(x_0) - m(x_0))^2] = -\boldsymbol{\mu}' \mathbf{f}_0 = \mathbf{f}_0' (\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F})^{-1} \mathbf{f}_0$$

Notice that the drift estimation system (3.17) coincides with a UK system (3.14) in which all covariances $\sigma_{\alpha 0}$ on the right-hand side are zero. The UK estimator merges into the drift estimator when the estimated point x_0 is at a distance from all data points greater than the range.

An alternative derivation of these results is as follows: In the finite sample case the UK model can be regarded as a linear regression model with correlated

“residuals,” whose expression in matrix terms is

$$\mathbf{Z} = \mathbf{F}\mathbf{a} + \mathbf{Y}$$

The optimal estimator of the vector \mathbf{a} of a_l coefficients is classically obtained by generalized least squares (GLS), namely by minimizing

$$(\mathbf{Z} - \mathbf{F}\hat{\mathbf{a}})' \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \mathbf{F}\hat{\mathbf{a}})$$

over all choices of $\hat{\mathbf{a}}$ (e.g., Rao, 1973). The GLS solution is

$$\mathbf{a}^* = (\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F})^{-1} \mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{Z} \quad (3.18)$$

It is unbiased and has minimum variance. The variance-covariance matrix of this estimator is

$$\text{Cov}(\mathbf{a}^*, \mathbf{a}^*) = (\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F})^{-1}$$

That this solution coincides with that found by the other method is seen by using these drift coefficients to form the drift estimator at a given point x_0 ,

$$\hat{m}(x_0) = \sum_l a_l^* f^l(x_0) = \mathbf{f}_0' \mathbf{a}^* = m^*(x_0)$$

Turning to residuals, for the optimal choice \mathbf{a}^* the covariance of the estimated residuals is

$$\mathbf{E}(\mathbf{Z} - \mathbf{F}\mathbf{a}^*)(\mathbf{Z} - \mathbf{F}\mathbf{a}^*)' = \boldsymbol{\Sigma} - \mathbf{F}(\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F})^{-1} \mathbf{F}'$$

or, explicitly

$$\text{Cov}(Z_\alpha - m_\alpha^*, Z_\beta - m_\beta^*) = \sigma_{\alpha\beta} - \text{Cov}(m_\alpha^*, m_\beta^*)$$

The covariance of the estimated residuals is a biased estimate of the covariance of the true residuals \mathbf{Y} ; in particular, the variances are systematically underestimated:

$$\text{Var}(Z_\alpha - m_\alpha^*) = \text{Var}(Z_\alpha) - \text{Var}(m_\alpha^*)$$

One property of the GLS regression solution (3.18) worth noticing is its invariance under a change of coordinates in the linear subspace generated by the columns of \mathbf{F} . If we use a new set of basis functions $\varphi^l(x)$ that are linearly

related to the $f^l(x)$ by

$$\varphi^l(x) = \sum_s B_s^l f^s(x)$$

the drift coefficients estimates will of course be different but the value of $m^*(x_0)$ itself will not change. This is clear from (3.18): if \mathbf{F} becomes $\mathbf{F}\mathbf{B}$, with \mathbf{B} invertible, then \mathbf{a}^* becomes $\mathbf{B}^{-1}\mathbf{a}^*$ so that $\mathbf{F}\mathbf{a}^*$ remains invariant.

Case with the Variogram Only

A difficulty appears here. Since an intrinsic RF is defined only through its increments, the drift coefficient a_0 (associated with the constant function $f^0 \equiv 1$) is fundamentally indeterminate, and it is in principle impossible to estimate $m(x_0)$. In fact we are unable to calculate the variance of $m^*(x_0) - m(x_0)$ because it is not an allowable linear combination of the data: the weights of the Z data involved in this error add up to one instead of zero as they do for a kriging error.

We can work around this difficulty by considering that, over a bounded domain, the function $A - \gamma(h)$ is a covariance for some large positive constant A (this is true for all variogram models of practical interest; see “locally equivalent covariances” in section 4.6.2). Thus we can reexpress the equations (3.17) and discover that the drift estimator does not depend on the constant A (which is merged with μ_0), but its variance does.

For definiteness let us write the system for drift estimation at point x_0 in the continuous case:

$$\begin{cases} \int \lambda(dy) \gamma(y-x) + \sum_l \mu_l f^l(x) = 0 & \forall x \in S \\ \int \lambda(dx) f^l(x) = f^l(x_0) & l = 0, \dots, k \end{cases} \quad (3.19)$$

While it is not possible to calculate the variance of the drift, we can calculate that of a drift *increment* $m^*(y_0) - m^*(x_0)$:

$$\begin{aligned} E[m^*(y_0) - m^*(x_0) - (m(y_0) - m(x_0))]^2 \\ = \sum_{l=1}^L [\mu_l(y_0) - \mu_l(x_0)] [f^l(y_0) - f^l(x_0)] \end{aligned}$$

The variogram of residuals from the optimum drift estimate is found to be

$$\begin{aligned} \gamma_{\text{Res}}(x_0, y_0) &= \frac{1}{2} \text{Var}[(Z(y_0) - m^*(y_0)) - (Z(x_0) - m^*(x_0))] \\ &= \gamma(y_0 - x_0) - \frac{1}{2} \text{Var}[m^*(y_0) - m^*(x_0)] \quad (x_0, y_0 \in S) \end{aligned}$$

It systematically underestimates the theoretical variogram, and the bias can be severe (cf. Section 2.7.3).

Example

6. *Linear variogram with a linear drift* (Matheron, 1970). In \mathbb{R}^1 consider an RF $Z(x)$ with the linear variogram $\gamma(h) = b|h|$ and a linear drift $a_0 + a_1x$. A realization of $Z(x)$ is known continuously over the segment $S = [-R, +R]$. What is the optimum estimator of the drift at the point x_0 ?

To solve (3.19), we note that the Dirac measures at the points $+R$ and $-R$ satisfy

$$\frac{1}{2} \int [\delta_R(dy) + \delta_{-R}(dy)]|y - x| = R \quad \frac{1}{2} \int [\delta_R(dy) - \delta_{-R}(dy)]|y - x| = -x$$

for any $-R \leq x \leq R$ so that the measure defined by

$$\lambda(dx) = \frac{1}{2} [\delta_R(dx) + \delta_{-R}(dx)] + \frac{x_0}{2R} [\delta_R(dx) - \delta_{-R}(dx)]$$

satisfies the equations (3.19) with $\mu_0 = -bR$ and $\mu_1 = bx_0/R$. Consequently the optimum drift estimator is given by

$$m^*(x_0) = \frac{1}{2}(Z_R + Z_{-R}) + \frac{1}{2R}(Z_R - Z_{-R})x_0$$

This is simply the equation of the line joining the two end points of the interval $[-R, +R]$. We had already obtained a similar estimator of the slope in Example 5 of Section 3.4.2. The variance of a drift increment is then

$$\text{Var}[m^*(y_0) - m^*(x_0)] = \frac{b}{R}(y_0 - x_0)^2$$

3.4.6. Additivity Relationship

Coming back to the implicit nature of drift estimation in universal kriging consider the UK and the SK systems, in matrix form, the latter being distinguished by the subscript K:

$$\begin{array}{ll} \Sigma \lambda + \mathbf{F} \mu = \sigma_0 & \Sigma \lambda_K = \sigma_0 \\ \mathbf{F}' \lambda = \mathbf{f}_0 & \text{---} \end{array}$$

Subtracting the second system from the first and letting $\lambda_D = \lambda - \lambda_K$ yields

$$\begin{cases} \Sigma \lambda_D + \mathbf{F} \mu = 0 \\ \mathbf{F}' \lambda_D = \mathbf{f}_0 - \mathbf{F}' \lambda_K \end{cases} \quad (3.20)$$

The UK estimator can be decomposed into the sum

$$Z^* = Z_K^* + Z_D^* \quad (3.21)$$

of the SK estimator Z_K^* calculated as if the mean were known and subtracted from the data ($\lambda_K = \Sigma^{-1}\sigma_0$ only involves covariances), and a corrective term Z_D^* . Solving (3.20) and taking into account (3.18), it is found that Z_D^* is of the form

$$Z_D^* = \sum_l a_l^* \left(f_0^l - \sum_\alpha \lambda_{K\alpha} f_\alpha^l \right) = m^*(x_0) - \sum_\alpha \lambda_{K\alpha} m_\alpha^*$$

It is a *drift correction* involving the optimal drift estimates at point x_0 and at the x_α . Recombining this result with (3.21) gives

$$Z^*(x_0) = m^*(x_0) + \sum_\alpha \lambda_{K\alpha} (Z_\alpha - m_\alpha^*) \quad (3.22)$$

Formula (3.22) is exactly the same as (3.1) for simple kriging except that the mean m is replaced by its optimal estimator m^* . In other words, universal kriging is equivalent to optimum drift estimation followed by simple kriging of the residuals from this drift estimate, as if the mean were estimated perfectly. This property only holds when the mean is estimated in a statistically consistent manner, that is, by generalized least squares and not by ordinary least squares.

The additivity relationship (3.21) extends to variances as well. The UK kriging error is

$$Z^* - Z_0 = (Z_K^* - Z_0) + Z_D^*$$

and by the characteristic orthogonality property of SK the error $(Z_K^* - Z_0)$ has zero covariance with all Z_α , and thus with Z_D^* . Hence

$$\sigma_{UK}^2 = \sigma_{SK}^2 + \text{Var}(Z_D^*) \quad (3.23)$$

When only the variogram exists the SK estimator is not defined. Additivity relations similar to (3.21) and (3.23) can be written with the OK estimator, except that the constant drift term a_0 cancels out from Z_D^* , since OK weights add up to one (see Example 5). In all cases the drift correction variance is the price to pay for imposing unbiasedness constraints on the UK estimator. Though implicit, drift estimation is not free.

Invariance under Linear Transformation of the f^l

A direct consequence of the additivity relationship (3.21) is the invariance of the UK estimator and variance under a linear transformation of the basis drift

functions. By definition, Z_K^* does not involve the drift at all, and Z_D^* is invariant under a linear transformation of the f^l because the GLS estimator \mathbf{a}^* is itself invariant, as seen above. Now monomials of degree $\leq k$ satisfy a relationship of the form (binomial formula)

$$f^l(x+h) = \sum_s B_s^l(h) f^s(x)$$

showing that a shift of the points is equivalent to a linear transformation of the basis drift functions. Therefore, when the drift is a polynomial function of degree k , with all monomials included,⁶ the UK solution remains invariant under shifts. Notice, however, that the μ_l do change. To keep the complete kriging matrix invariant under shifts, one must tie the origin of coordinates to the kriging neighborhood, and for enhanced numerical accuracy, its center is a good place.

The shift invariance property is sometimes taken for granted. Yet it attached to the limited class of functions that are closed under translations, namely the exponential polynomials, as will be seen in the context of IRF- k theory (Section 4.3.4). In applications it is advisable to use such functions only, unless physical conditions highlight a particular point of space that must be selected as the absolute origin of coordinates.

3.4.7. Wonderful Properties Revisited

In the zero mean case three wonderful properties were established: orthogonality of the error and the data, the smoothing relationship, and conditional unbiasedness in the Gaussian case. None of these properties holds anymore when the mean is unknown because of “the curse of the μ_l .” But similar results can be stated.

Orthogonality Properties

When unbiasedness constraints are introduced, the kriging estimator is selected within a restricted class of linear combinations, and $Z^* - Z_0$ is no longer orthogonal to all Z_α . It is therefore no longer orthogonal to *all* linear combinations of Z_α but only to a *restricted class* of linear combinations, namely those which annihilate the basis drift functions f_α^l . Specifically

$$\left\langle Z^* - Z_0, \sum_\alpha \nu_\alpha Z_\alpha \right\rangle = 0 \quad (3.24)$$

for any set of weights ν_α satisfying

$$\sum_\alpha \nu_\alpha f_\alpha^l = 0 \quad l = 0, 1, \dots, L \quad (3.25)$$

Formula (3.24) follows from a straightforward reformulation of the kriging equations (3.14) or (3.16). Like for simple kriging these equations have a geometric interpretation in terms of projection in Hilbert spaces (Matheron, 1969a; Journé and Huijbregts, 1978).

The constraints (3.25) generalize the permissibility condition $\sum \nu_\alpha = 0$ encountered with the variogram and play a central role in the theory of intrinsic random functions of order k . Suffice it to say here that any linear unbiased estimator of a residual $Z(y) - m(y)$ is of the form (3.25) so that, for any linear unbiased estimator $\hat{m}(x)$ of the mean, one has

$$\langle Z^* - Z_0, Z_\alpha - \hat{m}_\alpha \rangle = 0 \quad \alpha = 1, \dots, N$$

Smoothing Relationship

When the mean is unknown there is no guarantee that $\text{Var} Z^* \leq \text{Var} Z$ because $\text{Var} Z^*$ also carries the imprecision about the estimation of the mean. But a similar inequality holds between estimated residuals. By virtue of the above, the decomposition

$$Z_0 - \hat{m}_0 = (Z_0 - Z^*) + (Z^* - \hat{m}_0)$$

is an orthogonal one for any linear unbiased estimator \hat{m}_0 of the mean so that

$$\text{Var}(Z^* - \hat{m}_0) = \text{Var}(Z_0 - \hat{m}_0) - \sigma_{\text{UK}}^2$$

which generalizes (3.4). (Incidentally we find that $\sigma_{\text{UK}}^2 \leq \text{Var}(Z_0 - \hat{m}_0)$, which simply means that Z^* is a better estimator of Z_0 than \hat{m}_0 .)

Toward Conditional Unbiasedness

Unfortunately, it is no longer true that $E(Z_0 | Z^*) = Z^*$ when the mean is estimated from the data, even in the Gaussian case. However, the following relation shows that any minimization of mean square error tends also to minimize the conditional bias

$$E(Z_0 - Z^*)^2 = E[\text{Var}(Z_0 | Z^*)] + E[E(Z_0 | Z^*) - Z^*]^2 \quad (3.26)$$

This formula is completely general (no Gaussian assumption) and follows directly from (1.5) and the fact that $E(Z_0 - Z^*) = 0$. Thus kriging, by design, tends to reduce conditional bias.

Due to mining applications conditional unbiasedness has received attention in the case of OK. The criterion used most is the slope β of the linear regression of Z_0 on Z^* , which is found to be

$$\beta = \frac{\text{Cov}(Z_0, Z^*)}{\text{Var}(Z^*)} = 1 - \left(1 - \sum_{\alpha} \lambda_{\text{K}\alpha} \right) \frac{\text{Var}(m^*)}{\text{Var}(Z^*)}$$

where the λ_{K_α} are the SK weights and $\text{Var}(m^*)$ the variance of the optimal drift estimator (3.17). One solution to achieve $\beta \simeq 1$ is to select a kriging neighborhood that is large enough for a good (implicit) estimation of the mean, thus making $\text{Var}(m^*)$ small.

It is also interesting to note that $1 - \sum_\alpha \lambda_{K_\alpha} = \lambda_m$ represents the weight placed on the mean, whether the true mean as in (3.1) or the estimated mean as in (3.22). Now, necessarily, $\text{Var}(m^*) \leq \text{Var}(Z^*)$, since m^* is by design the minimum-variance linear combination of the data subject to the constraint that weights add up to 1. Hence we have the inequalities

$$\begin{aligned} 1 - \lambda_m &\leq \beta \leq 1 & \text{if } \lambda_m &\geq 0 \\ 1 &\leq \beta \leq 1 - \lambda_m & \text{if } \lambda_m &\leq 0 \end{aligned}$$

When the weight on the mean is small, the slope β is close to 1, even if the neighborhood is not large. Thus λ_m is a criterion for selecting the size of the kriging neighborhood (Rivoirard, 1987).

3.4.8. Dual Kriging

Another interesting property to revisit is the behavior of the UK estimator as an interpolator. Using the matrix notations of (3.14), but switching to lowercase for z which is now considered deterministic, the UK estimator z_0^* can be written as

$$z_0^* = \lambda'z = [\lambda' \mu'] \begin{bmatrix} z \\ \mathbf{0} \end{bmatrix} = [z' \quad \mathbf{0}] \begin{bmatrix} \lambda \\ \mu \end{bmatrix}$$

But

$$\begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} \Sigma & \mathbf{F} \\ \mathbf{F}' & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_0 \\ \mathbf{f}_0 \end{bmatrix}$$

so that

$$z_0^* = [z' \quad \mathbf{0}] \begin{bmatrix} \Sigma & \mathbf{F} \\ \mathbf{F}' & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_0 \\ \mathbf{f}_0 \end{bmatrix} = [\mathbf{b}' \quad \mathbf{c}'] \begin{bmatrix} \sigma_0 \\ \mathbf{f}_0 \end{bmatrix}$$

Denoting here the interpolated point by x , the estimator is a linear combination of the covariances $\sigma(x_\alpha, x)$ and the drift functions $f^l(x)$:

$$z^*(x) = \sum_\alpha b_\alpha \sigma(x_\alpha, x) + \sum_l c_l f^l(x) \quad (3.27)$$

with coefficient b_α and c_l that are linear functions of the z_α defined by

$$\begin{bmatrix} \Sigma & \mathbf{F} \\ \mathbf{F}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{c} \end{bmatrix} = \begin{bmatrix} \mathbf{z} \\ \mathbf{0} \end{bmatrix} \quad (3.28)$$

This system is called the *dual kriging* system and is strictly equivalent to the UK system (Matheron, 1970). Equations can also be written in terms of the variogram provided that $f^0(x) \equiv 1$. The dual kriging system *characterizes* the UK estimator as the only interpolator of the form (3.27) satisfying

$$\begin{cases} z^*(x_\alpha) = z_\alpha & \text{for all } \alpha \\ \sum_{\alpha} b_{\alpha} f_{\alpha}^l = 0 & \text{for all } l \end{cases} \quad (3.29)$$

The term “dual” originates from an alternative derivation of these equations by minimization in a functional space, similar to splines (Matheron, 1981a,c). The usefulness of this formulation appears in connection with spline interpolation and with kriging under inequality constraints (Sections 3.9.2 and 4.6.4). Formula (3.27) shows the mixed nature of the UK approach, a functional interpolation carried by the $\sum c_l f^l(x)$ term and a probabilistic interpolation carried by the covariance term $\sum b_{\alpha} \sigma(x_{\alpha}, x)$. At one extreme ($L + 1 = 0$), the interpolation is purely probabilistic (simple kriging); at the other ($L + 1 = N$), it is purely functional.

From a practical point of view the dual kriging system makes estimation faster in the case of a global neighborhood system—and has been used unknowingly for a long time. Since the coefficients b_{α} and c_l do not change with the location x , it suffices to compute them once for all, and the estimates $z^*(x)$ fall out as simple scalar products. But the kriging variance cannot be obtained in this manner.

The interpolating function approach has been used by Galli et al. (1984b) to generate contour maps directly without going through the intermediate step of a grid. The principle is to track contours by solving $z^*(x) = \text{constant}$ numerically, starting from initial points determined along an “exploration path.” By contrast, when a grid is used, the contours are drawn from the grid values and the initial data points are forgotten, leaving the possibility that some of them fall on the wrong side of a contour. The tracking algorithm eliminates that problem but creates a new one: the risk of missing some contours.

3.4.9. A Bayesian Bridge between Simple and Universal Kriging

This title of a paper by Omre and Halvorsen (1989) conveys the idea very clearly. Simple and universal kriging can be viewed as extreme cases in which we either know the drift perfectly or else know nothing about it. A Bayesian model introduced by Omre (1987) assumes that some prior knowledge is available about the drift and establishes a continuum between these two approaches.

As a background it is interesting to first reconsider a variant of the UK model, the *random drift* model, proposed early on by Matheron (1970). Specifically

$$Z(x) = Y(x) + M(x)$$

where $E[Y(x)] = 0$ and $M(x)$ is now a random function, not necessarily independent of $Y(x)$ but much smoother than $Y(x)$. So the dichotomy is physically meaningful, and $M(x)$ may be expanded

in the usual way as

$$M(x) = \sum_l A_l f^l(x) \quad (3.30)$$

except that now the coefficients are *random*. Define the first two moments

$$E(A_l) = a_l \quad \text{and} \quad \text{Cov}(A_l, A_s) = K_{ls}$$

and the noncentered covariances

$$\begin{cases} E[Y(x)Y(y)] = \sigma(x, y) \\ E[Y(x)M(y)] = R(x, y) \\ E[M(x)M(y)] = K(x, y) + \sum_l \sum_s a_l a_s f^l(x) f^s(y) \end{cases}$$

The (nonstationary) covariance of $M(x)$ is then simply

$$K(x, y) = \sum_l \sum_s K_{ls} f^l(x) f^s(y) \quad (3.31)$$

and the cross covariance between $Y(x)$ and $M(y)$ is of the form

$$R(x, y) = \sum_l E[Y(x)A_l] f^l(y) = \sum_l R_l(x) f^l(y)$$

If $Y(x)$ is a stationary RF and A_l a random variable, $R_l(x) = R_l$ is a constant, and assuming as usual that $f^0 \equiv 1$, we have the expansion

$$R(x, y) = \sum_l \sum_s R_{ls} f^l(x) f^s(y) \quad (3.32)$$

where $R_{ls} = 0$ for $s > 0$. If the mean is not assumed known, we consider an estimator of Z_0 of the form

$$Z^* = \sum_\alpha \lambda_\alpha Z_\alpha = \sum_\alpha \lambda_\alpha Y_\alpha + \sum_\alpha \lambda_\alpha M_\alpha$$

whose mean square error, using the foregoing relationships, can be written as

$$\begin{aligned} E(Z^* - Z_0)^2 &= E \left(\sum_\alpha \lambda_\alpha Y_\alpha - Y_0 \right)^2 + \sum_l \sum_s (K_{ls} + a_l a_s + 2R_{ls}) \\ &\quad \times \left(\sum_\alpha \lambda_\alpha f_\alpha^l - f_0^l \right) \left(\sum_\beta \lambda_\beta f_\beta^s - f_0^s \right) \end{aligned}$$

Here Matheron argues that it is not possible to estimate the terms $K_{ls} + a_l a_s$ from the data Z_α because the noncentered covariance of $M(x)$ is not stationary, and even if it were, its inference would be very poor because of a too high regularity, as was shown in Section 2.9.2. Therefore the only solution is to cancel the other terms and set

$$\sum_{\alpha} \lambda_{\alpha} f_{\alpha}^l - f_0^l = 0 \quad \forall l$$

Minimizing the mean square error under these conditions leads to the usual UK system.

Note that this result really hinges on the expansions (3.31) and (3.32) of the covariance functions rather than on (3.30) and would hold as an approximation whenever the functions $K(x, y)$ and $R(x, y)$ vary slowly in space.

Omre and Halvorsen (1989) start from a similar model except that they assume Y and M to be independent RFs (hence $R(x, y) \equiv 0$) and that the means a_l and covariances K_{ls} of the random drift coefficients are known from prior knowledge. Thus

$$\left\{ \begin{array}{l} E[Z(x) | A_l : l = 0, \dots, L] = M(x) = \sum_l A_l f^l(x) \\ \text{Cov}[Z(x), Z(y) | A_l : l = 0, \dots, L] = \sigma(x, y) \\ E[Z(x)] = m(x) = \sum_l a_l f^l(x) \\ \text{Cov}[Z(x), Z(y)] = \sigma(x, y) + \sum_l \sum_s K_{ls} f^l(x) f^s(y) \end{array} \right.$$

The last formula results from the general expression

$$\text{Cov}[Z(x), Z(y)] = E[\text{Cov}(Z(x), Z(y) | M)] + \text{Cov}[E(Z(x) | M), E(Z(y) | M)]$$

where for simplicity the conditions have been denoted by M .

Consider the following unbiased estimator of Z_0 where prior means are involved:

$$Z^* = \sum_{\alpha} \lambda_{\alpha} (Z_{\alpha} - m_{\alpha}) + m_0$$

It is optimized by unrestricted minimization of the variance

$$E(Z^* - Z_0)^2 = E \left(\sum_{\alpha} \lambda_{\alpha} Y_{\alpha} - Y_0 \right)^2 + \sum_l \sum_s K_{ls} \left(\sum_{\alpha} \lambda_{\alpha} f_{\alpha}^l - f_0^l \right) \left(\sum_{\beta} \lambda_{\beta} f_{\beta}^s - f_0^s \right)$$

This is the bridge between simple and universal kriging. At one end we have exact prior knowledge of the drift, so $K_{ls} = 0$; at the other end we have complete prior ignorance, so $K_{ls} \rightarrow \infty$ and the weights must satisfy the unbiasedness constraints of UK to keep the expression finite. In between, the "Bayesian kriging" equations are just those of SK but with the nonstationary covariance $\text{Cov}[Z(x), Z(y)]$ defined above. As with standard SK the solution is an exact interpolator, and no minimum number of points is required; in particular, we may have $N < L + 1$. Evidently the associated error variance is comprised between the SK variance and the UK variance.

The central question now is this: How can one make a “qualified guess” of the drift, that is, come up with an estimate and its associated uncertainty before any Z data become available? In the applications presented, this guess is always obtained from a different but related data set. Abrahamsen (1993), for example, builds a comprehensive multilayer reservoir model in which the Z variables are depths of geological surfaces measured in wellbores and the random drift is a function of seismic reflection time. The prior means and covariances of the coefficients of this function are derived from the analysis of velocity information at the wellbores. This technique uses seismic times and squared times as basis drift functions and is akin to the external drift method presented in Section 5.7.3, except that some knowledge is assumed about the drift coefficients.

Pilz (1994) extends the theory to the case where only partial prior knowledge of the first two moments of the drift coefficients is available. This knowledge in effect restricts the possible prior distributions of the drift coefficients to a subfamily, and a “Bayes robust” estimator is derived by minimizing the maximum mean square estimation error over all possible prior distributions in this subfamily. For example, in the case of ordinary kriging we may know that the expected value of M , namely the true mean, lies in a given interval $[a, b]$ and that the variance of M is less than some value K_0 . Then, using the notations of the additivity formula (3.22) for the UK estimator, the Bayes robust estimator derived by Pilz is

$$\hat{Z}_B = Z_K^* + \left(1 - \sum_{\alpha} \lambda_{K\alpha}\right) \hat{m}_B$$

where \hat{m}_B is related to the optimum estimator of the mean m^* given by (3.18), the interval midpoint $m_0 = (a + b)/2$, the uncertainty variance K_0 , and the largest interval variance $K_1 = (b - a)^2/4$ by

$$\hat{m}_B = \frac{m^* + \theta m_0}{1 + \theta} \quad \text{with} \quad \theta = \frac{\text{Var} m^*}{K_0 + K_1}$$

(in the above m^* and $\text{Var} m^* = (\mathbf{1}' \Sigma^{-1} \mathbf{1})^{-1}$ are conditional on $M = m$). In general, \hat{m}_B and \hat{Z}_B are biased estimators, but \hat{Z}_B achieves a smaller mean square error σ_B^2 than the UK estimator. Specifically⁷

$$\sigma_{UK}^2 - \sigma_B^2 = \frac{\left(1 - \sum_{\alpha} \lambda_{K\alpha}\right)^2 (\text{Var} m^*)^2}{\text{Var} m^* + K_0 + K_1}$$

The Bayesian approach can also be used to study the effect of uncertainty on the covariance parameters (e.g., Handcock, 1994), but the theory becomes more complex because of nonlinearities. The noticeable difference is that incorporating the uncertainty on the covariance tends to increase the uncertainty on the final results, whereas modeling the uncertainty on the drift through a Bayesian analysis tends to *reduce* it.

Under the impulse of the “Norwegian school” led by H. Omre, the Bayesian approach in geostatistics has seen a new wave of applications to petroleum problems where data are initially scarce and replaced by explicit prior guesses (see Section 7.9.3).

3.4.10. Lognormal Kriging and Generalization

The linear estimators considered so far involved no assumption on the finite-dimensional distribution of the data other than on first and second moments. We saw that this worked very well in the case of a Gaussian RF with a known mean because then the simple kriging estimator coincides with the regression function of $Z(x)$ on the data. If, on the other hand, we know that the regression

function is highly nonlinear, it would not be very smart to use linear estimators. The classic and important example is the lognormal case, namely when the logarithm of the random function $Z(x)$ is a Gaussian RF

$$Y(x) = \log Z(x) \sim \mathcal{N}(m(x), \sigma^2)$$

Assume first that the mean is known. The conditional expectation of Y_0 given Y_1, \dots, Y_N is a Gaussian with mean the simple kriging estimator Y^* and variance σ_{SK}^2 . It follows that the optimum estimator of $Z_0 = e^{Y_0}$ is given by

$$Z^* = E(e^{Y_0} | Y_1, \dots, Y_N) = \exp(Y^* + \frac{1}{2}\sigma_{SK}^2)$$

and that the conditional estimation variance is

$$\text{Var}(Z^* - Z_0 | Z_1, \dots, Z_N) = \text{Var}(e^{Y_0} | Y_1, \dots, Y_N) = (Z^*)^2 [\exp(\sigma_{SK}^2) - 1]$$

Note that this conditional variance depends on the data values but that the variance scaled by $(Z^*)^2$ does not.

In case of an unknown mean the problem becomes more complex, even when the mean is constant. It is possible to construct optimal linear estimators in the logarithmic scale and also devise a reverse transformation that ensures unbiasedness, but the optimality properties of such procedures are unclear. This question has been discussed in detail by Matheron (1974a). Considerable simplification occurs if we accept to work with quantiles rather than moments because they simply follow the transformation. Since the distribution of the error $Y^* - Y_0$ is symmetric (Gaussian), its median coincides with its mean and is thus zero; therefore $\exp(Y^*)$ is a *median unbiased* estimator of Z_0 :

$$\Pr\{\exp(Y^*) \geq Z_0\} = \Pr\{\exp(Y^*) \leq Z_0\} = \frac{1}{2}$$

Likewise any confidence interval can be reversed, and a 95% confidence interval for Z_0 is

$$[\exp(Y^* - 2\sigma_K), \exp(Y^* + 2\sigma_K)]$$

If one insists on having an unbiased estimator, then

$$Z^* = \exp(Y^* + \frac{1}{2}(\text{Var } Y_0 - \text{Var } Y^*)) \quad (3.33)$$

is the natural solution. The correction term $\text{Var } Y_0 - \text{Var } Y^*$ only involves the variogram of $Y(x)$ provided that $\sum \lambda_\alpha = 1$. Then, from (3.16),

$$\text{Var } Y_0 - \text{Var } Y^* = \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \gamma_{\alpha\beta} = \sigma_K^2 - 2 \sum_l \mu_l f_0^l$$

Even though this formula allows $\gamma(h)$ to be unbounded, the lognormal model requires, at least in concept, to use a covariance of the form $C(h) = A - \gamma(h)$ for a suitably large positive constant A adjusted for the domain of definition D of $Z(x)$ (cf. locally equivalent stationary covariances in Section 4.6.2).

The unbiased estimator (3.33) must be used with caution because it is non-robust against departures from the lognormal model. Also, contrary to linear kriging, the variogram sill or slope directly affects the estimator itself, making the correction formula very sensitive to variogram fluctuations. David (1988, p. 119) reports "horror stories" on this subject.⁸ To avoid gross errors, various schemes are used to "calibrate" the estimates on the untransformed data. For example, Journel and Huijbregts (1978, p. 572) propose, in the stationary case, to scale the kriging estimates by a constant factor such that the arithmetic mean of these estimates equals the estimate of the mean obtained directly from the Z data. But the new estimates no longer honor the data points (the technique was suggested for block estimation). To preserve the interpolation property, a modified estimator can be defined as

$$Z^{**} = \exp(Y^* + B \frac{1}{2}(\text{Var } Y_0 - \text{Var } Y^*))$$

with a calibration factor B to be determined by cross-validation (Delfiner, 1977). This procedure amounts to fine tuning the sill of the variogram of Y .

A common approach is to perform two studies in parallel, one on Z and one on $\log Z$, and compare the results in light of the properties of the lognormal model. For example, from (2.79) the variograms $\Gamma(h)$ of $Z(x)$ and $\gamma(h)$ of $Y(x)$ are related in the stationary case by

$$\Gamma(h) = M^2 e^A [1 - e^{-\gamma(h)}]$$

where M is the mean of Z and A the sill of $\gamma(h)$ so that typically, if $\gamma(h)$ is linear at the scale of the study, then $\Gamma(h)$ is exponential. A practitioner has confessed to locking himself behind thick doors to average the linear and lognormal kriging estimates.

The lognormal model is a natural for positive skewed data, such as low mineral grades, pollution levels, or permeability. Often it is used in the context of block estimation. See Switzer and Parker (1976), Dowd (1982), David (1988), and Rivoirard (1990) for further reading and references on the use of the lognormal model in mining. In principle, this model requires stationary data as do all models involving distributional assumptions, since residuals tend to be severely biased. But highly skewed data may exhibit a strong drift even after logarithmic transformation. For example, after a nuclear test, plutonium concentrations in the soil range over almost 5 orders of magnitude, with a large maximum at the detonation point and a rapid falloff in all directions (Delfiner and Gilbert, 1978).

General Transformation

The lognormal kriging approach can be generalized to an arbitrary transformation $Z = \varphi(Y)$ where Y is a Gaussian RF. In the stationary case the function φ is determined from the histogram of the data, and if this is continuous (no accumulation of frequency at discrete values), φ has an inverse. We can then work on the Y values for which linear estimators are well suited and transform the results back. Of course assuming that the RF $Y(x)$ is Gaussian as a whole just because its marginal distribution is Gaussian takes a big leap of faith, but if we go for it, the wonderful properties of simple kriging provide a general formula for reducing the bias in the reverse transformation.

Borrowing from a standard argument of bias reduction in statistical estimation (e.g., Cox and Hinkley, 1974, p. 260), suppose that φ is regular enough to have a Taylor expansion at order 2 and that the simple kriging error $Y^* - Y$ is relatively small, then

$$\varphi(Y) \simeq \varphi(Y^*) + (Y - Y^*)\varphi'(Y^*) + \frac{1}{2}(Y - Y^*)^2\varphi''(Y^*)$$

Taking the conditional expectation given Y^* and applying the results (3.11), we get

$$E[\varphi(Y) | Y^*] \simeq \varphi(Y^*) + \frac{1}{2}\sigma_{SK}^2\varphi''(Y^*)$$

So the estimator

$$Z^* = \varphi(Y^*) + \frac{1}{2}\sigma_{SK}^2\varphi''(Y^*) \quad (3.34)$$

approximately satisfies the conditional unbiasedness relationship

$$E(Z^* | Y^*) = E(Z | Y^*)$$

which is stronger than just unbiasedness. The correction formula hinges on the assumption of small kriging variances and the correct determination of the variogram sill.

In the lognormal case, (3.34) translates into

$$Z^* = \exp(Y^*)[1 + \frac{1}{2}\sigma_{SK}^2]$$

and does recover the leading term in the series expansion of the exact solution $\exp(Y^* + \frac{1}{2}\sigma_{SK}^2)$.

3.5. ESTIMATION OF A SPATIAL AVERAGE

Kriging was originally developed not for estimation of point values but of average grades over mining panels. Its main goal was to avoid the systematic overestimation that takes place when high-grade panels are selected solely on the basis of internal samples. To understand this important problem, it is interesting to look at it in its original form and explain the approach proposed by Krige (1951), one of the pioneers of geostatistical methods.

3.5.1. Krige's Regression Effect

When a panel is selected because it contains high-grade samples, the samples surrounding it tend, by definition, to have lower grades. These data should also be included in the estimation of the panel or else the sampling is biased,

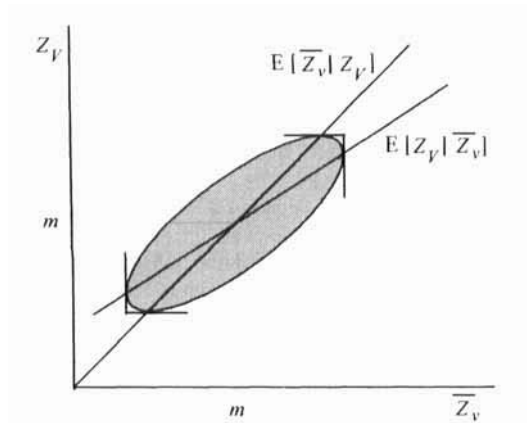


FIGURE 3.8. The regression effect. On the average the true panel grade is less than the sample mean grade for $\bar{Z}_v > m$ and greater for $\bar{Z}_v < m$.

and this is a cause of severe systematic error (here a constant mean is considered). On the average, high-grade panels are poorer than their internal samples suggest, and low-grade panels are richer.

Gold miners in South Africa were aware of this fact and used empirical correction factors. Krigé provided a theoretical justification for these corrections in the scope of regression theory. He started from the assumption that the expected value of the mean grade \bar{Z}_v of samples taken inside a panel is equal to the panel's true mean grade. This is an exact property if the samples are selected at random within the panel, or an approximate one if a subgrid is used. For the sake of simplicity, let us assume that grades are Gaussian (Krigé considered the lognormal case). Then the regression line giving the conditional expectation of \bar{Z}_v as a function of the panel grade is simply the first bisector (Fig. 3.8). Necessarily the other regression line

$$E(Z_v | \bar{Z}_v) = m + \beta(\bar{Z}_v - m) \quad (3.35)$$

relating the expected panel grade to the mean sample grade has a slope β less than one (the product of slopes is the square of the correlation coefficient). This is the correction formula used by Krigé: it pulls the estimate toward the mean.

Now, in practice, the overall mean m is not known and is replaced by the mean \bar{Z} of the samples in the ore body so that (3.35) is evaluated by

$$\hat{Z}_v = \bar{Z} + \beta(\bar{Z}_v - \bar{Z})$$

This is a linear combination of data of the form

$$\hat{Z}_v = \sum_{\alpha} \lambda_{\alpha} Z_{\alpha}$$

that assigns the same weight to each sample inside the panel and another constant weight $(1 - \beta)/N$ to each outside sample, the weights adding up to one.

Kriging generalizes this approach by personalizing the weight assigned to each sample.

3.5.2. Kriging Equations

We want to estimate the mean value of $Z(x)$ over a block v , using a linear combination of “point” data $Z(x_\alpha)$. We will consider directly the case of an unknown mean. The theory follows exactly the steps explained in Section 3.4.1, the only differences arising from the objective which is now

$$Z_0 = \frac{1}{|v|} \int_v Z(x) dx \quad (= Z_v)$$

The derivation involves the mean values of the drift functions over the block

$$f_0^l = \frac{1}{|v|} \int_v f^l(x) dx \quad (= f_v^l)$$

the covariances $\sigma_{\alpha 0}$ between each sample Z_α and the block, and the variance σ_{00} of the block

$$\sigma_{\alpha 0} = \frac{1}{|v|} \int_v \sigma(x_\alpha, x) dx \quad (= \sigma_{\alpha v})$$

$$\sigma_{00} = \frac{1}{|v|^2} \int_v \int_v \sigma(x, y) dx dy \quad (= \sigma_{vv})$$

With these values for f_0^l , $\sigma_{\alpha 0}$, and σ_{00} , the kriging system remains the same as (3.14), and the kriging variance is (3.15). (In parentheses are the standard geostatistical notations for block averages.)

As before these results can be re-expressed in terms of the variogram leading to the system (3.16) with the appropriate f_0^l and $\gamma_{\alpha 0}$ ($= \gamma_{\alpha v}$). For the kriging variance, however, the term $-\gamma_{00}$ corresponding to σ_{00} is no longer zero as in the case of point kriging but

$$\gamma_{00} = \frac{1}{|v|^2} \int_v \int_v \gamma(x, y) dx dy \quad (= \gamma_{vv})$$

and the kriging variance is

$$E(Z^* - Z_0)^2 = -\gamma_{00} + \sum_\alpha \lambda_\alpha \gamma_{\alpha 0} + \sum_l \mu_l f_0^l$$

Likewise the theory can be developed for the estimation of an arbitrary moving average

$$Z_0 = \int p_0(x)Z(x)dx \quad \text{with} \quad \int p_0(x)dx = 1$$

That just changes the expressions of f_0^l , $\sigma_{\alpha 0}$, and σ_{00} , which become

$$f_0^l = \int f^l(x)p_0(x)dx$$

$$\sigma_{\alpha 0} = \int \sigma(x_\alpha, x)p_0(x)dx$$

$$\sigma_{00} = \iint \sigma(x, y)p_0(x)p_0(y)dx dy$$

3.5.3. Piecing Together Local Estimates

If the domain V to be estimated is the union of several nonoverlapping blocks v_i , the mean grade Z_0 of V is related to the mean grades Z_i of the v_i ,

$$Z_0 = \frac{1}{|V|} \sum_i |v_i|Z_i \quad \text{with} \quad |V| = \sum_i |v_i|$$

If the estimation of Z_0 and the Z_i is carried out from the same data it follows immediately from the linearity of the kriging system (3.14) that the estimators are in the same relationship, namely

$$Z^* = \frac{1}{|V|} \sum_i |v_i|Z_i^*$$

This property allows a global estimation by piecing together local kriging estimates. In practice, the local estimates are not calculated from all the data but only those of a neighborhood. If this is well designed, the result is close to the optimum that would be obtained if all data were used. Note, however, that there is no similar relationship between kriging variances (the covariances between kriging errors would need to be introduced).

3.5.4. A Case Study in Forest Inventory

We consider again the tropical forest example introduced in Section 2.9.1 (from Narboni, 1979), in which an exhaustive survey of trees of the “gaboon” species (a variety of mahogany) was conducted over an area of 20,000 ha (1 hectare = 10,000 m² = 2.471 acres). This involved the analysis of 80,000 sampling units of 50 m × 50 m. Such considerable work was motivated by the

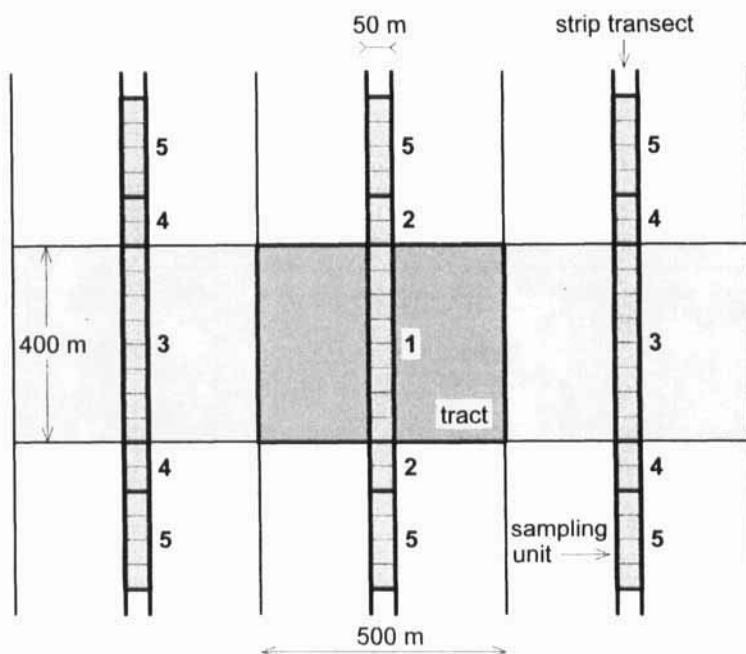


FIGURE 3.9. Kriging configuration for a tract (hatched area). The numbers define the rings. From Narboni (1979), © Bois et Forêts des Tropiques.

desire to validate the geostatistical approach by comparison with the traditional estimation method, and most important, with *reality*.

To ensure stationarity, the area was divided into four approximately equal zones of 5000 ha (Ngolo I through IV). Inventory estimation was carried out on 500 m × 400 m rectangular tracts using 10% of the data. Tracts are centered on 50-m wide strip transects further divided into the basic sampling units (Fig. 3.9). The classic estimate \hat{Z}_1 is simply the mean of the sampling units inside the tract. Kriging, on the other hand, also uses information from outside the tract. Data that are close or symmetric with respect to the center of the estimated tract play a similar role in the kriging equations and receive the same weights. For computational simplicity they are aggregated within rings, and an overall weight is applied to the mean value within the ring (this is sometimes called *random kriging*). Using a pattern of five rings as depicted in Figure 3.9, we have a loss of precision on variance of about 1%. In case of systematic sampling, the same kriging configuration can be used for all tracts, and the weights $\lambda_1, \dots, \lambda_5$ assigned to each ring are calculated once for all.

Table 3.2 compares the mean estimates over the four zones; multiplication by the areas gives the total number of trees. Notice that outside samples are weighted more than inside samples ($\lambda_1 < 0.5$). This is due to the presence of a large nugget effect (see Figure 2.30).

TABLE 3.2. Classic and kriging estimates compared with reality

NGOLO Zones	\bar{Z}_v	\bar{Z}_1	Kriging Weights					\bar{Z}^*	σ_K^2	$2\sigma_K/\bar{Z}^*$
			λ_1	λ_2	λ_3	λ_4	λ_5			
I	1.18	1.25	0.38	0.10	0.18	0.08	0.26	1.25	0.20	72%
II	1.26	1.29	0.38	0.08	0.18	0.08	0.28	1.28	0.23	75%
III	1.53	1.58	0.44	0.10	0.16	0.06	0.24	1.53	0.34	76%
IV	1.69	1.61	0.41	0.10	0.17	0.07	0.25	1.62	0.36	74%

Note: \bar{Z}_v : true value; \bar{Z}_1 : mean of classic estimates; \bar{Z}^* : mean of kriging estimates; σ_K^2 : kriging variance. From Narboni (1979).

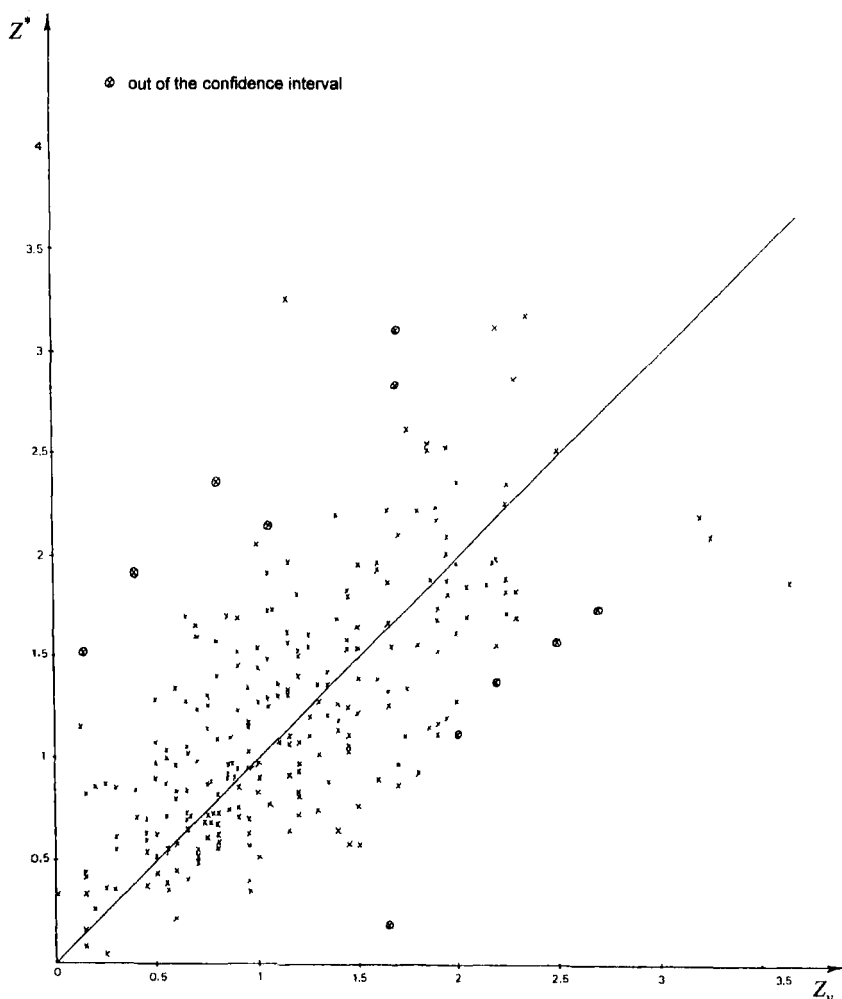


FIGURE 3.10. Kriging estimates (Y-axis) and true values (X-axis). Each point represents a 20-ha tract (10% sampling on Ngolo I). From Narboni (1979), © Bois et Forêts des Tropiques.

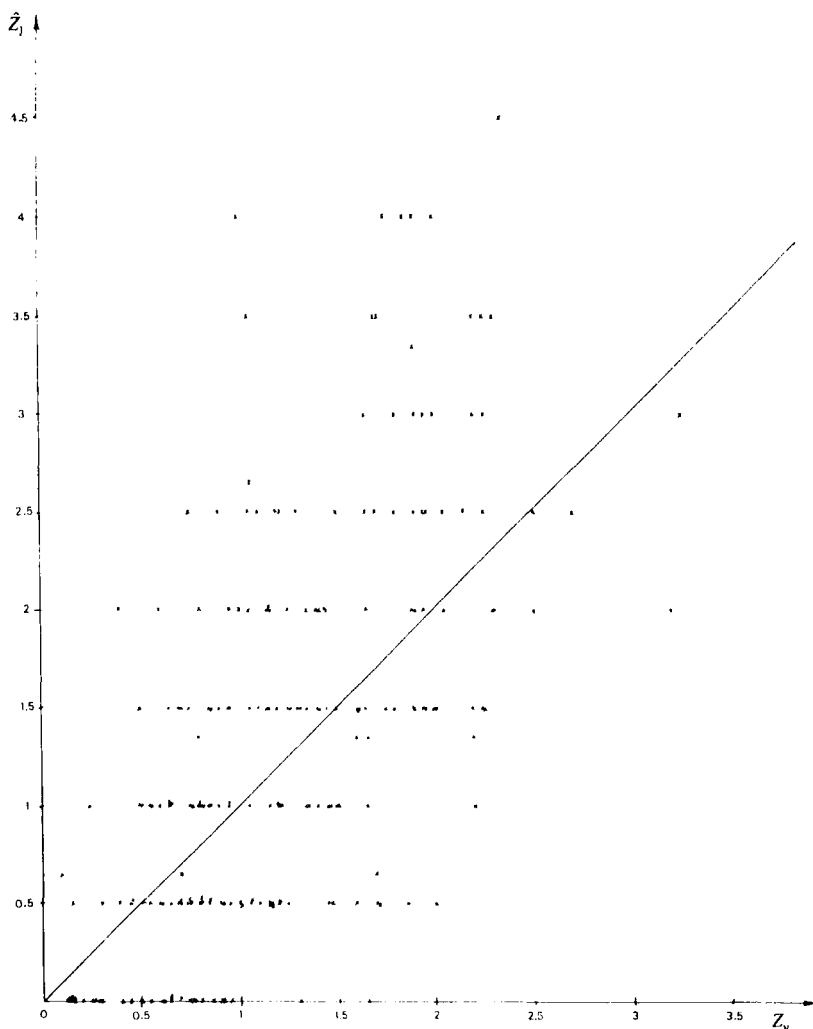


FIGURE 3.11. Classic estimates (Y -axis) and true values (X -axis). Each point represents a 20-ha tract (10% sampling on Ngolo I). From Narboni (1979), © Bois et Forêts des Tropiques.

If we only compare the means of the classic and the kriging estimates, we find very little difference, except for the third zone where kriging provides a better estimate. The two methods are nearly equivalent for *global estimation*. But these aggregates average out a large disparity of *local* situations. Figures 3.10 and 3.11 show the difference in a striking manner. The scatterplot of true values (plotted horizontally) versus kriging estimates is nicely centered about the unit slope line, with 11 points outside the 95% confidence interval (we expect about $250 \times 0.05 = 12.5$ points). The local kriging estimator satisfies the conditional unbiasedness property $E(Z_v | Z^*) = Z^*$, even though ordinary

kriging rather than simple kriging is used. Compare this with the classic estimator \hat{Z}_1 in Figure 3.11: tracts such that $\hat{Z}_1 > m$ are overestimated, whereas tracts such that $\hat{Z}_1 < m$ are underestimated. This is exactly the regression effect discussed earlier.

3.5.5. Estimation over a Random Support

It may be that the variable of interest is not suitable to a direct kriging but is expressed as the product or the ratio of two variables which are. A typical example is the estimation of a layer with variable thickness in a domain whose lateral extent only is known. The study is then usually carried out in two dimensions. The operational variables are vertical integrals of point-properties (they are sometimes called *service variables*). For example, in an assessment of underground pollution concentration, one defines the polluted *thickness* $H(x)$ ($x \in \mathbb{R}^2$) and the pollutant *accumulation* $A(x)$, which represents the product of thickness by the mean vertical pollutant concentration $Z(x)$ over this thickness. Denoting by H_0 and A_0 the spatial means of $H(x)$ and $A(x)$ in the domain, and S the area of the domain, the products $T = \rho S H_0$ and $Q = \rho S A_0$ represent the tonnage of contaminated ground and the quantity of pollutant it contains (ρ is mass density). The mean contaminant concentration is then $Q/T = A_0/H_0$.

Finding kriging estimates H^* of H_0 and A^* of A_0 does not raise any particular problem. Estimation of mean areal concentration, however, is tricky because the $Z(x)$ are defined over different supports (thicknesses), so direct kriging is not possible. In practice, one just takes the ratio $\hat{m} = A^*/H^*$. If the variograms of H and A are proportional, one has

$$\hat{m} = \frac{A^*}{H^*} = \frac{\sum \lambda_\alpha A_\alpha}{\sum \lambda_\beta H_\beta} = \frac{\sum \lambda_\alpha H_\alpha Z_\alpha}{\sum \lambda_\beta H_\beta} = \sum \nu_\alpha Z_\alpha \quad \text{with} \quad \nu_\alpha = \frac{\lambda_\alpha H_\alpha}{\sum \lambda_\beta H_\beta}$$

The ν_α have a unit sum, and \hat{m} is therefore a weighted average. This is not the case if A_0 and H_0 are estimated with nonproportional variograms, which may create artifacts when thicknesses vary a lot. If so, it may be better to consider nonoptimal estimates of H_0 and A_0 but ensuring that \hat{m} is a weighted average. Such practice concerns positive variables and should not be used, for example, in the case of a variable with zero mean. The estimation variance of \hat{m} can be calculated exactly or approximately under certain conditions (e.g., see Journel et Huijbregts, 1978, sec. V.C.3).

3.5.6. Filling-in the Gaps

In applications such as mining or forestry where data are collected on a regular grid, considerable saving is achieved by always using the same pattern of data, since the kriging weights can then be computed once for all (i.e., if weights are shift invariant). But often the data grid has gaps, which ruins this plan. The Three Perpendiculars then come to rescue: if the gaps are sparse, one can replace the missing data by their kriging estimates and proceed as if all data were present—kriging variances, however, are underestimated.

The Three Perpendiculars

If

$$Z_{N+M}^* = \sum_{\alpha=1}^{N+M} \lambda_\alpha Z_\alpha$$

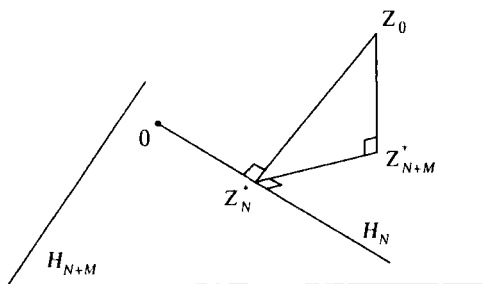


FIGURE 3.12. The three perpendiculars.

is the kriging estimator based on $N + M$ data of some quantity Z_0 , then the kriging estimator of Z_0 based on the first N data points is

$$Z_N^* = \sum_{\alpha=1}^N \lambda_{\alpha} Z_{\alpha} + \sum_{\alpha=N+1}^{N+M} \lambda_{\alpha} Z_{\alpha}^*$$

where Z_{α}^* is the kriging estimator of Z_{α} from the N data. Kriging variances are related by

$$\text{Var}(Z_N^* - Z_0) = \text{Var}(Z_{N+M}^* - Z_0) + \text{Var}(Z_N^* - Z_{N+M}^*)$$

In the case of a known mean there is a simple geometric proof illustrated in Figure 3.12: the projection of Z_0 onto the Hilbert space $H_N \subset H_{N+M}$ can be accomplished by first projecting onto H_{N+M} and then projecting this projection onto H_N . A similar proof can be given in the case of an unknown mean using the orthogonality relationship (3.24).

In our case the $N + M$ data points represent the complete kriging configuration and M the number of gaps. For the property to apply the missing data should be reconstructed only from the N data present in the configuration. Intuitively though, the final results should be even better if the gaps are filled using the best possible neighborhoods.

3.6. SELECTION OF A KRIGING NEIGHBORHOOD

The selection of data points to be included in the estimation is a key problem in the application of kriging. In theory, the minimum mean square error is achieved when all points are included, since any smaller neighborhood can be viewed as a constrained optimization with weights zero placed on the discarded points. But a global neighborhood may result in a kriging matrix that is too large to be inverted numerically. Typically standard algorithms work well for up to 100 points, and special purpose algorithms suited for band matrices can accommodate up to 400 points (Davis and Grivet, 1984). Another important consideration is the geostatistical model itself which may only be valid over short distances.

The solution is to restrict the point selection to a subset of the data, changing with the estimated point, and thus called a *moving neighborhood*. In doing so, however, we render the results dependent on the particular data selection algorithm, and tend to create spurious discontinuities in regions where control points are scarce due to the sudden change of sample points from one neighborhood to the next. This difficulty is of course not specific to kriging but is a feature of all neighborhood methods.

What is the optimum design of a moving neighborhood? This question turns out to be rather complex. Short of a rigorous theory we can only give some guidelines.

3.6.1. Screening Effect and Relay Effect

Screening Effect

In a mathematical sense the “screening effect” describes a situation in which nonzero kriging weights are concentrated on a subset of samples in the immediate vicinity of the estimated point or block. These samples screen off the influence of all other data. In 1D we have seen that with a linear variogram the ordinary kriging estimator only depends on the two adjacent samples, or on the nearest end point (Examples 4 and 5). For simple kriging the same circumstance occurs for the exponential covariance due to its associated Markov property: if $x_0 < x_1 < x_2$, then given Z_1 , Z_0 and Z_2 are uncorrelated.

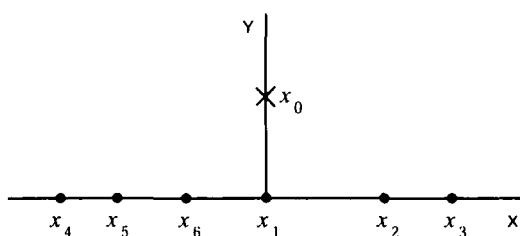
In higher dimensions there are some special covariance models for which any closed contour acts as a perfect screen between internal and external data. Let us consider for simplicity the isotropic case (a geometric anisotropy does not alter the conclusion). For SK these special models, expressed as functions of $r = |h|$, are in 2D the K -Bessel model (2.56) with $\nu = 1$, namely $C(r) = (r/a)K_1(r/a)$ (Whittle, 1954), and in 3D the covariance measure $C(r) = \exp(-r/a)/(r/a)$ (e.g., Arfken, 1985) ($a > 0$ in both cases). For OK the solutions are variogram measures: in 2D the de Wijs model $\gamma(r) = \log r$, and in n -D $\gamma(r) = 1/r^{n-2}$ (Matheron, 1965, p. 252).

In 2D a particular screening effect occurs for simple kriging when the covariance can be factorized along the components h_x , h_y of the separation vector h as

$$C(h_x, h_y) = C_1(h_x) \times C_2(h_y)$$

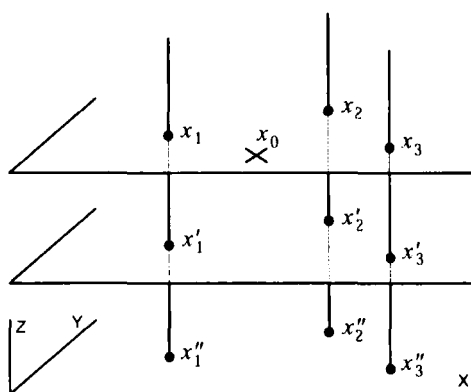
Consider points x_1, x_2, x_3, \dots , aligned along a parallel to one coordinate axis and x_0 on the perpendicular to that line through x_1 (Fig. 3.13a). Then the SK estimator of Z_0 from Z_1, Z_2, Z_3, \dots , assigns a zero weight to every point other than Z_1 . Indeed, if $\lambda_i = 0$, $i > 1$, the SK equations reduce to

$$\lambda_1 \rho_{1j} = \rho_{0j} = \rho_{01} \rho_{1j} \quad \forall j$$



$$C(h_x, h_y) = C_1(h_x) \times C_2(h_y)$$

(a)



$$C(h_x, h_y, h_z) = C_1(h_x, h_y) \times C_2(h_z)$$

(b)

FIGURE 3.13. Configurations producing a screening effect for simple kriging under a factorized covariance model: (a) points other than x_1 receive a zero weight; (b) points other than x_1, x_2, x_3 receive a zero weight.

and are satisfied when $\lambda_1 = \rho_{01}$. Similar results hold in 3D. For example, if data lie in horizontal planes (Fig. 3.13b) and if the covariance factorizes as

$$C(h_x, h_y, h_z) = C_1(h_x, h_y) \times C_2(h_z)$$

then the plane at the same elevation as the estimated point screens off the other data planes.

An example of factorized model is the Gaussian covariance $C(h) = \exp(-|h|^2/a^2)$. Another model, widely used for digital image compression, is the factorized exponential covariance

$$C(k, l) = \sigma^2 \rho_v^{|k|} \rho_h^{|l|}$$

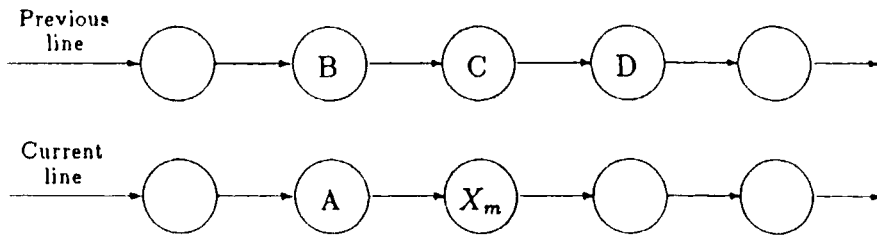


FIGURE 3.14. Differential Pulse Code Modulation predictor configuration. From Rabbani and Jones (1991).

where ρ_v and ρ_h , respectively, denote the vertical and horizontal correlation coefficients between the pixels of an image, and k and l are vertical and horizontal separations (Rabbani and Jones, 1991). Techniques of linear predictive coding exploit the high correlation between adjacent pixels of an image (ρ_v and ρ_h typically exceed 0.9) to reduce the quantity of bits transmitted over communication lines. Figure 3.14 illustrates a configuration used in forming the optimum linear predictor (i.e., the SK estimator) of a pixel X_m based on four previous pixels A through D. For the above factorized model the weights are

$$\lambda_A = \rho_h \quad \lambda_B = -\rho_v \rho_h \quad \lambda_C = \rho_v \quad \lambda_D = 0$$

Pixel D receives a weight zero: its influence is completely screened off by the other pixels.⁹

Aside from special mathematical cases, the screening effect represents a physical approximation: in the presence of strong spatial autocorrelation, the data of the first “ring” matter most, and very little additional information is gained by considering sample points beyond that. To take a specific example, consider the problem depicted in Figure 3.15: the estimation by ordinary kriging of a square panel from its 12 nearest samples in a square grid (from David, 1977, p. 258). The symmetry of the pattern reduces the problem to the determination of a single weight λ . In the case of a linear variogram, the first ring comprising the four vertices of the square receives a total weight of about 95%, while the eight samples of the second ring share only 5%. The kriging neighborhood can be limited to the first ring. However, when a nugget effect C_0 is added, the weight $\lambda/8$ on each outer sample increases up to the limiting value of $1/12 = 1/N$. The nugget effect destroys autocorrelation and in the limit makes all samples equivalent. Hence the celebrated geostatistical saying: *the nugget effect lifts the screening effect* (Matheron, 1968a).

In practice, the screening effect should not be gauged by the weights but rather by the variance. In a study of the behavior of kriging weights, Rivoirard (1984) showed that when the neighborhood is enlarged, negative weights ap-

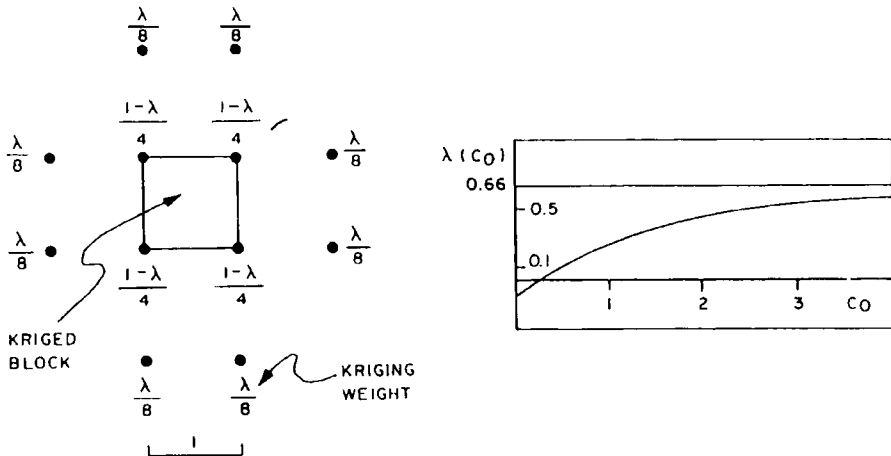


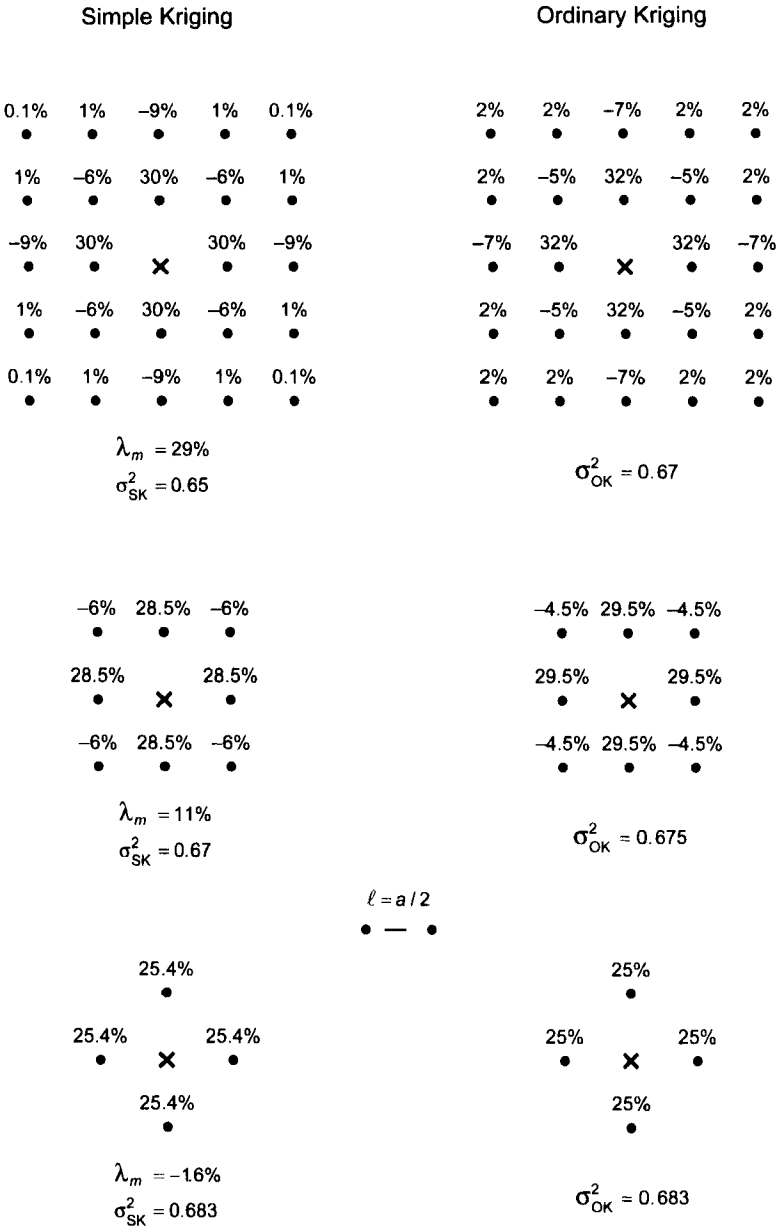
FIGURE 3.15. Variation of kriging weights with C_0 (with a linear variogram). All weights go to $1/12$. From David (1977).

pear, which are not small even for the most common variogram models, but do not improve precision significantly. Figure 3.16 displays the SK and OK weights for the spherical model with range twice the grid spacing and the weight on the mean $\lambda_m = 1 - \sum \lambda_{K\alpha}$ (see Section 3.4.7). Observe that all variances are comparable: very little is gained by extending the kriging neighborhood beyond the four nearest neighbors. Haslett (1989) proposes a stepwise algorithm for optimizing subset selection.

Relay Effect

It is often believed that when the variogram has a finite range, it is not necessary to include sample points beyond that range in the kriging neighborhood. This assertion is certainly false when the mean is unknown, since the mean is not local information, but it is also false in the case of simple kriging, as noted earlier. The reason is that points beyond the range may exert an influence through their correlation with points within the range by virtue of the so-called *relay effect*. In Figure 3.16, for example, all 16 points of the second ring have zero correlation with the estimated point, but the first ring acts as a relay. Similarly in Figure 3.14 pixel *B* would receive a weight zero in the absence of pixel *A*. Notice that relays tend to produce small or negative weights.

In a different setting, the autoregressive and moving average models (ARMA) of Box and Jenkins (1976) provide another angle at this. Autoregressive models, which by design depend on a finite number of past values, have correlation functions that extend to infinity, whereas the prediction of moving average models, which have a finite range, involves an infinite number of past values.



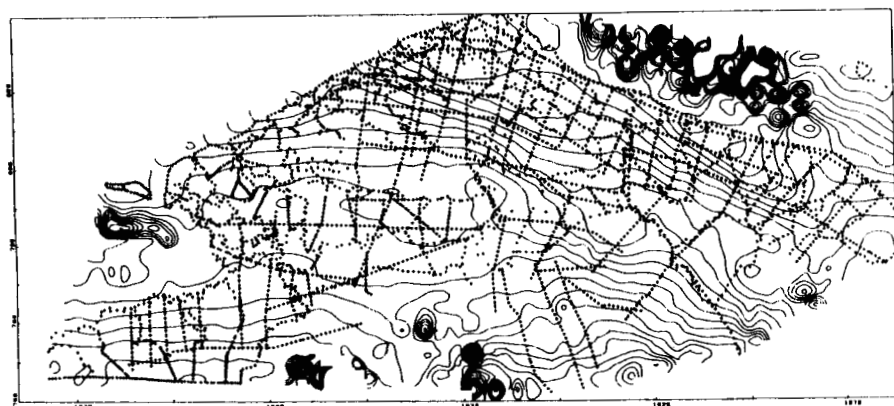
3.6.2. Practical Implementations of Neighborhood Selection

The additivity relationship (3.22) portrayed universal kriging as a two-stage procedure: one is the optimum estimation of the drift and the other is simple kriging of the residuals. Unfortunately, these two stages raise conflicting demands on the neighborhood: if simple kriging can benefit from the screening effect, drift estimation, on the contrary, as we know from least squares theory, is best accomplished with sample points as far apart and well scattered as possible. The obvious solution of estimating the drift (optimally) with a global neighborhood and subtracting it from the data is often impractical for the same reasons that preclude the use of a global neighborhood in the first place: too many data points and/or, more important, the poor validity of a simple drift model over the whole domain.

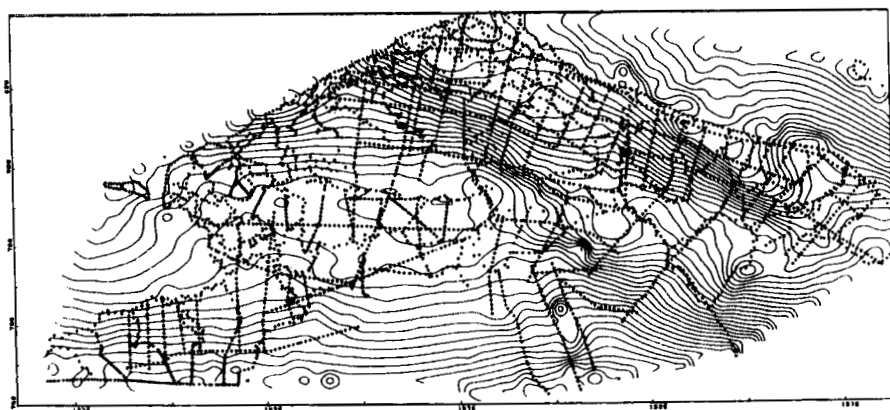
Sophisticated neighborhood selection algorithms have been devised to reach a compromise between near and far sample points. They usually include all points of the first ring and then more distant points, following a strategy that attempts to sample all directions as uniformly as possible while keeping the number of points as low as possible (octant search). Obviously good azimuthal coverage is important to avoid the risk of bias. Typically from 8 to 16 points are retained, from at least five octants or four noncontiguous octants; the search is accelerated by initially classing the data points in the cells of a coarse grid. For contour mapping purposes, where continuity is important, larger neighborhoods may be considered to provide more overlap. Another approach is to divide the gridded domain into overlapping sections, estimate each of them with a global neighborhood, and splice the resulting subgrids.

A useful technique to improve continuity without spoiling the accuracy of the estimates is selective low-pass filtering. The idea is to run a smoothing filter on the kriged grid but modify the estimates only within a confidence interval defined by the kriging standard deviation so as to modulate the degree of smoothing by the uncertainty about the estimates. The following low-pass filter has good spectral properties: a moving average with weights $(1/4, 1/2, 1/4)$ along the rows and then along the columns of the grid, followed by a moving average with weights $(-1/4, 3/2, -1/4)$ to restore some power in the high frequencies.

Ad hoc procedures have been developed to handle “difficult” data sets. For example, seismic data are densely sampled along seismic lines and sparsely across. A standard neighborhood search algorithm will tend to grab its data points from the nearest seismic line thereby producing quasi-singular kriging matrices and spurious “patterning” in extrapolated areas. Figure 3.17 from Renard (1983) shows the kriged map obtained from a seismic data set of 3753 points using a standard neighborhood search of 20 points, and the results obtained with an improved procedure based on blending global and moving neighborhood estimates (“pseudounique neighborhoods”; Renard and Yancey, 1984).



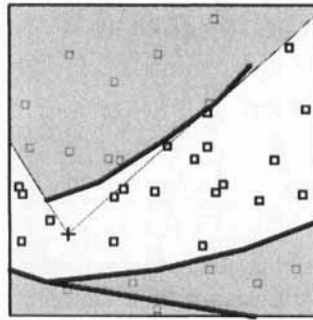
(a)



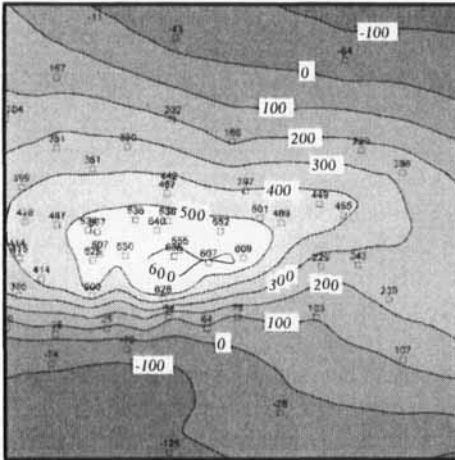
(b)

FIGURE 3.17. A difficult test case for gridding packages: (a) standard neighborhood search; (b) enhanced ("pseudounique") neighborhood search. From Renard (1983).

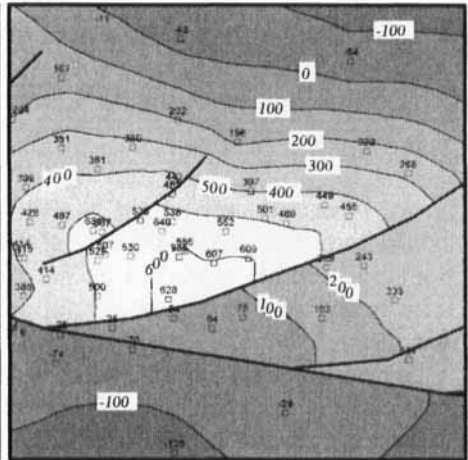
Another ad hoc technique is employed to deal with discontinuities such as geological faults. It is illustrated in Figure 3.18: the fault is considered as a screen and only the sample points directly "seen" from the estimated point are included in the kriging neighborhood. This technique has the merit of existence—though it is limited by the availability of sample points in the "fault blocks"—but it is very crude since it ignores the geological information associated with a fault. Maréchal (1984) proposes a better approach to deal with subvertical faults: estimate the vertical displacement (throw function), remove it from the data, interpolate the unfaulted surface which is expected to be smooth, and reintroduce the fault in the end (see also the external drift method in Section 5.7.3).¹⁰



(a)



(b)



(c)

FIGURE 3.18. Kriging with faults: (a) the fault as a screen: the data points located in the shaded area cannot be seen from the target point marked by a cross; (b) result of kriging without fault; (c) result of kriging with faults.

3.6.3. Global Neighborhood Cross-Validation

A special form of kriging neighborhood is involved in model cross-validation (Section 2.6.3): each sample point x_α is estimated from all others, excluding x_α itself. Essentially the same global neighborhood is considered, but not quite. In a configuration with N data points and $L + 1$ drift functions, this involves, in principle, the resolution of N linear systems of size $(N + L) \times (N + L)$, which may be prohibitive. Dubrule (1983) shows that all solutions can in fact be obtained by inversion of a single matrix, the matrix \mathbf{A} of system (3.14). If $[b_{\alpha\beta}]$ denotes the inverse of \mathbf{A} and if one considers the kriging estimator of $Z(x_\alpha)$ from $\{Z(x_\beta) : \beta \neq \alpha\}$, the kriging weights λ_β (depending of course on

x_α) and the kriging variance $\sigma_{K\alpha}^2$ are given by

$$\lambda_\beta = -\frac{b_{\alpha\beta}}{b_{\alpha\alpha}} \quad (\beta \neq \alpha)$$

$$\sigma_{K\alpha}^2 = \frac{1}{b_{\alpha\alpha}}$$

The second result assumes that the kriging matrix is written in covariance terms (if it is written in variogram terms, the sign must be changed on the right-hand side of the second formula). A generalization of the above results allows an estimation of $Z(x_\alpha)$ excluding more than a single observation.

3.7. MEASUREMENT ERRORS AND OUTLIERS

3.7.1. Filtering Nonsystematic Errors

Three types of error can affect the data: uncertainty on the exact positions of the measurements, systematic errors, and nonsystematic, also called *random*, errors. Positioning uncertainty is typically associated with marine surveys and can be modeled statistically (cf. Section 2.4.4). The advent of satellite positioning (GPS) has made this uncertainty very small¹¹ but perhaps still significant in applications where extreme precision is required, such as marine 3D seismic.

Systematic errors are the most dangerous. They usually go unnoticed, do not cancel out and can ruin a whole analysis. They may have several origins: a drift of the instrument (as in gravimetric surveys), acquisition problems (tool malfunction, insufficient dynamic range, quantization error), model inadequacies with *computed data* (parameters, e.g., porosity or saturation are computed from wireline logs by inversion of a petrophysical model), and so on. Systematic errors, if suspected, can be corrected by multivariate methods and are discussed with them. We will focus here on random errors.

The simplest case is that of uncorrelated errors with the same amplitude. They turn up in the variogram as an additional nugget effect equal to the error variance. But errors may also be unequal (e.g., data from different sources) or even correlated. For example, in a bathymetric survey data are acquired along profiles, and it can be assumed that errors are the same along a profile and independent across two different profiles. Further the standard deviation of the error is proportional to the local average depth (Chiès, 1977). If Z_ε denotes the measurements and Z the underlying RF, a comprehensive model is

$$Z_\varepsilon(x_\alpha) = Z(x_\alpha) + \varepsilon_\alpha$$

where ε_α are random errors defined only at the sample points, and subject to the following assumptions:

- Errors are nonsystematic,

$$E[\varepsilon_\alpha] = 0 \quad \alpha = 1, \dots, N$$

- Errors are uncorrelated with the studied RF

$$E[\varepsilon_\alpha Z(x)] = 0 \quad \forall x, \quad \alpha = 1, \dots, N$$

- Errors may be correlated among themselves

$$E[\varepsilon_\alpha \varepsilon_\beta] = S_{\alpha\beta} \quad \alpha, \beta = 1, \dots, N$$

We want to estimate the error-free value $Z_0 = Z(x_0)$ from observations $Z_\varepsilon(x_\alpha)$ “corrupted” by noise. In time series this is a standard problem known as *filtering* (the signal from the noise), and its solution in the frequency domain relies on the possibility of separating the spectral characteristics of the signal and of the noise. In the geostatistical terminology we regard this problem as a particular case of *cokriging*—estimating values of one variable on the basis of another. But the correlation structure is so simple that the result is only a slight modification of the standard kriging system. Our estimator is now

$$Z^* = \sum_{\alpha} \lambda_{\alpha} (Z_{\alpha} + \varepsilon_{\alpha})$$

Since errors have zero mean the unbiasedness conditions do not change. The m.s.e. becomes

$$E(Z^* - Z_0)^2 = E \left(\sum_{\alpha} \lambda_{\alpha} Z_{\alpha} - Z_0 \right)^2 + \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} S_{\alpha\beta}$$

Minimizing the complete expression leads to the cokriging system

$$\begin{cases} \sum_{\beta} \lambda_{\beta} (\sigma_{\alpha\beta} + S_{\alpha\beta}) + \sum_l \mu_l f_{\alpha}^l = \sigma_{\alpha 0} & \alpha = 1, \dots, N \\ \sum_{\alpha} \lambda_{\alpha} f_{\alpha}^l = f_0^l & l = 0, \dots, L \end{cases} \quad (3.36)$$

and the cokriging variance

$$\sigma_{CK}^2 = E(Z^* - Z_0)^2 = \sigma_{00} - \sum_{\alpha} \lambda_{\alpha} \sigma_{\alpha 0} - \sum_l \mu_l f_0^l$$

Equations in terms of the variogram remain the same as (3.16) but with $\gamma_{\alpha\beta} - S_{\alpha\beta}$ replacing $\gamma_{\alpha\beta}$ (so the first N diagonal terms are no longer zeros but $-S_{\alpha\alpha}$).

The cokriging system (3.36) differs from the UK system (3.14) by the presence of the error covariance terms $S_{\alpha\beta}$. It relies on the assumption that

the covariance $\sigma_{\alpha\beta}$ and $S_{\alpha\beta}$ are known separately. The cokriging estimator is no longer an exact interpolator: it does its job of filtering measurement errors.

In the particular case of uncorrelated errors, the system (3.36) is identical to the standard kriging system with the covariance $\sigma + S$ except at data points (where the kriging system has $\sigma_{\alpha 0} + S_{\alpha 0}$ on the right-hand side). This shows why the kriging estimator also filters errors. It can also be seen directly by noting that at any point x_0 other than a sample point the kriging mean square error

$$E[Z^* - Z_\varepsilon(x_0)]^2 = E(Z^* - Z_0)^2 + E(\varepsilon_0^2)$$

differs from the cokriging error only by the addition of a constant term, and thus has the same minimizer. At a sample point this relationship breaks down, and both the kriging estimate and its variance have a discontinuity (see Fig. 3.3c). For contour mapping purposes it is preferable to grid the continuous component Z and filter measurement errors and the rest of the nugget effect. As for which variance should be reported, it makes sense to exclude genuine measurement error variances (since we are not interested in reconstructing them) but to include the nugget effect variance due to micro-structures.

Figure 3.19 illustrates the effect of filtering error variances in mapping CO_2 concentration in soil. Straightforward estimation with a continuous variogram model produces the typical “fat in the soup” effect in which contour lines circle around data points because, due to errors, most of them are local extrema (Fig. 3.19a). We know there are indeed uncertainties in the data because gas emanation is subject to very local variations of soil properties and to large measurement errors. By contrast, the filtered map shows a physically more meaningful picture of the CO_2 concentration distribution (Fig. 3.19b).

3.7.2. Robust Kriging

Kriging estimators being linear in the data are sensitive to the presence of a few unusually large (or sometimes small) values called “outliers.” Since the pioneering work of Tukey (1960) and Huber (1964) an abundance of methods have been developed that are *resistant* against large changes in a few of the values, and this has become a field of its own called Robust Statistics (e.g., see Huber, 1981). We already had to deal with the robustness preoccupation in Section 2.2.5 while estimating the variogram, and now we have to face it again with kriging.

This turns out to be a very tricky issue, if we set aside the easy case of gross acquisition errors. The first step in a geostatistical study is to make sure that the data under consideration are real. When this verification is done, we are left with good data that can still be outliers. The problem typically arises with highly skewed distributions such as ore grades in gold deposits, where a small number of samples can be responsible for a large proportion of the metal

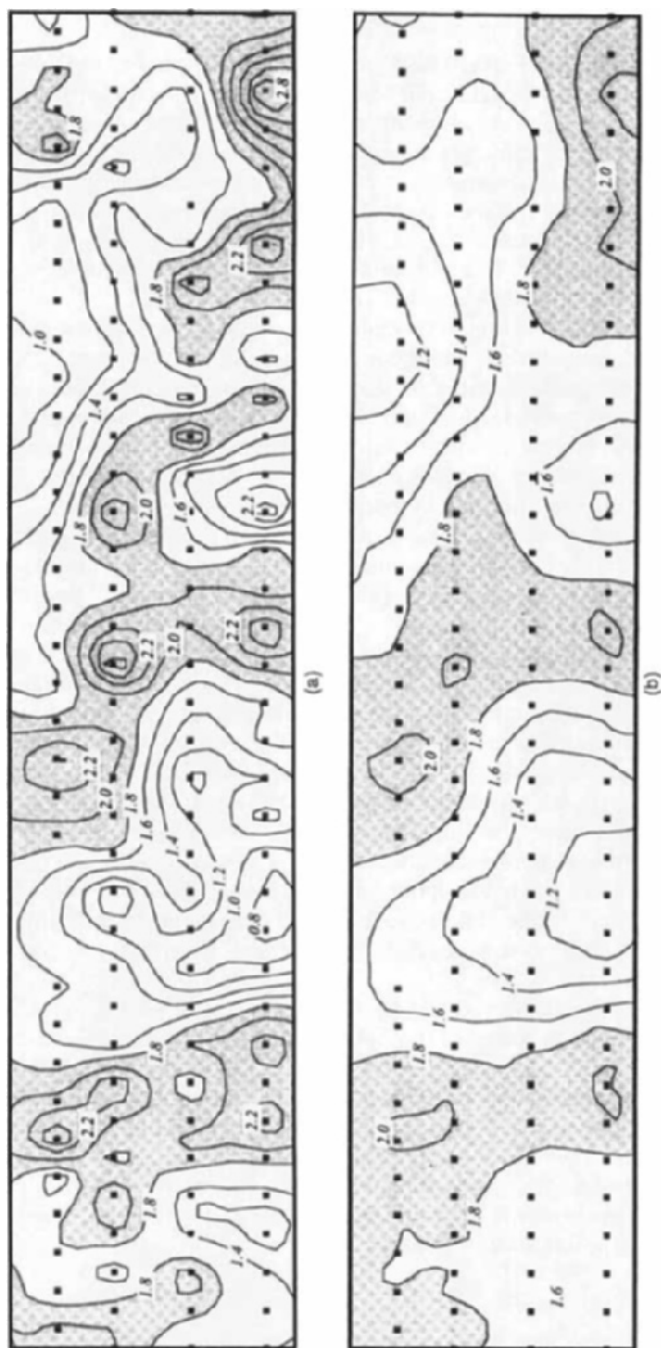


FIGURE 3.19. Filtering measurement errors (CO₂ concentration in soil): (a) contour map using a continuous variogram model; (b) contour map with noise filtered.

content in the deposit. Parker (1991) reports a case where 4 out of 34 samples are responsible for 70% of the gold content, and 7 out of 34 for 90% of the gold content. Clearly this is a very uncomfortable situation because, as Parker notes, "even small changes in the grade near these assays or the proportion of high-grade material may be responsible for large differences between estimated and recovered reserves. If on the downside, these differences could have adverse consequences." On the other hand, discarding the high grades completely is economically nonsensical. The fortune of companies may be due just to the occurrence of inordinately large values.

There is no single optimum procedure to handle outliers but only *engineering solutions*, namely reasonable compromises. Parker for example suggests to break the ore grade distribution into an ordinary-grade, well-behaved, population, and a strongly skewed upper tail. Block estimates are computed as a mixture of the kriging estimate from the ordinary-grade population and the mean high-grade value, derived by fitting a parametric model to the upper tail (a single- or double-truncated lognormal distribution).

In a milder case of skewness, it may suffice to clip the high grades. Consider, for example, the following situation encountered in a real gold deposit: in a data set of 62 average core grades, regularized over 10 m, 61 grades range from 0 to 15 gram/ton, with a mean of 3.66 g/t, and then one is equal to 25 g/t. This sample is responsible for an increase of about 10% of the reserves. Should we include it in the estimation? If we do, by consistency, we must also include it in the calculation of the variogram, thereby increasing the nugget effect. The net result is higher reserves but with a higher uncertainty. Instead of reporting a $\pm 2\sigma$ confidence interval for the mean grade of (3.66 ± 0.24) g/t, we would report (4.00 ± 0.72) g/t. The greater uncertainty is not good for mine planning, and the lower bound for the reserves is now less than before. A reasonable compromise is to reduce the value of the high grade to the more moderate upper limit of 15 g/t and report a mean grade of (3.84 ± 0.42) g/t. In the end pleasant surprises are naturally welcome, but for planning one prefers estimates that do not depend too much on a few extreme values.

Now, if we had observed ten 25 g/t samples out of 620 samples, say, and if these values were scattered throughout the mine, it would be reasonable to assume that the ore body comprises about 1.6% of high grades. For global estimation we might just include them without editing. Local block estimates, however, would still be completely distorted by the high grades.

What is reasonable for mine planning might be totally foolish in another context. For example, it might be that we *never* want to discard an extreme observation because it is indicative of an extreme danger (e.g., a spike on the seafloor, or a lethal dose of pollutant).¹²

Winsorizing

A different approach emphasizing local consistency has been proposed by Hawkins and Cressie (1984) and is discussed in Cressie (1991, pp. 144–150). Its main steps are as follows:

1. Compute a robust variogram, and fit a model to it.
2. Use this robust variogram to compute the kriging weights for the estimation of each sample point from all other data $Z_{(\alpha)}^* = \sum_{j \neq \alpha} \lambda_{\alpha,j} Z_{\alpha,j}$, and the kriging standard deviation $\sigma_{K\alpha}$.
3. As a robust estimate of Z_{α} consider the weighted median $\tilde{Z}_{(\alpha)}$, defined with the above kriging weights as the midpoint solution of $\sum_{j \neq \alpha} \lambda_{\alpha,j} \text{sign}(Z_{\alpha,j} - \tilde{Z}_{(\alpha)}) = 0$.
4. Edit each sample Z_{α} , and replace it by the Winsorized version

$$Z_{\alpha}^{(e)} = \begin{cases} \tilde{Z}_{(\alpha)} + c\sigma_{K\alpha} & \text{if } Z_{\alpha} - \tilde{Z}_{(\alpha)} > c\sigma_{K\alpha} \\ Z_{\alpha} & \text{if } |Z_{\alpha} - \tilde{Z}_{(\alpha)}| \leq c\sigma_{K\alpha} \\ \tilde{Z}_{(\alpha)} - c\sigma_{K\alpha} & \text{if } Z_{\alpha} - \tilde{Z}_{(\alpha)} < -c\sigma_{K\alpha} \end{cases} \quad (\text{usually } 1.5 \leq c \leq 2.5)$$

5. Now proceed with kriging, still with the robust variogram, but instead of the original data use the *edited data* $Z_{\alpha}^{(e)}$.

The crux of the method, it seems, is the weighted median which, as the authors suggest, could be used as a robust estimator by itself. If all weights are positive, it can be computed as follows: sort the data by ascending order and interpreting the weights as frequencies cumulate them until reaching $\frac{1}{2}$, then select the corresponding Z . The weighted median minimizes the quantity

$$\sum_{j \neq \alpha} \lambda_{\alpha,j} |Z_{\alpha,j} - \tilde{Z}_{(\alpha)}|$$

and generalizes in an understandable manner the usual median for which all weights are equal to one. But unlike a kriging variance, this criterion has no obvious significance, especially with the possibility of negative weights.

Still this technique could provide an interesting tool for robust cross-validation. Indeed the subtlety is to detect outliers on the basis of their local distribution (we avoid the term “conditional distribution”), rather than just their global distribution. It is important to make sure that this reference distribution is determined in a robust manner; otherwise, good data in the vicinity of an outlier could themselves be classified as outliers.

Another method proposed by Dowd (1984) detects outliers as in steps 1–3 (using $2.5 \leq c \leq 3$), but instead of editing the data, attaches an error variance to them and proceeds with the cokriging system (3.36). Treating outliers as unreliable data would not be appropriate when the validity of the samples has been demonstrated. Rather, the uncertainty lies with the spatial extent of the outliers.

3.8. CASE STUDY: THE CHANNEL TUNNEL

The Channel tunnel project may well be the ideal geostatistical case study. It is an important application, it is simple to understand, and—a rare event—it is possible to compare the geostatistical predictions with reality. Of course, as one would expect for a case study, the approach worked well. Boring of the tunnel was completed one month ahead of the initial schedule despite the fact that most of the tunnel boring machines only came into service several months late. This is in contrast with the long delays usually encountered in

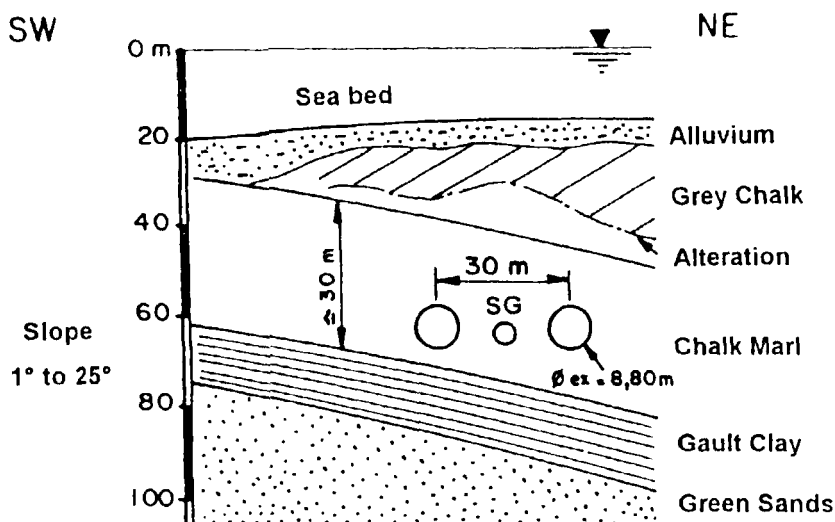


FIGURE 3.20. Channel Tunnel: typical geological cross-section showing the three tunnels. Reprinted from Blanchin et al. (1989), with kind permission from Kluwer Academic Publishers; Blanchin and Chilès (1992, Fig. 1), with kind permission from Springer-Verlag.

tunnel boring around the world. It was the result of the excellent performance of men and machines, but also of a careful assessment of the geological risk through geostatistical analysis.

The first results of the study were published by Blanchin et al. (1989), and an overview of the Channel tunnel project with a historical perspective is given by Blanchin and Chilès (1993a). We will borrow heavily from these two papers with an emphasis on methodological aspects.

3.8.1. Objectives and Methodology

The geological setting of the tunnel shown in Figure 3.20 can be summarized as follows: a favorable layer, the Cenomanian Chalk Marl, made of soft, generally impermeable and homogeneous rock, overlain by the Grey Chalk, a highly porous layer of typically fractured and altered rocks, and underlain by the Gault Clay, which cannot be penetrated without serious civil engineering problems. The Chalk Marl and the Gault Clay are in fact separated by a thin regular layer of Tourtia Chalk, but here it is lumped with the Chalk Marl for simplicity. Notice that what is called the tunnel comprises really three parallel tunnels 15 m apart, two for transport and a smaller one for servicing.

Although optimization of the tunnel alignment had to take into consideration the slope and curvature constraints imposed by a high-speed railway, the primary constraints were geotechnical and geological:

- No geophysical borehole could be intersected for fear of water inflow.

- The tunnel had to be at least 20 m deep below the seafloor to preserve the mechanical strength of the overlying formations.
- The tunnel could not be bored much deeper than 100 m below sea level because of the characteristics of the tunnel boring machines.
- The faults had to be intersected as orthogonally as possible.
- Most important, the tunnel had to be bored within the Chalk Marl formation.

The objective of the geostatistical study was to provide an accurate determination of the geometry of the Chalk Marl, which is only 30 m thick, and dipping, in order to prevent the risk of tunneling into the Gault Clay. It focused on the most critical variable, the top of the Gault Clay. A first estimation was made by kriging on the basis of the data available before the construction of the tunnel, with a careful evaluation of uncertainties. The results led the engineers to revise the initial layout. It was also realized that a better precision was needed in certain sections of the tunnel, and a complementary survey was designed by geostatistical analysis. Finally data acquired during drilling of the service tunnel allowed a comparison with the geostatistical predictions.

3.8.2. Contour Mapping

In this phase the goal is to calculate a reliable digital model of the top of the Gault Clay and to produce meaningful contour maps and cross sections. The basis is 1500 km of bathymetric and reflection seismic surveys, recorded continuously (every 3 m). The data include (1) 5 longitudinal seismic profiles running parallel to the tunnel, 25 m apart on the French side and 250 m apart on the British side; (2) 83 transverse seismic profiles, at variable intervals between 250 m and 1000 m; (3) 10 boreholes drilled in 1986 plus 90 old boreholes.

The depth $G(x)$ from sea level to the top of the Gault Clay is given by

$$G(x) = S(x) + \frac{V(x)T(x)}{2}$$

where $S(x)$ is the depth to the seafloor, computed from bathymetric data, $T(x)$ represents seismic two-way time from the seafloor to the top of the Gault Clay, computed from seismic profile data, and $V(x)$ is the average velocity obtained by various geophysical methods, such as sonic logs in some of the wells. The variables $S(x)$, $T(x)$, and $V(x)$ are estimated independently and combined through this equation to produce the final result $G(x)$.

Kriging variances are computed for each interpolated variable and as the errors are independent the kriging variance attached to G is given by

$$\sigma_G^2 = \sigma_S^2 + \frac{\sigma_V^2 T^2 + V^2 \sigma_T^2 + \sigma_V^2 \sigma_T^2}{4}$$

where σ_S^2 , σ_V^2 , and σ_T^2 are the kriging variances associated with S , V , and T and all values depend on the location x .

Given the importance of risk assessment in this application and the geological heterogeneity from one part of the Channel to the other, a global structural analysis would be meaningless because not only the variogram parameters could change but also the variogram shape. The approach taken by Blanchin et al. (1989) is to divide the area into successive 1000-m long units (37 units) and in each one compute the histogram and statistical parameters, and the raw and residual variograms in the two profile directions (that coincide with the main directions of the anticline). To avoid sampling bias, some data in over-represented areas are discarded. The outcome of this preliminary study is the definition of 16 homogeneous zones 1 to 5 km long, obtained by merging similar successive units.

To complete the analysis, the various sources of measurement errors are identified (e.g., tide correction, migration of seismic reflectors, velocity calculations) and included in the final structural model. The latter is validated in each zone and also globally, using standard cross-validation techniques.

The area of study is defined as a 1-km wide and 40-km long strip centered along the main axis of the first planned alignment. The variables are interpolated by kriging to the nodes of a 40-m \times 20-m rectangular grid (20 m in the transverse direction), using for each variable a moving variogram model fitted zone by zone, with some smoothing between zones. Known faults are included as screens for the estimation of seismic times and velocities. Measurement errors that are locally constant along a profile are taken into account by adding a specific variance term to all covariances between two points on the profile.

Figures 3.21 and 3.22 show contour maps of depths and standard deviations for part of the French side of the tunnel layout. The general pattern of the top of the Gault Clay reflects the regional geological trend, an anticline whose axis is parallel to the alignment; the boundary of the Gault Clay outcrop is shown on the map as a dotted line. The standard deviation increases from south to north, reflecting higher uncertainties on velocity with increasing thickness from seafloor to top of the Gault. It shows well-marked minima in the vicinity of the profiles (good knowledge of seismic times) and the boreholes (consistent knowledge of seismic times and average velocities). Throughout the study area the standard deviation lies between 2 and 6 m, never exceeds 4 m along the underwater section of the tunnel route, and generally falls between 2 and 3 m.

3.8.3. Risk Assessment

We now have to answer the initial civil engineering question: Will the planned alignment intersect the Gault Clay? Or rather, given our incomplete knowledge, What is the *risk* that the planned alignment intersects the Gault Clay? This is where kriging standard deviations prove to be useful.

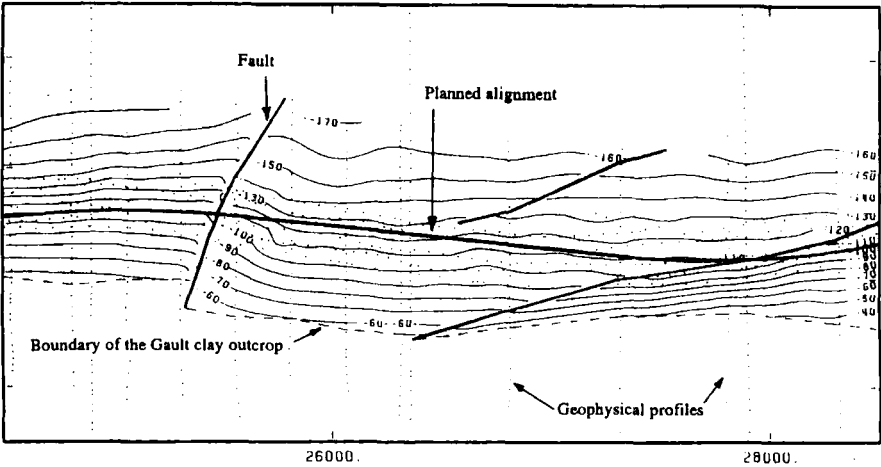


FIGURE 3.21. Channel Tunnel: contour map of the top of the Gault Clay on the French side, in meters. Reprinted from Blanchin and Chilès (1992, Fig. 2), with kind permission from Springer-Verlag; Blanchin and Chilès (1993a), with kind permission of the International Association for Mathematical Geology; Blanchin and Chilès (1993b), with kind permission from Kluwer Academic Publishers.

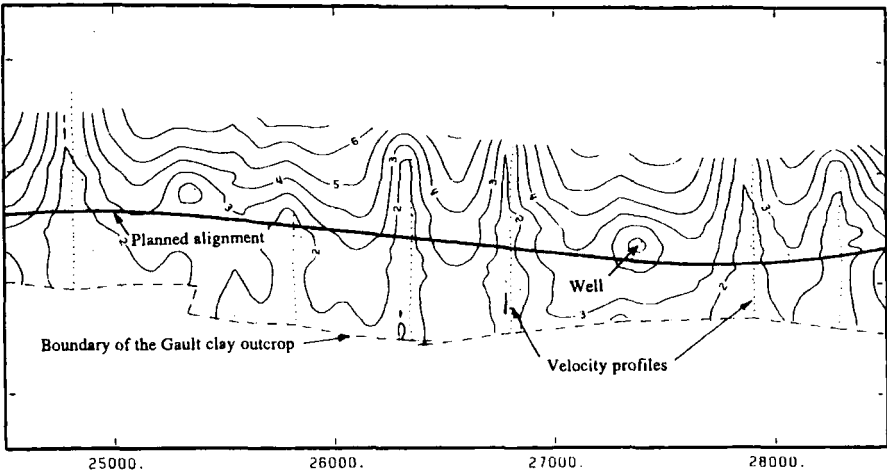


FIGURE 3.22. Channel Tunnel: kriging standard deviation for the top of the Gault Clay on the French side, in meters. Reprinted from Blanchin and Chilès (1992, Fig. 4), with kind permission from Springer-Verlag; Blanchin and Chilès (1993a), with kind permission of the International Association for Mathematical Geology; Blanchin and Chilès (1993b), with kind permission from Kluwer Academic Publishers.

Cross sections are the tool of choice to visualize the geometry of the tunnel project in the vertical plane. They can be generated from the grid or directly by kriging at points along the three tunnel galleries. Figure 3.23 shows the results obtained with a spacing of 20 m along a section of the south tunnel (vertical scale exaggeration: 20). The seafloor and the top of the Gault Clay are represented with their nominal 68% confidence intervals (± 1 standard deviation). For bathymetry the estimation is so precise that the three lines are indistinguishable on the graph (the kriging standard deviation does not exceed 0.5 m). When considering the estimated top minus one standard deviation, one can see that the first alignment could intersect the Gault Clay in several places.

This led the engineers to revise the layout so as to maintain the tunnels nearly everywhere at least one standard deviation above the estimated top of the Gault, as shown in the cross section. The risk of penetrating the Gault Clay was thus reduced, but of course not entirely eliminated. Notice that the engineers chose to use σ rather than the statistician's sacred 2σ because, in sections where it mattered, they were ready to assume a 16% risk of hitting the Gault Clay (one-sided interval).

3.8.4. Optimum Design of a Complementary Survey

The tunnel project also included crossover excavations at two locations to enable the trains to pass from one tunnel to the other if necessary. Their construction required more accurate geological predictions than those used for the main tunnel, which required a complementary geophysical survey. How should the survey be designed to achieve a standard error of less than 1 m on the top of the Gault Clay?

Since kriging variances can be computed without knowing the values of the variables, it suffices to simulate the surveying process by adding fictitious data until the required precision is achieved. The result is a recommendation to place the transverse seismic profiles 25 m apart over the French crossover and 100 m apart over the British crossover, and calibrate the seismic velocity by at least four boreholes at each crossover. With this new survey, and a new pass of variogram analysis and kriging, maps and cross sections are redrawn with improved precision.

3.8.5. Geostatistical Predictions versus Reality

As boring of the tunnel progressed, dual boreholes, one dipping to the north and the other to the south, were drilled downward from the central service tunnel to determine the actual depth and dip of the Gault Clay. Fifty-four dual boreholes were drilled along the first 13 km on the French side, and 31 were drilled along the first 51 km on the British side. The objective was to check the accuracy of the current estimates and, if needed, revise them to get a reliable geometric model for the continuation of the project.

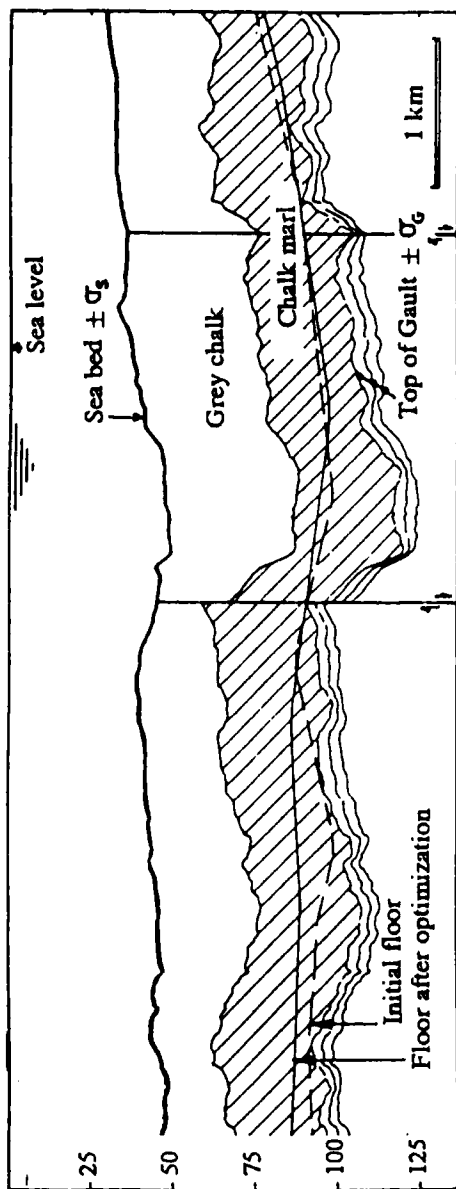


FIGURE 3.23. Channel Tunnel: cross-section of kriged results along the profile of the south tunnel. Vertical scale exaggeration: 20. Reprinted from Blanchin et al. (1989), with kind permission from Kluwer Academic Publishers; Blanchin and Chiles (1992, Fig. 1), with kind permission from Springer-Verlag.

TABLE 3.3. "Reality – prediction" differences in British and French sides

	French Side (13 km)	British Side (15 km)
Number of borehole pairs	54	31
Minimum difference	–5.00 m	–7.00 m
Maximum difference	+3.90 m	+8.00 m
Mean difference	+0.48 m	+1.70 m
Standard deviation from zero	2.02 m	3.40 m
Kriging standard deviation	2.85 m	2.55 m

Source: Blanchin and Chilès (1993a,b), with kind permission of the International Association for Mathematical Geology and Kluwer Academic Publishers.

The differences "reality – prediction" are plotted in Figure 3.24, and their statistical characteristics are summarized in Table 3.3, separately for the British side and the French side because their data patterns and spatial characteristics are rather different. Typical of real case studies, the results are not as straightforward as one would wish. On the British side the mean difference is 1.70 m and the standard deviation 3.4 m. The distribution of errors along the service gallery shows a good agreement between the actual and predicted depths for the first 20 points (dotted lines are $\pm\sigma_K$), but a systematic effect is apparent for the last ten boreholes (average difference of 5.6 m). This prompted a review of the initial interpretation in light of the new data. A key factor is that these ten discrepancies are clustered in a zone where the density of seismic profiles and borehole data is the lowest for the whole tunnel, that is, where the accuracy of the geostatistical mapping is low ($\sigma_K = 3.25$ m) compared with the average value for the British sector ($\sigma_K = 2.55$ m). A careful reinterpretation of the seismic data in that zone pinpointed two systematic errors: (1) an error in the time pick, the Tourtia Chalk was mistaken for the Gault Clay horizon lying in reality 3.5 m below; (2) errors in the calibration of the velocity data, due to poor positioning of old (1964–1965) geophysical boreholes in an area with a strong dip.

On the French side the positioning uncertainty on the geophysical boreholes was taken into account in the kriging process by a highly correlated error component in the velocity data. The observed discrepancies are small throughout, even apparently too small. But caution, the errors are far from independent! If they were, they would fluctuate back and forth around the zero line; on the contrary, they tend to stay on the same side of the line. Blanchin and Chilès (1993b) did a variogram analysis of the errors and concluded that on the French side the 54 observations are in fact equivalent to only 13 independent samples, while on the British side the 31 values are worth 21 independent samples.

In conclusion, the observations in the service tunnel were generally in good agreement with the geostatistical model and its predicted accuracy. When discrepancies occurred, they could be traced to systematic interpretation errors

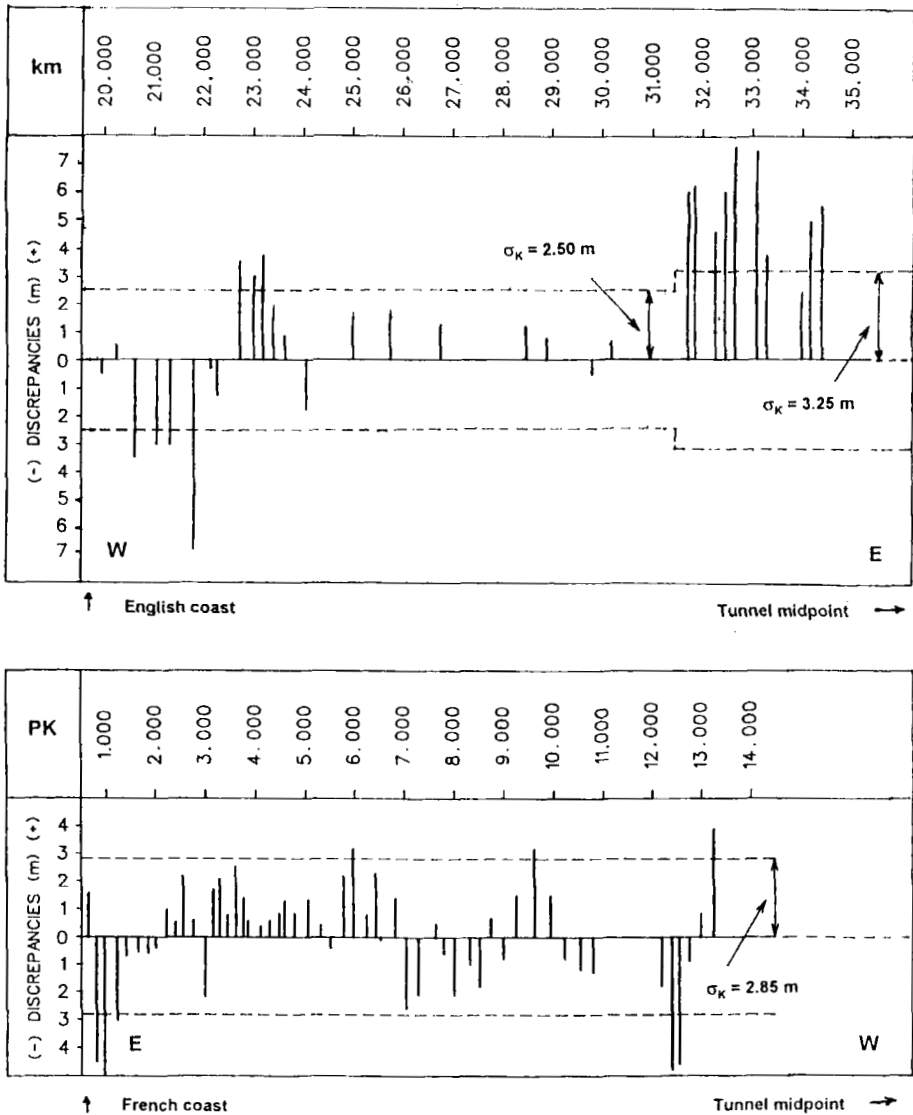


FIGURE 3.24. Channel Tunnel: discrepancies observed in the service tunnel (reality – prediction): on the British side (above); on the French side (below). Both graphs go from the coast (left) toward the tunnel midpoint (right). The distance axes have opposite directions, and their origins differ. The two parts are separated by a zone of about 10 km without control borehole. From Blanchin and Chiles (1993a,b), with kind permission of the International Association for Mathematical Geology and Kluwer Academic Publishers.

localized in sparsely sampled zones. This type of error is always a risk with geophysical (i.e., indirect) measurements, so it requires calibration data. The main objective of the geostatistical study, avoid penetrating the Gault Clay, was achieved: it never happened on the French side and happened twice on the British side, but where expected.

3.9. KRIGING UNDER INEQUALITY CONSTRAINTS

In natural phenomena one encounters inequalities of two types. The first are *global constraints* due to the very definition of the variables. For example, mineral grades, thicknesses of geological layers are positive quantities, and rock properties such as porosity and fluid saturations vary between 0 and 1. The second type are *local constraints* providing bounds on the results, and these can be regarded as *data*. For example, if drilling was stopped at a depth z_0 without hitting a given geological surface, we know that the depth of the surface at that point is $Z(x) > z_0$. Naturally the information carried by interval data $a \leq Z(x) \leq b$ depends on the tightness of the bounds.

Three different approaches have been considered to handle inequalities: one places constraints on the kriging weights, another constrains the estimates, and the third treats constraints as data. These methods should be regarded as engineering solutions to practical problems, rather than true kriging under constraints, which is not possible within a linear theory. Nonlinear approaches are presented in the scope of simulations (Sections 7.7.1 and 7.7.3).

3.9.1. Nonnegative Kriging Weights

The reason why known global constraints may be violated by kriging estimates is the possibility of weights < 0 or > 1 (cf. Example 4 with a power $\alpha = 1.5$; Figs. 3.14 and 3.16). The coincidence of a negative weight with a large sample value can produce negative estimates, or simply puzzling ones such as an estimated grade lower than every sample in the neighborhood. A sufficient but not necessary condition to preclude this is to constrain all kriging weights to be nonnegative. If in addition weights add up to one, the kriging estimates automatically lie within the minimum and maximum of the estimating data.

Constraints on Covariance Model

The ideal solution would be to use covariance or variogram models that automatically generate positive weights. Matheron (1986) investigated whether such models exist and stated the following conclusions.

Consider a set $I = \{1, 2, \dots, N\}$ of indexes that we arbitrarily partition into sample points and estimated points. The random variables Z_i have a covariance matrix Σ , assumed to be strictly positive definite. We want to find Σ such that kriging weights are ≥ 0 for all partitions of the set I , namely all kriging configurations. The results are as follows:

1. The problem has no solution for universal kriging except with a constant mean (OK).
2. SK and OK weights are ≥ 0 if and only if all off-diagonal terms of the inverse matrix Σ^{-1} are ≤ 0 .
3. Only in 1D can we find covariance or variogram functions ensuring positive kriging weights.

We can verify that the conditions on the off-diagonal terms of Σ^{-1} ensure positive cross-validation weights in a global neighborhood (Section 3.6.3). In the case of OK, the covariance matrix Σ in statement (2) can be replaced by the generalized covariance matrix \mathbf{K} where $K_{ij} = -\gamma_{ij}$, or $K_{ij} = C - a_i - a_j - \gamma_{ij}$, and \mathbf{K} is assumed strictly conditionally positive definite.

In 1D we already know that the exponential covariance ensures positive SK weights and the linear variogram positive OK weights. In fact the exponential covariance also ensures positive OK weights, and this property extends to completely monotone covariances (2.28). The (simple) kriging weights satisfy $\sum \lambda_\alpha \leq 1$ for all finite configurations and the optimal weights for the estimation of the mean are also positive. Likewise, positive OK weights are obtained with any variogram of the form

$$\gamma(h) = A|h| + C(0) - C(h) \quad (A \geq 0)$$

where $C(h)$ is a completely monotone covariance. Unfortunately, these results are only for 1D. In higher dimensions the conditions on Σ^{-1} stated above are not instrumental because in the geostatistical approach we model Σ and not Σ^{-1} . Also the fact that negative weights appear with the geostatistician's best friend, the spherical covariance, even in 1D, leaves little hope of finding useful direct models.

Positive Kriging

Barnes and Johnson (1984) propose a direct elimination of negative weights in ordinary kriging (the interesting case) by casting it as a quadratic programming problem

$$\begin{aligned} &\text{Minimize} \quad E \left(\sum_{\alpha} \lambda_{\alpha} Z_{\alpha} - Z_0 \right)^2 \\ &\text{subject to} \quad \sum_{\alpha} \lambda_{\alpha} = 1 \quad \text{and} \quad \lambda_{\alpha} \geq 0 \quad \forall \alpha \end{aligned}$$

Developing the equations of what they call "positive kriging," the authors find simple results that can be summarized as follows. Call "active constraints" (or "nonbasic weights") the weights that at the optimum are set to zero. Once the set of active constraints is known, the optimal values of the positive weights are the same as those obtained by OK using only those samples associated with the positive weights. The corollary is that if the OK solution only includes nonnegative weights, then the solution is also optimal for the positive kriging problem. This does not mean, however, that the active constraints coincide with the negative weights obtained by an initial execution of the OK procedure.

Note that a solution always exists for OK (e.g., the one sample estimators), and for SK (where the only solution might be the mean of the SRF), whereas for UK the solution does not necessarily exist.

Arguing that general purpose quadratic optimization algorithm do not exploit the particular structure of the positive kriging problem, Barnes and Johnson propose an approximate stepwise algorithm that can be implemented with minor effort in a standard kriging program: solve the OK system, eliminate the point with the most negative weight, recompute the weights, and repeat this until all weights are positive. Computational tricks are provided to avoid duplicate matrix inversions. Herzfeld (1989a) claims that this algorithm, which does not ensure convergence to the optimal solution, is unnecessary and that general purpose quadratic optimization programs would do just fine.

In any event, positivity constraints on the weights fit the kriging procedure like a straitjacket. One should check that the estimation variance does not suffer too much.

3.9.2. Minimization under Inequality Constraints

First, let us mention two simple methods for constraining estimates. One is prior *transformation* of the data, as with lognormal kriging, which automatically produces positive estimates. It can be a useful method even if the reverse transformation poses a bias problem. The other method is *censoring*: estimation is carried out without consideration of the constraints, and the results are modified as needed to satisfy the constraints. Of course we are concerned here with explicit supervised editing of the results, not burying a fix in a computer program. For example, when a geological layer “pinches out,” negative thickness estimates appear by continuity as the estimated point moves away from the layer boundary; it is legitimate to set these negative estimates to zero. In the following the constraints are part of the algorithm itself.

The general problem is to find an estimator $Z^*(x)$ such that

$$\begin{cases} Z^*(x_\alpha) = z_\alpha & \text{for } \alpha = 1, \dots, N \\ z_{\inf_\alpha} \leq Z^*(x_\alpha) \leq z_{\sup_\alpha} & \text{for } \alpha = N + 1, \dots, N + M \end{cases}$$

where the first N data are exact data and the remaining M are inequality data.

In two papers Dubrule and Kostov (1986) and Kostov and Dubrule (1986) approach this problem by first noting that a direct constrained minimization of the quadratic form $E(Z^* - Z_0)^2$ leads to an impasse because a constraint at the point x_α only affects estimation at x_α itself, without any lateral continuity, producing jumps just like kriging of a sample point in the presence of a nugget effect. They reformulate the problem under dual kriging as follows: Find an interpolating function of the form

$$z^*(x) = \sum_{\alpha=1}^{N+M} b_\alpha K(x - x_\alpha) + \sum_l c_l f^l(x)$$

where the coefficients b_α and c_l satisfy the conditions

$$\begin{cases} z^*(x_\alpha) = z_\alpha & \text{for } \alpha = 1, \dots, N \\ z_{\inf_\alpha} \leq z^*(x_\alpha) \leq z_{\sup_\alpha} & \text{for } \alpha = N + 1, \dots, N + M \\ \sum_{\alpha=1}^{N+M} b_\alpha f_\alpha^l = 0 & \text{for } l = 0, \dots, L \end{cases} \quad (3.37)$$

The first and third equalities are similar to the usual dual kriging equations (3.29) for exact data. The interval constraints can be broken down into two one-sided inequalities.

These conditions on the form of the interpolating function do not suffice to determine a unique solution. By analogy with spline theory, the authors propose to select the solution minimizing the quadratic form

$$Q = \sum_{\alpha=1}^{N+M} \sum_{\beta=1}^{N+M} b_\alpha b_\beta K_{\alpha\beta}$$

When $M = 0$ (no inequalities), the solution is the usual (dual) kriging estimator; when $M > 0$ and the spline covariance model is used for $K(x, y)$, Q is interpreted as the mean curvature of the interpolated surface, and the solution is then a thin plate spline.

A technical difficulty appears here because the function $K(x, y)$ used is not a genuine covariance but a “generalized covariance” (see Section 4.5). The quadratic form is not positive definite (strictly or not), and therefore not a convex function, which is bothersome for minimization. The

difficulty is only apparent because in reality we only deal with vectors \mathbf{b} that satisfy the conditions $\mathbf{F}'\mathbf{b} = \mathbf{0}$ (allowable linear combinations), and the restriction of the quadratic form to these vectors is a convex function. The following explicit expression is given by Langlais (1989): consider the inverse of the $(N + M)$ point kriging matrix defined by

$$\begin{bmatrix} \mathbf{K} & \mathbf{F}' \\ \mathbf{F}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}' & \mathbf{W} \end{bmatrix} = \mathbf{I}$$

and an arbitrary $(N + M)$ vector \mathbf{z} , then $\mathbf{b} = \mathbf{U}\mathbf{z}$ satisfies $\mathbf{F}'\mathbf{b} = \mathbf{0}$, and therefore $\mathbf{b}'\mathbf{K}\mathbf{b} \geq 0$ by definition of generalized covariances. So

$$\mathbf{z}'\mathbf{U}\mathbf{z} = \mathbf{b}'\mathbf{K}\mathbf{b} \geq 0 \quad \forall \mathbf{z}$$

The matrix \mathbf{U} is singular ($\mathbf{U}'\mathbf{F} = \mathbf{0}$) but nonnegative definite. The problem (3.37) may be cast in the equivalent form

$$\begin{cases} \text{Minimize } \mathbf{z}'\mathbf{U}\mathbf{z} & \text{subject to} \\ z^*(x_\alpha) = z_\alpha & \text{for } \alpha = 1, \dots, N \\ z_{\inf_\alpha} \leq z^*(x_\alpha) \leq z_{\sup_\alpha} & \text{for } \alpha = N + 1, \dots, N + M \end{cases}$$

Once \mathbf{z} is determined, thus the values at the M constraint points, one computes the dual kriging solutions $\mathbf{b} = \mathbf{U}\mathbf{z}$ and $\mathbf{c} = \mathbf{V}'\mathbf{z}$. The constrained minimization problem has a unique solution if the UK problem for the N exact data has itself a unique solution (i.e., there are at least $L + 1$ exact data points over which the drift functions are linearly independent); this condition is sufficient but not necessary (Langlais, 1989).

Figure 3.25 from Dubrule and Kostov illustrates the method in 1D using cubic spline interpolation: first, only the 9 exact data are used; then 18 inequality constraints are added, consisting of $M_1 = 12$ lower bounds, $M_2 = 6$ upper bounds, two points ($x = 4$, $x = 6$) having two-sided inequalities. In the first pass 12 inequalities are violated; in the second pass they are all satisfied of course, but notice the “clamping effect” at the bounds: the function is exactly equal to some of the bounds. This is how quadratic programming works: it selects some of the constraints (the “active” ones) and satisfies them at the bounds; then all the other constraints are automatically satisfied as well. If we knew which constraints are the active ones, it would be possible to introduce them as equality data, ignore the others, and proceed with normal kriging. Just as with positive kriging the problem is to determine the active set, for it does not coincide with the constraints that are violated by an initial execution of the kriging procedure based only on exact data.

Minimization under constraints requires special care when moving neighborhoods are used because the algorithm may assign different values at the same inequality point depending on the neighborhood considered. The recommended procedure is to select exact data first, using the standard neighborhood search algorithm, and then include the constraint points that would have been selected by the search algorithm operating without distinction between data and constraints. To summarize, inequality data are treated as secondary information.

3.9.3. Inequalities as Data

An alternative approach proposed by Langlais (1990) is to regard inequalities as data and replace them by exact values. The procedure is to simulate exact data satisfying the given inequalities, proceed to kriging from both actual and generated data, and finally average the results over several

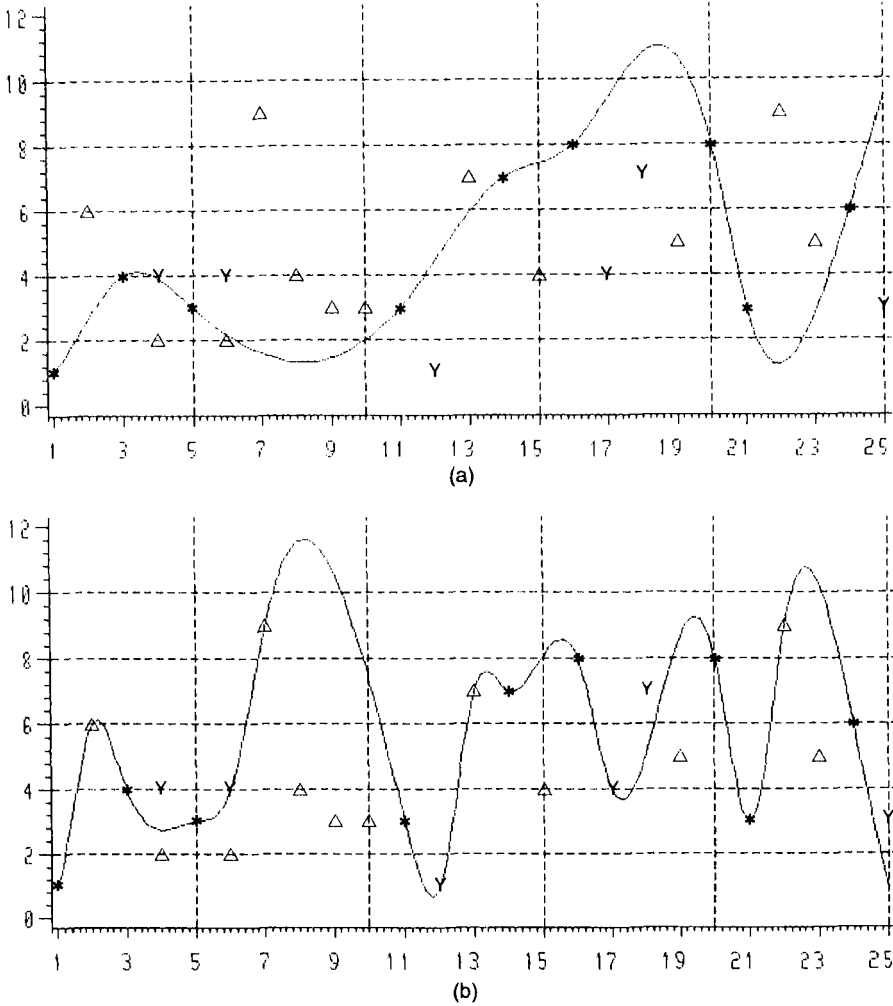


FIGURE 3.25. Interpolation (a) based on exact data only; (b) with constrained optimization. The function is a cubic spline obtained with $K(h) = |h|^3$ and a linear drift. Asterisks are exact data, triangles are lower bounds, and “Y” are upper bounds. From Dubrule and Kostov (1986), with kind permission of the International Association for Mathematical Geology.

simulations of the inequality data. We will keep her approach but present a new implementation based on the use of the Gibbs sampler.

To simplify notations, we consider here that all data are of the form $Z_{\alpha} \in B_{\alpha}$, where B_{α} denotes an interval, or more generally a Borel set (e.g., a finite union of intervals), with the understanding that it may be reduced to a single point in the case of exact data. If Z is a Gaussian RF with known mean, consider the conditional expectation

$$E[Z(x) | Z_{\alpha} \in B_{\alpha}, \forall \alpha]$$

In principle, this expectation can be calculated by integration of the Gaussian p.d.f., but in practice, this is an intractable problem. Instead, note that the simple kriging estimator is

$$E[Z(x) | Z_\alpha = z_\alpha, \forall \alpha] = \sum_{\alpha} \lambda_{\alpha} z_{\alpha}$$

so that

$$E[Z(x) | Z_\alpha \in B_\alpha, \forall \alpha] = \sum_{\beta} \lambda_{\beta} E[Z_{\beta} | Z_\alpha \in B_\alpha, \forall \alpha]$$

Thus it suffices to perform simple kriging after replacing the inequality data by their conditional expectations $E(Z_{\beta} | Z_\alpha \in B_\alpha, \forall \alpha)$. We compute these means empirically by generating samples from the joint conditional distribution of the Z_α given all data $\{Z_\alpha \in B_\alpha\}$, using an algorithm known as the Gibbs sampler (see Section 7.7.2). This is implemented by repeating the following sequence:

1. Select an index α in the set of inequality data.
2. Simulate Z_α conditionally on $Z_\alpha \in B_\alpha$ and $Z_\beta = z_\beta$ for all $\beta \neq \alpha$ (β ranges over *all* data).

The procedure can be initialized by generating each Z_α separately, conditionally on $Z_\alpha \in B_\alpha$. The index α may be scanned either periodically or using an irreducible Markov chain, the important point being that, in theory, each index is almost surely drawn infinitely often. It is possible to generalize the method to more complex constraints defined as a Borel set of \mathbb{R}^N (e.g., ellipsoids instead of parallelepipeds; N is the number of data).

This approach finds its theoretical justification in the ideal case of a Gaussian RF with known mean. It can be used more generally, *as an algorithm*, by assuming that at each step the conditional distribution is Gaussian with mean the kriging estimate and variance the kriging variance. The algorithm ensures that the inequality data are accounted for in a consistent manner, but of course its optimality properties are unknown. The same approach is used effectively to generate conditional simulations constrained by inequality data (Section 7.7.3).

NOTES

1. To make the presentation more accessible, the convenient and concise tensor notation $\lambda^\alpha Z_\alpha$ used by Matheron (1969a, 1970) will not be employed here.
2. Conditional unbiasedness has been established here for point estimation but remains valid in the case of blocks.
3. The term “drift” is standard for a stochastic process (e.g., a Brownian motion with a drift). The residual considered here is the *true* residual, by contrast with the calculated residual which is what remains after subtraction of a fit; *fluctuation* would be a more neutral term.
4. The systems (3.16) and (3.14) have the same solution except that μ becomes $-\mu$.
5. The most popular model for the evolution of stock prices, which underlies the derivation of the famous Black–Scholes formula for options pricing, has the logarithm of stock price as a Wiener process with a linear drift (e.g., Jarrow and Rudd, 1983). But that does not make forecasting easier, as Example 5 shows.
6. This condition can be weakened; the true requisite is closedness of the set under translations (see Section 4.3.4).
7. A misprint in Pilz’s article has been corrected in this formula.
8. Michel David once described lognormal kriging as “riding a wild horse” (quoted in Snowden, 1994).

9. On the subject of digital image processing, see the article by Yfantis et al. (1994) and comment by P. Delfiner (same reference).
10. Structural geologists have developed the general concept of *retrodeformability*. Let us quote Suppe (1985, p. 57): "The fact that the rocks were originally deformed may seem trivial to mention, but it is actually an important key to the solution of many structural problems. It must be geometrically possible to undeform any valid cross section to an earlier less deformed or undeformed state; the cross section must be *retrodeformable*." Retrodeformable cross sections are popular in the petroleum industry under the name of *balanced cross sections*.
11. In real time the GPS system accuracy is within 100 m for commercial use and well within 10 m for military use; it can be further improved to the 1 m range using differential corrections.
12. We remember being told a story about a devastating hurricane that could have been forecast, had the data from a meteorological ship not been discarded by a statistical procedure based on principal components.

CHAPTER 4

Intrinsic Model of Order k

*A trend is a trend is a trend
but the question is will it bend?
Will it alter its course
through some unforeseen force
or come to a premature end?*

From an unknown author, quoted by
R. C. Selley, a sedimentologist

4.1. IRF-0 AND IRF- k

The notion of intrinsic random function of order k (henceforth abbreviated as IRF- k) constitutes a natural generalization of the intrinsic random functions (i.e., with stationary increments) of traditional geostatistics. These correspond to the particular case $k = 0$ and will now be called IRF-0. In passing from stationary random functions (SRF) to IRF-0, the following changes take place:

- The basic working tool of the stationary case, the covariance $C(h)$ is replaced by the variogram $\gamma(h)$. This extends generality, since the class of valid variogram functions is broader than the class of covariance functions ($C(h)$ must be positive definite whereas it is only required that $-\gamma(h)$ be conditionally positive definite). In contrast to the covariance, the variogram may be unbounded, and this enables the description of phenomena with a potentially unlimited dispersion (theoretically infinite variance) such as the Brownian motion.
- In the stationary case there exists a mean value m about which the SRF fluctuates. The phenomenon remains “controlled” in the sense that the deviations from the mean are never too large nor last too long. In the case of an IRF-0 with an unbounded variogram, no such regulation exists. The Brownian particle has no memory. It bounces from its current position as if it were a new origin and shows no tendency to revert to its starting point. There is no constant mean value m . In effect such a process is

defined up to an arbitrary constant and is generally studied only through its increments.

- In the case of an SRF, any linear combination

$$Z(\lambda) = \sum_i \lambda_i Z(x_i)$$

has the variance

$$\text{Var} Z(\lambda) = \sum_i \sum_j \lambda_i \lambda_j C(x_j - x_i)$$

where $C(h)$ is a centered covariance. In the case of an IRF-0, only special linear combinations have a finite variance, the allowable ones, satisfying the condition $\sum \lambda_i = 0$. The variance is then calculated using $C(h) = -\gamma(h)$ as if it were a covariance function

$$\text{Var} Z(\lambda) = - \sum_i \sum_j \lambda_i \lambda_j \gamma(x_j - x_i)$$

Other combinations do not, in general, have a finite variance. Thus, at the cost of a relatively minor operating restriction (only use linear combinations summing up to zero), we gain the possibility of dealing with a large class of phenomena that, like the Brownian motion, cannot be represented by a stationary model.

Is it possible to go even further in the same direction? In other words, can we get access to broader classes of nonstationary phenomena at the price of restrictions more severe than $\sum \lambda_i = 0$? The theory of IRF- k provides a positive answer to this question. The idea is to define models through increments of a sufficiently high order for stationarity to be reached, an approach generalizing the ARIMA models of time series analysis (Box and Jenkins, 1976). We discover that these models are characterized by a new structure function that is completely free of the influence of polynomial drifts and turns out to be the real minimum prerequisite for universal kriging. The IRF- k approach gets us around the statistical inference problem caused by the biases affecting the raw variogram and the variogram of residuals.

More fundamentally it also provides the correct conceptual model to represent the nonstationary solutions of stochastic partial differential equations. Consider, for example, the classic Poisson equation $\Delta Z(x) = Y(x)$, where Δ is the Laplacian operator in \mathbb{R}^n . It is intuitively obvious that if $Y(x)$ is stationary, the solutions $Z(x)$ will not, in general, be stationary. In fact the solution is a twice differentiable IRF-1. It follows, for example, that if the log-transmissivity $Y(x)$ over an aquifer is stationary, the correct theoretical model for the piezometric-head perturbation $Z(x)$ is an IRF-0 (Section 8.4.3). The

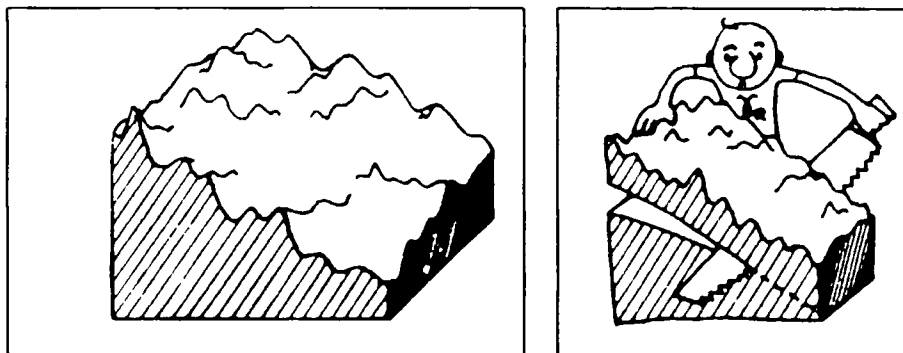


FIGURE 4.1. A favorable case for universal kriging. Author: J. P. Delhomme.

IRF- k theory may sometimes appear mathematically awesome, but it builds a connection between geostatistics and physics.

4.2. A SECOND LOOK AT THE MODEL OF UNIVERSAL KRIGING

The basic model of universal kriging is the dichotomy

$$Z(x) = m(x) + Y(x) \quad (4.1)$$

where $Z(x)$ is the variable under study, $m(x)$ is the drift, and $Y(x)$ is the fluctuation, or residual, about this drift. From a mathematical point of view, the drift is well-defined as the expected value $m(x) = EZ(x)$. But $m(x)$ is not an observable, except if there are repetitions allowing us to actually compute $m(x)$ as an average across several realizations of the same phenomenon.

When the phenomenon is unique, as it is in geological applications, this $m(x)$ is a purely theoretical construct. Its modeling is inspired by observations of “trends,” namely systematic patterns of variations in the data. But in reality the drift is an elusive concept, sometimes unclear or very complex, sometimes clear but spurious.

4.2.1. Questioning the Dichotomy into Drift and Fluctuation

Except for replications, the most meaningful case for the dichotomy (4.1) is that of a phenomenon showing small local fluctuations about a clear overall trend liable to be modeled by a simple and smooth mathematical function (Fig. 4.1). In the model the trend is treated deterministically because it is simple, whereas the fluctuations are too complex to be described in detail and are captured by probabilistic means. Specifically $Y(x)$ is a stationary random function with a zero mean and a covariance whose range is short at the scale

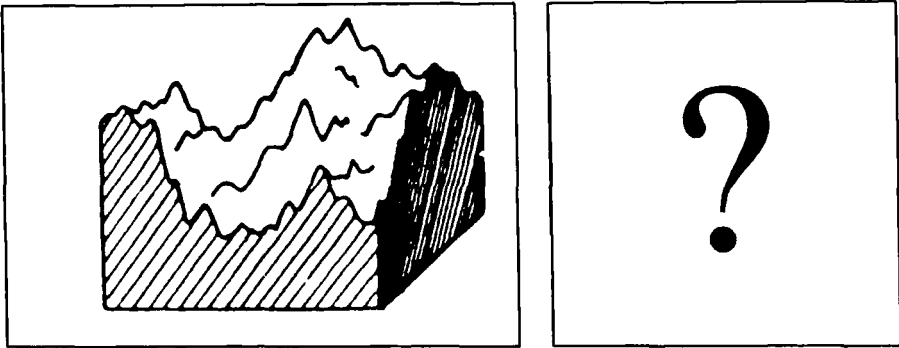


FIGURE 4.2. A puzzling case for drift modeling. Author: J. P. Delhomme.

of the study. The drift may be estimated reliably and subtracted from the data to restore stationarity. The underlying covariance can be determined from the variogram of residuals which, at short distances, differs little from the true variogram. The drift then possesses an objective meaning in the sense of Section 1.2.

The situation is very different when the covariance of fluctuations does not have a short range. As we have seen in Section 2.7, the drift causes both the experimental variogram and the variogram of residuals to be considerably biased, which makes statistical inference of the covariance very difficult and thereby precludes the use of statistical tests aimed precisely at deciding whether there is a drift or not.

Assuming that it can be defined, the drift is not necessarily simple enough to be modeled by an analytic expression valid over the whole studied domain (Fig. 4.2). In such cases one can turn to a local model of the form

$$m(x) = \sum_l a_l(x_0) f^l(x) \quad (4.2)$$

valid only in a neighborhood $V(x_0)$ of each point x_0 . As usual, the $f^l(x)$ are given functions (typically monomials), but the unknown coefficients a_l vary with the neighborhood. Their estimation is even more difficult than before, since there are fewer data points in each neighborhood, and we are confronted with the problem of piecing together local estimates. The map obtained by moving neighborhood drift estimation often looks much less smooth than the kriging map, which is rather disturbing (Chilès, 1979a). In reality, in the absence of a clear separation of scales between $m(x)$ and $Y(x)$, the dichotomy is simply arbitrary, and the drift is not an objective parameter.

4.2.2. Examples of Zero-Mean Processes with Apparent Drifts

Chance fluctuations may produce clear trends. A well-known and striking example is shown in Figure 4.3 taken from Feller (1968) displaying a record

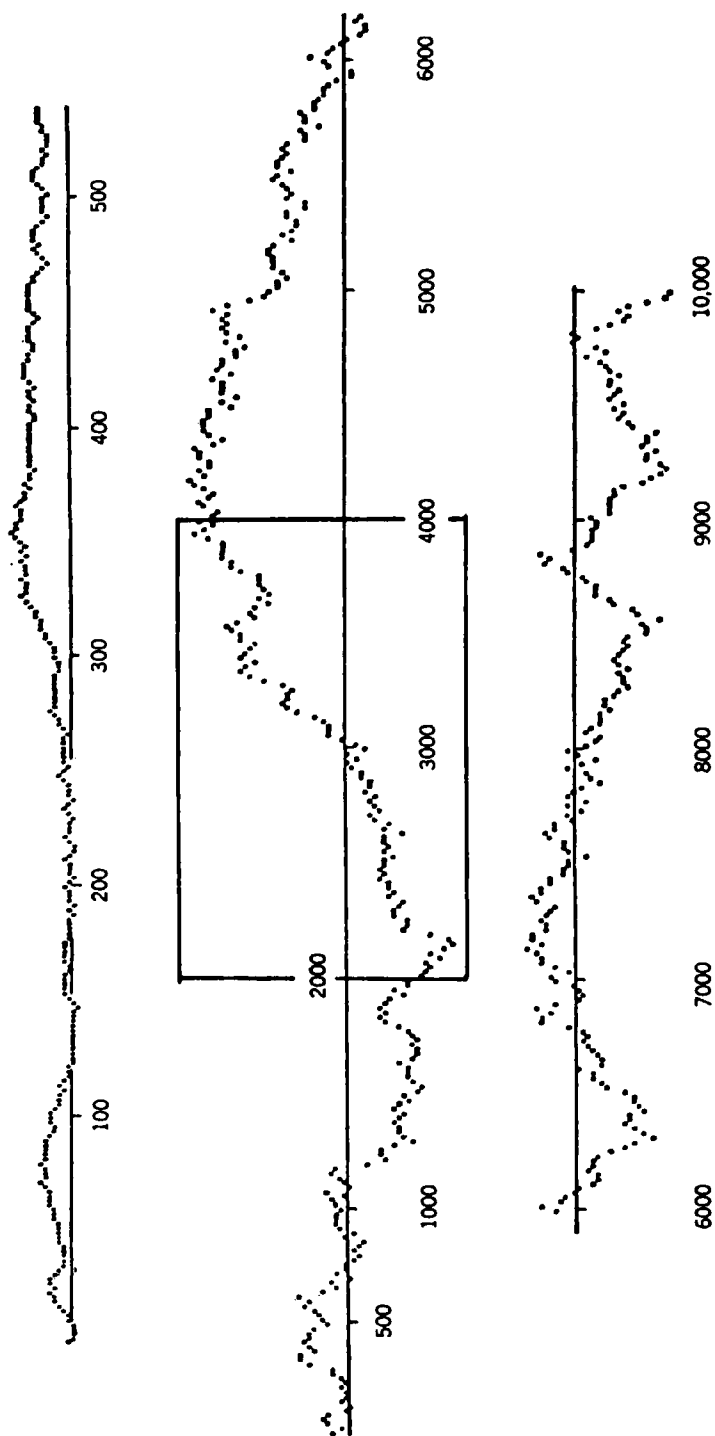


FIGURE 4.3. The record of 10,000 tosses of an ideal coin. From Feller (1968, p. 87), with permission of John Wiley & Sons, Inc.

of a coin-tossing experiment. The function graphed is

$$S_n = X_1 + X_2 + \cdots + X_n$$

where $X_i = +1$ or $X_i = -1$ according to the outcome of the trial with a fair coin.

Given the string between $n = 2000$ and $n = 4000$ and told that these are daily oil prices, an “expert” would fit an upward trend and extrapolate it for the next decade. Alas, the trend turns and now points downward for the next 2000 samples! So the next time the expert tries to forecast the “turning point.” Of course these trends are mere fluctuations of the random walk which has mean zero for each n .

Large fluctuations such as these can occur due to the lack of any regulating mechanism in the process. By contrast, a stationary process is subject, so to speak, to an elastic force pulling it back to its mean. For example, consecutive increments $Z(x_3) - Z(x_2)$ and $Z(x_2) - Z(x_1)$ at three aligned points x_1, x_2, x_3 have a correlation coefficient of $-\frac{1}{2}$ when the point interdistance is larger than the range: a move up tends to be compensated by a move down. But the variogram of S_n does not have a range since it is unbounded ($\text{Var}(S_n - S_m) = |n - m|$).

Examples like this are not limited to one dimension. We can generate surfaces that show clear systematic patterns even though, by design of the simulation algorithm, $EZ(x) = 0$ for all x . Figure 4.4a shows a simulation of a Brownian RF in 2D. The process has a mean zero and a linear variogram; it was simulated by turning bands (see Section 7.4). Figures 4.4b and c were obtained similarly by radial integrations of 4.4a (Section 7.6.3); they exhibit long-range patterns that could be interpreted as drifts, though more complicated ones than planes or quadratics. But we know that no drift was incorporated in the generation of the process.

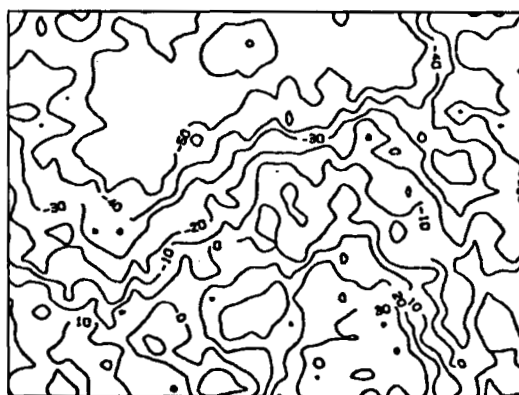
These examples illustrate the existence of zero-mean processes whose fluctuations have an aspect usually attributed to the presence of a trend. There is an *apparent* drift but no genetic drift. Any interpretation of these pseudodrifts as systematic effects is totally spurious. One has to accept the fact that chance fluctuations may lead to results that contradict common intuition.

4.3. ALLOWABLE LINEAR COMBINATIONS OF ORDER k

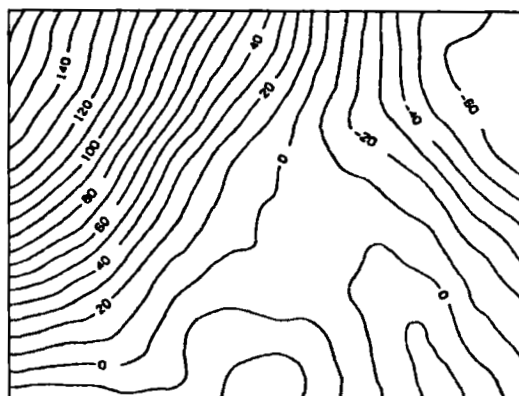
4.3.1. Allowable Measures and Generalized Increments of Order k

A set of weights λ_i applied to m points x_i of \mathbb{R}^n defines a discrete measure λ of the form

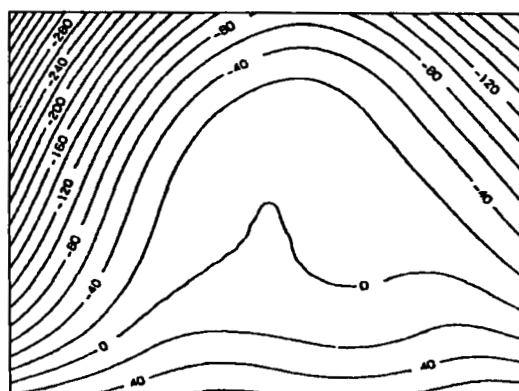
$$\lambda = \sum_{i=1}^m \lambda_i \delta_{x_i} \quad (4.3)$$



(a)



(b)



(c)

FIGURE 4.4. Simulations of IRFs of order $k = 0, 1$, and 2 . Panels b and c suggest drifts that were not present in the construction of the models. From Orfeuil (1972).

where δ_{x_i} is the Dirac measure at the point x_i (Dirac delta function). The action of λ on a function $f(x)$ defines a linear combination that will be synthetically denoted by $f(\lambda)$:

$$f(\lambda) = \int f(x)\lambda(dx) = \sum_{i=1}^m \lambda_i f(x_i)$$

Definition. A discrete measure λ is allowable at the order k if it annihilates polynomials of degree less than or equal to k

$$P(\lambda) = 0 \quad \text{whenever} \quad \text{degree } P \leq k \quad (4.4)$$

We call Λ_k the class of such allowable measures.

It is clear that (4.4) is achieved if and only if λ annihilates separately all monomials of degree up to k . In one-dimensional space there are $k + 1$ conditions, one for each power of x . In n -dimensional space there are

$$k_n = \binom{n+k}{k}$$

monomials of degree less than or equal to k and thus as many conditions. To avoid quadruple indexes, it is convenient to use the following condensed notations:

$x_i = (x_{i1}, \dots, x_{in})$	for a point in \mathbb{R}^n (but in 2D we use (x_i, y_i) as the coordinates)
$l = (l_1, \dots, l_n)$	for a set of nonnegative integers
$x_i^l = x_{i1}^{l_1} x_{i2}^{l_2} \cdots x_{in}^{l_n}$	for a monomial
$ l = l_1 + \cdots + l_n$	for the degree of x^l

Then (4.4) is equivalent to the set of conditions

$$\sum_{i=1}^m \lambda_i x_i^l = 0 \quad |l| = 0, 1, \dots, k \quad (4.5)$$

that is, all moments of order up to k , inclusive, are zero. Obviously we have $\Lambda_{k+1} \subset \Lambda_k$.

An allowable measure $\lambda \in \Lambda_k$ defines an *allowable linear combination of order k* (abbreviated as *ALC- k*) also called a *generalized increment of order k* ,

$$Z(\lambda) = \sum_i \lambda_i Z(x_i) \quad (4.6)$$

Link with Error Contrasts

The concept of ALC- k is of algebraic nature and can be presented in the framework of linear models and "error contrasts" familiar to statisticians. Suppose that in the standard notations of linear models

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U} \quad (4.7)$$

where $\mathbf{Y} = (Y_1, \dots, Y_N)'$ is a vector of observations ($N > k_n$), $\mathbf{X} = (x_i^l)$ is the $N \times k_n$ matrix of all monomials evaluated at points x_1, \dots, x_N , $\beta = (\beta_0, \dots, \beta_{k_n-1})'$ is a vector of coefficients and $\mathbf{U} = (U_1, \dots, U_N)'$ a vector of residuals. Formula (4.7) is exactly the discrete formulation of the universal kriging model with a polynomial drift of degree k . Then a linear combination $\sum \lambda_i Y_i$ is an ALC- k if the relation (4.5) is satisfied, namely if the vector of weights $\lambda = (\lambda_1, \dots, \lambda_N)'$ satisfies

$$\lambda' \mathbf{X} = \mathbf{0} \quad (4.8)$$

Combining (4.7) and (4.8) gives

$$\lambda' \mathbf{Y} = \lambda' \mathbf{U}$$

and it is seen that β (the drift coefficients) has been completely eliminated. Generalized increments are linear functions of \mathbf{U} only; they are *error contrasts*. Conversely, if a function $\Phi(\mathbf{Y})$ of the data does not depend on β at all, namely satisfies the invariance property

$$\Phi(\mathbf{Y} + \mathbf{X}\mathbf{a}) = \Phi(\mathbf{Y}) \quad \forall \mathbf{a}$$

then $\Phi(\mathbf{Y})$ depends on \mathbf{Y} only through $N - k_n$ error contrasts (Delfiner, 1977). This justifies why in a theory where the drift is to be bypassed, it is necessary to allow some linear combinations and forbid others.

Before turning to examples, it is worth mentioning that the notion of ALC- k can be extended to the continuous case by considering

$$f(\mu) = \int \mu(dx) f(x)$$

for measures μ in the larger class \mathcal{M}_k of measures vanishing outside a compact set (measures with compact support) and satisfying

$$\int \mu(dx) x^l = 0 \quad |l| = 0, \dots, k$$

4.3.2. Examples

Finite Differences on the Line

The forward finite difference of order $(k+1)$ of a function $f(x)$ is defined by

$$\Delta_a^1 f(x) = f(x+a) - f(x)$$

$$\Delta_a^{k+1} f(x) = \Delta_a^1 \Delta_a^k f(x) = (-1)^{k+1} \sum_{p=0}^{k+1} (-1)^p \binom{k+1}{p} f(x+pa)$$

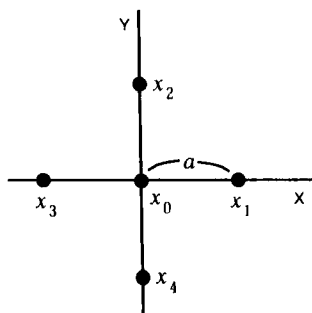


FIGURE 4.5. $f(\lambda) = f(x_1) + f(x_2) + f(x_3) + f(x_4) - 4f(x_0)$ is an allowable linear combination of order 1 but not of order 2.

As Δ_a^1 decreases by one the degree of any polynomial, it is easy to see by induction that finite differences of order $k + 1$ annihilate polynomials of degree k . So they are ALC- k .

Five-Point Laplacian Approximation

In 2D consider 5 points x_0, x_1, \dots, x_4 arranged as in Figure 4.5. Let $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1$ and $\lambda_0 = -4$, and let x_i and y_i denote the two coordinates of the point x_i . The measure λ is allowable at order 1, since

$$\sum_i \lambda_i = \sum_i \lambda_i x_i = \sum_i \lambda_i y_i = 0$$

Since $x_i y_i = 0$ for all i , λ also annihilates the monomial xy . But it does not annihilate x^2 nor y^2 :

$$\sum_i \lambda_i x_i^2 = \sum_i \lambda_i y_i^2 = 2a^2$$

Therefore λ is not allowable at the order 2. (As a matter of fact $f(\lambda)/a^2$ is a finite difference approximation of the Laplacian Δf of f at the point x_0 .)

Increments on a Circle

$2k + 2$ points on a circle at regular angular intervals $\pi/(k + 1)$ and with alternating weights $+1$ and -1 define an allowable measure of order k . The proof follows from the orthogonality of the complex exponentials over the set of points and De Moivre's theorem $(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$ (Chilès, 1977). Note for example that for $k = 3$ there are 8 points for 10 monomials to annihilate.

Errors from Unbiased Estimation

A simple way of constructing an ALC- k is to take the difference between a value and a linear unbiased estimate of this value calculated under the assumption that the mean is a full polynomial of degree k . Indeed

$$\hat{Z}_0 = \sum_i \lambda_i Z_i$$

is an unbiased estimator of $Z_0 = Z(x_0)$ if and only if

$$\sum_i \lambda_i x_i^l = x_0^l \quad |l| = 0, \dots, k$$

and therefore $\sum \lambda_i Z_i - \hat{Z}_0$ is an ALC- k .

The manner in which \hat{Z}_0 is obtained does not matter: it can be kriging (whatever the variogram) or least squares. For this purpose least squares have the advantage of being the simplest, and they are used most. Note that several ALC- k may be obtained from the same least squares fit: all the residuals at the points used for the fit plus estimation errors at arbitrary other points not used for the fit.

4.3.3. Minimum Number of Points

According to Equation (4.8) an ALC- k based on N points has weights λ_i which are a nontrivial solution of $\mathbf{X}'\mathbf{X} = \mathbf{0}$. If the columns of \mathbf{X} are linearly independent, necessarily $N \geq \text{rank}(\mathbf{X}) + 1$. Since all monomials of \mathbb{R}^n are used, the rank of \mathbf{X} is k_n so that $k_n + 1$ is the minimum number of points required. Minimum values of N according to the dimensionality n of the space are

2	points for $k = 0$
$n + 2$	points for $k = 1$
$(n + 1)(n + 2)/2 + 1$	points for $k = 2$

However, if the points are such that there is linear dependence between the columns of \mathbf{X} , it is possible to construct increments with fewer points because some of the constraints are automatically satisfied. For example, on a circle, the fact that $x^2 + y^2 = 1$ implies that there are only $2k + 1$ linearly independent monomials of degree up to k instead of $(k + 1)(k + 2)/2$ for the whole plane; so $2k + 2$ points suffice to define an ALC- k as was indicated above. Geometrically this means that all points lie on a curve or surface defined by an algebraic equation of the form $\mathbf{X}\mathbf{a} = \mathbf{0}$, such as a line or plane if $k = 1$, a conic or quadric if $k = 2$.

In practice, this possibility of taking advantage of the location of the points is interesting when data are along lines: increments of order k on the line, such as finite differences, are also valid increments in the plane or space. Indeed a change of coordinate system transforms a polynomial of degree k into another polynomial of degree k . Taking the line as one of the axes makes all other coordinates zero and shows that the conditions involving these coordinates are automatically satisfied.

4.3.4. Why Polynomials Are Used

We define the translate $\tau_h \lambda$ of the measure λ by the vector h as the measure with the same weights as λ but applied to a point configuration shifted by h , namely

$$f(\tau_h \lambda) = \sum_i \lambda_i f(x_i + h)$$

A problem is to know whether, when shifted, an allowable measure remains allowable; in other words, Is Λ_k closed under translations? The answer is in the affirmative owing to the binomial formula

$$(x + h)^l = \sum_{s=0}^l \binom{l}{s} x^s h^{l-s}$$

since

$$\sum_i \lambda_i (x_i + h)^l = \sum_{s=0}^l \binom{l}{s} h^{l-s} \sum_i \lambda_i x_i^s = 0$$

for $|l| \leq k$ and $\lambda \in \Lambda_k$. In \mathbb{R}^n , $s = 0, \dots, l$ is short for $s_j = 0, \dots, l_j \forall j = 1, \dots, n$, and

$$\binom{l}{s} = \frac{l!}{s!(l-s)!}$$

where $l! = l_1! \dots l_n!$. This result ensures that the translate $Z(\tau_h \lambda)$ of an ALC- k $Z(\lambda)$ is also an ALC- k , a property without which stationarity assumptions on generalized increments would not make sense. It is a direct consequence of the property of polynomials to be themselves closed under translations. Are polynomials special, or can other functions be used to construct generalized increments?

Mathematically the problem is the following: find $p + 1$ functions $f^l(x)$ such that the finite-dimensional vector space \mathcal{F} generated by the $f^l(x)$ is closed under translations. In other words, we want that

$$f(x) = \sum_{l=0}^p a_l f^l(x) \Rightarrow f(x + h) = \sum_{l=0}^p a_l(h) f^l(x)$$

Obviously this will hold if and only if it holds for each $f^l(x)$. So the $f^l(x)$ are solutions of the functional equation

$$f^l(x + h) = \sum_{s=0}^p B_s^l(h) f^s(x) \quad (4.9)$$

A general theorem (Matheron, 1979a) states that the only continuous (and even the only measurable) solutions of (4.9) are finite sums of functions of the form

$$f(x) = P(x) \exp(c_1 x_1 + \cdots + c_n x_n)$$

where $P(x)$ is a polynomial in $x = (x_1, \dots, x_n)$ and $c = (c_1, \dots, c_n)$ are real or complex coefficients.

For $n = 1$, \mathcal{F} is generated by families of exponential monomials

$$\{x^l e^{cx}; l = 0, \dots, k\} \quad (4.10)$$

Since we only consider real functions, if $\omega \neq 0$, the complex coefficient $c = \alpha + i\omega$ must have its conjugate counterpart, and \mathcal{F} is generated by the functions

$$\{x^l e^{\alpha x} \cos \omega x, x^l e^{\alpha x} \sin \omega x; l = 0, \dots, k\}$$

There are three remarkable subsets of this family:

1. Pure polynomials $P(x)$ of degree $\leq k$
2. Pure trigonometric functions $\{\cos \omega x, \sin \omega x\}$
3. Pure exponentials $\{\exp(\alpha x)\}$

Note that for exponentials or trigonometric functions, the rate α or the frequency $\omega/2\pi$ must be selected, whereas for polynomials there is no scaling parameter.

The same results hold formally when $n > 1$. It suffices to regard ωx as a scalar product $\langle \omega, x \rangle = \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n$. However, when $n > 1$, \mathcal{F} does not necessarily contain all terms of (4.10). For example, for $n = 2$ and $x = (x, y)$, the space \mathcal{F} generated by the functions: $1; x; y; x^2 + y^2$ is invariant under shifts but does not contain x^2 and y^2 , nor xy .

From a theoretical point of view, it is possible to develop a theory of intrinsic random functions based on all the solutions of (4.9), including the exponential terms (Matheron, 1979a). This extension will not be presented here. Indeed in more than one dimension there is a genuine difficulty to use functions including an exponential component because they are scale and orientation dependent. Except perhaps in very specific cases it is desirable that allowable measures remain so under scaling or rotations. If this is also required, the polynomial solutions of (4.9) are the only possible ones.

In conclusion, polynomials are not used just for the sake of convenience but because they satisfy fundamental geometric invariance requirements.

4.4. INTRINSIC RANDOM FUNCTIONS OF ORDER k

4.4.1. Illustration of the Concept

Figure 4.4b shows clear patterns and we can qualify this phenomenon as nonstationary. The raw directional variograms plotted in Figure 4.6a support this view: they exhibit a parabolic increase and a violent anisotropy whose main axes are understandably in the NE–SW and NW–SE directions.

From the grid of values we can construct a new grid assigning to each grid node (i, j) —except on the edges—the value of the increment of order 1

$$I(i, j) = Z(i - 1, j) + Z(i + 1, j) + Z(i, j - 1) + Z(i, j + 1) - 4Z(i, j)$$

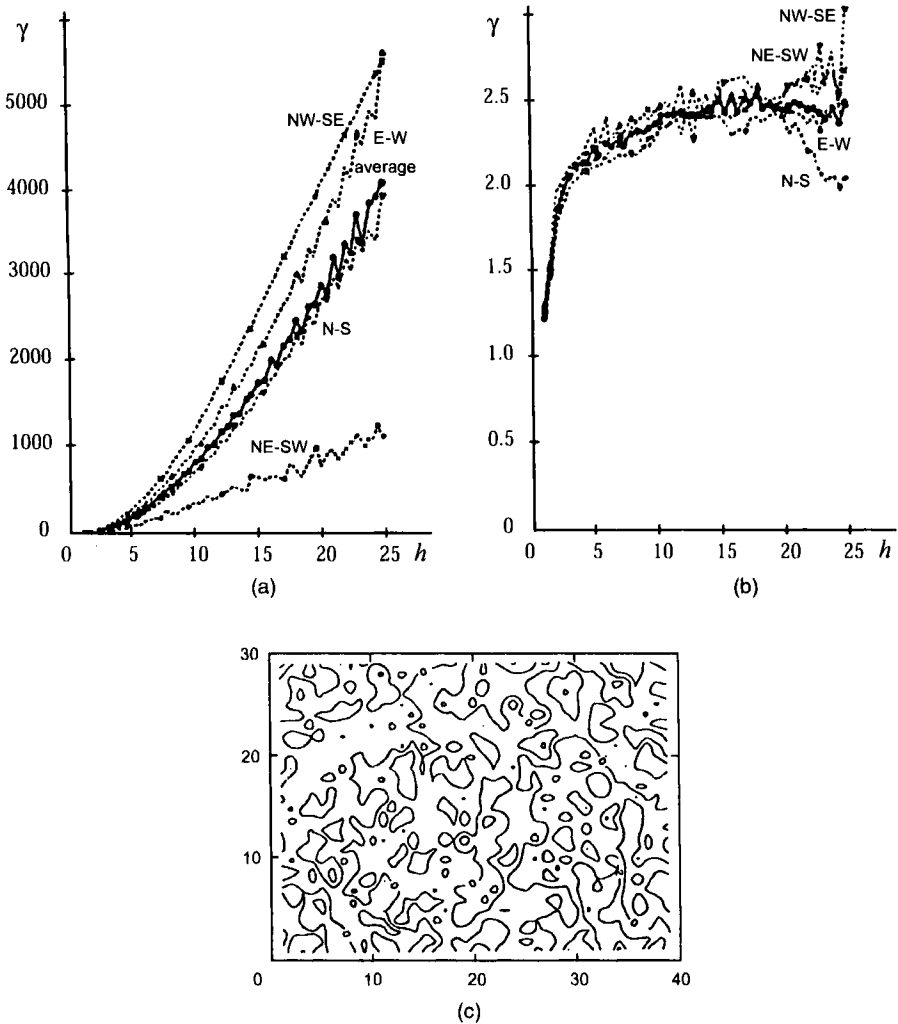


FIGURE 4.6. A case where increments of order 1 render data stationary: (a) raw directional variograms of grid displayed in Fig. 4.4b; (b) directional variograms of increments of order 1; (c) map of increments of order 1.

The grid of increments is displayed in Figure 4.6c. The “systematic” effects have been removed; the increments of order 1 appear stationary. This impression is confirmed by the existence of a clear sill on the variograms of $I(i, j)$ in Figure 4.6b. Furthermore the anisotropy has disappeared, which indicates that it was due to a (local) polynomial component that the increments have filtered. In this case, taking increments of order 1 turns nonstationary data into stationary ones.

This is not always the case. Starting from Figure 4.4c, which is highly non-stationary (Fig. 4.7a), we can see that the increments of order 1 still exhibit a drift (Fig. 4.7b). However, taking increments of order 1 again, which amounts to taking increments of order 2 of the initial data, establishes stationarity (Fig. 4.7c and d). Again this is evidenced by a clear sill on the variograms and the restoration of isotropy. Note that by construction of the simulations, the noise level is zero so that the low correlation of increments of order 2, indicated by the flat variogram, is a genuine feature of the variable $Z(x)$ and not a mere reflection of a deteriorated signal-to-noise ratio.

The examples of Figures 4.6 and 4.7 are meant to illustrate the stationarity property which defines the intrinsic random functions of order k . They should not convey the false impression that the data need to be on a regular grid. As will be seen, the IRF- k approach is applicable to scattered data, even though an explicit computation and display of increments is no longer possible.

Two mathematical definitions of IRF- k will be given: one is Matheron's original definition (1973a) which is abstract but mathematically more profound, and the other perhaps is formally less satisfactory but simpler. We start with the latter.

4.4.2. Ordinary IRF- k

Definition 1. A random function $Z(x)$ is intrinsic of order k if for any allowable measure $\lambda \in \Lambda_k$ the random function

$$Z_\lambda(x) = Z(\tau_x \lambda) = \sum_i \lambda_i Z(x_i + x)$$

is second-order stationary in $x \in \mathbb{R}^n$ and has a zero mean.

This is equivalent to

$$\begin{cases} E[Z_\lambda(x)] = 0 \\ E[Z_\lambda(x)Z_\lambda(y)] = K_\lambda(y-x) \quad \forall x, y \in \mathbb{R}^n, \quad \lambda \in \Lambda_k \end{cases}$$

An IRF- k is simply a random function with stationary increments of order k . The usual intrinsic model of geostatistics corresponds to $k = 0$. Clearly an IRF- k is also an IRF- $(k+1)$ and of any higher order, since $\Lambda_{k+1} \subset \Lambda_k$. For example, an SRF is intrinsic at all orders—formally it would correspond to the case $k = -1$.

The condition that increments of order k have a zero mean is introduced for a simpler presentation and does not restrict generality. If these increments

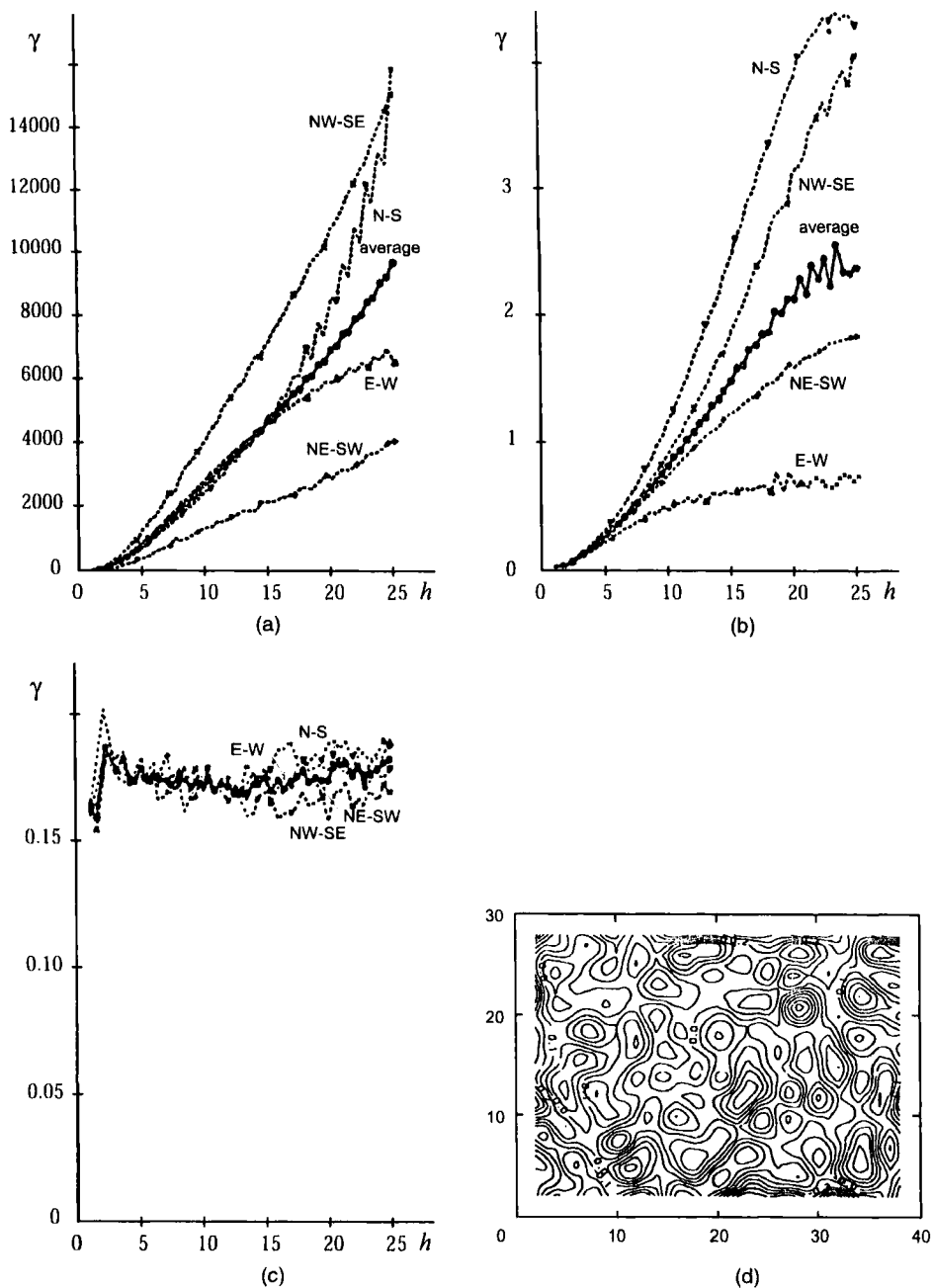


FIGURE 4.7. A case where increments of order 2 are needed to make data stationary. (a) raw directional variograms of grid displayed in Fig. 4.4c; (b) directional variograms of the increments of order 1; (c) directional variograms of the increments of order 2; (d) map of increments of order 2.

are stationary, their mean is necessarily a polynomial of degree $k + 1$ at most (Matheron, 1973a)¹ which is eliminated by regarding $Z(x)$ as an IRF $-(k + 1)$. For example, we have seen in Section 2.1.2 that the mean of stationary increments is necessarily of the form

$$E[Z(x + h) - Z(x)] = \langle a, h \rangle$$

If $a \neq 0$ we should regard $Z(x)$ as an IRF -1 rather than as an IRF -0 .

By definition an IRF satisfies $EZ_\lambda(x)^2 < \infty$ for any $\lambda \in \Lambda_k$. But $EZ(x)^2$ may be infinite, or at least depend on x ; in the introductory coin tossing example we had $ES_n^2 = n$. Stationarity of increments allows for both nonstationarity in the mean and in the variance.

As usual with random functions, it will be assumed that $Z(x)$ is continuous in the mean square sense. This property is mathematically essential to extend the theory from the space Λ_k of discrete measures to the space \mathcal{M}_k of measures with compact supports. If a discontinuity were present (nugget effect) it should be handled separately.

Examples of IRF $-k$

1. For $n = 1$ the integral of a zero-mean SRF is an IRF -0 . Indeed let $X(x)$ be the SRF in \mathbb{R}^1 and $Y_0(x)$ the integral

$$Y_0(x) = \int_0^x X(t) dt$$

For any h the increment

$$Y_0(x + h) - Y_0(x) = \int_x^{x+h} X(t) dt$$

is a moving average of the stationary process $X(t)$ and is therefore stationary.

2. The $(k + 1)$ -th integral of a zero-mean SRF is an IRF $-k$. We show this by induction. Assume that Y_k is an IRF $-k$, and consider the integral

$$Y_{k+1}(x) = \int_0^x Y_k(t) dt$$

Introducing the indicator function of the summation interval, we can write equivalently

$$Y_{k+1}(x) = \int Y_k(t) 1_{0 \leq t \leq x} dt$$

Thus for any discrete measure λ we have

$$\begin{aligned} Y_{k+1}(\lambda) &= \sum_i \lambda_i Y_{k+1}(x_i) \\ &= \int Y_k(t) \left(\sum_i \lambda_i 1_{0 \leq t \leq x_i} \right) dt \\ &= \int Y_k(t) \mu(dt) = Y_k(\mu) \end{aligned}$$

where

$$\mu(dt) = \left(\sum_i \lambda_i 1_{0 \leq t \leq x_i} \right) dt$$

Now $\lambda \in \Lambda_{k+1}$ implies that $\mu \in \Lambda_k$ (or more precisely $\in \mathcal{M}_k$), since

$$\int t^l \mu(dt) = \sum_i \lambda_i \int_0^{x_i} t^l dt = \frac{1}{l+1} \sum_i \lambda_i x_i^{l+1} = 0 \quad l = 0, \dots, k$$

For any x the translate $\tau_x \mu$ of the measure μ is defined by

$$\begin{aligned} Y_k(\tau_x \mu) &= \sum_i \lambda_i \int_x^{x_i+x} Y_k(t) dt \\ &= \sum_i \lambda_i [Y_{k+1}(x_i + x) - Y_{k+1}(x)] \\ &= \sum_i \lambda_i Y_{k+1}(x_i + x) \end{aligned}$$

the last equality being a consequence of $\sum \lambda_i = 0$. Finally we have

$$Y_k(\tau_x \mu) = Y_{k+1}(\tau_x \lambda)$$

Since Y_k is an IRF- k and $\mu \in \Lambda_k$, $Y_k(\tau_x \mu)$ is stationary in x . Therefore $Y_{k+1}(\tau_x \lambda)$ is also stationary in x , and this is true for any $\lambda \in \Lambda_{k+1}$. Consequently Y_{k+1} is an IRF- $(k+1)$.

Conversely, if an IRF- k is differentiable $(k + 1)$ times, its $(k + 1)$ -th derivative is stationary, being the limit of an ALC- k .

3. The same results hold if we start with an IRF-0. By integrating k times a Brownian motion $W_0(x)$, we obtain the IRF- k :

$$W_k(x) = \int_0^x \frac{(x-t)^{k-1}}{(k-1)!} W_0(t) dt$$

where $W_k(x)$ vanishes at $x = 0$ as well as its first $(k - 1)$ derivatives (just like Y_k above).

4. Conversely, now in \mathbb{R}^n , if a random function $Z(x)$ is differentiable $(k + 1)$ times and if all its partial derivatives of order $(k + 1)$ are stationary and with zero mean, $Z(x)$ is an IRF- k . This property characterizes differentiable IRF- k . Of course there exist nondifferentiable IRF- k (e.g., an IRF-0 with a linear variogram), but it is shown below that any continuous IRF- k is the sum of an SRF and a differentiable IRF- k .

5. An ARIMA process (autoregressive integrated moving average process) is defined as a process whose finite difference of order d is a stationary ARMA process (Box and Jenkins, 1976; see Section 7.5.1 for a definition of ARMA models). Since a finite difference of order d is an ALC- $(d - 1)$, an ARIMA process is an IRF- $(d - 1)$. However, the ARIMA and IRF- k approaches differ in the following aspects:

- ARIMA models are completely specified, whereas IRFs are only second-order models.
- ARIMA models are one-dimensional, whereas IRFs are defined in \mathbb{R}^n .
- ARIMA models are essentially discrete, whereas IRFs are continuous or discrete.

4.4.3. Abstract IRF and Its Representations

If $Z(x)$ is an IRF- k and A_l random variables—-independent or not of $Z(x)$ —the new random function

$$Z_1(x) = Z(x) + \sum_{|l| \leq k} A_l x^l$$

is also an IRF- k : by definition, $\lambda \in \Lambda_k$ cancels all monomials x^l , and thus

$$Z_1(\lambda) = Z(\lambda)$$

$Z_1(x)$ and $Z(x)$ are indistinguishable on the basis of ALC- k only.

In reality the concept of IRF- k relates to an *equivalence class* rather than a single function, namely the class of all random functions generating the same increments of order k . This motivated

Matheron's definition of the IRF- k as a family of increments. To avoid any confusion, we will call this *an abstract IRF- k* and denote it with a tilde.

Definition 2. An abstract intrinsic random function \tilde{Z} is a linear mapping of Λ_k into a Hilbert space H of zero mean, finite variance random variables, such that for any $\lambda \in \Lambda_k$ the random function $\tilde{Z}(\tau_\lambda \lambda)$ is second-order stationary in x :

$$\tilde{Z} : \Lambda_k \rightarrow H$$

such that

$$Z_\lambda(x) = \tilde{Z}(\tau_\lambda \lambda) \quad \text{is an SRF} \quad \forall \lambda \in \Lambda_k$$

An ordinary IRF- k is a random function $Z(x)$ in the usual sense, whereas the abstract IRF \tilde{Z} is not a function of x but of λ . Any ordinary IRF- k $Y(x)$ generating the same increments as \tilde{Z} , namely satisfying

$$Y(\lambda) = \tilde{Z}(\lambda) \quad \forall \lambda \in \Lambda_k$$

is called a *representation* of \tilde{Z} . So, from this new perspective, what we defined as an ordinary IRF- k was in fact a representation of the abstract IRF \tilde{Z} . But it is simpler to reason with representations because they "materialize" the equivalence class defined by the abstract IRF. The following two properties make it possible to identify \tilde{Z} with the class of all its representations:

1. Any abstract IRF- k \tilde{Z} has representations.
2. If one representation is known, all the others are deduced by addition of a polynomial of degree k with random coefficients.

Proof. It was shown in Section 4.3.2 that errors from linear unbiased estimation are ALC- k . We will use this property to construct a representation that has the structure of a residual. To this end consider a collection of measures $\lambda_l(dx)$ satisfying for all $|l| \leq k$ the conditions

$$\int \lambda_l(dx) x^s = \delta_l^s$$

($\delta_l^s = 1$ if $l = s$ and $= 0$ otherwise). One may think of these measures as defining unbiased estimators of the coefficients of a polynomial of degree k in the space \mathbb{R}^n considered. For any $x \in \mathbb{R}^n$ the measure

$$\varepsilon_x(dt) = \delta_x(dt) - \sum_l x^l \lambda_l(dt) \quad (4.11)$$

belongs to Λ_k , since

$$\int \varepsilon_x(dt) t^s = \int \delta_x(dt) t^s - \sum_l x^l \int \lambda_l(dt) t^s = x^s - \sum_l x^l \delta_l^s = 0$$

Now we claim that the RF $Y(x)$ defined by

$$Y(x) = \tilde{Z}(\varepsilon_x)$$

is a representation of \tilde{Z} . Indeed, for any $\lambda = \sum \lambda_i \delta_{x_i} \in \Lambda_k$ the linearity of \tilde{Z} entails

$$Y(\lambda) = \sum_i \lambda_i Y(x_i) = \sum_i \lambda_i \tilde{Z}(\varepsilon_{x_i}) = \tilde{Z} \left(\sum_i \lambda_i \varepsilon_{x_i} \right)$$

But

$$\sum_i \lambda_i \varepsilon_{x_i} = \sum_i \lambda_i \delta_{x_i} - \sum_l \left(\sum_i \lambda_i x_i^l \right) \lambda_l = \lambda$$

the last equality being a consequence of $\lambda \in \Lambda_k$ (the coefficients of λ_l are all zeros). Thus we have

$$Y(\lambda) = \tilde{Z}(\lambda) \quad \forall \lambda \in \Lambda_k$$

which proves point 1. In particular, for $\lambda = \varepsilon_x$ we obtain $Y(\varepsilon_x) = Y(x)$.

For point 2 suppose that $Z(x)$ is another representation of \tilde{Z} . We have

$$Z(\lambda) - Y(\lambda) = 0 \quad \forall \lambda \in \Lambda_k$$

So $Z_0(x) = Z(x) - Y(x)$ is a representation of the identically nil abstract IRF- k

$$\tilde{Z}_0(\lambda) = 0 \quad \forall \lambda \in \Lambda_k$$

In particular, with the measure ε_x defined in (4.11),

$$Z_0(\varepsilon_x) = Z_0(x) - \sum_l Z_0(\lambda_l) x^l = 0$$

and thus $Z_0(x)$ is of the form

$$Z_0(x) = \sum_l A_l x^l$$

□

With this presentation we have three levels of abstraction instead of two as usual: the abstract IRF- k \tilde{Z} , its representations $Y(x)$, which are random functions, and numerical realizations of $Y(x)$. We regard our regionalized variable as a realization of a representation of an abstract IRF \tilde{Z} . This allows us to distinguish two kinds of properties:

- Those that do not depend on the representation: they are *intrinsic properties*.
- Those that depend on the representation.

The estimation of intrinsic properties only requires the specification of the abstract IRF- k model. To estimate other properties, it is necessary to qualify the representation being considered.

Internal Representations

The particular representation $Y(x) = Y(\varepsilon_x)$ that we have just constructed has the remarkable property of being itself an ALC- k . Therefore $Y(x)$ satisfies $EY(x) = 0$ and $EY(x)^2 < \infty$.

To appreciate the specificity of $Y(x)$, consider another representation $Y_1(x)$. We have

$$Y(x) = Y_1(\epsilon_x) = Y_1(x) - \sum_l x^l \int \lambda_l(dy) Y_1(y)$$

We see that $Y(x)$ is an *additive renormalization* of $Y_1(x)$ obtained by subtracting a polynomial of degree k whose coefficients $A_l = \int \lambda_l(dy) Y_1(y)$ are linear functionals² of $Y_1(x)$. A simple and standard example for an IRF-0 is the representation $Y(x) = Y_1(x) - Y_1(0)$. By construction, $Y(x)$ has a finite variance, while $Y_1(x)$ may not. Now, considering a compact domain D (with a nonempty interior), if all measures λ_l have their support included in D , the representation $Y(x)$ only depends on values of $Y_1(x)$ within D . We say that $Y(x)$ is an *internal representation* over D . This notion is useful for the solution of partial differential equations of the type $\Delta Z(x) = Y(x)$ (see Dong, 1990).

An internal representation is generally not stationary although it is sometimes possible to make it locally stationary by an astute choice of the measures λ_l . We will revisit this subject later.

General Form of Representations

From the above we can conclude that all representations are the sum of an internal representation and an arbitrary polynomial of degree k at most

$$Y(x) = Y\left(\delta_x - \sum_l x^l \lambda_l\right) + \sum_l A_l x^l \quad (4.12)$$

From now on we will leave the abstract IRF- k \tilde{Z} in the background and, following the simpler definition, will refer to $Z(x)$ and an IRF- k (instead of a representation of an abstract IRF- k).

4.5. GENERALIZED COVARIANCE FUNCTIONS

The correlation structure of a stationary random function $Z(x)$ is defined by its ordinary covariance function $C(h)$. If only the (ordinary) increments $Z(x+h) - Z(x)$ of the function are assumed stationary, the variogram $\gamma(h)$ is the structural tool. We saw that $\gamma(h)$ only allows the calculation of the variances of linear combinations whose sum of weights is zero. In the same manner, when the stationarity assumptions are limited to generalized increments of order k , what characterizes the correlation structure of $Z(x)$ is a new function called a generalized covariance function, abbreviated as GC and denoted by $K(h)$. Just as there were more models for variograms than for covariances, there are more models for generalized covariances than for both ordinary covariances and variograms. But there are also more restrictions attached to their use in variance calculations: they only work on allowable linear combinations. Figure 4.8 shows a family picture of C , γ , and K taken by J. P. Delhomme.

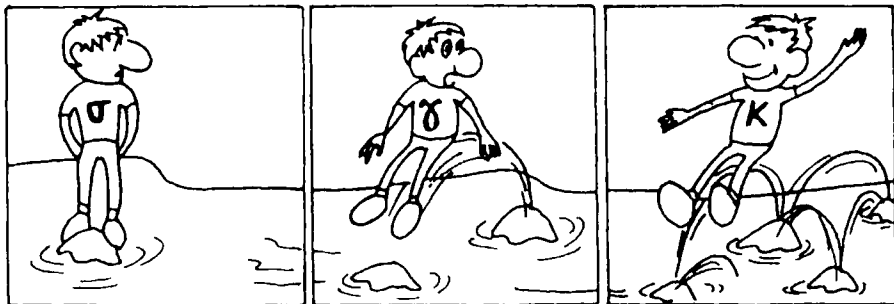


FIGURE 4.8. Covariance, variogram, and generalized covariance: a family picture. Author: J. P. Delhomme.

4.5.1. Existence and Uniqueness

Definition. Let Z be an IRF- k . A symmetric function $K(h)$ defined on \mathbb{R}^n is called a generalized covariance (GC) of Z if

$$E[Z(\lambda)Z(\mu)] = \sum_i \sum_j \lambda_i \mu_j K(y_j - x_i) \quad (4.13)$$

for any pair of measures $\lambda, \mu \in \Lambda_k$.

In fact it is sufficient that the above condition be satisfied in the case $\lambda = \mu$ (it is seen easily by expanding $[Z(\lambda + \mu)]^2$). In other words, it suffices to verify that

$$EZ(\lambda)^2 = \sum_i \sum_j \lambda_i \lambda_j K(x_j - x_i) \quad (\lambda \in \Lambda_k) \quad (4.14)$$

Formally $K(h)$ is used just as an ordinary covariance $C(h)$, but (4.14) only holds for $\lambda \in \Lambda_k$.

There is an existence and uniqueness theorem for any continuous IRF- k (i.e., whose representations are continuous in the mean square sense).

Theorem. Any continuous IRF- k has a continuous GC $K(h)$. This GC $K(h)$ is unique as an equivalence class, in the sense that any other GC is of the form $K(h) + Q(h)$, where $Q(h)$ is an even polynomial of degree $2k$ or less.

This theorem holds in \mathbb{R}^n whatever the dimension n , and for any IRF- k . But the proof in the general case encounters technical difficulties which obscure the almost intuitive aspect of the result (Gel'fand and Vilenkin, 1961; Matheron, 1973a). Therefore we will restrict ourselves here to the simple case $n = 1$ and Z differentiable.

Proof of the theorem for a differentiable Z in \mathbb{R}^1 .

- *Existence*

Let then $Z(x)$ be an IRF- k in \mathbb{R}^1 , differentiable $k+1$ times (in the mean square). Its $(k+1)$ -th derivative $Z^{(k+1)}(x)$ is stationary and has a stationary covariance $C(h)$. If we denote by $\sigma(x, y)$ the nonstationary covariance of $Z(x)$, it is related to $C(h)$ by

$$\frac{\partial^{2(k+1)}}{\partial x^{k+1} \partial y^{k+1}} \sigma(x, y) = C(y - x)$$

Integrating $(k+1)$ times in x gives

$$\frac{\partial^{k+1}}{\partial y^{k+1}} \sigma(x, y) = (-1)^{k+1} \int_0^{y-x} \frac{(y-x-u)^k}{k!} C(u) du + \sum_{l=0}^k a_l(y) x^l$$

Integrating now $(k+1)$ times in y for fixed x gives

$$\sigma(x, y) = (-1)^{k+1} \int_0^{y-x} \frac{(y-x-u)^{2k+1}}{(2k+1)!} C(u) du + \sum_{l=0}^k b_l(y) x^l + \sum_{l=0}^k c_l(x) y^l$$

where $b_l(y)$ is a definite integral of order $k+1$ of $a_l(y)$. Let

$$K(h) = (-1)^{k+1} \int_0^h \frac{(h-u)^{2k+1}}{(2k+1)!} C(u) du$$

Verify that $K(h) = K(-h)$ so that $K(h)$ is a symmetric function just as $C(h)$, and since $\sigma(x, y)$ is symmetric in x and y , the $b_l(\cdot)$ and $c_l(\cdot)$ functions are identical. Thus the general form of the covariance $\sigma(x, y)$ is

$$\sigma(x, y) = K(y - x) + \sum_{l=0}^k c_l(y) x^l + \sum_{l=0}^k c_l(x) y^l \quad (4.15)$$

If λ is an allowable linear combination

$$\sum_i \lambda_i x_i^l = 0 \quad l = 0, \dots, k$$

and from (4.15),

$$\sum_i \sum_j \lambda_i \lambda_j \sigma(x_i, x_j) = \sum_i \sum_j \lambda_i \lambda_j K(x_j - x_i)$$

which proves that $K(h)$ is a generalized covariance.

- *Uniqueness*

Let us assume that $Z(x)$ has two distinct GCs, $K_1(h)$ and $K_2(h)$. From (4.15) the difference $K_0 = K_1 - K_2$ is of the form

$$K_0(y - x) = \sum_{l=0}^k c_l(y) x^l + \sum_{l=0}^k c_l(x) y^l$$

Because $Z(x)$ is differentiable $(k + 1)$ times, $K_0(h)$ is differentiable $2(k + 1)$ times. Letting $h = y - x$, we have

$$\frac{d^{2k+2}}{dh^{2k+2}} K_0(h) = (-1)^{k+1} \frac{\partial^{k+1}}{\partial x^{k+1}} \frac{\partial^{k+1}}{\partial y^{k+1}} K_0(y - x)$$

But this is zero identically, since the x^l and y^l are monomials of degree strictly less than $(k + 1)$. So $K_0(h)$ is a polynomial of degree $2k + 1$ at most. Because $K_0(h)$ is a symmetric function, it is necessarily an even polynomial of degree $\leq 2k$ (i.e., with even powers only). \square

Conversely for any even polynomial $Q(h)$ of degree $\leq 2k$, $K(h) + Q(h)$ is indeed a GC. It suffices to note that $Q(h)$ can be written just as $K_0(y - x)$, where $c_l(\cdot)$ are polynomials of degree up to $2k$.

Examples of GCs

1. If $Z(x)$ is an SRF with ordinary stationary covariance $C(h)$, then $C(h)$ is clearly also a GC. If $Z(x)$ is an IRF-0, then

$$K(h) = -\gamma(h) + \text{constant}$$

The GC is equal to $-\gamma$ up to an arbitrary constant (which is an even polynomial of degree $k = 0$).

2. We have seen earlier that if $X(t)$ is a zero-mean SRF on \mathbb{R}^1 its $(k + 1)$ -th integral

$$Y_k(x) = \int_0^x \frac{(x-t)^k}{k!} X(t) dt$$

is an IRF- k . The proof of the existence theorem given above shows that the GC of $Y_k(x)$ is

$$K_k(h) = (-1)^{k+1} \int_0^h \frac{(h-u)^{2k+1}}{(2k+1)!} C(u) du \quad (h > 0)$$

where $C(h)$ is the stationary covariance of $X(t)$.

4.5.2. Link between Generalized and Ordinary Covariance Functions

The covariance and the variogram are easy to understand because they are directly related to the variable $Z(x)$. The GC is more abstract. To “materialize” it, let us see how it relates to the ordinary covariance. To this end, we consider

the general form of representations (4.12),

$$Z(x) = Z \left(\delta_x - \sum_l x^l \lambda_l \right) + \sum_l A_l x^l$$

and calculate the centered covariance $\sigma(x, y)$. The first term being of the form $Z(\varepsilon_x)$ with $\varepsilon_x \in \Lambda_k$, its covariance can be calculated with the GC $K(h)$ using the general formula (4.13):

$$\begin{aligned} E[Z(\varepsilon_x)Z(\varepsilon_y)] &= \iint \left(\delta_x(dt) - \sum_l x^l \lambda_l(dt) \right) K(t' - t) \left(\delta_y(dt') - \sum_s y^s \lambda_s(dt') \right) \\ &= K(y - x) - \sum_l x^l \int K(y - t) \lambda_l(dt) - \sum_l y^l \int K(t' - x) \lambda_l(dt') \\ &\quad + \sum_l \sum_s x^l y^s \iint \lambda_l(dt) K(t' - t) \lambda_s(dt') \end{aligned} \quad (4.16)$$

Now the covariance of $Z(x)$ is

$$\begin{aligned} \text{Cov}(Z(x), Z(y)) &= E[Z(\varepsilon_x)Z(\varepsilon_y)] + \sum_l x^l \text{Cov}(A_l, Z(\varepsilon_y)) \\ &\quad + \sum_l y^l \text{Cov}(Z(\varepsilon_x), A_l) + \sum_l \sum_s x^l y^s \text{Cov}(A_l, A_s) \end{aligned}$$

The covariance $\text{Cov}(A_l, \cdot)$ cannot be calculated with the GC because A_l is not an ALC- k . Anyhow, collecting the coefficients of x^l and y^s with those in (4.16), we get a covariance of the form

$$\sigma(x, y) = K(y - x) + \sum_{|l|=0}^k c_l(y) x^l + \sum_{|l|=0}^k c_l(x) y^l + \sum_{|l|=0}^k \sum_{|s|=0}^k T_{ls} x^l y^s \quad (4.17)$$

This is the same formula as (4.15) obtained by integration of a differentiable GC (the terms $x^l y^s$ can be distributed equally between the x^l and y^s terms). It exposes the impact of the stationarity of increments of order k , an attenuated form of stationarity, on the covariance of $Z(x)$. There is a stationary part $K(y - x)$ and a nonstationary part involving polynomial terms separately in x and y . In general, the functions $c_l(x)$ are not polynomials, nor can they be determined from the data on the basis of a single realization because they involve nonstationary features of $Z(x)$. If we consider an arbitrary linear combination $Z(\lambda)$ its variance depends on the $c_l(x)$ and cannot be evaluated. But, if $\lambda \in \Lambda_k$, then these coefficients are filtered out.

Examples

1. The simplest illustration is with a GC-0 $K(h) = -\gamma(h)$. Then

$$\sigma(x, y) = -\gamma(y - x) + \frac{\sigma(x, x) + \sigma(y, y)}{2}$$

which is of the form (4.17) with $k = 0$ and $c_0(x) = \sigma(x, x)/2$.

To take a specific example, consider in 1D a Brownian motion $X(t)$ without drift and the representation $Y(t) = X(t) - X(0)$. This is a nonstationary RF with mean zero and a variance proportional to t . We assume a scaling such that the variogram of X has a slope of 1. Then

$$\sigma(t, t') = -|t' - t| + |t| + |t'|$$

When both t and $t' > 0$ this takes the classic form $\sigma(t, t') = 2 \min(t, t')$.

2. Now suppose that $X(t)$ has a linear drift. We then model it as an IRF-1 and consider the representation defined by

$$Y(t) = X(t) - X(0) - \frac{1}{2R}[X(R) - X(-R)]t$$

which by construction satisfies $EY(t) = 0$. It is therefore an ALC-1 of X and has a finite variance that can be calculated with the GC $K(h) = -|h|$. The nonstationary covariance of $Y(t)$ is found to be

$$\begin{aligned} \sigma(t, t') &= -|t' - t| + |t| + |t'| + \frac{1}{2R}(|R - t'| - |R + t'|)t \\ &\quad + \frac{1}{2R}(|R - t| - |R + t|)t' + \frac{1}{R}tt' \end{aligned}$$

which is of the form (4.17). It is interesting to note that for t and $t' \in [-R, +R]$, the raw variogram of $Y(t)$ is

$$\frac{1}{2}E[Y(t') - Y(t)]^2 = |t' - t| - \frac{1}{2R}(t' - t)^2$$

and depends only on $|t' - t|$, a result already encountered in Section 2.7.1 (due to a different scaling of X the variogram slope there was $\frac{1}{2}$ instead of 1).

4.5.3. Spectral Theory

The existence and uniqueness theorem for GCs is a direct consequence of Gel'fand and Vilenkin's theory of generalized random fields (1961, ch. 3, sec. 5.2). The result gives the general form of the correlation functional $B(\varphi, \psi)$ of a generalized random field acting on functions φ, ψ of the class of infinitely differentiable functions vanishing outside a compact set. This correlation functional is characterized by a "slowly growing measure" satisfying certain requirements. Matheron (1971b)

established this result directly for the case of random functions (as opposed to random generalized functions).

The spectral theory of ordinary processes with stationary increments of arbitrary order (case $n = 1$) was established by Yaglom and Pinsker (1953). A comprehensive presentation of the theory is also given by Yaglom (1987, vol. 1, ch. 4).

The class of GCs coincides with the class of continuous and symmetric functions $K(h)$ on \mathbb{R}^n satisfying

$$\sum_i \sum_j \lambda_i \lambda_j K(x_j - x_i) \geq 0 \quad (4.18)$$

for any real allowable measure $\lambda \in \Lambda_k$. This condition ensures that $EZ(\lambda)^2 \geq 0$. A real function satisfying (4.18) is said to be *k-conditionally positive definite*. Such functions are characterized by a certain spectral representation, just as ordinary covariances in the Bochner-Khinchin theory. The formula looks awesome but it provides insight into the physical significance of IRFs.

Theorem. A continuous and symmetric function $K(h)$ on \mathbb{R}^n is a GC of an IRF- k if and only if it is of the form

$$K(h) = \int \frac{\cos(2\pi\langle u, h \rangle) - 1_B(u)P_k(2\pi\langle u, h \rangle)}{(4\pi^2|u|^2)^{k+1}} \chi(du) + Q(h) \quad (4.19)$$

where $P_k(x) = 1 - x^2/2 + \dots + (-1)^k x^{2k}/(2k)!$, $1_B(u)$ is the indicator function of an arbitrary neighborhood of $u = 0$, and $\chi(du)$ is a positive symmetric measure, with no atom at the origin and satisfying

$$\int \frac{\chi(du)}{(1 + 4\pi^2|u|^2)^{k+1}} < \infty \quad (4.20)$$

$Q(h)$ is an arbitrary even polynomial of degree $\leq 2k$.

$u = (u_1, \dots, u_n)$	denotes a frequency vector
$ u ^2 = u_1^2 + \dots + u_n^2$	denotes its squared modulus
$\langle u, h \rangle = u_1 h_1 + \dots + u_n h_n$	denotes a scalar product

To understand this formula, first note that the term $1_B(u)P_k(2\pi\langle u, h \rangle)$ under the integral represents the expansion at the order k of $\cos(2\pi\langle u, h \rangle)$ in the neighborhood B of $u = 0$. It is an even polynomial of degree $2k$ in the argument h and represents exactly what must be subtracted from the cosine to make the integral converge at $u = 0$ given (4.20) (the difference is of the order $|u|^{2k+2}$). The value of the integrand is not defined at $u = 0$, but since $\chi(du)$ has no atom there, the integral converges.

The neighborhood B is arbitrary: if B_1 and B_2 are two different neighborhoods of $u = 0$ the difference between the associated $K_1(h)$ and $K_2(h)$ is an even polynomial of degree $2k$ in h and may thus be incorporated in the arbitrary $Q(h)$ polynomial.³

When $K(h)$ is differentiable $2k$ times, the term $1_B(u)$ in formula (4.19) can be dropped (i.e., replaced by 1). Thus for $k = 0$ this term is never needed and formula (4.19) is seen to coincide with the spectral representation (2.21) of $-\gamma(h)$.

Formula (4.19) associates to each $K(h)$ the spectral measure

$$F(du) = \frac{\chi(du)}{(4\pi^2|u|^2)^{k+1}}$$

defined on the space $\mathbb{R}^n - \{0\}$. This measure is the same for all GCs in the equivalence class and thus appears as the fundamental information on the correlation structure of an IRF- k . The synthetic condition (4.20) on the measure $\chi(du)$ is equivalent to the following two conditions on $F(du)$:

$$\int_{|u|<\varepsilon} |u|^{2k+2} F(du) < \infty \quad \int_{|u|>\varepsilon} F(du) < \infty \quad (4.21)$$

where $\varepsilon > 0$ is arbitrary. Unlike for ordinary covariances it is not required that the integral of the spectrum converges but only that near-zero frequency $\int_{|u|<\varepsilon} |u|^{2k+2} F(du)$ converges while $\int_{|u|>\varepsilon} F(du)$ must converge at infinity.

In the case of stationary processes, $F(du)$ is interpreted as the power in the frequency interval $(u, u + du)$ and the integral $\int F(du) = C(0)$ as the total power of the process. With IRF- k the integral $\int F(du)$ may become infinite because the first condition (4.21) allows $F(du)$ to tend rapidly to infinity as $u \rightarrow 0$: there may be an infinite power at low frequencies, a phenomenon referred to as an *infrared catastrophe*. Such effect is observed with Brownian and fractional Brownian motions in one dimension (Mandelbrot, 1967, 1982, p. 389)—the spectral measure is then proportional to $du/u^{1+\alpha}$ ($0 < \alpha < 2$).

It is this high power at low frequencies that is responsible for the apparent long-term “trends.” It also explains why the restriction to allowable measures is necessary. Indeed consider $\lambda \in \Lambda_k$, and denote by

$$\tilde{\lambda}(u) = \int \exp(-2\pi i \langle u, x \rangle) \lambda(dx)$$

the Fourier transform of λ (i is the unit pure imaginary number). From (4.19) we obtain

$$E[Z(\lambda)]^2 = \int \frac{|\tilde{\lambda}(u)|^2}{(4\pi^2 |u|^2)^{k+1}} \chi(du)$$

Since λ has a compact support, the function $\tilde{\lambda}(u)$ is infinitely differentiable, and $\lambda \in \Lambda_k$ implies that $\tilde{\lambda}(u)$ and its first k derivatives vanish at $u = 0$. This ensures the convergence of the above integral given the fact that $\chi(du)$ itself is integrable near 0 (first condition 4.21). In other words, the Fourier transform $\tilde{\lambda}(u)$ neutralizes catastrophes.

In the case that $\int F(du)$ converges, the integral in (4.19) may be written as the difference between two convergent integrals, and since the second integral is a polynomial of degree $2k$ in h , $K(h)$ finally takes the form

$$K(h) = C(h) + Q(h)$$

where $C(h)$ is an ordinary covariance function. This is sufficient to assert that the IRF- k $Z(x)$ possesses a stationary representation $Y_{S_1}(x)$ whose covariance is $C(h)$. From Section 4.4.3 we then know that $Z(x)$ differs from $Y_{S_1}(x)$ by a random polynomial of degree k : this is the universal kriging model.

Another special case is when $\int \chi(du) < \infty$. Then $K(h)$ is differentiable $(2k+2)$ times, which means that $Z(x)$ is differentiable $(k+1)$ times in the sense that all its partial derivatives of order $k+1$,

$$\frac{\partial^{k+1} Z(x)}{\partial^{l_1} x_1 \cdots \partial^{l_n} x_n} \quad l_1 + \cdots + l_n = k+1$$

exist and are stationary. Differentiating (4.19) under the integral sign (which is valid) yields the equation

$$\Delta^{k+1} K(h) = (-1)^{k+1} C(h) \quad (4.22)$$

where Δ^{k+1} is the iterated Laplacian operator and $C(h)$ is the stationary covariance associated with the spectral measure $\chi(du)$.

Finally let us note the following decomposition of the integral in (4.19),

$$K(h) = \int_{|u| \leq u_0} \frac{\cos(2\pi\langle u, h \rangle) - P_k(2\pi\langle u, h \rangle)}{(4\pi^2|u|^2)^{k+1}} \chi(du) + \int_{|u| > u_0} \cos(2\pi\langle u, h \rangle) \frac{\chi(du)}{(4\pi^2|u|^2)^{k+1}}$$

The integral over $|u| \leq u_0$ corresponds to an IRF- k with no high frequencies, an infinitely differentiable IRF- k , while the integral over $|u| > u_0$ is an ordinary stationary covariance function $C(h)$ —thanks to the second relation (4.21). We can write the phenomenological “equation”:

$$\text{continuous IRF-}k = \text{infinitely differentiable IRF-}k + \text{stationary random function}$$

In principle, this dichotomy into low and high frequencies could be used as a definition of a “drift + residual” model. However, the arbitrariness of the cutoff frequency u_0 highlights once again the elusive character of the notion of drift. In practice it is also nearly impossible to estimate these two components. This decomposition is thus mainly of theoretical interest.

4.5.4. Majorization of Generalized Covariances

From the spectral formula (4.19) and the majorization $|\cos x - P_k(x)| \leq x^{2k+2}/(2k+2)!$ it is seen readily that a GC- k must satisfy the following inequality:

$$|K(0) - K(h)| \leq a + b|h|^{2k+2} \quad \forall h \in \mathbb{R}^n \quad (4.23)$$

for some positive constants a and b . Likewise an IRF- k is differentiable (once) if and only if its GCs satisfy an inequality of the form ($a', b' \geq 0$),

$$|K(0) - K(h)| \leq a'|h|^2 + b'|h|^{2k+2} \quad \forall h \in \mathbb{R}^n$$

The following two results are proved in Matheron's 1973a paper:⁴

•
$$\lim_{|h| \rightarrow \infty} \frac{K(h)}{|h|^{2k+2}} = 0 \quad (4.24)$$

- An IRF- k is the restriction to Λ_k of a stationary random function if and only if one of its GCs is bounded on \mathbb{R}^n . When $k = 0$, this means that the variogram must be bounded.

4.5.5. The $|h|^\alpha$ Model

As the order k increases, the class of IRF- k expands, and so does the class of GC models. For example, when $\alpha > 0$,

$$\gamma(h) = |h|^\alpha$$

is a variogram (i.e., $-\gamma$ is a GC-0) whenever $0 < \alpha < 2$. Likewise when

$$0 < \alpha < 2k + 2 \quad \alpha \text{ noneven}$$

the function

$$K(h) = \Gamma\left(-\frac{\alpha}{2}\right) |h|^\alpha$$

is a GC of order k . $\Gamma(\cdot)$ denotes the usual Euler gamma function (A.1). The sign of the coefficient $\Gamma(-\alpha/2)$ alternates with the value of α : it is negative for $0 < \alpha < 2$, positive for $2 < \alpha < 4$, negative for $4 < \alpha < 6$, and so on. When α is an even integer, $\Gamma(-\alpha/2)$ is not defined, but in this case $K(h)$ is an even degree polynomial and thus equivalent (in the same GC class as) the identically nil covariance $K(h) \equiv 0$.

It is interesting to relate this result to the general spectral representation conditions (4.21). In \mathbb{R}^n we can start from the relationship

$$A_{-\alpha} |h|^\alpha = - \int_{\mathbb{R}^n} [1 - \cos(2\pi \langle u, h \rangle)] A_{\alpha+n} |u|^{-\alpha-n} du$$

where $A_t = \Gamma(t/2)/\pi^{t/2}$. Thus the spectral measure $F(du) \propto |u|^{-\alpha-n}$. Applying (4.21) after a change to polar coordinates in which $\rho = |u|$ and $du = \rho^{n-1} d\rho$ (up to a multiplicative factor), we find the conditions

$$\int_{\rho < \varepsilon} \rho^{2k+1-\alpha} d\rho < \infty \quad \int_{\rho > \varepsilon} \rho^{-\alpha-1} d\rho < \infty$$

These conditions are satisfied if and only if $0 < \alpha < 2k + 2$.

4.5.6. Polynomial Models

The interesting elements of the $|h|^\alpha$ family are the terms $\alpha = 2p + 1$. Then $(-1)^{p+1} |h|^{2p+1}$ is a GC- k provided that $p \leq k$. More generally, the function

$$K(h) = \sum_{p=0}^k (-1)^{p+1} b_p |h|^{2p+1} \quad (4.25)$$

is a GC of order k under conditions on the coefficients b_p , which are obviously satisfied if $b_p \geq 0 \forall p$.

Covariances of this form are called polynomial GCs. This is a slight misnomer because the functions are polynomials with respect to the modulus $|h| = (h_1^2 + \dots + h_n^2)^{1/2}$ of the vector h and not with respect to the components h_1, \dots, h_n of this vector. Since they depend only on the modulus of h polynomial GCs are *isotropic* models. (However, the techniques presented to deal with variogram anisotropies can be used here as well.)

The physical meaning of the exponent α is this: the higher α , the more regular the model at a local scale and the more fluctuating at a global scale. A phenomenon with a linear covariance $-b_0|h|$ is continuous but not differentiable, with $b_1|h|^3$ it is differentiable once but also has very large fluctuations, and with $-b_2|h|^5$ it is differentiable twice but fluctuates wildly at large distances. Figure 4.4 illustrates the aspects of phenomena associated with the pure terms $-|h|$, $|h|^3$, $-|h|^5$. The last one represents a slowly varying twice differentiable component that in the universal kriging terminology would be called a *drift*.

The exact conditions to be placed on the coefficients b_p in (4.25) for $K(h)$ to be a valid GC- k in \mathbb{R}^n are obtained by requiring that the measure $\chi(du)$ in the general spectral representation (4.19) is positive, leading to

$$\sum_{p=0}^k \frac{(2p+1)!}{p!} \Gamma[p + \frac{1}{2}(n+1)] b_p x^{k-p} \geq 0 \quad \text{for } x \geq 0$$

For $k = 0, 1$, and 2 , and according to the space dimensionality n , the conditions are

$$k = 0: \quad b_0 \geq 0 \quad b_1 = 0 \quad b_2 = 0$$

$$k = 1: \quad b_0 \geq 0 \quad b_1 \geq 0 \quad b_2 = 0$$

$$k = 2: \quad b_0 \geq 0 \quad b_2 \geq 0 \quad b_1 \geq -\sqrt{\frac{20}{3} \left(1 + \frac{2}{n+1}\right)} \sqrt{b_0 b_2}$$

Note that the condition on b_1 becomes more severe as n increases, the lower bound being -3.651 , -3.333 , -3.162 for $n = 1, 2, 3$, respectively.

Table 4.1 summarizes the polynomial models for $k = 0, 1$, and 2 ; the nugget effect has been added to take into account microstructures or measurement

TABLE 4.1. Polynomial GC models for $k \leq 2$ with nugget effect added

Filtered Polynomial (Drift)	k	Polynomial Generalized Covariance Model + Nugget Effect
Constant	0	$K(h) = C_0 \delta(h) - b_0 h $
Linear	1	$K(h) = C_0 \delta(h) - b_0 h + b_1 h ^3$
Quadratic	2	$K(h) = C_0 \delta(h) - b_0 h + b_1 h ^3 - b_2 h ^5$
<i>Constraints</i>		
In \mathbb{R}^1	$C_0 \geq 0, b_0 \geq 0, b_2 \geq 0, b_1 \geq -2\sqrt{10/3} \sqrt{b_0 b_2}$	
In \mathbb{R}^2	$C_0 \geq 0, b_0 \geq 0, b_2 \geq 0, b_1 \geq -(10/3) \sqrt{b_0 b_2}$	
In \mathbb{R}^3	$C_0 \geq 0, b_0 \geq 0, b_2 \geq 0, b_1 \geq -\sqrt{10} \sqrt{b_0 b_2}$	

errors. The main advantage of these models for applications is that they depend linearly on their parameters, which facilitates their statistical inference.

Construction of an IRF- k with a Polynomial GC in \mathbb{R}^1

Successive integrations of a process $W_0(x)$ with GC $K_0(h) = -|h|$ (e.g., a Wiener-Lévy process) lead to a process

$$W_k(x) = \int_0^x \frac{(x-t)^{k-1}}{(k-1)!} W_0(t) dt$$

that is an IRF- k , with the polynomial GC

$$K_k(h) = \frac{(-1)^{k+1} |h|^{2k+1}}{(2k+1)!}$$

Now consider a process of the form

$$Y(x) = \sum_{p=0}^k c_p W_p(x) \quad (4.26)$$

where all $W_p(x)$ are integrals of the *same process* $W_0(x)$. Then $Y(x)$ can be expressed as a sum of derivatives of $W_k(x)$:

$$Y(x) = \sum_{p=0}^k c_p D^{k-p} W_k(x)$$

where D^p denotes derivation of order p , from which the GC of $Y(x)$ is found to be

$$K(h) = \left(\sum_{p=0}^k c_p D^{k-p} \right) \left(\sum_{q=0}^k (-1)^{k-q} c_q D^{k-q} \right) \frac{(-1)^{k+1} |h|^{2k+1}}{(2k+1)!}$$

It is a polynomial GC. Conversely, any IRF- k in 1D with a polynomial GC has a representation of the form (4.26), which means that for any given set of covariance coefficients b_p defined in (4.25), we can always find the matching set of coefficients c_p (Matheron, 1973a). For $k = 2$ the relationships are

$$b_0 = c_0^2 \quad b_1 = \frac{1}{6}(c_1^2 - 2c_0c_2) \quad b_2 = \frac{1}{120}c_2^2$$

This provides a simple algorithm for simulating in one dimension an IRF- k with a polynomial GC, starting from a simulation of an IRF-0 with a linear variogram, which is very easy to construct.

4.5.7. Spline Covariance Model

Another very important model is hidden in the family (4.25). This is the GC-1

$$K(h) = |h|^2 \log |h|$$

which is associated with biharmonic splines in 2D, as will be seen later. This covariance is the limit

$$K(h) = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} (|h|^{2+\alpha} - |h|^2)$$

Indeed $|h|^{2+\alpha}$ is a GC-1 if $0 < \alpha < 2$ and certainly for small α , and so is $|h|^{2+\alpha} - |h|^2$, since $|h|^2$ is an even polynomial of degree $2k = 2$. The limit of a GC- k is still a GC- k , and therefore $|h|^2 \log |h|$ is a GC-1. It is not a GC of lower order because $K(h)/|h|^2 = \log |h|$ is not bounded so that (4.24) is not satisfied. Likewise

$$K(h) = (-1)^{k+1} |h|^{2k} \log |h|$$

is a GC of order k and not of any lower order.

An $|h|^2 \log |h|$ term can thus be added to the polynomial covariance model (4.25). In doing so, it is preferable to drop the $-|h|^5$ covariance so as to keep the number of parameters as low as possible for better statistical inference. (The $-|h|^5$ covariance leads to poorly conditioned kriging matrices.) The resultant covariance with terms arranged by increasing regularity is

$$K(h) = C_0 \delta(h) - b_0 |h| + b_S |h|^2 \log |h| + b_1 |h|^3$$

$K(h)$ is a valid GC of order $k = 1$ —and of course of higher order as well—if and only if the coefficients satisfy the following inequalities (from Dubrule, 1981):

In \mathbb{R}^1 ,

$$C_0 \geq 0, \quad b_0 \geq 0, \quad b_1 \geq 0, \quad b_S \geq -\frac{\sqrt{24}}{\pi} \sqrt{b_0 b_1}$$

In \mathbb{R}^2 ,

$$C_0 \geq 0, \quad b_0 \geq 0, \quad b_1 \geq 0, \quad b_S \geq -\frac{3}{2} \sqrt{b_0 b_1}$$

In \mathbb{R}^3 ,

$$C_0 \geq 0, \quad b_0 \geq 0, \quad b_1 \geq 0, \quad b_S \geq -\frac{8}{\pi\sqrt{3}} \sqrt{b_0 b_1}$$

4.6. ESTIMATION IN THE IRF MODEL

4.6.1. Intrinsic Kriging

The IRF- k model was designed to extend the scope of kriging to nonstationary cases. It remains to show that it indeed does so. The rule of the game is to derive the equations using only allowable linear combinations, which are the only ones to have computable variances. This is done below in a straightforward manner. A geometric derivation of the equations in terms of projections in Hilbert spaces can be found in Matheron (1981c).

In our model the variable under study $Z(x)$ is regarded as a realization of an IRF- k of known order k and known GC $K(h)$ —the inference problem will be considered later. To keep things simple, let us suppose that we want to estimate the value $Z_0 = Z(x_0)$ at a point x_0 using a linear combination of observed data $Z(x_\alpha)$. By placing appropriate conditions on the weights, we can ensure that the *estimation error* $Z^* - Z_0$ is an allowable linear combination of order k . Then its variance can be expressed in terms of $K(h)$, and the kriging equations are derived exactly as in the universal kriging approach.

Specifically if

$$\sum_{\alpha} \lambda_{\alpha} x_{\alpha}^l - x_0^l = 0 \quad |l| = 0, \dots, k \quad (4.27)$$

then $Z_0^* - Z_0 = Z(\sum \lambda_{\alpha} \delta_{x_{\alpha}} - \delta_{x_0})$ is an ALC- k , and by formula (4.14) its variance is

$$E(Z^* - Z_0)^2 = \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} K(x_{\beta} - x_{\alpha}) - 2 \sum_{\alpha} \lambda_{\alpha} K(x_0 - x_{\alpha}) + K(0)$$

Minimizing this subject to (4.27) leads to the system

Intrinsic Kriging System

$$\begin{cases} \sum_{\beta=1}^N \lambda_{\beta} K(x_{\beta} - x_{\alpha}) + \sum_{|l|=0}^k \mu_l x_{\alpha}^l = K(x_0 - x_{\alpha}) & \alpha = 1, \dots, N \\ \sum_{\alpha=1}^N \lambda_{\alpha} x_{\alpha}^l = x_0^l & |l| = 0, \dots, k \end{cases} \quad (4.28)$$

The kriging variance is as usual

Intrinsic Kriging Variance

$$\sigma_K^2 = K(0) - \sum_{\alpha=1}^N \lambda_{\alpha} K(x_0 - x_{\alpha}) - \sum_{|l|=0}^k \mu_l x_0^l$$

This system, sometimes named “intrinsic kriging,” is exactly the same as in the universal kriging model (3.14) except that $K(h)$ is substituted for $C(h)$. The kriging estimator and the kriging variance only depend on the GC class and not on the particular version used—namely not on the particular representation. They are intrinsic properties.

The system (4.28) may easily be adapted to block estimation as is done for ordinary covariances in Section 3.5.2. Likewise filtering of random measurement errors is achieved by using $K(x_\beta - x_\alpha) + C_0\delta_{\alpha\beta}$ on the left-hand side and $K(x_0 - x_\alpha)$ on the right-hand side. More generally, the properties of kriging extend to the IRF case: existence and uniqueness of a solution, interpolation property, superposition theorem, kriging of gaps, and so on.

Coming back to the existence and uniqueness, in the derivation of (4.28), it has been implicitly assumed that the constraints (4.27) can always be satisfied. In other words, denoting by $\mathbf{X} = [x_\alpha^l]$ the $N \times k_n$ matrix of the monomials evaluated at the points x_α and by $\mathbf{x}_0 = [x_0^l]$ the k_n -dimensional vector of monomials evaluated at x_0 , we assumed that there exists a vector of weights $\lambda = (\lambda_1, \dots, \lambda_N)'$ such that

$$\mathbf{X}'\lambda = \mathbf{x}_0$$

It can be so only if \mathbf{x}_0 belongs to the linear subspace generated by the columns of \mathbf{X}' , which in turn can be true for all \mathbf{x}_0 if, and only if, the columns of \mathbf{X}' generate the whole k_n -dimensional space, that is, if \mathbf{X}' has rank k_n . This is equivalent to the familiar condition of linear independence of the k_n columns $\mathbf{x}^l = [x_\alpha^l]$ of \mathbf{X} encountered in universal kriging

$$\sum_{|l|=0}^k c_l x^l = 0 \quad \forall x \in S \quad \Rightarrow \quad c_l = 0, \quad |l| = 0, \dots, k$$

where S is the set of the data points. This plus the use of a strictly k -conditionally positive definite GC $K(h)$ and distinct data points ensure that the kriging system (4.28) has a unique solution.

Note that in this approach the conditions (4.27) are not introduced as unbiasedness conditions but as constraints to make the estimation error an allowable measure. Unbiasedness is achieved nevertheless because, by definition of an IRF- k , the ALC- k $Z^* - Z_0$ has a zero mean.

In fact there is more to $Z^* - Z_0$ than simply a zero mean. The conditions (4.27) ensure *numerical invariance* in the sense that an arbitrary polynomial $\sum a_l x^l$ of degree k may be added to $Z(x)$ without changing the value of $Z^* - Z_0$ (which is stronger than not changing the expected value). If moving neighborhoods are used for the estimation, then (4.27) may be interpreted physically as *filtering conditions*. They eliminate the effect of a local “drift,” if we mean by this a smooth component *locally* approximable by a polynomial of degree k .

4.6.2. Locally Equivalent Stationary Covariances

From a computational point of view, it is interesting to replace the $\mathbf{K} = [K_{\alpha\beta}]$ matrix, which is only k -conditionally positive definite, by a genuinely positive definite matrix $\mathbf{C} = [C_{\alpha\beta}]$.

In principle, this is always possible by picking an internal representation $Y(x) = Z(\varepsilon_x)$ and using the nonstationary covariance $\text{Cov}(Y(x), Y(y))$ in the kriging system (4.28) in lieu of $K(y - x)$. Since $\varepsilon_x \in \Lambda_k$, this covariance can be computed from the GC $K(h)$ using (4.16).

But this method is cumbersome and computationally slow. It would be much nicer if the IRF- k had a stationary representation: the stationary covariance $C(h)$ of this representation could then be used as a version of the GC $K(h)$. But in Section 4.5.4 we have seen that only an IRF- k with a bounded GC can have a stationary representation. This excludes the most important case of IRFs with polynomial covariances.

Fortunately we can be less demanding because in fact only *local stationarity* is needed. For any practical purpose we work in a restricted domain D —the studied area, or a moving neighborhood—and it is enough if the IRF- k $Z(x)$ has a representation $Y(x)$ that coincides within D with an SRF $Y_1(x)$ (being understood that $Y(x)$ may differ from $Y_1(x)$ outside the working domain D). This leads to the following definition:

Definition. An IRF- k is locally stationary over a bounded domain D if it has a representation $Y(x)$ that coincides on D with a stationary random function $Y_1(x)$.

The relationship between $Y_1(x)$, $Y(x)$, and $Z(x)$ is the following:

$$Y_1(x) = Y(x) \quad \forall x \in D$$

$$Y(\lambda) = Z(\lambda) \quad \forall \lambda \in \Lambda_k$$

Hence

$$\sum_i \lambda_i Y_1(x_i) = \sum_i \lambda_i Z(x_i)$$

for any λ in the space $\Lambda_k(D)$ of allowable measures with support in D . The ordinary covariance $C(h)$ of $Y_1(x)$ is equivalent to the GC $K(h)$ over D in the sense that for any $\lambda \in \Lambda_k(D)$,

$$\sum_i \sum_j \lambda_i \lambda_j C(x_j - x_i) = \sum_i \sum_j \lambda_i \lambda_j K(x_j - x_i)$$

$C(h)$ is called a *locally equivalent stationary covariance*.

Not every IRF- k is locally stationary,⁵ but the IRF- k of practical interest are. In particular, any IRF- k with a polynomial GC is locally stationary on any bounded open set (Matheron, 1973a).

A locally equivalent stationary covariance $C(h)$ differs from its parent GC- k $K(h)$ by an even polynomial of degree $2k$ that depends on the dimensions of the domain D (it must since the IRF is not globally stationary)

$$C(h) = K(h) + Q(h)$$

Even for a fixed D the covariance $C(h)$ is not unique. However, Matheron (1974b) showed that there is at most one stationary *internal* representation over D —but it is usually difficult to find it!

Examples

1. We return to the Brownian motion without drift of Example 1 in Section 4.5.2 but restrict our attention to the interval $[-R, +R]$. If instead of $X(t) - X(0)$ we consider the representation

$$Y(t) = X(t) - \frac{1}{2}[X(-R) + X(R)]$$

(which satisfies $Y(-R) = -Y(R)$), we obtain the covariance

$$\sigma(t, t') = R - |t' - t|$$

$C(h) = R - |h|$ is therefore locally equivalent to the GC $K(h) = -|h|$ for $|h| \leq 2R$. (It corresponds to the stationary internal representation on $[-R, +R]$.) An arbitrary constant $A > 0$ may be added so that the class of locally stationary covariances is a one-parameter family.

2. If $X(t)$ has linear drift, it is advantageous to consider the representation

$$Y(t) = X(t) - \frac{1}{2R} \int_{-R}^{+R} X(s) ds - \frac{1}{2R} [X(R) - X(-R)]t$$

By construction, $EY(t) = 0$ and the covariance is found to be

$$\sigma(t, t') = \frac{R}{3} - |t' - t| + \frac{(t' - t)^2}{2R}$$

This is in fact the covariance of the stationary internal representation on $[-R, +R]$. The complete class of locally equivalent stationary covariances (for the IRF-1) is of the form (Matheron, 1974b)

$$C(h) = A - |h| + Bh^2 \quad \text{with} \quad B \leq \frac{1}{2R} \quad \text{and} \quad A \geq R(1 - 2BR) + \frac{1}{3}R(2BR)^2$$

This class contains the covariance $(1/2R)(R - |h|)^2$ (quadratic model, valid in 3D).

3. An example of a locally stationary representation in 2D is obtained using the construction based on Poisson lines presented in Section 7.6.2: the convex set D is intersected by a network of Poisson lines. Each line cuts D in two parts; random values are assigned to each part and cumulated. The resulting RF has the stationary covariance of the form

$$\sigma(x, y) = A[\frac{1}{2}|\partial D| - |y - x|]$$

where $|\partial D|$ is the perimeter of D and A is a positive constant.

4. Table 4.2 gives a particular family of covariances locally equivalent to polynomial GCs for $k \leq 2$. C_2 and C_3 are deduced by the turning bands method from C_1 which satisfies $C_1(2R) = -C_1(0)$ when $A = 0$. The formulas are written in terms of the modulus $r = |h|$ and hold for $r \leq 2R$. Adding a strictly positive constant A makes the covariances strictly positive definite. Note that for ALC- k the results of variance calculations do not depend on R or A because R and A are only involved in even powers of r .

5. Table 4.3 gives the locally equivalent stationary covariances for the GC model $K(h) = |h|^2 \log |h|$. Note that this is a valid model in \mathbb{R}^n even though the equivalence with splines is only true in \mathbb{R}^2 . $C_2(r)$ is graphed in Figure 4.9. It becomes negative at $r = R$, indicating a "hole effect" which is not surprising when one thinks of the physics of a flexed plate. As long as $r \leq 2R$, the function $C_2(r)$ may be used just as an ordinary covariance even without restrictions on the weights, but it is equivalent to the spline GC only if the three conditions ensuring that $k = 1$ are imposed. For $r \geq 2R$ the function $C_2(r)$ loses its positive-definiteness property.

TABLE 4.2. Stationary covariances locally equivalent to the GC

$$K(h) = -b_0|h| + b_1|h|^3 - b_2|h|^5 \text{ for } r = |h| \leq 2R$$

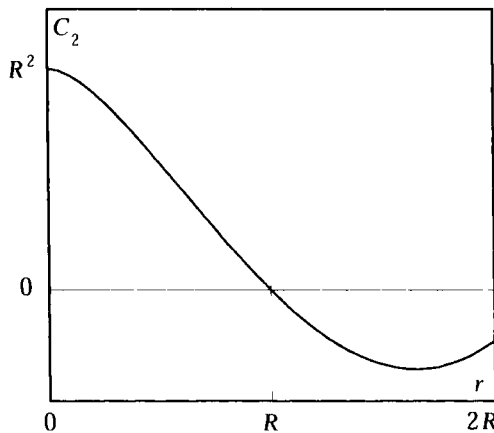
In \mathbb{R}^1	$C_1(r) = A - b_0(r - R) + b_1(r^3 - 3r^2R + 2R^3) - b_2(r^5 - 5r^4R + 20r^2R^3 - 16R^5)$
In \mathbb{R}^2	$C_2(r) = A - b_0\left(r - \frac{\pi}{2}R\right) + b_1\left(r^3 - \frac{9\pi}{8}r^2R + \frac{3\pi}{2}R^3\right) - b_2\left(r^5 - \frac{225\pi}{128}r^4R + \frac{75\pi}{8}r^2R^3 - 15\pi R^5\right)$
In \mathbb{R}^3	$C_3(r) = A - b_0(r - 2R) + b_1(r^3 - 4r^2R + 8R^3) - b_2(r^5 - 6r^4R + 40r^2R^3 - 96R^5)$

TABLE 4.3. Stationary covariances locally equivalent to the GC

$$K(h) = |h|^2 \log |h| \text{ for } r = |h| \leq 2R$$

$C_1(r) = R^2/2 - (3/2 - \log 2)r^2 + r^2 \log(r/R)$
$C_2(r) = R^2 - r^2 + r^2 \log(r/R)$
$C_3(r) = 3R^2/2 - (11/6 - \log 2)r^2 + r^2 \log(r/R)$

Source: Matheron (1981a).

**FIGURE 4.9.** Stationary covariance locally equivalent to $K(r) = r^2 \log r$ in 2D for $r \leq 2R$.

The r^2 and r^4 terms in the covariances are transparent to the kriging equations: in theory, they have no influence on either the kriging weights or the kriging variance. Numerically, however, these terms will dominate the other terms if R is large.

These numerical considerations suggest to choose for R the smallest possible value. R is the radius of a circle (or a sphere) large enough to include all points involved in the estimation problem considered, that is, both the observations and the points to be estimated. For example,

R may be the radius of the largest moving neighborhood or of the current neighborhood (hence the maximum distance involved in covariance calculations is $2R$).

4.6.3. Status of the Drift in IRF- k Theory

Since the concept of IRF- k encompasses a class of random functions (the representations) that are defined up to a polynomial, one could argue that this polynomial implicitly captures the concept of drift—whatever that means. But the drift itself can only be defined in an exceptional case: when the class of representations of the abstract IRF- k \tilde{Z} contains a special random function $Y(x)$ that is distinguishable from all others by some remarkable property and can thus serve as an absolute reference. Then any other representation $Z(x)$ would differ from $Y(x)$ by a polynomial drift.

A case when this circumstance occurs is that of an IRF- k with a stationary representation $Y_{St}(x)$. Any other representation is then of the form

$$Z(x) = Y_{St}(x) + \sum_l A_l x^l$$

This is nothing but the universal kriging model (drift + stationary residual) with random coefficients. This case, however, is an exceptional one requiring that one of the GCs be bounded. In general, an IRF- k has no stationary representation and therefore no uniquely definable drift (IRF- k with polynomial GCs never have a stationary representation). Far from being an objection against the IRF model, this circumstance is consistent with our initial goal which was precisely to deal with phenomena for which the dichotomy into drift and residual is meaningless, or at least arbitrary. The important point is that spatial estimation (kriging) remains possible within this model.

A Fallacy

At the early times of the development of the IRF theory, it was believed that locally equivalent stationary covariances could be used to perform drift estimation. But practical examples soon showed that it did not work. The drift estimates obtained did not reflect at all the general behavior of the phenomenon. Furthermore these estimates differed with the radius R of the domain. The reason is now well understood. Locally equivalent stationary covariances are not unique even over a fixed domain D . As a consequence the estimated drift is arbitrary and depends on the particular choice of the covariance. To some extent the terminology “locally equivalent stationary representation” is misleading, for one expects the equivalent $Y(x)$ to really look stationary over D . But that cannot be the case. If $Y_1(x)$ and $Y_2(x)$ are two distinct locally equivalent stationary representations, their difference

$$Y_1(x) - Y_2(x) = \sum_l A_l x^l$$

is necessarily a polynomial of degree k , and at least one of these functions has locally the aspect of a drift.

The locally equivalent stationary $Y(x)$ are no more than mathematical constructs, useful to derive covariance models but unable to give a meaning to the drift.

Estimation of a Regional Trend

For some reason users may want nevertheless to estimate a spatial trend. Since our information on the phenomenon is most of the time very fragmentary, we must carefully distinguish two problems: that of the conceptual definition of the trend proposed by the user, and that of the estimation of the trend so defined.

In light of the discussion on the objectivity of probabilistic parameters (see Section 1.2), we are led to ask the user the following question: Assuming that instead of knowing the variable at a limited number of points x_α you knew values $z(x)$ at *all* the points x in the domain of interest, which algorithm would you use to calculate your trend function $t(x)$? If the notion of trend as portrayed in the user's mind has an objective meaning, such an algorithm must exist.

Most times the user will respond with a purely conventional definition of the trend that translates into a linear algorithm, for example, fit a least squares polynomial, or apply a low-pass filter. We will then be able, on the basis of the available data z_α , to estimate *that* specific trend using a kriging estimator.

Suppose, for example, that the user defines the trend as the polynomial of degree k that best fits $Z(x)$ over a specified domain D . The coefficients β_l defined by

$$\int_D \left(Z(x) - \sum_l \beta_l x^l \right)^2 dx \quad \text{minimum}$$

are *spatial averages* of $Z(x)$, of the form

$$\beta_l = \int_D Z(x) \varphi_l(x) dx \quad (4.29)$$

where φ_l are ad hoc polynomials. It is therefore possible to estimate these β_l by kriging on the basis of the available data Z_1, \dots, Z_N . By the linearity of kriging, the estimators are simply the same weighted averages of the kriging estimator $Z^*(x)$ at each point x :

$$\beta_l^* = \int_D Z^*(x) \varphi_l(x) dx \quad (4.30)$$

and the estimator of the trend is

$$t^*(x) = \sum_l \beta_l^* x^l$$

This is equivalent to minimizing the integral

$$\int_D \left(Z^*(x) - \sum_l \beta_l x^l \right)^2 dx$$

in which the true $Z(x)$ are replaced by their kriging estimates $Z^*(x)$. In principle, $Z^*(x)$ must be computed from all the data, but (4.30) certainly provides a good estimate of (4.29) even if $Z^*(x)$ is computed locally.

It must be emphasized that this technique differs from a direct polynomial fit to the observations. We want the fit to be representative of the domain D and not only of the data points. The closest we can come to this objective is by reconstructing the values of $Z(x)$ as accurately as possible over the whole domain D , or a discretized form of it, and then fit a polynomial. In the process we make use of the structure information contained in the GC, which the plain least squares fit at the observations does not.

4.6.4. Kriging and Splines

Interpolating Splines

Just as in the model of universal kriging the estimator $Z^*(x)$ derived in the IRF model can be regarded as an interpolating function of the form

$$z^*(x) = \sum_{\alpha} b_{\alpha} K(x - x_{\alpha}) + \sum_l c_l x^l \quad (4.31)$$

with coefficients b_{α} and c_l defined by the conditions (dual kriging system)

$$\begin{cases} z^*(x_{\alpha}) = z_{\alpha} & \alpha = 1, \dots, N \\ \sum_{\alpha} b_{\alpha} x_{\alpha}^l = 0 & l = 0, \dots, k \end{cases}$$

(Note: In this section we switch to lowercase notations for the interpolators to emphasize their deterministic character; for the same reason we denote the interpolated point by x rather than x_0 .)

Consider now in 1D the IRF-1 $Z(x)$ with GC $K(h) = |h|^3$, and let us examine the behavior of the kriging interpolator $z^*(x)$ based on N consecutive points x_1, \dots, x_N . From (4.31),

$$\begin{cases} z^*(x) = \sum_{\alpha=1}^N b_{\alpha} |x - x_{\alpha}|^3 + c_0 + c_1 x \\ z^*(x_{\alpha}) = z_{\alpha} & \alpha = 1, \dots, N \\ \sum_{\alpha=1}^N b_{\alpha} = 0 & \text{and} \quad \sum_{\alpha=1}^N b_{\alpha} x_{\alpha} = 0 \end{cases}$$

$z^*(x)$ is a cubic polynomial within each interval $[x_\alpha, x_{\alpha+1}]$, assumes the values z_α and $z_{\alpha+1}$ at its boundaries, and is continuous in the first and second derivatives across boundaries, including at x_1 and x_N where the second derivative is zero. In the intervals $]-\infty, x_1]$ and $[x_N, +\infty[$, $z^*(x)$ is linear as can be seen by expansion of the polynomial $(x - x_\alpha)^3$ and the application of the two constraints on b_α . Therefore $z^*(x)$ coincides with the cubic interpolating spline going through z_1, \dots, z_N , which is the function $f(x)$ minimizing

$$J(f) = \int_{-\infty}^{+\infty} [f''(x)]^2 dx$$

subject to the N constraints $f(x_\alpha) = z_\alpha$. Note that the linear behavior of $f(x)$ on the outer intervals cancels $f''(x)$ and ensures that the integral is finite.

A physical interpretation of the problem is to consider a metal strip clamped at the N points $(x_1, z_1), \dots, (x_N, z_N)$. The strip will adopt the shape that minimizes its flexing energy, which is proportional to the square of the curvature of the strip, a quantity approximated by $J(f)$.

The analog of this problem in 2D is to minimize the flexing energy of a thin metal plate, which is proportional to

$$J(f) = \iint \left[\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right] dx dy$$

under the N constraints $f(x_\alpha, y_\alpha) = z_\alpha$ (e.g., Gonzalez-Casanova and Alvarez, 1985). (As usual in 2D we use (x, y) as the coordinates of a point.) Duchon (1975) derived the explicit solution of this minimization problem

$$f(x, y) = \sum_{\alpha} b_{\alpha} K(r_{\alpha}) + c_0 + c_1 x + c_2 y \quad (4.32)$$

where

$$K(r) = r^2 \log r \quad \text{and} \quad r_{\alpha}^2 = (x - x_{\alpha})^2 + (y - y_{\alpha})^2$$

with coefficients satisfying

$$\sum_{\alpha} b_{\alpha} = 0 \quad \sum_{\alpha} b_{\alpha} x_{\alpha} = 0 \quad \sum_{\alpha} b_{\alpha} y_{\alpha} = 0$$

The spline interpolating function (4.32) turns out to have the same form as the kriging interpolator (4.31) with $k = 1$ and a generalized covariance function $K(h) = |h|^2 \log |h|$.

The functions f presented above are known as *biharmonic splines* because each in its space of definition satisfies the equation $\Delta^2 f(x) = 0$ at every point $x \neq x_{\alpha}$ (Δ^2 : iterated Laplacian). The basis function K itself satisfies

$\Delta^2 K = \delta$. In 3D the solution is $K(r) = -r$. Higher-order splines can be defined by considering higher-order derivatives, but biharmonic splines are used most.⁶ The reader is referred to Wahba (1990) for further reading.

It is remarkable that biharmonic splines and kriging lead to the same results, considering that they proceed from two very different approaches. With splines one postulates a property of the interpolating surface, whereas with kriging one focuses on modeling the underlying random function itself. A common ground is the property of invariance under translations that is shared by IRF theory and by the differential operators with constant coefficients generally used to define splines.

The notion that spline interpolation is in certain cases equivalent to minimum variance linear prediction was shown by Kimeldorf and Wahba (1970) within the scope of a Bayesian analysis, or Duchon (1976) who related splines with conditional expectations of random fields. Matheron (1981a) proved that the converse is also true: any kriging problem can be cast into a spline problem defined as the minimization of a norm associated with a linear operator. Such operator, however, is not necessarily with constant coefficients. Dubrule (1981) gives the following example: In 1D, kriging an IRF-I with GC $K(h) = |h|^3$ is equivalent to splines with the differential operator $T = \partial^2 / \partial x^2$. But if the covariance is $K(h) = -|h| + |h|^3$, the operator T cannot be differential with constant coefficients (the proof is not obvious). Likewise for a fixed order k , if we consider polynomial GCs

$$K(h) = \sum_{p=0}^k (-1)^{p+1} b_p |h|^{2p+1}$$

only one of them $(-1)^{k+1} |h|^{2k+1}$ corresponds to the minimization of the norm of a simple differential operator ($T = \partial^{k+1} / \partial x^{k+1}$).

Spline and kriging interpolation are thus equivalent in a formal way but not so in a practical way. Given the operator defining the spline it is a tractable problem to find the equivalent kriging formulation, whereas it can be extremely difficult to identify the minimization problem associated with a given kriging solution. In this perspective biharmonic splines constitute a very special and isolated case.

Smoothing Splines

When the data are subject to measurement errors—or when interpolating splines produce nonsensical results—one relaxes the constraints of exact fit at the sample points to request only that the fitted surface pass not too far from the data. Assuming equal error variances, one minimizes a criterion of the form

$$\sum_{\alpha} [f(x_{\alpha}) - z_{\alpha}]^2 + \rho J(f) \quad (\rho > 0) \quad (4.33)$$

within a class of functions f with continuous derivatives up to the appropriate order. The parameter ρ controls the trade-off between the smoothness of the curve and the fit at the data point. As ρ tends to zero, the solution tends to an interpolating spline; as ρ increases, the solution approaches a line or plane fit by least squares. Such interpolating functions are known as *smoothing splines*.

The formal equivalence with kriging extends to smoothing splines but in a slightly modified setup. The model is now

$$Z_\alpha = Y_\alpha + \varepsilon_\alpha$$

where the underlying $Y(x)$ is a *smooth* random function with covariance $K(h)$ and the ε_α are errors, uncorrelated with the RF $Y(x)$ and satisfying

$$E\varepsilon_\alpha = 0 \quad E\varepsilon_\alpha \varepsilon_\beta = S_{\alpha\beta}$$

The estimation of the smooth component $Y(x)$ from noisy data Z_α is a cokriging problem (analogous to Equations (3.36)) whose solution, in dual kriging terms, is of the form

$$\begin{cases} y^*(x) = \sum_{\alpha} b_{\alpha} K(x - x_{\alpha}) + \sum_l c_l x^l \\ \sum_{\beta} b_{\beta} (K_{\alpha\beta} + S_{\alpha\beta}) + \sum_l c_l x_{\alpha}^l = z_{\alpha} & (\alpha = 1, \dots, N) \\ \sum_{\alpha} b_{\alpha} x_{\alpha}^l = 0 & (l = 0, \dots, k) \end{cases} \quad (4.34)$$

Biharmonic smoothing splines are a particular case of such cokriging, with a diagonal error covariance matrix $S_{\alpha\beta} = C_0 \delta_{\alpha\beta}$ and a covariance $K(h)$ proportional to $|h|^3$ in 1D and to $|h|^2 \log |h|$ in 2D (Matheron, 1981a; Dubrule, 1983). A simple parallel is established in 1D by Watson (1984).

A key problem in the application of smoothing splines is the choice of the smoothing parameter ρ . It is achieved by cross-validation techniques, in particular, the generalized cross-validation method of Craven and Wahba (1979). This is where splines and kriging really meet because the endeavor to determine ρ from the data is a form of structure identification, a step central to kriging. Along these lines Dubrule (1983) proposed to model the generalized covariance of $Z(x)$ with precisely the function that leads to smoothing splines in 2D:

$$K(h) = C_0 \delta(h) + b_S |h|^2 \log |h|$$

The problem of selecting ρ subsumes into the determination of the structural parameters C_0 and b_S . In fact, by identification with the solution of (4.33) (e.g., Wahba, 1990, p. 12), we find simply $\rho = C_0/b_S$, which makes sense when interpreted as a “noise-to-signal ratio.”

In conclusion of this section we note that the identification of splines with a particular form of kriging is a tangible benefit of the IRF- k theory. It allows kriging to borrow techniques developed for splines, such as estimation under linear inequality constraints. From an operational perspective, however, the two approaches remain different because kriging leaves open the possibility of selecting a covariance model other than of standard spline type if the data so suggest. An empirical comparison of the predictive performance of kriging and splines, with a discussion, can be found in Laslett (1994). We give an excerpt from his conclusion: "It is when data are not sampled on a grid that kriging has the potential to outpredict splines, because it involves a translation of information on covariation of data values from intensely sampled regions to sparsely sampled regions; splines and some other nonparametric regression procedures do not appear to exploit such information."

4.7. GENERALIZED VARIOGRAM

Now that the theory is laid out, we are left with the crucial problem of structure identification: to determine the order k and the generalized covariance function $K(h)$. We will start with the particular case of regularly spaced data along lines and introduce the generalized variogram, a new structural tool that inherits the advantages of the ordinary variogram. It can be calculated "non-parametrically" (i.e., without presuming a parametric model), and displayed and fitted graphically. The first definition of the generalized variogram can be found in Matheron (1972b), and it was first used by Orfeuil (1972) to check simulations of IRF- k . Applications include the analysis of microgravimetry data (Chilès, 1979b), geothermal data (Chilès and Gable, 1984), and fracture data (Chilès and Gentier, 1993). A similar tool, though more difficult to interpret, has been proposed by Cressie (1987).

4.7.1. Definition

The generalization of the simple increment or forward finite difference

$$\Delta_h Z(x) = Z(x + h) - Z(x)$$

is the increment or forward finite difference of order $k + 1$, which is the simplest ALC- k in 1D:

$$\Delta_h^{k+1} Z(x) = (-1)^{k+1} \sum_{p=0}^{k+1} (-1)^p \binom{k+1}{p} Z(x + ph) \quad (4.35)$$

By definition, the generalized variogram of order k , denoted by $\Gamma(h)$ and abbreviated as GV, is the appropriately scaled variance of the increment of

order $k + 1$:

$$\Gamma(h) = \frac{1}{M_k} \text{Var}[\Delta_h^{k+1} Z(x)] \quad (4.36)$$

The scaling factor

$$M_k = \binom{2k+2}{k+1}$$

is introduced to ensure that in the case of a pure nugget effect, $\Gamma(h) = C_0 \delta(h)$ as for an ordinary variogram.⁷ Explicitly we have

$$\begin{cases} k = 1: & \Gamma(h) = \frac{1}{6} \text{Var}[Z(x+2h) - 2Z(x+h) + Z(x)] \\ k = 2: & \Gamma(h) = \frac{1}{20} \text{Var}[Z(x+3h) - 3Z(x+2h) + 3Z(x+h) - Z(x)] \\ k = 3: & \Gamma(h) = \frac{1}{70} \text{Var}[Z(x+4h) - 4Z(x+3h) + 6Z(x+2h) \\ & \quad - 4Z(x+h) + Z(x)] \end{cases}$$

4.7.2. Relationship between Generalized Variogram and Generalized Covariance

Since the increment of order $k + 1$ is an ALC- k , its variance can be expressed in terms of the generalized covariance. Applying the definition (4.14) of a generalized covariance to (4.36) gives

$$\Gamma(h) = \frac{1}{M_k} \sum_{p=-(k+1)}^{k+1} (-1)^p \binom{2k+2}{k+1+p} K(ph) \quad (4.37)$$

Explicitly

$$\begin{cases} k = 1: & \Gamma(h) = K(0) - \frac{4}{3}K(h) + \frac{1}{3}K(2h) \\ k = 2: & \Gamma(h) = K(0) - \frac{3}{2}K(h) + \frac{3}{5}K(2h) - \frac{1}{10}K(3h) \\ k = 3: & \Gamma(h) = K(0) - \frac{8}{5}K(h) + \frac{4}{5}K(2h) - \frac{8}{35}K(3h) + \frac{1}{35}K(4h) \end{cases}$$

While being simple, the relationship between $\Gamma(h)$ and $K(h)$ is more complex than for the ordinary variogram. In particular, whether the GV determines the GC (up to the even polynomial) is still an unsolved problem (Chauvet, 1987, pp. 215–231). Indeed $\Gamma(h)$ only represents the variance of the increments associated with h , while what determines the GC is the set of all covariances of these increments for all h .

However, there are two important cases where the GV does determine the GC: when the GC is bounded and when the GC is of polynomial form. This is readily seen from (4.37) by observing that if $K(h)$ behaves like $|h|^a$, so does $\Gamma(h)$, only the coefficient changes. On the other hand, if $K(h)$ is an ordinary

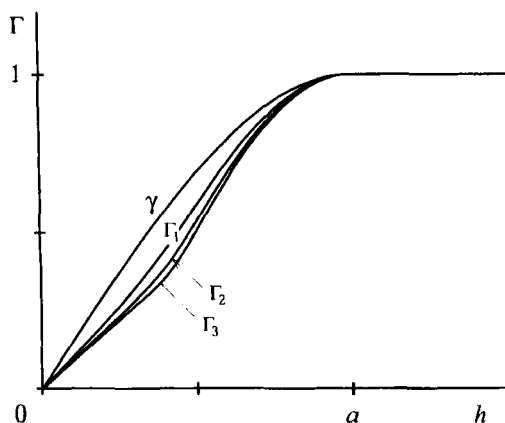


FIGURE 4.10. Generalized variograms associated with the spherical variogram model for $k = 0, 1, 2, 3$.

covariance such that $K(h) = 0$ for $|h| \geq a$, then by (4.37) $\Gamma(h) = K(0)$ for $|h| \geq a$. The GV has the same range and sill as the ordinary variogram $K(0) - K(h)$; the nugget effect remains the same. All common covariance models have a polynomial behavior near the origin, and the GV inherits this but with changes in the coefficients; in particular, the even terms of degree $\leq 2k$ disappear. Figure 4.10 displays the GVs associated with the spherical model (2.47).

From a practical point of view, it should be kept in mind that the objective is to identify the GC function $K(h)$, since it is the quantity involved in estimation problems. So, after computing an experimental GV, one will try to fit a linear combination of GV models associated with known GC models. The above-mentioned models are the only ones used in practice. Since the GV models remain close to the GC models, fitting a model to an experimental GV is no more difficult than fitting a model to an ordinary variogram.

The statistical inference problems associated with the GV are essentially the same as for the ordinary variogram. The conclusions are also similar except that the GV can only be computed at shorter distances (the length of the region of interest divided by $k + 1$) and fluctuation variances are higher. In the Gaussian case the regional GV $\Gamma_R(h)$ still converges to $\Gamma(h)$ when $h \rightarrow 0$ (micro-ergodicity) provided that $\Gamma(h)$ is not too regular near the origin ($\Gamma(h) \sim |h|^\alpha$ with $0 < \alpha < 2k + 2$). More on statistical properties of the GV can be found in Chilès (1979b).

Polynomial Covariances

If $K(h) = \varepsilon(-\alpha/2)b_\alpha|h|^\alpha$, where $\varepsilon(\cdot)$ is the Euler gamma function which controls the sign (we cannot use the symbol $\Gamma(\cdot)$ here), then

$$\Gamma(h) = \varepsilon\left(-\frac{\alpha}{2}\right)B_\alpha b_\alpha |h|^\alpha \quad \text{with} \quad B_\alpha = \frac{1}{M_k} \sum_{p=-k-1}^{k+1} (-1)^p \binom{2k+2}{k+1+p} |p|^\alpha$$

Explicit results are given below for the polynomial/logarithmic family model (where h stands for $|h|$; notice that b_p is here the coefficient of the monomial or logarithmic-monomial of degree p)

$$\begin{aligned}
 k = 1 & \quad \begin{cases} K(h) = C_0 \delta(h) - b_1 h + b_2 h^2 \log h + b_3 h^3 \\ \Gamma(h) = C_0[1 - \delta(h)] + \frac{2}{3} b_1 h + \frac{4}{3} (\log 2) b_2 h^2 + \frac{4}{3} b_3 h^3 \end{cases} \\
 k = 2 & \quad \begin{cases} K(h) = C_0 \delta(h) - b_1 h + b_2 h^2 \log h + b_3 h^3 - b_4 h^4 \log h - b_5 h^5 \\ \Gamma(h) = C_0[1 - \delta(h)] + \frac{3}{5} b_1 h + \frac{3}{10} (\log 256 - \log 27) b_2 h^2 + \frac{3}{5} b_3 h^3 \\ \quad + \frac{3}{10} (27 \log 3 - 32 \log 2) h^4 + \frac{33}{5} b_5 h^5 \end{cases} \\
 k = 3 & \quad \begin{cases} K(h) = C_0 \delta(h) - b_1 h + b_2 h^2 \log h + b_3 h^3 - b_4 h^4 \log h - b_5 h^5 + b_6 h^6 \log h + b_7 h^7 \\ \Gamma(h) = C_0[1 - \delta(h)] + \frac{4}{7} b_1 h + \frac{72}{35} (\log 4 - \log 3) b_2 h^2 + \frac{16}{35} b_3 h^3 \\ \quad + \frac{24}{35} (27 \log 3 - 40 \log 2) b_4 h^4 + \frac{16}{7} b_5 h^5 \\ \quad + \frac{8}{35} (1248 \log 2 - 729 \log 3) b_6 h^6 + \frac{2416}{35} b_7 h^7 \end{cases}
 \end{aligned}$$

$\Gamma(h)$ is a polynomial of degree $2k + 1$ in h , where all the logarithmic terms of $K(h)$ have turned into even degree terms.

4.7.3. An Application to the Topography of Fractures

Chilès and Gentier (1993) describe an application of generalized variograms to the study of the morphology of natural rock fractures. A precise determination of the topography of the two surfaces bordering a fracture is performed in the laboratory on cores by sampling the surfaces along profiles, using a special tool called a “profilometer.” The goal is to understand the roughness of the fracture surfaces and the spatial variations of the aperture, which control the mechanical behavior of fractures under stress and the flow and transport within the fractures.

Figure 4.11 shows a profile recorded across a fracture surface, its raw variogram, its variogram of residuals, and its GV for $k = 1$. The raw variogram has a parabolic shape that mainly reflects the bias due to the presence of a strong linear drift. By contrast, the GV, displayed with a vertical exaggeration of about 10, exhibits a clear sill reached at a range of about 10 mm. It agrees with the variogram of residuals in the sense that they have the same range and sill, and the slope at the origin of $\Gamma(h)$ is two-thirds of the slope of $\gamma_R(h)$, as predicted by theory.

The excellent agreement of the two variograms in this example is due to the simplicity of the drift, which allows a global linear fit. In more complex cases a simple global fit of the drift would still leave bias in the variogram of residuals, whereas the GV remains unbiased at short distances because it filters the drift locally. In addition $\Gamma(h)$ can reflect the unbounded behavior of the underlying variogram, while by construction a variogram of residuals is always bounded.

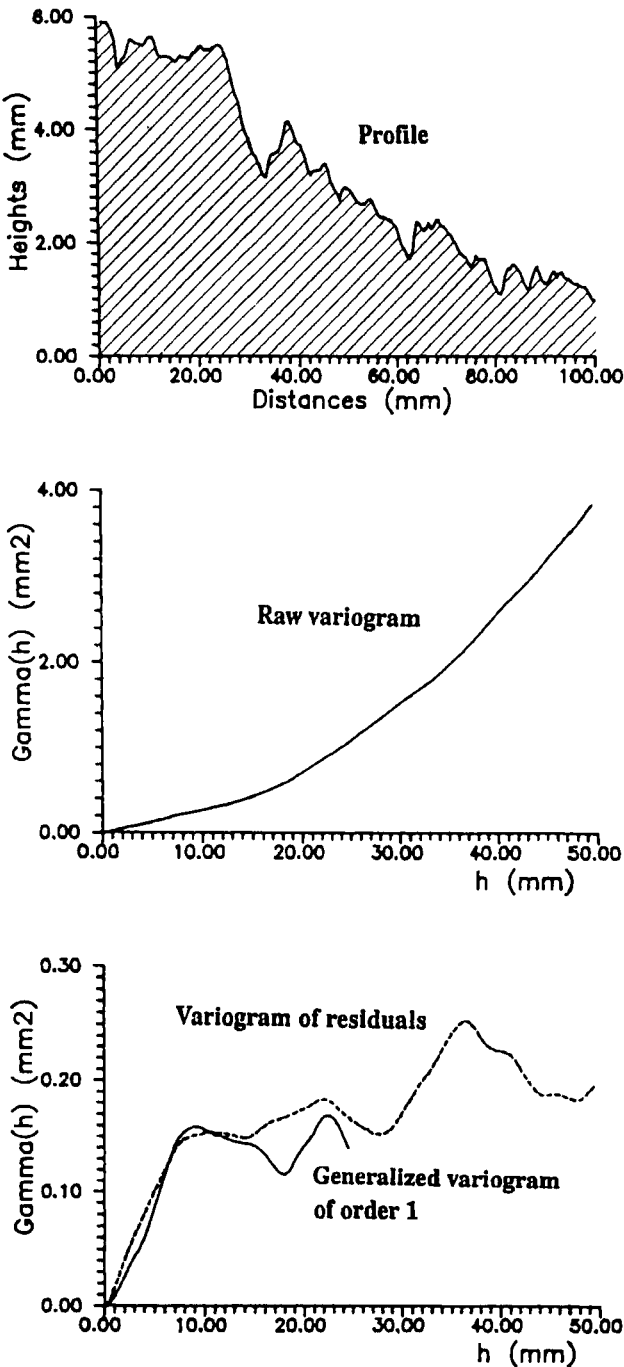


FIGURE 4.11. Fracture profile shown on cross-section along core axis, and associated variograms. From Chilès and Gentier (1993), with kind permission from Kluwer Academic Publishers.

4.8. AUTOMATIC STRUCTURE IDENTIFICATION IN THE GENERAL CASE

In the general case of scattered data, it is not possible to compute a structure function such as the generalized variogram and graph it to suggest a model. We have to operate “blindly” and rely on an algorithm to pick the best set of parameters within a prespecified family of models. This is a parametric approach, whereas the generalized variogram is nonparametric. The design of that automatic procedure is the critical and also the most difficult part of the application of the IRF theory. We will present here a method that claims no optimality properties but has the advantage of being completely general, since it does not require the data points to be evenly distributed nor the underlying RF to be Gaussian.

Since this method was introduced (Delfiner, 1976) numerous attempts have been made to improve on it and find estimators endowed with optimality properties. One approach (Delfiner, 1977; Kitanidis, 1983, 1985; Marshall and Mardia, 1985; Stein, 1986a) consists in treating the covariance identification problem as one of estimation of variance components using Rao’s minimum norm quadratic unbiased estimator (MINQUE) theory (Rao, 1970, 1971a, b, 1973 p. 303). Another approach is restricted maximum likelihood (REML) estimation based on maximizing the loglikelihood associated with a vector of linearly independent ALC- k (Kitanidis, 1983). The latter method relies heavily on the Gaussian assumption and is computationally intensive. Zimmerman (1989) proposes an efficient implementation for data on a lattice—but then, why not use generalized variograms?

4.8.1. General Principles

The method is based on the assumption that the covariance $K(h)$ is a linear combination of elementary models $K_p(h)$

$$K(h) = \sum_p b_p K_p(h) \quad (4.38)$$

For a polynomial GC model of order k we take the elementary models $K_p(h) = (-1)^{p+1} |h|^{2p+1}$, $p = 0, \dots, k$, plus the unit nugget effect $\delta(h)$ that we will associate with $p = -1$ (but we will maintain the usual notation C_0 rather than b_{-1} for the nugget effect value). It is also possible to include the spline model $|h|^2 \log |h|$, but then $-|h|^5$ should be dropped to keep the number of terms as low as possible. We could also consider for $K_p(h)$ models with given ranges and estimate the sills b_p .

Let us define the synthetic notation

$$K(\lambda) = \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} K(x_{\beta} - x_{\alpha}) = EZ(\lambda)^2$$

It results from (4.38) that

$$EZ(\lambda)^2 = K(\lambda) = \sum_p b_p K_p(\lambda) \quad (4.39)$$

This is a *linear regression* of $Z(\lambda)^2$ on the predictor variables $K_p(\lambda)$. To determine the coefficients of this regression, we construct, based on the data points, a large number of ALC- k $Z(\lambda_i)$ and minimize

$$Q(\mathbf{b}) = \sum_i w_i^2 \left[Z(\lambda_i)^2 - \sum_p b_p K_p(\lambda_i) \right]^2 \quad (4.40)$$

where $\mathbf{b} = [b_p] = (C_0, b_0, \dots, b_k)'$ is the vector of the unknown coefficients. The weights w_i^2 are introduced to equalize the variances of $Z(\lambda_i)^2$ and therefore should, in theory, be equal to the reciprocals of $\text{Var}[Z(\lambda_i)^2]$. But these variances are unknown. They involve the fourth-order moments of $Z(\lambda_i)$ or at least, in the Gaussian case, the second-order moments, which just depend on the covariance to be estimated. An iterative procedure could be considered, but it is not worth pursuing too far in this direction, since another basic assumption of least squares, the noncorrelation of the $Z(\lambda_i)^2$, is violated anyway. Thus one simply uses empirical weights. For example, one can take the weights associated with the elementary models $\delta(h)$, $-|h|$, or $|h|^3$ in the Gaussian case where

$$\text{Var}[Z(\lambda_i)^2] = 2K(\lambda_i)^2$$

Just as a variogram is better known at short distances than at large ones, it is preferable to assign a larger weight to increments that are more "packed." Therefore the weighting derived from $|h|^3$ is used whenever possible. For $k = 0$ a weighting based on $-|h|$ is used.

In fact the problem departs from a plain least squares regression when we require the fit to constitute a valid GC at the order k considered. The coefficients are not free but must satisfy the inequalities of a polynomial GC. This would call for a constrained minimization of (4.40) using the techniques of quadratic programming. But this procedure would be computationally intensive, and the easier way is to simply fit all possible regressions based on all subsets of the predictor variables $K_p(\lambda)$ and discard those that do not satisfy the constraints. Valid models are guaranteed to exist: the solutions associated with a single elementary model lead to a positive coefficient. If the search is limited to models for $k \leq 2$, there are $2^4 - 1 = 15$ possible regression equations to consider.

At this stage two questions remain:

- How to select the order k ?
- What is the best valid regression equation (4.39)?

The answers are unfortunately largely empirical due to the absence of solid goodness-of-fit criteria. For the second question least squares suggest that one ought to consider the residual sum of squares $Q(\mathbf{b})$ or its normalized version

$$\frac{Q(\mathbf{b})}{Q(\mathbf{0})} = \frac{Q(C_0, b_0, \dots, b_k)}{Q(0, 0, \dots, 0)} = \frac{\text{residual sum of squares}}{\text{total sum of squares}}$$

This ratio is always less than 1. A value close to 1 indicates a bad fit, and a value close to $\frac{2}{3}$ indicates a good fit. This number $\frac{2}{3}$ is the theoretical value $E[Q(\mathbf{b})]/E[Q(\mathbf{0})]$ in the Gaussian case with the correct model and a perfect fit. The problem is that no significance test is available such as the F -test of standard least squares theory, and we must therefore base our selection on mere sample fluctuations. In practice, the quality of the fit is gauged not from the simple criterion Q but from several criteria considered simultaneously. The same holds true for the determination of the order k .

To illustrate the above, we now sketch the procedure implemented in a program named BLUEPACK developed at the Center for Geostatistics, Fontainebleau.⁸

4.8.2. An Implementation Example

BLUEPACK fits a polynomial GC model plus a nugget effect. (There is a variant, not presented here, including a spline term). The procedure comprises three steps:

1. Determine the order k (0, 1 or 2).
2. Compute all possible regressions (3, 7, or 15 according to the value of k), eliminate those that are not valid, and perform a first selection.
3. Compare the best regressions of step 2 to make a final selection.

Step 1. To determine the order k the idea is to cross out known points and estimate them from neighboring data points, while varying only the number of unbiasedness conditions (1, 3, or 6), that is, the order k of the IRF.

For each estimated data point Z_α the errors $\hat{Z}_\alpha - Z_\alpha$ are ALC- k of order $k = 0, 1$ or 2 . These errors are ranked by ascending magnitude, and the corresponding orders k are thus ranked by performance. The mean rank is computed for each k over all estimated Z_α , and the value selected is that with minimum mean rank. The advantage of this method is its robustness against outliers.

The mean square error criterion is also output by the program but only as a secondary check. In ambiguous cases it is advisable to select the lowest possible k .

Since at this stage of structure identification the covariance is not yet known, the above estimators \hat{Z}_α are simply determined by least squares. One design rule is to avoid using configurations of data points that have a high symmetry about the estimated point. Indeed a symmetric neighborhood tends to filter polynomials by itself, which leads to underestimate k . To introduce asymmetry, the data points are split into two concentric rings: data from the inner ring are used to estimate points from the outer ring, and vice versa.

Step 2. Once k is determined, the possible regressions are computed. Those that are not valid are discarded (i.e., for polynomial GCs, those that do not fulfill the constraints expressed in Table 4.1). A selection is then made among the remaining regressions. The operational criterion involves

a ratio of mean square errors. For any ALC- k , $Z(\lambda_i)$ we have

$$EZ(\lambda_i)^2 = K(\lambda_i) = \sum_p b_p K_p(\lambda_i)$$

Let

$$\hat{K}(\lambda_i) = \sum_p \hat{b}_p K_p(\lambda_i)$$

be its estimator. Since

$$E \left[\sum_i Z(\lambda_i)^2 \right] = \sum_i K(\lambda_i)$$

the ratio

$$\rho = \frac{E \left[\sum_i Z(\lambda_i)^2 \right]}{E \left[\sum_i \hat{K}(\lambda_i) \right]}$$

should be close to 1. The estimator

$$r = \frac{\sum_i Z(\lambda_i)^2}{\sum_i \hat{K}(\lambda_i)} \quad (4.41)$$

is a biased estimator of ρ . To reduce the bias, we can split the sample and use the jackknife estimator

$$\hat{\rho} = 2r - \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2}$$

where r_1 and r_2 are the ratios (4.41) computed separately in ring 1 and ring 2, comprising n_1 and n_2 measures, and r the ratio computed with all $n_1 + n_2$ measures (this eliminates a bias of the form b/n). The regression whose $\hat{\rho}$ is closest to 1 is selected.

Step 3 (Optional cross-validation). The method is similar to the leave-one-out described in Section 2.6.3 for the stationary case, and it consists in evaluating candidate models through their performance in actual kriging situations. This is to avoid basing the selection on marginal differences in the jackknife statistics. The sample points, or a subset of them, are estimated from their neighbors as if they were unknown, using the order k determined at Step 1 and the best covariance models from Step 2 (ranked by increasing distance of $\hat{\rho}$ from 1). If the model is correct, the error $Z_{(\alpha)}^* - Z_{\alpha}$ obtained when leaving the point x_{α} out has a variance equal to $\sigma_{K\alpha}^2$ so that the mean standardized square error

$$\text{m.s.e.} = \frac{1}{N} \sum_{\alpha} \frac{(Z_{(\alpha)}^* - Z_{\alpha})^2}{\sigma_{K\alpha}^2}$$

should be close to 1. As a reference, note that the variance of this ratio is equal to $2/N$ if the variables $Z_{(\alpha)}^* - Z_{\alpha}$ are independent and Gaussian.

The operational criterion is to minimize the mean square error

$$\text{m.s.e.} = \frac{1}{N} \sum_{\alpha} (Z_{(\alpha)}^* - Z_{\alpha})^2$$

under the constraint that the mean standardized square error defined above is not too far from 1. (Typically a tolerance of $\pm 3\sqrt{2/N}$ is used).

More sophisticated procedures may be used, such as pairwise ranking of models. Better yet, an analysis of the kriging errors $Z_{(\alpha)}^* - Z_{\alpha}$ or their standardized version $(Z_{(\alpha)}^* - Z_{\alpha})/\sigma_{K\alpha}$ can be performed using graphical statistical procedures. This procedure also allows one to spot suspect data points in the context of their surrounding neighbors.

An Example

Figure 4.12 shows an output of the BLUEPACK Automatic Structure Identification options illustrating the first two steps. The data are 573 topographic elevations of the Noirétable area presented in Section 2.7.4. In the first section each line indicates the performance of an order k . Results relative to measures based on polynomials fitted in rings 1 are reported under the heading "Ring 1"; they correspond to estimation of points of rings 2, and vice versa for "Ring 2." The means of the results obtained in Ring 1 and Ring 2 are given in the last column "Total." Naturally these means are weighted by the count of measures in each ring which are printed below the table ($n_1 = 805$ and $n_2 = 877$). A cutoff is applied on the sum of squares of the weights of the measures to avoid strong ill-conditioning; this explains why n_1 and n_2 may be different. The "mean neighborhood radius" (257 m) is also printed and tells us the scale at which the degree selection is made.

In the present case the degree $k = 1$ is the best with an average rank of 1.83. Note that the sum of the ranks is $1 + 2 + 3 = 6$. The mean square error criterion also indicates $k = 1$ as the best choice, which increases our confidence. Such agreement of criteria for drift identification is not always achieved, but it is not uncommon.

Only five out of the seven possible fits were found valid and are listed in the second section of Figure 4.12. The discarded models are $C_0 \delta(h) - b_0|h|$ and $C_0 \delta(h) - b_0|h| + b_1|h|^3$.

The right-hand part of Figure 4.12 displays several criteria for selecting a model among the five candidates. The **Q** column shows the ratio $Q(\mathbf{b})/Q(\mathbf{0})$, while r_1 , r_2 , r , and $\hat{\rho}$ are given in the other columns. Here the model

$$C_0 = 0 \quad b_0 = 0.230 \quad b_1 = 0.960 \cdot 10^{-5}$$

stands out with a jackknife of 0.994, which turns out to also have the smallest ratio $Q(\mathbf{b})/Q(\mathbf{0}) = 0.761$. Figure 4.13 shows the results of cross-validation of the best four covariance models of Figure 4.12 using all data points. These tests confirm the above selection with $\text{m.s.e.} = 21.12 \text{ m}^2$ and $\text{m.s.s.e.} = 0.83$. Note that the model selected is not the one with the minimum m.s.e. because a good estimation of error variances is also required.⁹

Additional insight into the behavior of the selected model can be gained by inspection of the diagnostic plots shown in Figure 4.14. These plots are meant to detect the presence of residual structure not accounted for by the selected model and should be used *qualitatively* (e.g., the independence of Z^* and $Z - Z^*$ is established only for simple kriging and for a Gaussian SRF). Here we verify that posted standardized errors exhibit no particular clustering of large errors in specific areas; the scatterplot of Z versus Z^* shows no systematic effect hinting at a bias; the scatterplot of Z^* and standardized error indicates no major dependency of the error on the Z^* value (except for a decrease of the apparent scatter). Finally the histogram of standardized errors is symmetric about zero, with a normal shape but more squeezed in the middle and more tail-stretched than the Gaussian (for some reason, this shape is very frequent).

AUTOMATIC STRUCTURE IDENTIFICATION

1) IDENTIFICATION OF THE ORDER k

Degree	MEAN SQUARED ERRORS			RANKING OF THE TRIALS AVERAGE RANKS		
	Ring 1	Ring 2	Total	Ring 1	Ring 2	Total
2	342.3	349.1	345.8	1.92	1.85	1.88
1	119.7	137.1	128.8	1.82	1.85	1.83
0	298.2	293.3	295.7	2.26	2.31	2.28

Degree of the drift = 1

Counts: Ring 1 = 805 Ring 2 = 877 Total = 1682

Mean neighborhood radius = 256.7

2) IDENTIFICATION OF THE COVARIANCE

COVARIANCE COEFFICIENTS			EXPL/THEOR VARIANCE RATIOS				
C_0	$-b_0$	b_1	Q	r_1	r_2	r	ρ
3.060	0	0	0.8461	13.5092	12.4363	12.9867	13.0047
0	-0.3392	0	0.7637	1.8649	1.7122	1.7904	1.7929
0	0	0.3904 E-4	0.9241	0.4692	0.4348	0.4525	0.4531
0	-0.2296	0.9595 E-5	0.7609	1.0321	0.9518	0.9931	0.9944
2.715	0	0.2798 E-4	0.8091	0.6276	0.5814	0.6052	0.6061

Provisional covariance fit

Order $k = 1$ $C_0 = 0$ $b_0 = 0.2296$ $b_1 = 0.9595E-5$

FIGURE 4.12. An example of automatic structure identification by BLUEPACK.

Discussion

One of the limitations of the above procedure is the isotropy of polynomial GCs. The inference of anisotropic models was considered by Chilès (1978) but is impractical because it requires too many parameters. While in the stationary case it is not acceptable to ignore anisotropy, in the nonstationary case one can postulate that the anisotropies are captured by the filtered polynomials (the "drift"). For $k = 0$ the only filtered terms are constants which is a good reason, among others, to revert to a standard structural analysis. A stronger reason yet is that a model with a range may be more appropriate than a linear variogram.

The weakness of the automatic structure identification procedure lies in its lack of robustness against variables that do not fit the intrinsic hypotheses well: presence of a few isolated peaks, mixture of flat and jagged areas, erroneous or "outlying" data. These heterogeneities have a strong influence on the criterion $Q(C_0, \dots, b_k)$ to be minimized. As we have seen there are methods to

MODEL
CROSS-VALIDATION

COVARIANCE COEFFICIENTS			JACKKNIFE	ERROR STATISTICS	
C_0	$-b_0$	b_1	ρ	M.S.E.	M.S.S.E.
0	-0.2296	0.9595 E-5	0.9944	21.12	0.8276
2.715	0	0.2798 E-4	0.6061	20.50	1.3450
0	0	0.3904 E-4	0.4531	20.22	2.4520
0	-0.3392	0	1.7929	24.41	0.9697

Final-selection criteria:

- 1. M.S.E. close to its minimum value; and
- 2. M.S.S.E. close to 1, possibly in the interval $1 \pm 3 \text{ Sqrt}(2/N)$, i.e., [0.82, 1.18].

Final covariance fit

Order $k = 1$ $C_0 = 0$ $b_0 = 0.2296$ $b_1 = 0.9595\text{E-}5$

FIGURE 4.13. Cross-validation of the best four covariance models (ranked by increasing deviation of the jackknife from 1). Every data point has been estimated from 12 neighbors. m.s.e. (Mean Square Error) shows that the first three fits outperform the fourth one. m.s.s.e. (Mean Square Standardized Error) is the closest to 1 for fit 1, and lies in the tolerance interval for this fit only. Fit 1 is selected.

check the model, but only after the fact. Techniques of *robust regression* such as proposed by P. Huber (e.g., 1981) seem worth trying.

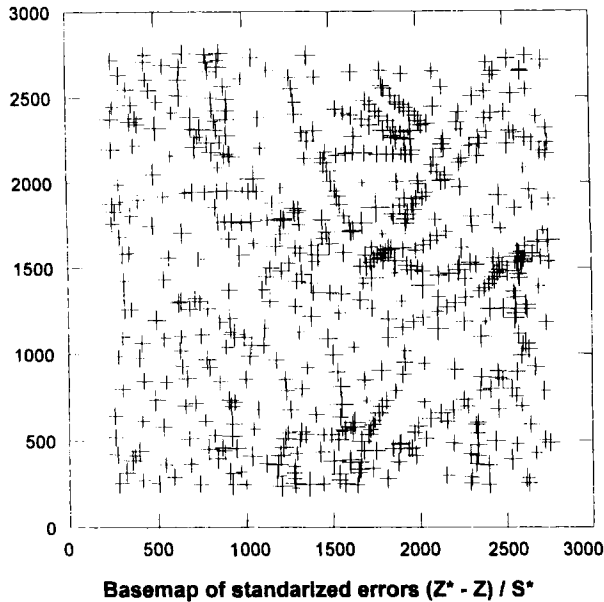
4.8.3. Variability of Polynomial Covariance Coefficients Estimates

Starks and Fang (1982) pointed out that in the case of the basic polynomial covariances $K_0(h) = -|h|$, $K_1(h) = |h|^3$, and $K_2(h) = -|h|^5$, there is a great deal of linear dependency between the predictor variables $K_0(\lambda)$, $K_1(\lambda)$, and $K_2(\lambda)$. This can cause instability in the estimation of the regression coefficients b_p . To evaluate the importance of the effect, they calculated the amount by which the variance of a regression coefficient is inflated due to correlation between the predictor variables. If the $Z(\lambda_i)^2$ were uncorrelated and with equal variance σ^2 , standard least squares theory says that the variance-covariance matrix of the regression estimator **b** is

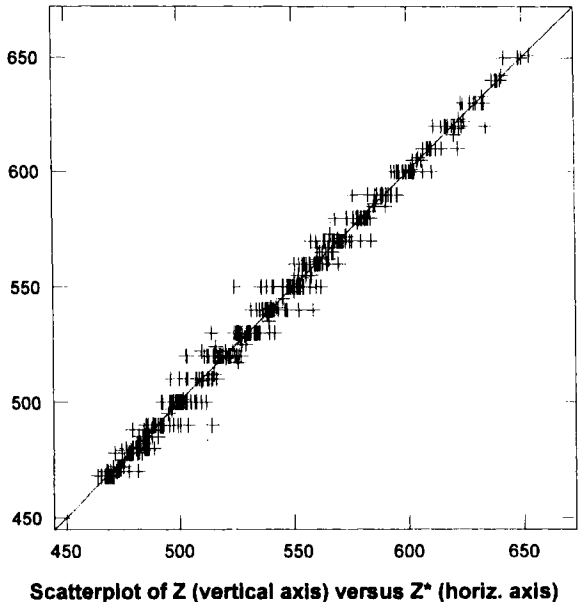
$$\text{Cov}(\mathbf{b}, \mathbf{b}) = \sigma^2 [T_{pq}]^{-1}$$

where $[T_{pq}]$ is the matrix with general term

$$T_{pq} = \sum_i w_i^2 K_p(\lambda_i) K_q(\lambda_i)$$



(a)



(b)

FIGURE 4.14. Diagnostic plots (topographic elevation, Noirétable): (a) posted standardized errors (the symbol size is proportional to the absolute value of the standardized error); (b) scatterplot of Z versus Z^* ; (c) histogram of standardized errors; (d) scatterplot of Z^* versus standardized error.

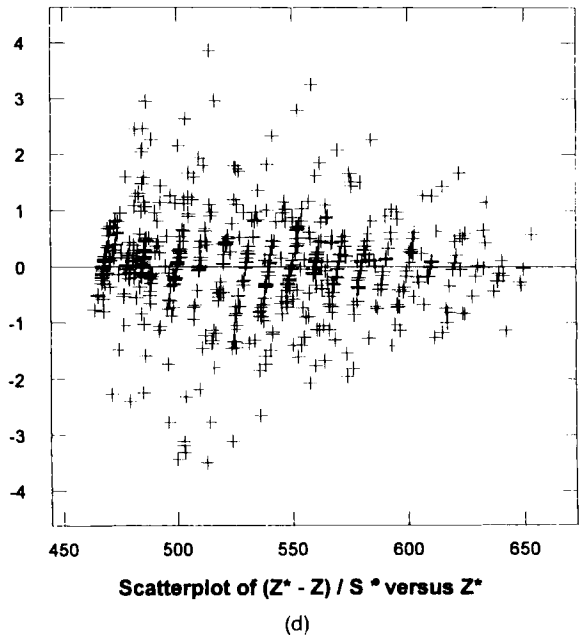
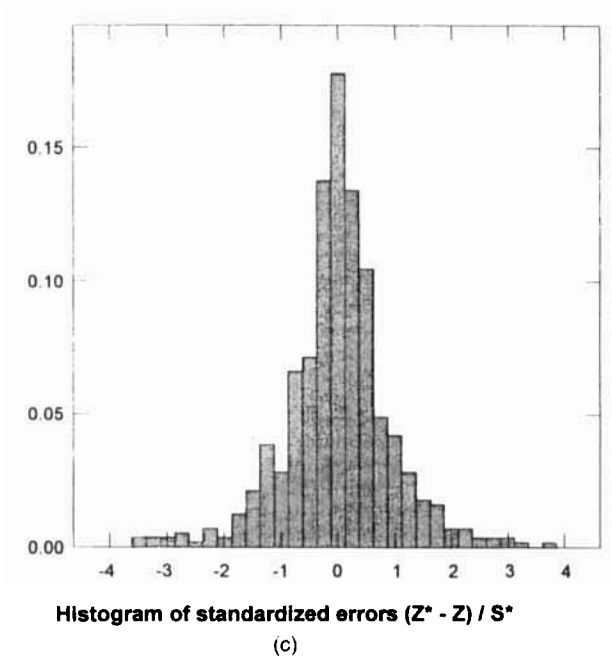


FIGURE 4.14. (Continued).

If the predictor variables $K_p(\lambda)$ have been centered and normalized so that $[T_{pq}]$ is in the form of a correlation matrix, one can compare the actual variance of \hat{b}_p with what it would be if all off-diagonal terms of $[T_{pq}]$ were zeros, that is, σ^2 . From an experiment based on a square array of 225 points and 200 measures for $k = 2$ involving 12 points each, Starks and Fang obtained the following "variance inflation factors"

$$\text{Var} \hat{b}_0 = 202 \quad \text{Var} \hat{b}_1 = 805 \quad \text{Var} \hat{b}_2 = 231$$

These figures correspond to unweighted least squares. If the weights $w_i^2 = 1/[K_0(\lambda_i)]^2$ are applied, the variances become

$$\text{Var} \hat{b}_0 = 49 \quad \text{Var} \hat{b}_1 = 279 \quad \text{Var} \hat{b}_2 = 115$$

Although weighting reduces the variances by a factor of about 4, 3, or 2, these variances remain much larger than 1.

These results should not lead one to conclude, however, that estimation is hopeless. Even if the estimation of individual coefficients b_p is difficult, and if numerically the estimators b_p "bounce back and forth," the covariance $K(h) = \sum b_p K_p(h)$ as a whole may be correctly estimated. This is a classic situation in regression: the fit has a meaning, whereas the coefficients do not (e.g., see Mosteller and Tukey, 1977, p. 319).

One should also keep in mind that the variance inflation factors result from conventional calculations involving only the design of the measures λ_i . In reality the dependent variables $Z(\lambda_i)^2$ are not uncorrelated, nor do they have equal variances. To compute the variance of the coefficient estimator \hat{b}_p , which is linear in the $Z(\lambda_i)^2$, the actual covariance structure of these $Z(\lambda_i)^2$ must be used.

This brings us back to the usual paradox of needing fourth-order moments while the second-order ones are not known. Yet it is not the most serious problem in the present case; the Gaussian assumption and the replacement of $K(h)$ by its estimator $\hat{K}(h)$ would provide orders of magnitude. What precludes the calculation of $\text{Var}(\hat{b}_p)$ in each application is the computational burden: typically several hundreds, sometimes thousands, of measures are used.

As a substitute an empirical study was made in one dimension with equi-spaced data (Chilès, 1978). Its conclusions cannot a priori be extended to scattered two-dimensional data, but they are consistent with the experience accumulated on automatic structure identification. The conclusions of this study are the following:

- If the covariance is predominantly linear with a nugget effect, all models can be fitted with fair precision.
- If the covariance is predominantly of type $|h|^3$, only models with a single other term may be fitted with acceptable precision.
- If the covariance model is predominantly of type $-|h|^5$, estimation of anything else than the coefficient of $|h|^5$ is meaningless.

NOTES

1. Matheron names *intrinsic drift* of the IRF- k the part of degree exactly $k + 1$ of this polynomial.
2. For true generality we must include the limits of all $Y_1(\varepsilon_x)$ where $\varepsilon_x \in \Lambda_k$ and with support in D .
3. If the assumption $\text{EZ}(\lambda) = 0$ were relaxed, $Z(\lambda)$ remaining of course stationary, the polynomial $Q(h)$ would be of degree $\leq 2k + 2$ instead of $2k$, the coefficient of h^{2k+2} not depending on 1_B .

For example, $\gamma(h) + a|h|^2/2$ is an admissible variogram if $E[Z(x+h) - Z(x)] = \langle a, h \rangle$. But we exclude such possibility by requiring $a = 0$.

4. All inequalities in this section assume an IRF- k with no intrinsic drift (i.e., $EZ(\lambda) = 0$ for $\lambda \in \Lambda_k$).
5. An analytic IRF-0 with an unbounded variogram, such as a Brownian motion convolved with a Gaussian density, cannot be locally stationary because it would then be stationary on the whole space, which is impossible since its variogram is not bounded.
6. A fallacious argument sometimes heard to justify the use of splines for mapping geological structures goes roughly like this: since geological materials have elastic properties and since thin-plate splines are a solution to the minimum stress problem, they are clearly physically correct. In reality, even if nature did create minimum stress surfaces, it did not force them to go through boreholes drilled millions of years later, namely the boundary conditions were completely different. The fact that beds are necessarily minimum stress surfaces is also dubious, geologically. As shown by Suppe (1985), and simplifying to the extreme, the shape of a fold is largely a function of its thickness, thicker layers having a longer wavelength. On the other hand, structural geologists have developed algorithms that take into account the geometric constraints imposed by the physics of deformation, most notably *balanced cross sections* (see note 10 of Section 3.6.2).
7. Here we find further justification for naming $\gamma(h)$ a variogram rather than a *semi*-variogram; else $\Gamma(h)$ would not be the generalized variogram, even less the generalized semi-variogram, but the M_k th of a generalized variogram!
8. The predecessor of their current geostatistical package named ISATIS.
9. The stationary covariance locally equivalent to the selected GC for $|h| \leq 525$ m (approximately the neighborhood diameter, cf. Table 4.2) allows an excellent fit of the sample variogram shown in Figure 2.27b, c. This reconciles the variographic and IRF- k approaches.

CHAPTER 5

Multivariate Methods

There are mountains here and hairy gooseberries don't grow on mountains.

D. H. McLain

5.1. INTRODUCTION

The information available on a natural phenomenon is rarely limited to the values assumed by a single variable over a set of sample points. In earth sciences applications there are at least four additional sources of information:

1. Numerical values of other variables possibly at other locations
2. Numerical values of the same variable at other time points
3. Physical relationships between variables
4. Application-specific knowledge

Most real studies involve more than one variable. Taking examples from the petroleum industry, we note that a borehole intersects several geological horizons, that dipmeter tools can measure dip and strike in addition to depth, that reservoirs are described by several parameters such as porosity, permeability, and fluid saturations, all more or less related and dependent on depth. Examples like these can easily be found in all domains of application and call for a multivariate generalization of kriging which is developed in the present chapter under the name of *cokriging*. Despite an increased complexity mainly due to notations, there is little novelty from a theoretical standpoint. It suffices to pool all the random variables representing the data into a single Hilbert space and project the estimated quantity onto that space. When drifts are present, one must consider the possible functional dependencies between them, but this is rather straightforward once the mechanism for deriving the equations is understood. We will not attempt to be exhaustive. We will explain the general principles and concentrate on a few interesting problems. The real theoretical challenge is this: How can we model the relationships between the different

variables in a globally coherent manner? And the practical challenge is: How can we infer this model from an often undersampled data set?

A particular class of phenomena, mostly those related to the atmosphere (meteorology, pollution), are studied from data distributed in both space and time. While the general cokriging framework would be applicable, some specific models have been developed that account for the particular nature of the indexing quantity, that is, time instead of just some nominal index i . This is still very much a research area, stimulated by questions such as: Do our data show evidence of a global warming? A few representative space-time models are presented in the last section.

Earth sciences variables are often related by equations of varying degrees of empiricism. At one extreme are the purely empirical fits, such as porosity-permeability formulas. In the middle are empirical laws with wide (though discussed) applicability, such as Archie's formula relating the resistivity of a clean water bearing formation with its porosity. At the other extreme are the laws of physics, such as those governing fluid flow in a porous medium, or wave propagation in a stratified earth model. Measurements are related through stochastic partial differential equations whose coefficients (the transmissivities) are regionalized variables. The study of random functions subject to physical equations is a frontier area explored in Chapter 8.

Knowledge of the subject matter is obviously important to fill in the gaps between the samples. To take a geological example, knowing that we are dealing with a sand bar, a channel, or a reef increases our ability to predict the location, geometry, and orientation of a petroleum reservoir. But such knowledge is largely qualitative rather than quantitative and difficult to formalize and integrate in a geostatistical model. Object-based simulations (Section 7.8) are a step in this direction, and here we will take two more: the use of dip information and the input of a guess field or an external drift. Short of random genetic models that would capture the application-specific knowledge, the multivariate methods presented in this chapter allow one to constrain geostatistical estimates by all numerical information already available.

5.2. NOTATIONS AND ASSUMPTIONS

Notations are the main difficulty with cokriging theory. One quickly becomes short of space and names for indexes. We have tried to emphasize clarity by using a mix of Latin and Greek subscripts and matrix notations.

- We are considering p simultaneous random functions $Z_i(x)$, indexed by Latin subscripts i ranging in the set $I = \{1, \dots, p\}$ and defined over a domain D of \mathbb{R}^n

$$\{Z_i(x) : x \in D \subset \mathbb{R}^n\}$$

Each $Z_i(x)$ is sampled over a set S_i of $N_i > 0$ points, referenced as usual with Greek indexes—which makes the distinction with function indexes

very clear

$$S_i = \{x_\alpha \in D : Z_i(x_\alpha) \text{ known}\}$$

For simplicity we use the same notation x_α for generic data points, but *the sample sets S_i are in general different for the different indexes*. They may be pairwise disjoint and some may even be empty. In fact the relationships between the different sets is one of the key aspects of a multivariate estimation problem.

- As with universal kriging we assume that each function $Z_i(x)$ may have a drift $m_i(x)$ that can be represented as a linear combination with unknown coefficients of known basis drift functions $f_i^l(x)$, $l = 0, \dots, L_i$, which in general may be different for the different RFs:

$$EZ_i(x) = m_i(x) = \sum_l a_{il} f_i^l(x)$$

Cross-covariances between $Z_i(x_\alpha)$ and $Z_j(x_\beta)$ will be denoted by $\sigma_{ij}(x_\alpha, x_\beta)$

$$\sigma_{ij}(x_\alpha, x_\beta) = \text{Cov}[Z_i(x_\alpha), Z_j(x_\beta)] = E[Z_i(x_\alpha)Z_j(x_\beta)] - m_i(x_\alpha)m_j(x_\beta)$$

At this point we make no assumption on covariances $\sigma_{ij}(x, y)$ other than the fact that they exist and are mathematically consistent. In the stationary case these covariances are of the form $C_{ij}(h)$ with $h = y - x$.

- To avoid carrying quadruple indexes such as above, it is useful to define vector notations. There are two ways of doing that. The natural one is to consider a vector-valued random function $\mathbf{Z}(x)$ (denoted with a bold \mathbf{Z}) whose components are the p RFs, namely

$$\mathbf{Z}(x) = (Z_1(x), Z_2(x), \dots, Z_p(x))'$$

However, this formulation implicitly assumes an equal sampling of all components, which is usually not the case. The other way, which we prefer to use here, is to pool all observations relative to the i th variable into an N_i -vector

$$\mathbf{Z}_i = (Z_i(x_1), Z_i(x_2), \dots, Z_i(x_{N_i}))'$$

Accordingly an N_i -vector of weights is defined by $\lambda_i = (\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{iN_i})'$ so that

$$\lambda_i' \mathbf{Z}_i = \sum_\alpha \lambda_{i\alpha} Z_i(x_\alpha)$$

the summation in α being short for “sum over all points $x_\alpha \in S_i$.”

The covariance matrix between \mathbf{Z}_i and \mathbf{Z}_j data is the $N_i \times N_j$ matrix with general term $\sigma_{ij}(x_\alpha, x_\beta)$ — $C_{ij}(x_\beta - x_\alpha)$ in the stationary case— $x_\alpha \in S_i$,

$x_{\beta} \in S_j$. The letter **C** will be used for covariance matrices instead of Σ to avoid confusion with the summation sign:

$$\mathbf{C}_{ij} = \text{Cov}(\mathbf{Z}_i, \mathbf{Z}_j) = \mathbf{E}(\mathbf{Z}_i \mathbf{Z}_j') - \mathbf{E}(\mathbf{Z}_i) \mathbf{E}(\mathbf{Z}_j')$$

In general, \mathbf{C}_{ij} is not a symmetric matrix (even when it is a square matrix), however, from the above $\mathbf{C}_{ji} = \mathbf{C}_{ij}'$.

The vector of mean values can be written as

$$\mathbf{E} \mathbf{Z}_i = \mathbf{F}_i \mathbf{a}_i$$

where $\mathbf{F}_i = [f_i^l(x_\alpha)]$ is the matrix of drift functions arranged by columns l and rows α ($x_\alpha \in S_i$) and $\mathbf{a}_i = (a_{il})$ is the column vector of drift coefficients for the RF Z_i , just as in standard UK theory.

- The estimated variable corresponds to the value $i = 1$ of the subscript and will be called the *primary variable*, by opposition to the *secondary variables* with indexes $i \neq 1$. We will focus here on the estimation of the primary variable at a single target point x_0 , knowing that any linear functional can be linearly deduced from such point estimates. Our objective is thus

$$Z_0 = Z_1(x_0)$$

Its mean value is

$$\mathbf{E} Z_0 = \sum_l a_{1l} f_1^l(x_0) = \mathbf{a}_1' \mathbf{f}_{10}$$

where $\mathbf{f}_{10} = (f_1^l(x_0))$ is the vector of drift functions values at the point x_0 and $\mathbf{a}_1 = (a_{1l})$ is the vector of drift coefficients for the RF Z_1 .

The covariance between the data vector \mathbf{Z}_i and the objective Z_0 is the N_i -vector (lowercase)

$$\mathbf{c}_{i0} = \text{Cov}(\mathbf{Z}_i, Z_0) = \mathbf{E}(\mathbf{Z}_i Z_0) - \mathbf{E}(\mathbf{Z}_i) \mathbf{E}(Z_0)$$

and the variance of the objective is

$$c_{00} = \text{Cov}(Z_0, Z_0) = \text{Var}(Z_0)$$

- Finally the cokriging estimator is defined as an optimum linear combination of *all* available data

$$Z^{**} = \sum_{i \in I} \sum_{\alpha \in S_i} \lambda_{i\alpha} Z_i(x_\alpha) = \sum_{i \in I} \lambda_i' \mathbf{Z}_i$$

The double asterisk (**) is introduced to emphasize the difference with the kriging estimator; and the subscript CK will be used for the variance.

5.3. SIMPLE COKRIGING

Simple cokriging corresponds to the case where the means of the estimating and estimated RFs are known. Clearly to eliminate bias, it suffices to subtract these means and work with the zero mean RFs

$$Y_i(x) = Z_i(x) - m_i(x)$$

In the rest of this section we will consider that all RFs have zero means.

5.3.1. Derivation of the Equations

In order to estimate $Z_0 = Z_1(x_0)$ from realizations of the RFs $Z_i(x)$ known over the sample sets S_i , $i \in I$, we consider a linear estimator of the form

$$Z^{**} = \sum_i \sum_{\alpha} \lambda_{i\alpha} Z_i(x_{\alpha}) = \sum_i \lambda_i' \mathbf{Z}_i$$

Since the mean is zero, the mean square error is equal to the error variance

$$\begin{aligned} E(Z^{**} - Z_0)^2 &= E \left[\sum_i \sum_j \lambda_i' \mathbf{Z}_i \mathbf{Z}_j' \lambda_j - 2 \sum_i \lambda_i' \mathbf{Z}_i Z_0 + Z_0 Z_0 \right] \\ &= \sum_i \sum_j \lambda_i' \mathbf{C}_{ij} \lambda_j - 2 \sum_i \lambda_i' \mathbf{c}_{i0} + c_{00} \end{aligned}$$

Canceling the partial derivatives with respect to λ_i yields the cokriging system:

Simple Cokriging System

$$\sum_{j=1}^p \mathbf{C}_{ij} \lambda_j = \mathbf{c}_{i0} \quad i = 1, \dots, p \quad (5.1)$$

The cokriging variance is given by

Simple Cokriging Variance

$$\sigma_{CK}^2 = E(Z^{**} - Z_0)^2 = c_{00} - \sum_{i=1}^p \lambda_i' \mathbf{c}_{i0} \quad (5.2)$$

5.3.2. Cokriging as a Projection

Just like its univariate brother, simple cokriging admits of a geometric interpretation in terms of projection onto a Hilbert space. Following Matheron (1970), we consider RFs defined over the product space

$$E = \mathbb{R}^n \times I$$

of \mathbb{R}^n by the set $I = \{1, \dots, p\}$ of indexes. An element of E is a pair (x, i) where $x \in \mathbb{R}^n$ and $i \in I$. The p random functions $Z_i(x)$ can be regarded as a single RF $Z(x, i)$ on this product space. Similarly the sampling set of the RF $Z(x, i)$ is the collection of pairs (x, i) such that the point x belongs to S_i :

$$S = \{(x, i) : x \in \mathbb{R}^n, i \in I, x \in S_i\}$$

Now consider the Hilbert subspace H_S generated by linear combinations of zero mean finite variance random variables $\{Z(x, i) : (x, i) \in S\}$ (i.e., the sampled ones) and all their L^2 limits, with the usual scalar product $\langle X, Y \rangle = EXY$. In explicit form the equations (5.1) can be written as

$$\text{Cov}(Z_i(x_\alpha), Z^{**}) = \text{Cov}(Z_i(x_\alpha), Z_0) \quad \forall i \in I, \quad x_\alpha \in S_i$$

which translates into

$$\langle Z^{**} - Z_0, X \rangle = 0 \quad \forall X \in H_S \quad (5.3)$$

showing that the cokriging estimator is the projection of Z_0 onto H_S . Since $Z^{**} \in H_S$, the cokriging error is orthogonal to the cokriging estimator

$$\langle Z^{**} - Z_0, Z^{**} \rangle = 0$$

and the smoothing relationship holds

$$\text{Var} Z_0 = \text{Var} Z^{**} + \sigma_{CK}^2$$

This geometric formulation may appear as an unnecessary theoretical complication, but it is not. Geometric reasoning supports intuition and straightforward orthogonality results can shortcut cumbersome algebra. To take a specific example, consider, following Myers (1983), the problem of estimating a given linear combination of variables

$$W(x) = \sum_{i=1}^p w_i Z_i(x)$$

(e.g., the sum over several geological horizons or the sum of values for different minerals). When all variables are equally sampled, we have the choice of

either direct estimation of $W_0 = W(x_0)$ by kriging or estimation by cokriging. How do they compare? For simplicity let us assume here that all means are zero. For every variable i the cokriging estimator of $Z_i(x_0)$ satisfies (5.3) so that by linearity the linear combination

$$W^{**} = \sum_{i=1}^p w_i Z_i^{**}(x_0)$$

also satisfies (5.3) and is therefore *the* cokriging estimator of W_0 . Consider now the straight kriging estimator W^* . The difference $W^* - W^{**} \in H_S$ so that

$$\langle W^{**} - W_0, W^* - W^{**} \rangle = 0$$

and

$$\|W_0 - W^*\|^2 = \|W_0 - W^{**}\|^2 + \|W^{**} - W^*\|^2$$

The cokriging estimator has a smaller mean square error than the straight kriging estimator. Note, in passing, that cokriging ensures consistency: the estimate of the sum is the sum of the estimates, which would not be true if each variable Z_i were estimated separately by kriging.

5.4. UNIVERSAL COKRIGING

The simplicity and generality of simple cokriging is marred by the introduction of drifts because, to derive the equations, we have to specify if and how the means $m_i(x)$ of the different variables are related. Three cases will be considered (1) algebraically independent drifts, (2) linearly dependent drifts, or (3) mixed case. The fourth possible case, nonlinear functional relationships between drifts, cannot be handled by universal cokriging; transformations and reparametrizations are required that are not discussed here.

Given the diversity of multivariate situations our intent here is not to be exhaustive but to lay out the general principles that will enable the user to develop the cokriging equations for a particular problem. We will focus on point estimation. The equations for block estimation and for drift estimation can be established in a similar straightforward manner.

5.4.1. Theory for Algebraically Independent Drifts

Here we assume that each RF $Z_i(x)$ has a drift of its own and that the drift coefficients of the different variables are not related. As before, we wish to estimate $Z_0 = Z_1(x_0)$ using an estimator of the form

$$Z^{**} = \sum_i \lambda'_i Z_i$$

We have

$$E(Z^{**} - Z_0) = \sum_i \lambda'_i \mathbf{F}_i \mathbf{a}_i - \mathbf{f}'_{10} \mathbf{a}_1$$

and want to make this zero for all drift coefficient vectors \mathbf{a}_i and \mathbf{a}_1 . There are two cases to consider:

1. $S_1 = \{\emptyset\}$. Then the drifts of Z^{**} and Z_0 have no coefficient in common and the above can hold as an identity only if $\mathbf{f}_{10} = \mathbf{0}$. But, in general, $\mathbf{f}_{10} \neq \mathbf{0}$ (typically there is at least a 1) so that, except for particular cases, estimation is impossible if there is no observation of the primary variable.
2. $S_1 \neq \{\emptyset\}$. Then universal unbiasedness is achieved if and only if

$$\lambda'_i \mathbf{F}_i = \mathbf{0} \quad \text{for } i \neq 1 \quad \lambda'_1 \mathbf{F}_1 = \mathbf{f}'_{10}$$

These conditions can be rewritten in the synthetic form

$$\lambda'_i \mathbf{F}_i = \delta_{i1} \mathbf{f}'_{10} \quad \text{where} \quad \delta_{i1} = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{if } i \neq 1 \end{cases}$$

Now

$$\text{Var}(Z^{**} - Z_0) = \sum_i \sum_j \lambda'_i \mathbf{C}_{ij} \lambda_j - 2 \sum_i \lambda'_i \mathbf{c}_{i0} + c_{00}$$

Minimizing this variance subject to the unbiasedness constraints leads to the system with Lagrange parameters vector $\boldsymbol{\mu}_i = (\mu_{il})'$, $l = 0, \dots, L_i$:

Universal Cokriging System

$$\begin{cases} \sum_{j=1}^p \mathbf{C}_{ij} \lambda_j + \mathbf{F}_i \boldsymbol{\mu}_i = \mathbf{c}_{i0} & i = 1, \dots, p \\ \mathbf{F}'_i \boldsymbol{\mu}_i = \mathbf{f}_{10} \delta_{i1} & i = 1, \dots, p \end{cases} \quad (5.4)$$

The cokriging variance is given by

Cokriging Variance

$$\sigma_{\text{CK}}^2 = E(Z^{**} - Z_0)^2 = c_{00} - \sum_{i=1}^p \lambda'_i \mathbf{c}_{i0} - \boldsymbol{\mu}'_1 \mathbf{f}_{10} \quad (5.5)$$

(Note: If $E Z_l(x) = 0$ the corresponding matrix \mathbf{F}_i is absent from the system.)

To take a specific example, if $p = 2$, the structure of the cokriging system is as follows:

$$\underbrace{\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{F}_1 & \mathbf{0} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \mathbf{0} & \mathbf{F}_2 \\ \dots & \dots & \dots & \dots \\ \mathbf{F}'_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}'_2 & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \mu_1 \\ \mu_2 \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} \mathbf{c}_{10} \\ \mathbf{c}_{20} \\ \dots \\ \mathbf{f}_{10} \\ \mathbf{0} \end{bmatrix}}_{\mathbf{B}} \quad (5.6)$$

where on the right-hand side

$$\mathbf{c}_{10} = (\text{Cov}[Z_1(x_\alpha), Z_1(x_0)], x_\alpha \in S_1)'$$

$$\mathbf{c}_{20} = (\text{Cov}[Z_2(x_\beta), Z_1(x_0)], x_\beta \in S_2)'$$

are the vectors of covariances between \mathbf{Z}_1 and Z_0 , and \mathbf{Z}_2 and Z_0 . Also

$$\sigma_{CK}^2 = c_{00} - \mathbf{X}'\mathbf{B} \quad (5.7)$$

Unlike kriging weights which do not depend on the variance level, cokriging weights depend on the *relative* scale of the different variables. If \mathbf{Z}_1 is the variable of interest and a change of units turns \mathbf{Z}_i into $\alpha_i \mathbf{Z}_i$, the weights λ_1 remain unchanged while λ_i turns into $(\alpha_1/\alpha_i)\lambda_i$. This shows the importance of an accurate fit of the relative levels of the different covariances (all of them being possibly defined up to a multiplicative factor).

Existence and Uniqueness of a Solution

With obvious notations the cokriging matrix \mathbf{A} in (5.6) is of the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{F} \\ \mathbf{F}' & \mathbf{0} \end{bmatrix}$$

Notice that \mathbf{C} is symmetric owing to the symmetry property $\mathbf{C}_{ji} = \mathbf{C}'_{ij}$. In fact the cokriging system has the same structure as the standard UK system.

If the multivariate covariance model is strictly positive definite (discussed later) and if there is no data duplication, the matrix \mathbf{C} is nonsingular. Then, just as in the UK theory, the matrix \mathbf{A} is nonsingular if and only if the columns of \mathbf{F} are linearly independent, namely if

$$\mathbf{F}\mathbf{a} = \mathbf{0} \quad \Rightarrow \quad \mathbf{a} = \mathbf{0}$$

Given the block diagonal structure of \mathbf{F} , this is equivalent to

$$\begin{cases} \mathbf{F}_1 \mathbf{a}_1 = \mathbf{0} & \Rightarrow & \mathbf{a}_1 = \mathbf{0} \\ \vdots & & \vdots \\ \mathbf{F}_p \mathbf{a}_p = \mathbf{0} & \Rightarrow & \mathbf{a}_p = \mathbf{0} \end{cases}$$

that is, every \mathbf{F}_i is of full rank equal to its number of columns.

Each matrix \mathbf{F}_i has N_i rows (the number of data points in S_i) and $L_i + 1$ columns (the number of basis drift functions). For the rank of \mathbf{F}_i to be $L_i + 1$, it is required that $N_i \geq L_i + 1$. For a secondary variable the practical minimum is in fact $L_i + 2$, or else the variable plays no role. Indeed, if $N_i = L_i + 1$, the matrix \mathbf{F}_i is then a square nonsingular matrix, and the constraint $\mathbf{F}_i' \boldsymbol{\lambda}_i = \mathbf{0}$ entails $\boldsymbol{\lambda}_i = \mathbf{0}$.

Discussion

1. It is intuitively obvious that if all variables are mutually uncorrelated (orthogonal) and if their means are algebraically independent, the secondary variables are of no help for the estimation of the primary variable. Indeed in this case the cokriging solution coincides with the UK estimator based on the primary variable alone.
2. At the same time it is difficult to conceive of an actual situation in which two variables are allowed to drift away independently and yet carry useful information about each other: the variations of the unknown drifts would dominate those of the residuals. The only case of practical interest for algebraically independent drifts is when all the means are constants, namely *ordinary cokriging*.
3. There must be at least enough data points of the primary (i.e., estimated) variable to be able to estimate its mean. However, if the primary variable has a zero mean, or equivalently if its mean is known, there is no minimum of points. In this particular case estimation of the primary variable is possible on the basis of secondary variables only.
4. Being constrained by $\mathbf{F}_i' \boldsymbol{\lambda}_i = \mathbf{0}$, the secondary variables contribute to the estimator only as corrections (they only involve the true residuals of Z_i). Estimation will be improved only if the secondary variables are strongly correlated with the objective. Also, as noted above, the weights on secondary variables are nonzero only if their number exceeds the number of basis drift functions. For example, the weight assigned to a single secondary data point with an unknown mean is necessarily zero (but not so if the mean is known).
5. If we can construct the cokriging estimators of a collection of variables $\{Z_k(x_u)\}$ from some fixed data set S , the cokriging estimator of a linear

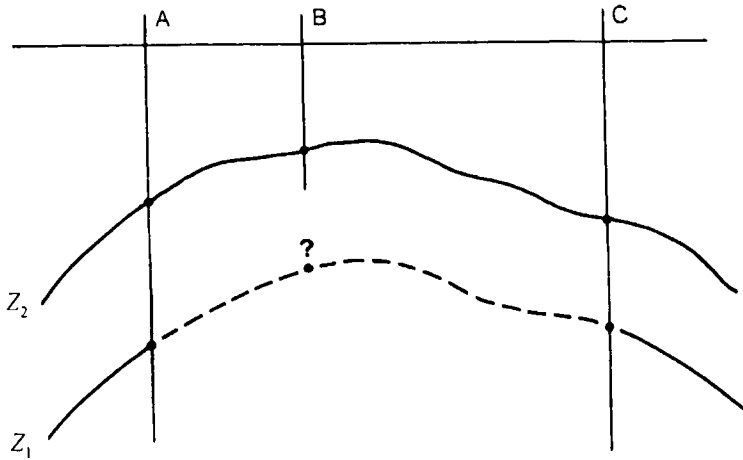


FIGURE 5.1. Estimation of the depth of Z_1 at well B from known depths of Z_1 and Z_2 . Wells A and C have been drilled through both surfaces, but well B is incomplete.

combination

$$Z_0 = \sum_k \sum_u w_{ku} Z_k(x_u)$$

is simply the linear combination of the individual cokriging estimators

$$Z^{**} = \sum_k \sum_u w_{ku} Z_k^{**}(x_u)$$

This solution is obtained directly by solving the system (5.4) for a right-hand side, which is the same linear combination of the individual right-hand sides. The converse, however, is not true: it may be possible to estimate a linear combination (e.g., one that filters the drift) without being able to estimate its components individually.

5.4.2. A Worked Example of Ordinary Cokriging: Incompletely Drilled Wells

In order to demonstrate the use of ordinary cokriging equations, we will work out a synthetic example that is also of practical interest. We consider two variables Z_1 and Z_2 with constant but unknown means m_1 and m_2 , for example, the depths (counted positively downward) of two geological surfaces where $Z_1 \geq Z_2$. If depths are determined from vertical wells, the coordinates of samples points are the same for the two surfaces, except that in some of the wells drilling may have been interrupted before hitting the deeper surface.

The sample sets are related by $S_1 \subset S_2$. If wells are deviated, the sampling points are different for the two surfaces, and we may have $S_1 \cap S_2 = \emptyset$. For the example we consider the simple situation of Figure 5.1 with three vertical wells located at points A, B, C and well B drilled only

down to the shallower surface. We wish to estimate the depth Z_1 at location x_B using all available depth values.

The cokriging estimator of $Z_1(x_B)$ is of the form

$$Z_1^{**}(x_B) = \lambda_{1A}Z_1(x_A) + \lambda_{1C}Z_1(x_C) + \lambda_{2A}Z_2(x_A) + \lambda_{2B}Z_2(x_B) + \lambda_{2C}Z_2(x_C)$$

Letting $p = 2$ and $x_0 = x_B$ the general cokriging equations (5.4) become

$$\begin{bmatrix} C_{11}(0) & C_{11}(AC) & C_{12}(0) & C_{12}(AB) & C_{12}(AC) & 1 & 0 \\ C_{11}(CA) & C_{11}(0) & C_{12}(CA) & C_{12}(CB) & C_{12}(0) & 1 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ C_{21}(0) & C_{21}(AC) & C_{22}(0) & C_{22}(AB) & C_{22}(AC) & 0 & 1 \\ C_{21}(BA) & C_{21}(BC) & C_{22}(BA) & C_{22}(0) & C_{22}(BC) & 0 & 1 \\ C_{21}(CA) & C_{21}(0) & C_{22}(CA) & C_{22}(CB) & C_{22}(0) & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_{1A} \\ \lambda_{1C} \\ \dots \\ \lambda_{2A} \\ \lambda_{2B} \\ \lambda_{2C} \\ \dots \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} C_{11}(AB) \\ C_{11}(CB) \\ \dots \\ C_{21}(AB) \\ C_{21}(0) \\ C_{21}(CB) \\ \dots \\ 1 \\ 0 \end{bmatrix} \quad (5.8)$$

where AB is shorthand for $x_B - x_A$.

Direct covariances $C_{11}(h)$ and $C_{22}(h)$ are symmetric, whereas cross-covariances are not but satisfy $C_{12}(AB) = C_{21}(BA)$ so that the matrix of the cokriging system is indeed symmetric.

To go further and get interpretable results, we must now particularize the covariance models. We will consider two cases: the additive model and the proportional covariance model.

Additive Model

In this model the thickness $H(x)$ of the layer delimited by the two surfaces is assumed uncorrelated with the top depth $Z_2(x)$:

$$Z_1(x) = Z_2(x) + H(x) \quad \text{with} \quad \text{cov}[Z_2(x), H(y)] = 0 \quad \forall x, y$$

The covariance structure becomes

$$\begin{cases} C_{12}(h) = C_{21}(h) = C_{22}(h) = C(h) \\ C_{11}(h) = C(h) + K(h) \end{cases}$$

where $K(h)$ is the covariance of $H(x)$. Substituting into (5.8), we find that the solution satisfies

$$\lambda_{2A} = -\lambda_{1A} \quad \lambda_{2B} = 1 \quad \lambda_{2C} = -\lambda_{1C} \quad \mu_2 = 0$$

so that

$$Z_1^{**}(x_B) = Z_2(x_B) + \lambda_{1A}[Z_1(x_A) - Z_2(x_A)] + \lambda_{1C}[Z_1(x_C) - Z_2(x_C)]$$

where λ_{1A} and λ_{1C} are solutions of the system

$$\begin{cases} K_{AA}\lambda_{1A} + K_{AC}\lambda_{1C} + \mu_1 = K_{AB} \\ K_{CA}\lambda_{1A} + K_{CC}\lambda_{1C} + \mu_1 = K_{CB} \\ \lambda_{1A} + \lambda_{1C} = 1 \end{cases}$$

which coincides with the ordinary kriging system of $H(x_B)$ from $H(x_A)$ and $H(x_C)$.

In this particular case cokriging amounts to the obvious solution of estimating the thickness $H(x_B)$ by univariate kriging from $H(x_A)$ and $H(x_C)$ and adding the estimate to $Z_2(x_B)$

$$Z_1^{**}(x_B) = Z_2(x_B) + H^*(x_B)$$

The cokriging variance computed from (5.5),

$$\sigma_{CK}^2 = K_{00} - \lambda_{1A}K_{AB} - \lambda_{1C}K_{CB} - \mu_1$$

coincides with the OK variance of $H(x_B)$.

Proportional Covariance Model

In this model all direct and cross-covariances are proportional to the same covariance function $C(h)$:

$$\begin{cases} C_{11}(h) = C_{22}(h) = C(h) \\ C_{12}(h) = C_{21}(h) = \rho C(h) \end{cases} \quad -1 < \rho < 1$$

That this is a valid multivariate covariance model will be shown later. Substituting into (5.8), we see that the last three covariance equations are satisfied identically with $\mu_2 = 0$ whenever

$$\lambda_{2A} = -\rho\lambda_{1A} \quad \lambda_{2B} = \rho \quad \lambda_{2C} = -\rho\lambda_{1C}$$

The remaining equations then become

$$\begin{cases} C_{AA}\lambda_{1A} + C_{AC}\lambda_{1C} + \frac{\mu_1}{1-\rho^2} = C_{AB} \\ C_{CA}\lambda_{1A} + C_{CC}\lambda_{1C} + \frac{\mu_1}{1-\rho^2} = C_{CB} \\ \lambda_{1A} + \lambda_{1C} = 1 \end{cases}$$

One recognizes the OK system for estimating $Z_1(x_B)$ from $Z_1(x_A)$ and $Z_1(x_C)$ and also, since $C_{11}(h) = C_{22}(h)$, for estimating $Z_2(x_B)$ from $Z_2(x_A)$ and $Z_2(x_C)$ without using $Z_2(x_B)$, of course. Finally the cokriging estimator can be written in the form

$$Z_1^{**}(x_B) = Z_1^*(x_B) + \rho[Z_2(x_B) - Z_2^*(x_B)] \quad (5.9)$$

where $Z_1^*(x_B)$ and $Z_2^*(x_B)$ are the OK estimators of $Z_1(x_B)$ and $Z_2(x_B)$ based on the observations of Z_1 alone and Z_2 alone at points A and C.

The result is strikingly simple and intuitive: cokriging improves the ordinary kriging estimator of Z_1 by adding a correction that is ρ times the kriging error observed for Z_2 . The improvement

of the estimator is evidenced by the smaller variance

$$\sigma_{CK}^2 = \sigma_K^2(1 - \rho^2)$$

Under this model cokriging estimation can be decoupled into two identical kriging problems. The same results are found if simple, instead of ordinary, kriging and cokriging are used.

5.4.3. Collocated Cokriging

In the previous example both variables are sampled equally (are co-located) except that the secondary variable is also available at the estimated point—where the primary variable of course is not. A simplified implementation of cokriging, introduced by Xu et al. (1992) under the name of *collocated cokriging*, is to retain only the secondary datum that is co-located with the estimated point, thus using a total of $N + 1$ values. The other method uses $2N + 1$ values and we propose to call it *multi-collocated cokriging*. Both methods have become popular as techniques to simplify cokriging.

Collocated cokriging is implemented as simple cokriging, for otherwise the weight on the secondary data point would be zero (although the primary variable could have an unknown mean). The method is usually applied in conjunction with the special covariance structure

$$\rho_{12}(h) = \rho_{12}(0)\rho_{11}(h) \quad (5.10)$$

where $\rho_{11}(h)$ is the correlogram of Z_1 and $\rho_{12}(h)$ the cross-correlogram of Z_1 and Z_2 . No assumption is made for the correlogram of Z_2 because this is not needed when a single secondary value is used. A property of this model, in the Gaussian case with known means, is the conditional independence of $Z_1(x_0)$ and $Z_2(x)$ when $Z_1(x)$ is given, so that $Z_1(x)$ “screens” $Z_2(x)$ from $Z_1(x_0)$. For this reason it has been named a “Markov-type” model.

Under the “Markov” model the cokriging estimator is a variance-weighted linear combination of the simple kriging and linear regression estimators of $Z_1(x_0)$ from Z_1 and Z_2 data alone

$$Z_1^{**}(x_0) = \frac{(1 - \rho^2)Z_1^*(x_0) + \sigma_{SK}^2 \rho Z_2(x_0)}{(1 - \rho^2) + \rho^2 \sigma_{SK}^2} \quad (5.11)$$

$$\sigma_{CK}^2 = \sigma_{SK}^2 \frac{(1 - \rho^2)}{(1 - \rho^2) + \rho^2 \sigma_{SK}^2}$$

In this formula $\rho = \rho_{12}(0)$ and Z_1 and Z_2 have zero mean and unit variance. Necessarily $\sigma_{SK}^2 \leq 1$ with equality when the kriging estimate is zero (the mean), in which case the cokriging estimate reduces to the linear regression estimate $\rho Z_2(x_0)$. Formula (5.11) can also be established by Bayesian analysis (Doyen et al., 1996).

The screening property of (5.10) should not lead to the false conclusion that the estimation of $Z_1(x_0)$ cannot benefit from the knowledge of Z_2 at locations where Z_1 is known. This screening takes place when Z_1 and Z_2 data are available *at the same locations* (see Section 5.6.4) but, surprisingly, vanishes when a secondary datum is added at x_0 (the cause is a “relay effect” with the other Z_2 values). In formula (5.9), which conforms to model (5.10), we can see that all data do contribute to the estimation. Retaining only one secondary point would amount to neglecting thickness data, a clear loss of information.

The decoupling of estimations observed in (5.9) for multi-located cokriging turns out to be always valid in the case of a proportional covariance model, even when means are unknown, provided that the same basis drift functions (with algebraically independent coefficients) are used for Z_1 and Z_2 . Making the dependence on scales explicit the formulas are

$$\begin{aligned} Z_1^{**}(x_0) &= Z_1^*(x_0) + \rho \frac{\sigma_1}{\sigma_2} [Z_2(x_0) - Z_2^*(x_0)] \\ E[Z_1^{**}(x_0) - Z_1(x_0)]^2 &= \sigma_{K1}^2 (1 - \rho^2) \end{aligned} \quad (5.12)$$

where $Z_1^*(x_0)$ and $Z_2^*(x_0)$ are kriging estimators (same weights), σ_{K1}^2 the kriging variance of $Z_1(x_0)$, σ_1 and σ_2 the standard deviations of Z_1 and Z_2 , and ρ their correlation coefficient. Cokriging improves on kriging by the addition of a correction term equal to the linear regression estimate of the kriging error $Z_1 - Z_1^*$ from $Z_2 - Z_2^*$. Cokriging “learns” from kriging errors. Naturally, multi-located cokriging always outperforms single located cokriging, as can be checked on variances in the particular cases of (5.11) and (5.12) applied with known drifts.

In order to verify that (5.12) is the solution of the general cokriging equations (5.4) we write them explicitly in the form

$$\text{Cov}[Z_1^{**}(x_0) - Z_1(x_0), Z_i(x_\alpha)] + \sum_l \mu_{il} f^l(x_\alpha) = 0 \quad i = 1, 2 \quad \forall x_\alpha \in S_i$$

The proportional covariance structure entails that

$$\begin{aligned} \text{Cov}[Z_1^{**}(x_0) - Z_1(x_0), Z_1(x_\alpha)] &= (1 - \rho^2) \text{Cov}[Z_1^*(x_0) - Z_1(x_0), Z_1(x_\alpha)] \\ &= -(1 - \rho^2) \sum_l \mu_{Kl} f^l(x_\alpha) \quad \forall x_\alpha \in S_1 \end{aligned}$$

$$\text{Cov}[Z_1^{**}(x_0) - Z_1(x_0), Z_2(x_\alpha)] = 0 \quad \forall x_\alpha \in S_2$$

where the μ_{Kl} are the Lagrange parameters associated with ordinary (univariate) kriging of Z_1 . Therefore the cokriging equations are satisfied with $\mu_{1l} = (1 - \rho^2)\mu_{Kl}$ and $\mu_{2l} = 0$.

The decoupling is “nice to have” but not essential; one just has to solve a bigger cokriging system. The more important issue is the possibility to simplify the inference and modeling of the joint correlation structure in a physically justified manner (see Example 4 in Section 5.6.1).

5.4.4. Ordinary Cokriging in the Presence of a Nugget Effect

An early application of cokriging dealt with the estimation of uranium reserves on the basis of chemical analyses from drill-hole cores and of radioactivity measurements (Guarascio, 1976). Although the relationship between radioactivity and uranium is not simple, in particular due to the fact that radioactivity integrates a much larger volume than cores, these data are still useful because they are sampled much more densely. Also the nugget effect in the uranium grades is integrated out by radioactivity measurements.

In order to evaluate the benefit of cokriging as a function of the relative importance of the nugget effect, let us consider a controlled configuration in which the uranium grade (variable 1) of a block is estimated from a central drill-hole and four drill-holes symmetrically located far from the block, that is, at a distance greater than the range. Radioactivity measurements (variable 2) are also assumed known at these locations. The two means are constant but unknown, and to simplify calculations, we assume that variable 2 and the continuous component of variable 1 conform to a proportional covariance model. Our complete model is therefore

$$\begin{cases} C_{11}(h) = \sigma_0^2 \delta(h) + \sigma_1^2 \rho(h) \\ C_{22}(h) = \sigma_2^2 \rho(h) \\ C_{12}(h) = \sigma_{12} \rho(h) \end{cases}$$

where $\rho(h)$ is an isotropic correlogram. The covariance between the central drill-hole and the mean block value is $\sigma_1^2 \rho_{0v}$, while the covariance between the other four samples and that block is zero. Likewise the covariance between the central radiometry sample and the block is $\sigma_{12} \rho_{0v}$, and the other four covariances are zero. Finally the variance of the block itself is $\sigma_1^2 \rho_{vv}$, assuming that the nugget effect is averaged out. For the example we use the values $\rho_{0v} = 0.718$ and $\rho_{vv} = 0.624$ obtained by integration of a spherical model of unit sill over a square of length half the range. These numbers can also be read from graphs (e.g., Journel and Huijbregts, 1978, pp. 127–128), though with less precision.

Due to the symmetry of the configuration, it is possible to reduce the problem to the determination of the two weights λ_1 and λ_2 assigned to the central sample of Z_1 and Z_2 , respectively, the weights assigned to the other four samples being equal and, respectively, equal to $(1 - \lambda_1)/4$ and $-\lambda_2/4$. The cokriging system (5.6) comprises four equations

$$\begin{cases} \lambda_1(\sigma_0^2 + \sigma_1^2) + \lambda_2\sigma_{12} + \mu_1 = \sigma_1^2\rho_{0v} \\ \left(\frac{1-\lambda_1}{4}\right)(\sigma_0^2 + \sigma_1^2) - \left(\frac{\lambda_2}{4}\right)\sigma_{12} + \mu_1 = 0 \\ \lambda_1\sigma_{12} + \lambda_2\sigma_2^2 + \mu_2 = \sigma_{12}\rho_{0v} \\ \left(\frac{1-\lambda_1}{4}\right)\sigma_{12} - \left(\frac{\lambda_2}{4}\right)\sigma_2^2 + \mu_2 = 0 \end{cases}$$

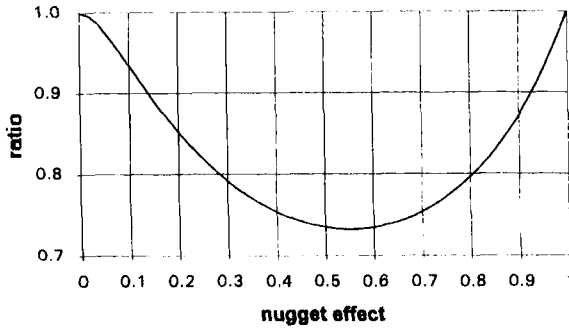


FIGURE 5.2. Normalized cokriging variance of a square panel as a function of the nugget effect. Vertical axis is σ_{CK}^2/σ_K^2 and horizontal axis is σ_0^2 , where $\sigma_0^2 + \sigma_1^2 = 1$.

and its solution is

$$\lambda_1 = \frac{1}{5} + \frac{4}{5} \frac{(1 - \rho_{12}^2)\sigma_1^2}{\sigma_0^2 + (1 - \rho_{12}^2)\sigma_1^2} \rho_{0v}$$

$$\lambda_2 = \frac{4}{5} \frac{\sigma_0^2}{\sigma_0^2 + (1 - \rho_{12}^2)\sigma_1^2} \rho_{12} \frac{\sigma_1}{\sigma_2} \rho_{0v}$$

The cokriging variance is then

$$\sigma_{CK}^2 = \sigma_1^2 \rho_{vv} - \lambda_1 \sigma_1^2 \rho_{0v} - \lambda_2 \sigma_{12} \rho_{0v} - \mu_1$$

The solutions for ordinary kriging are obtained by letting $\rho_{12} = 0$ in the above results. Figure 5.2 shows a graph of the ratio of the cokriging to kriging variances for $\sigma_0^2 + \sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho_{12} = 0.9$. The minimum ratio of 0.73 is reached for a nugget effect $\sigma_0^2 = 0.56$, representing a reduction in standard deviation $(\sigma_K - \sigma_{CK})/\sigma_K = 15\%$. This figure gives an idea of the maximum gain that can be obtained by cokriging. The weight λ_2 reaches a maximum of about 31% for $\sigma_0^2 = 0.42$. Contrary to intuition, this maximum is not achieved for the highest nugget effect. For a pure nugget effect, Z_1 and Z_2 become uncorrelated and the weight λ_2 is zero. The fact that this weight is also zero when the nugget effect vanishes is a property of the proportional covariance model. Notice in passing that λ_1 is insensitive to scale changes in either Z_1 or Z_2 , whereas λ_2 is scale dependent through the ratio σ_1/σ_2 .

When both primary and secondary variables have a nugget effect, it is the continuous components that contribute to cross-correlations. Marbeau (1976) made an extensive application of ordinary cokriging in forestry, where nugget effects are strong, and found a reduction in error standard deviation of about 10% with a maximum of 15%. Pan et al. (1993) compared cokriging and OK by cross-validation in a case study of a polymetallic deposit and report a standard deviation reduction of about 20%.

5.4.5. Filtering a Systematic Error

Suppose that we have two sets of data for a variable of interest Z_1 , where the second set comprises a *systematic* error $\varepsilon(x)$ that is uncorrelated with Z_1 but spatially structured. Can the biased data help estimate Z_1 ?

Our model is

$$\begin{cases} Z_2(x) = Z_1(x) + \varepsilon(x) \\ E\varepsilon(x) = m_\varepsilon \neq 0 \quad \text{and} \quad \text{Cov}(Z_1, \varepsilon) = 0 \end{cases}$$

In the absence of indications on the bias m_ε , the means of Z_1 and Z_2 are algebraically independent and the cokriging equations (5.6) apply. We can therefore use the information in the second data set provided that the decomposition $C_{22}(h) = C_{11}(h) + C_{\varepsilon\varepsilon}(h)$ is known, and most important, that we can tell which data are biased and which are not.

As usual, this requires the availability of at least enough data of Z_1 to estimate the mean, that is, evaluate the bias; clearly, biased data alone would be worthless unless the bias is very small. At the other extreme, if Z_1 and Z_2 are simultaneously present at most of the points, the practical solution is to fill in the gaps by interpolation of the error field and then base the estimation on the Z_1 data only, both actual and derived from Z_2 . In intermediate sample configurations, cokriging can provide a useful solution.

Journel and Huijbregts (1978, p. 342) compared kriging with cokriging in a mining case of block estimation where cuttings are used in addition to a central core analysis, and they report a gain in standard deviation of about 20%.¹

5.4.6. Universal Cokriging and Linear Regression

Cokriging exploits both correlations due to proximity in space and correlation among variables. We have already noted that when considering independent drifts, cokriging reduces to kriging if correlations among variables are zero, and we will now show that cokriging coincides with linear regression when means are known and all point to point correlations are zero. In this sense cokriging constitutes a generalization of both approaches.

In linear regression theory one considers independent samples of a random vector $\mathbf{Z} = (Z_1, \dots, Z_p)'$ drawn from a multivariate distribution with mean vector $\mathbf{m} = (m_1, \dots, m_p)'$ and a covariance matrix defined by $\sigma_{ij} = \text{Cov}(Z_i, Z_j)$. The linear regression of Z_1 on Z_2, \dots, Z_p is the function of the form

$$\hat{Z}_1 = \lambda_1 + \lambda_2 Z_2 + \dots + \lambda_p Z_p$$

minimizing the expected mean square error. The solution is easily found to be

$$\hat{Z}_1 = m_1 + \sum_{i=2}^p \lambda_i (Z_i - m_i) \quad (5.13)$$

with coefficients λ_i that satisfy the linear system

$$\sum_{j=2}^p \sigma_{ij} \lambda_j = \sigma_{i1} \quad i = 2, \dots, p$$

These equations coincide with the simple cokriging equations (5.1) for estimating Z_1 from Z_2, \dots, Z_p , where vectors are replaced by scalars (one sample per variable) and $i_0 = 1$, the set S_1 being empty. In the Gaussian case the estimator also coincides with the regression equation (conditional expectation). Note that the means are assumed constant and known and that no spatial aspect is involved so far.

In linear regression theory the multivariate samples are regarded as *independent* realizations of a parent random vector so that the estimation of a vector component Z_1 only involves the other components of the *same* vector. The equivalent geostatistical model where vectors \mathbf{Z} are indexed by location x is to assume a multivariate nugget effect

$$\text{Cov}[Z_i(x), Z_j(y)] = \sigma_{ij} \delta(y - x)$$

or that all ranges are small compared with distances between samples (large grids). The simple cokriging estimator of Z_1 at a point x_0 , where the other variables are known, only involves the values at this point and is similar to (5.13) except that the means are possibly location dependent:

$$Z_1^{**}(x_0) = m_1(x_0) + \sum_{i=2}^p \lambda_i [Z_i(x_0) - m_i(x_0)] \quad (5.14)$$

This remains the simple cokriging estimator of Z_1 when ranges are not small but x_0 is so far from the other data points that correlations have vanished (isolated point).

Universal cokriging generalizes this by allowing for unknown means defined in linear parametric form. The result is obtained by simply replacing all means in (5.14) by their cokriging estimates, by virtue of an additivity relationship between simple and universal cokriging similar to that between SK and UK:

$$Z_1^{**}(x_0) = m_1^{**}(x_0) + \sum_{i=2}^p \lambda_i [Z_i(x_0) - m_i^{**}(x_0)] \quad (5.15)$$

Note that in general $m_i^{**}(x_0)$ is *not* the least squares estimator based on the values of the i th variables only but may involve all variables at all points. For example, if the data consist of N complete vectors in addition to the $(p-1)$

values at x_0 , the means for $i > 1$ coincide with the least squares estimates based on the $N + 1$ values of the i th variable, but for $i = 1$ the optimum estimate of the mean turns out to be the least squares estimate from the $N + 1$ values $Z_1^{**}(x_0), Z_1(x_1), \dots, Z_1(x_N)$, where the unknown $Z_1(x_0)$ is replaced by its optimal estimate (5.15). (This looks circular but it isn't!) In practice, the means and the coefficients of the regression are established once and for all by least squares from a subset of complete vectors, and a fixed equation is used thereafter.

The next generalization is to introduce spatial correlation among the vectors $\mathbf{Z}(x)$. But the price to pay for that extra power is the burden of statistical inference of $p(p + 1)/2$ covariance functions.

5.4.7. Simultaneous Estimation of Several Variables

Instead of estimating a single variable we may want to estimate a few or all of them simultaneously. Typical examples are the evaluation of different elements, such as Pb, Zn, and Ag in a polymetallic mineral deposit, or the joint reconstruction of geological surfaces to define a consistent three-dimensional model. This can be achieved by cokriging each variable from all the others, one at a time, but it would be a very inefficient algorithm. The better way is to perform all calculations at once and solve one linear system with p right-hand sides rather than p linear systems with one right-hand side. Denote the cokriging estimator of the k th variable at the point x_0 by

$$Z_k^{**}(x_0) = \sum_i \lambda'_{ik} Z_i$$

where the second subscript k refers to the k th variable, then the joint cokriging equations for the vector

$$\mathbf{Z}^{**}(x_0) = (Z_1^{**}(x_0), Z_2^{**}(x_0), \dots, Z_p^{**}(x_0))'$$

is obtained by writing the equations (5.4) for the different values of the index of the estimated variable

$$\begin{cases} \sum_{j=1}^p \mathbf{C}_{ij} \lambda_{jk} + \mathbf{F}_i \mu_{ik} = \mathbf{c}_{ik} & i = 1, \dots, p; \quad k = 1, \dots, p \\ \mathbf{F}_i' \lambda_{ik} = \mathbf{f}_{k0} \delta_{ik} & i = 1, \dots, p; \quad k = 1, \dots, p \end{cases} \quad (5.16)$$

The μ_{ik} and the \mathbf{f}_{k0} are, respectively, the vector of Lagrange parameters and vector of drift functions relative to the k th variable evaluated and \mathbf{c}_{ik} is the covariance between Z_1 and $Z_k(x_0)$. In (5.16) the index k ranges from 1 to p , but it could as well vary within a subset if one wanted to estimate a subset of the p variables.

The covariance of cokriging errors is given by

$$\mathbf{E}[Z_k^{**}(x_0) - Z_k(x_0)][Z_{k'}^{**}(x_0) - Z_{k'}(x_0)] = c_{kk'} - \sum_i \lambda'_{ik'} \mathbf{c}_{ik} - \mu'_{kk'} \mathbf{f}_{k0} \quad (5.17)$$

where $c_{kk'}$ is the covariance of $Z_k(x_0)$ and $Z_{k'}(x_0)$. In the simple case $p = 2$, the awesome system (5.16) is simply written as

$$\begin{bmatrix} C_{11} & C_{12} & F_1 & 0 \\ C_{21} & C_{22} & 0 & F_2 \\ \vdots & \vdots & \vdots & \vdots \\ F'_1 & 0 & 0 & 0 \\ 0 & F'_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \vdots & \vdots \\ \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ \vdots & \vdots \\ f_{10} & 0 \\ 0 & f_{20} \end{bmatrix}$$

generalizing (5.6).

The problem of simultaneous cokriging of several variables was first considered by Myers (1982) who solved it by minimizing

$$\sum_{k=1}^p E[Z_k^{**}(x_0) - Z_k(x_0)]^2 = \text{Trace } \mathbf{M}_B$$

where \mathbf{M}_B is the covariance matrix of cokriging errors for a given matrix of weights \mathbf{B} ensuring unbiasedness. Carr et al. (1985) published a computer program implementing this approach. Alternatively, Ver Hoef and Cressie (1993) proposed to find the matrix \mathbf{B} such that for any other matrix of weights \mathbf{A} ensuring unbiasedness, the matrix $\mathbf{M}_A - \mathbf{M}_B$ is positive definite, implying that any linear combination of variables $Z_i(x_0)$ is estimated better using \mathbf{B} than \mathbf{A} . The equations turn out to be the same and coincide with (5.16). The error covariance matrix \mathbf{M}_B is given by (5.17) and can be used to define confidence ellipsoids for the vector $\mathbf{Z}(x_0) = (Z_1(x_0), \dots, Z_p(x_0))'$.

5.4.8. Algebraic Dependence between Drifts

We now turn to the case where the drift coefficients of the different variables are related. Each case is different, but the principle is the same: the unbiasedness conditions must be modified to reflect the algebraic dependencies. In this section we will just examine the case of two variables with the same mean. Other cases will be encountered later.

Consider the cokriging estimator of $Z_1(x_0)$ based on Z_1 and Z_2 .

$$Z^{**} = \lambda'_1 Z_1 + \lambda'_2 Z_2$$

If $Z_1(x)$ and $Z_2(x)$ have the same mean, the unbiasedness conditions are of the form

$$EZ^{**} = (\lambda'_1 F_1 + \lambda'_2 F_2) \mathbf{a} = \mathbf{f}'_{10} \mathbf{a} \quad \forall \mathbf{a}$$

that is,

$$F'_1 \lambda_1 + F'_2 \lambda_2 = \mathbf{f}_{10}$$

and the cokriging system (5.6) becomes

$$\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{F}_1 \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \mathbf{F}_2 \\ \mathbf{F}'_1 & \mathbf{F}'_2 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{11} \\ \mathbf{c}_{21} \\ \mathbf{f}_{10} \end{bmatrix}$$

Here there is a single set of unbiasedness conditions and therefore a single μ . An application of this system is the estimation of point or block values from data with different supports. If the mean is constant, all points and blocks have the same mean, and the result applies directly. If it isn't the system is still valid, simply the basis drift functions for Z_1 or Z_2 are replaced by their block average equivalents.

The structure of the above system allows for the estimation of Z_1 from Z_2 alone using

$$\begin{bmatrix} \mathbf{C}_{22} & \mathbf{F}_2 \\ \mathbf{F}'_2 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \lambda_2 \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{21} \\ \mathbf{f}_1 \end{bmatrix}$$

which is essentially a UK system in which the right-hand side has been modified to include the cross-covariances between the estimating and the estimated variables instead of direct covariances. We have already encountered one application of this: filtering a nonsystematic error (cf. Equation (3.36)). Another could be the estimation of block values from point values or the reverse—provided that the point-to-point covariance is known.

Instead of postulating that all variables have the same mean, one could let the drifts differ by a free constant as when mapping subparallel horizons. This would be a *mixed case*. The resulting changes in the equations are left to the reader.

5.5. CASE OF GRADIENT INFORMATION

A case of algebraic dependence that is particularly interesting for applications is when the auxiliary variables are derivatives of the primary variable. Typically we may know the depth of a geological surface and also its dip magnitude and azimuth. In the petroleum industry those measurements are performed by dipmeter logging tools, which are resistivity devices that record several curves along the borehole wall or even a complete image. The dipping plane is determined from the displacements of these curves or by tracking the trace of the plane along the image (e.g., Hepp and Dumestre, 1975; Antoine and Delhomme, 1993). Another familiar example of gradient information is atmospheric pressure or geopotential and wind. In the so-called geostrophic approximation wind is the gradient of atmospheric pressure or geopotential, up to a multiplicative factor and a 90° rotation due to the Coriolis force.

Two estimation problems may be considered, one is estimation *of* gradients, and the other is estimation *with* gradients. For a general treatment it is useful to introduce the notion of *directional derivative* of the RF $Z(x)$ with respect to the unit vector $u \in \mathbb{R}^n$ (e.g., Rockafellar, 1970). This is a random variable defined as the following limit in the mean square sense

$$\frac{\partial Z}{\partial u}(x) = \lim_{\rho \downarrow 0} \frac{Z(x + \rho u) - Z(x)}{\rho} \quad (5.18)$$

If $Z(x)$ is (m.s.) differentiable and $\nabla Z(x)$ is its gradient at the point x , one has

$$\frac{\partial Z}{\partial u}(x) = \langle \nabla Z(x), u \rangle$$

The mathematical conditions for the existence of directional derivatives in all directions u are the same as the differentiability of $Z(x)$, namely that the (generalized) covariance of $Z(x)$ be twice differentiable at the origin (see inequality (4.23)). Unfortunately, the variogram functions most used in practice—namely the spherical, the exponential, or the linear models—are not differentiable. The derivatives of $Z(x)$ are then simply not defined mathematically (infinite variance) and can in principle neither be estimated nor be used for estimation. Does it mean that slope information is essentially useless? We will discuss this question later; for now let us assume that derivatives exist.

5.5.1. Estimation of Derivatives

This part is straightforward. By linearity of the kriging system, the optimal estimator of the difference $Z(x + \rho u) - Z(x)$ is $Z^*(x + \rho u) - Z^*(x)$. Passing to the limit as in (5.18), we get

$$\left(\frac{\partial Z}{\partial u} \right)^* = \frac{\partial Z^*}{\partial u} \quad (5.19)$$

That is, the cokriging estimator of the derivative from Z alone is equal to the derivative of the kriging estimator. Hence the system for cokriging the derivative from Z is simply obtained by differentiating the right-hand side of the UK system (3.14), using the formula in the first line of (5.22). For the variance the term σ_{00} becomes $\text{Var}[(\partial Z / \partial u)(x_0)]$. The variogram equations (3.16) can be modified in the same manner, but the above variance term must be added. (Note: A derivative is a linear combination whose sum of weights is zero.)

The relation (5.19) assumes that both the derivatives of Z and Z^* exist, but in fact only Z^* may have a derivative; for example, if the variogram of Z is linear near the origin, Z^* is differentiable nevertheless except at the sample points. By extension, we may accept $\partial Z^* / \partial u$ as the derivative estimator, but of course no estimation variance can be attached. Note that differentiating the kriging estimator does not lead to noise enhancement in the derivatives as

would the differentiation of the Z data themselves because Z^* is a *smoothed* version of Z , except at the sample points.²

5.5.2. Estimation with Derivatives

Now we turn to the more difficult and also the more interesting problem. Our cokriging estimator is of the form

$$Z^{**}(x_0) = \sum_{x_\alpha \in S_1} \lambda_{1\alpha} Z(x_\alpha) + \sum_{x_j \in S_2} \lambda_{2j} \frac{\partial Z}{\partial u_j}(x_j) + \sum_{x_j \in S_3} \lambda_{3j} \frac{\partial Z}{\partial v_j}(x_j) \quad (5.20)$$

S_1 is the sampling set of Z data, and the directional derivatives are taken along directions u and v which are allowed to vary with the point in the sampling sets S_2 and S_3 of derivatives. Typically u and v are simply the directions of the coordinate axes, but they can also be the gradient direction (e.g., downdip) and the orthogonal direction (strike). This general form of the estimator allows for the case where only one of the components is known, such as only the strike direction (the horizontal line of the bedding plane).

The functional relationship between $Z(x)$ and its derivatives carries over to the drift and covariances. From (5.20) the unbiasedness conditions are immediately found to be

$$\sum_{x_\alpha \in S_1} \lambda_{1\alpha} f^l(x_\alpha) + \sum_{x_j \in S_2} \lambda_{2j} \frac{\partial f^l}{\partial u_j}(x_j) + \sum_{x_j \in S_3} \lambda_{3j} \frac{\partial f^l}{\partial v_j}(x_j) = f^l(x_0) \quad \forall l$$

The cokriging system is then of the form

$$\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{C}_{13} & \mathbf{F}_1 \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \mathbf{C}_{23} & \mathbf{F}_2 \\ \mathbf{C}_{31} & \mathbf{C}_{32} & \mathbf{C}_{33} & \mathbf{F}_3 \\ \mathbf{F}'_1 & \mathbf{F}'_2 & \mathbf{F}'_3 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{11} \\ \mathbf{c}_{21} \\ \mathbf{c}_{31} \\ \mathbf{f}_0 \end{bmatrix} \quad (5.21)$$

and the cokriging variance is as in (5.7). On the left-hand side, \mathbf{F}_1 holds the drift function values over S_1 , while \mathbf{F}_2 and \mathbf{F}_3 hold their derivatives in u and v over S_2 and S_3 . All \mathbf{F}_i have the same number of columns, that is, the number of basis drift functions, and a number of rows equal to the number of data points of each variable (the first column of \mathbf{F}_1 is usually composed of 1's, and those of \mathbf{F}_2 and \mathbf{F}_3 of 0's). Assuming the covariance model to be positive definite and to have no duplication of points, a sufficient condition for the cokriging system to have a unique solution is that \mathbf{F}_1 be of full rank, namely that kriging based on the Z values alone is possible. That condition is not necessary, however, since knowledge of derivative information may remove indeterminacy (e.g., specifying the normal to a plane in addition to one line in that plane).

The C_{ij} terms are the direct and cross-covariances between the Z data and the derivatives, and the c_{i1} terms are the covariances between, respectively, the vectors of Z , $\partial Z/\partial u$ and $\partial Z/\partial v$ data, and $Z(x_0)$. All these can be obtained by differentiation of the covariance $\sigma(x, y)$ of $Z(\cdot)$ using (5.18) provided that all derivatives involved exist and that differentiation and expectation may be interchanged. We get

$$\begin{aligned} \text{Cov} \left[Z(x), \frac{\partial Z}{\partial u}(y) \right] &= \lim_{\rho \downarrow 0} \frac{\sigma(x, y + \rho u) - \sigma(x, y)}{\rho} \\ \text{Cov} \left[\frac{\partial Z}{\partial u}(x), \frac{\partial Z}{\partial v}(y) \right] &= \lim_{\rho, \rho' \downarrow 0} \frac{\sigma(x + \rho u, y + \rho' v) - \sigma(x, y + \rho' v) - \sigma(x + \rho u, y) + \sigma(x, y)}{\rho \rho'} \end{aligned} \quad (5.22)$$

These formulas are general and do not require stationarity of the covariance. Explicit results can be obtained if the covariance is stationary and isotropic. For example, in 2D, denoting the components of $h = y - x$ by h_x and h_y , if

$$\sigma(x, y) = C(r) \quad \text{with} \quad r = \sqrt{h_x^2 + h_y^2}$$

then letting

$$h_u = \langle h, u \rangle = h_x \cos \theta + h_y \sin \theta \quad h_v = \langle h, v \rangle = h_x \cos \varphi + h_y \sin \varphi$$

we have

$$\begin{aligned} \text{Cov} \left[Z(x), \frac{\partial Z}{\partial u}(y) \right] &= h_u \frac{C'(r)}{r} \\ \text{Cov} \left[\frac{\partial Z}{\partial u}(x), \frac{\partial Z}{\partial v}(y) \right] &= -h_u h_v \frac{1}{r^2} \left[C''(r) - \frac{C'(r)}{r} \right] - \cos(\theta - \varphi) \frac{C'(r)}{r} \end{aligned} \quad (5.23)$$

Here u and v are arbitrary vectors not assumed orthogonal. Therefore (5.23) can be applied when u and v represent u_α and v_β , u_α and u_β , and v_α and v_β . Notice the anisotropy of the covariance of derivatives. These equations as well as (5.22) remain valid when the covariance $C(h)$ is replaced by $-\gamma(h)$, and the cokriging system (5.21) can be written in terms of the variogram provided that the constant function 1 is included in the set of drift basis functions.

The foregoing approach has been applied to meteorological problems in the scope of a broader study involving about 40 variables related by physical laws (Chilès, 1976; Chauvet et al., 1976). The example shown in Figure 5.3 is a 500-millibar geopotential map over Western Europe and the Atlantic Ocean obtained by using both geopotential and wind observations. The variogram of geopotential was modeled as the sum of two Gaussian variograms of type (2.54), one with a scale parameter of about 1400 km and the other, representing

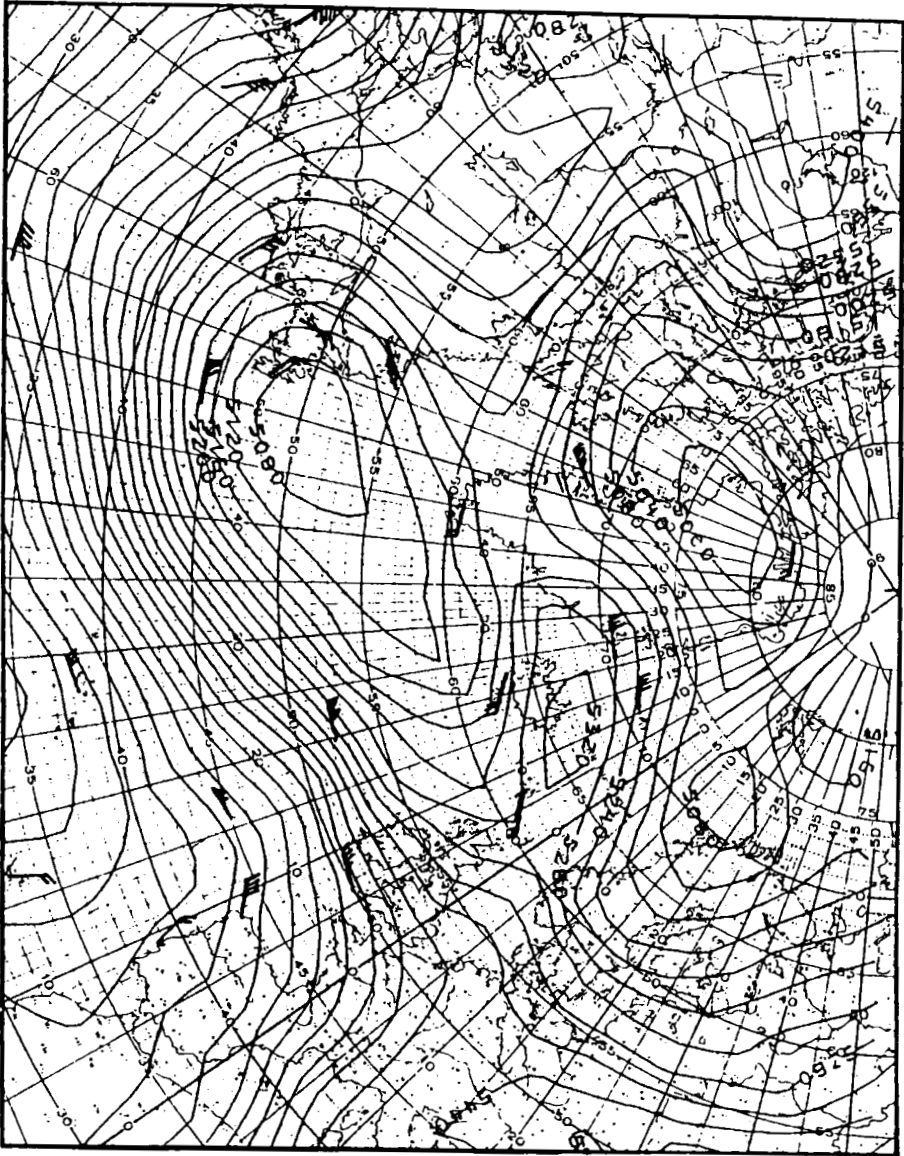


FIGURE 5.3. Cokriging of the 500-millibar geopotential field on 15/03/74—12:00. Unit: gpm. The symbol *F* represents wind data (direction and intensity). A guess field was used. From Chilès (1976), with kind permission from Kluwer Academic Publishers.

the main component of the variogram of derivatives, with a scale parameter of 500 km. With a grid spacing of about 200 km the short-scale variogram allows winds to influence nearby grid nodes, but this influence decreases rapidly with distance. The map shows wind data in sparsely sampled areas. Notice

that as a result of the Coriolis force, these winds are approximately *tangent* to geopotential contours.

5.5.3. Physical Validity of Derivatives

The singularity of common variogram models at the origin forces us to question the physical validity of representing gradient information by derivatives. In a mathematical sense a derivative is punctual information, but in the real world this is usually not what is measured. Consider, for example, geological dips. With a dipmeter tool there is a minimum distance between dip computation points corresponding to the diameter of the borehole, typically 20 to 25 cm. Even so there can be considerable variability in the magnitude and azimuth of these dips due to local sedimentary features such as ripples or cross-bedding. Since we are not interested in representing these local fluctuations, we consider only the dips that are likely to have significant lateral extension. Typically this is what geologists call the *structural dip*. It is determined by statistical averaging within depth intervals with approximately constant dip magnitude and azimuth and is supposed to represent beds that were deposited on a nearly horizontal surface, their present dips being the result of tectonic stresses (e.g., Serra, 1985). Thus the slope information to be used for geological mapping is not a point gradient but rather an average over some support actually difficult to specify.³ Renard and Ruffo (1993) modeled this dip as the derivative of the depth convolved by a Gaussian weighting function with main axes lengths (a_1, a_2) , which were assigned arbitrary but plausible values ($a_1 = a_2 = 50$ m for a depth variogram range of 1400 m).

Another solution is to model derivative information by finite increments, which is where we started from in the definition (5.18), and this is also a form of averaging. In 2D, when both value and gradient are known at the same data point P_0 , a simple procedure is to create *dummy points* P_1 and P_2 that, together with P_0 , recreate the correct tangent plane. For example, P_1 and P_2 are placed at a distance ρ from P_0 , where P_1 is along the direction of steepest slope (e.g., downdip) with the value $Z_1 = Z_0 + \rho \|\nabla Z\|$ and P_2 is normal to this direction (e.g., along strike) with the value $Z_2 = Z_0$. The value of ρ depends on the lateral validity of the dip and is an *interpretation decision*. A practical difficulty may be encountered when placing the dummy points: that they fall in the immediate vicinity of actual points, thereby causing the kriging matrix to be singular and creating possible inconsistencies in the data. This is resolved by interactive editing of the dummy points. The main advantage of the method is that it can be implemented with a standard kriging program. When only the gradient is known, a similar method can be applied, except that the dummy points are involved only through the differences $Z_1 - Z_0$ and $Z_2 - Z_0$, which requires a genuine cokriging program. The difficulty raised by the placement of the dummy points does not go away, unfortunately. Notice that it is strictly equivalent to specify two values or one of them and their difference: in the kriging system one row is just replaced by the difference between two rows.

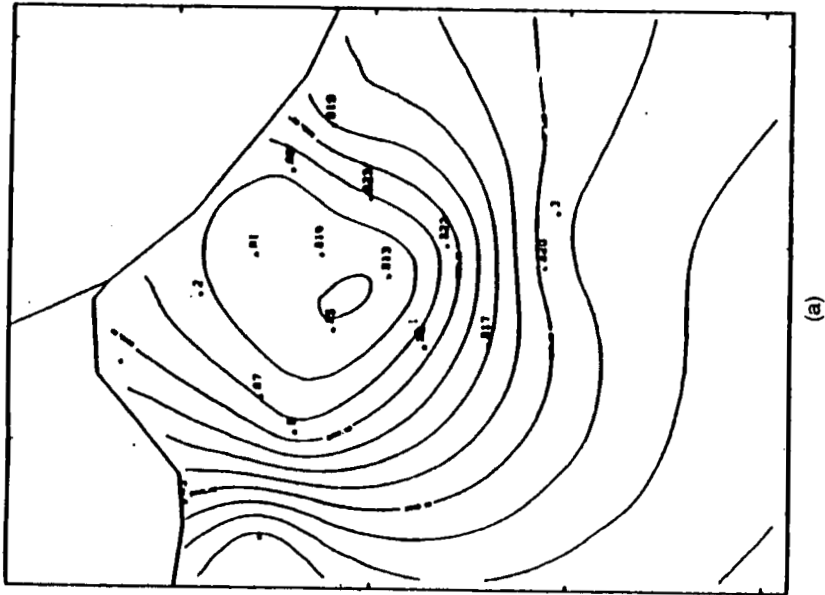
When the RF $Z(x)$ is differentiable, the question of the lateral extent (the distance ρ) seems irrelevant, at first glance. But we know from experience that the influence of punctual derivatives may be so local as to have virtually no effect on the interpolated grid. In reality there is a difference between a *mathematical derivative* and a *physical derivative*. The mathematical derivative at the point x in the direction u is the limit of the ratio $[Z(x + \rho u) - Z(x)]/\rho$ as ρ becomes infinitely small. It suffices to modify a function only very slightly, for example, by convolving it with a smooth window of arbitrarily small width, to make it differentiable and even infinitely differentiable. The mathematical derivative has a physical existence only if this ratio remains approximately constant over a distance commensurate at least with the scale of observation, and preferably the grid spacing. That distance represents the lateral extent of the gradient. Phrased differently, derivative data have an impact only if they exhibit enough spatial correlation at the scale of the study; otherwise, they appear like a nugget effect.

A final remark. The presence of a nugget effect on the variogram of $Z(x)$, which is the most adverse case of singularity at the origin, does not exclude the possibility of useful derivative information. If the derivative is computed from measurements of Z recorded by the same process, the measurement errors are the same and cancel out in derivatives. A case in point is again the dipmeter, where the uncertainty on the absolute depth of a single resistivity curve can reach several feet but the displacement between curves can be determined within a fraction of an inch.

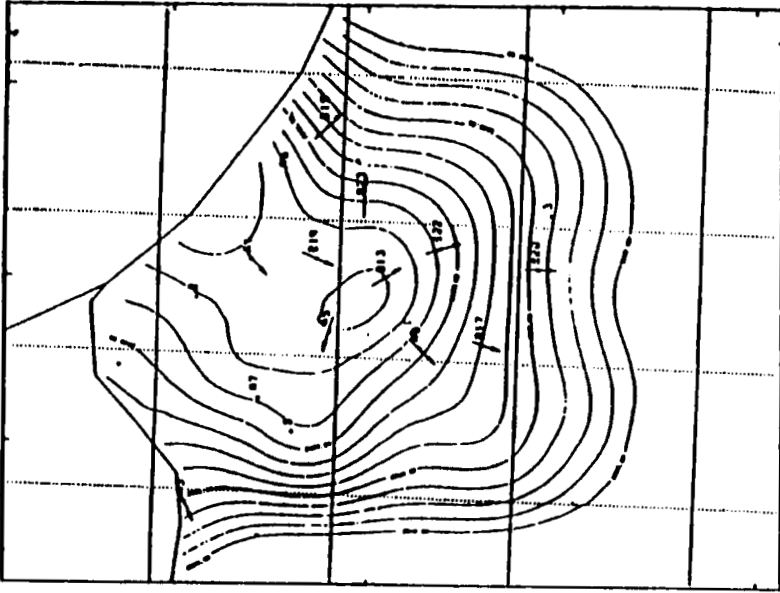
5.5.4. Contour Mapping and Kriging under Boundary Conditions

The first example compares the maps of a petroleum reservoir top obtained with and without dipmeter data (Fig. 5.4, from Delfiner et al., 1983). The dip directions measured by dipmeter are essentially the same as those obtained by interpolation of plain well data, except at the top of the structure where the shape of the 2750 contour line is changed due to a local azimuth oriented SW at well A1. But the dip values of 17 and 15 degrees measured in peripheral wells B19 and B20 ensure closure that could not be obtained without them.

The second example is extracted from a fascinating study that J. P. Delhomme (1979b) presented at a conference but never published, where he used gradient information as a means of constraining estimates to honor known physical boundary conditions. The work is about an aquifer in which water, for geological reasons, cannot flow through the boundaries marked with a thick solid line in Figure 5.5a and is released through the western outlet. For the purpose of numerical modeling of the aquifer, it is desired to map the hydraulic head on the basis of 49 available piezometric measurements. Figure 5.5b shows the map produced by usual kriging; it is not acceptable for a hydrogeologist. First, the no-flow constraints are violated: since the flow is orthogonal to hydraulic head contour lines, these line should be perpendicular



(a)



(b)

FIGURE 5.4. A comparison of maps obtained with and without dipmeter data. (a) map from well data; (b) map from well and dipmeter data. From Delfiner et al. (1983).

to the no-flow boundary, but in the southern part they are not. Second, to the west the water release is not in front of the actual outlet. Boundary conditions are now introduced by specifying that the gradient component normal to the no-flow boundary is zero. In principle, since the no-flow contour is continuous, one should consider a continuous cokriging estimator, but in practice, it suffices to discretize the problem and replace the orthogonal gradient component by the differences between pairs of dummy points shown in Figure 5.5a, considering

$$Z^{**} = \sum_{\alpha} \lambda_{1\alpha} Z(x_{\alpha}) + \sum_{\beta} \lambda_{2\beta} [Z(x_{\beta} + \rho v_{\beta}) - Z(x_{\beta} - \rho v_{\beta})]$$

where the differences here are zeros. Why then consider these differences at all since their contribution to the estimator is nil? Because, and that is key, the weights $\lambda_{1\alpha}$ are different from kriging weights based on the Z_{α} alone. The resulting map in Figure 5.5c is the “ideal” one in the sense that it is consistent with everything that is already known. In order to evaluate the strength of the additional boundary conditions, the map was redrawn using only nine wells plus these constraints and the result is pictured in Figure 5.5d, which also shows the locations of the dummy wells. The map is not very different from the ideal one, indicating that boundary conditions constitute very strong constraints.

Lajaunie et al. (1997) use a similar approach to interpolate geological interfaces by measuring orientation data (foliation or stratification planes) on the interfaces—where they coincide with the interface orientation—and within the geological layers. The model considers that the geological interfaces are iso-surfaces of some scalar 3D potential field Z . There is no direct measurement of Z , but three other types of data are used: increments of Z with a zero value (pairs of points known to belong to the same interface), gradient data (direction orthogonal to the stratification, and modulus derived from the compression of the layers), and orientation data (gradient component in a direction parallel to the stratification, i.e., with a zero value). These data define Z up to a constant so that the estimated quantity is the relative value $Z(x_0) - Z(0)$.

5.6. MULTIVARIATE RANDOM FUNCTIONS

5.6.1. Cross-Covariances

The cross-covariance functions of a multidimensional stationary RF $\mathbf{Z}(x) = (Z_1(x), \dots, Z_p(x))'$ with mean vector $\mathbf{m}(x) = (m_1(x), \dots, m_p(x))'$ are defined by

$$C_{ij}(h) = E[Z_i(x) - m_i][Z_j(x + h) - m_j] \quad (5.24)$$

and only depend on the separation h between the two points. For simplicity and without loss of generality, we will assume here that all variables are centered to zero means.

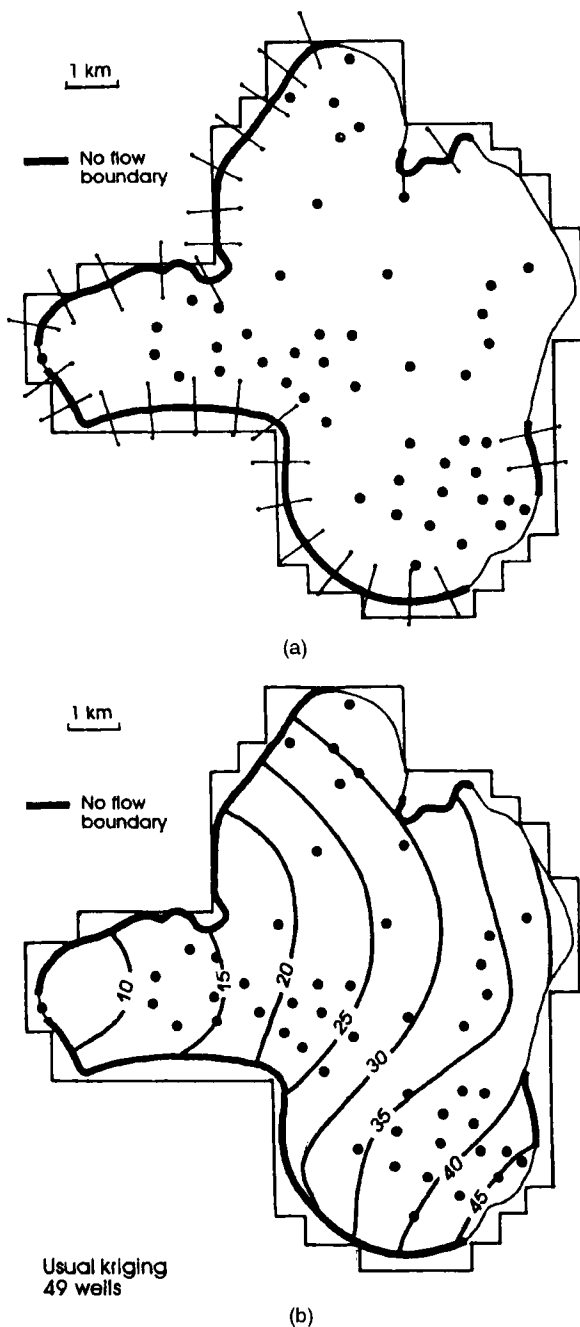
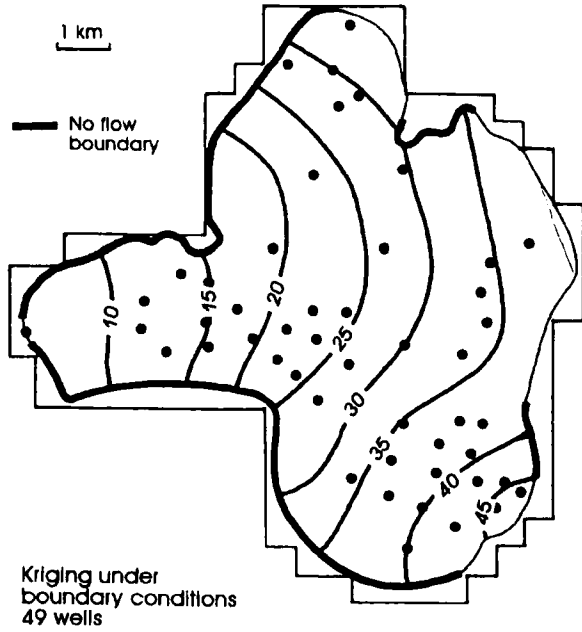
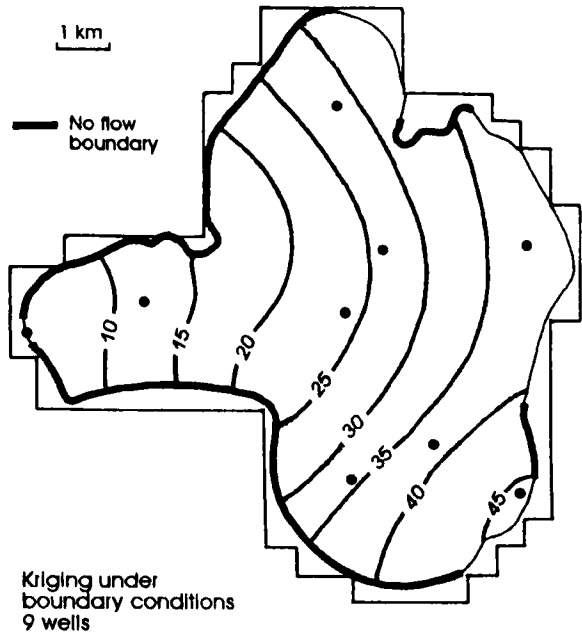


FIGURE 5.5. Kriging under boundary conditions: (a) layout of the aquifer and locations of the wells (thick lines are no-flow boundaries); (b) usual kriging from the 49 wells; (c) kriging under boundary conditions and the 49 wells; (d) kriging under boundary conditions using 9 wells only. From Delhomme (1979b).



(c)



(d)

FIGURE 5.5. (Continued).

The order in which variables are considered matters: Z_i is at the origin x and Z_j at the end point $x + h$. If we permute the variables, we get another cross-covariance

$$E[Z_j(x)Z_i(x+h)] = C_{ji}(h)$$

Trivially

$$C_{ij}(h) = E[Z_i(x)Z_j(x+h)] = E[Z_j(x+h)Z_i(x)] = C_{ji}(-h)$$

but in general

$$C_{ij}(-h) \neq C_{ij}(h)$$

By Cauchy-Schwarz's inequality

$$|C_{ij}(h)| \leq \sqrt{C_{ii}(0)C_{jj}(0)}$$

so that $C_{ij}(h)$ is bounded. Applying the same inequality to one variable and an increment of another shows that $C_{ij}(h)$ is continuous at $h = 0$ whenever $Z_i(x)$ or $Z_j(x)$ is continuous in the mean square. A cross-covariance behaves very differently than a direct covariance function $C_{ii}(h)$. It is not necessarily symmetric, and its maximum may lie at $h = \tau \neq 0$. Such lagged correlation is common in time series due to a delay between the input and the output signals. It is less frequent in spatial phenomena, perhaps because it is incompatible with isotropy. But there are examples: offsets caused by faulting, by selective migration of a chemical element (in uranium mines, maximum radioactivity does not necessarily coincide with maximum uranium grades), or by "man-made" offsets between two images.

Example: The Pure Offset Model

A fracture is studied by means of a core drilled perpendicular to the average fracture plane. The topography of the lower and upper surfaces of the fracture are described by two RFs $Z_1(x)$ and $Z_2(x)$, where x is the position of a point in the average fracture plane. If the two surfaces simply differ by a displacement of horizontal and vertical components τ and m (idealized case of a shear fracture)

$$Z_2(x) = Z_1(x + \tau) + m$$

so that

$$C_{12}(h) = C_{11}(h + \tau)$$

and is a maximum for $h = -\tau$, allowing the detection of τ (Chilès and Gentier, 1993). The same principle is used for locating patterns in an image using templates.

Characterization of Covariance Function Matrices

Since a positive definite function has its maximum at zero, a cross-covariance is generally not positive definite. Under which conditions is $C_{ij}(h)$ a valid cross-covariance function, or rather, under which conditions does the set of $C_{ij}(h)$ constitute a valid model? The criterion is in the frequency domain; it is due to Cramér (1940) who mentions in a footnote of his paper that the same result was found independently by Kolmogorov.

A continuous cross-covariance function $C_{ij}(h)$ has the spectral representation

$$C_{ij}(h) = \int e^{2\pi i \langle u, h \rangle} F_{ij}(du) \quad (5.25)$$

where $u = (u_1, u_2, \dots, u_n)$ is frequency in \mathbb{R}^n and $\langle u, h \rangle = u_1 h_1 + \dots + u_n h_n$. The cross-spectral measure $F_{ij}(du)$ represents the common power of Z_i and Z_j in the infinitesimal spectral interval du . It satisfies the following symmetry relations for any Borel set B :

$$F_{ij}(-B) = \overline{F_{ij}(B)} = F_{ji}(B)$$

Criterion. (Cramér, 1940; Yaglom, 1987, vol. 1, p. 314) The continuous functions $C_{ij}(h)$ are the elements of the covariance matrix of a multi-dimensional stationary RF of order 2 if and only if the cross-spectral matrix $\mathbf{M}(B) = [F_{ij}(B)]$ is positive definite for any (Borel) set B of \mathbb{R}^n , namely $\sum_i \sum_j \lambda_i \bar{\lambda}_j F_{ij}(B) \geq 0$ for any set of complex coefficients $\lambda_1, \dots, \lambda_p$.

This concise criterion ensures that given any linear combination with complex coefficients

$$Y(x) = \lambda_1 Z_1(x) + \dots + \lambda_p Z_p(x)$$

its covariance function

$$E[Y(x)\overline{Y(x+h)}] = \sum_{i=1}^p \sum_{j=1}^p \lambda_i \bar{\lambda}_j C_{ij}(h) = \int e^{2\pi i \langle u, h \rangle} \left[\sum_{i=1}^p \sum_{j=1}^p \lambda_i \bar{\lambda}_j F_{ij}(du) \right]$$

is the Fourier transform of a nonnegative measure and is therefore positive definite.⁴ More generally, all variances of linear combinations calculated with the model are nonnegative. Note, however, that the criterion does *not* exclude the possibility of singular cokriging matrices due to linearly dependent variables. But the cokriging estimator is always unique (being a projection).

If $C_{ij}(h)$ decreases fast enough for $C_{ij}(h)^2$ to be integrable, there exists a spectral density function $f_{ij}(u)$ such that $F_{ij}(du) = f_{ij}(u)du$. If in addition $|C_{ij}(h)|$ is integrable, $f_{ij}(u)$ is continuous and bounded and can be computed by

inversion of the Fourier transformation (5.25). If all cross-covariances admit a spectral density function the validity criterion is a positive definite spectral density matrix

$$\mathbf{M}(u) = \begin{bmatrix} f_{11}(u) & \cdots & f_{1p}(u) \\ \vdots & \cdots & \vdots \\ f_{p1}(u) & \cdots & f_{pp}(u) \end{bmatrix}$$

In general, $f_{ij}(u)$ is a complex function with the Hermitian symmetry (since $C_{ij}(h)$ is real)

$$f_{ij}(-u) = \overline{f_{ij}(u)} = f_{ji}(u)$$

The positive definiteness property entails in particular that for any pair (i, j)

$$|f_{ij}(u)| \leq \sqrt{f_{ii}(u)f_{jj}(u)} \quad (5.26)$$

Examples

1. *Independent RFs.* In this case $\mathbf{M}(u)$ is diagonal. Since all $f_{ii}(u)$ are real and nonnegative, being spectral densities of ordinary covariances $C_{ii}(h)$, $\mathbf{M}(u)$ is positive definite.

2. *Derivative.* In \mathbb{R}^1 consider a RF $Z(x)$ and its derivative $Z'(x)$, and let $f(u)$ be the spectral density associated with the covariance of $Z(x)$. The spectral density matrix is

$$\mathbf{M}(u) = \begin{bmatrix} 1 & 2\pi iu \\ -2\pi iu & 4\pi^2 u^2 \end{bmatrix} f(u)$$

We have $f(u) \geq 0$ and $\det \mathbf{M}(u) = 0$ for all u so that $\mathbf{M}(u)$ is positive definite (= nonnegative definite). The fact that the determinant is identically zero reflects the functional dependence between the variables $Z(x)$ and $Z'(x)$. The same reasoning applies to the pure offset model, with

$$\mathbf{M}(u) = \begin{bmatrix} 1 & e^{2\pi iu\tau} \\ e^{-2\pi iu\tau} & 1 \end{bmatrix} f(u)$$

3. *Proportional covariances.* Let $C_{ij}(h) = \sigma_{ij}\rho(h)$ and $f(u)$ the spectral density associated with $\rho(h)$. Then $\mathbf{M}(u) = [\sigma_{ij}]f(u)$, and since $f(u) \geq 0$, $\mathbf{M}(u)$ is positive definite if and only if the matrix $[\sigma_{ij}]$ is also positive definite. Note that if $\rho(h)$ assumes negative values, the matrix $[C_{ij}(h)]$ is *negative* definite for those values of h .

4. *"Markov-type models".* Consider two RFs Z_1 and Z_2 with unit variance and correlation coefficient ρ . Assume only that their cross-covariance is proportional to the covariance of Z_1

$$C_{12}(h) = \rho C_{11}(h)$$

If C_{11} and C_{22} (which are correlograms) have spectral densities f_1 and f_2 we have

$$\mathbf{M}(u) = \begin{bmatrix} f_1(u) & \rho f_1(u) \\ \rho f_1(u) & f_2(u) \end{bmatrix} \Rightarrow \det \mathbf{M}(u) = f_1(u)[f_2(u) - \rho^2 f_1(u)]$$

The multivariate model is valid if and only if $f_2(u) \geq \rho^2 f_1(u)$ for every frequency u . In particular $f_2(u)$ cannot tend to zero faster than $f_1(u)$ as $u \rightarrow \infty$, showing that Z_2 can be no smoother than Z_1 . Equivalently

$$C_{22}(h) = \rho^2 C_{11}(h) + (1 - \rho^2) C_R(h)$$

for any arbitrary correlogram $C_R(h)$. A particular case is $C_{22}(h) = C_{11}(h) + C_0 \delta$. A similar condition is obtained if we start from the "reverse Markov" model $C_{12}(h) = \rho C_{22}(h)$, which may be more appropriate if Z_2 is smoother than Z_1 , as is typically the case when the secondary data are of an integrated nature.

5. *Nugget effects.* In the multivariate case it is not sufficient (nor necessary) for nugget effect constants to be positive; there are consistency constraints, a fact sometimes overlooked. We consider here the case where discontinuities are at the origin only, keeping in mind that cross-covariances can have discontinuities elsewhere (e.g., the pure offset model). If we admit that the phenomenon can be decomposed into two uncorrelated multivariate components, a pure nugget effect and a continuous component, each must be a valid model, and this condition is also clearly sufficient for the sum to be valid. A little technical difficulty appears if we want to apply the Cramér criterion to the nugget effect because it is applicable only to *continuous* covariances. A simple way to work around this difficulty is to consider the pure nugget effect as a case of proportional covariance model with a continuous correlogram $\rho(h)$ of arbitrarily small range (see Section 2.4.2)

$$C_{ij}(h) = \sigma_{ij} \rho(h)$$

We conclude immediately that the nugget effect matrix $[\sigma_{ij}]$ must be positive definite. For example, in a bivariate model we must have $\sigma_{11} \geq 0$, $\sigma_{22} \geq 0$, and $\sigma_{12}^2 \leq \sigma_{11} \sigma_{22}$: the magnitude of the nugget effect on the cross-covariance cannot exceed the geometric mean of the nugget effects on the two direct covariances. Again, if one of the variables is continuous, the cross-covariance must be continuous too.

Coherency and Phase Spectrum

The spectral density function $f_{ij}(u)$ is a very informative tool to study the relationship between two stationary RFs $Z_i(x)$ and $Z_j(x)$ and is used extensively for the analysis of time signals (e.g., Jenkins and Watts, 1968; Koopmans, 1974). We do not know of such use in a spatial geostatistical context, but it is interesting to briefly mention the approach. Since the letter i is needed to denote the pure imaginary number, we will use the indexes 1 and 2 to denote the two generic variables.

Due to the absence of symmetry of $C_{12}(h)$ the spectral density $f_{12}(u)$ is a complex function

$$f_{12}(u) = c_{12}(u) - iq_{12}(u)$$

The real part $c_{12}(u)$ —not be confused with the covariance $C_{12}(h)$ —is the Fourier transform of the even part of the cross-covariance and is usually called the *cospectrum*, while the imaginary part $q_{12}(u)$ is the Fourier transform of the odd part of the cross-covariance, and is called the *quadrature spectrum*.

The Hermitian symmetry of $f_{12}(u)$ entails

$$c_{12}(-u) = c_{12}(u) \quad \text{and} \quad q_{12}(-u) = -q_{12}(u)$$

The polar representation

$$f_{12}(u) = |f_{12}(u)| e^{i\varphi_{12}(u)}$$

yields another set of spectral parameters that are perhaps the most useful because they can be interpreted quantitatively. Owing to the Cauchy-Schwarz inequality (5.26) the ratio

$$\rho_{12}(u) = \frac{|f_{12}(u)|}{\sqrt{f_{11}(u)}\sqrt{f_{22}(u)}}$$

when defined, is always between 0 and 1. It is named the *coherency spectrum* and provides a nondimensional measure of the correlation between two RFs as a function of frequency. The term “coherency” (or “coherence”) is borrowed from the study of light, and an interesting explanation of this idea can be found in Koopmans (1974, pp. 138ff.). Coherency remains invariant under scale changes in the frequency domain, namely under the application of independent linear filters to $Z_1(x)$ and $Z_2(x)$, which is a very important property when the data have been passed through linear filters with possibly unknown characteristics. The *phase spectrum* is defined by

$$\varphi_{12}(u) = -\arctan \left(\frac{q_{12}(u)}{c_{12}(u)} \right)$$

It is interpreted as the average phase difference between $Z_1(x)$ and $Z_2(x)$ at frequency u . When there is no phase difference at any frequency the cross spectral density $f_{12}(u)$ is real and the cross-covariance is an even function.

5.6.2. Cross-Variograms

The cross-variogram was introduced by Matheron (1965) as the natural generalization of the variogram

$$\gamma_{ij}(h) = \frac{1}{2}E[Z_i(x+h) - Z_i(x)][Z_j(x+h) - Z_j(x)] \quad (5.27)$$

for multivariate intrinsic random functions (of order 0), namely satisfying

$$\begin{cases} E[Z_i(x+h) - Z_i(x)] = 0 & \text{for } i = 1, \dots, p \\ \text{Cov}[Z_i(x+h) - Z_i(x), Z_j(x+h) - Z_j(x)] = 2\gamma_{ij}(h) & \text{exists and depends only on } h \end{cases}$$

The cross-variogram has two advantages over the cross-covariance: (1) it does not assume finite variances, and (2) the estimation of the cross-variogram is not corrupted by the estimation of the means. The relationship between the cross-variogram and the cross-covariance, when it exists, is the following:

$$\gamma_{ij}(h) = C_{ij}(0) - \frac{1}{2}[C_{ij}(h) + C_{ij}(-h)] \quad (5.28)$$

The cross-variogram satisfies $\gamma_{ij}(0) = 0$ and is an *even* function of h , whereas the cross-covariance in general is not. Here lies the shortcoming of the cross-variogram: there is a potential loss of information. The decomposition of the cross-covariance into even and odd parts

$$C_{ij}(h) = \frac{1}{2}[C_{ij}(h) + C_{ij}(-h)] + \frac{1}{2}[C_{ij}(h) - C_{ij}(-h)]$$

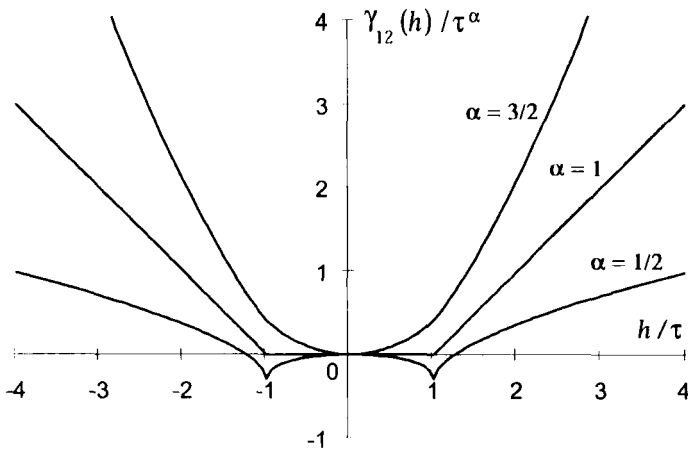


FIGURE 5.6. Cross-variograms of $Z_1(x)$ and $Z_2(x) = Z_1(x + \tau)$ for different power law variograms.

shows exactly what is lost, namely the odd part of the cross-covariance. For definiteness let us consider a case where covariances do not exist.

Example: The Pure Offset Model

Let $Z_1(x)$ and $Z_2(x)$ be two intrinsic random functions where $Z_2(x) = Z_1(x + \tau) + m$. Their cross-variogram satisfies

$$2\gamma_{12}(h) = \gamma_{11}(h + \tau) + \gamma_{11}(h - \tau) - 2\gamma_{11}(\tau)$$

and shows how the offset by τ is symmetrized. Figure 5.6 pictures cross-variograms for $\gamma_{11} = |h|^\alpha$ models in 1D. When $\alpha \leq 1$, the presence of an offset may be detected but its direction cannot be determined; when $\alpha > 1$, detection of an offset is not possible.

In the extreme case where the cross-covariance is odd the cross-variogram is identically zero. For example, the cross-covariance between $Z(\cdot)$ and $Z'(\cdot)$ is the odd function $C'(h)$, and as a consequence $Z(y) - Z(x)$ and $Z'(y) - Z'(x)$ are always uncorrelated. Note in passing that the variogram derivative $\gamma'(h)$ is not a cross-variogram.

As a covariance of increments, $\gamma_{ij}(h)$ is subject to the Cauchy-Schwarz inequality

$$|\gamma_{ij}(h)| \leq \sqrt{\gamma_{ii}(h)\gamma_{jj}(h)} \quad \forall h \quad (5.29)$$

This inequality ensures the continuity of $\gamma_{ij}(h)$ at $h = 0$ provided that $Z_i(x)$ or $Z_j(x)$ is continuous in the mean square, and it also majorizes the growth of

$\gamma_{ij}(h)$. Further the *coefficient of codispersion*

$$R_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(h)\gamma_{jj}(h)}} \quad (5.30)$$

introduced by Matheron (1965) provides an interpretive tool to analyze the correlation between the variations of $Z_i(x)$ and those of $Z_j(x)$.

Expansion of $\text{Var} \sum \lambda_i [Z_i(x+h) - Z_i(x)]$ shows that $[\gamma_{ij}(h)]$ is a positive definite matrix for every h (unlike $[C_{ij}(h)]$). The Cauchy-Schwarz inequality (5.29) is a particular consequence of this, which is by no means sufficient to ensure the validity of the set of cross-variograms, as the following example demonstrates. The true validity criterion is again in the frequency domain.

Example: An Invalid Cross-Variogram Model Satisfying the Cauchy-Schwarz Inequality

Consider the exponential cross-variogram model

$$\gamma_{11}(h) = \gamma_{22}(h) = 1 - e^{-b|h|} \quad \gamma_{12}(h) = -(1 - e^{-c|h|}) \quad (b, c > 0)$$

Whenever $b > c$, these variogram functions satisfy the Cauchy-Schwarz inequality

$$\gamma_{11}(h)\gamma_{22}(h) - \gamma_{12}(h)^2 = (2 - e^{-b|h|} - e^{-c|h|})(e^{-c|h|} - e^{-b|h|}) \geq 0$$

Yet the model is not valid. Indeed under a valid multivariate model the ordinary variogram of any linear combination of $Z_1(x)$ and $Z_2(x)$, expressed in terms of the model, must be a valid variogram function. Consider the sum $Y(x) = Z_1(x) + Z_2(x)$; its variogram is

$$\gamma_{YY}(h) = \gamma_{11}(h) + \gamma_{22}(h) + 2\gamma_{12}(h) = 2[e^{-c|h|} - e^{-b|h|}]$$

Since $b > c$, this function is positive, but it is a suspicious looking variogram that tends to 0 as $h \rightarrow \infty$. A direct proof (not by Fourier) is to consider that the covariance of $Y(x+\tau) - Y(x)$ for an arbitrary constant τ ,

$$\gamma_{YY}(h+\tau) + \gamma_{YY}(h-\tau) - 2\gamma_{YY}(h)$$

must be a positive definite function of h , and therefore bounded in absolute value by $2\gamma_{YY}(\tau)$. But for fixed h and large τ , the first two terms tend to zero so that this function is equivalent to $-2\gamma_{YY}(h)$ which can be greater in magnitude than $2\gamma_{YY}(\tau)$. So $\gamma_{YY}(h)$ is not a variogram and the multivariate model is invalid. The only case⁵ where the model can be valid is $b = c$ with the strange consequence that $\gamma_{YY}(h) \equiv 0$. A simple example of such model is $Z_1(x)$ with an exponential covariance and $Z_2(x) = -Z_1(x)$.

Cokriging Equations in Terms of Cross-Variograms

In the special case where the RFs satisfy the symmetry condition

$$E[Z_i(x) - Z_i(x')][Z_j(y) - Z_j(y')] = E[Z_j(x) - Z_j(x')][Z_i(y) - Z_i(y')]$$

the covariance of any two increments is given by

$$\begin{aligned} E[Z_i(x) - Z_i(x')][Z_j(y) - Z_j(y')] \\ = \gamma_{ij}(y' - x) + \gamma_{ij}(y - x') - \gamma_{ij}(y - x) - \gamma_{ij}(y' - x') \end{aligned}$$

(Matheron, 1965, p. 146) so that in particular

$$\begin{aligned} E[Z_i(x) - Z_i(x_0)][Z_j(y) - Z_j(x_0)] \\ = \gamma_{ij}(x - x_0) + \gamma_{ij}(y - x_0) - \gamma_{ij}(y - x) \end{aligned}$$

From this is it easy to see that the variance of any linear combination of increments

$$\sum_i \sum_\alpha \lambda_{i\alpha} Z_i(x_\alpha) \quad \text{with} \quad \sum_\alpha \lambda_{i\alpha} = 0 \quad \forall i$$

and in particular, error variances, can be expressed in terms of the cross-variogram by simply substituting $-\gamma_{ij}(h)$ for $C_{ij}(h)$, the same rule as used in the univariate case. Thus ordinary and universal cokriging systems can be written in terms of cross-variograms provided that for each variable the sum of weights is explicitly constrained to zero (i.e., even if the variable has mean zero) and of course that the symmetry condition is satisfied.

Generalized Cross-Covariance

What would be the completely general equivalent of the variogram for a cross structure function? In the case of an intrinsic model (of order 0) this would be a function $K_{ij}(h)$ enabling us to calculate the covariance of any pair of increments of the variables $Z_i(x)$ and $Z_j(x)$. Using the notations of IRF- k theory let us define such increments in the concise integral form

$$\begin{aligned} Z_i(\lambda) &= \int Z_i(x) \lambda(dx) & \text{where} & \int \lambda(dx) = 0 \\ Z_j(\mu) &= \int Z_j(y) \mu(dy) & \text{where} & \int \mu(dy) = 0 \end{aligned}$$

Then the structure function should satisfy

$$E[Z_i(\lambda)Z_j(\mu)] = \iint \lambda(dx)K_{ij}(y-x)\mu(dy)$$

When the ordinary cross-covariance $C_{ij}(h)$ exists, it is obviously a solution. But so is the family

$$K_{ij}(h) = C_{ij}(h) - C_{ij}(0) - \langle c_1, h \rangle$$

for any constant vector c_1 . This leads to a multivariate generalization of the IRF-0 theory, in which the structural tool is a class of generalized cross-covariances $K_{ij}(h)$ with the spectral representations (Matheron, personal communication)

$$K_{ij}(h) = \int \frac{e^{2\pi i \langle u, h \rangle} - 1 - i \langle u, h \rangle 1_B(u)}{4\pi^2 |u|^2} \chi_{ij}(du) + c_0 + \langle c_1, h \rangle \quad (5.31)$$

where B is an arbitrary symmetric neighborhood of the origin, $\chi_{ij}(du)$ is a complex measure with the Hermitian symmetry, no atom at the origin, satisfying $\int \chi_{ij}(du)/(1 + 4\pi^2 |u|^2) < \infty$, and the matrix $[\chi_{ij}(B)]$ is positive definite for all (Borel) sets $B \subset \mathbb{R}^n$. The constant c_0 and vector c_1 are arbitrary. A roman "i" is used here to represent the unit pure imaginary number.

Separating the real and imaginary parts of χ_{ij} in (5.31) yields a decomposition of $K_{ij}(h)$ into even and odd parts

$$K_{ij}(h) = [c_0 - \gamma_{ij}(h)] + \left[\langle c_1, h \rangle - \int \frac{\sin(2\pi \langle u, h \rangle) - \langle u, h \rangle 1_B(u)}{4\pi^2 |u|^2} Q_{ij}(du) \right]$$

where Q_{ij} is the imaginary part of χ_{ij} . When this is zero, $K_{ij}(h)$ is reduced to $c_0 - \gamma_{ij}(h)$.

In the case of the pure offset model with variogram $\gamma(h)$ and spectral measure $\chi(du)$, the generalized cross-covariance is of the form $-\gamma(h + \tau)$ for a fixed offset τ , and its spectral measure is $e^{2\pi i u \tau} \chi(du)$.

It does not seem possible to express $K_{ij}(h)$ as the expected value of increments and compute an experimental version of it. One must resort to postulating a parametric model and estimate its parameters like for generalized covariances. Dowd (1989) approaches this in the scope a linear model of coregionalization. The definition of interesting models of generalized cross-covariances deserves further research.

5.6.3. An Example of Structural Analysis with Cross-Variograms

The following example illustrates the use of cross-variograms and codispersion graphs as analysis tools, independently of cokriging. It is extracted from a study of the hydrothermal behavior of the ocean crust in a subduction zone located in the eastern Pacific Ocean (Chilès et al., 1991). Simply put, the objective of the work is to determine if the heat transfer in the ocean crust proceeds from a conductive or a convective regime. A conductive regime, because of thermal refraction, is associated with high heat transfer in deep water and low heat transfer in shallow water, that is, a positive correlation between variations of ocean depth and of heat flux. A convective regime (influx of cold water in deep permeable seabeds, followed by percolation and heating through the overlaying sedimentary layers up to shallower seabeds) results in the opposite effect. Although the site under consideration has been extensively studied, in particular in the scope of the ODP program (Ocean Drilling Program), its hydrothermal system is still not well understood.

The most basic statistical analysis consists of a scatterplot of depth (in m) versus heat flux (in mW/m²), displayed in Figure 5.7a for the western part of the site. It shows a clear negative correlation, but the correlation coefficient is only -0.68 . Geostatistical tools enable us to improve this correlation and also to analyze it more subtly.

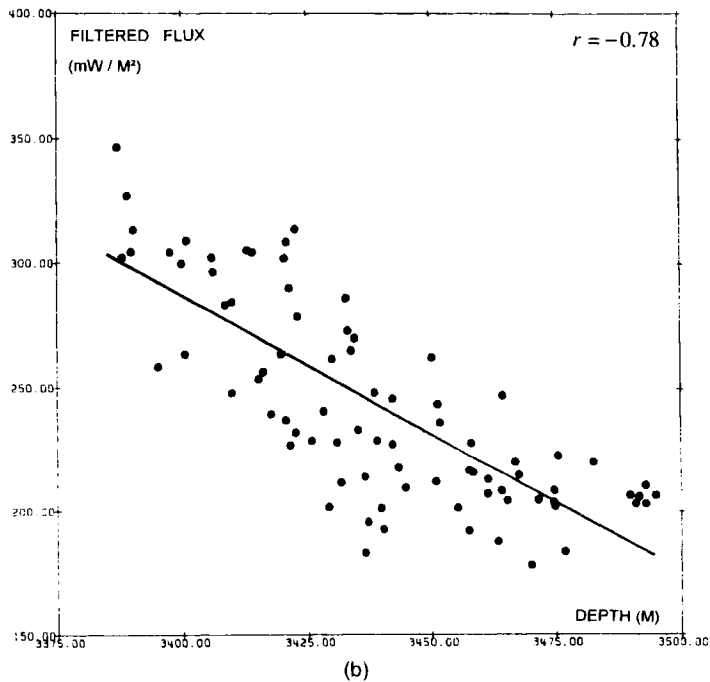
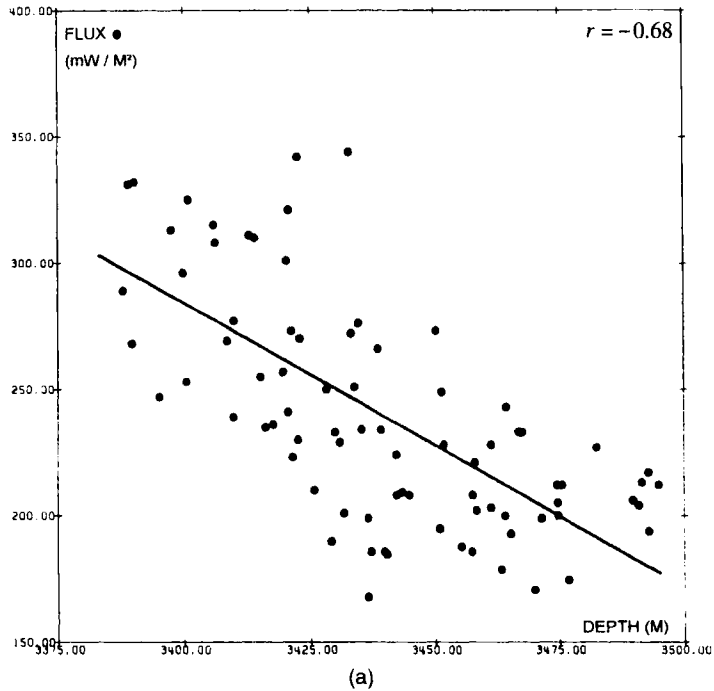


FIGURE 5.7. Scatterplots of heat flux versus depth and regression lines, western zone: (a) unfiltered flux; (b) filtered flux. From Chilès et al. (1991).

The experimental cross-variogram between two variables is computed from all pairs of data points where both variables are known using the unbiased estimator

$$\hat{\gamma}_{ij}(h) = \frac{1}{2N(h)} \sum_{x_\beta - x_\alpha \approx h} [Z_i(x_\beta) - Z_i(x_\alpha)][Z_j(x_\beta) - Z_j(x_\alpha)]$$

where $N(h)$ is the number of pairs (x_α, x_β) with separation h . Data points where only one of the variables is present are simply ignored. If the variables have no or too few samples in common to compute a cross-variogram, one may turn to the cross-covariance, but it is risky. Remember that for a single variable the covariance is biased by the estimation of the mean; for two variables things may be worse as we subtract different means computed from data at different places. To ensure numerical consistency and avoid correlation coefficients greater than one, it is also recommended to compute the cross and direct variograms from the same set of points.

In the present case study the sampling requirement is fulfilled with the help of a little trick, a preliminary kriging interpolation to determine the depth at every point where the flux is known. This estimation having a good precision the kriging error may be neglected. Figure 5.8a to c shows the direct and cross-variograms for depth (bathymetry) and heat flux in the western zone. These variograms are strongly anisotropic: the variations are more important in the N–S direction than in the E–W direction. This apparent anisotropy can also be interpreted as the effect of a N–S drift, but for our purpose we can leave the choice of an interpretation open.

The depth variogram is continuous at the origin. The heat flux variogram has the same shape but exhibits a nugget effect due to uncertainties on flux measurements. The flux measurement error variance has a value of about 300 to 400 (mW/m²)², which corresponds to an error standard deviation of 17 to 20 mW/m², for a mean flux value of about 220 mW/m². The cross-variogram between depth and flux confirms the negative correlation between these variables. The interesting new feature is that the correlation is weak, if not negligible, along the E–W direction; at the same time the correlation in the N–S direction is stronger than what the correlation coefficient indicates. The coefficient of codispersion $R(h)$, defined by (5.30), is graphed in Figure 5.8d. At medium to large distances the good correlation between depth variations and flux variations in the N–S direction shows up clearly.

Improvement of Correlation by Filtering of Measurement Error

The measurement errors on heat flux do not affect the covariance between flux and depth nor the cross-variogram. But they increase the scatter in the cross-plots and inflate the variance and variogram of the flux, and therefore reduce the correlation and codispersion coefficients, especially at short distances. There are two possibilities to correct for measurement errors:

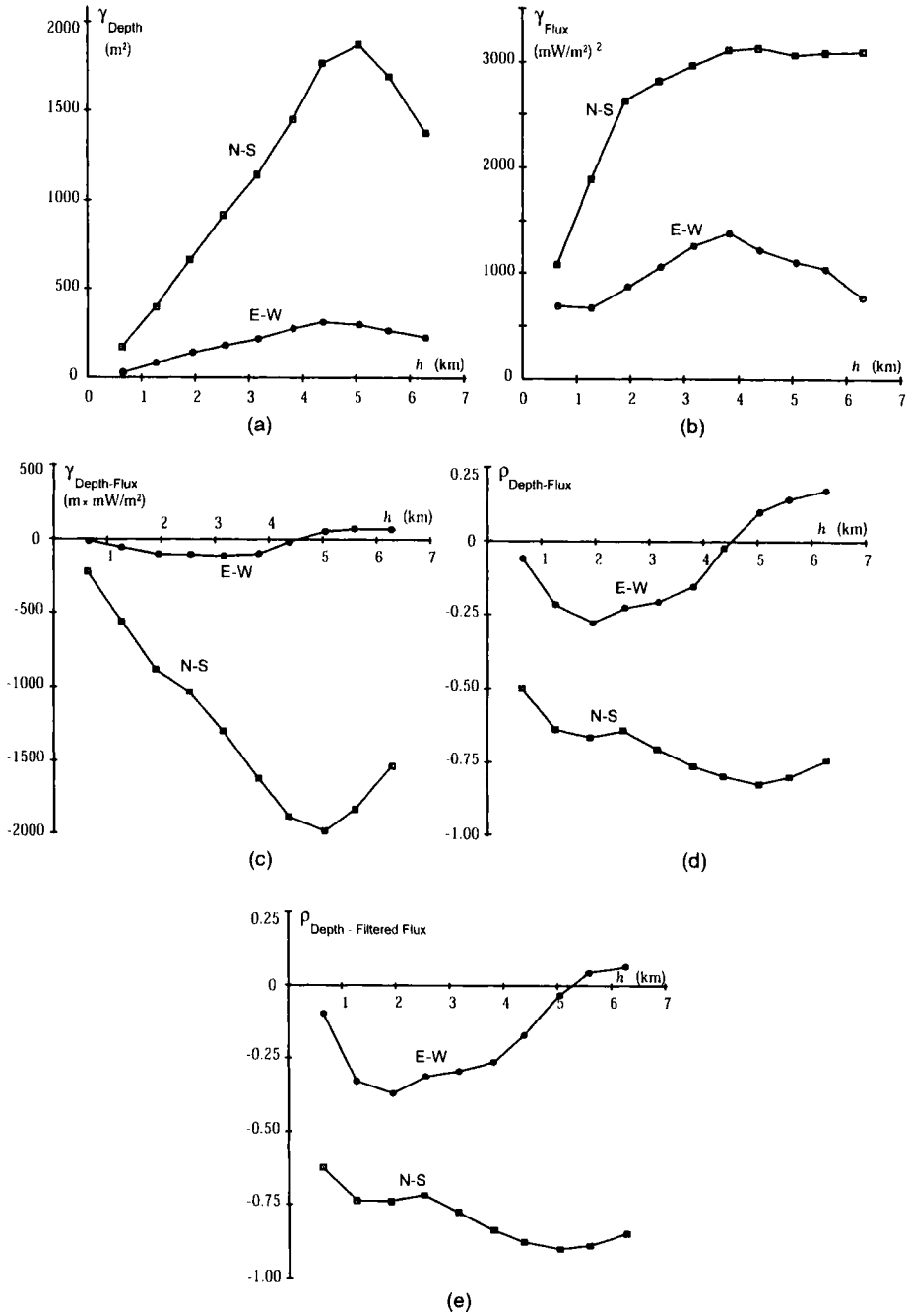


FIGURE 5.8. Direct and cross-variograms of depth and heat flux, western zone: (a) depth variogram; (b) flux variogram; (c) cross-variogram of depth and flux; (d) codispersion graph; (e) codispersion graph of depth and filtered heat flux (in *ibid.*).

- Subtract the measurement error variance from the variance and variogram of the flux, and recompute the correlation coefficient and the codispersion function.
- Filter the flux data themselves by the method of kriging with filtering of the nugget effect, and recompute the statistical parameters and variograms.

In the second approach the effect of measurement errors is considerably reduced, without being completely eliminated because the neighboring points used for kriging are only “pseudoreplicates” and they involve some lateral variations. But the first approach is very dependent on the fit of the nugget effect, and so the second method is preferred.

Figure 5.7b displays the new depth-flux scatterplot for the western zone. Observe that it is more packed along the regression line. The correlation coefficient improves from -0.68 to -0.78 . The variogram of the filtered flux (not shown here) shows that the effect of the correction is to practically eliminate the nugget effect. Thus the new codispersion function $R(h)$ can be considered unaltered by measurement errors, at least at large distances; at short distances, where variations remain small, residual errors, even small ones, can produce an instability. The graph of $R(h)$, shown in Figure 5.8e, confirms the preceding analysis: low correlation in the E–W direction and high correlation in the N–S direction where $R(h)$ stabilizes around -0.86 . This correlation of -0.86 between variations of depth and of flux when the original correlation coefficient of the raw data is -0.68 confirms the convective behavior in the western zone, at least along the N–S direction.

In the example presented only one of the variables was subject to noise. But the technique of spatial filtering provides a solution to a more challenging statistical problem: determine the relationship between two variables when both are subject to measurement errors. The raw data scatterplot may be completely blurred by noise while the plot of filtered data may be focused enough to be interpretable.

5.6.4. Proportional Covariance Model and Screening Property

The proportional covariance model is the simplest multivariate model used in geostatistics. All covariances (variograms) are proportional to the same covariance (variogram) function

$$C_{ij}(h) = b_{ij}C(h) \quad \text{or} \quad \gamma_{ij}(h) = b_{ij}\gamma(h) \quad (5.32)$$

with symmetric coefficients b_{ij} that define a positive definite matrix $\mathbf{B} = [b_{ij}]$. The conditions on \mathbf{B} result directly from the symmetry $C_{ij}(h) = C_{ji}(-h)$ built in this model (hence $C_{ij}(h) = C_{ji}(h)$) and the Cramér criterion of Section 5.6.1. It is convenient to scale the basic structure function to $C(0) = 1$ so that \mathbf{B} is the matrix of covariances of the variables at the same point, or of correla-

tion coefficients if all variables are scaled to unit variance. If the matrix \mathbf{B} is *strictly* positive definite ($\lambda' \mathbf{B} \lambda > 0$ for $\lambda \neq 0$), it has an inverse, and the cokriging system is always nonsingular. If \mathbf{B} were singular, it would suffice to consider only a subset of linearly independent variables and deduce the others through the linear relationships. This model reduces the determination of $p(p+1)/2$ covariance (variogram) *functions* to the determination of one covariance (variogram) function and $p(p+1)/2$ *numbers*.

Matheron (1965) introduced the proportional covariance model (which he named “intrinsic correlation” model) not for the purpose of cokriging but to validate the usual statistical correlation coefficient in a geostatistical context. Why is there a problem? Because, when variables are spatially correlated, their variance or covariance within a finite domain V depends on V . We saw, for example, that the empirical variance s^2 is an estimate of $\sigma^2(0|V)$, the variance of a point within V , which is smaller than the variance of $Z(x)$ if it exists (Section 2.8.2). The general formula for the covariance between the point values $Z_i(x)$ and $Z_j(y)$ within V , including the case $i = j$, is

$$\sigma_{ij}(0|V) = \frac{1}{V^2} \int_V \int_V \gamma_{ij}(y-x) dx dy$$

and the correlation coefficient of the variable i and j within V is

$$\rho_{ij} = \frac{\sigma_{ij}(0|V)}{\sqrt{\sigma_{ii}(0|V)\sigma_{jj}(0|V)}} \quad (5.33)$$

This coefficient depends on V . It may tend to a limit when V becomes infinite, but this limit in general depends on the particular way V tends to infinity. Now, when all variograms are proportional, (5.33) becomes

$$\rho_{ij} = \frac{b_{ij}}{\sqrt{b_{ii}b_{jj}}}$$

and this coefficient reflects the relationship between the two variables independently of the domain V . In this case the coefficient of codispersion $R_{ij}(h)$ does not vary with h and is equal to ρ_{ij} . By a similar argument we can see that the proportional covariance structure is preserved when the same change of support is applied to all variables.

A proportional covariance model can be generated by a linear combination of p mutually orthogonal RFs $Y_j(x)$ with the same covariance $C(h)$, or the same variogram $\gamma(h)$

$$Z_i(x) = \sum_{j=1}^p A_{ij} Y_j(x)$$

where the matrix $\mathbf{A} = [A_{ij}]$ satisfies $\mathbf{B} = \mathbf{A}\mathbf{A}'$ and is determined by principal component analysis (Section 5.6.6). This expression is established for zero-mean RFs but remains valid when drifts are present, provided that the same set of basis drift functions is used for all variables Z_i and Y_j (i.e., $f_i^l(x) = f^l(x)$) and that coefficients are algebraically independent (\mathbf{A} being nonsingular). This decomposition is often used to co-simulate correlated Z_i by simulating p independent RFs Y_j .

The proportional covariance model has the following screening property: *if all variables are measured at the same points the cokriging estimator of a variable coincides with the kriging estimator based on that variable alone.* The secondary variables are “hidden” by the primary variable and receive zero weights. Indeed, by linearity of the cokriging estimator

$$Z_i^{**}(x_0) = \sum_{j=1}^p A_{ij} Y_j^{**}(x_0)$$

Since all RFs $Y_j(x)$ are orthogonal and have algebraically independent drifts, the cokriging estimator of Y_j coincides with the kriging estimator based on that variable alone

$$Y_j^{**}(x_0) = Y_j^*(x_0) = \sum_{\alpha} \lambda_{j\alpha} Y_j(x_{\alpha})$$

Now, and this is the crux of the argument, if all covariances (variograms), drift functions, and point configurations are identical, the weights $\lambda_{j\alpha}$ are also identical and do not depend on j . So

$$Z_i^{**}(x_0) = \sum_{\alpha} \lambda_{\alpha} \sum_{j=1}^p A_{ij} Y_j(x_{\alpha}) = \sum_{\alpha} \lambda_{\alpha} Z_i(x_{\alpha}) = Z_i^*(x_0)$$

The cokriging estimator of Z_i coincides with the kriging estimator based on that variable alone. The same conclusion would hold for the estimation of the means.

The requirement that all variables be sampled equally is essential to ensure that the kriging weights are the same. However, as noted by Helterbrand and Cressie (1994), the screening property persists if an additional observation of the primary variable is available without observations of the other variables. There is a simple explanation for this: whenever a secondary variable is missing, we can always by thought give it an arbitrary value; once equal sampling is achieved, the screening property applies, making values of secondary variables irrelevant.

The screening property also persists under the much weaker conditions

$$\mathbf{C}_{i1}(h) = b_{i1} \mathbf{C}_{11}(h) \quad \text{or} \quad \gamma(h) = b_{i1} \gamma_{11}(h)$$

which generalize the “Markov-type” model of Section 5.4.3. Note that these do not involve the correlation structure between secondary variables. (Proof from general cokriging equations (5.4) with $C_{i1} = b_{i1}C_{11}$ and $F_i = F$, letting $\lambda_i = 0$ for $i \neq 1$). But the screening property is usually destroyed by the presence of a nugget effect on the primary variable (see Section 5.4.4).

In applications the secondary variables are often sampled more densely than the primary variable. A common practice is to use a regression equation to transform data of the secondary variables into pseudo-observations of the primary variable. A better procedure proposed by Maréchal (1970) is the following: first, at the secondary data locations where they are missing reconstruct the primary data by (possibly collocated) cokriging, next use the screening property to carry out kriging at other points on the basis of the primary data alone. This is essentially a multivariate version of the filling-in the gaps procedure seen in Section 3.5.6, exploiting the screening property of the proportional covariance model. Naturally the kriging variance thus obtained underestimates the actual cokriging variance.

The considerable simplifications brought by the proportional covariance model should not hide the fact that it is a very special model. For example, it is practically incompatible with the presence of a nugget effect, because then it must be present also in cross-covariances, while microstructures usually do not cross-correlate. Another limitation of the model is its behavior under a change of support. Since the correlation between variables does not vary with the support it cannot be improved by averaging over larger volumes, while such improvement may be observed. In view of this, the proportional covariance model is often used as part of another model presented next.

5.6.5. Linear Model of Coregionalization

The linear model of coregionalization is a sum of proportional covariance models. In matrix notations where $C(h) = [C_{ij}(h)]$ is the $p \times p$ covariance matrix and similarly $\Gamma(h) = [\gamma_{ij}(h)]$, this model takes the simple form

$$C(h) = \sum_{k=1}^s \mathbf{B}_k C_k(h) \quad \text{or} \quad \Gamma(h) = \sum_{k=1}^s \mathbf{B}_k \gamma_k(h) \quad (5.34)$$

The explicit form is somewhat clumsy because triple indexes are involved but hopefully the following will be clear

$$C_{ij}(h) = \sum_{k=1}^s b_k(i, j) C_k(h) \quad \text{or} \quad \gamma_{ij}(h) = \sum_{k=1}^s b_k(i, j) \gamma_k(h)$$

In this model all covariances (variograms) are linear combinations of the same basic structures, indexed by k . If a particular elementary model is not present,

its coefficient is set to zero; with this convention all covariances (variograms) comprise the same number s of structures. A sufficient condition for the model to be valid is that for each k the matrix \mathbf{B}_k of coefficients be positive definite.

The name "linear model" originates from the fact that the above covariance structure can be obtained by linearly combining multivariate random functions $\mathbf{Y}_k(x) = (Y_{1k}(x), \dots, Y_{pk}(x))'$ (the coregionalizations), whose components have a common structure function $C_k(h)$ or $\gamma_k(h)$ and where all RFs with different indexes are orthogonal. Indeed

$$\mathbf{Z}(x) = \sum_{k=1}^s \mathbf{A}_k \mathbf{Y}_k(x) \Rightarrow (6.34) \quad \text{with} \quad \mathbf{B}_k = \mathbf{A}_k \mathbf{A}_k'$$

The functions $C_k(h)$ or $\gamma_k(h)$ are supposed known, and the $p \times p$ matrices \mathbf{B}_k are to be estimated. Each component $C_k(h)$ or $\gamma_k(h)$ is associated with a certain structure scale, typically a nugget effect, a short- or medium-range model, and a long-range model. In practice, the number of structures does not exceed three because of the difficulty of clearly separating scales from an empirical variogram. When the elementary structures are scaled to unity ($C_k(0) = 1$), the sum of all \mathbf{B}_k represents the covariance matrix of the variables at the same point.

By construction of the model, all cross-covariances are symmetric, and cross-variograms are therefore the correct structural tool. The same ranges are present in all variograms, which constitutes a diagnostic criterion for considering a linear model. Due to the positivity assumption, for every (i, j) and every k ,

$$|b_k(i, j)| \leq \sqrt{b_k(i, i) b_k(j, j)} \quad (5.35)$$

in words, every basic structure present in the cross-covariance (variogram) of the variables i and j must also be present in the two direct covariances (variograms) of i and j , but the converse is not true.

Due to its simplicity and relative versatility, the linear model of coregionalization has received the most attention. Many applications can be found in the geostatistical literature. Journel and Huijbregts (1978, pp. 256ff.), after a study by P. Dowd, 1971, show how the cross-variograms of lead, zinc, and silver in the Broken Hill mine can be nicely modeled by two basic structures, a nugget effect, and a spherical variogram with a range of 60 feet (so here $p = 3$ and $s = 2$). Wackernagel (1985, 1988) was able to fit the 120 variograms from 15 geochemical variables using just two structures (nugget effect and spherical model with a 5-km range). Daly et al. (1989) model microprobe X-ray images of six chemical elements with mixtures of spherical and cubic variograms and use this to optimize linear filtering of the noisy images. But the most exotic of this incomplete list must be the study by Steffens (1993) of animal abundance in the Kruger National Park in South Africa, where he analyzes variograms and cross-variograms of giraffe, impala, kudu, warthog, blue wilde-

beest, and zebra counts. We learn that zebra variograms have a very high nugget effect, while warthogs and kudus are negatively correlated. A spherical model (range 4 km) with a nugget effect fits these wildlife variograms reasonably well.

The general procedure for fitting the model is to postulate the number and the shapes of the elementary models, including their ranges, and attempt a fit of the sills/slopes either by trial and error or by some numerical optimization technique. A difficulty is to ensure the positivity of the coefficient matrices \mathbf{B}_k . Goulard (1989) and Goulard and Voltz (1992) propose an iterative algorithm to fit these coefficients by least squares under positivity constraints. Defining the empirical and model cross-variogram matrices

$$\hat{\Gamma}(h) = [\hat{\gamma}_{ij}(h)] \quad \text{and} \quad \Gamma(h) = [\gamma_{ij}(h)] = \sum_k \mathbf{B}_k \gamma_k(h)$$

the goodness-of-fit criterion is a weighted sum of squares (WSS) of all terms of the error matrix $\hat{\Gamma}(h) - \Gamma(h)$, summed over the set of lags J used for the fit. Specifically, it is the Euclidean norm

$$\text{WSS} = \sum_{h \in J} w(h) \text{Trace}[\mathbf{V}(\hat{\Gamma}(h) - \Gamma(h))^2]$$

The weights $w(h)$ are positive and typically equal to the number of pairs used for variogram estimation at lag h . The matrix \mathbf{V} is a positive-definite matrix designed to equalize the influence of variables, typically the diagonal matrix of inverse variances—or the identity. The idea is to minimize the criterion by optimizing one \mathbf{B}_k at a time and to repeat this until no improvement is possible. The residual for the current fit less the k th term is

$$d\Gamma_k(h) = \hat{\Gamma}(h) - \sum_{u \neq k} \mathbf{B}_u \gamma_u(h)$$

In the absence of positivity constraint, the optimal fit of $d\Gamma_k$ by $\mathbf{B}_k \gamma_k(h)$ is obtained with

$$\mathbf{B}_k = \left(\frac{1}{\alpha_k} \right) \sum_{h \in J} w(h) \gamma_k(h) d\Gamma_k(h) \quad \text{where} \quad \alpha_k = \sum_{h \in J} w(h) \gamma_k(h)^2$$

The constrained solution $\mathbf{B}_k^+ \geq 0$ is the positive definite matrix nearest to \mathbf{B}_k . Being symmetric the matrix \mathbf{B}_k has a spectral decomposition of the form

$$\mathbf{B}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k' \quad \text{with} \quad \mathbf{U}_k' \mathbf{V} \mathbf{U}_k = \mathbf{I}_p$$

where \mathbf{U}_k is a matrix of eigenvectors of $\mathbf{B}_k \mathbf{V}$ and Λ_k is the diagonal matrix of its eigenvalues. The constrained solution is then

$$\mathbf{B}_k^+ = \mathbf{U}_k \Lambda_k^+ \mathbf{U}_k'$$

where Λ_k^+ is the matrix Λ_k in which all negative eigenvalues are replaced by zeros. This algorithm converges to a unique solution which is always reached and is independent of the starting point. This results from the strict convexity of the minimized criterion under a natural design condition, namely the linear independence of the variogram functions $\gamma_k(h)$ over the set of lags used for the fit

$$\sum_k c_k \gamma_k(h) = 0 \quad \forall h \in J \quad \Rightarrow \quad c_k = 0 \quad \forall k$$

Usually the basic variogram models are not collinear, and this condition is satisfied automatically.

5.6.6. Factorial Kriging Analysis

We have seen that a linear model of coregionalization is associated with an orthogonal decomposition of the form $\mathbf{Z}(x) = \sum_k \mathbf{A}_k \mathbf{Y}_k(x)$. A method known as factorial kriging analysis, (or factorial kriging for short) permits the determination of the components $\mathbf{Y}_k(\cdot)$ on the basis of $\mathbf{Z}(\cdot)$, in the hope that these components represent different genetic structures. For example, in magnetic or gravimetric surveys one wants to distinguish long wavelengths, associated with deep sources, from short wavelengths, reflecting shallower sources. This separation is traditionally performed on the basis of the two-dimensional power spectrum but a direct spatial determination is possible and even advantageous.

In the multivariate case, factorial kriging analysis separates a p -dimensional RF $\mathbf{Z}(x)$ into several p -dimensional components, each representing a different scale (or wavelength) of the phenomenon, and then reduces the dimensionality of each component to a few composite variables.

Matheron (1982a) developed the general theory of factorial kriging analysis within the IRF- k framework (using generalized covariances), but we will present it with ordinary covariances.

Univariate Case

The model is of the form

$$Z(x) = \sum_{k=1}^s a_k Y_k(x) \quad (5.36)$$

where the a_k are known coefficients⁶ and the $Y_k(x)$ mutually orthogonal and scaled RFs with covariances $C_k(h)$. None of the $Y_k(x)$ is directly observable

but can be estimated by cokriging from observations of $Z(x)$ provided that we know the cross-covariance between $Z(x)$ and $Y_k(x)$, which we do: it is $a_k C_k(h)$. The only open question is how to distribute the drift between the different components. Two options make the most sense: (1) assume that all $Y_k(x)$ have mean zero and that the decomposition (5.36) holds for the residual $Z(x) - m(x)$, or (2) assign the drift to the component with the largest range, say the last one. Only the unbiasedness conditions change. The resulting cokriging system for the estimation of $Y_k(x)$ is given below for these two cases

$$\begin{cases} \sum_{\beta} \lambda_{\beta} C(x_{\beta} - x_{\alpha}) + \sum_l \mu_l f_{\alpha}^l = a_k C_k(x_0 - x_{\alpha}) & \alpha = 1, \dots, N \\ \sum_{\alpha} \lambda_{\alpha} f_{\alpha}^l = 0 \text{ or } = f_0^l / a_s \text{ for } k = s \text{ in case (2)} & l = 0, \dots, L \end{cases}$$

If the mean of Z were known and subtracted out, there would of course be no unbiasedness condition at all, and no Lagrange parameter. Constraining the weights to zero has the effect of filtering out this mean without requiring an explicit estimation. Note that the left-hand side of the system remains the same for the estimation of all components k , since the estimators are based on the same data, so that the work may be done in parallel. A standard kriging program may be used with a modified right-hand side—just as for filtering a nonsystematic error, which is a particular case of factorial kriging. In case (2), since $C(x_0 - x_{\alpha}) = \sum_k a_k^2 C_k(x_0 - x_{\alpha})$ the consistency relationship $Z^* = \sum a_k Y_k^{**}$ automatically holds.⁷

Applications of this approach include geochemical prospecting (Sandjiv, 1984), geophysical prospecting (Galli et al., 1984a; Chilès and Guillen, 1984), remote sensing (Ma and Royer, 1988), and petroleum exploration (Jaquet, 1989; Yao and Mukerji, 1997). We will extract some results from the above study of gravity data by Chilès and Guillen to illustrate the method. The experimental variogram of gravity data in this basin is shown in Figure 5.9. An excellent fit is obtained with a sum of two generalized Cauchy models (2.55) with shape parameter $\frac{1}{2}$, especially suited to represent gravimetry data: a structure with scale parameter 5 km and sill 5 mgal² and another with scale parameter 11.4 km and sill 53 mgal², plus a very small measurement error nugget effect. In this variogram model the scale parameter represents twice the average depth of the sources, which are therefore located at 2.5 km and 5.7 km, in agreement with local geological knowledge. Figure 5.10 shows the logarithm of the isotropic power spectrum and the fit made by the geophysicist—don't ask how! The slopes of the lines indicate the depths of the sources: 1.5 and 4.5 km, consistent with the variogram fit. In order to compute this spectrum, the data had first to be interpolated to a regular grid, which was achieved by kriging using the fitted variogram. The kriged map and standard deviation are displayed in Figure 5.11. The decomposition of the kriged map into deep and shallow components is shown in Figures 5.12 and 5.13, on the right. Note in 5.13 that the anomaly in the upper right-hand corner coincides with

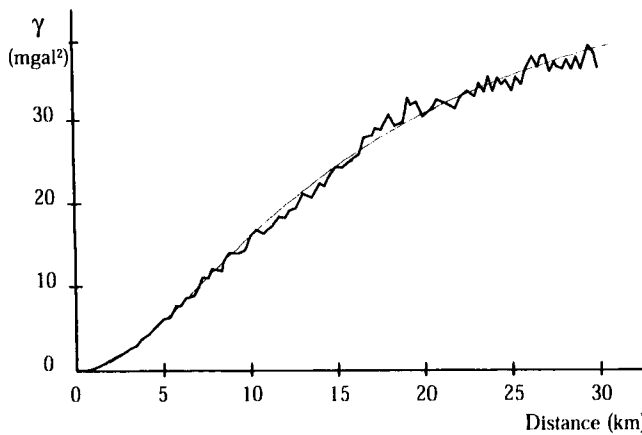


FIGURE 5.9. Sample variogram of gravity data and fit with two nested generalized Cauchy models. From Chilès and Guillen (1984).

a maximum of the kriging standard deviation, and is therefore dubious. In performing the cokriging estimation, the neighborhood search algorithm must be customized to include near and far data so as to allow the separation of scales.

The results obtained with the spectral method are shown on the left-hand sides of Figures 5.12 and 5.13. The spectral and geostatistical maps are equivalent. The deep-field spectral map is smoother because of the global character of the spectral approach. Note that the separation between deep and shallow components, as well as the interpretation of gravimetric anomalies, are up to a constant. But the two methods do not correspond to the same constant. For geostatistical filtering the mean of the shallow component has been set to zero while it is a positive value for the spectral method.

The spectral method has some practical limitations that factorial kriging does not have:

1. It requires a complete rectangular grid of data; if these data are not acquired on a grid, or if the rectangle is not complete, a preliminary interpolation, or worse, extrapolation, is needed which already involves kriging (or similar) and alters the spectral characteristics.
2. It requires tedious tapering or padding to make the input grid periodic (wrap around in both directions).
3. It smears a local anomaly over all frequencies.

On the other hand, some phenomena, typically acoustic waves, are better understood in the frequency domain.

Regardless of the method used, the separation of a field into spatial components must be supported by a physical model.

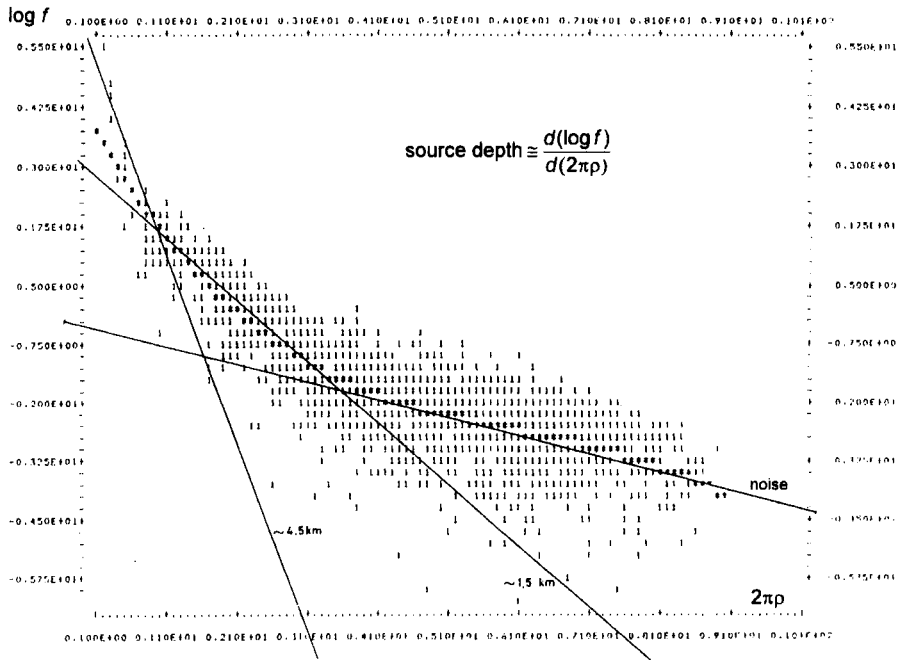


FIGURE 5.10. Logarithm of the isotropic power spectrum of gravity data (in *ibid.*).

Multivariate Case

In the multivariate version of the model we assume that the means of the \mathbf{Y}_k vectors are all zeros and write the model as $\mathbf{Z} = \mathbf{m} + \sum \mathbf{A}_k \mathbf{Y}_k$, or in explicit notations

$$Z_i(x) = m_i(x) + \sum_{k=1}^s \sum_{j=1}^p A_k(i, j) Y_{jk}(x) \quad (i = 1, \dots, p) \quad (5.37)$$

where the $Y_{jk}(x)$ are mutually uncorrelated RFs called *factors*. In reality only $\mathbf{B}_k = \mathbf{A}_k \mathbf{A}_k'$ is directly estimable and \mathbf{A}_k is not uniquely defined. Given a symmetric positive definite matrix \mathbf{B} , one can find an infinite number of matrices \mathbf{A} such that $\mathbf{B} = \mathbf{A} \mathbf{A}'$. Principal component analysis, however, provides a natural determination of the matrices \mathbf{A}_k . Then the individual components $Y_{jk}(x)$ can be estimated by cokriging. When the drifts $m_i(x)$ are algebraically independent, the general equations (5.4) apply with the index pair (j, k) in the role of the index $i = 1$. The vectors \mathbf{c}_{i0} are the N_i -vectors of the covariances between $Z_i(x_\alpha)$ and $Y_{jk}(x_0)$ for $x_\alpha \in S_i$, namely the vectors $(A_k(i, j) C_k(x_0 - x_\alpha)) : x_\alpha \in S_i$, and the unbiasedness conditions are $\mathbf{F}_i' \boldsymbol{\lambda}_i = \mathbf{0}$ for all i (special case of a primary variable with no observation of it but a zero mean).

Now we need to elaborate on the definition of these “factors” that we are attempting to estimate. To this end let us focus on a single scale and drop

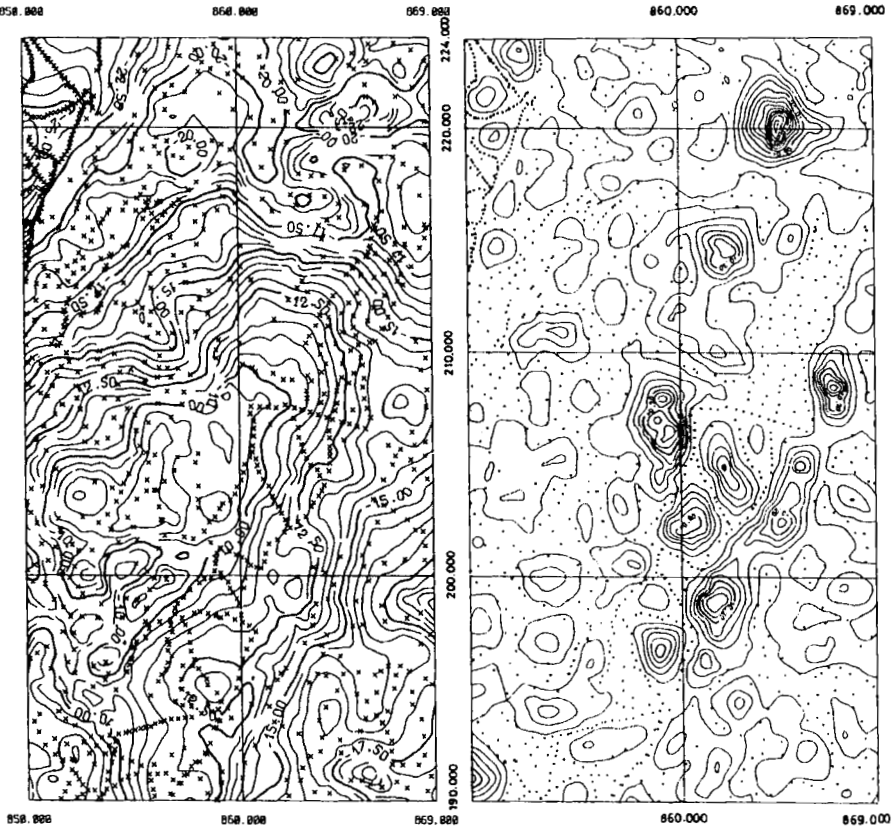


FIGURE 5.11. Kriged gravity field (left) and error standard deviation (right) (in *ibid.*).

the index k . Thus we are now dealing with a pure proportional covariance model. Let $\lambda_1 \geq \dots \geq \lambda_p > 0$ be the eigenvalues of the covariance matrix \mathbf{B} and $\mathbf{u}_1, \dots, \mathbf{u}_p$ the associated eigenvectors, where $\mathbf{u}_i' \mathbf{u}_j = \delta_{ij}$. The \mathbf{u}_i , called the *principal axes*, define a new orthonormal coordinate system diagonalizing \mathbf{B} :

$$\mathbf{B} = \sum_{j=1}^p \lambda_j \mathbf{u}_j \mathbf{u}_j' \quad \mathbf{I} = \sum_{j=1}^p \mathbf{u}_j \mathbf{u}_j' \quad \mathbf{u}_i' \mathbf{B} \mathbf{u}_j = \lambda_j \delta_{ij} \quad (5.38)$$

The *principal components* are the coordinates of \mathbf{Z} in this new system. They are uncorrelated linear combinations $\mathbf{u}_j' \mathbf{Z}$ with variances λ_j . The scaled quantities

$$Y_j = \frac{\mathbf{u}_j' \mathbf{Z}}{\sqrt{\lambda_j}}$$

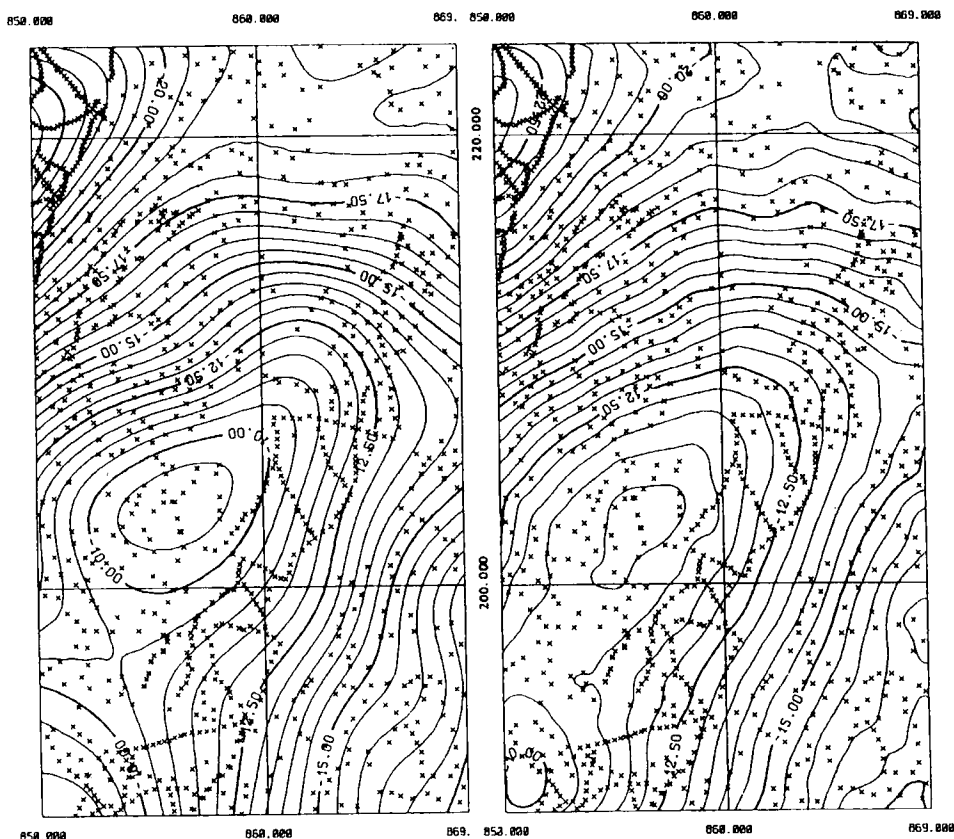


FIGURE 5.12. Estimates of deep gravity field: spectral (left), geostatistical (right) (in *ibid.*).

are then uncorrelated random variables with a unit variance. From (5.38), \mathbf{Z} can be expressed as

$$\mathbf{Z} = \sum_{j=1}^p (\mathbf{u}_j \mathbf{u}_j') \mathbf{Z} = \sum_{j=1}^p (\mathbf{u}_j' \mathbf{Z}) \mathbf{u}_j = \sum_{j=1}^p \sqrt{\lambda_j} Y_j \mathbf{u}_j$$

or, explicitly reintroducing the dependence on x and letting u_{ij} be the i th coordinate of \mathbf{u}_j ,

$$Z_i(x) = \sum_{j=1}^p A(i, j) Y_j(x) \quad \text{with} \quad A(i, j) = \sqrt{\lambda_j} u_{ij}$$

The interesting feature of this decomposition is the possibility to reduce it to the first $q < p$ terms without losing much information, that is, if the first q eigenvalues account for most of the dispersion in the data (represent a high

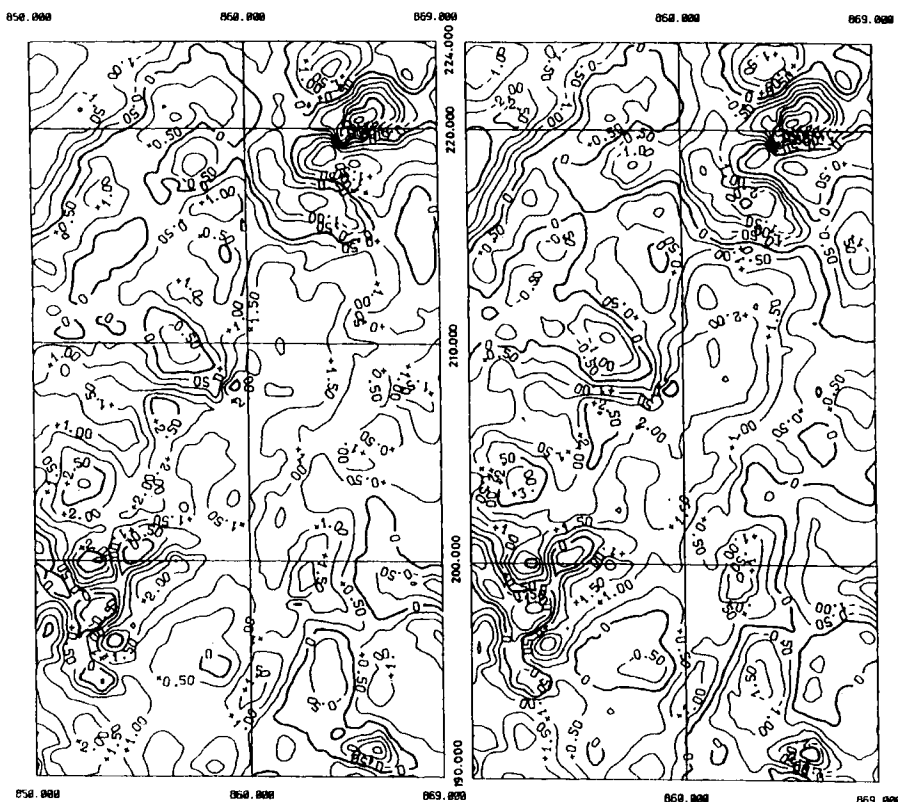


FIGURE 5.13. Estimates of shallow gravity field: spectral (left), geostatistical (right) (in *ibid.*).

fraction of the trace of \mathbf{B}). Typically the first two terms will account for 80% of the variance and the first three for 90% or more. Thus the Y_j can be interpreted as the *common factors* underlying the different observed variables.

To summarize the approach, we first compute direct and cross-covariances (variograms), fit a linear model of coregionalization, perform a PCA on each matrix \mathbf{B}_k and deduce the coefficients $A_k(i, j)$ in (5.37), and finally estimate the leading factors $Y_{jk}(x)$ by cokriging.

The proportional covariance model has the remarkable property that the factors remain uncorrelated at any distance h and not just at the same location. Indeed

$$\text{Cov}[Y_i(x), Y_j(x+h)] = \frac{\mathbf{u}_i' \mathbf{C}(h) \mathbf{u}_j}{\sqrt{\lambda_i \lambda_j}} = C(h) \frac{\mathbf{u}_i' \mathbf{B} \mathbf{u}_j}{\sqrt{\lambda_i \lambda_j}} = C(h) \delta_{ij}$$

The factors are orthogonal as *functions* and achieve a spatially consistent decomposition of $\mathbf{Z}(x)$. Such decomposition is not possible with a general covariance matrix because the eigenvectors of $\mathbf{C}(0)$ do not remain orthogonal

with respect to $C(h)$. A standard principal component analysis (PCA) produces factors that are only uncorrelated at zero distance. The novelty of factorial kriging analysis is to enable a *scale-dependent* analysis of the correlation structure among variables. A separate PCA is carried out for each scale component rather than the raw observations, namely for each matrix \mathbf{B}_k rather than the variance-covariance matrix. Thus some of the spatial information revealed by the variograms is introduced in the PCA and the factors are expected to be more interpretable. The paper of Goovaerts et al. (1993) provides a good illustration of the method. Other applications to geochemical data include, for example, Dousset and Sandjiv (1987), Wackernagel (1988), Bourgault and Marcotte (1993), Goovaerts and Sonnet (1993), and to soil science Wackernagel et al. (1988).

5.6.7. Other Models

In this section we give a brief overview of three more generic models that can be of interest. Another one, indicator cokriging, is covered in Section 6.2.4.

Cross-Covariances Derived from Complex Models

The linear model of coregionalization has the following limitations:

- Direct and cross-covariances are even functions.
- There is limited flexibility for modeling cross-covariances: by the Cauchy-Schwarz inequality (5.35) any term included in the cross-covariance must also be present in the two direct covariances.

An innovative method to extend the model has been proposed by Grzebyk (1993), based on work by Lajaunie and Béjaoui (1991) to adapt kriging to the case of complex variables, or directional data in \mathbb{R}^2 (complex kriging). The idea is to consider the real part of a complex covariance model. If $Z(x)$ is a complex RF, with mean zero to simplify the presentation, its covariance

$$C(h) = E Z(x) \overline{Z(x+h)}$$

is a complex function with the Hermitian symmetry $C(-h) = \overline{C(h)}$, whose Fourier transform is real and nonnegative. Consider now a complex proportional covariance model

$$C_{ij}(h) = b_{ij} C(h)$$

where $\mathbf{B} = [b_{ij}]$ is a complex positive definite matrix with the Hermitian symmetry $b_{ji} = \overline{b_{ij}}$. That this is a valid model follows directly from the characterization criterion in Section 5.6.1; for the same reason the complex conjugate model is also valid and therefore the sum

$$\frac{1}{2} [C_{ij}(h) + \overline{C_{ij}(h)}] = \text{Re}(C_{ij}(h))$$

So the real part of the complex covariance matrix $\{C_{ij}(h)\}$ is a valid covariance model. The expression of these covariances in terms of the real and imaginary parts of $C(h)$ and b_{ij} is given by

$$\text{Re}(C_{ij}(h)) = \text{Re}(b_{ij})\text{Re}(C(h)) - \text{Im}(b_{ij})\text{Im}(C(h))$$

To go further some explicit model of complex covariance must be selected. The above authors propose the following:

$$\operatorname{Re}(C(h)) = \sigma(h) \quad \operatorname{Im}(C(h)) = \frac{1}{2} \sum_{\theta} p_{\theta} [\sigma(h - \tau_{\theta}) - \sigma(h + \tau_{\theta})]$$

where $\sigma(h)$ is a real covariance function, $\{\tau_{\theta}\}$ a family of translation vectors in \mathbb{R}^n and $\{p_{\theta}\}$ a set of positive constants satisfying $\sum p_{\theta} \leq 1$ (sufficient but not necessary condition). Dropping the Re sign and renaming the real and imaginary parts of \mathbf{B} as \mathbf{G} and \mathbf{H} , respectively, the cross-covariance model is of the form

$$C_{ij}(h) = g_{ij}\sigma(h) - h_{ij} \frac{1}{2} \sum_{\theta} p_{\theta} [\sigma(h - \tau_{\theta}) - \sigma(h + \tau_{\theta})] \quad (5.39)$$

Diagonal terms h_{ii} equal zero by the Hermitian symmetry of \mathbf{B} . In this formula the parametric form of $\sigma(h)$ is selected and so are the vectors $\{\tau_{\theta}\}$, which may have different directions. The coefficients g_{ij} , h_{ij} , p_{θ} are fitted, for example, by a least squares procedure described in Grzebyk (1993).

The model can be further generalized by introducing several scales of structures and lends itself to a decomposition into orthogonal factors leading to a "bilinear model of coregionalization." The reader is referred to Grzebyk and Wackernagel (1994), and Wackernagel (1995, sec. 27) for the theory. The theory of complex kriging itself is presented in Wackernagel (1995, sec. 26). Basically this approach minimizes the sum of the errors on the real and imaginary components $(Z_{\text{Re}}^* - Z_{\text{Re}})^2 + (Z_{\text{Im}}^* - Z_{\text{Im}})^2$ and spares us the necessity of modeling the cross-covariance between Z_{Re} and Z_{Im} , which cokriging of the components would require. But if that cross-covariance is an even function, a separate kriging of the components would perform better.

Compositional Data

These are vector RFs whose components are proportions; they are all positive and add up to one. We would like the estimates to have the same properties. Pawlowsky et al. (1994) suggest to use the additive logratio transform of Aitchison (1986) and define

$$Y_i(x) = \log \left[\frac{Z_i(x)}{Z_p(x)} \right]$$

where $Z_p(x)$ is the last component of $\mathbf{Z}(x)$. In this formulation all components $Z_i(x)$ must be strictly positive. The first $(p - 1)$ transforms $Y_i(x)$ are estimated by cokriging, or kriging, and the transformation is reversed to yield the estimates

$$Z_i^*(x) = \frac{\exp(Y_i^{**}(x))}{1 + \sum_{j=1}^{p-1} \exp(Y_j^{**}(x))}$$

$$Z_p^*(x) = \frac{1}{1 + \sum_{j=1}^{p-1} \exp(Y_j^{**}(x))}$$

By construction, all estimates lies in the interval $]0, 1[$ and add up to one, as desired. But the statistical properties of the estimators are unclear. In particular, unbiasedness is a problem.

Random Sets Models

Multivariate RFs can be derived from random sets, thereby automatically ensuring the consistency of models when several probabilistic functions are considered (histogram, covariance, bivariate

distributions). Jeulin (1992) defines various classes of multivariate RF models: Boolean, dead leaves, dilution, and alternate sequential RFs. They combine the choice of a family of primary random functions, of Poisson linear varieties in n -dimensional space, and of various combinations of the primary random functions: addition, supremum, infimum, first or last occurrence, and the like. The statistical correlation between the components of a given model may be progressively changed from independence to functional dependence. The resulting models are proposed to describe multivariate images.

5.7. SHORTCUTS

Collocated cokriging provides us with simplified implementations of cokriging. In this section we will consider particular but useful models that allow us to bypass cokriging altogether. The main emphasis is on what is often referred to as the *undersampled* case. There is a variable of prime interest $Z(x)$, sampled sparsely, and a secondary (or *background*) variable $T(x)$ carrying direct or indirect information about $Z(x)$, and sampled densely. We can think of $Z(x)$ as the depth of a geological horizon at the point x and $T(x)$ as the travel time of a seismic wave from the surface to that horizon and back. Geophysical reconnaissance surveys provide, at relatively low cost, a fine-mesh grid of the geometry of the subsurface, but only actual drilling can confirm the validity of the model derived from geophysical data. In boreholes, geological depths can be measured with high vertical resolution and absolute accuracy, while geophysical methods, though less precise, ensure a far superior lateral coverage. By and large it may be considered that seismic measurements describe the *shape* of a geologic object, whereas borehole data provide *depth control*.

We wish to use both Z and T data without going through the effort of statistical inference and modeling required by cokriging.

Another motivation for these models is the introduction of external knowledge in the interpolation of nonstationary phenomena. McLain (1980) poses the problem in a striking manner with an example about mapping the growth rate of hairy gooseberries in Ireland. In an area where there are no data points, a blind run of a computer mapping package produces a high growth rate, which the user knows to be incorrect because gooseberries don't grow on mountains. McLain diagnoses three causes for the computer's lack of success: "the computer didn't know where the mountains are, the computer didn't know that hairy gooseberries don't grow on mountains, and the computer wouldn't know how to use such information in any case." Is there a way to introduce intelligence in the computer mapping process?

5.7.1. Random Error Model

The simplest procedure is to transform the T data into estimates of Z , for example, by linear regression, and to use these estimates as Z data but with lower weights to account for the uncertainty attached to predicting Z from T . An uncertainty variance σ_o^2 is associated with each of the pseudo-data and

the kriging system for filtering random errors is applied (Section 3.7.1). As σ_α^2 increases, the information contributed by Z_α vanishes (Delhomme, 1974). An evaluation of the error variances can be derived from regression models, though based on the incorrect assumption of independence between samples.

This approach has the advantage of simplicity and makes no assumption on the relative sampling patterns of $Z(x)$ and $T(x)$ —if they have no point in common their relationship must be known from other sources. But since T is sampled more densely than Z , there are areas where only values of T are present. Consequently, if moving neighborhoods are used, the estimation is based on secondary data only, without any calibration. This is not desirable and a special neighborhood search algorithm, always including some real observations of Z , may be required.

This approach, however, suffers from being based on the inadequate model

$$Z_T(x) = Z(x) + \varepsilon(x)$$

where $Z_T(x)$ is the estimate of $Z(x)$ from $T(x)$ and $\varepsilon(x)$ is a random error uncorrelated with $Z(\cdot)$. This implies that $Z_T(x)$ has greater spatial variability than $Z(x)$, which is usually false given the integrated nature of geophysical data. The random error model is justified when the secondary variable is really of the same kind as the primary variable but corrupted by noise; otherwise, the model is just too crude.

5.7.2. Guess Field Model

In order to generate a grid of some atmospheric parameter at time t , meteorologists run a numerical weather forecast model based on observations at time $t - 1$ and use the result as an initial *guess field*, which they update locally to match the observations at the current time point t (Cressman, 1959; Rutherford, 1972; Schlatter, 1975; Chauvet et al., 1976). The guess field integrates data from the recent past as well as the laws of physics to produce plausible estimates over areas where sampling is sparse (oceans and deserts). In Figure 5.3, for example, the 12-hour-lead forecast of the 500-mb geopotential surface was used as a guess field; if it were not, the results would be similar over Europe where observations are dense, but nonsensical gradients would be obtained over the Southern Atlantic, because extrapolation is unstable especially in the presence of a drift.

In the meteorological application the guess field is known at every grid node and can be interpolated to any data location with negligible error. So the procedure can be described as simple kriging. From another perspective it also qualifies as a space-time model. We can also relate it to regression. Consider a set of observations $T(x_1), T(x_2), \dots$, of secondary variables—in the meteorological case, observations of various atmospheric parameters at time $t - 1$ —and define the guess field $Z_T(x)$ as the best possible estimator of $Z(x)$

from these observations, namely the conditional expectation

$$Z_T(x) = E(Z(x) \mid T(x_1), T(x_2), \dots)$$

The regression error $\eta(x) = Z(x) - Z_T(x)$ is uncorrelated with all (measurable) functions of the T data, and in particular with $Z_T(y)$ for any point y . Therefore the statistical model is

$$\begin{cases} Z(x) = Z_T(x) + \eta(x) \\ \text{Cov}[Z_T(x), \eta(y)] = 0 \end{cases} \quad \forall x, y \quad (5.40)$$

In this decomposition $Z_T(x)$ summarizes the part of $Z(x)$ that can be “explained” by the values of T , not necessarily at the point x only, and $\eta(x)$ captures the random component of spatial variability. Equation (5.40) is in a sense the “opposite” of the random error model, the error term being on the other side of the equal sign.

Now we take (5.40) as a *definition* of the guess field model: the RF $Z(x)$ is the sum of a component $Z_T(x)$ that accounts for the systematic behavior of $Z(x)$, and in particular carries the drift, and a residual (or correction) $\eta(x)$, uncorrelated with the guess field. Depending on the context $\eta(x)$ is assumed to have a zero mean and a stationary covariance $C(h)$, or a constant mean and a stationary variogram $\gamma(h)$. The zero mean assumption ensures graceful blending with the guess field over sparsely sampled regions, whereas the constant mean assumption allows for local adjustments of the estimates when moving neighborhoods are used. The model has the following covariance structure:

$$\begin{cases} \text{Cov}[Z_T(x), Z_T(y)] = K(y - x) \\ \text{Cov}[Z_T(x), Z(y)] = K(y - x) \\ \text{Cov}[Z(x), Z(y)] = K(y - x) + C(y - x) \end{cases}$$

The interesting case is when the Z_T values are available at all the locations where the Z values are given, allowing the determination of residuals and their variogram. Special attention must be paid to the nugget effect because the additive model (5.40) is not compatible with independent measurement errors on Z and Z_T (whose variances would *add up* in $Z - Z_T$).

Under the above sampling condition the cokriging procedure can be decoupled into separate kriging estimation of the guess field and of the residuals; estimation variances are also additive:

$$\begin{cases} Z^*(x_0) = Z_T^*(x_0) + \eta^*(x_0) \\ E(Z^*(x_0) - Z(x_0))^2 = \sigma_{KT}^2(x_0) + \sigma_{K\eta}^2(x_0) \end{cases}$$

In essence the foregoing approach is analogous to regression followed by interpolation of the residuals. It is almost McLain’s solution for the gooseberry

problem: establish a relationship between gooseberry growth rate and location altitude, use the regression equation to predict the growth rate anywhere without interpolation, and finally correct the results by adding interpolated residuals. Why almost? Because the guess field model allows for the fact that it may not be possible to evaluate the regression equation everywhere, such as if the altitude has itself been sampled.

Examples of application of the guess field model in conjunction with kriging include the study of soil contamination by plutonium using both soil sample analyses and radioactivity measurements (Delfiner and Gilbert, 1978), and mapping water saturation in petroleum reservoirs taking into account the physical relationship between water saturation and elevation above the oil-water contact (Delfiner et al., 1983).

From an implementation point of view the guess field approach has many virtues: complex dependencies may be rendered, only standard kriging software is required, and different kriging neighborhoods may be used to account for the different sampling densities of $Z_T(x)$ and $\eta(x)$ without running into drift consistency problems. A difficulty arises when the guess field is not available at locations where Z is known. For example, meteorological observations do not coincide with grid nodes, and in the case of seismic and well data, geophysical surveys are usually available only in a small fraction of the wells. The practical solution is to interpolate the guess field to the sampling locations of Z and create pseudoresiduals. To achieve a zero variance at data points we must assume that this interpolation is without error.

The strongest assumption of the guess field model is the stationarity of the residuals. When, as in meteorology, the transformation from T to $Z_T(x)$ is known independently the residuals have the status of genuine *data*, whereas when the transformation has to be established by regression, their theoretical status is unclear due to the random and correlated character of both dependent and independent variables. In other words, empirical residuals can be misbehaved and show poor spatial stationarity. The key requirement for the use of the guess field model is therefore the existence of enough data or physical background to establish a stable and meaningful guess field.

5.7.3. External Drift Model

We will present the approach in the context of the application that motivated it in the first place, namely mapping the geometry of a petroleum reservoir from borehole and seismic data. The traditional method is more or less the guess field approach: a depth map is derived from seismic measurements and corrected locally to match the wells. To convert seismic reflection times to depths, geophysicists establish a time-velocity curve which is a best fit on a scatterplot of average velocity v against travel time T . Both quantities are measured between the mapped seismic horizon and an arbitrary reference plane (the datum plane), to correct for anomalous effects induced by significant elevations and/or near surface velocity changes. As a result of higher compaction,

at greater depth velocity tends to increase with time and is often modeled, within zones, by a linear relationship. Average velocity and seismic derived depth are therefore

$$v = a_2 T + a_1$$

$$Z_T = vT + a_0 = a_2 T^2 + a_1 T + a_0$$

The coefficients a_0, a_1, a_2 have to be determined from depth, time, and velocity information; they may vary from one zone to another. Due to the many uncertainties involved in seismic computations, it is unreasonable to expect a perfect match between the seismic depth and the depth measurements in boreholes. There are misties at the wells.

In the geostatistical model a residual fluctuation $U(x)$ is allowed, accounting for details that cannot be captured at the resolution of surface seismic, and also for local errors on datum plane determination. The depth at a point x is finally expressed as

$$Z(x) = a_2 T^2(x) + a_1 T(x) + a_0 + U(x)$$

where $U(x)$ is random and $T(x)$ deterministic. Statistically this is a regression model with correlated residuals. From a geostatistical perspective this equation is similar to a universal kriging decomposition into drift and fluctuation in which the basis drift functions are $(1, T, T^2)$ instead of monomials in the geographic coordinates of x . The drift of $Z(x)$ is defined *externally* through an auxiliary variable $T(x)$ rather than from some smooth version of $Z(x)$ itself. The UK equations (3.16) apply where the unbiasedness conditions become

$$\sum_{\alpha} \lambda_{\alpha} = 1 \quad \sum_{\alpha} \lambda_{\alpha} T(x_{\alpha}) = T(x_0) \quad \sum_{\alpha} \lambda_{\alpha} T^2(x_{\alpha}) = T^2(x_0)$$

Physically the first condition eliminates the influence of a constant shift due to an error in picking the seismic reflector; the other two conditions ensure that our estimator interpolates $T(x)$ and $T^2(x)$ exactly. The striking consequence of this formulation is that $Z(x)$ is estimated directly without an explicit time-to-depth conversion. Furthermore the use of moving neighborhoods allows for lateral variations of these conversion coefficients. Kriging interpolation being exact there are no misties at well locations.

As usual with UK, the variogram to be used is the variogram of the residuals $U(x)$. If there are not enough data points to compute it, its shape can be hypothesized and the parameters fitted by cross-validation.

An application of this model was first published by Delhomme et al. (1981). We will consider here a different example that is extreme but real, where only seven wells are available, five of them on top of a dome-shaped structure and two on the flanks (Delfiner et al., 1983). To make the problem more fun, a fault is present to the NE. Naturally it is not possible to do any statistical

inference with so few data, but we can still use kriging equations as *algorithms* and gauge the results on their merit. Figure 5.14a shows the time contour map of a marker close to the top of the formation of interest; this marker does not coincide with the reservoir top, and there is a small but unknown shift. We do not know which method was used to establish that map, but it does not matter. We accept it as the geophysicist's best guess, and of course it incorporates the fault information. On the basis of the seven wells alone, the best we can do is shown in Figure 5.14b; it is obtained by kriging with a forced quadratic drift, to achieve closure, and a spline covariance model. A fault screen is used: note that no estimation is performed behind the fault due to the absence of wells.

Now the seismic times are introduced as an external drift and kriging is performed without a fault screen. The resulting map is shown in Figure 5.14c. It matches the well data exactly—see, for example, contour line 800 m—and its contours have the same shape as those of the seismic map. In particular, the influence of a secondary dome to the west marked by a seismic line at 960 m (Fig. 5.14a) is reflected in the final depth contours, with some attenuation. Note that now there are contours behind the fault, thanks to the seismic information.

The method also allows the computation of a standard error map (Fig. 5.14d), a conventional one in this case since no real inference was performed. As expected, the uncertainty is smaller near the wells (labeled less than 10 m), and one can see a sharp increase behind the NE fault. That increase takes place even though this time no fault was introduced in the estimation process. It seems that the method is intelligent and “knows” that a fault is there. Granted, we expect this to be the case for the map itself, since the fault information is present in the external drift, but standard errors do not depend on values. The explanation lies in the additivity relationship (3.23),

$$\sigma_{\text{UK}}^2 = \sigma_{\text{SK}}^2 + \sum_{l=0}^2 \sum_{s=0}^2 \text{Cov}(a_l^*, a_s^*) \times \sum_{\alpha} \sum_{\beta} [\lambda_{K\alpha} T^l(x_{\alpha}) - T^l(x_0)] [\lambda_{K\beta} T^s(x_{\beta}) - T^s(x_0)] \quad (5.41)$$

When $T(x_0)$ is very different from the calibration values $T(x_{\alpha})$, the variance of the drift correction is large. In other words, the uncertainty increases as one moves away from the range of validity of the (implicit) velocity function—as a geophysicist would expect.

The external drift approach is easy to implement but requires the knowledge of $T(x)$ at every point involved in the kriging estimation, namely every Z sampling location and every grid node. Generally $T(x)$ is not measured everywhere and must be interpolated, but in this model we neglect the interpolation error on $T(x)$ and treat it as deterministic.

$T(x)$ is not necessarily smooth. For example, Maréchal (1984) represents faulting along a profile with the discontinuous external drift function

$$m(x) = a_0 + a_1x + a_2H_c(x) + a_3(x - c)H_c(x)$$

where $H_c(x)$ is the step function valued 1 if $x > c$ and zero otherwise. Similar models are used in two dimensions and allow interpolation of faulted surfaces without the artificial recourse to fault screens.⁸ Contouring the residuals enables one to visualize the original unfaulted surface.

When using a nonsmooth external drift, however, one must be sure that the details are real and relevant. Speaking of relevance, note that, in principle, nothing bars us from using a seismic map from South America to improve the estimation of a map in Asia!

It is also possible to regard the external drift as random. One way is to treat the coefficients as random; this poses no difficulty and leads to Bayesian kriging as developed in Section 3.4.9. The other way is to regard the secondary variable itself as random, but then the mathematical properties of the procedure become unclear. If we consider two RFs $Z(x)$ and $Y(x)$, the natural thing to do is to condition Z on Y and define the external drift by a linear relation such as

$$E[Z(x) | Y(x)] = aY(x) + b$$

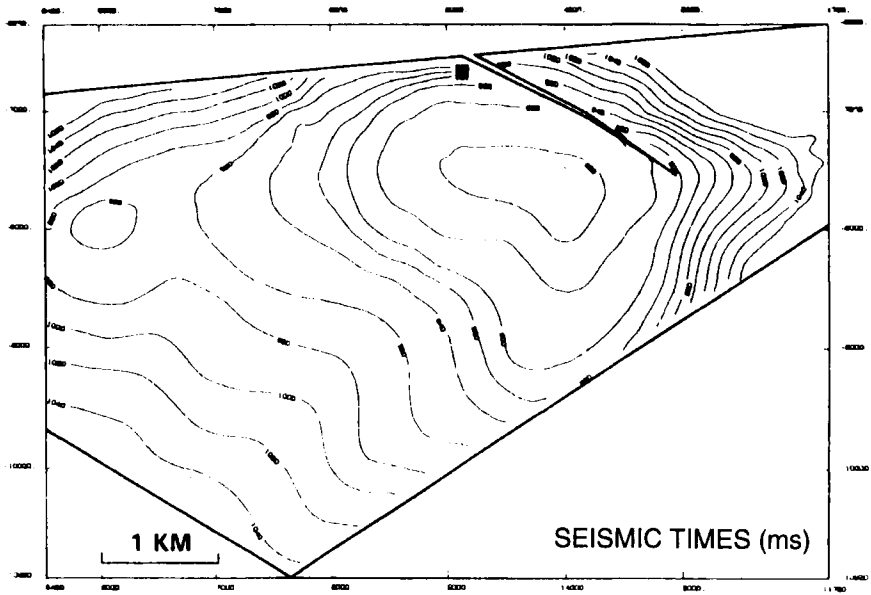
A first difficulty appears: Does this model make sense when Y comprises an erratic component? Do we really want to condition Z on the nugget effect of Y ? Certainly not for the part that is caused by measurement errors. It seems more natural to condition Z on some filtered value of Y having a decent spatial extent. Next, if we have to interpolate Y , which is often the purpose of treating it as random, it becomes questionable to use estimates as exact constraints on the weights with no allowance for uncertainty. It is also questionable to neglect the interpolation error on Y . To illustrate the issue, let us assume that Y is available at all locations x_α where Z is known, but not at x_0 . We write the unbiasedness constraints as $\sum \lambda_\alpha Y_\alpha = \hat{Y}_0$ and $\sum \lambda_\alpha = 1$, where \hat{Y}_0 is an unbiased estimate of $Y(x_0)$. The conditional bias given all relevant values of Y is then

$$E(Z^* - Z_0 | Y_0, Y_\alpha : \alpha = 1, \dots, N) = a(\hat{Y}_0 - Y_0)$$

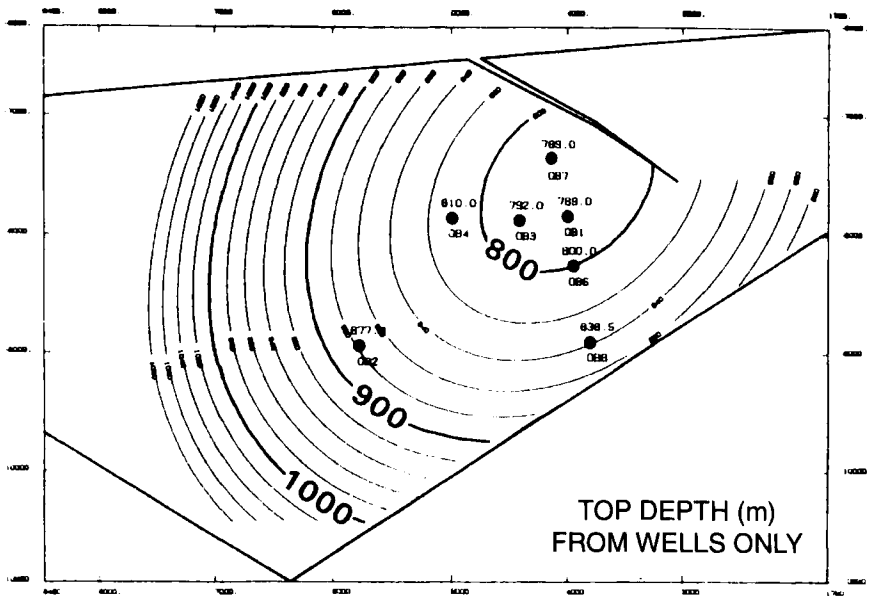
and by (1.5) the unconditional variance is

$$E(Z^* - Z_0)^2 = a^2 E(\hat{Y}_0 - Y_0)^2 + E[\text{Var}(Z^* - Z_0 | Y_0, Y_\alpha : \alpha = 1, \dots, N)]$$

The second term on the right-hand side is of a form similar to (5.41): it is highly nonlinear in the Y_α and intractable but depends only on the variogram (of the residuals) of Z . Now in certain situations the residuals may be very small in comparison with the interpolation error on Y which is the dominant

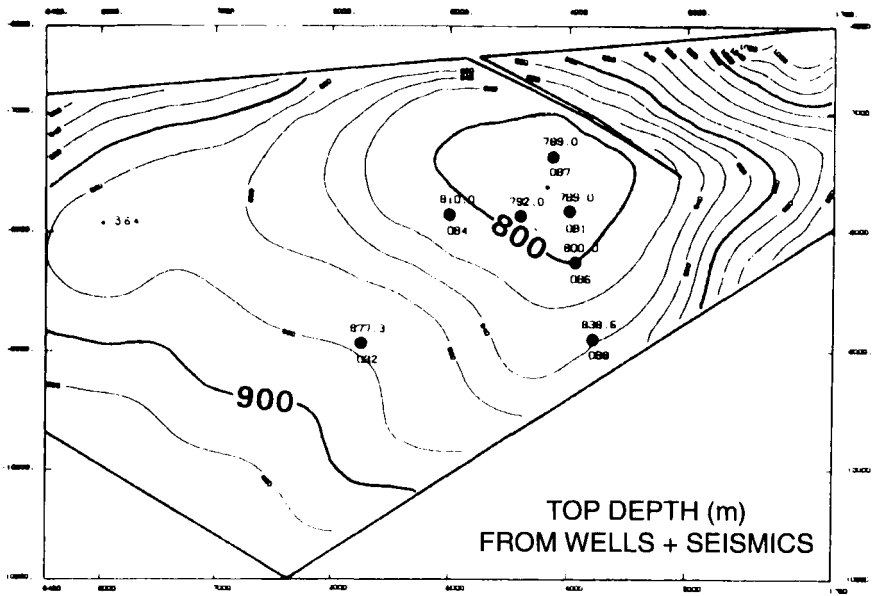


(a)

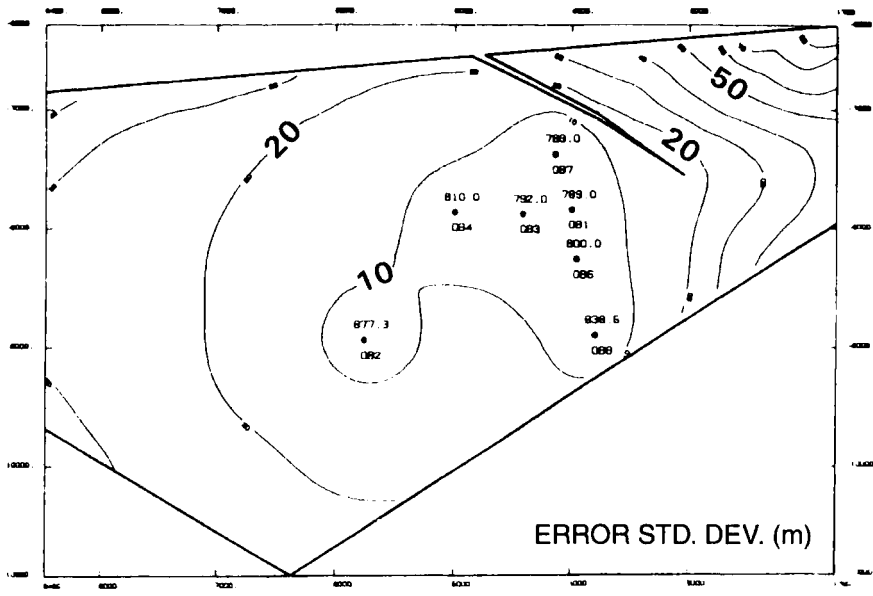


(b)

FIGURE 5.14. Use of seismic data in combination with well data to improve external structure definition: (a) seismic two-way time map; (b) top of reservoir from the wells alone; (c) map combining seismic times and well depths; (d) standard error map. From Delfiner et al. (1983).



(c)



(d)

FIGURE 5.14. (Continued).

source of uncertainty. The important goal is then to optimize the interpolation of Y .

This discussion shows that the external drift shortcut is not adapted to the case where, for some reason, it is desirable to introduce some randomness in the secondary variable. Then cokriging is preferable, possibly in the simplified form of multi-located cokriging (see Section 5.4.3). Both techniques use the same information (Z and Y at all estimating points and Y only at the estimated point) but they are *not* equivalent. Cokriging is linear in Z and Y whereas the external drift method does not use the Y data linearly. By and large, the external drift method introduces a tighter relationship between primary and secondary variables than does multi-located cokriging.

5.7.4. Layer Cake Estimation

The subsurface is often modeled as a stack of layers characterized by their thicknesses $Z_i(x)$, for x varying in the horizontal plane (layer cake model). We are interested in estimating individual layer thicknesses of which there is a large number p and also the total thickness

$$Z(x) = \sum_{i=1}^p Z_i(x)$$

It is assumed that individual and total thicknesses are simultaneously available from vertical boreholes at locations x_α (no missing data). We want to find linear estimators $\hat{Z}_i(x)$ and $\hat{Z}(x)$ that satisfy the obvious consistency relationship

$$\hat{Z}(x) = \sum_{i=1}^p \hat{Z}_i(x) \quad (5.42)$$

Naturally the complete cokriging solution satisfies this requirement, but it is impractical given the large number of layers. Further, modeling errors on cross-covariances would make the improved precision largely spurious. Hence the need for simpler estimators with the consistency property.

The simplest of all is to use the same kriging weights for all layers, but this disregards the spatial structure of individual layers. A more elaborate approach proposed by M. Riguidel (quoted in Matheron, 1979) is to carry out independent kriging estimations of each layer and of the total thickness and combine the estimates linearly so as to satisfy (5.42). This requires only the determination of p direct covariances $C_{ii}(h)$ and that of total thickness $C(h)$. Denoting by Z_i^* and Z^* the kriging estimators of Z_i and Z , the proposed formula is

$$\hat{Z}_i = Z_i^* + a_i(Z^* - Z_i^*) \quad \text{where} \quad Z_i^* = \sum_j Z_j^* \quad (5.43)$$

and where a_i are coefficients satisfying $\sum_{i=1}^p a_i = 1$.

The difference in estimates of total thickness is simply distributed among individual layers. Matheron (1979b) shows that this formula results automatically from the unbiasedness conditions when the following two conditions are imposed:

1. The estimators \hat{Z}_i are linear combinations of Z_i^* and Z^* .

2. $\sum_{i=1}^p \hat{Z}_i = Z^*$, the computed total thickness coincides with its direct kriging estimate.

While the first condition is a natural one, saying basically that we use what we can reasonably get, the second condition is questionable. It presumes that the direct kriging estimate of Z is more reliable than any other computed by combining individual layer thicknesses. Yet, if thicknesses are independent, the Z_i^* are in fact cokriging estimators; their sum is the best estimator of total thickness and does not coincide with Z^* . But, Matheron argues, since the number of RFs involved is large, the covariance $C(h)$ of total thickness, $\sum_i \sum_j C_{ij}(h)$, can be sensibly different from the sum $\sum_i C_{ii}(h)$ of direct covariances even if each rectangular term $C_{ij}(h)$ considered in isolation is not significantly different from zero. So estimating the total thickness by the sum of the individual layer thickness estimates, may lead to an accumulation of errors, except perhaps in the case where it is possible to check that the approximate relationship $C(h) = \sum_i C_{ii}(h)$ does hold empirically.

Having justified formula (5.43), there remains to select the coefficients a_i . A good choice is

$$a_i = \frac{\text{Var}(Z_i^* - \tilde{Z}_i)}{\sum_j \text{Var}(Z_j^* - \tilde{Z}_j)} \quad \text{where} \quad \tilde{Z}_i = \sum_{\alpha} \lambda_{\alpha} Z_{i\alpha} \quad (5.44)$$

\tilde{Z}_i is the estimator of Z_i obtained using the weights λ_{α} of the kriging estimator Z^* of total thickness. The interpretation of (5.44) is as follows: the goal of the correction associated with the weight a_i is to make $\sum Z_i^*$ coincide with the sum $Z^* = \sum \tilde{Z}_i$. If for the index i , Z_i^* is already close to \tilde{Z}_i , there is no need for a correction, or it ought to be very small.

Matheron established this result in the scope of a model of the form

$$\text{Cov}[Z_i(x), Z(x+h)] = \sum_j C_{ij}(h) = \varpi_i C(h)$$

for which the cokriging estimate of total thickness coincides with the kriging estimate.

A particular case is when relative thicknesses $\Pi_i(y)$ are independent of the total thickness $Z(x)$ so that $Z_i(x)$ is the product of two independent random functions

$$Z_i(x) = \Pi_i(x)Z(x)$$

Another particular case is of course the proportional covariance model, or the slightly more general model

$$C_{ij}(h) = \varpi_i \varpi_j C(h) + K_{ij} R(h)$$

with

$$\sum_j \varpi_j = 1 \quad \text{and} \quad \sum_j K_{ij} = 0 \quad \forall i$$

In practice, however, we are not able to judge whether the model fits or not, given the large number of cross-covariances $C_{ij}(h)$. But it does not matter much because the coefficients a_i represent correction terms and do not need to be determined with a high precision.

A simple and powerful variant has been proposed by Haas et al. (1998). Here the total thickness is provided by a seismic grid but with some uncertainty. Wells need not be vertical and each layer is

estimated independently and optimally from its own data. Layer thickness estimates are combined by formula (5.43) where Z^* is replaced by the seismic estimate and the weights a_i are left free. Assuming that kriging errors for individual layers are uncorrelated among themselves and with the seismic prediction error leads to the weights

$$a_i = \frac{\sigma_{Ki}^2}{\sum_j \sigma_{Kj}^2 + \sigma_T^2}$$

When the seismic prediction variance σ_T^2 is zero, the weights a_i add up to one and the total thickness is honored exactly. As the uncertainty on the seismic prediction increases, the constraint becomes looser (note that this uncertainty may vary with location). The correction is weighted by the kriging variance of each layer, ensuring that no correction is applied at a point where the thickness is known. Error variances are found to be

$$E(\hat{Z}_i - Z_i)^2 = \sigma_{Ki}^2 (1 - a_i) \quad E \left[\sum_i \hat{Z}_i - Z \right]^2 = \frac{(\sum_i \sigma_{Ki}^2) \sigma_T^2}{\sum_i \sigma_{Ki}^2 + \sigma_T^2}$$

The combination of well and seismic data reduces error variances over estimates obtained from well or seismic data alone. Errors between two different layers are negatively correlated and with covariances

$$E[\hat{Z}_i - Z_i][\hat{Z}_j - Z_j] = - \frac{\sigma_{Ki}^2 \sigma_{Kj}^2}{\sum_l \sigma_{Kl}^2 + \sigma_T^2}$$

5.8. SPACE-TIME MODELS

In this section we are considering variables $Z(x, t)$ of both space location x and time t . The reason for presenting this topic with multivariate methods is purely formal: if observations are available at different time points t_i , we can define a collection of RFs $Z_i(x) = Z(x, t_i)$ and apply the general cokriging formalism. But it is not advantageous to do so. We will discuss general models found in the geostatistical literature to deal with space-time phenomena and present a few specific models illustrating particular aspects. Physical equations will not be treated here.

The case that received most attention is when the variable of interest, say temperature or pollution, is recorded at regular time intervals at a fixed set of monitoring stations. These records do not necessarily have the same length, and there may be missing values, but they usually are, or can be, synchronized. Thus we can regard the data either as a multiple time series, or as several realizations of a spatial random field for which time is a replication index. The existence of such replications allows the inference of nonstationary models, which lead to better spatial estimates and standard errors than those obtained using a spatially averaged covariance structure. This is typically useful for studying atmospheric pollution insofar as paths of pollutant transportation are affected by varying topography or landscape. New problems can also be

considered where time comes into play, such as: estimate the average change in a region between two different times, or estimate a time trend and quantify the associated uncertainty. This sort of question arises in the discussion on global warming or the effect of human activity on health.

At the opposite of monitoring data is the case of completely asynchronous data. This situation occurs when recording data along the trajectory of a mobile sensor. At any one time only one value is measured, but the location of that measurement changes—else there would be no spatial aspect. In order to map the phenomenon in space, we need to eliminate the effect of time variations, in other words, synchronize the data. A prototypical example is a magnetic survey from a research ship sailing back and forth over a mid-oceanic ridge. The objective is to detect a series of bands parallel to the ridge and indicating the pattern of reversals in the earth's magnetic field. Near the equator the data are perturbed by a strong periodic component caused by diurnal effects created by the rotation of the earth and by solar storms. This diurnal effect must be removed to allow detection of the magnetic alignments.

Mathematically $Z(x, t)$ may be regarded as a random function in the $(n + 1)$ -dimensional space $\mathbb{R}^n \times \mathbb{T}$, that is, n space dimensions plus one time dimension (for a formal treatment, see Christakos, 1992, ch. 5). If one can build a meaningful model of the covariance structure and the drift in this space, then it is kriging as usual. The difficulty, however, is that phenomena are never isotropic in the space-time domain. Time plays a very special role.

5.8.1. General Models in Space-Time Domain

The straightforward extension of kriging is to consider that each point has space-time coordinates (x, t) . For example, there may be two space coordinates, easting and northing measured in meters, and time which is measured, say, in hours. In view of statistical inference, it is assumed that the RF $Z(x, t)$, or its increments of some order, are second-order stationary in space and time. Thus, for example, the spatio-temporal variogram is a function of space and time separations

$$\frac{1}{2} E[Z(x', t') - Z(x, t)]^2 = \gamma(x' - x, t' - t)$$

When $t = t'$, the variogram reflects purely spatial variations; these are assumed statistically the same for all fixed times. When $x = x'$, the variogram captures the time variations only, and they are assumed statistically similar for all locations. Intermediate “directions” represent pairs separated both in space and time, for example, points 5 km and 3 hours apart; these directions are interesting for model discrimination and should not be overlooked. Naturally there is a scaling problem between the time and the space dimensions, but it is not different, in principle, from the scaling problem encountered in 3D between vertical and horizontal variations. In case of “geometric anisotropy,”

a generalized distance such as $[(x' - x)^2/a^2 + (t' - t)^2/b^2]^{1/2}$ would correct for unequal spatial and temporal scale parameters a and b . The real difference is that phenomena are often periodic in time (diurnal period, yearly period, moon cycle, etc.), whereas they are not in space. This is the essential reason why going from \mathbb{R}^2 to $\mathbb{R}^2 \times \mathbb{T}$ is more difficult than from \mathbb{R}^2 to \mathbb{R}^3 .

Great simplifications occur when the spatial and temporal variations can be completely separated. Two models share this nice property. The first one is a “zonal anisotropy” where the time direction plays the same role as the vertical direction in 3D problems (cf. Section 2.5.2)

$$\gamma(h, \tau) = \gamma_S(h) + \gamma_T(\tau) \quad (5.45)$$

Structural analysis can be performed independently in space and time by considering pairs in the same spatial planes or in the same time series, possibly with tolerance limits. One way of checking this model is to verify that in all intermediate directions the sill is the sum of the space and the time sills. The additive model (5.45) can also be written in terms of generalized covariances, in which case, as noted by Rouhani and Hall (1989) who proposed it for groundwater data, the separation extends to the unbiasedness conditions as well—that is, there is no need for mixed space-time monomials. A known problem with model (5.45) is the possibility of singular kriging systems. More important, this model may be too crude. A more useful variant is the sum of an “isotropic” model in the space-time domain and a zonal term corresponding to the periodic time component of the phenomenon.

The second case of simplification occurs with the factorized covariance model

$$C(h, \tau) = C_S(h) \times C_T(\tau) \quad (5.46)$$

Here the sill is the same in all directions. A quite useful property of this model is its associated screening effect (Section 3.6.1): given observations at time t , all other observations at the same space locations but different times are irrelevant for estimation at time t by simple kriging. In particular, *where the present is known, there is no information in the past*. However, this is only true in the case of a known mean.

The above models are for the covariance structure. We have said nothing about the means. In general, we expect anisotropy there too; that is, the spatial mean for fixed t will differ from the temporal mean for fixed x .

Once the covariance structure is determined and a drift model is postulated, kriging proceeds as usual by optimizing a linear estimator of the form

$$Z^* = \sum_{\alpha} \lambda_{\alpha} Z(x_{\alpha}, t_{\alpha})$$

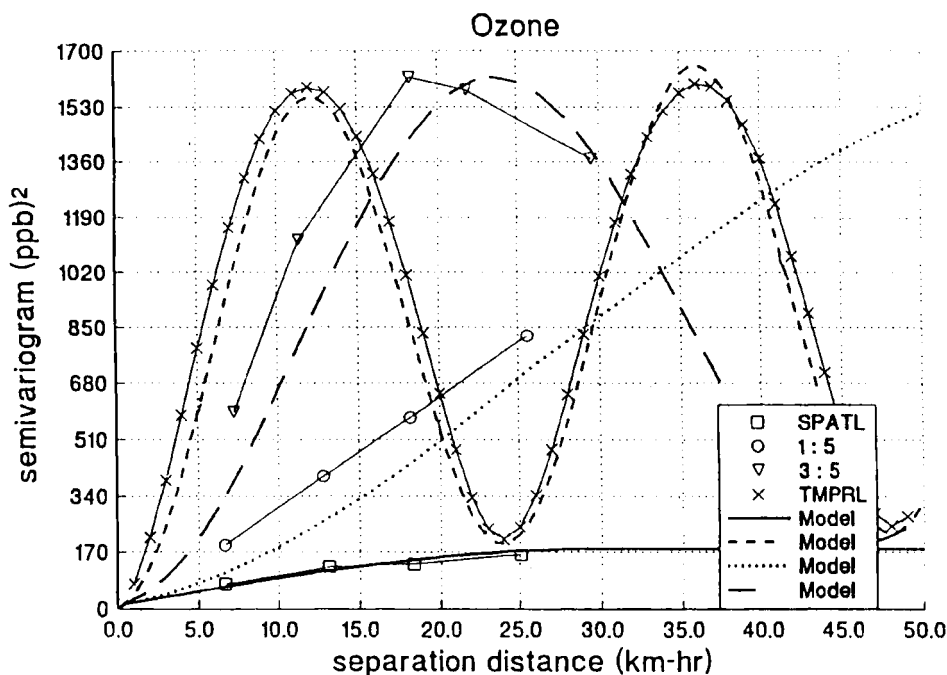


FIGURE 5.15. Experimental variograms and fitted model (5.47) for ozone ($C_0 = 10$, $C_1 = 170$, $C_2 = 1$, $C_3 = 720$, $a = 30$ km-h). Four directions are shown, within the spatial plane (SPATL), in the temporal direction (TMPRL), time shift of 1 hour for every 5 km of spatial shift (1 : 5), and time shift of 3 hours for every 5 km of spatial shift (3 : 5). From Buxton and Pate (1994), with kind permission from Kluwer Academic Publishers.

Sample points are now $\{x_o, t_o\}$, and the kriging neighborhood is defined in both space and time.

Direct applications of kriging in the space-time domain are limited. One example is a study by Buxton and Pate (1994) of concentrations of hazardous pollutants in urban air. The objective is to determine how reliably concentrations measured at one or more times and locations can be used to estimate concentrations at other times and locations, and thus the actual exposures which may occur. Figure 5.15 shows experimental variograms of ozone concentrations along four "directions." A strong periodic component with a period of 24 hours is observed along the time direction (TMPRL): it corresponds to diurnal variations that can be related to daily temperature inversions and other meteorological and chemical processes. In the spatial plane, (SPATL) variations are much milder and reach a sill of about 180 ppb^2 , compared with an oscillation between 180 ppb^2 and 1620 ppb^2 for the temporal variogram. The other two variograms correspond to the intermediate directions 1 hour–5 km and 3 hours–5 km. The authors fitted a sum of a geometric and zonal

anisotropy models of the form

$$\gamma(h, \tau) = C_0 + C_1 \text{Sph}(r, a) + C_2 |\tau| + C_3 \left[1 - \cos \left(\frac{2\pi\tau}{24} \right) \right] \quad (5.47)$$

where $\text{Sph}(r, a)$ is an isotropic spherical model and r is a generalized distance corrected for a geometric anisotropy, in km-hour. (Apparently the units are such that 24 km is equivalent to 24 hours.) Among other things the study concludes, not surprisingly, that a greater precision in the estimation of ozone at the current hour is obtained by adding a single measurement at the current hour at several sites than by adding continuous hourly monitoring at a single site.

On this subject of trading time versus space, let us mention an interesting study by Rodríguez-Iturbe and Mejía (1974). Here the goal is the design of a hydrologic network for the estimation of the long-term areal mean rainfall (an integral over space *and* time). Using a stationary model in time and space and a factorized covariance of type (5.46), they show that the time interval cannot be reduced too much: "no miracles can be expected in short times even from the most dense of all possible networks."⁹

5.8.2. Completely Asynchronous Data

Recording Z along a trajectory $x(t)$ yields a raw time signal of the form $Z(x(t), t)$. In the magnetic survey application of interest, the recordings are the effect of an underlying magnetic field which is independent of time, and a component depending on time only. Thus the following simple additive model is considered

$$Z(x, t) = Y(x) + X(t)$$

where $X(t)$ represents diurnal variations and $Y(x)$ the space component, modeled for generality as an IRF- k . Our purpose is to estimate $Y(x)$. Since in this model the time effect is the same everywhere, the ideal solution would be to monitor $Z(x_0, t)$ at some fixed location x_0 and subtract the signal from $Z(x, t)$ to obtain a spatial field $Y(x) - Y(x_0)$. But we work here under the assumption that such reference station is not available. So we must *estimate* the component $X(t)$, or equivalently $Y(x)$, from the observations of $Z(x, t)$, a problem akin to factorial kriging analysis.

The key element is how we model $X(t)$ and its relationship with $Z(x, t)$. In the approach called *trigonometric kriging*, designed to solve the magnetic survey problem, $X(t)$ is modeled as a drift (Séguret, 1989; Séguret and Huchon, 1990). Specifically, the pseudoperiodic diurnal variations of $X(t)$ are represented by a *finite* trigonometric expansion of the form

$$X(t) = \sum_i A_i \cos(\omega_i t) + \sum_i B_i \sin(\omega_i t) \quad (5.48)$$

where the ω_i are fixed angular frequencies (e.g., $2\pi/24$ for a daily cycle and t in hours) and A_i and B_i unknown (possibly random) coefficients. Note that, by choice, $X(t)$ is centered at zero, the baseline level of $Z(x, t)$ being included in the spatial component $Y(x)$.

The estimation of $Y(x_0)$ can be performed in a single step, pretty much as in universal kriging (i.e., with an *implicit* estimation of $X(t)$), by imposing appropriate linear constraints on the weights;

these filter out the time component completely. Consider the linear estimator

$$Y^*(x_0) = \sum_{\alpha} \lambda_{\alpha} Z(x_{\alpha}, t_{\alpha}) = \sum_{\alpha} \lambda_{\alpha} Y_{\alpha} + \sum_i A_i \sum_{\alpha} \lambda_{\alpha} \cos(\omega_i t_{\alpha}) + \sum_i B_i \sum_{\alpha} \lambda_{\alpha} \sin(\omega_i t_{\alpha})$$

If for all frequencies ω_i of the expansion the weights satisfy the unbiasedness conditions

$$\sum_{\alpha} \lambda_{\alpha} \cos(\omega_i t_{\alpha}) = 0 \quad \sum_{\alpha} \lambda_{\alpha} \sin(\omega_i t_{\alpha}) = 0 \quad (5.49)$$

then the estimation error only depends on $Y(x)$

$$Y^*(x_0) - Y(x_0) = \sum_{\alpha} \lambda_{\alpha} Y_{\alpha} - Y_0$$

and the derivation of the estimator proceeds as in standard UK or IRF- k theory, with the additional constraints (5.49) and their associated Lagrange parameters.

The estimation of $X(t_0)$ can be carried out directly in a similar manner, the conditions (5.49) being now

$$\sum_{\alpha} \lambda_{\alpha} \cos(\omega_i t_{\alpha}) = \cos(\omega_i t_0) \quad \sum_{\alpha} \lambda_{\alpha} \sin(\omega_i t_{\alpha}) = \sin(\omega_i t_0)$$

It is interesting to estimate $X(t)$ even if the primary focus is on $Y(x)$ because it provides objective criteria to verify the quality of the results. At a crossing point x the difference $Z(x, t') - Z(x, t)$ between observations is due solely to the diurnal variation $X(t') - X(t)$, and it therefore constitutes a control point for $X^*(t') - X^*(t)$. In theory, trigonometric kriging interpolates cross-track differences exactly, but in reality, errors on the determination of x and the use of moving neighborhoods cause discrepancies whose magnitude are indicative of the quality of the results. In addition the estimated $X^*(t)$ may be compared with time records from areas adjacent to the survey.

Figure 5.16 illustrates the application of the method to magnetic anomaly data recorded during the Krakatau cruise off the Sunda Strait of Indonesia, one of the most volcanically active region on earth (from Séguret and Huchon, 1990). This is a difficult case because the amplitude of the daily variation is of the order of magnitude of the recorded magnetic anomalies. The matter is complicated further by the interaction between space and time effects induced by the ship's track pattern: sailing along 12-hour-long profiles, the ship revisits the same spatial neighborhood after 24 hours, creating a spurious 24-hour period in the measured signal $Z(x(t), t)$. Thus the map of raw data (Fig. 5.16a) is impossible to interpret; the comparison with the map obtained by trigonometric kriging (Fig. 5.16b) speaks for itself.

This successful application should not hide the difficulties of putting trigonometric kriging into practice. We are walking a tightrope here. Essentially we are attempting to reconstruct a three-dimensional domain, the space-time cube, from a one-dimensional curve. The most significant implementation problem is the selection of the kriging neighborhood. Indeed, given the density of points along profiles, namely one per minute and more than 10,000 for a typical survey, it is not possible to use a global neighborhood. Then one is confronted with the conflicting requirements of sampling in space and in time. A spatially isotropic neighborhood is desirable for the estimation of the spatial component $Y(x)$, but the time component is then sampled in an uncontrolled manner. On the other hand, a regular sampling in time means data along one or a few profiles and a poor choice of neighborhood for spatial estimation. These issues are discussed at length in Séguret's thesis (1991), as well as the multiple causes of numerical instability (collinearity of drift functions, strong sampling anisotropy, crossing points, etc.).

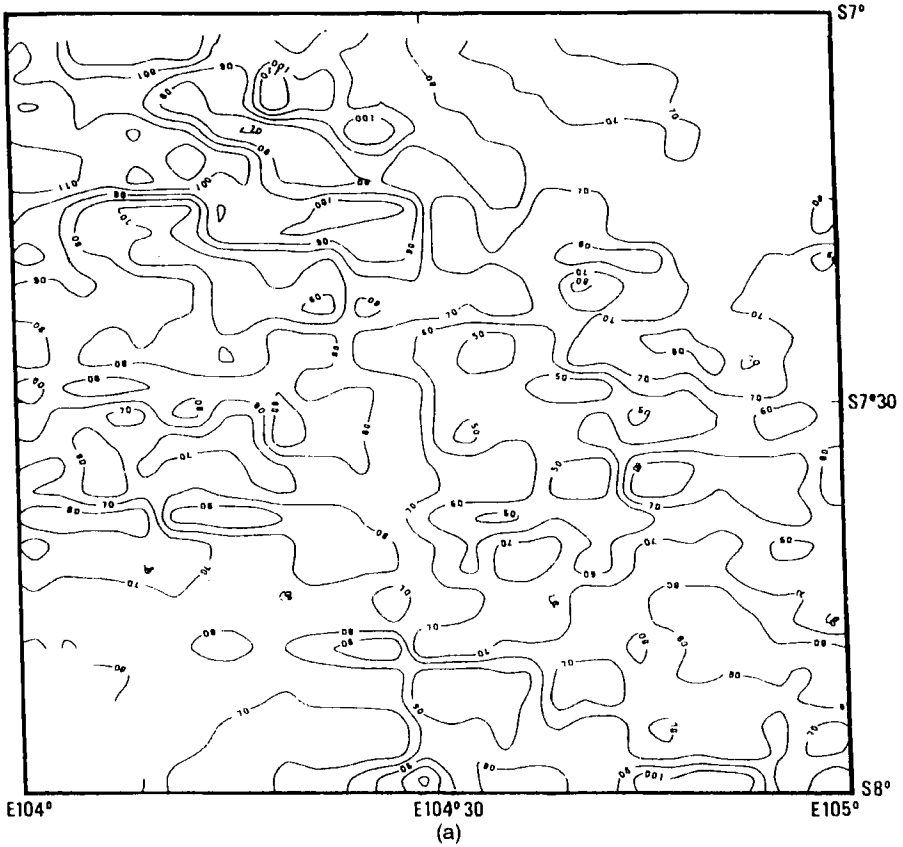


FIGURE 5.16. Contour maps of the magnetic anomaly data recorded during the Krakatau cruise: (a) map of raw data; (b) map of trigonometric kriging estimates, filtering diurnal variations. Reprinted from Séguret and Huchon (1990), © American Geophysical Union.

A positive aspect of using moving neighborhoods is the ability to deal with time effects that are not really periodic. Indeed, since $A \cos(\omega t) + B \sin(\omega t) = C \cos(\omega t + \Phi)$, fitting A and B locally (in time) amounts to considering a model with time-varying amplitude and phase. In addition results appear to be stable in relation to the frequency used; that is, it makes no difference if 23, 24, or 25 hours are used as work frequency.

5.8.3. Estimation of Nonstationary Spatial Covariances

Now consider that the variable of interest is monitored at N sites x_α and p time points t_i so that the data are on a rectangular space-time lattice $\{Z(x_\alpha, t_i), \alpha = 1, \dots, N; i = 1, \dots, p\}$. Typically there are much fewer sites than time points. If we postulate that the process $Z(x, t)$ is temporally stationary, each time section $Z_i(x) = Z(x, t_i)$ may be regarded as a realization of the same parent random function $Z(x)$. Time plays no role except that the multiple realizations of $Z(x)$ may exhibit serial correlation if the lag between observation times is short. In the real world the assumption of temporal stationarity is never a trivial one and requires a careful selection of the data to ensure that they originate from the same underlying physical process, and in particular that the mean

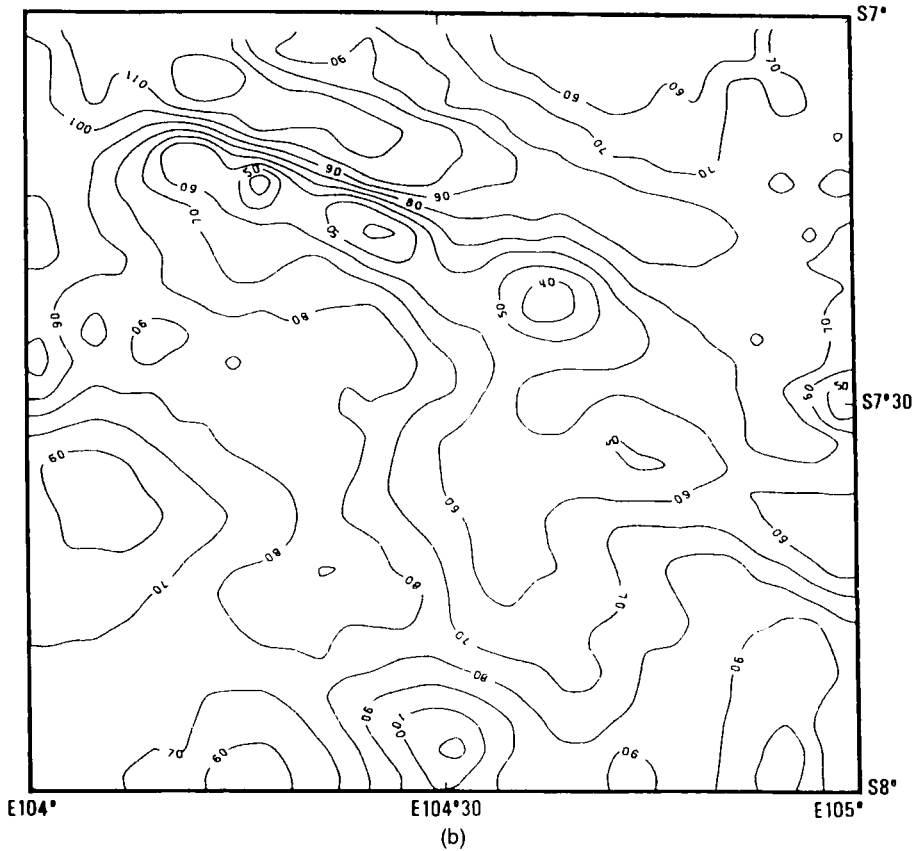


FIGURE 5.16. (Continued).

does not depend on t . For example, if we are interested in January temperatures, we may want to work only from January temperature data rather than a mix of all months.

The model is of the form

$$Z(x, t) = m(x) + Y(x, t)$$

where $Y(x, t)$ is a zero-mean residual process that is stationary over time but nonstationary over space. In this model it is possible to estimate the mean at any monitoring site and the spatial covariance between two sites as time averages, without forcing spatial stationarity assumptions. For example,

$$\hat{m} = \bar{Z} = \frac{1}{p} \sum_{i=1}^p \mathbf{Z}_i \quad \hat{\Sigma} = \frac{1}{p} \sum_{i=1}^p (\mathbf{Z}_i - \bar{Z})(\mathbf{Z}_i - \bar{Z})' \quad (5.50)$$

where as usual $\mathbf{Z}_i = (Z_i(x_1), \dots, Z_i(x_N))'$.

Note that $\hat{\Sigma}$ is numerically positive definite, and is singular if and only if there exists a linear combination such that $\sum w_\alpha [Z_i(x_\alpha) - \bar{Z}(x_\alpha)] = 0$ for all i . Also $\hat{\Sigma}$ is a valid estimate of the

true Σ even in the presence of serial correlation, but of course the estimate is more stable if the realizations are independent. (In this case dividing by $p - 1$ eliminates the slight bias in the covariance estimate due to the estimation of the mean.)

Now suppose that we want to estimate $Z(x_0, t_0)$ at one of the observed time point $t = t_0$, using contemporaneous data only. We can subtract the means from the data and interpolate the residuals by simple kriging. It is possible to plug the covariance estimate (5.50) directly into the left-hand side of the simple kriging equations (3.2), which were established without stationarity assumptions. (This amounts to minimizing the time-averaged squared error). By doing so, we introduce what Switzer calls *location-specific* information. However, we also need the covariances $\sigma_{\alpha 0}$ between the stations and the unobserved location x_0 and the variance σ_{00} .

First, in order to mitigate the dependence of the empirical covariance $\hat{\Sigma}$ between monitored locations on sampling variations, Loader and Switzer (1992) propose the alternative estimate

$$\tilde{\Sigma} = \nu \hat{\Sigma} + (1 - \nu) \mathbf{C} \quad (5.51)$$

where $0 < \nu < 1$ is a constant and \mathbf{C} is computed from a parametric model of the covariance function, typically obtained by forcing a stationary model. This is called a *shrinkage* estimate because it "shrinks" $\hat{\Sigma}$ toward \mathbf{C} . The value of the parameter ν is chosen from the data. A formal optimization formula (based on Wishart distributions) is provided but looks awesome; trial and error may be simpler, if not optimal.

Next the covariance vector $\sigma_0 = (\sigma_{10}, \dots, \sigma_{N0})'$ between the stations and an arbitrary unobserved point x_0 is estimated by interpolation of the columns of $\tilde{\Sigma}$, using as interpolation coefficients the simple kriging weights for estimation at the location x_0 calculated with the parametric covariance \mathbf{C} (Switzer, 1989)

$$\tilde{\sigma}_0 = \tilde{\Sigma} \mathbf{C}^{-1} \mathbf{c}_0 \quad (5.52)$$

where $\mathbf{c}_0 = (c_{10}, \dots, c_{N0})'$. With this choice the SK system (3.2) becomes

$$\tilde{\Sigma} \lambda = \tilde{\Sigma} \mathbf{C}^{-1} \mathbf{c}_0$$

which when $\tilde{\Sigma}$ is nonsingular is equivalent to the SK system $\mathbf{C} \lambda = \mathbf{c}_0$. The kriging estimator obtained using the location-specific $\tilde{\Sigma}$ is no different than that obtained using the parametric model \mathbf{C} . The improvement lies in the kriging variance, which reflects local covariance properties. This is given by

$$\sigma_{SK}^2 = \text{Var} Z_0 - \text{Var} Z^* = \tilde{\sigma}_{00} - \lambda_K' \tilde{\Sigma} \lambda_K = \tilde{\sigma}_{00} - \tilde{\sigma}_0' (\tilde{\Sigma})^{-1} \tilde{\sigma}_0 \quad (5.53)$$

where $\tilde{\sigma}_{00}$ is an estimate of $\text{Var} Z_0$ and λ_K is the vector of simple kriging weights. The procedure suggested for estimating $\tilde{\sigma}_{00}$ is spatial smoothing of the diagonal elements of $\tilde{\Sigma}$. From (5.53) it is seen that the estimate is constrained by

$$\tilde{\sigma}_{00} \geq \tilde{\sigma}_0' (\tilde{\Sigma})^{-1} \tilde{\sigma}_0$$

More generally, the consistency of the covariance model requires that the *augmented* covariance matrix obtained by adding rows and columns for q unobserved locations be positive definite. A necessary and sufficient condition is the positive definiteness of the matrix

$$\tilde{\Sigma}_q - \tilde{\sigma}_q' (\tilde{\Sigma})^{-1} \tilde{\sigma}_q$$

where $\tilde{\Sigma}_q$ is the estimate of the $q \times q$ covariance matrix between the new sites and $\tilde{\sigma}_q = \tilde{\Sigma} \mathbf{C}^{-1} \mathbf{c}_q$ the $N \times q$ covariance matrix between observed and unobserved sites. Loader and Switzer show

that the estimation of $\hat{\Sigma}_g$ can be carried out sequentially by adding one site at a time. The order in which points are added has no influence on the estimates of covariances between observed and unobserved locations but does influence the covariances between unobserved locations. The sequential algorithm only requires fresh estimation of the diagonal terms of the matrix, the off-diagonal terms are obtained automatically from (5.52).

The limitation of the foregoing approach is the estimation of spatial covariances for pairs of unobserved locations. An alternative approach has been proposed by Sampson and Guttorp (1992) based on multidimensional scaling. In a nutshell, the empirical $\sqrt{\gamma_{\alpha\beta}}$ is a distance $d_{\alpha\beta}$ between Z_α and Z_β ; the idea is to represent the data as points in an image space such that the Euclidean interpoint distances match the order relationships among the $d_{\alpha\beta}$. The distances $|y_\beta - y_\alpha|$ in the image space and the corresponding $d_{\alpha\beta}$ are related by a monotone step function g such that $d_{\alpha\beta}^2 = g(|y_\beta - y_\alpha|)$. The function g is analogous to the usual variogram and is both stationary and isotropic. The extension to unobserved locations is achieved by fitting a parametric variogram model to this g function and modeling the mapping $y = f(x)$ between the geographic space and the image space. A discussion of this approach and an alternative rescaling procedure can be found in Monestiez and Switzer (1991).

5.8.4. A Model for Monitoring Data

The assumption of temporal stationarity may be unreasonable or may force us to ignore too much data. A milder form is to assume that this is only true for residuals. The above covariance estimation technique is then applied to residuals. There remains to define the model. A very detailed one has been proposed by Høst et al. (1995), based on the decomposition

$$Z(x, t) = M(x, t) + S(x, t)U(x, t)$$

where $M(x, t)$ is the mean, $S(x, t)$ the standard deviation or scale, and $U(x, t)$ a temporally stationary residual random function with zero mean and unit variance. The authors decide to also regard M and S as random functions and assume mutual independence among M , S , and U .

Next the mean field is modeled as the sum of a "space effect" $F(x)$ and a "time effect" $X(t)$

$$M(x, t) = F(x) + X(t)$$

where $X(t)$ has zero mean over the discrete set of observation times. Similarly the scale field is decomposed as

$$S(x, t) = H(x)\kappa(t)$$

where $H(x)$ is a spatial scale field and $\kappa(t)$ is a temporal modulation such that $\kappa^2(t)$ has mean one. $F(x)$ and $H(x)$ are considered as second-order stationary random functions.

It is convenient to abbreviate $Z(x_\alpha, t_j)$ as $Z_{\alpha j}$ and write the model in the condensed notation

$$Z_{\alpha j} = F_\alpha + X_j + \kappa_j H_\alpha U_{\alpha j}$$

Taking means along rows and columns as is done for two-way tables yields the following estimates of the model components

$$\begin{aligned} \hat{F}_\alpha &= \frac{1}{p} \sum_{j=1}^p Z_{\alpha j} & \hat{X}_j &= \sum_{\alpha} \theta_\alpha (Z_{\alpha j} - \hat{F}_\alpha) & \hat{H}_\alpha^2 &= \frac{1}{p} \sum_{j=1}^p (Z_{\alpha j} - \hat{F}_\alpha - \hat{X}_j)^2 \\ \hat{\kappa}_j^2 &= \frac{\sum_{\alpha} \theta_\alpha (Z_{\alpha j} - \hat{F}_\alpha - \hat{X}_j)^2}{\sum_{\alpha} \theta_\alpha \hat{H}_\alpha^2} & \hat{U}_{\alpha j} &= \frac{Z_{\alpha j} - \hat{F}_\alpha - \hat{X}_j}{\hat{H}_\alpha \hat{\kappa}_j} \end{aligned}$$

where θ_α are spatial weights adding up to one, for example, the optimal weights in the estimation of the constant mean of $F(x)$ (see Section 3.4.5). Once computed, these estimates are considered as exact.

The estimate at an unobserved location x_0 and a monitoring time t_j is obtained by combining the separate kriging estimates

$$Z_{0j}^* = F_0^* + \hat{X}_j + \hat{\kappa}_j H_0^* U_{0j}^*$$

Here F_0^* and H_0^* are ordinary kriging estimates from the values \hat{F}_α and \hat{H}_α at monitoring stations, derived with fitted spatially stationary variogram models. U_{0j}^* is the simple kriging estimate from empirical residuals $\hat{U}_{\alpha j}$, based on a nonstationary covariance matrix obtained as indicated above. There is a difficulty here because the estimation of the residuals forces the linear constraint $\sum_\alpha \theta_\alpha \hat{H}_\alpha \hat{U}_{\alpha j} = 0$ for all j making the empirical covariance matrix singular (cf. (5.50)). To overcome this problem, Høst et al. suggest to replace the time effect estimate \hat{X}_j by the sum of annual and seasonal effects and modify the residuals accordingly. Note that interpolation is only possible at monitored times, since no model is assumed for the temporal modulation $X(t)$ and $\kappa(t)$.

The interpolation variance is given by¹⁰

$$\begin{aligned} \text{Var}(Z_{0j}^* - Z_{0j}) &= \text{Var}(F_0^* - F_0) + \kappa_j^2(\nu^2 + \sigma_H^2)\text{Var}(U_{0j}^* - U_{0j}) \\ &\quad + \kappa_j^2 \text{Var}(H_0^* - H_0) \text{Corr}^2(U_{0j}^*, U_{0j}) \end{aligned}$$

where $\nu = EH(x)$ and $\nu^2 + \sigma_H^2$ is approximated by $\sum_\alpha \theta_\alpha \hat{H}_\alpha^2$. The squared correlation is equal to $\text{Var}(U_{0j}^*)$ because simple kriging is used and the U field has unit variance. This formula shows the contributions of the interpolation variance of the mean field F , the residual field U , and the scale component H .

The authors apply the approach to sulfur dioxide concentrations recorded as monthly averages over six years, and they compare the results with those obtained using standard kriging based only on data from the same time section. They conclude that the estimates are virtually identical but that the space-time model allows a better assessment of standard errors.

There are simpler versions of this very general model, for example, the scale can be made constant (e.g., Stein, 1986b).

5.8.5. Estimating a Time-Trend

In all the preceding models the focus was on spatial estimation, and time was secondary. We used the time observations merely to strengthen spatial estimation. The priorities may be reversed, and the time evolution of the process may on the contrary be the primary focus of interest. A spectacular example is the discussion on global warming in relation with the greenhouse effect. Is there a systematic increase of temperature liable to change the life conditions on earth within the next century? Here we have no interest in the spatial heterogeneity among monitoring stations.

Recent statistical contributions on this subject are those of Handcock and Wallis (1994) and of Switzer and Sølna (1996). The latter authors propose the following temperature model:

$$Z(x, t) = a(x) + b(x)t + U(x, t)$$

where $a(x)$ is a deterministic baseline temperature field, $b(x)$ a spatially varying time-trend rate modeled as a second-order stationary RF, t is a year index, and $U(x, t)$ is a zero-mean residual

random function, second-order stationary in both space and time. The random functions $b(x)$ and $U(x, t)$ are assumed uncorrelated. Two related quantities are of interest:

- Estimation of the time-trend rate averaged over a geographic region A

$$m_A = \int_A b(x) d\mu(x)$$

- Estimation of an annual regional temperature change

$$M_A(t_1, t_2) = \frac{1}{t_2 - t_1} \int_A [Z(x, t_2) - Z(x, t_1)] d\mu(x)$$

The idea is to estimate these quantities from the available records and compare them with the associated standard errors in order to test whether the rate of temperature change is significantly different from zero. It is immediately apparent that these two quantities depend on the year-to-year differences

$$d(x, t) = Z(x, t) - Z(x, t - 1) = b(x) + [U(x, t) - U(x, t - 1)]$$

rather than on temperatures themselves. The derivation of the space-time variogram of $d(x, t)$ and of the kriging estimates and variances of m_A and M_A would take too long, and can be found in the Switzer and Søltna paper. An application to temperature data in a climatically homogeneous region, the steppe of eastern Europe, shows a very strong spatial dependence allowing a precise estimation of regionally averaged temperature changes. On the other hand, the time series of these estimates show little temporal structure. On the basis of 40 years of temperature observations the authors conclude that "a warming trend cannot easily be discerned from a 'nonwarming' scenario," but it might with a record of 90 years. Naturally these conclusions pertain to a relatively small geographic region and cannot be extended to the global scale.

NOTES

1. The authors normalize the gain by the cokriging variance, whereas here we normalize it by the kriging variance so that the relative gain is always less than one.
2. In frequency terms, differentiation amounts to a multiplication of the signal frequency response $X(u)$ by the frequency u , thereby enhancing the high frequencies present in the signal. The smoothing nature of kriging attenuates high frequencies, at least away from the data points.
3. The depth of investigation of the dipmeter tool should also be taken into account and that in turns depends on the resistivity contrasts and also the geometry of the beds and the inclination of the wellbore.
4. As in Chapter 2, positive definite is taken synonymously to *nonnegative definite*, and to exclude the value zero, we refer to *strict* positive definiteness.
5. A characterization of cross-covariances in \mathbb{R}^1 with exponential autocovariances is given by Yaglom (1987, vol. 1, p. 315). In the example we have postulated that the cross-covariance is exponential.
6. They are just scaling factors and can be set to one if the variances are included in the $Y_k(x)$ terms. Of course they have nothing to do with drift coefficients usually denoted by a_j .
7. In case (1) the relationship is $Z^* = m^* + \sum a_k Y_k^{**}$.

8. The conditions of linear independence of the drift functions still entail that empty fault blocks cannot be interpolated.
9. Incidentally this is the first use of factorized time-space covariance that we know of.
10. This formula rectifies an error in formula (5) of the Høst et al. (1995) paper. Note that the parameter ν has nothing to do with the parameter of formula (5.51).

CHAPTER 6

Nonlinear Methods

We do not send estimates to the mill.

Michel David

6.1. INTRODUCTION

Kriging is designed for the estimation of a variable Z at any point, or of its average value over any domain. But for a number of applications the real issue is whether $Z(x)$ exceeds a given threshold. The navigator gives special attention to the shallow parts of a bathymetric map. The environmentalist looks for places where contaminant concentration may exceed a given critical level. The meteorologist focuses on weather situations that may result in cooling below 0°C . The mining engineer attempts to predict which elementary blocks have a mean grade above the economic cutoff grade and the quantity of metal these blocks contain. When varying the threshold it is seen that the basic problem is the determination of the cumulative distribution function $F(z) = \Pr\{Z(x) < z\}$. There are, however, specific spatial aspects that distinguish this problem from standard estimation of a frequency distribution:

1. *Domain.* The distribution is not necessarily estimated at a precise location but over a given domain. This may be the whole domain of interest D , using all available data (global estimation). Or it may be a *local* distribution based only on a subset of neighboring values (local estimation), in which case the distribution is also conditional on these values. The typical application is the determination of the grade distribution of small blocks in a mining panel. If the spacing between samples is large compared with the dimensions of a block, it is unrealistic to attempt predicting individual blocks—precision would be poor, and estimates would change very little from block to block. A more reasonable goal is to predict the *number* of blocks above the threshold. An equivalent way of expressing this, which may be meaningful for a variety of applications, is to say that rather than estimating the probability of exceeding

the threshold at a given location x , we are estimating that probability at a *random* location \underline{x} within the region of interest.

2. *Change of support.* Although measurements are usually made on a point support—or at least small enough to be considered as such—one is often interested in larger supports. Thus in selective mining only the high-grade ore is processed for metal recovery. But a deposit is not mined with a teaspoon. In the absence of geological guidance, selection is performed on “small” blocks, which are in fact huge compared with the core samples. The distribution of these blocks is the quantity of interest and must be determined from the core samples, a challenging task which requires change-of-support models.
3. *Information effect.* The classification of Z as above or below threshold is necessarily made on the basis of incomplete information, which results in a loss of efficiency. In selective mining, a selection of the blocks based on estimates Z^* will yield inferior economic results compared to a selection based on true values. This effect is named the *information effect*. Its evaluation requires modeling of the joint distribution of Z and Z^* .

In a first part of this chapter, we will examine the estimation of the point distribution. We will first review the solutions provided by simple methods whose implementation only requires modeling second-order moments. We will in particular discuss the capabilities and the limits of indicator kriging and its variants. We will then present a nonlinear estimator that is intermediate between simple or ordinary kriging and conditional expectation, *disjunctive kriging*. Since this requires modeling bivariate distributions, we will give a glimpse of the main models available and criteria for selecting one.

In the second part, we will turn to the change-of-support problem. There lies the full worth of these bivariate models, for they provide change-of-support models whose domain of validity far exceeds that of classic models, which will be presented too. The information effect will be discussed only briefly.

6.2. SIMPLE METHODS FOR ESTIMATING A POINT DISTRIBUTION

6.2.1. Global Estimation of the Point Distribution

Regional Distribution and Theoretical Distribution

Global estimation of the point distribution covers in fact two conceptually different problems:

1. Considering a regionalized variable $z(x)$, where x belongs to a domain D , one is interested in the regional histogram of the values of $z(x)$, $x \in D$,

or in the regional c.d.f.

$$F_R(z) = \frac{1}{|D|} \int_D 1_{z(x) < z} dx$$

$F_R(z)$ is an ordinary function which exists independently of any probabilistic interpretation. However, as $z(x)$ is known only at a limited number of points $\{x_\alpha : \alpha = 1, \dots, N\}$, one is led to consider $z(x)$ as a realization of a RF $Z(x)$, and the problem becomes that of the estimation of

$$F_R(z) = \frac{1}{|D|} \int_D 1_{Z(x) < z} dx$$

for any value of z from the data $\{Z(x_\alpha) : \alpha = 1, \dots, N\}$. We will keep the same notation $F_R(z)$ for a c.d.f. which is now random (through Z). The problem is meaningful for general random function models, but in practice, it is only considered under an assumption of global or local stationarity.

2. Given that one turns to the model of a stationary and ergodic RF $Z(x)$, the global problem may be regarded as the estimation of the theoretical marginal distribution $F(z) = E[1_{Z(x) < z}]$.

This distinction generalizes the distinction between the theoretical mean, or mathematical expectation, of an SRF $Z(\cdot)$ and the regional mean of $Z(x)$ in the domain of interest D . In practice, we can at best determine the regional distribution, and most of the time only its approximation by an empirical distribution. We will often have to consider a framework in which the theoretical distribution is known. We will then assume that it can be identified with the empirical distribution, which amounts to assuming that we have a fairly large number of data and that the domain of interest is large with respect to the range (cf. the discussion in Section 2.3.5 on the integral range and the conditions under which the regional mean and the theoretical mean can be identified). Of course we will try to walk away as much as possible from the relatively strict stationarity which is implicit in this hypothesis.

Determination of the Empirical Distribution

Consider a stationary and ergodic RF $Z(x)$, with marginal distribution $F(dz)$. This distribution is generally unknown. We assume ourselves to be dealing with a realization $z(x)$ of the RF $Z(x)$, whose values $z_\alpha = z(x_\alpha)$ are known at N data points $\{x_\alpha : \alpha = 1, \dots, N\}$. It is further assumed that we can construct an empirical distribution \hat{F} that is a good approximation of the true F by assigning an appropriate weight to each sample. The empirical distribution \hat{F}

is of the form

$$\hat{F}(z) = \sum_{\alpha=1}^N \lambda_{\alpha} 1_{z_{\alpha} < z} \quad (6.1)$$

where the λ_{α} are nonnegative weights summing to one. The simplest is to give all weights the same value $1/N$. This solution is acceptable if the sampling pattern is nearly uniform. But data are often in clusters or aligned along profiles, or else the sampling grid is locally tighter, for example, in shallow areas, near pollution peaks, or in well-mineralized zones. The raw data histogram then provides a biased image of the mean characteristics. We have to seek a *declustering* technique to weight each sample according to its location relative to the other samples and to the domain D of interest. A natural way is to use kriging weights.

Several solutions are possible according to the type of kriging considered. Switzer (1977) proposes to minimize a weighted sum of relative variances of the differences $\hat{F}(z) - F(z)$ for different values of z . If one is mainly interested in values near the median, he suggests to simply take the kriging weights of the estimation of the mean value in D of the indicator associated with the median. Another solution is to take the ordinary kriging weights of Z_D . This method has the merit of being consistent with the estimation of Z_D . Indeed the mean that can be deduced from the histogram thus estimated coincides by construction with the direct kriging of Z_D . The histogram obtained is a genuine histogram if all weights are positive, which is generally the case for global estimation. In both cases D should not exceed too much the “envelope” of the sample points. Other methods are discussed by Cressie (1991, sec. 5.8).

6.2.2. Local Estimation of the Point Distribution: Goals

Consider a subdomain V of the domain of interest D . The proportion of V occupied by the points where $Z(x)$ is below a given threshold z is

$$I(z) = \frac{1}{|V|} \int_V 1_{Z(x) < z} dx \quad (6.2)$$

Our goal will be its estimation for one or more values of z , or even for all possible values of z , from observations $Z(x_1), \dots, Z(x_N)$ at data points x_1, \dots, x_N . The optimum estimator is the conditional expectation $E[I(z) | \text{data}]$, where “data” represents the conditioning data $Z(x_1), \dots, Z(x_N)$. The function $I(\cdot)$ is a random c.d.f. (through Z), and determining the conditional expectation of $I(z)$ amounts to determining the conditional c.d.f.

$$F_{\underline{z}}(z | \text{data}) = \Pr\{Z(\underline{x}) < z | \text{data}\} \quad (6.3)$$

where \underline{x} represents a uniform random point of V . Indeed

$$\Pr\{Z(\underline{x}) < z \mid \text{data}\} = E[1_{Z(\underline{x}) < z} \mid \text{data}] = E[I(z) \mid \text{data}]$$

Any other estimator $I^*(z)$ can be considered as an approximation $\hat{F}_{\underline{x}}(z \mid \text{data})$ to the conditional distribution. The distinction between the conditional distribution and its approximation is important, especially when the objective is to define confidence intervals. Indeed, if we know the conditional c.d.f. $F_{\underline{x}}(z \mid \text{data})$, a conditional confidence interval for $Z(\underline{x})$ at level α is

$$Z(\underline{x}) \in \left[F_{\underline{x}}^{-1} \left(\frac{\alpha}{2} \mid \text{data} \right), F_{\underline{x}}^{-1} \left(1 - \frac{\alpha}{2} \mid \text{data} \right) \right]$$

This interval covers a proportion $1 - \alpha$ of actual situations. But if the estimator $I^*(z)$ is only an approximation to the conditional probability, the interval $[I^{*-1}(\alpha/2), I^{*-1}(1 - \alpha/2)]$ does not necessarily cover a proportion $1 - \alpha$ of situations. Switzer and Xiao (1988) present a simple example (Gaussian SRF, four data points at the vertices of a square, $I^*(z)$ obtained by kriging of the indicator $1_{Z(x) < z}$) for which the interval $[I^{*-1}(0.05), I^{*-1}(0.95)]$, which could be expected to cover the true value in 90% of the cases, contains it only in 79% of the cases (results obtained by Monte Carlo simulation): extreme situations are more likely than expected on the basis of $I^*(z)$. Additional results can be found in Lajaunie (1990) who started from this example.

It is pointed out that our stated goal is different from the estimation of the probability that $Z(x)$ remains less than z at every point of V (i.e., that the *maximum* of $Z(x)$ in V remains less than z): this problem is a sensitive and difficult one and it will not be tackled in this book.¹ Our goal also differs from the estimation of the probability that the mean value of $Z(x)$ in V is less than z , a problem that will be examined later. However, formula (6.2) shows that the estimation of $I(z)$ derives directly from the estimation of $1_{Z(x) < z}$ at any point x of V . In terms of the conditional distribution, (6.3) can be written as

$$F_{\underline{x}}(z \mid \text{data}) = \frac{1}{|V|} \int_V F_x(z \mid \text{data}) dx$$

where $F_x(\cdot \mid \text{data})$ is the conditional c.d.f. of $Z(x)$ at a fixed point x . Generally, the difference between the case where V is reduced to a point and that where V is really a volume will be of the same nature as point kriging and block kriging. To simplify the presentation, we will only consider here the case where the studied domain V is reduced to a point x_0 . Rather than (6.2), the function $I(z)$ that we try to estimate is simply

$$I(z) = 1_{Z(x_0) < z}$$

Phrased differently, we will seek (an approximation to) the conditional distribution

$$F_{x_0}(z \mid \text{data}) = \Pr\{Z(x_0) < z \mid \text{data}\}$$

rather than the distribution (6.3). Before reviewing a few classic estimators I^* that constitute approximations to the conditional distribution, let us examine the conditional distribution itself. In its most general form it is poles apart from the simple estimators that we consider here. But it is applied only to Gaussian RFs (up to a transform), namely in a framework in which all multivariate distributions are entirely determined by their second-order moments, in other words, by the structural characteristics of linear geostatistics.

6.2.3. Conditional Expectation

General Case

Consider N sample points $\{x_\alpha : \alpha = 1, \dots, N\}$ and the target point x_0 , and denote by $F(z, z_1, \dots, z_N)$ the $(N + 1)$ -dimensional distribution defined by

$$F(z, z_1, \dots, z_N) = \Pr\{Z(x_0) < z, Z(x_1) < z_1, \dots, Z(x_N) < z_N\}$$

From this distribution we can deduce

- the marginals $F_0(z) = \Pr\{Z(x_0) < z\}$, $F_1(z) = \Pr\{Z(x_1) < z\}, \dots$, $F_N(z) = \Pr\{Z(x_N) < z\}$;
- the conditional distribution $F_0(z \mid z_1, \dots, z_N)$ of $Z(x_0)$ when the $Z(x_\alpha)$ are set to values z_α .

The optimum estimator of $I(z)$ at the point x_0 , that is, the conditional expectation of $1_{Z(x_0) < z}$, is precisely $F_0(z \mid z_1, \dots, z_N)$.

This method is not practical in the general case because the inference of the multivariate distribution is beyond reach (if the interval of variation of Z is discretized in p classes, there are p^{N+1} probabilities to estimate, e.g., 10^{11} values for only ten classes and ten samples; the conditional expectation itself is an element of a functional space of dimension 10^{10} , and we cannot seriously claim to be optimizing in a space of such dimension!). Hence the absolute need to specify a model despite the risks inherent to this.

Gaussian Transform

In practice the Gaussian model is sought for its good properties. While it is rare to encounter a variable with a Gaussian spatial distribution, it is usually possible at least to transform it into one with a Gaussian *marginal*. If $Z(\cdot)$ is an SRF with a continuous marginal distribution $F(dz)$ and if $G(y)$ stands for the c.d.f. of the standard Gaussian, the transformation $Y = G^{-1}(F(Z))$ trans-

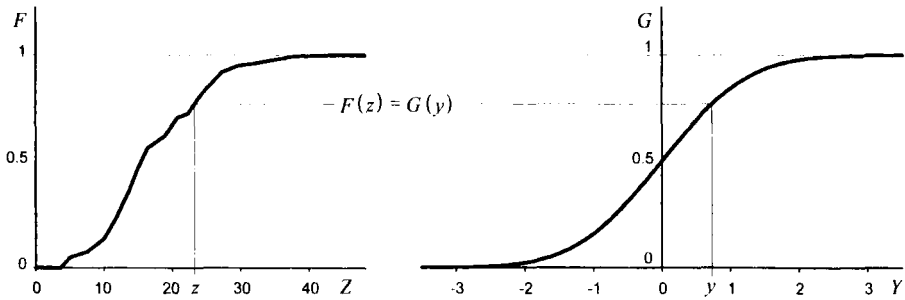


FIGURE 6.1. An example of graphical Gaussian transformation.

forms $Z(\cdot)$ into an SRF $Y(\cdot)$ with a standard Gaussian marginal (recall that $F(Z(x)) = G(Y(x))$ in distribution, being both uniform deviates). This *Gaussian transformation* can be defined “graphically” from the empirical histogram: the cumulative histogram is fitted to a classic c.d.f., or more simply to a piecewise linear function $F(z)$; each point $(z, F(z))$ of the graph is associated with the point $(y, G(y))$ such that $F(z) = G(y)$ (Fig. 6.1). Conversely, Z can be considered as the transform of the Gaussian Y by $Z(x) = \varphi(Y(x))$, and this will turn out to be the useful formulation. The function $\varphi = F^{-1} \circ G$ is called the *transformation* (or *anamorphosis*) *function*. It can also be obtained by its expansion into Hermite polynomials (Section 6.3.4). The Gaussian transformation is of course the function φ^{-1} .

The Multi-Gaussian Model

Suppose that we transform the variables $Z(x_0), Z(x_1), \dots, Z(x_N)$ to $N + 1$ standard normal variates $Y(x_0), Y(x_1), \dots, Y(x_N)$ by means of an appropriate transformation. The crucial hypothesis of the multi-gaussian model is to assume that the $(N + 1)$ -dimensional vector $(Y(x_0), Y(x_1), \dots, Y(x_N))'$ has a multivariate Gaussian distribution, and is therefore characterized by the matrix of covariances σ_{ij} ($i, j = 0, 1, \dots, N$). The conditional distribution of $Y(x_0)$ is then Gaussian. More specifically

$$Y(x_0) \mid \{Y(x_\alpha) : \alpha = 1, \dots, N\} = Y_{SK}^* + \sigma_{SK} U$$

where Y_{SK}^* is the simple kriging estimator of $Y(x_0)$ from the $\{Y(x_\alpha) : \alpha = 1, \dots, N\}$; σ_{SK} is the associated kriging standard deviation; U is an independent standard normal deviate (standardized kriging error).

We then have

$$\Pr\{Z(x_0) < z \mid Z(x_\alpha) : \alpha = 1, \dots, N\} = G\left(\frac{y - Y_{SK}^*}{\sigma_{SK}}\right) \quad \text{with } y = G^{-1}(F(z)) \quad (6.4)$$

All conditioning information is concentrated in the kriging estimate Y_{SK}^* . The multi-gaussian model strongly determines the form of the estimator, and hardly allows the data “to speak” for themselves. The assumption on which it relies is practically unverifiable, but one can at least verify that bivariate distributions are Gaussian (on this topic, see Section 6.3.4) and implement a validation procedure.

If $Z(x)$ is only locally stationary, with a slowly varying mean in the zone of interest, the same is generally true for $Y(x)$. So one can think of substituting ordinary kriging for simple kriging of $Y(x)$. But then the kriging estimate no longer coincides with the conditional expectation. This introduces a bias which can be corrected, approximately, only in special cases such as a lognormal RF.

The multi-gaussian model is usually used under the assumption that the spatial distribution of the RF $Y(x)$ is Gaussian (i.e., *all* finite-dimensional distributions are Gaussian), so that we could simply call it “the Gaussian model.” We will, however, use its usual denomination, which emphasizes that we are working under a stricter assumption than with the bi-gaussian model presented later.

Case of a Lognormal SRF

If $Z(x)$ is an SRF with a lognormal spatial distribution,² let us consider $Y(x) = \log Z(x)$, which is a nonstandard Gaussian SRF. Y_{SK}^* and σ_{SK}^2 now denoting the simple kriging estimator and variance of this Gaussian, relation (6.4) becomes

$$\Pr\{Z(x_0) < z \mid Z(x_\alpha), \alpha = 1, \dots, N\} = G\left(\frac{\log z - Y_{SK}^*}{\sigma_{SK}}\right)$$

In other words the conditional distribution of $Z(x_0)$ remains lognormal. As we have seen in Section 3.4.10, its arithmetic mean is the simple lognormal kriging estimator

$$Z_{SLK}^* = \exp(Y_{SK}^* + \frac{1}{2}\sigma_{SK}^2)$$

and its logarithmic variance is σ_{SK}^2 .

If $Y(x)$ is only locally stationary and if the local mean is unknown, we cannot determine the conditional distribution of $Y(x_0)$ nor the conditional expectation of $Z(x_0)$, but we are still able to calculate σ_{SK}^2 . An approximation to the conditional distribution of $Z(x_0)$ is obtained under the assumption that it is lognormal, with arithmetic mean the *ordinary* lognormal kriging estimator Z_{OLK}^* (3.33) and logarithmic variance the *simple* kriging variance σ_{SK}^2 (Journel, 1980).

Gallichand et al. (1992) present an application to the drainage design of agricultural lands, where the effective hydraulic conductivity is considered as lognormal.

6.2.4. Indicator Methods

Indicator Kriging

We still wish to estimate $I(z) = 1_{Z(x_0) < z}$ from data $\{Z(x_\alpha) : \alpha = 1, \dots, N\}$. For a given threshold z , the indicator $1_{Z(x) < z}$, regarded as a function of x , is a random function, and our objective can be restated as the estimation of this RF at the

point x_0 from the data $Z(x_\alpha)$. Indicator kriging, proposed by Journel (1982), consists in estimating $I(z) = 1_{Z(x_0) < z}$ by kriging the corresponding indicator RF $1_{Z(x) < z}$. We are thus back to a classic problem, one of simple kriging, given two successive simplifications:

- Replacement of the initial data $Z(x_\alpha)$ by indicator data $1_{Z(x_\alpha) < z}$.
- Replacement of conditional expectation by kriging.

While the second step may be regarded as an inevitable approximation to solve the problem within the scope of two-point statistics, the first step constitutes a clear loss of information.

In theory, if $Z(x)$ can be considered as an SRF with known marginal $F(dz)$, since $F(z)$ is by definition the mean of the RF $1_{Z(x) < z}$ one can use simple kriging with an additional term for the mean

$$I^*(z) = \left(1 - \sum_{\alpha=1}^N \lambda_\alpha\right) F(z) + \sum_{\alpha=1}^N \lambda_\alpha 1_{Z(x_\alpha) < z}$$

without introducing an unbiasedness constraint. In practice, however, the marginal is not always known, and one uses ordinary kriging

$$I^*(z) = \sum_{\alpha=1}^N \lambda_\alpha 1_{Z(x_\alpha) < z}$$

under the condition $\sum_{\alpha=1}^N \lambda_\alpha = 1$.

Note that if the objective is the c.d.f. of $Z(x)$ in a volume V rather than at a point x_0 (i.e., if $I(z)$ is defined by an expression such as (6.2)), it suffices to modify the right-hand sides of the kriging systems as indicated in Section 3.5.2. In both cases kriging must of course be done using the covariance of the indicator at the given threshold (which in general changes with z). Indicator kriging has a number of advantages:

- It takes into account the structure of each indicator $1_{Z(x) < z}$.
- It produces an estimation variance.
- It does not require prior modeling of the theoretical distribution F .
- It does not require global stationarity, but only local stationarity.

It has, however, some drawbacks:

- There are as many variograms to model as there are levels z considered, and of course as many kriging systems to solve.
- Since kriging does not guarantee, except in special circumstances, that the weights λ_α are nonnegative, we are exposed to getting estimates $I^*(z)$

that are negative or greater than 1. Even if for each level z the weights λ_n are positive, there is no assurance that the estimates of $I(z)$ at the various levels satisfy the order relations of a c.d.f. ($z < z' \Rightarrow I(z) \leq I(z')$). This problem must be resolved, for example, by using a quadratic correction (Sullivan, 1984): from initial estimates $I_n^* = I^*(z_n)$ at the different levels z_n , define consistent estimators I_n^{**} by

$$\sum_n w_n [I_n^{**} - I_n^*]^2 \quad (\text{minimum})$$

under the condition that monotonicity is achieved. The positive weighting coefficients w_n are selected according to the relative importance of the various levels z_n .

A simplification often made is to use the same variogram for all levels, for example, the variogram of $Z(x)$ or, when the levels z_n remain close to the median z_0 , the variogram of $I_{Z(x) < z_0}$, as suggested by Journel (1983). In reality this is justified only if all variograms are proportional. Matheron (1982b) showed that such property is true for all levels only in the very special case of the mosaic model.

Mosaic Model

In this model $Z(x)$ and $Z(x')$ are either equal with probability $\rho(x' - x)$, and the distribution of their common value is F , or else are independent and have this same distribution F . In other words, the space is partitioned into random compartments, and the values in each compartment are independent random variables with the same distribution F . It is immediately seen, denoting by m and σ^2 the mean and variance of F , that the centered covariances of $Z(x)$ and $I_{Z(x) < z}$ are

$$\text{Cov}[Z(x), Z(x+h)] = \sigma^2 \rho(h)$$

$$\text{Cov}[I_{Z(x) < z}, I_{Z(x+h) < z}] = F(z)(1 - F(z))\rho(h)$$

In this very special model indicator kriging is equivalent to kriging with the covariance of $Z(x)$. Further yet, in this model indicator cross-covariances for two different levels z and z' are also proportional to $\rho(h)$:

$$\text{Cov}[I_{Z(x) < z}, I_{Z(x+h) < z'}] = [F(\min(z, z')) - F(z)F(z')]\rho(h)$$

The set of indicators associated with various threshold levels thus conforms to a proportional covariance model, and we have seen in Section 5.6.4 that kriging is then equivalent to cokriging.

But aside from the mosaic case the shape of an indicator variogram usually changes with the level. A “destructuring” of the indicator variogram is often observed as the threshold departs from the median (e.g., see Fig. 2.21). The indicator variogram tends to the nugget effect model so that at extreme levels local estimation of an indicator by kriging becomes impossible.

Estimation of a Conditional Variable

Once we have an approximation $I^*(z)$ to the conditional c.d.f. of $Z(x_0)$, it is possible to approximate any function $\psi(Z(x_0))$ by

$$\psi^* = \int \psi(z) I^*(dz)$$

Consider, for example, for a fixed threshold z_0 , the variables

$$T_0 = 1_{Z(x_0) \geq z_0} \quad \text{and} \quad Q_0 = Z(x_0) 1_{Z(x_0) \geq z_0}$$

In mining, if z_0 is the cutoff grade, these variables represent the indicator of ore at point x_0 and the corresponding quantity of metal (counted zero when the grade is less than z_0). They can be estimated by

$$T^* = 1 - I^*(z_0) \quad \text{and} \quad Q^* = \int_{z_0}^{\infty} z I^*(dz)$$

In the scope of a mosaic model, or simply if we are content to use the same variogram model (up to a multiplicative factor) for all thresholds z , it can be seen that these estimators are of the form

$$T^* = \sum_{\alpha=1}^N \lambda_{\alpha} 1_{Z(x_{\alpha}) \geq z_0} \quad \text{and} \quad Q^* = \sum_{\alpha=1}^N \lambda_{\alpha} Z(x_{\alpha}) 1_{Z(x_{\alpha}) \geq z_0}$$

where the λ_{α} are kriging weights. We can estimate T_0 and Q_0 by direct kriging of the variables $T(x) = 1_{Z(x) \geq z_0}$ and $Q(x) = Z(x) 1_{Z(x) \geq z_0}$ and the corresponding ore grade (the conditional variable $Z(x_0) | Z(x_0) \geq z_0$) by the ratio $\hat{m} = Q^*/T^*$. Just as in Section 3.5.5, \hat{m} is a weighted average (i.e., with weights adding up to one) of Z data above the threshold z_0 . This would not be the case if different variograms were used for the estimation of T_0 and Q_0 .

Indicator-Rank Cokriging

Indicator kriging takes into account the position of a value relative to the threshold but not its proximity: a value just a bit above the threshold is not distinguished from a very large value. This led Sullivan (1984) to propose the following procedure (which he named *probability kriging*):

- Use as an auxiliary variable the rank (normalized to 1) of each value $Z(x)$ in the data set, specifically the variable $R(x) = F(Z(x))$.

- Estimate the indicator $1_{Z(x)<z}$ by cokriging based on the data $1_{Z(x_\alpha)<z}$ and $R(x_\alpha)$.

The estimator is then of the form

$$I^*(z) = \lambda_0 + \sum_{\alpha=1}^N \lambda_\alpha 1_{Z(x_\alpha)<z} + \sum_{\alpha=1}^N \lambda'_\alpha R(x_\alpha)$$

In the scope of a model with a known mean, here

$$E[1_{Z(x)<z}] = F(z) \quad \text{and} \quad E[R(x)] = \frac{1}{2},$$

simple cokriging is used and

$$\lambda_0 = F(z) \left(1 - \sum_{\alpha=1}^N \lambda_\alpha \right) - \frac{1}{2} \sum_{\alpha=1}^N \lambda'_\alpha$$

In practice, it is better to place unbiasedness conditions on each variable and therefore set

$$\lambda_0 = 0 \quad \sum_{\alpha=1}^N \lambda_\alpha = 1 \quad \sum_{\alpha=1}^N \lambda'_\alpha = 0$$

The solution does not involve $F(z)$, which spares us the task of its determination and the related stationarity problems.

As with indicator kriging, if several levels are considered, the order relations are not necessarily satisfied and are restored by quadratic programming. This method requires the determination of the covariance of $R(x)$, and for each level, the covariance of the indicator and the cross-covariance of the indicator and the rank. As shown by Lajaunie (1990), it is bad idea to consider, for simplification, only the direct and cross-covariances associated with the median of z and to use them for the other levels, since the correlation between $1_{Z(x)<z}$ and $R(x)$ is usually maximal for a value z near the median.

Indicator Cokriging

Considering several levels and the estimation of $I(z_n)$ in this multivariate context, it is logical to seek to estimate $I(z_n)$ by cokriging using not only the indicators $1_{Z(x_\alpha)<z_n}$ but also the indicators at all other levels $1_{Z(x_\alpha)<z_m}$. This is indicator cokriging. The estimator is then of the form

$$I^*(z_n) = \lambda_0 + \sum_m \sum_{\alpha=1}^N \lambda_{m\alpha} 1_{Z(x_\alpha)<z_m}$$

Again one can either consider cokriging with a known mean, that is, find the $\lambda_{m\alpha}$ by simple cokriging and let

$$\lambda_0 = F(z_n) - \sum_m \left[F(z_m) \sum_{\alpha=1}^N \lambda_{m\alpha} \right]$$

or use ordinary cokriging with the unbiasedness constraints

$$\lambda_0 = 0 \quad \sum_{\alpha=1}^N \lambda_{n\alpha} = 1 \quad \sum_{\alpha=1}^N \lambda_{m\alpha} = 0 \quad \text{if } m \neq n$$

From a theoretical standpoint this method is clearly more powerful than indicator kriging, since it incorporates the information that a sample value is slightly above the threshold z_n ($1_{Z(x_n) < z_{n+1}} = 1$) or well above the threshold (e.g., $1_{Z(x_n) < z_n} = 0$ for all levels considered), while in both cases $1_{Z(x_n) < z_n} = 0$. But it has a downside: it requires estimation and modeling of covariances and cross-covariances of all the different levels, not to mention solving very large systems.

One might be tempted to take the same model for all these covariances, up to a scaling factor. But this leads to a proportional covariance model where, all variables being sampled at the same points, cokriging is equivalent to kriging. It is therefore seen that the advantage of indicator cokriging over plain kriging relies on a fine modeling of all direct and cross-covariances. Unfortunately, we do not have theoretical models available for representing a set of indicators in a consistent manner. Journel and Posa (1990) give necessary but not sufficient conditions, and we have seen in Section 2.5.3 that the problem is already not simple when a single indicator is considered. The implementation of indicator cokriging thus remains difficult, which explains why the method has been presented many times since mentioned by Journel (1983), but apparently seldom applied to estimation problems, except in studies of a methodological nature, such as by Liao (1990); Goovaerts (1994, 1997), studying concentrations of chemical elements in the soil, obtains inferior results with indicator cokriging than with plain indicator kriging, which underlines this difficulty. This has motivated the development of indicator-rank cokriging. In fact true cokriging of the whole set of indicators is achieved by disjunctive kriging. (Indicator cokriging seems to be more used to build conditional simulations, as will be seen in Section 7.7.1.)

Elementary Comparisons

Let $Z(x)$ be an SRF whose marginal distribution is continuous and with median zero. We are interested in the indicator RF $1_{Z(x) \geq 0}$. More specifically, we want to evaluate the indicator of the event $Z(x_0) \geq 0$ at the point $x_0 = x_1 + h$ given the value $Z(x_1) = z$. Indicator kriging proposes to approximate this indicator by

$$I^* = \frac{1}{2} + r(1_{z \geq 0} - \frac{1}{2})$$

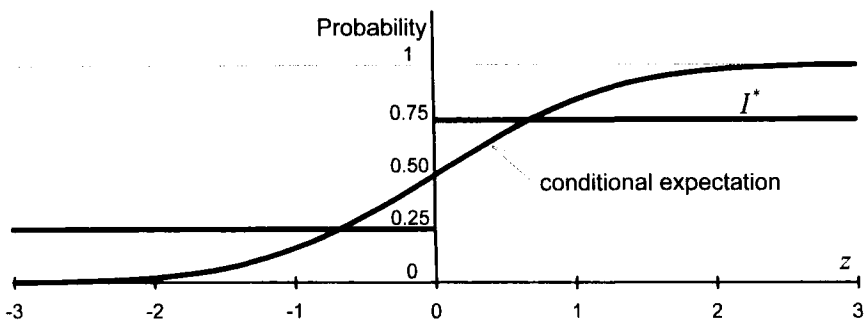


FIGURE 6.2. Estimates of the probability that $Z(x_0) \geq 0$ conditionally on $Z(x_1) = z$ as a function of z when the bivariate distribution of $Z(x_0)$ and $Z(x_1)$ is a standard normal with correlation coefficient 0.707: comparison between conditional probability and indicator kriging.

where r is the covariance of the indicator values at x_0 and x_1 . This problem can cover very different situations with the following extremes:

1. $Z(x)$ is a mosaic SRF. Indicator kriging coincides with conditional expectation and is the appropriate answer.
2. $Z(x)$ is a Gaussian SRF. The correlation coefficient ρ of $Z(x_0)$ and $Z(x_1)$ is related to the indicator covariance r by $\rho = \sin(\pi r/2)$ (see (2.77)). The distribution of $Z(x_0)$ conditionally on $Z(x_1) = z$ is Gaussian, with mean ρz and variance $(1 - \rho^2)\sigma^2$, where σ^2 denotes the variance of Z , and we have

$$E\{1_{Z(x_0) \geq 0} \mid Z(x_1) = z\} = \Pr\{Z(x_0) \geq 0 \mid Z(x_1) = z\} = 1 - G\left(\frac{-\rho z}{\sigma \sqrt{1 - \rho^2}}\right)$$

Figure 6.2 depicts the conditional expectation in the Gaussian case as a function of z and its approximation by indicator kriging for $\sigma = 1$, $\rho = 0.707$, and $r = \frac{1}{2}$, which is seen to be rather crude, and hence the interest of indicator cokriging and more precisely of disjunctive kriging (which would here coincide with conditional expectation).

Note that neither the conditional expectation nor the kriged indicator are 0/1 functions. This is inherent to the mean square error minimization criterion. If we force the classification by using a decision rule based on the sign of the simple kriging estimator $Z^*(x_0)$, the resulting probability of misclassification is given by

$$E\{[1_{Z^*(x_0) \geq 0} - 1_{Z(x_0) \geq 0}]^2\} = \frac{1}{2} - \frac{1}{\pi} \arcsin \sqrt{1 - \frac{\sigma_K^2}{\sigma^2}}$$

This formula holds for a Gaussian RF, with any number and configuration of sample points, and follows from (2.77) and properties of the simple kriging estimator (Section 3.3.4). It is well approximated by $\frac{1}{2} \sigma_K^2 / \sigma^2$.

6.3. LOCAL ESTIMATION OF A POINT DISTRIBUTION BY DISJUNCTIVE KRIGING

We have just seen that it is not easy to estimate the indicator $I(z) = 1_{Z(x_0) < z}$ by cokriging of indicators associated with a number of thresholds, if only be-

cause of the difficulty to represent direct and cross covariances in a consistent way. But as mentioned in Section 2.5.3, the knowledge of direct and cross covariances of the indicators of an SRF for all possible levels is equivalent to the knowledge of the bivariate distributions of the SRF, since

$$\Pr\{Z(x) < z \text{ and } Z(x+h) < z'\} = E[1_{Z(x)<z} 1_{Z(x+h)<z'}] = C_{zz'}(h)$$

where $C_{zz'}(h)$ is the noncentered cross-covariance between the indicators associated with the thresholds z and z' . We are now going to examine the approach based on a direct modeling of bivariate distributions, namely disjunctive kriging (Matheron, 1973b, 1976a).

6.3.1. Disjunctive Kriging

Objective

Let $Z(x)$ be a random function which for the time being we do not assume necessarily stationary. We have N sample points $\{x_\alpha : \alpha = 1, \dots, N\}$, and therefore N data $Z_\alpha = Z(x_\alpha)$. We wish to estimate $1_{Z(x_0) < z}$, or more generally a quantity $Z_0 = f_0(Z(x_0))$. The approach is to carry out the estimation of Z_0 by cokriging of the indicators. If the domain of variation of $Z(x)$ is partitioned into classes B_m , we can associate with $Z(x)$ the indicators $1_{B_m}(Z(x))$. Cokriging Z_0 from indicators amounts to seeking an estimator of the form

$$Z^* = \lambda_0 + \sum_m \sum_\alpha \lambda_{m\alpha} 1_{B_m}(Z_\alpha) \quad (6.5)$$

If the line is partitioned into infinitely small intervals B_m , Z^* is in fact of the form

$$Z^* = \sum_{\alpha=1}^N f_\alpha(Z_\alpha) \quad (6.6)$$

where the f_α are measurable functions. The problem thus amounts to the estimation of Z_0 by a sum of univariate measurable functions of the data. This is an intermediate objective between kriging ($Z^* = \lambda_0 + \sum_{\alpha=1}^N \lambda_\alpha Z_\alpha$, linear functions f_α only) and conditional expectation ($Z^* = f(Z_1, \dots, Z_N)$, a single function of all the data). This objective is called the *disjunctive kriging* of Z_0 (abbreviation DK), since in the expression (6.5) of Z^* the indicators $1_{B_m}(Z_\alpha)$ achieve a disjunctive coding of information, in the sense that at a point x_α one and only one of the indicators is nonzero. Such type of coding is classic in correspondence analysis (Benzécri et al., 1973). As we will see, knowledge of the bivariate distributions of $\{Z_0, Z_1, \dots, Z_N\}$ is sufficient to solve the problem. The determination of these distributions is not out of reach when we have

enough data. Revisiting the example of Section 6.2.3, if the interval of variation is discretized into p classes and if disjunctive kriging is carried out from N data, we must determine $(N+1)p[(N+1)p+1]/2$ probabilities compared with p^{N+1} for conditional expectation and $(N+1)(N+2)/2$ for simple kriging. So, still with $p = 10$ and $N = 10$, we find that disjunctive kriging requires the inference of 6605 values, against 66 for simple kriging and 10^{11} for conditional expectation. With regularly spaced data, stationarity assumptions reduce the required numbers for kriging by a factor of the order of $N/2$, and this is further reduced to a few parameters by modeling the covariance functions.

General Equations of Disjunctive Kriging

We assume that the random variables Z_α , $\alpha = 1, \dots, N$, and Z_0 have second-order moments. Thus they belong to a Hilbert space equipped with the scalar product $\langle X, Y \rangle = E[XY]$. We denote by H_α the subspace generated by random variables of the form $f_\alpha(Z_\alpha)$ with a finite second-order moment, where f_α is a measurable function. Let H be the subspace of random variables of the form $\sum_{\alpha=1}^N f_\alpha(Z_\alpha)$. The disjunctive kriging of Z_0 , in other words, the optimal estimator of the form (6.6), is simply the projection of Z_0 onto this subspace H . By the projection theorem (cf. Section 1.1.2), the DK estimator Z^* is characterized by the orthogonality property

$$\langle Z^* - Z_0, X \rangle = 0 \quad \forall X \in H \quad (6.7)$$

Any random variable of the subspace H being of the form $X = \sum_{\alpha=1}^N f_\alpha(Z_\alpha)$ orthogonality must be satisfied in particular for any random variable belonging to one of the subspaces H_α , thus of the form $X = f_\alpha(Z_\alpha)$, and on the other hand, the linearity of the scalar product ensures that this condition is also sufficient to achieve (6.7).³ Consequently (6.7) is equivalent to

$$\langle Z^*, X \rangle = \langle Z_0, X \rangle \quad \text{for all } X \text{ of the form } f_\alpha(Z_\alpha)$$

By the characteristic formula (1.3) for conditional expectation this condition is equivalent to

$$E[Z^* | Z_\alpha] = E[Z_0 | Z_\alpha] \quad \alpha = 1, \dots, N$$

Therefore the functions f_α corresponding to the DK $Z^* = \sum_{\alpha=1}^N f_\alpha(Z_\alpha)$ are characterized by

$$\sum_{\beta=1}^N E[f_\beta(Z_\beta) | Z_\alpha] = E[Z_0 | Z_\alpha] \quad \alpha = 1, \dots, N \quad (6.8)$$

This system involves the conditional distributions of Z_0 and of Z_β given Z_α and only requires knowledge of bivariate distributions.

It follows from (6.7) that the DK estimation variance is

$$\sigma_{DK}^2 = E[Z^* - Z_0]^2 = \text{Var}[Z_0] - \text{Var}[Z^*] = \text{Var}[Z_0] - \text{Cov}[Z_0, Z^*]$$

These results are interpreted like those of simple kriging (Section 3.3.2).

Disjunctive Kriging Equations for a Transformed Isofactorial Model

There are SRFs for which the solution of the system (6.8) is greatly simplified, namely those whose bivariate distributions are isofactorial, the best known example being SRFs with Gaussian bivariate distributions. We will see shortly that this circumstance is far from being general. It can be extended, however, to the case where $Z(\cdot)$ is the *transform* of an SRF $Y(\cdot)$ with isofactorial bivariate distributions. We will present later the general properties of isofactorial models and the main models used in geostatistics.

The main property of an isofactorial model is that there exists a set of functions $\{\chi_n(y) : n = 0, 1, \dots\}$ (the *factors*) such that any transform $\psi(Y(x))$ can be decomposed into a sum $\sum_{n=0}^{\infty} \psi_n \chi_n(Y(x))$ whose coefficients ψ_n are given by

$$\psi_n = \int \psi(y) \chi_n(y) G(dy) \quad (6.9)$$

and factors $\chi_n(Y(x))$ are uncorrelated SRFs, namely

$$E[\chi_m(Y(x)) \chi_n(Y(x+h))] = \delta_{mn} T_n(h) \quad (6.10)$$

where $T_n(h)$ is the covariance of the SRF $\chi_n(Y(x))$. From the characteristic property (1.3) it follows that (6.10) is equivalent to

$$E[\chi_n(Y(x+h)) | Y(x)] = T_n(h) \chi_n(Y(x)) \quad (6.11)$$

One of the factors, ascribed the index 0, is the constant function $\chi_0(y) \equiv 1$. The corresponding (noncentered) covariance is $T_0(h) \equiv 1$.

Let us now reformulate our problem in the DK framework:

1. The SRF $Z(x)$ of interest is of the form $Z(x) = \varphi(Y(x))$ where $Y(x)$ is an SRF with isofactorial bivariate distributions.
2. We wish to estimate a function of $Z(x_0)$ that we consider of the form $Z(x_0) = \psi(Y(x_0))$. Typically the objective is the estimation of $1_{Z(x_0) < z}$. It is equivalent to the estimation of $1_{Y(x_0) < y}$ where $y = \varphi^{-1}(z)$, since applying the threshold z on Z is the same as thresholding Y at $y = \varphi^{-1}(z)$.
3. We have data $\{Z(x_\alpha) : \alpha = 1, \dots, N\}$, or equivalently, data $\{Y(x_\alpha) : \alpha = 1, \dots, N\}$.

To simplify notations, let $Y_0 = Y(x_0)$ and $Y_\alpha = Y(x_\alpha)$, and let us rewrite our objective as being the estimation of $Z_0 = \psi(Y_0)$ by an estimator of the form

$$Z^* = \sum_{\alpha=1}^N f_\alpha(Y_\alpha)$$

Z_0 and the $f_\alpha(Y_\alpha)$ can be decomposed on the basis of the factors χ_n as

$$\begin{aligned}\psi(Y_0) &= \sum_{n=0}^{\infty} \psi_n \chi_n(Y_0) \\ f_\alpha(Y_\alpha) &= \sum_{n=0}^{\infty} f_{\alpha n} \chi_n(Y_\alpha)\end{aligned}\tag{6.12}$$

The coefficients ψ_n , defined by (6.9), are known, since the function ψ is given. The coefficients $f_{\alpha n}$ are to be determined. The disjunctive kriging system (6.8) is written here as

$$\sum_{\beta=1}^N E[f_\beta(Y_\beta) | Y_\alpha] = E[\psi(Y_0) | Y_\alpha] \quad \alpha = 1, \dots, N$$

As a consequence of (6.11) and (6.12), and by interchanging the summation order,

$$\sum_{n=0}^{\infty} \sum_{\beta=1}^N f_{\beta n} T_n(x_\beta - x_\alpha) \chi_n(Y_\alpha) = \sum_{n=0}^{\infty} \psi_n T_n(x_0 - x_\alpha) \chi_n(Y_\alpha) \quad \alpha = 1, \dots, N$$

Since the $\chi_n(Y(\cdot))$ are uncorrelated SRFs, the system can be broken up into a distinct system for each value of n : the coefficients $\{f_{\alpha n} : \alpha = 1, \dots, N\}$ are solutions of the system

$$\sum_{\beta=1}^N f_{\beta n} T_n(x_\beta - x_\alpha) = \psi_n T_n(x_0 - x_\alpha) \quad \alpha = 1, \dots, N \tag{6.13}$$

$T_n(h)$ being the covariance of the SRF $\chi_n(Y(x))$, this is the simple kriging system of $\psi_n \chi_n(Y_0)$ from the $\chi_n(Y_\alpha)$. For $n = 0$ this system is degenerate, and it suffices to take any $f_{\alpha 0}$ such that $\sum_{\alpha=1}^N f_{\alpha 0} = \psi_0$.

The advantage of the isofactorial model is that cokriging of factors reduces to kriging of each factor separately. We could say in an incorrect but illustrative manner that disjunctive kriging of $Z_0 = \psi(Y_0)$ is equivalent to a combination of simple kriging of Y_0 from the Y_α , of Y_0^2 from the Y_α^2 , and so on (the image

is incorrect because the factors are usually not monomials). Note that if the objective is the mean value of $\psi(Y(x))$ in a volume V and not at a point x_0 (e.g., if for example it is the c.d.f. $I(z)$ defined in (6.2)), it suffices in the systems (6.13) to replace the point covariance $T_n(x_0 - x_\alpha)$ on the right-hand side by the mean value of $T_n(x - x_\alpha)$ over V (cf. Section 3.5.2).

At this point let us recapitulate the DK procedure in the typical case of the estimation of $I(z_c) = 1_{Z(x_0) < z_c}$ where for clarity the index c is used to denote a fixed cutoff or threshold:

1. Transform the data $Z(x_\alpha)$ into isofactorial $Y(x_\alpha)$ defined by $Z(x_\alpha) = \varphi(Y(x_\alpha))$.
2. Translate the objective $I(z_c)$ into a function of $Y(x_0)$: $\psi(Y(x_0)) = 1_{Y(x_0) < y_c}$ with $y_c = \varphi^{-1}(z_c)$.
3. Compute the coefficients ψ_n of the expansion of the function ψ using (6.9); in this case this simplifies to $\psi_n = \int_{-\infty}^{y_c} \chi_n(y) G(dy)$.
4. Solve the systems (6.13) to obtain the weights $f_{n\alpha}$.
5. Compute the estimate as $I^*(z_c) = \sum_{\alpha=1}^N \sum_{n=0}^{\infty} f_{n\alpha} \chi_n(Y(x_\alpha))$.

Expression of Disjunctive Kriging from Estimates of the Factors

If we write (6.13) with $\psi_n = 1$ and weights $\lambda_{n\alpha}$ (we deliberately reverse the order of the indexes) we obtain the system

$$\sum_{j=1}^N \lambda_{nj} T_n(x_j - x_\alpha) = T_n(x_0 - x_\alpha) \quad \alpha = 1, \dots, N \quad (6.14)$$

whose solution defines the estimate

$$\chi_n^* = \sum_{\alpha=1}^N \lambda_{n\alpha} \chi_n(Y_\alpha) \quad (6.15)$$

of the factor $\chi_n(Y_0)$. By linearity, we can then form the DK estimate of *any* function $\psi(Y_0)$ using the formula

$$\psi^* = \psi_0 + \sum_{n=1}^{\infty} \psi_n \chi_n^* \quad (6.16)$$

Pseudo Conditional Distribution

Starting from (6.16) and replacing the coefficients ψ_n by their expression in (6.9), we get

$$\psi^* = \psi_0 + \int \psi(y) \left(\sum_{n=1}^{\infty} \chi_n(y) \chi_n^* \right) G(dy)$$

which amounts to the expected value of $\psi(Y_0)$ with respect to the pseudo probability distribution

$$G^*(dy) = \left(1 + \sum_{n=1}^{\infty} \chi_n(y) \chi_n^* \right) G(dy)$$

This distribution is an approximation to the conditional distribution of Y_0 given the Y_α . However, there is no guarantee that $G^*(dy)$ is positive for any value of y .

Disjunctive Kriging Variance

The estimation variance of $\chi_n(Y_0)$ is the simple kriging variance

$$\sigma_n^2 = \text{Var}[\chi_n^* - \chi_n(Y_0)] = 1 - \sum_{\alpha} \lambda_{n\alpha} T_n(x_0 - x_{\alpha}) \quad (6.17)$$

It is of course zero for $n = 0$, since the factor $\chi_0(Y_0) \equiv 1$ is estimated without error. The DK variance of $Z_0 = \psi(Y_0)$ is the sum of the kriging variances of the factors $\psi_n \chi_n(Y_0)$:

$$\sigma_{\text{DK}}^2 = \text{Var}[Z^* - Z_0] = \sum_{n=1}^{\infty} \psi_n^2 \sigma_n^2 \quad (6.18)$$

6.3.2. Definition and General Properties of an Isofactorial Model

The isofactorial representation of the bivariate normal distribution has been known for a long time (e.g., see Cramér, 1945, p. 290). More generally, isofactorial bivariate models have first been introduced in quantum mechanics; they appeared in the field of stochastic processes with the study of Markov chains (e.g., Karlin and McGregor, 1957, 1960); they are at the basis of correspondence analysis (Benzécri et al., 1973). Their introduction in geostatistics and the systematic study of a large number of models are due to Matheron (a presentation of most models, along with references to the original technical reports can be found in Armstrong and Matheron, 1986a, b). Let us examine in more detail the properties of the isofactorial models that will be useful to us. Our presentation will concern the bivariate distributions of the $(Y(x), Y(x+h))$ pairs from an SRF (of order 2).

Hilbertian Basis

Let $Y(x)$ denote an SRF with marginal distribution G (not necessarily Gaussian). By definition, the system of functions $\{\chi_n(y) : n = 0, 1, \dots\}$ is a Hilbertian basis for the space $L^2(G)$ if

1. The functions $\chi_n(y)$ form a complete countable system for the marginal distribution G : any measurable function ψ such that $\int \psi(y)^2 G(dy) < \infty$

can be represented by the series

$$\psi(y) = \sum_{n=0}^{\infty} \psi_n \chi_n(y) \quad (6.19)$$

with

$$\psi_n = \int \psi(y) \chi_n(y) G(dy) \quad (6.20)$$

2. The functions $\chi_n(y)$ are orthonormal for the distribution G , namely

$$\int \chi_m(y) \chi_n(y) G(dy) = \delta_{mn} \quad (6.21)$$

or, in probabilistic terms,

$$E[\chi_m(Y(x)) \chi_n(Y(x))] = \delta_{mn}$$

The functions χ_n are called the factors. For most models in use one of the factors, ascribed the index 0, is a constant function. Since its norm is 1, its value is either -1 or $+1$. The value $+1$ is usually selected:

$$\chi_0(y) \equiv 1$$

We will restrict ourselves to this case, as all the isofactorial models we will present satisfy this condition.⁴ Letting $m = 0$ in (6.21), it is seen that all the other factors have zero expectation:

$$\int \chi_n(y) G(dy) = 0 \quad \text{or} \quad E[\chi_n(Y(x))] = 0 \quad (n > 0)$$

Note that relation (6.21) simply expresses that *at the same point* x , $\chi_m(Y(x))$ and $\chi_n(Y(x))$ are uncorrelated ($m \neq n$).

Definition of an Isofactorial Model

By definition, the bivariate distributions of an SRF $Y(x)$ with marginal distribution $G(dy)$ constitute an isofactorial model if the following property is satisfied:

- For any two points x and $x + h$, the bivariate distribution of the pair $(Y(x), Y(x + h))$ can be factorized in the form

$$G_h(dy, dy') = \sum_{n=0}^{\infty} T_n(h) \chi_n(y) \chi_n(y') G(dy) G(dy') \quad (6.22)$$

for some Hilbertian basis $\{\chi_n : n = 0, 1, \dots\}$ of the space $L^2(G)$.

The factors χ_n are the same for all h , hence the denomination “isofactorial model.”

Properties of an Isofactorial Model

The following properties can be verified:

- Formula (6.22) expresses that the conditional distribution of $Y(x+h)$ given $Y(x) = y$ is $\sum_{n=0}^{\infty} T_n(h) \chi_n(y) \chi_n(y') G(dy')$. As a consequence we find again (6.11)

$$E[\chi_n(Y(x+h)) | Y(x)] = T_n(h) \chi_n(Y(x))$$

which can be taken as an alternative definition of an isofactorial model. It is a little more general than definition (6.22), for it allows one, for example, to incorporate the case where $Y(x)$ and $Y(x+h)$ are equal and have a correlation of +1.

- From (6.22) we also have

$$\iint \chi_m(y) \chi_n(y') G_h(dy, dy') = \delta_{mn} T_n(h)$$

or, in probabilistic terms,

$$E[\chi_m(Y(x)) \chi_n(Y(x+h))] = \delta_{mn} T_n(h)$$

which is (6.10). In other words, when $m \neq n$, not only are $\chi_m(Y(x))$ and $\chi_n(Y(x))$ orthogonal, as (6.21) expresses it, but also $\chi_m(Y(x))$ and $\chi_n(Y(x+h))$ are orthogonal for all h . The *random functions* $\chi_n(Y(x))$ are therefore stationary (at order 2) and uncorrelated, and they have as respective covariances the $T_n(h)$ functions. These covariances in fact are correlograms, since the factors are normalized. In particular, $T_0(h) \equiv 1$, since $\chi_0(y) \equiv 1$ (the other factors have zero expectation, so T_0 is the only noncentered covariance).

- Given that $\chi_0(y) \equiv 1$, that the other factors have zero expectation, and that all these factors constitute an orthonormal system, it is easy to show that the mean and variance of the SRF $\psi(Y(\cdot))$ can be related to the ψ_n defined by (6.20) by

$$E[\psi(Y(x))] = \psi_0 \quad \text{Var}[\psi(Y(x))] = \sum_{n=1}^{\infty} \psi_n^2$$

and that its covariance takes the form

$$\text{Cov}(\psi(Y(x)), \psi(Y(x+h))) = \sum_{n=1}^{\infty} \psi_n^2 T_n(h) \quad (6.23)$$

Likewise, if φ is another measurable function whose expansion has coefficients φ_n , the cross-covariance of $\varphi(Y(x))$ and $\psi(Y(x+h))$ can be written in the form

$$\text{Cov}(\varphi(Y(x)), \psi(Y(x+h))) = \sum_{n=1}^{\infty} \varphi_n \psi_n T_n(h) \quad (6.24)$$

Special Case of Polynomial Factors

Some models have polynomial factors and share the following additional properties:

- χ_n is a polynomial of degree n .
- Denoting the mean and standard deviation of $Y(x)$ by m_Y and σ_Y , the factor χ_1 is (with an arbitrary choice of sign)

$$\chi_1(y) = \frac{y - m_Y}{\sigma_Y}$$

- As a consequence $T_1(h)$ coincides with the correlogram $\rho(h)$ of $Y(x)$.
- Since the above expression of χ_1 can be inverted in $y = m_Y + \sigma_Y \chi_1(y)$, the conditional expectation of $Y(x+h)$ is a linear regression:

$$E[Y(x+h) | Y(x)] = m_Y + \rho(h)[Y(x) - m_Y]$$

- The coefficients of the polynomial χ_n are related to the moments of order k of the distribution G , $k \leq 2n$.
- The expansion (6.19) truncated to any order n_{\max} is the best approximation of $\psi(y)$ by a polynomial of degree n_{\max} , in the least squares sense measured by $G(dy)$, that is, such that $\int [\psi(y) - \sum_{n=0}^{n_{\max}} \psi_n \chi_n(y)]^2 G(dy)$ is a minimum. However, if $\psi(y)$ is bounded, the expansion limited at the order n_{\max} deviates indefinitely from $\psi(y)$ for $|y| \rightarrow \infty$: the approximation of $\psi(y)$ by a finite degree polynomial can only be valid over a bounded domain of y .

For more information on families of orthogonal polynomials the reader is referred to Szegő (1939), Hochstrasser (1972), Beckman (1973), and Askey (1975).

Additional Remarks

It seems easy to construct isofactorial models, since it is easy to construct a system of orthonormal polynomials χ_n over the distribution G provided

that this distribution has moments of all orders. This apparent simplicity is misleading:

1. If functions χ_n constitute an orthonormal system, nothing proves, even for fixed h , that there exist values T_n such that a bivariate distribution can be expanded as in (6.22).
2. Conversely, there can exist a system of nonpolynomial functions χ_n such that the bivariate distribution has the expansion (6.22). A simple example is that of a truncated bivariate Gaussian distribution: starting from two independent standard normals Y and Y' , we keep only the part of the bivariate distribution where Y and Y' have the same sign. The marginals remain standard normal and the bivariate p.d.f. is

$$g_{\text{trunc}}(y, y') = (1 + \text{sign}(y)\text{sign}(y'))g(y)g(y')$$

where g is the univariate p.d.f. of the standard normal. The two factors of the bivariate density g_{trunc} are $\chi_0(y) \equiv 1$ and $\chi_1(y) = \text{sign}(y)$, which is not a polynomial.

3. Even if we find a system of functions χ_n enabling the expression of the bivariate distribution for fixed h as (6.22), the model is isofactorial only if the factors χ_n do not depend on h , which cannot be taken for granted (cf. discussion in Matheron, 1989).

6.3.3. Main Isofactorial Models

The class of bivariate distributions of an SRF coincides with the class of (non-centered) direct and cross-covariances of indicators obtained by thresholding the random function. Since there does not exist a generalization of Bochner's theorem characterizing direct and cross-covariances of indicators, the same is true for bivariate distributions, and this remains the case in the limited scope of isofactorial models. The main source of models is derived from special random functions whose bivariate distributions can be written down in closed form: when these are isofactorial, the model obtained is isofactorial by construction. Unsurprisingly there is a large number of special models that more or less overlap. It is not possible to give a complete inventory here (Matheron alone devoted over 600 pages to these models). We will present one model akin to the mosaic model and the classic models with polynomial factors, which have been used in most applications. Since in practice the isofactorial model does not directly concern the raw variable Z but a transform Y , we will assume, without loss of generality, that the transformed variable is standardized, and we will only consider this type of marginal distribution (standard normal for example). Let us start with two simple examples that correspond to random functions we have already encountered, the bi-gaussian and the mosaic models.

Bi-Gaussian Model

The normalized Hermite polynomials $\chi_n(y) = H_n(y)/\sqrt{n!}$ form an orthonormal basis with respect to the standard normal distribution. Hermite polynomials H_n are defined by Rodrigues's formula⁵

$$H_n(y)g(y) = \frac{d^n}{dy^n} g(y)$$

and are easily calculated by the recurrence relation

$$H_{n+1}(y) = -yH_n(y) - nH_{n-1}(y)$$

The bivariate normal distribution $G_\rho(dy, dy')$ with correlation coefficient ρ ($-1 < \rho < 1$) has the bivariate p.d.f.

$$g_\rho(y, y') = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{y^2 - 2\rho yy' + y'^2}{2(1-\rho^2)}\right)$$

It can be represented by the series

$$g_\rho(y, y') = \sum_{n=0}^{\infty} \rho^n \chi_n(y) \chi_n(y') g(y) g(y') \quad (6.25)$$

It is an isofactorial model, and the covariances T_n are equal to ρ^n . If $Y(x)$ is an SRF with correlogram $\rho(h)$, whose pairs $(Y(x), Y(x+h))$ are Gaussian, their bivariate distributions are of the form (6.22) with $T_n(h) = \rho(h)^n$.

The possibility to develop the bivariate p.d.f.s as well as the functionals of $Y(\cdot)$ can be exploited for covariance calculations. An interesting example is the calculation of the covariance of indicators obtained by thresholding Y (Matheron, 1975b). From (A.10) the indicator $1_{Y(x)<y}$ can be represented by the series

$$1_{Y(x)<y} = G(y) + g(y) \sum_{n=1}^{\infty} \frac{\chi_{n-1}(y)}{\sqrt{n}} \chi_n(Y(x)) \quad (6.26)$$

By application of (6.24), the cross-covariance of the indicators $1_{Y(x)<y}$ and $1_{Y(x+h)<y'}$ associated with thresholds y and y' is

$$C_{yy'}(h) = g(y)g(y') \sum_{n=1}^{\infty} \frac{\chi_{n-1}(y)\chi_{n-1}(y')}{n} \rho(h)^n \quad (6.27)$$

This expression can be used directly for numerical computations. From an analytical point of view, for fixed y and y' this is a function of $\rho \equiv \rho(h)$, whose

derivative is

$$\frac{dC_{yy'}}{d\rho} = g(y)g(y') \sum_{n=1}^{\infty} \chi_{n-1}(y)\chi_{n-1}(y')\rho^{n-1}$$

We recognize the isofactorial development of the bivariate density $g_{\rho}(y, y')$. By integration from 0 to $\rho(h)$, we obtain

$$C_{yy'}(h) = \int_0^{\rho(h)} g_u(y, y') du$$

In the special case $y = y'$, the direct covariance of the indicator is found to be

$$C_y(h) = \frac{1}{2\pi} \int_0^{\rho(h)} \exp\left(-\frac{y^2}{1+u}\right) \frac{du}{\sqrt{1-u^2}}$$

which establishes formula (2.76).

Mosaic Model

In the case of a mosaic random function, the random variables $Y(x)$ and $Y(x+h)$ are equal with probability $\rho(h)$, or uncorrelated with probability $1 - \rho(h)$. Their bivariate distribution is

$$G_h(dy, dy') = \rho(h)G(dy)\delta_y(dy') + [1 - \rho(h)]G(dy)G(dy')$$

where $G(y)$ is the marginal c.d.f. of $Y(\cdot)$ and $\rho(h)$ its correlogram. If $\varphi(\cdot)$ is a measurable function such that $\int \varphi(y)G(dy) = m_{\varphi}$, we have

$$E[\varphi(Y(x+h)) | Y(x)] = \rho(h)\varphi(Y(x)) + [1 - \rho(h)]m_{\varphi}$$

In particular, if $\{\chi_n(y), n = 0, 1, \dots\}$ constitutes an orthonormal family of factors for the distribution G , the relations (6.11) are satisfied with $T_n(h) = \rho(h)$ for $n > 0$ (and $T_0(h) \equiv 1$). The mosaic model is therefore an isofactorial model. However, its isofactorial representation brings nothing new, at least as long as it is not complemented by a change-of-support model.

Why Other Models?

The preceding two models correspond to extreme cases. For the same exponential covariance, the first one can represent a Gaussian diffusion process with continuous realizations, whereas the second one represents a random partition into Poisson polygons. One can think of describing other situations by introducing a transformation to an SRF with a Gaussian marginal. But experience has shown that a Gaussian transformation is not always desirable nor even possible. Besides, even if we can obtain an SRF with a Gaussian

marginal, it does not follow that its bivariate distributions are Gaussian or mosaic. Therefore we need to broaden the class of possible models.

We will only give a glimpse of the two approaches that have received most attention (there are many others). The first one is to extend the mosaic model to less systematic effects, leading to the model with orthogonal indicator residuals, whose factors are not polynomials. On the contrary, the second approach is to develop models with polynomial factors. First, by a systematic study of diffusion processes we will identify all the classic marginal distributions that can be considered to construct isofactorial models with polynomial factors. Next, for each of these marginals we will characterize the class of isofactorial models that can be associated with it, in other words the general form of covariances $T_n(h)$.

Model with Orthogonal Indicator Residuals

A mosaic random function has no edge effects in the sense that it can change abruptly from a low to a high value. It is therefore suitable to represent a deposit where we pass without transition from clear-cut waste to rich ore, or to model a pollution that remains concentrated near the contamination sources. However, even among models without edge effects, the mosaic model is an extreme case. The model with orthogonal indicator residuals, which belongs to the same class, has been developed by Rivoirard (1988, 1989) to capture more varied spatial structures.

This model includes notably some Boolean integer-valued random functions like those presented in Figure 7.35 (Section 7.8.1). In these examples $Y(x)$ is obtained by superposition of independent random sets A_1, \dots, A_N . Starting from a situation where all points are assigned a zero value, we set to 1 all points that belong to the random set A_1 , then to 2 all points that belong to A_2 , and so on, until we set to the value N the points that belong to A_N (at each update the preceding values are overridden). Realizations of the random function are therefore piecewise constant and display jumps whose amplitude can be greater than one. It can be shown that $Y(\cdot)$ has isofactorial bivariate distributions whose factors are, up to a normalizing constant

$$\chi_n(y) = \frac{1_{y < n}}{F_n} - \frac{1_{y < n+1}}{F_{n+1}} \quad n = 1, 2, \dots, N-1$$

$$\chi_0(y) = 1$$

where the $F_n = \Pr\{Y(x) < n\} = \prod_{m=n}^N \Pr\{x \notin A_m\}$ constitute the marginal c.d.f. of $Y(\cdot)$. The factors are therefore linear combinations of successive indicators. The model has been named after this property. Note that this model has a finite number of factors.⁶ It can be shown that the direct and cross-variograms of indicators are no longer all identical, up to a multiplicative factor, as is the case for a mosaic RF. But the cross-variogram of two indicators is proportional to the direct variogram of the higher indicator, which allows an easy check of the suitability of the model for a given data set. This model captures a destructuring of low grades. In applications a destructuring of high grades is more common, and the model is constructed in the reverse manner; that is, all points are initialized with the value N and random sets associated with decreasing values are superposed. The cross-variogram of two indicators is then proportional to the direct variogram of the lower indicator.

Diffusion Models

Diffusion models are the prototype of models with edge effect, in the sense that to go from a value to another all intermediate values must be visited. Such model is appropriate, for example, in the case of a diffuse pollution or a deposit where there is no clear-cut discontinuity between the

waste and the ore. From a theoretical standpoint, diffusion processes are symmetric continuous-time Markov processes and as such have isofactorial bivariate distributions (Matheron, 1989). In the continuous case, namely when they take on their values in \mathbb{R} , they satisfy a diffusion equation similar to the heat equation (Feller, 1968, sec. XIV.6; Feller, 1971, secs. X.4 and X.5; Matheron, 1975b; Armstrong and Matheron, 1986a). Discrete diffusion processes, also named birth-and-death processes (Feller, 1971, sec. XIV.6) take on their values in \mathbb{N} and only progress by unit jumps (+1 or -1). Their general study from the point of view of isofactorial properties is due to Matheron (1975c, 1984e, 1989). The system being in state i at time t (which means that $Y(t) = i$), in the interval $[t, t + dt]$ goes to state $i + 1$ with probability $a_i dt$, to state $i - 1$ with probability $b_i dt$, and stays in state i with probability $1 - (a_i + b_i)dt$. The birth rates a_i and death rates b_i are called the diffusion coefficients of the process. The marginal distribution G of the stationary process $Y(t)$ associated with these diffusion coefficients, defined by the probabilities $p_i = \Pr\{Y(t) = i\}$, satisfies

$$a_i p_i = b_{i+1} p_{i+1} \quad i = 0, 1, 2, \dots$$

It is shown in the general case that the covariances $T_n(h)$ of the factors of a diffusion process are of the form $\exp(-\lambda_n |h|)$. The process itself therefore has for covariance an exponential function (when the factors are polynomials) or a sum of exponentials (in the general case). The parameter λ_0 is zero, and the others are positive and increase with n . Their more or less rapid growth characterizes the destructuring of the factors with n . In the case where the marginal is Gaussian, $\lambda_n = n$, and we find again covariances $T_n(h)$ of the form $\rho(h)^n$; however, this circumstance is not general.

Diffusion processes are defined in one dimension. They are extended to higher dimensions by the substitution method. If $Y(t)$, $t \in \mathbb{R}$ is a diffusion process and if $T(x)$, $x \in \mathbb{R}^p$, is a random function with strictly stationary increments and a bounded variogram, the RF $\tilde{Y}(x) = Y(T(x))$ is an SRF whose bivariate distributions are isofactorial. It has the same marginal and same factors λ_n as the diffusion process $Y(t)$, but their covariances change: the covariances $\tilde{T}_n(h)$ ($h \in \mathbb{R}^p$) of the factors of the RF $\tilde{Y}(x)$ are

$$\tilde{T}_n(h) = \text{Cov}(\chi_n(\tilde{Y}(x)), \chi_n(\tilde{Y}(x+h))) = E[\exp(-\lambda_n |T(x+h) - T(x)|)]$$

In practice, the studied SRF is assumed to be of the form $\tilde{Z}(x) = \varphi(\tilde{Y}(x))$. If the function $\varphi(y)$ can be expanded as $\sum_{n=0}^{\infty} \varphi_n \chi_n(y)$, the covariance of \tilde{Z} is

$$\begin{aligned} C_{\tilde{Z}}(h) &= \sum_{n=1}^{\infty} \varphi_n^2 \tilde{T}_n(h) = E \left[\sum_{n=1}^{\infty} \varphi_n^2 \exp(-\lambda_n |T(x+h) - T(x)|) \right] \\ &= E[C_Z(|T(x+h) - T(x)|)] \end{aligned}$$

where $C_Z(\cdot)$ is the covariance of the stationary process $Z(t) = \varphi(Y(t))$. Additional results on random functions obtained by substitution can be found in Section 7.7.6.

It can also be shown that any set of covariances of the form $\tilde{T}_n(h) = \exp(-\lambda_n \gamma(h))$, where $\gamma(h)$ is any unbounded variogram in \mathbb{R}^p , leads to a valid model. This type of model can be used to define numerically the bivariate distributions of discrete diffusive SRFs. But the inference of the parameters of the underlying diffusion process and of the variogram $\gamma(h)$ requires specific tools and is very tricky (Lajaunie and Lantuéjoul, 1989).

For applications the preferred approach is to consider diffusion processes with polynomial factors known explicitly (Rodrigues formula definition, calculation of $\chi_{n+1}(y)$ from $\chi_n(y)$ and $\chi_{n-1}(y)$ using a recurrence relation whose parameters are known explicitly). Only eight classes of diffusion processes meet this condition. Three of them have a continuous marginal:

- The Gaussian model with Hermite polynomials.

- The gamma model with Laguerre polynomials. The gamma distribution, defined for positive variables, broadens the scope of the Gaussian model.
- The beta model with Jacobi polynomials, which has hardly ever been used in geostatistics so far (the beta distribution with parameters α and β is that of $X_1/(X_1 + X_2)$ where X_1 and X_2 are two independent gamma random variables, with the same scale and respective parameters α and β).

The five other classes of models are associated with a discrete marginal distribution:

- The binomial model with Krawtchouk polynomials, which is the discrete equivalent of the Gaussian model.
- The negative binomial model with Meixner polynomials, which is the discrete equivalent of the gamma model.
- The Poisson model with Charlier polynomials, which can be obtained as limit of the preceding ones.
- The discrete Jacobi type model with discrete Jacobi polynomials, which is the discrete equivalent of the continuous model with a beta distribution.
- The discrete anti-Jacobi type model which is related to the Jacobi one but has no continuous equivalent.

Other well-known models which are special cases of the above are of course to be included. Thus for $\alpha = \beta$ the Jacobi polynomials are ultraspherical polynomials, and in particular the Chebyshev polynomials for $\alpha = \beta = \frac{1}{2}$ and the Legendre polynomials for $\alpha = \beta = 1$; in this latter case the beta distribution is the uniform over $[0, 1]$. Diffusion models with a binomial, negative binomial, and Poisson distribution can be obtained as limits of the Jacobi type model. Continuous diffusion models (Gaussian, gamma, beta) can also be obtained as limits of this model.

For the Gaussian, gamma, binomial, negative binomial, and Poisson models, λ_n is proportional to n , so that the $T_n(h)$ are of the form $\rho(h)^n$. For the beta and discrete Jacobi models the decrease of $T_n(h)$ with n is faster than that of $\rho(h)^n$, whereas it is slower for the anti-Jacobi model.

The main properties of the polynomials associated with the Gaussian, gamma, or negative binomial distributions, which are the most common target marginals, are summarized in the Appendix. The reader is referred to Beckman (1973) for additional results on most of these families of polynomials, as well as Armstrong and Matheron (1986b) and Matheron (1984a, 1984c, 1984e, 1989) for isofactorial models.

Pure Models with Isofactorial Factors

We call a pure model a model for which the $T_n(h)$ are of the form $\rho(h)^n$. We have seen that diffusion processes with a Gaussian marginal are of this type with an exponential correlogram $\rho(h)$. By substitution they provide SRFs which remain of this type with a broader choice of correlograms satisfying $\rho(h) > 0 \forall h$. We have likewise seen that random functions with a Gaussian spatial distribution, and thus Gaussian bivariate distributions, also have $T_n(h)$ of type $\rho(h)^n$ but without restriction on the correlogram $\rho(h)$. What for other marginal distributions? Grouping the results obtained by Beckman (1973, sec. 6.2 and app. I) for continuous variables and by Matheron (1980) for discrete distributions, we obtain the following result.

Given a marginal distribution and its associated orthonormal polynomials, the coefficients $T_n = \rho^n$ do define a bivariate distribution

- for any $\rho \in [-1, 1]$ if the marginal is Gaussian, binomial, or beta with $\alpha = \beta$;
- provided that $\rho \in [0, 1]$ if the marginal is gamma, beta ($\alpha \neq \beta$), negative binomial, or Poisson.

For a random function a model $T_n(h) = \rho(h)^n$ is valid provided that $\rho(h)$ satisfies the above condition for all h and is a correlogram (this being sufficient to ensure that $\rho(h)^n$ is a covariance for

all n). Such a model shows a rapid destructuring of the factors as n increases, commensurate with the decrease of $\rho(h)^n$ with n .

Mixture of Pure Models: Hermitian, Laguerre, and Meixner Models

The univariate distribution of $Y(x)$ may be Gaussian and not its bivariate distributions. If the SRF $Y(x)$, with marginal $G(dy)$, has a system of orthonormal polynomials $\chi_n(y)$ and isofactorial bivariate distributions, this does not imply that the covariances $T_n(h)$ of the factors are of the form $\rho(h)^n$ (i.e., assuming that such a model is valid, which we do not know in the discrete Jacobi or anti-Jacobi case). For a pair of variables, when the pure model associated with $T_n = \rho^n$ exists, it can be generalized by randomization of the correlation coefficient ρ , namely by taking T_n of the form

$$T_n = \int r^n \varpi(dr) \quad (6.28)$$

The support of the distribution ϖ must be included in the interval of permissible values of ρ , that is, the interval $[-1, 1]$ if the marginal is Gaussian or binomial and $[0, 1]$ if the marginal is gamma, negative binomial, or Poisson. Conversely, at least in the case of a Gaussian, gamma, or negative binomial marginal, any isofactorial distribution with polynomial factors can be expressed with coefficients T_n of the form (6.28) (Matheron, 1976a; Sarmanov, 1968; Matheron, 1984a, respectively). The corresponding isofactorial models are called by the name of their associated polynomials: *Hermitian model* (Gaussian marginal), *Laguerre model* (gamma marginal), *Meixner model* (negative binomial marginal). It can be shown that if T_1 is a given positive value ρ , T_n is minimal and equal to ρ^n when the distribution ϖ is concentrated at the value ρ . In other words, pure models have the property of corresponding to the maximum destructuring.

Considering an SRF and not only a pair of random variables, the distribution ϖ associated with the $T_n(h)$ usually depends on h . We will denote it by ϖ_h . Restrictions must be placed on the family of distributions ϖ_h to guarantee that $T_n(h)$ is a covariance function for all values of n . In applications, in order to keep statistical inference simple, one limits the choice to distributions ϖ_h such that the $T_n(h)$ can be expressed as functions of the correlogram $\rho(h)$.

Example of Bi-Gamma Distributions

By analogy with the bi-gaussian case, there are two ways of defining a bi-gamma distribution. But unlike in the bi-gaussian case these two definitions do not coincide. The first way is to define the bi-gamma distribution as a pure isofactorial model with T_n of the form ρ^n . We have seen that the bivariate distributions of gamma diffusion processes are of this type. The second way is to start from three independent unit-scale gamma random variables X_0 , X_1 , and X_2 , with respective parameters $\rho\alpha$, $(1 - \rho)\alpha$ and $(1 - \rho)\alpha$, with $0 < \rho < 1$, $0 < \alpha < 1$, and define

$$Y_1 = X_0 + X_1$$

$$Y_2 = X_0 + X_2$$

Y_1 and Y_2 are two unit-scale gamma random variables, with the same parameter α and correlation coefficient ρ . Matheron (1973b) shows that their bivariate distribution can be written in the isofactorial form (6.22) with normalized Laguerre polynomials and coefficients

$$T_n = \frac{\Gamma(\alpha)}{\Gamma(\alpha + n)} \frac{\Gamma(\alpha\rho + n)}{\Gamma(\alpha\rho)}$$

This quantity is the n th-order moment of the beta distribution with parameters $\rho\alpha$ and $(1 - \rho)\alpha$, mean ρ , and is therefore of the form (6.28). This result can be immediately transposed

to negative binomial random variables (e.g., see Feller, 1968, p. 285). SRFs with this type of bivariate distribution can be obtained by regularization of a stationary orthogonal random measure with gamma or negative binomial distribution. When applied to random measures with a Gaussian or Poisson distribution, this procedure leads, however, to pure bivariate models (Matheron, 1973b; Armstrong and Matheron, 1986a).

Other Isofactorial Models

Let us recall that the factors of isofactorial models are not necessarily polynomials. For example, the uniform distribution over $[0, 1]$ can be associated with various bivariate distributions with different factors (Matheron, 1975c):

- Legendre polynomials.
- Trigonometric factors $\cos(2\pi n\theta)$ and $\sin(2\pi n\theta)$ (e.g., Brownian motion on the circle).
- Walsh functions (nonpolynomial); this model exists in continuous and in discrete versions.
- Factors of the form $\chi_n(G^{-1}(z))$, obtained by the change of variable $Z(\cdot) = G(Y(\cdot))$, starting from a model with factors $\chi_n(y)$ and an absolutely continuous marginal $G(dy)$. These factors are generally not polynomial functions.

6.3.4. Practice of Disjunctive Point Kriging

Isofactorial models are very diverse, but in fact only a few have been used. The methodology has, naturally, first been developed for the Hermitian model (Maréchal, 1976). After a preliminary transformation to normal, this model can represent a large class of variables with a continuous, possibly mildly skewed, distribution. Nevertheless, we can deplore that so many applications have used the bi-gaussian model, which is quite special indeed, without questioning the validity of this choice. The application methodology for the Laguerre model is essentially due to Hu and Lantuéjoul (1988) and Hu (1988). This model extends the capabilities of the bi-gaussian model to distributions that exhibit a cluster of values near the origin and a long distribution tail. As for the Meixner model, it makes it possible, at the price of a discretization, to properly account for a large proportion of zero values, or to study a discrete variable directly (Demange et al., 1987; Kleingeld, 1987). The Hermitian, Laguerre, and Meixner models comprise a range of bivariate distributions that go from the mosaic to the pure diffusive model. Thus we have elected to limit our presentation of the practice of disjunctive kriging to these three models. Table 6.1 presents their main features: the left-hand columns give for each type of marginal distribution the model used and the corresponding family of polynomials, and the right-hand columns give the names of the bivariate distributions that can be associated with them. Other than that, the main model that has been applied, to our knowledge, is the model with orthogonal indicator residuals (Bordessoule et al., 1988, 1989; Rivoirard, 1994, in particular the application of ch. 13).

Choice of the Type of Transformation

In applications the data are seldom obliging enough to have a histogram matching one of the classic marginals. We have to use a transformation in which

TABLE 6.1. Names of main isofactorial models with polynomial factors classified by type of marginal and bivariate distributions

Marginal Distribution			Bivariate Distribution		
Type	Distribution	Polynomials	$T_n = \rho^n$	Intermediate	$T_n = \rho$
Continuous symmetric	Gaussian	Hermite	Bi-gaussian	Hermitian model	Mosaic Gaussian
Continuous skewed	Gamma	Laguerre	Pure gamma	Laguerre model	Mosaic gamma
Discrete	Negative binomial	Meixner	Pure neg. binomial	Meixner model	Mosaic neg. binomial

the variable of interest $Z(\cdot)$ is assumed of the form $Z(x) = \varphi(Y(x))$, and $Y(\cdot)$ is supposed to conform to an isofactorial model whose marginal is one of the classic distributions. As will be shown further, the transformation function φ is used through the coefficients φ_n of its expansion in the basis of orthonormal polynomials $\chi_n(y)$. The more similar the shapes of the marginal distributions G of $Y(\cdot)$ and F of $Z(\cdot)$, the fewer terms this expansion can be truncated to and the more robust it is. This explains why a transformation to a uniform, which at first glance would seem appealing, is rarely advisable, since it would “pack” extreme data too much, with the consequence that a large number of Jacobi polynomials would be required to represent the transformation function.

When the data histogram reflects a continuous, not too skewed, distribution, a Gaussian transformation is appropriate. But if the data histogram is very skewed, with a large proportion of low values and a long tail, a Gaussian transformation will limit the amplitude of large values, which may be the goal, but also magnify unimportant differences between low values. To resolve this problem, it is better to transform to a skewed variable. The gamma distribution is particularly interesting for its varied behaviors. A positive random variable Y follows a (standard) gamma distribution with parameter $\alpha > 0$ if its density is

$$g_\alpha(y) = \frac{1}{\Gamma(\alpha)} e^{-y} y^{\alpha-1} \quad (y > 0)$$

The shape of the density g_α depends on α : if $\alpha < 1$, g_α is a decreasing function, unbounded at the origin; if $\alpha > 1$, g_α is a bell-shaped curve that tends to the Gaussian for large values of α ; in the intermediate case $\alpha = 1$, g_1 is the density of the exponential distribution.

The data histogram can have an atom at the origin. For example, if Z represents the thickness of a geological formation, this may be absent, which is reflected in a proportion $p_0 > 0$ of zero values. If Z is a mineral grade, there may be a nonnegligible proportion p_0 of zero values (or at least considered as such because below the detection limit). This is called the *zero effect*. In the case of a Gaussian (or gamma) transformation, φ still exists but is not

one to one, so we do not know which value of Y should be associated with a zero value of Z ; we only know that Y belongs to the interval $]-\infty, G^{-1}(p_0)]$. This problem disappears if we choose for Y a discrete distribution such that $\Pr\{Y = 0\} = p_0$. On the other hand, we must discretize the continuous part of the distribution of Z , which is not a major hurdle if we can master this discretization. The negative binomial distribution is a good candidate, for it enables the representation of both a zero effect and a strong skewness of the rest of the distribution that generally comes with it. The negative binomial distribution with parameters $\alpha > 0$ and $p \in [0, 1]$ is defined by

$$p_i = (1 - p)^\alpha \frac{\Gamma(\alpha + i)}{\Gamma(\alpha)} \frac{p^i}{i!} \quad i \geq 0$$

The case $\alpha = 1$ corresponds to the geometric, or Pascal, distribution $p_i = (1 - p)p^i$, which is the discrete version of the exponential distribution. In general, the negative binomial distribution appears as a discrete version of the corresponding gamma distribution with the same parameter α .

In these last two cases one problem is the choice of the parameter α (in the case of the negative binomial the choice of the parameter p results from the probability p_0 to be matched). It seems reasonable to choose a low value for α when the histogram is skewed and the relative variance high. Demange et al. (1987) choose for α the value that also preserves the ratio of ore/waste cutoff to the mean. A similar criterion is the preservation of the ratio of the median to the mean. Hu (1988) proposes to preserve a parameter that characterizes the relative dispersion of the distribution, for example, the relative variance or the selectivity index. The latter, which is studied in length in Section 6.4.1, is defined for a distribution $F(dz)$, as $s = S/m$ where S is the dispersion indicator

$$S = \frac{1}{2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(dz) |z' - z| F(dz') \quad (6.29)$$

and m the mean. The dispersion indicator is a more robust parameter than the variance, which has a definition similar to (6.29) with $(z' - z)^2$ instead of $|z' - z|$. One can take as an estimator of the dispersion indicator of the variable Z the empirical dispersion indicator \hat{S} presented below. Concerning Y , the expression of the dispersion indicator of the gamma and negative binomial distributions is given by (A.13) and (A.19).

Determination of the Transformation Function

Having selected the marginal distribution $G(dy)$ of $Y(x)$, we have to determine the transformation function φ such that $Z(x) = \varphi(Y(x))$. We will need $\varphi(y)$ in the form of its expansion into factors $\chi_n(y)$:

$$\varphi(y) = \sum_{n=0}^{\infty} \varphi_n \chi_n(y)$$

The function φ is simply equal to $F^{-1} \circ G$, where F is the marginal of the SRF $Z(x)$. In practice, we usually only know an approximation of F , namely the empirical \hat{F} given by (6.1), obtained from the data z_n appropriately weighted by λ_n . This empirical distribution can be fitted to a c.d.f. given in closed form, or be fitted graphically as is usually done for conditional simulations (see Section 7.3.2 and the example in Section 7.11.1). But it is not necessary to go through these steps for the purpose of a disjunctive kriging: in practice we derive the coefficients φ_n from the empirical distribution \hat{F} , and we truncate the expansion to an order n_{\max} (Lantuéjoul and Rivoirard, 1984; Hu and Lantuéjoul, 1988).

To this end, the N data z_n are sorted by ascending values (assuming no ties to simplify the presentation). We denote the sorted data by z_i ($i = 1, \dots, N$), the associated weights by λ_i (e.g., $\lambda_i = 1/N$), and by convention let $z_0 = -\infty$ and $z_N + 1 = +\infty$. Thus we have

$$-\infty = z_0 < z_1 < z_2 < \dots < z_{i-1} < z_i < z_{i+1} < \dots < z_N < z_{N+1} = +\infty$$

\hat{F} is constant over each interval $]z_{i-1}, z_i]$ and takes on the value

$$\hat{F}(z) = \sum_{j=1}^{i-1} \lambda_j = F_i \quad z \in]z_{i-1}, z_i] \quad i = 1, \dots, N+1$$

with the understanding that $F_1 = 0$ and $F_{N+1} = 1$. Let us denote by y_i the value of the variable $Y(x)$ such that $\hat{F}(z_i) = G(y_i)$, namely

$$y_i = G^{-1}(F_i) \quad i = 1, \dots, N+1$$

with in particular $y_1 = -\infty$ (or 0 for a gamma distribution) and $y_{N+1} = +\infty$. If the function $\hat{F}(z)$ were strictly monotone like $G(y)$, the transformation function would relate the values y and z such that $G(y) = \hat{F}(z)$. But here $\hat{F}(z)$ only increases at the points z_i , jumping from F_i (at z_i) to F_{i+1} (at $z_i + 0$). The value z_i must therefore be associated with the values of y such that $F_i < G(y) \leq F_{i+1}$, namely with the interval $]y_i, y_{i+1}]$. Hence

$$\varphi(y) = z_i \quad y \in]y_i, y_{i+1}] \quad i = 1, \dots, N$$

and

$$\varphi_n = \int \varphi(y) \chi_n(y) G(dy) = \sum_{i=1}^N z_i \int_{y_i}^{y_{i+1}} \chi_n(y) G(dy)$$

In the case of a Gaussian or gamma transformation, the integral in this expression can be calculated analytically using relation (A.9) or (A.17).

Since the expansion is truncated to an order n_{\max} , the transformation function φ obtained is a polynomial of degree n_{\max} . It is usually not monotone and

thus takes on meaningless values for extreme values of y . We select n_{\max} so that φ is monotone in the actual domain of variation of y and the sum of the φ_n^2 , for n varying from 1 to n_{\max} , is close to the variance of $Z(x)$. Usually one to a few dozen terms are sufficient.

While computing the empirical distribution \hat{F} , we can also compute the corresponding empirical dispersion indicator \hat{S} using the formula (Matheron, 1981b)

$$\hat{S} = \sum_{i=1}^N \lambda_i (2F_i + \lambda_i - 1) z_i$$

Like the empirical variance it is biased. Denoting the dispersion indicator of the distribution F by S , one has $E[\hat{S}] = (1 - 1/N)S$ in the case of independent random variables Z_i (which they usually are not).

Choice of Isofactorial Model

The isofactorial model is specified once the covariances $T_n(h)$ of the factors $\chi_n(Y(x))$ are known for all $n > 0$ (since in all cases $T_0(h) \equiv 1$). Since the models considered here have polynomial factors, the covariance $T_1(h)$ is simply the correlogram $\rho(h)$ of the SRF $Y(x)$. It is determined by a classic structural analysis of the transformed data $Y(x_n)$, since it is related to the variogram $\gamma(h)$ of $Y(x)$ by

$$\gamma(h) = C[1 - \rho(h)]$$

where the sill C is equal to 1 in the Gaussian case, to α in the gamma case, and to $\alpha p/(1-p)^2$ for a negative binomial transformation.

We have seen in Section 6.3.3 (relation (6.28)) that for the Hermitian, Laguerre, and Meixner models the covariances of the factors are of the form

$$T_n(h) = \int r^n \varpi_h(dr)$$

where ϖ_h is a probability distribution concentrated on the interval $[-1, 1]$ in the Hermitian case and on the interval $[0, 1]$ for Laguerre and Meixner models. In practice, one limits the choice to isofactorial models where the covariances $T_n(h)$ are functions of $\rho(h)$ and where $\rho(h)$ is nonnegative. Four isofactorial distribution models are used most (Fig. 6.3):

1. *The pure diffusive model*, with a distribution ϖ_h concentrated at the single value $\rho(h)$:

$$T_n(h) = \rho^n(h) \quad n \geq 0$$

The covariances of the factors tend to a pure nugget effect when n increases.

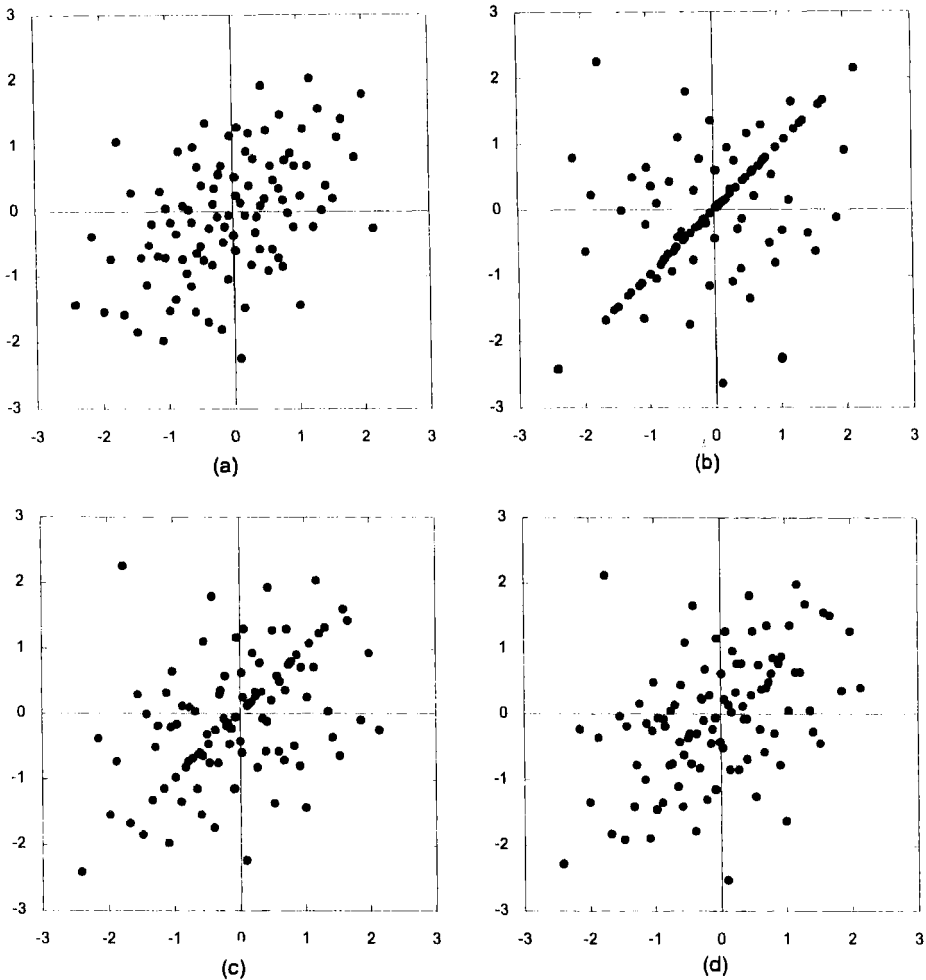


FIGURE 6.3. Scatterplot of 100 pairs $(Z(x), Z(x+h))$ with standard normal marginal distributions for a value h such that $\rho(h) = 0.5$: (a) Gaussian bivariate distribution; (b) mosaic model; (c) barycentric model with $\beta = 1/2$; (d) beta model with $\beta = 2$.

2. *The mosaic model*, which corresponds to a distribution ϖ_h concentrated at the values 0 and 1, with respective probabilities $1 - \rho(h)$ and $\rho(h)$:

$$T_n(h) = \rho(h) \quad n \geq 0$$

All factors have the same covariance.

3. *The barycentric model*, which is a mixture of the preceding two, in proportions β and $1 - \beta$: the distribution ϖ_h is concentrated at the values 0, $\rho(h)$ and 1, respectively, with probabilities $(1 - \beta)(1 - \rho(h))$, β and

$(1 - \beta)\rho(h)$, where β is a parameter between 0 and 1, so that

$$T_n(h) = \beta\rho^n(h) + (1 - \beta)\rho(h) \quad n > 0$$

4. *The beta model*, which is a mixture of all pure models associated with a positive correlation coefficient, since ϖ_h is a beta distribution with parameters $\beta\rho(h)$ and $\beta(1 - \rho(h))$, where β is a positive parameter; it leads to

$$T_n(h) = \frac{\Gamma(\beta)}{\Gamma(\beta + n)} \frac{\Gamma(\beta\rho(h) + n)}{\Gamma(\beta\rho(h))} \quad n \geq 0$$

For an SRF with gamma marginal this model generalizes the second type of bi-gamma distribution presented in Section 6.3.3. It has variants (see Hu, 1988).

In principle, the justification of the use of an isofactorial model and the choice of the model itself must proceed from the study of the regression curves $\chi_n(Y(x + h))$ as a function of $Y(x)$ or of $\chi_n(Y(x))$ (cf. relation (6.11)). In practice, it would be tedious to examine lagged scatterplots (or “ h -scattergrams”) for all values of n until n_{\max} and for various classes of lag h , although it is advisable to inspect a few for validation (e.g., Goovaerts, 1997). These models have the advantage of being distinguishable simply by inspection of the variogram of order 1 of $Y(x)$, defined as we have seen in Section 2.5.3 by

$$\gamma_1(h) = \frac{1}{2}E[|Y(x + h) - Y(x)|]$$

Its sill C_1 is in all cases the dispersion indicator S of the marginal distribution of Y , as results from definition (6.29), but its shape depends on the model. Taking as structure functions the normalized variograms of order 1 $\tilde{\gamma}_1(h) = \gamma_1(h)/C_1$ and of order 2 (the usual variogram) $\tilde{\gamma}(h) = \gamma(h)/C = 1 - \rho(h)$, we have

$$\tilde{\gamma}_1(h) = \begin{cases} \sqrt{\tilde{\gamma}(h)} & \text{for the pure diffusive model} \\ \tilde{\gamma}(h) & \text{for the mosaic model} \\ \beta\sqrt{\tilde{\gamma}(h)} + (1 - \beta)\tilde{\gamma}(h) & \text{for the barycentric model with} \\ & \text{parameter } \beta \\ \frac{\Gamma(\beta)}{\Gamma(\beta + \frac{1}{2})} \frac{\Gamma(\beta\tilde{\gamma}(h) + \frac{1}{2})}{\Gamma(\beta\tilde{\gamma}(h))} & \text{for the beta model with parameter } \beta \end{cases}$$

These relations express very simple links between the variogram of order 1 and the usual variogram of $Y(x)$. In particular, by plotting $\tilde{\gamma}$ as a function of $\tilde{\gamma}_1$, we obtain a straight line for the mosaic model, a parabola for the pure diffusive model, and intermediate behaviors for the other models (Fig. 6.4).

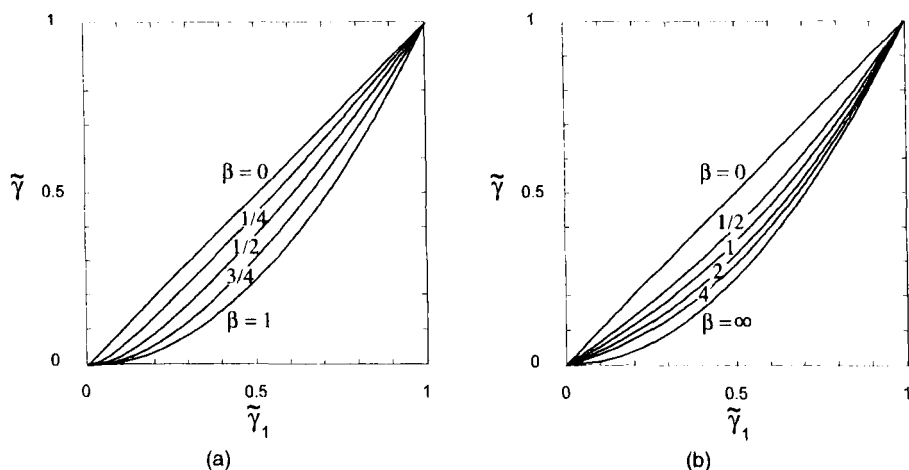


FIGURE 6.4. Relationship between the ordinary variogram and the variogram of order 1 of an SRF with Gaussian, gamma, or negative binomial marginal for the main isofactorial models (both variograms are normalized): (a) barycentric model; (d) beta model. The extremes correspond to the pure diffusive model (parabola) and the mosaic model (straight line).

The final choice of the model can be validated by inspection of the variogram of the initial variable $Z(x)$, whose theoretical value can be calculated from the covariances $T_n(h)$ using relation (6.23) with $\psi_n = \varphi_n$, at least if this variogram is robust enough. It can also be validated with indicator variograms whose theoretical expression is likewise given by (6.23) where the ψ_n are now the coefficients of the expansion of the indicator (these are given below).

For completeness, let us mention another tool proposed by Journel and Deutsch (1993) to summarize bivariate distributions, the bivariate entropy function $H(h)$ giving the entropy of the bivariate p.d.f. $g_h(y, y')$ of the $(Y(x), Y(x+h))$ pair as a function of h

$$H(h) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g_h(y, y') \log g_h(y, y') dy dy'$$

(the integral is in fact limited to the support of g_h). Among unbounded SRFs with the same covariance, Gaussian random functions correspond to maximum entropy, or “maximum disorder” (relation (6.28) has already shown that for a given marginal, the maximum destructuring is obtained with the pure diffusive models). But this entropy function cannot be determined experimentally, except if we have a complete image.

An Example of Structural Analysis

In a vein of the open-pit Salsigne gold mine (France), zones to be mined are delineated on the basis of boreholes on a $5 \text{ m} \times 5 \text{ m}$ grid. The data from

1-m long cores have been studied in detail (Liao, 1990; Chiès and Liao, 1993). Figure 6.5 displays the histogram of core gold grades. It has 30% of zero values and a long tail (grades spread out to 400 g/t, for a mean of 10 g/t). This suggests a gamma transformation. Choosing the parameter α so that the percentage of waste is matched (59% of data below 5 g/t $\simeq 0.5m_Z$), we obtain $\alpha = 0.34$, which corresponds to a very skewed distribution. If we only kept the first term in the expansion of the transformation function, which amounts to assuming that grades follow a gamma distribution (up to a shift), we would already account for 89% of the variance ($\varphi_1^2/\sigma_Z^2 = 0.89$). In practice, the expansion is truncated to $n_{\max} = 16$ Laguerre polynomials so that over 99% of the variance is reproduced. As a comparison, to achieve an equivalent precision with a Gaussian transformation requires 27 Hermite polynomials, and the first term accounts for only 42% of the variance (i.e., the histogram of grades is far from Gaussian).

Figure 6.6 shows the ordinary sample variogram (of order 2) $\hat{\gamma}$ and the variogram of order 1 $\hat{\gamma}_1$ of the gamma variable Y associated with Z , the graph of the relation between the normalized variograms $\tilde{\gamma}_1$ and $\tilde{\gamma}$ (limited to zones where the variograms do not fluctuate too much), as well as the variogram $\hat{\gamma}_Z$ of grades. An empirical bivariate distribution of Y data is displayed in Figure 6.6a. It is not easy to interpret. By contrast, Figure 6.6b suggests an intermediate model between the diffusive and mosaic models. A beta model with $\beta = 1$ has been chosen. The theoretical variogram γ has then been selected to fit the three sample variograms (Fig. 6.6c,d). Similar variograms can in fact be obtained starting from a Gaussian transformation. But a comparison of the performance of the two models (Hermitian and Laguerre) for predicting change-of-support effects demonstrated, as expected, the clear superiority of the model based on the gamma transformation. Other examples can be found in Hu and Lantuéjoul (1988).

Implementation of Kriging

The disjunctive kriging estimate of $Z_0 = \psi(Y(x_0))$ and the DK variance are deduced from the simple kriging estimates of the factors $\chi_n(Y(x_0))$ by relations (6.16) and (6.18). The generalization to the DK of the mean value of $\psi(Y(x))$ in a volume v is straightforward (cf. Section 3.5.2). The coefficients ψ_n are obtained by application of (6.20). Let us mention a few classic cases (for clarity the index c denotes a fixed cutoff or threshold):

- *Disjunctive kriging of $Z(x_0)$.* We simply have $\psi_n = \varphi_n$. However, disjunctive kriging is rarely used as a replacement of ordinary kriging because the improvement is generally minimal (Puente and Bras, 1986).
- *Disjunctive kriging of $I(z_c) = 1_{Z(x_0) < z_c}$.* Since $1_{Z(x_0) < z_c} = 1_{Y(x_0) < y_c}$ with $y_c = \varphi^{-1}(z_c)$, we have

$$\psi_n = \int_{-\infty}^{y_c} \chi_n(y) G(dy)$$

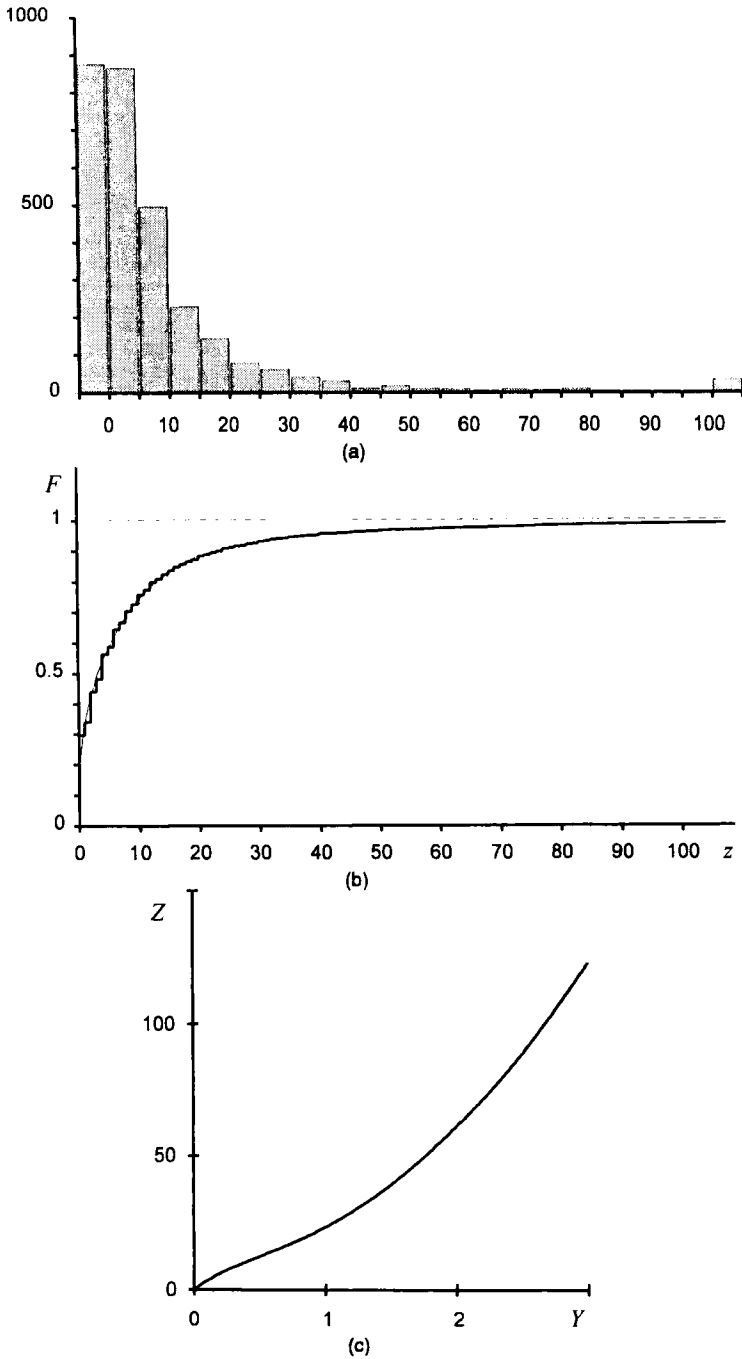


FIGURE 6.5. Salsigne deposit: gold grade distribution and gamma transformation: (a) histogram; (b) empirical c.d.f. and fit $G \circ \varphi^{-1}$; (c) transformation function φ obtained with 16 Laguerre polynomials. From Liao (1990).

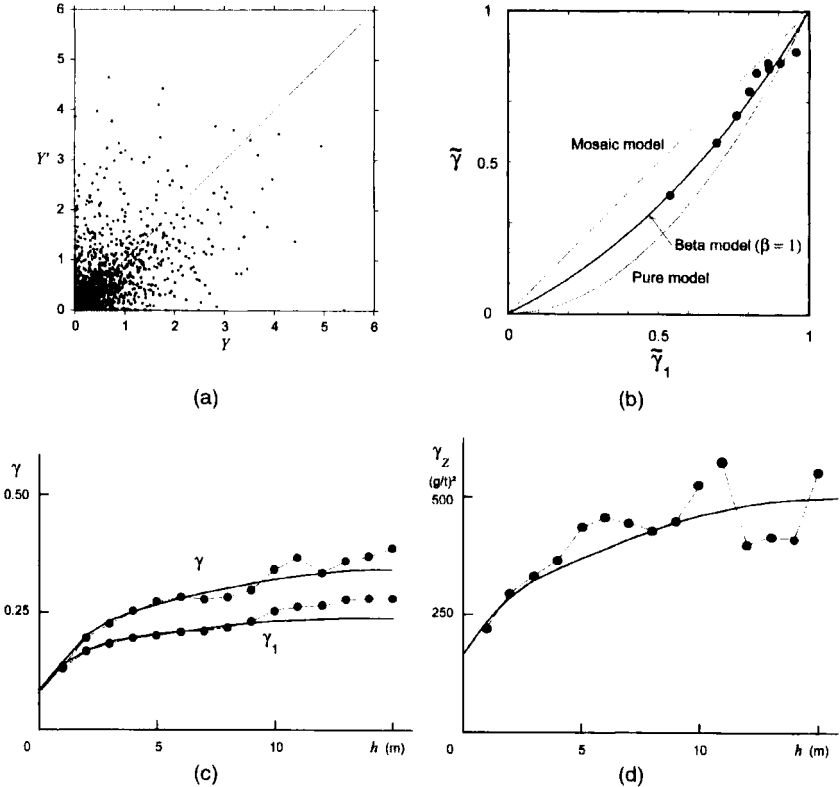


FIGURE 6.6. Variogram of gold grades and of gamma transformed data: (a) scatterplot of gamma transformed pairs $Y(x)$ and $Y' = Y(x + h)$ for $h = 1$ m vertically; (b) relationship between normalized ordinary and order 1 variograms of gamma transformed data; (c) ordinary variogram and variogram of order 1 of the gamma transformed data; (d) variogram of grades. From Liao (1990) and Chilès and Liao (1993), with kind permission from BRGM and Kluwer Academic Publishers.

The explicit form of the integral is deduced from (A.9) in the case of a Gaussian distribution and from (A.17) for a gamma distribution. For example, in the case of a Gaussian transformation we have

$$\psi_0 = G(y_c) \quad \psi_n = \frac{1}{\sqrt{n}} \chi_{n-1}(y_c) g(y_c) \quad n = 1, 2, \dots$$

- where G and g are, respectively, the c.d.f. and p.d.f. of the standard Gaussian and χ_n are the normalized Hermite polynomials.
- *Disjunctive kriging of point-support recovery functions.* By analogy with the (deterministic) recovery functions defined in Section 6.4.1, these are the (random) quantities

$$T(z_c) = 1_{Z(x_0) \geq z_c} \quad \text{and} \quad Q(z_c) = Z(x_0) 1_{Z(x_0) \geq z_c}$$

which can be rewritten as

$$T(z_c) = 1_{Y(x_0) \geq y_c} \quad \text{and} \quad Q(z_c) = \varphi(Y(x_0)) 1_{Y(x_0) \geq y_c}$$

with $y_c = \varphi^{-1}(z_c)$. Thus we have

$$\psi_n = \int_{y_c}^{\infty} \chi_n(y) G(dy) \quad \text{for } T(z_c)$$

$$\psi_n = \int_{y_c}^{\infty} \varphi(y) \chi_n(y) G(dy) = \sum_{m=0}^{\infty} \varphi_m \int_{y_c}^{\infty} \chi_m(y) \chi_n(y) G(dy) \quad \text{for } Q(z_c)$$

The Appendix provides elements for calculating the partial integrals (relations (A.9) and (A.11) in the Gaussian case, (A.17) and (A.18) in the gamma case).

The kriging systems (6.14) of the factors are constructed and solved iteratively, and the corresponding DK variances (6.17) are calculated along; since two successive systems are very similar, the solution of system n can be taken as an initial solution of system $n + 1$, and an iterative improvement of the solution quickly converges to the exact solution. In the case where $T_n(h) = \rho^n(h)$, the off-diagonal terms of the left-hand side matrix as well as the right-hand side terms tend rapidly to zero as n increases so that the estimator of $\chi_n(Y_0)$ also tends rapidly to zero with increasing n . This allows us to limit the expansion to a finite number of terms n_0 . For the more complex models presented above, the left-hand side matrix tends toward a matrix that is no longer the identity matrix. Still simplifications can be implemented, which are of the same nature as that presented below for the introduction of unbiasedness constraints. In any case it is not necessary to exceed the order at which the expansion of ψ is truncated. Note that this order generally differs from the truncation point of the transformation function φ . The expansion of ψ must include a relatively large number of terms if the objective is a strongly nonlinear quantity, as is the case for an indicator. Figure 6.7 shows the approximation of an indicator function by 20 and 50 Hermite polynomials. Even with 50 terms this approximation is not acceptable at the extremes (an indicator cannot be represented exactly by a finite number of polynomials).

Lognormal Disjunctive Kriging

If $Z(\cdot)$ is of the form $Z(x) = \exp(m' + \sigma'Y(x))$ and the marginal distribution of $Y(\cdot)$ is standard normal, the coefficients of the expansion of Z into normalized Hermite polynomials are (see relation (A.12))

$$\varphi_n = (-1)^n e^{m' + \sigma'^2/2} \frac{\sigma'^n}{\sqrt{n!}}$$

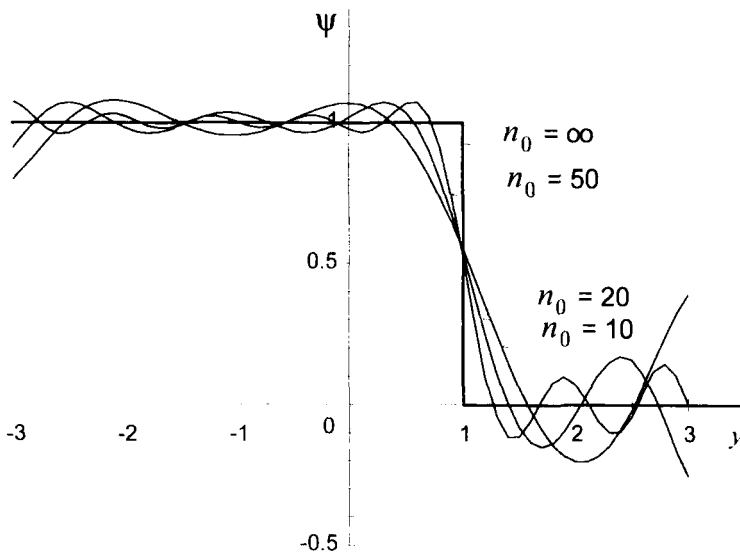


FIGURE 6.7. Approximation of the indicator function $\psi(y) = 1_{y < 1}$ by a finite number of Hermite polynomials.

This allows the definition of a lognormal disjunctive kriging. Note that contrary to the usual lognormal models, this one does not require a full Gaussian assumption for $Y(\cdot)$ but simply that its bivariate distributions be Hermitian.

Taking into Account Stationarity Defects

So far we have been assuming that $Z(x)$ is an SRF with a known mean, and even with a known marginal distribution. In practice, we only know an estimate of it, and it may also be that stationarity is only local. It is tempting, then, for calibration on the local characteristics of $Z(x)$, to impose the unbiasedness condition of ordinary kriging to the kriging of factors, and to substitute for all $n > 0$ the ordinary kriging system to the simple kriging system (6.14):

$$\begin{cases} \sum_{\beta=1}^N \lambda_{n,\beta} T_n(x_{\beta} - x_{\alpha}) + \mu_n = T_n(x_{\alpha} - x_0) & \alpha = 1, \dots, N \\ \sum_{\beta=1}^N \lambda_{n,\beta} = 1 \end{cases}$$

In this case one must beware that if the off-diagonal covariances on the left-hand side as well as those on the right-hand side tend to 0 when n increases, the estimator χ_n^* no longer tends to 0 but to the mean value of the $\chi_n(Y_{\alpha})$. Thus, in principle, it is necessary to compute all χ_n^* . Rivoirard (1994) proposes an elegant trick to limit the expansion to the same n_0 systems as in the

stationary case. By application of (6.16), we have to an excellent approximation

$$\psi^* \simeq \psi_0 + \sum_{n=1}^{n_0} \psi_n \chi_n^* + \sum_{n=n_0+1}^{\infty} \psi_n \hat{\chi}_n \quad \text{with} \quad \hat{\chi}_n = \frac{1}{N} \sum_{\alpha=1}^N \chi_n(Y_\alpha)$$

or equivalently

$$\psi^* \simeq \psi_0 + \sum_{n=1}^{\infty} \psi_n \hat{\chi}_n + \sum_{n=1}^{n_0} \psi_n (\chi_n^* - \hat{\chi}_n)$$

Now substituting $\hat{\chi}_n$ by its definition, we observe that the estimator ψ^* can be written as the finite sum

$$\psi^* \simeq \frac{1}{N} \sum_{\alpha=1}^N \psi(Y_\alpha) + \sum_{n=1}^{n_0} \psi_n (\chi_n^* - \hat{\chi}_n)$$

This estimator no longer depends on ψ_0 , which was our objective. However, the replacement of simple kriging of factors by ordinary kriging requires caution. Assuming that locally the $\chi_n(Y(x))$ are no longer with zero mean ($n > 0$) amounts to questioning the marginal distribution, thus the orthogonality of the factors and the bivariate distributions, and therefore the optimal character of disjunctive kriging.

6.3.5. Applications

The first implementations of disjunctive kriging were made in the scope of a bi-gaussian isofactorial model (Jackson and Maréchal, 1979; Young, 1982). In mining, predictions obtained by disjunctive kriging from exploration surveys have been compared with the reality of exploitation as known from numerous blast-holes. The following conclusions can be drawn:

- Concerning the estimation of the grade at a point, or the mean grade of a panel, disjunctive kriging brings little improvement over simple or ordinary kriging. This is confirmed by various case studies, including electrical conductivity data (Yates et al., 1986), contamination of underground water (Yates and Yates, 1988), and soil geochemistry (Webster and Oliver, 1989). Yates et al. observe, for example, a reduction of estimation variance of about 5% for point kriging and 10% for block kriging. Therefore disjunctive kriging is seldom used as a substitute for ordinary kriging to estimate the observed value Z itself.
- If we look for zones where Z has a chance greater than 50% to exceed a threshold z_c , it is better to estimate the indicator $1_{Z > z_c}$ by DK than select the zones where the kriging estimate of Z exceeds the threshold, especially

- if the distribution of Z is skewed (Webster and Oliver, 1989, for grades in agricultural soils; Wood et al., 1990, for physical properties of soils).
- Point-support recoverable reserves in a large panel corresponding to the zone of influence of a borehole can be predicted with an error of about 10%, which is considered as quite good compared to the results obtained using conventional methods (Jackson and Maréchal, 1979).
 - Disjunctive kriging and indicator cokriging are theoretically equivalent, up to discretization effects (number of factors in one case, number of thresholds in the other). Liao (1990) verified this empirically on a data set from a gold deposit by applying on the one hand indicator cokriging with fits of all direct and cross-covariances, and disjunctive kriging on the other, using a model fitted from the data. Let us mention that various studies have compared the performance of bi-gaussian disjunctive kriging and indicator kriging by trying them on real or synthetic data sets. Some have concluded that indicator kriging could give better results than bi-gaussian disjunctive kriging (e.g., Carr and Deng, 1987, for earthquake ground motion data). This should not be surprising, since none of these two methods is optimal. DK loses its optimality character if used with an arbitrary isofactorial model (in this case the bi-gaussian model), and indicator kriging is poorer than indicator cokriging. Since these two methods are not equivalent, either one may outperform the other in given circumstances (e.g., diffusive or mosaic model).
 - This shows the importance of an appropriate model selection. However, if the variable under study is poorly structured (strong nugget effect, short range compared with the data spacing) the choice of a model and a method (Gaussian or gamma transformation, disjunctive or indicator kriging) is of limited importance. Predictions will be poor anyhow and meaningful only after aggregation of a large number of point results.

Disjunctive kriging can be easily generalized to disjunctive cokriging if a consistent set of bivariate distributions can be modeled. It is the case, for example, when all bivariate distributions are Gaussian. Muge and Cabeçadas (1989) developed such a model to study the short- and medium-term evolution of the pollution of a lake by eutrophication. The model takes into account the time delay of some variables and allows the estimation of the “probability” to exceed a threshold value in a given period of time.

6.4. SIMPLE METHODS FOR ESTIMATING A BLOCK DISTRIBUTION

We now turn to change-of-support problems. These are always tricky problems, especially when only point-support data, or considered as such, are available which will be our working assumption here. We will adopt the terminology used in mining, which historically has been the most important field

of application. But selectivity problems may be posed in similar terms for the management of fish stock, agricultural land, or the remediation of contaminated areas. Before considering estimation problems, we will first define our objectives more clearly and investigate which (partial) answers linear geostatistics already makes available. This presentation summarizes results presented by Matheron (1984b) and Lantuéjoul (1990) to which the reader is referred for more details.

6.4.1. Selectivity

Let us first review a number of tools characterizing probability distributions and their dispersion, which will be quite useful to study support and information effects.

Selectivity Curves

Consider a nonnegative random variable Z representing, for definiteness, the grade of a block v selected uniformly among identical blocks discretizing a panel V . The distribution of block grades is defined by the cumulative distribution function

$$F(z) = E[1_{Z < z}]$$

Let us assume that for mining we only select blocks whose mean grade exceeds a cutoff value z . $F(z)$ represents the proportion of blocks regarded as waste (i.e., whose mean grade is below the cutoff grade). In practice, one is more interested in the blocks selected and in the mineral they contain, and the following auxiliary functions named *recovery functions* or *selectivity curves* are defined (Fig. 6.8):

- The *ore tonnage* at cutoff z , normalized by the total tonnage (i.e., with no cutoff)

$$T(z) = E[1_{Z \geq z}] = 1 - F(z)$$

- The *quantity of metal* residing in this ore, normalized by the total tonnage (of ore)

$$Q(z) = E[Z 1_{Z \geq z}] = \int_z^\infty u F(du)$$

- The *mean grade* of selected ore

$$\bar{m}(z) = E[Z | Z \geq z] = \frac{Q(z)}{T(z)}$$

- The *conventional income*

$$B(z) = E[(Z - z) 1_{Z \geq z}] = Q(z) - zT(z)$$

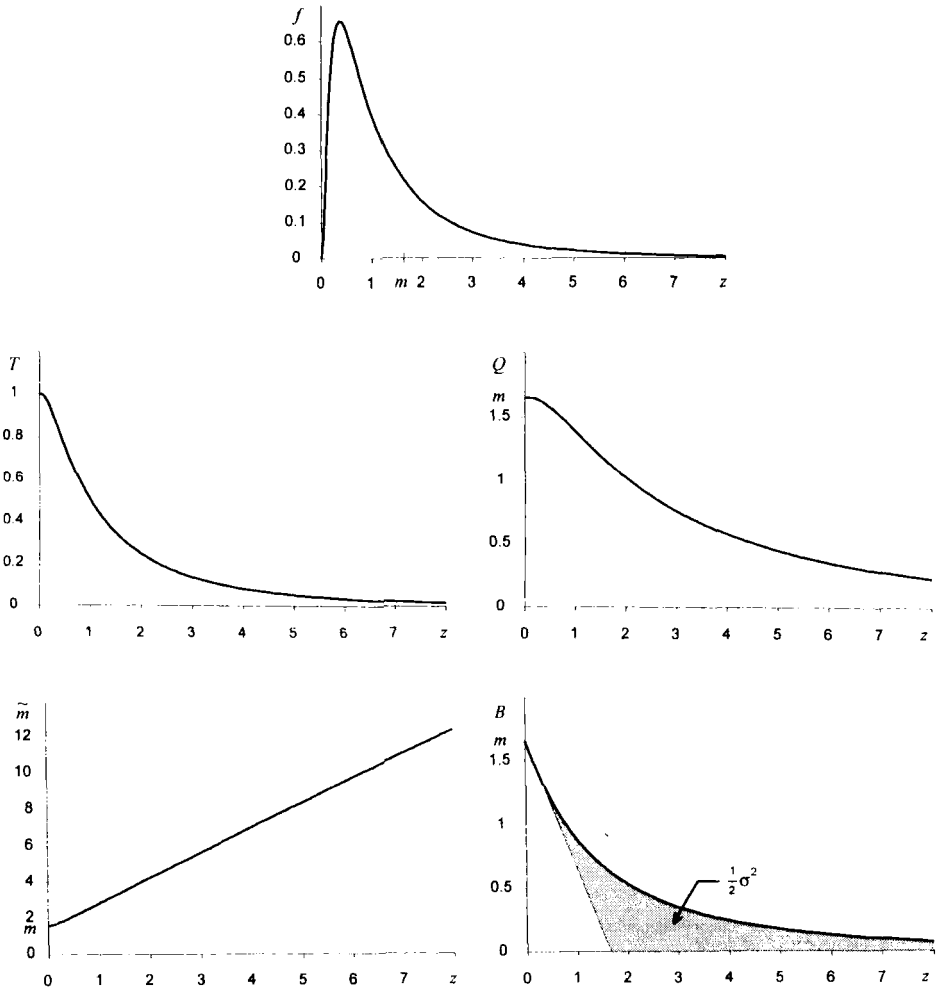


FIGURE 6.8. Probability density $f(z)$ and selectivity curves $T(z)$, $Q(z)$, $\tilde{m}(z)$, and $B(z)$. Example of a lognormal distribution with logarithmic mean 0 and logarithmic variance 1.

This last function has an economic significance. Let us assume that the selection of blocks above the cutoff grade z is entirely free; that is, we can leave in place any block with a mean grade below the cutoff and mine any block with grade above the cutoff. Let us also assume that the cutoff grade z is chosen so that the quantity of metal recovered from a block of grade z pays for its marginal mining and processing costs. In other words, z is the marginal cutoff; it can be considered as a variable in the sense that it depends on the selling price of the recovered metal and on available technology. Denoting by c the selling price of metal (per unit weight) and by d the density of ore (assumed to be constant and thus independent of grade, making tonnage equivalent to

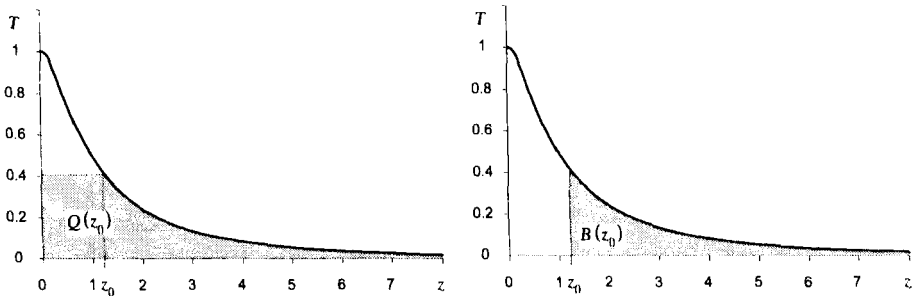


FIGURE 6.9. Graphical reading of $Q(z)$ and $B(z)$ from the graph of $T(\cdot)$.

volume), the marginal income from a block with mean grade Z is $cd|v|(Z - z)$ if Z is greater than or equal to z , and is zero otherwise (no stripping and removing of the block). Thus $B(z)$ represents, up to a normalization, the operating income before depreciation and fixed costs as a function of the marginal grade z . This function can also be written as

$$B(z) = \int_z^{\infty} T(u) du$$

The definitions of $Q(z)$ and $B(z)$ can be read on the graph of $T(\cdot)$, as shown in Figure 6.9: $Q(z)$ is the area of the domain under the tonnage curve defined by points with ordinates less than $T(z)$, whereas $B(z)$ is the area of the domain defined by points with abscissas greater than z .

We will present without proof the main properties of recovery functions. The functions T and Q each characterize the distribution F . They are not necessarily continuous, except if the distribution F has a density, and they are both nonincreasing. Their values at the origin are $T(0) = 1$ and $Q(0) = m$, mean of Z , and they vanish at infinity. The function \tilde{m} is not necessarily continuous and is nondecreasing. Its value at the origin is $\tilde{m}(0) = m$.

B is a more interesting function for it is convex, and therefore continuous on $]0, +\infty[$, and nonincreasing. It assumes the value m at zero and vanishes at infinity. Its slope at the origin is equal to $-(1 - p_0)$, where $p_0 = F(0+)$ represents the proportion of zero values. B also characterizes the distribution F .

T being a nonincreasing function of z , z is a nonincreasing function of T . It is thus possible to eliminate z and reparametrize Q , \tilde{m} , and B as functions of the ore tonnage T (Fig. 6.10). $Q(T)$ is a concave function, and thus continuous on $]0, m[$, and nondecreasing. $\tilde{m}(T)$ is a nonincreasing function. In particular, one has

$$\frac{dQ}{dT} = z(T) > 0 \quad \frac{d\tilde{m}}{dT} = -\frac{\tilde{m}(T) - z(T)}{T} = -\frac{B(T)}{T^2} \leq 0$$

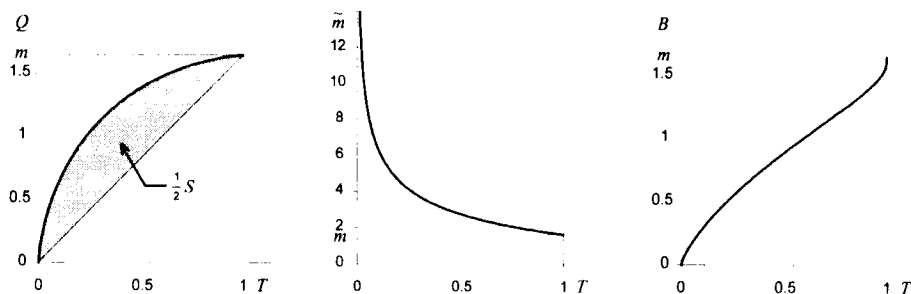


FIGURE 6.10. Selectivity curves $Q(T)$, $\tilde{m}(T)$ and $B(T)$. Example of a lognormal distribution with logarithmic mean 0 and logarithmic variance 1.

The functions $B(z)$ and $Q(T)$ are related by the duality formulas

$$B(z) = \sup_T \{Q(T) - zT\} \quad Q(T) = \inf_z \{B(z) + zT\}$$

As a consequence the function $Q(T)$, just as the function $B(z)$, characterizes the distribution F . Similar concepts are used in econometrics to study the concentration of income in the population (the fraction $T(x)$ of the people with income greater than x earns a fraction $Q(x)$ of the global income; e.g., Saporta, 1990).

Interpretation in Terms of Geochemistry

Recovery functions were introduced by Lasky (1950) to represent grade-tonnage relationships for porphyry copper deposits. Lasky used a law of the form $\tilde{m}(T) = \alpha(1 - \log T)$. It corresponds to an exponential distribution of grades, where $T(z) = \exp(-z/\alpha)$ and $\tilde{m}(z) = z + \alpha$. In other words, to get ore with mean grade above m_1 , it suffices to mine at the cutoff grade $z = m_1 - \alpha$.

Recovery functions have been used in geochemistry at the global scale to model grades in the earth's crust. Ahrens (1954), on the basis of numerous examples, stated a "law" according to which the distribution of grades is lognormal. We have seen with the de Wisjjan model that this type of distribution can be obtained by a process of enrichment of half of the ore and depletion of the other half, the phenomenon being repeated for each half over and over; in the language of fractals, this is a multifractal model (Cheng and Agterberg, 1996). Using a statistical approach Vistelius (1960) reached much more circumstantiated conclusions, interpreting some long-tailed distributions as mixtures of Gaussians. Unsurprisingly, Ahrens' model has inspired a number of generalizations (e.g., de Wijs, 1976). Turcotte (1986), working in the scope of fractal models, claimed that grades follow a Pareto distribution (power law) obtained by Rayleigh distillation where successive crystallizations enrich the still liquid fraction of the magma. Reexamining these different points of view in the light of geochemistry, Allègre and Lewin (1995) conjectured that distributions of fractal or multifractal type result from geochemical differentiation processes, whereas the other distributions can be explained by geochemical mixing processes. Nature is of course more complex than these conceptual models would suggest. While such models are important for the understanding of the origin of deposits and thus mineral exploration, in geo-statistics our objective is more modestly to fit the distribution actually observed in the area of interest.

Dispersion Indicator and Selectivity Index

The *dispersion indicator* of the distribution F of a random variable Z with values in \mathbb{R} (not necessarily positive) is defined by

$$S = \frac{1}{2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(dz) |z' - z| F(dz') = \int_{-\infty}^{+\infty} F(z)(1 - F(z)) dz$$

This can be compared with the variance

$$\sigma^2 = \frac{1}{2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(dz)(z' - z)^2 F(dz') = \int_{-\infty}^{+\infty} (z - m)^2 F(dz)$$

where $m = \int_{-\infty}^{+\infty} zF(dz)$ is the mean. S and σ^2 both summarize the dispersion of the distribution F , but S is more robust than σ^2 : up to the factor $\frac{1}{2}$, S is the mean of the magnitude of $Z' - Z$ when Z and Z' are two independent random variables with the same distribution F , while σ^2 is the mean square value of $Z' - Z$. The dispersion indicator S has a finite value if and only if the mean m exists (i.e., $E[|Z|] < \infty$), whereas this condition does not guarantee that the variance σ^2 is finite. It can be shown that S and σ are related by the inequality $S \leq \sigma/\sqrt{3}$. Equality is obtained when F is uniform over a bounded interval. For a Gaussian distribution $S = \sigma/\sqrt{\pi}$, a value close to the maximum. The *normality index* $\nu = \sqrt{\pi}S/\sigma$ can also be defined, and it is a dimensionless parameter with value 1 for a Gaussian distribution. S is related to the interquantile distance associated with p and $1 - p$ by the following inequality which justifies the name dispersion indicator given to S :

$$z_{1-p} - z_p \leq \frac{S}{p(1-p)} \quad \left(p < \frac{1}{2} \right)$$

In order to include the case of a distribution with discontinuities, it is understood that z_p is the smallest value such that $F(z_p) \geq p$ and z_{1-p} the largest value such that $F(z_{1-p}) \leq 1 - p$.

When Z is a nonnegative variable, which we assume from now on, it is shown that S is strictly less than m . The *selectivity index* $s = S/m$ can be defined and belongs to the interval $[0, 1[$. In the case of a lognormal distribution, for example, this index is related to the logarithmic standard deviation σ' by $s = 2G(\sigma'/\sqrt{2}) - 1$ and can take on any value in $[0, 1[$. The selectivity index s is more robust than the coefficient of variation σ/m , which can be infinite. In applications, however, the second-order moments need to be known—and therefore the coefficient of variation—since the variance is the only parameter whose evolution under a change of support can be predicted exactly. In this perspective Matheron (1985) examines, for a number of distributions, to which extent the coefficient of variation can be related to the selectivity index.

The interest of the dispersion indicator and the selectivity index, aside from their robustness, appears clearly when one considers the graph $Q(T)$. Selectivity is maximal if one can recover all the metal while mining only a minute fraction of the ore. This would be the case, for example, if we could pick the nuggets in a gold deposit. The graph of $Q(T)$ would reach its upper limit m already for a very low value of T . Conversely, no selectivity is possible if the ore has a uniform grade equal to the mean grade m . The function $Q(T)$ then coincides with the straight line of equation $Q = mT$. Generally, it is shown that the area between this line and the graph of $Q(T)$ is equal to $S/2$ (Fig. 6.10):

$$\int_0^1 (Q(T) - mT) dT = \frac{1}{2} S$$

The variance satisfies a similar property, but its interpretation in terms of selectivity is not so clear (Fig. 6.8): the area between the graph of $B(z)$ and the straight line $B = m - z$ (the tangent at the origin if there is no zero effect) is equal to $\sigma^2/2$:

$$\int_0^\infty (B(z) - (m - z)1_{z \leq m}) dz = \frac{1}{2} \sigma^2$$

Support Effect

Let us now consider a RF $Z(x)$, for example, representing a grade and the grades of blocks with support v . The distribution of the block grades depends of course on their size, the general tendency being that the dispersion decreases as the size of the support increases: very low as well as very high grades can be observed on cores, whereas the mean grades of large panels have a low contrast (e.g., see Fig. 6.11). At the same time selective mining is all the more difficult as the support v of the selection unit is large. In the limit, when v is very large, all blocks have grades very close to the mean m of Z , and no real selectivity is possible. From a probabilistic viewpoint these facts can be explained by the two following remarks:

1. The block mean grade (no cutoff) is independent of block size (it is the mean m of $Z(\cdot)$).
2. The block histogram gets narrower around the mean as the support gets larger, according to the variance (2.33), and it tends to become Gaussian (central limit theorem) provided that the covariance of Z has a finite integral range.

To improve the comparison of the two distributions, let us introduce a definition and a theorem.

Definition. If Z_1 and Z_2 have distributions F_1 and F_2 with the same mean m , we say that the distribution F_1 is more selective than F_2 if the associated recovery functions satisfy $Q_1(T) \geq Q_2(T)$ for any $T \in [0, 1]$, or $B_1(z) \geq B_2(z)$

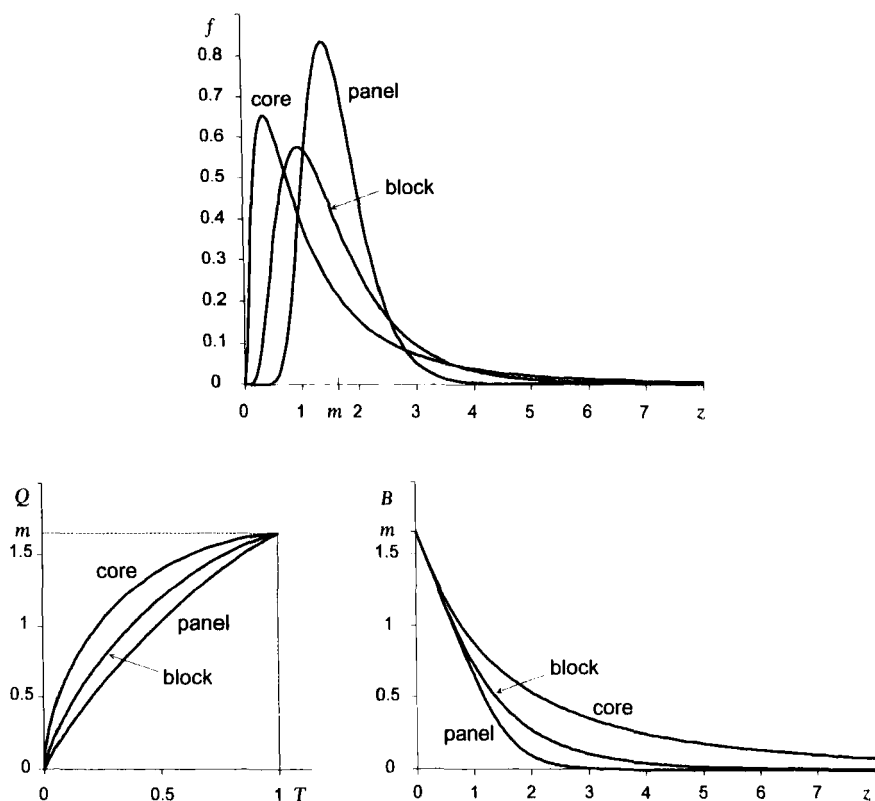


FIGURE 6.11. Evolution of the probability density $f(z)$ and the selectivity curves $Q(T)$ and $B(z)$ with the size of the support. From the cores to the blocks, and again from the blocks to the panels, the variance is reduced by a factor of 4.

for any $z \geq 0$ (the two definitions are equivalent given the duality between $B(z)$ and $Q(T)$).

To know whether a given distribution is more selective than another, we can use a general functional analysis theorem established by Cartier (see Alfsen, 1971, sec. 3):

Theorem. The c.d.f. F_1 is more selective than the c.d.f. F_2 if and only if there exists a bivariate distribution $F_{12}(dz_1, dz_2)$ with marginals F_1 and F_2 and such that

$$E[Z_1 | Z_2] = Z_2 \quad (6.30)$$

Relation (6.30) is called *Cartier's relation*. By application of (1.5) this implies that

$$\text{Var}[Z_1] = \text{Var}[Z_2] + \text{Var}[Z_1 - Z_2] \quad (6.31)$$

Coming back to the support effect, if \underline{x} is a uniform random point in a block v with grade Z_v , it is easy to show that

$$E[Z(\underline{x}) | Z_v] = Z_v \quad (6.32)$$

Likewise, if \underline{v} is a block randomly selected among blocks with support v forming a partition of a panel V with mean grade Z_V , we have

$$E[Z(\underline{v}) | Z_V] = Z_V$$

By Cartier's theorem the distribution F_V of panels is less selective than the distribution F_v of blocks, which is itself less selective than the distribution F of cores (considered as points). This phenomenon, which is fundamental in geostatistics, is known as the support effect. Selectivity relations between the distributions F_V , F_v , and F also entail that the dispersion indicator and the selectivity index decrease as the support increases (at least for nested supports).

Information Effect

In selective mining as elsewhere people have to make decisions based on incomplete information. When only point-support data or regarded as such are at hand (core samples), the decision to accept or reject a block v must be based on an estimated value since the block true grade is unknown. The best one can do is consider the conditional expectation $Z_v^* = E[Z_v | Z_\alpha : \alpha = 1, \dots, N]$. It is a conditionally unbiased estimator

$$E[Z_v | Z_v^*] = Z_v^*$$

It follows from Cartier's theorem that the distribution of Z_v^* is less selective than that of Z_v . In particular, by application of (6.31), estimated grades are less dispersed (smoother) than actual grades: for a high cutoff grade there will be far fewer blocks accepted on the basis of Z_v^* than on Z_v . Even if one considers a cutoff for which the bias on the ore tonnage is negligible, there will be a loss due to bad selection:

- Poor blocks are selected because estimated rich.
- Rich blocks are rejected because estimated poor.

In all cases this translates into a degradation of the value of the exploited ore, which increases with the dispersion of the scatterplot (Z_v, Z_v^*) around the first bisector. This degradation is called the *information effect*: indirect selection, that is, based on samples rather than the true block value, always results in a degradation of the conventional income.

In our reasoning we have overlooked a difficulty that deserves further explanation. Even though the selection is made on the basis of the estimated

value Z_v^* , the grade of a mined block is the true grade Z_v and not Z_v^* ("We do not send estimated values to the mill," David, 1977). Therefore the ore tonnage and the quantity of metal of the selected blocks are

$$T(z) = E[1_{Z^*(v) \geq z}] \quad Q(z) = E[Z(v)1_{Z^*(v) \geq z}]$$

However, when the estimator Z_v^* is the conditional expectation, we have that

$$E[Z^*(v) | Z^*(v) \geq z] = E[Z(v) | Z^*(v) \geq z] = Q(z)$$

This is an extremely desirable property: the true mean grade of the selected blocks is equal to the mean grade of their estimated values (in expectation). This justifies the calculation of selectivity curves directly from the distribution of Z_v^* . This simplification of course is no longer legitimate if we use an estimator that is not conditionally unbiased.

Selection on Kriged Grades

In practice, we cannot find the conditional expectation, and we do the selection on simpler estimators, such as the grade of a sample in the block or a kriging estimate from samples in the block and its surrounding. But kriging only guarantees unbiasedness, not conditional unbiasedness. This results in an additional loss. For example, if the regression $E[Z_v | Z_v^*]$ is linear, we have

$$E[Z_v | Z_v^*] = m + r[Z_v^* - m]$$

(i.e., actual increment = $r \times$ expected increment in view of the selection) where

$$r = \rho \frac{\sigma_v}{\sigma_v^*} \quad \sigma_v^2 = \text{Var}[Z_v] \quad \sigma_v^{*2} = \text{Var}[Z_v^*] \quad \rho = \text{Corr}(Z_v, Z_v^*)$$

If Z_v^* is simply the grade of a sample taken in the block v , one has $\sigma_v^* > \sigma_v$, and thus $r < 1$. For $Z_v^* > m$ one tends to select an ore that is less rich than expected. At the other extreme, for very scattered grades and a small support v , when each panel is kriged with a very large neighborhood, it is possible to get $\sigma_v^* \ll \sigma_v$ and $r > 1$ and reach the wrong conclusion that the deposit is uneconomic.

To evaluate the quantity of metal that is actually recovered, it is convenient to introduce the regression between true and estimated grades:

$$h(z) = E[Z_v | Z_v^* = z]$$

The quantity of metal is then

$$Q(z) = E[Z_v 1_{Z_v^* \geq z}] = E\{E[Z_v 1_{Z_v^* \geq z} | Z_v^*]\} = E[h(Z_v^*) 1_{Z_v^* \geq z}]$$

If the function h is increasing, which is a reasonable assumption, the tonnage and the quantity of metal can be written as

$$T(z) = E[1_{h(Z_v^*) \geq h(z)}] \quad Q(z) = E[h(Z_v^*) 1_{h(Z_v^*) \geq h(z)}]$$

Therefore

$$T(z) = \tilde{T}(h(z)) \quad Q(z) = \tilde{Q}(h(z))$$

where $\tilde{T}(z)$ and $\tilde{Q}(z)$ stand for the recovery functions associated with $h(Z_v^*)$. It can be shown that $Q(T) = \tilde{Q}(T)$. In other words, to quote a figurative expression of G. Matheron, "things are as though we were not mining the true grades but their conditional expectations."

The actual calculation of the information effect can be done simply in the scope of an extension of the discrete Gaussian model presented below (cf. Matheron, 1976b; Lantuéjoul, 1990).

A Comparison of Support and Information Effects

To compare these two effects, Matheron (1984b) presents a simple example that can be solved explicitly: the decision to mine the block v is made in view of the grade of a single sample, and the bivariate distribution of the block and the sample is lognormal. We denote by $\sigma_v'^2$ the logarithmic variance of the block grade Z_v , by σ'^2 the logarithmic variance of the sample grade Z , and by m the common arithmetic mean of Z and Z_v . Cartier's relation $E[Z | Z_v] = Z_v$ entails that the correlation coefficient of $\log Z$ and $\log Z_v$ is $\rho' = \sigma_v' / \sigma'$. The conditional expectation $H = E[Z_v | Z]$ is also lognormal with arithmetic mean m and logarithmic variance $\sigma_H'^2 = \rho'^2 \sigma_v'^2 = \sigma_v'^4 / \sigma'^2$. The different selectivity curves can then be calculated easily. For $m = 1$, $\sigma'^2 = 1$, and $\sigma_v'^2 = \frac{1}{2}$, for example, the selectivity index goes from 0.521 for point-support grades to 0.383 for actual block grades (degradation due to support effect) and 0.276 for blocks selected on the basis of a central sample (additional degradation due to the information effect). The support effect and the information effect can therefore be of comparable importance.

To summarize, one should not base the selection on single samples from the blocks and calculate the value of the deposit as if the grades of the selected blocks were equal to those of their samples—the implicit methodology behind an evaluation of recoverable reserves by polygons of influence. Indexing the recovery functions by

- "pol" for an evaluation by polygons of influence which would identify the distribution of block grades with the distribution of the samples,
- "ide" for the ideal case of a selection based on true block grades,
- "opt" for a selection based on conditional expectation,
- "sub" for the actual selection based on suboptimal estimates,

we obtain the following ranking of conventional incomes:

$$B_{\text{sub}}(z) \leq B_{\text{opt}}(z) \leq B_{\text{ide}}(z) \leq \hat{B}_{\text{pol}}(z) \quad \forall z$$

The hat over \hat{B}_{pol} recalls that it is a biased estimate (by excess) rather than an evaluation susceptible to be actually reached. The loss from B_{opt} to B_{sub} is due to the use of a suboptimal selection criterion; it could be reduced. The difference between B_{ide} and B_{opt} is due to the information effect, and the difference between \hat{B}_{pol} and B_{ide} represents the support effect; these can only be reduced at the cost of new data and a reduction of the volume v of the selection unit. The loss $B_{\text{sub}} - \hat{B}_{\text{pol}}$ shows how illusory our expected income is when it is calculated with polygons of influence and the histogram of the blocks is confused with that of the samples.

Some authors have introduced empirical or theoretical corrections to transform kriging estimators into conditionally unbiased estimators. David et al. (1984) propose to combine a local simple kriging estimator with a global estimator of the mean rather than use a local ordinary kriging estimator, and Guertin (1987) proposes a solution combining a classic approach and a modeling of bivariate distributions.

6.4.2. Global Change-of-Support Models

General Principles

Consider an SRF $Z(x)$, with marginal $F(dz)$. We wish to determine the global recovery functions associated with the support v . It will suffice to determine the marginal distribution $F_v(z)$ of the regularized SRF $Z_v(x)$, or of any other recovery function characterizing it. The other functions will be deduced from this.

In some cases we have data with support v (e.g., mean value of mined panels), but here we will assume that only point-support data are available (in mining, core samples). We cannot determine the distribution of block grades by nonparametric methods and have to consider change-of-support models. Indeed the marginal distribution of $Z_v(x)$ depends on the complete spatial distribution of the SRF $Z(x)$, and not only on its marginal. In practice, the rigorous calculation of this marginal distribution is only possible in a few special cases, and we have to resort to approximations. The general approach is to seek a transformation from the distribution of the point-support $Z(x)$ to the distribution of block grades $Z_v(x)$. Since no data with support v are available, we have no direct experimental check of the results obtained. Fortunately a few guidelines are available. As we have seen in the preceding section, the distribution of block grades $Z_v(x)$ must satisfy a number of constraints:

1. The distributions $F_v(\cdot)$ and $F(\cdot)$ must have the same mean m .

2. The variance of the distribution $F_v(\cdot)$ must be equal to the variance given by (2.33),

$$\sigma_v^2 = \frac{1}{|v|^2} \int_v \int_v C(x' - x) dx dx' \quad (6.33)$$

where $C(h)$ is the covariance of the SRF $Z(x)$.

3. The distribution $F_v(\cdot)$ must be less selective than $F(\cdot)$.

Four very simple change-of-support models are presented below: affine correction, discrete Gaussian model, permanence of lognormality (a special case of the preceding one), and mosaic model. These models satisfy the above conditions (we will say that they are *consistent*), which does not mean that they are always suitable. We will review in Sections 6.5.1 and 6.5.2 other models that proceed from the modeling of bivariate distributions and in Section 6.4.3 the multi-gaussian model (after transformation).

It should be noted that in all cases the change-of-support models give “exact” recoverable reserves, in the sense that there is no associated confidence interval. This apparent exactness only reflects the choice of a change-of-support model. The results are conditional on this choice and must be taken with a critical eye. Any validation, even if partial, is useful; for example, one can construct composite samples (group metric cores into 10-m composites) and verify that their histogram is correctly predicted from the point-support histogram (core samples).

Affine Correction

If $Z(x)$ is a Gaussian SRF, with mean m , variance σ^2 and covariance $C(h)$, then $Z_v(x)$ is Gaussian, with the same mean m , and variance σ_v^2 given by (6.33). Therefore the following equalities hold in distribution:

$$\frac{Z(x) - m}{\sigma} \stackrel{D}{=} \frac{Z_v(x) - m}{\sigma_v} \sim \mathcal{N}(0, 1)$$

Hence

$$Z_v(x) \stackrel{D}{=} m + \frac{\sigma_v}{\sigma} (Z(x) - m) \quad (6.34)$$

The affine correction is to consider that even if the distribution of Z is not Gaussian, the relation (6.34) remains valid, at least to a first approximation. The estimate \hat{F}_v of the block distribution can be derived from the estimate \hat{F} of the point distribution by

$$\hat{F}_v(z) = \hat{F}\left(m + \frac{\sigma_v}{\sigma}(z - m)\right) \quad (6.35)$$

This model has the advantage of simplicity: \hat{F}_v is immediately deduced from the histogram of samples without mathematical modeling. This change-

of-support model is consistent but only valid, besides the Gaussian case, if the distribution of Z is continuous and if v is small compared to the range (in particular, the variogram of Z should not have a nugget effect). Indeed:

1. It assumes a permanence of the distribution (F and F_v have the same type of distribution, only the scale parameters change). This permanence is only true in the Gaussian case. We also know that if v becomes far larger than the range, F_v must tend to a Gaussian distribution by virtue of the central limit theorem, which is not the case in this model (except again if the SRF is Gaussian).
2. The interval of definition of Z_v narrows around the mean in a ratio σ_v/σ . If the distribution of Z has an atom at 0, the distribution of Z_v will have one at $m(1 - \sigma_v/\sigma)$, which makes no physical sense.

Discrete Gaussian Model

Consider a block v and a uniform random point \underline{x} within v . The distribution of the random variable $Z(\underline{x})$ is the marginal distribution F of the SRF $Z(\cdot)$. We have seen before that if it is not Gaussian, we can at least consider that $Z(\underline{x})$ is the transform of a Gaussian of the form $Z(\underline{x}) = \varphi(Y)$ where Y is a standard normal and $\varphi = F^{-1} \circ G$ (G being the standard Gaussian c.d.f.). Similarly we can consider that the mean grade Z_v of the block v is of the form $Z_v = \varphi_v(Y_v)$ where Y_v is a standard normal and φ_v the block transformation function, which we want to determine. The crucial assumption of the discrete Gaussian model is that the bivariate distribution of the (Y, Y_v) pair is also Gaussian. It is characterized by a correlation coefficient r assumed positive. The block transformation function and its distribution are then derived.

Cartier's relation (6.32) is equivalent to

$$E[\varphi(Y) | Y_v] = \varphi_v(Y_v) \quad (6.36)$$

Since the (Y, Y_v) pair is Gaussian, the conditional distribution of Y given $Y_v = y_v$ is Gaussian with mean ry_v and variance $1 - r^2$. Hence

$$\varphi_v(y_v) = \int \varphi(ry_v + \sqrt{1-r^2}u)g(u)du \quad (6.37)$$

The correlation coefficient r is selected so that the distribution defined by φ_v has the variance given by (6.33). The variance of $\varphi_v(Y_v)$, taken as a function of r , increases from 0 for $r = 0$ to σ^2 for $r = 1$. The value of σ_v^2 defined by (6.33) thus uniquely determines r .

It is interesting to note that (6.37) has the form of a convolution product

$$\varphi_v(y_v) = \varphi * g_{1-r^2}(ry_v)$$

where g_{1-r^2} is the p.d.f. of a zero-mean normal with variance $1 - r^2$. There are remarkable families of functions φ for which this convolution can be expressed in a simple and exact manner. An example is the exponential transformation (see hereafter). In other cases this convolution can be computed numerically by running a weighted moving average on the empirical function φ . The block transformation function can also be calculated from the expansions of φ and φ_v into Hermite polynomials (see Section 6.5.1).

The discrete Gaussian model is by construction a consistent change-of-support model. Moreover it is compatible with the central limit theorem. It does not do well in the presence of a significant zero effect, but there are variants that will be presented in Sections 6.5.1 and 6.5.2 with isofactorial change-of-support models.

Permanence of Lognormality

When $Z(x)$ is a lognormal SRF the transformation function is

$$\varphi(y) = \exp(m' + \sigma' y)$$

where m' and σ'^2 are the mean and variance of $\log Z$. As the mean of Z is invariant in any change of support, it is interesting to take as main parameters the mean m of Z and its logarithmic variance σ^2 . By application of formula (2.79), the transformation function becomes

$$\varphi(y) = m \exp(\sigma' y - \frac{1}{2} \sigma'^2) \quad (6.38)$$

From (6.37) the block transformation function is then

$$\varphi_r(y) = m \exp(r \sigma' y - \frac{1}{2} r^2 \sigma'^2)$$

The distribution of Z_v is still lognormal: there is *permanence* of lognormality. This permanence cannot be strictly true: the average of two lognormal random variables is not a lognormal random variable. However, there is good empirical evidence of the permanence of lognormality, to a first approximation, as long as v remains small (for definiteness, as long as the variogram of $\log Z(x)$ between two points of v remains less than 1; e.g., see Taylor, 1982).

From a practical standpoint the correlation coefficient $r > 0$ is defined by

$$r^2 = \frac{\log(1 + \sigma_v^2/m^2)}{\log(1 + \sigma^2/m^2)}$$

The variance σ^2 of Z is given by (2.79) and the variance σ_v^2 of Z_v is calculated by applying formula (6.33) with the covariance of Z given by the third relation (2.79).

Mosaic Model

Let us close with a model especially suited for representing a zero effect (Mathéron, 1984d). This is obtained by change of support from a mosaic random function. Since the point distribution is more selective than the distribution of Z_v , which is itself more selective than the distribution associated with an

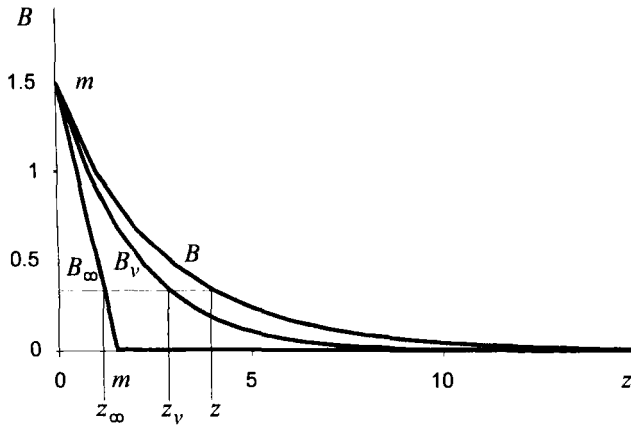


FIGURE 6.12. Definition of a mosaic change-of-support model: graph of the conventional income. The point distribution is negative binomial with $\alpha = 0.5$ and $p = 0.75$. The change-of-support parameter is $\beta = 0.4$.

infinite support, the corresponding conventional incomes satisfy

$$B_{\infty}(z) \leq B_v(z) \leq B(z) \quad \text{with} \quad B_{\infty}(z) = \max(m - z, 0)$$

that is, the graph of B_v is comprised between the graphs of B and B_{∞} . The mosaic change-of-support model can be defined by a method of construction of B_v from B and B_{∞} (Fig. 6.12): the B_v graph is assumed to be an horizontal weighted average of the B and B_{∞} graphs. More specifically, given the parameter β , $0 < \beta < 1$, let

$$z_v = (1 - \beta)z + \beta[m - B(z)]$$

and

$$B_v(z_v) = B(z)$$

For the relation (6.33) to hold, it is necessary to take $\beta = 1 - \sigma_v^2/\sigma^2$. Then the change-of-support model is consistent. This model captures the zero effect: if the point distribution comprises a proportion p_0 of zero values, the distribution F_v comprises the proportion

$$p_0^v = \frac{(1 - \beta)p_0}{1 - \beta p_0}$$

which is smaller than p_0 and decreases as the size of the support becomes larger. The mosaic model, however, is not compatible with the central limit theorem for large supports.

6.4.3. Local Estimation of Block Distribution

We now focus our interest on a precise block v to which we ascribe the index 0, and seek the distribution F_{v_0} of the grade Z_{v_0} of this block conditioned on the data $\{Z_\alpha = Z(x_\alpha) : \alpha = 1, \dots, N\}$ (in practice, only a subset of the data, inside the block or close to it are taken into account).

Multi-Gaussian Model

The multi-gaussian model, in which $Z(x)$ is assumed related to a Gaussian SRF $Y(x)$ by the transformation $\varphi = F^{-1} \circ G$, can be used if we approximate the block grade by the mean grade of M points x_i discretizing the block. Then

$$\begin{aligned} F_{v_0}(z) &= \Pr \left\{ \frac{1}{M} \sum_{i=1}^M Z(x_i) < z \mid Z(x_\alpha) : \alpha = 1, \dots, N \right\} \\ &= \Pr \left\{ \sum_{i=1}^M \varphi(Y(x_i)) < Mz \mid Y(x_\alpha) : \alpha = 1, \dots, N \right\} \end{aligned} \quad (6.39)$$

The SRF $Y(x)$ being assumed Gaussian, the vector of conditional expectations of the $Y(x_i)$ given the $Y(x_\alpha)$ is Gaussian and is obtained by linear regression. Its distribution is entirely determined by the vector of conditional means and the conditional variance-covariance matrix, which are obtained as by-products of these regressions.

The distribution (6.39) generally cannot be calculated analytically. It is calculated by the Monte Carlo method: we draw a large number of vectors of $Y(x_i)$ from the conditional distribution and evaluate $F_{v_0}(z)$ numerically as the proportion of simulated vectors satisfying the condition $\sum_{i=1}^M \varphi(Y(x_i)) < Mz$.

This method has been proposed by and applied Verly (1983, 1984). Just like for point-support data, it is advisable, before using this method, to verify at least that the data are compatible with the hypothesis of stationarity of the RF $Y(x)$ and that the bivariate distributions of this SRF can be considered Gaussian. Verly (1986) gives details on an application to a uranium deposit; predictions made from limited information are validated by the real grades of blocks (approximated by the mean of blast holes).

It should be noted that the nonconditional version of this model constitutes a global change-of-support model. It relies on stronger hypotheses than the discrete Gaussian model, though it is not clear whether they are really far more restrictive. Though rigorous, up to the discretization of the block and the computation by the Monte Carlo method, the multi-gaussian model is not much used because the discrete Gaussian model is much easier to use (let us remind, however, that the latter is not valid in case of an important zero effect).

Permanence of Lognormality

If $Z(x)$ is an SRF with a lognormal spatial distribution, namely if $Y(x) = \log Z(x)$ is a Gaussian SRF, the vector of transformed data $Y(x_\alpha)$, $\alpha = 1, \dots, N$, is Gaussian. Under the simplifying assumption of permanence of lognormality, Z_{v_0} is a lognormal random variable, or equivalently, $Y_{v_0} = \log Z_{v_0}$ has a Gaussian distribution. We can determine the conditional distribution of blocks by going one step further and assuming that the joint distribution of $Y_{v_0}, Y(x_1), \dots, Y(x_N)$ is multivariate Gaussian. The conditional distribution of Y_{v_0} is then Gaussian; its mean is the simple kriging estimate Y_{SK}^* from the data $Y(x_\alpha)$, and its variance is the simple kriging variance. When carrying out this (co)kriging, it should be reminded that Y_{v_0} and $Y(x)$ do not have the same mean: if m' is the mean of $Y(x)$, the mean of Y_{v_0} is

$$m'_v = m' + \frac{1}{2}(1 - r^2)\sigma'^2 \quad (6.40)$$

To implement these calculations, we need, in addition to the knowledge of the point covariance of $Y(x)$, the variance of Y_{v_0} and the covariances of Y_{v_0} and $Y(x_\alpha)$. These result from relation (2.79) between the covariance of a lognormal SRF and the covariance of its logarithm

$$\begin{aligned} \text{Var}[Y_{v_0}] &= \log \left(1 + \frac{\text{Var}[Z_{v_0}]}{m^2} \right) = \log \left(1 + \frac{1}{m^2 |v_0|^2} \int_{v_0} \int_{v_0} C_Z(x' - x) dx dx' \right) \\ &= \log \left(\frac{1}{|v_0|^2} \int_{v_0} \int_{v_0} e^{C_Y(x' - x)} dx dx' \right) \\ \text{Cov}(Y_{v_0}, Y(x)) &= \log \left(1 + \frac{\text{Cov}(Z_{v_0}, Z(x))}{m^2} \right) = \log \left(1 + \frac{1}{m^2 |v_0|} \int_{v_0} C_Z(x' - x) dx' \right) \\ &= \log \left(\frac{1}{|v_0|} \int_{v_0} e^{C_Y(x' - x)} dx' \right) \end{aligned}$$

If strict stationarity of $Y(x)$ is not assured, an approximate solution can be obtained, like in the point-support case, by substituting the ordinary lognormal (co)kriging estimate of Z_{v_0} for its simple lognormal kriging estimate. When performing the ordinary (co)kriging of Y_{v_0} (which is the basis of the lognormal estimator), the relation (6.40) between the (local) means of the RFs $Y(x)$ and $Y_{v_0}(x)$ must be taken into account so that m' only is considered unknown. Details of this method can be found in Journel (1980). This procedure must be used with caution because it lacks robustness if data depart from a lognormal distribution.

Parker et al. (1979) present an application to a uranium deposit, including a cross-validation of local point-support recovery functions. The uranium grade is not strictly stationary; the authors have thus used ordinary lognormal kriging and have carefully calibrated the sill of the variogram of $Y(x)$ (using a local calibration). Rendu (1979) presents a validation of lognormal kriging and of the calculation of the conditional distribution of blocks for a gold deposit in South Africa. The study shows an excellent agreement between the results predicted by geostatistics and reality. The mean being well known, the results obtained by simple kriging of Y_{v_0} are better than those obtained by ordinary kriging. However, forcing the lognormal formalism to data whose histogram is spread out but not lognormal can lead to real disasters (David et al., 1984).

Indicator Kriging and Local Affine Correction

If we cannot turn to the Gaussian case, the calculation of conditional expectation is generally impossible. So is kriging, or rather cokriging of the indicator

$1_{Z(v_0) < z}$ from indicator data $1_{Z_{x_0} < z}$, since the v_0 -support indicator is not the mean in v_0 of point-support indicators. A possible solution is to apply a local affine correction to the distribution defined by kriging a series of indicators (for different cutoffs) of the point variable $Z(x_0)$, where x_0 is the center of the block v_0 . As this local distribution has for mean the kriging estimate of $Z(x_0)$, (6.35) writes

$$\hat{F}_{v_0}(z) = \hat{F}_{x_0}(Z_{x_0}^* + r(z - Z_{x_0}^*))$$

where r^2 is the ratio of the conditional variance for support v_0 to the variance of the conditional point-support distribution at x_0 . This ratio is unknown, but we know the theoretical variances of $Z(x_0)$ and Z_{v_0} . Journel (1984) proposes to approximate the ratio of conditional variances by the ratio of theoretical variances. As for practical implementation, we note that the point cutoff grade to be associated with the cutoff grade z for the support v depends on the local kriging of $Z(x_0)$ and thus differs from block to block.

This method rests on a double approximation, the approximation to the conditional point-support distribution by indicator kriging estimates, and the approximation to the conditional block distribution by an affine correction. A weakness is that the local recovery functions obtained give a different overall result than global recovery functions. But, above all, this method very often produces results that are simply unacceptable (Liao, 1990; Rossi and Parker, 1994). Apart from exceptions, nonparametric methods such as indicator kriging cannot capture a change of support. This requires explicit modeling of bivariate distributions and the use of disjunctive kriging.

6.5. LOCAL ESTIMATION OF A BLOCK DISTRIBUTION BY DISJUNCTIVE KRIGING

Global estimation of the distribution of blocks amounts to the determination of the block transformation function. Local estimation of the block distribution, however, calls for a more completely specified model: in addition to bivariate distributions between points, we need the bivariate distribution between a point and a block and possibly between two blocks. Matheron developed a large number of possible isofactorial models. Though highly interesting, they will not be presented here for lack of space. We will focus on models with Gaussian marginals (up to a transformation) and will give a glimpse of other available models.

6.5.1. Gaussian Isofactorial Change-of-Support Models

Hermitian Change-of-Support Model

Consider for a start the simple case of point-support grades of the form $Z(x) = \varphi(Y(x))$ where $Y(x)$ is an SRF with standard Gaussian marginal and Hermitian

bivariate distributions. The transformation function φ can be expanded by (6.19) into normalized Hermite polynomials χ_n as $\varphi(y) = \sum_{n=0}^{\infty} \varphi_n \chi_n(y)$, and the bivariate p.d.f. of $Y(x)$ and $Y(x')$ is of the form

$$g_{xx'}(y, y') = \sum_{n=0}^{\infty} T_n(x, x') \chi_n(y) \chi_n(y') g(y) g(y') \quad (6.41)$$

where $T_n(x, x')$ is the covariance of the factor $\chi_n(Y(\cdot))$. This covariance in fact only depends on the vector $x' - x$, but a more explicit notation is needed here. It is assumed that enough data are available to determine the function φ , and hence the coefficients φ_n and the covariances $T_n(x, x')$.

Let us now turn to the support of the blocks. We denote by v_x the block v centered at point x and by $Z_v(x)$ the mean value of $Z(\cdot)$ in this block

$$Z_v(x) = \frac{1}{|v|} \int_v Z(x + u) du$$

As a function of x , $Z_v(x)$ is an SRF, which we can consider to be of the form $Z_v(x) = \varphi_v(Y_v(x))$, where $Y_v(x)$ is an SRF with a standard Gaussian marginal. An important remark: contrary to $Z_v(x)$, $Y_v(x)$ is *not* the mean value of $Y(x)$ in the block v_x (that would be too simple!). The transformation function φ_v of the blocks has an expansion that can be written in the form

$$\varphi_v(y) = \sum_{n=0}^{\infty} D_n \varphi_n \chi_n(y) \quad (6.42)$$

where the coefficients D_n are to be determined. Assuming that the bivariate distributions of $Y_v(\cdot)$ as well as the cross-bivariate distributions of $Y(\cdot)$ and $Y_v(\cdot)$ are Hermitian, Matheron (1974c, 1976b) shows that these coefficients are positive and given by

$$D_n^2 = \frac{1}{|v|^2} \int_v \int_v T_n(u, u') du du' \quad (6.43)$$

and also calculates the covariances $T_n(v, v')$ and $T_n(x, v')$ of the factors of the point-block and block-block isofactorial distributions.

A shortcoming of this model is that the hypothesis of isofactorial point-block and block-block distributions may not be mathematically consistent. In particular, there is no guarantee that the densities of the distributions thus obtained are always positive. However, one can expect this model to give acceptable results as long as v remains small.

Discrete Gaussian Model

It can be shown that in the above Hermitian model the covariances of the factors of the isofactorial point-block and block-block cannot be of the form

r^n . This means that, in this model, pairs formed by the random variables $Y(x)$, $Y(x')$, $Y_{v_p}(x)$, $Y_{v_p}(x')$ cannot all be Gaussian, and that the multi-gaussian model is of course excluded, even if $Y(\cdot)$ is itself a Gaussian RF. But bi-gaussian and multi-gaussian models are very convenient to proceed further in the calculation of recovery functions after a change of support, including for indirect recovery functions (i.e., when the selection will be based on an *estimate* of the true value rather than on the true value itself). Hence the motivation for an isofactorial model in which all bivariate distributions are Gaussian. Such a model can be obtained by generalizing the discrete Gaussian model presented in Section 6.4.2. To this end we must accept to discretize the deposit into a partition of small blocks of size v and forget the exact position of each sample in its block.

Problem Discretization

The deposit is regarded as the union of nonoverlapping blocks v_p which are identical up to a translation. To simplify, we denote by Z_{v_p} the mean grade of block v_p , and v the support of these blocks. The random variables Z_{v_p} thus constitute a discretized version of the deposit.

Concerning point-support grades, we will not attempt to consider their exact position inside the block, and will denote by Z_p the grade of a point selected randomly in the block v_p (uniform density) and independently of any other point. Experimental data do not necessarily exist for all indexes p , and will be denoted by Z_α , $\alpha \in A$, where $A = \{p_1, \dots, p_N\}$ is a subset of the set of indexes p .

As in the Hermitian model we will consider that the Z_p and Z_{v_p} are transforms of standard normals Y_p and Y_{v_p} through transformation functions φ and φ_v :

$$Z_p = \varphi(Y_p) \quad Z_{v_p} = \varphi_v(Y_{v_p})$$

Further we assume that point-support bivariate distributions of the SRF $Y(\cdot)$ are Gaussian, that is, of the form (6.41) with

$$T_n(x, x') = [\rho(x' - x)]^n$$

Block Transformation Function

As we have seen in Section 6.4.2, the main assumption of the discrete Gaussian model is that the bivariate distribution of Y_p and Y_{v_p} is Gaussian with a positive correlation coefficient r . The transformation function φ_v is then deduced from Cartier's relation (6.36). Its actual calculation is carried out using the expansions of the transformation functions into normalized Hermite polynomials, either under the usual form (6.19) for φ or under the form (6.42) for φ_v :

$$\varphi(y) = \sum_{n=0}^{\infty} \varphi_n \chi_n(y) \quad \varphi_v(y) = \sum_{n=0}^{\infty} D_n \varphi_n \chi_n(y)$$

The coefficients φ_n , which derive from φ , are assumed known, and the D_n are to be determined (they are different from those obtained in the Hermitian framework). Cartier's relation (6.36) is expanded into

$$\sum_{n=0}^{\infty} r^n \varphi_n \chi_n(Y_{v_p}) = \sum_{n=0}^{\infty} D_n \varphi_n \chi_n(Y_{v_p})$$

Since the χ_n constitute an orthonormal basis, we have $D_n = r^n$ for all n ; thus, if φ_n is the n th coefficient of the expansion of φ , that of φ_v is simply $r^n \varphi_n$.

The correlation coefficient r is selected so that the transformation φ_v defines a random variable whose variance is equal to the variance σ_v^2 given by (6.33). This is computed with the covariance $C(h)$ resulting from the application of relation (6.23). In other words, r is the solution of the equation

$$\sum_{n=1}^{\infty} \varphi_n^2 r^{2n} = \sigma_v^2 \quad \text{with} \quad \sigma_v^2 = \frac{1}{|v|^2} \int_v \int_v \sum_{n=1}^{\infty} \varphi_n^2 T_n(x, x') dx dx'$$

The left-hand side is an increasing function of r , going from 0 to σ^2 as r increases from 0 to 1, where $\sigma^2 = C(0)$ is the variance of $Z(x)$; since the variance σ_v^2 is itself comprised between 0 and σ^2 , this equation has indeed one and only one solution.

Specification of all Bivariate Distributions

For the model to be completely specified, in order for example to carry out a local estimation by disjunctive kriging, two additional hypotheses are required:

- Given Y_{v_p} , Y_p is independent of the other Y_q , $q \neq p$, as well as the other Y_{v_q} , $q \neq p$.
- The pairs (Y_{v_p}, Y_{v_q}) are bivariate Gaussian.

All pairs of variables chosen among the random variables Y_p and Y_{v_p} then have Gaussian bivariate distributions. These are characterized by the corresponding correlation coefficients. If R_{pq} denotes the correlation coefficient between Y_{v_p} and Y_{v_q} , the discrete Gaussian model is defined as follows:

- Block-block distribution,

$$\text{Cov}(Y_{v_p}, Y_{v_q}) = R_{pq} \quad (6.44)$$

- Point-block distribution,

$$\text{Cov}(Y_p, Y_{v_q}) = r R_{pq} \quad (6.45)$$

- Point-point distribution,

$$\text{Cov}(Y_p, Y_q) = r^2 R_{pq} \quad p \neq q \quad (6.46)$$

Proof. Indeed, since the point x_p is selected at random in the block v_p , the covariance between Z_p and Z_{v_q} is equal to the covariance of Z_{v_p} and Z_{v_q} . Likewise the covariance between Z_p and Z_q is equal to the covariance between Z_{v_p} and Z_{v_q} (this would remain true for two different points in the same block $p = q$, though of course not for the variances of Z_p or of Z_q , which remain equal to σ_z^2). But these covariances can be expressed from the covariances of the Gaussian. By transposing relation (6.24) to the transforms $\varphi(Y)$ and $\psi(Y')$ of two Gaussians Y and Y' whose bivariate distribution is Gaussian with correlation coefficient ρ , one has

$$\text{Cov}(\varphi(Y), \psi(Y')) = \sum_{n=1}^{\infty} \varphi_n \psi_n \rho^n$$

Taking into account that the transformation φ_v of the blocks is of the form (6.42), here with $D_n = r^n$, we get

$$\text{Cov}(Z_{v_p}, Z_{v_q}) = \sum_{n=1}^{\infty} \varphi_n^2 r^{2n} (\text{Cov}(Y_{v_p}, Y_{v_q}))^n$$

$$\text{Cov}(Z_p, Z_{v_q}) = \sum_{n=1}^{\infty} \varphi_n^2 r^n (\text{Cov}(Y_p, Y_{v_q}))^n$$

$$\text{Cov}(Z_p, Z_q) = \sum_{n=1}^{\infty} \varphi_n^2 (\text{Cov}(Y_p, Y_q))^n$$

The equality of these three covariances implies relations (6.44), (6.45), and (6.46). \square

In this model things are as though the main role was not held by the SRF $Z(x)$ but by a discretized version composed of the grades of the blocks v_p , or their Gaussian transforms Y_{v_p} . This justifies the name of the model. The samples, to be regarded as randomly located in each block, have grades whose Gaussian transforms are of the form

$$Y_p = r Y_{v_p} + \sqrt{1 - r^2} \varepsilon_p \quad (6.47)$$

where the ε_p are standard normals uncorrelated among themselves and with the Y_{v_q} (whether q is equal to p or not).

There remains to determine the correlation coefficients R_{pq} . This is achieved by inverting numerically the relation

$$\text{Cov}(Z_{v_p}, Z_{v_q}) = \sum_{n=1}^{\infty} \varphi_n^2 r^{2n} R_{pq}^n$$

since the left-hand side can be computed from the covariance $C(h)$ of the RF $Z(x)$ and thus from the covariances of the factors of the RF $Y(x)$:

$$\text{Cov}(Z_{v_p}, Z_{v_q}) = \frac{1}{|v|^2} \int_{v_p} \int_{v_q} \sum_{n=0}^{\infty} \varphi_n^2 T_n(x, x') dx dx'$$

There is no guarantee that the matrix $[R_{pq}]$ thus obtained is positive definite. If this is not the case, a covariance model is fitted to the set of these numerical covariances. The discrete Gaussian model is then consistent with respect to variances and covariances and provides all elements required by disjunctive kriging, as we will see in Section 6.5.3. Matheron (1976b) presents an extension of the model enabling the incorporation of the information effect.

The discrete Gaussian model can easily be extended to the case where point-support bivariate distributions are not Gaussian but Hermitian, assuming that the same is true for point-block and block-block distributions. But we are lacking true guidelines for choosing the type of Hermitian model of these distributions and their parameters, except of course if block data are available.

6.5.2. Other Isofactorial Change-of-Support Models

General Form of Isofactorial Change-of-Support Models

Let us consider that the variable of interest is directly the SRF $Y(x)$ represented by an isofactorial model. The case where $Y(x)$ is a Gaussian SRF is a very special one, for then $Y_v(x)$ is also a Gaussian SRF. Such distributional permanence does not generally hold for other types of random functions, even for the marginal. For example, if Y_1 and Y_2 are two i.i.d. gamma random variables with parameter α , their mean follows a gamma distribution with parameter 2α and is thus more regular. Therefore in the general case we need point-point, point-block, and block-block isofactorial models for SRFs $Y(x)$ and $Y_v(x)$ whose marginal distributions $G(dy)$ and $G_v(dy)$ are of a different form. Considering also that we work on transformed data, the typical model is the following:

- Point-support grades are of the form $Z(x) = \varphi(Y(x))$, where $Y(x)$ is an SRF with marginal distribution $G(dy)$.
- Average grades over a support v are of the form $Z_v(x) = \varphi_v(Y_v(x))$, where $Y_v(x)$ is an SRF with marginal $G_v(dy)$.
- The marginal distribution G has orthonormal factors $\chi_n(y)$, and $Z(x)$ can be expanded as

$$Z(x) = \sum_{n=0}^{\infty} \varphi_n \chi_n(Y(x))$$

- The marginal distribution G_v has orthonormal factors $\chi_n^v(y)$, and $Z_v(x)$ can be expanded as

$$Z_v(x) = \sum_{n=0}^{\infty} \varphi_n^v \chi_n^v(Y_v(x))$$

- The bivariate point-point distribution of $Y(x)$ and $Y(x')$ is of the form

$$G_{xx'}(dy, dy') = \sum_{n=0}^{\infty} T_n(x, x') \chi_n(y) \chi_n(y') G(dy) G(dy')$$

- The bivariate point-block distribution of $Y(x)$ and $Y_v(x')$ is of the form

$$G_{xv'}(dy, dy') = \sum_{n=0}^{\infty} T_n(x, v_{x'}) \chi_n(y) \chi_n^v(y') G(dy) G_v(dy')$$

- The bivariate block-block distribution of $Y_v(x)$ and $Y_v(x')$ is of the form

$$G_{vv'}(dy, dy') = \sum_{n=0}^{\infty} T_n(v_x, v_{x'}) \chi_n^v(y) \chi_n^v(y') G_v(dy) G_v(dy')$$

In the notations $G_{xx'}$ and $G_{vv'}$, the indexes x , v , and v' recall synthetically that the bivariate distributions concern, respectively, $Y(x)$, $Y_v(x)$ and $Y_v(x')$. The first problem is to find consistent models. Then, for applications, the second problem is the choice of the transformation φ_v , and the third the determination of the coefficients φ_n^v of the expansions of φ_v and of the covariances $T_n(x, v_{x'})$ and $T_n(v_x, v_{x'})$. Fortunately consistent models exist, and even a large number of them. Their presentation would take too long, so we will just give a glimpse of the case of a gamma marginal distribution and refer the reader to the literature for the rest.

Gamma Isofactorial Change-of-Support Models

We define a Laguerre change-of-support model exactly as for a Hermitian model, by choosing for Y a gamma marginal with parameter α and for Y_v a less selective gamma distribution, namely with parameter $\alpha_v \geq \alpha$. The functions χ_n and χ_n^v are thus normalized Laguerre polynomials associated, respectively, with α and α_v . As in the Hermitian model we seek coefficients φ_n^v of the form $D_n \varphi_n$. We find exactly the same expressions for D_n , $T_n(x, v_{x'})$ and $T_n(v_x, v_{x'})$ as in the Hermitian case.

We can also define a discrete gamma model exactly as in the Gaussian case (considering bivariate distributions only) for an SRF $Y(x)$ represented by a bi-gamma isofactorial model with factors of the form $T_n = \rho^n$. Under the assumption that point-point and block-block distributions are of this form and that the point-block pair (Y_p, Y_{vp}) follows a dissymmetric bi-gamma distribution,⁷ namely satisfies

$$T_n(\underline{x}_p, v_p) = T_n(\underline{x}, v) = C_n(\alpha, \alpha_v) r^n$$

with

$$C_n(\alpha, \alpha_v) = \sqrt{\frac{\Gamma(\alpha + n) \Gamma(\alpha_v)}{\Gamma(\alpha_v + n) \Gamma(\alpha)}} \quad \text{and} \quad 0 \leq r < 1$$

where \underline{x} denotes a random point in v , we obtain the following results:

- Block transformation:

$$D_n = T_n(\underline{x}_p, v_p) = C_n(\alpha, \alpha_v) r^n$$

- Block-block distribution:

$$\text{Cov}(Y_{v_p}, Y_{v_q}) = \alpha_v R_{pq}$$

- Point-block distribution:

$$\text{Cov}(Y_p, Y_{v_q}) = \alpha_v R_{pq}$$

- Point-point distribution:

$$\text{Cov}(Y_p, Y_q) = \frac{\alpha_v^2}{\alpha_v} r^2 R_{pq} \quad p \neq q$$

r being the solution of the equation

$$\sum_{n=1}^{\infty} C_n(\alpha, \alpha_v)^2 \varphi_n^2 r^{2n} = \frac{1}{|v|^2} \int_v \int_v C(x' - x) dx dx'$$

and R_{pq} the solution of the equation

$$\sum_{n=1}^{\infty} C_n(\alpha, \alpha_v)^2 \varphi_n^2 r^{2n} R_{pq}^n = \frac{1}{|v|^2} \int_{v_p} \int_{v_q} C(x' - x) dx dx'$$

A practical problem is the choice of the parameter α_v . The interval of possible values is in fact rather limited given that we must reach a solution $r \leq 1$ (which limits the possible maximum value of α_v), and that the block transformation must lead to a “reasonable” value for the minimum of Z_v (which sets the possible minimum value). On this subject the reader is referred to Hu (1988).

Some Other Models

1. Change of support for a mosaic model is touched on in Matheron (1982b, 1984a) and treated in detail in Matheron (1984d). Several models are proposed. For one of them disjunctive kriging of block recovery functions only requires solving a single kriging system, but the solution has nothing to do with an affine correction on point-support indicator kriging estimates.
2. For the model with orthogonal indicator residuals, a change-of-support model is proposed by Rivoirard (1988, 1989) and Matheron (1989).
3. All discrete diffusive models with polynomial factors (except probably the anti-Jacobi model) have change-of-support models (Matheron, 1980, 1984e, 1989). These models associate a discrete point-support variable with a block-support variable having the same distribution (but different parameters). But it is also interesting to be able to represent the block characteristics by continuous variables. Mixed change-of-support models have been developed to this effect, namely the negative binomial/gamma model (Matheron, 1984a) and the discrete Jacobi/beta model (Matheron, 1984e).

6.5.3. Local Estimation with Change of Support

We now focus on a particular block v , indexed by 0, and seek the distribution of the grade Z_{v_0} of this block conditionally on neighboring data $\{Z_\alpha = Z(x_\alpha) : \alpha = 1, \dots, N\}$.

Disjunctive Kriging

Isofactorial change-of-support models have the advantage of providing consistent solutions at the local and global scales. Gaussian transformations are generally used in conjunction with the discrete Gaussian model, but the Hermitian model can also be considered. These models provide all covariances needed for disjunctive kriging of any function of Y_{v_0} (and thus of Z_{v_0}) from data Y_α , $\alpha \in A$, where A is a subset of blocks, by duplicating the method described in Section 6.3.1 (we could even take into account the mean grades of blocks already mined, presented as Y_{v_β} for some indexes β).

The discrete Gaussian model gives good results provided that the bi-gaussian assumption holds reasonably well, that point-support grades in a block are well correlated with the average grade of the block (a strong nugget effect would be detrimental), and that the distribution of grades is continuous or comprises a small proportion of zeros (less than 10% to be specific). Like the Hermitian model its performance can deteriorate for low and high cutoffs.

Conditional Distribution in the Discrete Gaussian Model

Instead of using DK, the conditional distribution can be determined directly if we accept the framework of a full Gaussian model: assuming that the vector of the Y_{v_p} is multivariate Gaussian, so is also the vector of the Y_{v_p} and Y_p . The conditional distribution of Y_{v_0} is then a Gaussian whose conditional expectation and variance coincide with the simple kriging estimate and variance of Y_{v_0} . The simple kriging estimator of Y_{v_0} is

$$Y_{v_0}^* = \sum_{\alpha \in A} \lambda_\alpha Y_\alpha$$

with λ_α solutions of the system

$$\sum_{\beta \in A} \lambda_\beta \text{Cov}(Y_\alpha, Y_\beta) = \text{Cov}(Y_\alpha, Y_{v_0}) \quad \forall \alpha \in A$$

According to (6.46) and (6.45) we have

$$\text{Cov}(Y_\alpha, Y_\beta) = (1 - r^2) \delta_{\alpha\beta} + r^2 R_{\alpha\beta}$$

$$\text{Cov}(Y_\alpha, Y_{v_0}) = r R_{0\alpha}$$

The kriging system therefore writes

$$(1 - r^2) \lambda_\alpha + r^2 \sum_{\beta \in A} \lambda_\beta R_{\alpha\beta} = r R_{0\alpha} \quad \forall \alpha \in A$$

and the kriging variance is

$$\sigma_{SK}^2 = \text{Var}[Y_{v_0}^* - Y_{v_0}] = \text{Var}[Y_{v_0}] - \sum_{\alpha \in A} \lambda_{\alpha} \text{Cov}(Y_{\alpha}, Y_{v_0}) = 1 - r \sum_{\alpha \in A} \lambda_{\alpha} R_{0\alpha}$$

The conditional distribution of Y_{v_0} is Gaussian and given by

$$\Pr\{Y_{v_0} < y \mid Y_{\alpha} = y_{\alpha} : \alpha \in A\} = G\left(\frac{y - \sum_{\alpha \in A} \lambda_{\alpha} y_{\alpha}}{\sigma_{SK}}\right)$$

Since the estimator $Y_{v_0}^*$ and the kriging error $Y_{v_0}^* - Y_{v_0}$ are orthogonal and their sum has unit variance, the variance of the kriging estimator is

$$\sigma^{*2} = \text{Var}[Y_{v_0}^*] = 1 - \sigma_{SK}^2$$

and the covariance of Y_{v_0} and $Y_{v_0}^*$ is also equal to σ^{*2} . The pair $(Y_{v_0}, Y_{v_0}^*/\sigma^*)$ is distributed as a standard bivariate Gaussian with correlation coefficient σ^* . Its isofactorial representation is given by (6.25) with $\rho = \sigma^*$. It enables the calculation of local recovery functions. Further results on how to take into account not only the support effect as done here but also an information effect can be found in Matheron (1976b).

One may note that the assumptions that have been introduced along the way are on the whole not very different from those of the multi-gaussian model of Section 6.4.3. The discrete Gaussian model can thus appear as an approximation to the multi-gaussian model. However, this latter model involves calculations that are only approximate and much more cumbersome. Moreover the estimation of recovery functions of a set of small blocks (in mining, the selective mining units) can be obtained in a single run by disjunctive kriging, whereas the conditional expectation requires the combination of block by block calculations.

Uniform Conditioning

In general, we are not interested in the distribution of the mean grade of a single block v_0 but in the distribution of small blocks of size v making up a panel V . The result is obtained as a mixture of elementary distributions. If V is very large compared with v , the calculation can become prohibitive, so approximations are used, the most classic of which is uniform conditioning (Matheron, 1974c; Remacre, 1984; Marcotte and David, 1985; Rivoirard, 1994, ch. 11). The method consists in conditioning the grades of small blocks by the mean grade of the panel rather than by the samples in the blocks (working on the Gaussian transforms).

Taking into Account Stationarity Defects

When $Z(x)$ is not globally stationary, we have seen at the end of Section 6.3.4 that we could substitute ordinary kriging (with a constraint on the sum of

weights) for simple kriging in view of calibration on the local mean. This is not possible in the case of DK with change of support. Indeed DK of Y_r is then co-DK of Y_r from data $Y(x_\alpha)$ of a different kind. While the RFs $Z(x)$ and $Z_r(x)$ have the same local mean, this is not true of the RFs $Y(x)$ and $Y_r(x)$, and co-DK is possible only if the means of the two RFs are known. In other words, the practice of DK with change of support requires stationarity and not only local stationarity. An approximate solution to the problem exists, however, in the case of uniform conditioning (Guibal and Remacre, 1984; Remacre, 1987; Rivoirard, 1994, ch. 12).

Estimation of Local Multivariate Recovery

In a mine there are generally several mineral constituents $Z_1(x), \dots, Z_p(x)$. If the selection of mined blocks is performed on the basis of a particular mineral grade, one still wishes to know the quantities of metal that will be recovered for each constituent. To this end Maréchal (1982) proposes a generalization of conditional expectation under a multivariate Gaussian hypothesis, along with disjunctive cokriging estimators in a setting in which all bivariate distributions are Gaussian.

6.5.4. Applications and Discussion

The first implementations of disjunctive kriging with change of support were made in the scope of the discrete Gaussian model. In mining, predictions obtained by disjunctive kriging from exploration boreholes have been compared with reality (the actual grade of mined panels being seldom known, each panel was discretized by the blast holes it contains, Jackson and Maréchal, 1979; Young, 1982). A number of authors have compared the performance of different estimation methods of recoverable reserves on real or synthetic data sets. Liao (1990) and Chilès and Liao (1993) compared indicator cokriging (with affine correction for change of support) and disjunctive kriging with Gaussian or gamma transformation on data from three real deposits (including the Salsigne gold mine). These studies arrive at the following conclusions:

- Concerning global recoverable reserves, the discrete model (Gaussian or gamma) generally produces acceptable results, as opposed to the affine correction which is to be banned except if the nugget effect is small and if the change of support remains moderate.
- Local recoverable reserves can be predicted fairly well by disjunctive kriging, if applied carefully (e.g., straight elimination of isolated blocks and border areas if they are barren).
- Indicator kriging with affine correction can produce results close to those of disjunctive kriging in some limited cases (mosaic model, moderate support effect, and limited σ/m variability). It can even outperform disjunctive kriging if the model used for DK is selected arbitrarily (e.g.,

bi-gaussian model) and is ill-suited. It is not as good as disjunctive kriging with an appropriate isofactorial model. It often comes with a strong bias, especially regarding the estimation of the quantity of metal.

- Despite appearances the implementation of indicator kriging is no simpler than that of disjunctive kriging, except if the same variogram is used for all thresholds.

As emphasized earlier by Matheron (1985) and Lantuéjoul (1988), these works demonstrate the importance of a correct choice of the model. Indeed the advantages of nonlinear methods vanish if applied blindly. The value of modeling bivariate distributions lies in the availability of relatively varied models, of diagnostic tools for choosing among these models, of methods for fitting their parameters and checking whether these models provide a reasonable representation of the data. Compared with indicator kriging, for example, the DK approach has the advantage of providing consistent change-of-support models. Of course it is ideal if both point-support data (samples) and block data (grades of mined blocks) are available to validate the model—and they can also be used for estimation.

NOTES

1. Air pollution standards are often defined by reference to peak values, a typical rule being that at each station where it is monitored, the pollution level should not exceed a prescribed threshold. Naturally this is an incentive not to monitor. The statistician's recommendation to make the threshold dependent on the number of monitoring stations is regarded as unfair by politicians who contend that pollution standards should be the same for everyone. The dilemma could be resolved by including in the standards guidelines on what a good monitoring network should be like (Paul Switzer, personal communication).
2. A lognormal RF is a random function of the form $Z(x) = \exp(Y(x))$, where $Y(x)$ is a Gaussian SRF. Its marginal distribution has two parameters. In applications we sometimes have to add a translation parameter and to consider that $Z(x) = b + \exp(Y(x))$. As the properties of the three-parameter lognormal distribution are easily deduced from those of the two-parameter distribution we will only consider the latter.
3. Note that the H_α are not disjoint since all contain the constant random variables. Also, the projection theorem assumes that the space H is closed, which is not proven. However, this question is not blocking since in the isofactorial model the projection can be found and is of the correct form.
4. It is, however, easy from such a basis to derive a basis without constant function, such as by substituting $(\chi_0(\cdot) - \chi_1(\cdot))/\sqrt{2}$ and $(\chi_0(\cdot) + \chi_1(\cdot))/\sqrt{2}$ to χ_0 and χ_1 .
5. In most textbooks the definition includes a $(-1)^n$ factor, which leads to odd-order Hermite polynomials of the opposite sign. We have kept the present definition for consistency with the geostatistical literature.
6. The constant factor is usually ascribed the index N instead of 0.
7. Let U and V be two standard gamma variables with the same parameter α and a Laguerre bivariate distribution with covariances T_n . Let W be an independent standard gamma variable with parameter $\alpha' - \alpha > 0$. The variable $V + W$ is then a standard gamma with parameter α' and the pair $(U, V + W)$ has an isofactorial bivariate distribution with covariances $T_n^\lambda = \sqrt{\Gamma(\alpha + n)\Gamma(\alpha')/\Gamma(\alpha' + n)\Gamma(\alpha)} T_n$. This distribution is called a dissymmetric Laguerre distribution. Here we consider the dissymmetric bi-gamma distribution, namely the case $T_n = r^n$.

CHAPTER 7

Conditional Simulations

Je suis un mensonge qui dit toujours la vérité.
Jean Cocteau

7.1. INTRODUCTION AND DEFINITIONS

7.1.1. Introductory Example

As an introduction let us consider the somewhat simplified problem brought up by Alfaro (1979) and illustrated in Figure 7.1: a submarine cable is to be laid between two cities on both sides of an ocean, and we want to predict the length of the cable. Measurements of the depth of the sea floor are available every 100 m (Fig. 7.1a).

An easy solution is to estimate the depth by kriging at every point along the profile and to compute the length of this kriged curve (Fig. 7.1b). As an example, the length obtained for the profile is $L = 945$ m. But this result is not quite convincing, since the actual seafloor is far more complex than the kriged one (kriging has a smoothing effect). As a result the length computed from the kriged profile certainly underestimates, perhaps quite severely, the actual length.

To get the exact solution, we need a continuous, or at least very dense, survey of depth (Fig. 7.1c). In the present case a survey with samples taken every 10 m gives a length $L = 1182$ m. The previous estimate $L = 945$ m was thus indeed too low. Is there a way of avoiding the costs of the second survey? There is if we introduce a model.

The dependency between the length of a curve and the scale of investigation is a key issue in fractal theory (Mandelbrot, 1975b, 1977, 1982). It is easy to see that if the depth $Z(x)$ is a Gaussian IRF-0 with variogram $\gamma(h)$, and if the distance D is sampled at the interval d , the length L has expectation

$$E(L) = D \int \sqrt{1 + \sigma^2 u^2} g(u) du$$

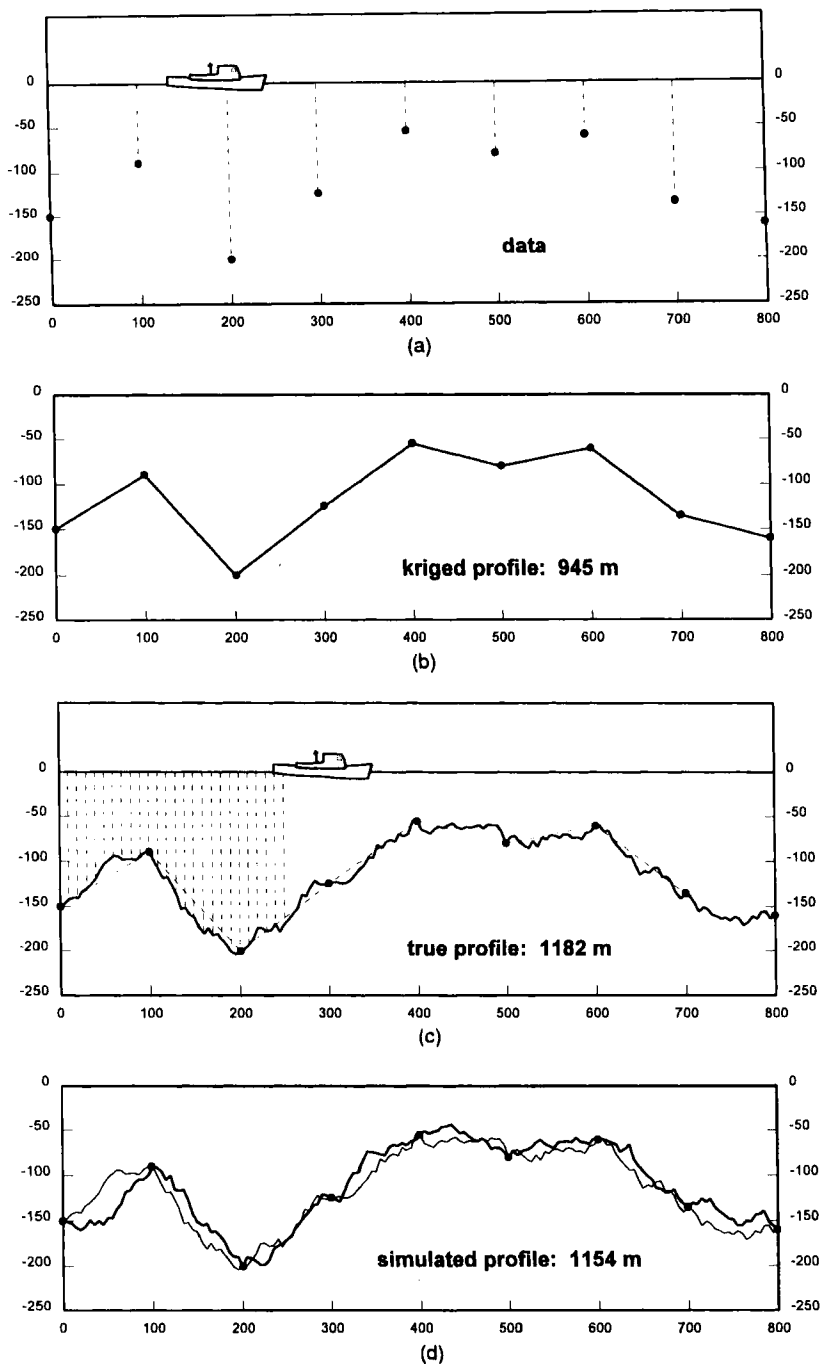


FIGURE 7.1. Submarine cable: (a) seafloor survey with 100-m spacing; (b) kriging estimate of the seafloor; (c) continuous survey (true profile); (d) conditional simulation based on the 100-m data, and true profile. From Alfaro (1979).

where $\sigma^2 = 2\gamma(d)/d^2$ and $g(u)$ denotes the standard Gaussian p.d.f. If the IRF is not differentiable, the length L increases as the sampling interval used to measure it becomes smaller, and it becomes so large as to be best considered infinite. If the variogram is of type $|h|^\alpha$, the relationship between the lengths $L(d)$ and $L(d_0)$ is

$$L(d) \simeq \left(\frac{d}{d_0} \right)^{(\alpha/2)-1} L(d_0)$$

provided that d and d_0 are small enough. In reality, even if the length of the seafloor along the profile were infinite, a cable has some rigidity. If we admit that it can hang above the sea bottom over distances of about 10 m, then the goal becomes an estimation of $L(d)$ for $d = 10$ m.

The idea of conditional simulations is to build a representation of the phenomenon that is consistent with the data observed at the 100-m sampling interval, as kriging is, and yet reproduces the local fluctuations at the scale of 10 m. Clearly a conditional simulation is not reality but only a possible version of it, among myriads of others. An example is shown in Figure 7.1d; the conditional simulation differs from the real seafloor, often more than kriging does, but has the same character. Not surprisingly the cable measured on this simulation with a sampling interval of 10 m is very close to the true length: $L = 1154$ m instead of $L = 1182$ m. Other features of the seafloor can also be seen from the simulation, for example, the occurrence of large slopes. This is what makes simulations interesting.

"You are trying to entangle us in your mathematical tricks!" shouted a bewildered attendant at a presentation of conditional simulations. Likewise, the reader may at this point have at least three questions in mind:

1. How can one claim to reproduce a profile at a 10-m sampling interval from data at a 100-m sampling interval? By building a model at the 10-m scale, either derived from similar cases, or from additional sampling at the 10-m interval. But this will not require a systematic survey of the seafloor every 10 m along the whole profile. A limited number of data at 10-m spacing will suffice to evaluate the variogram at distances of 10 m and more.
2. How confident can one be in the result of the simulation? Had another simulation been made the cable length would have been different. Indeed. Therefore several simulations are generated so as to determine the probability distribution of the cable length.
3. Why not estimate the length L directly? The answer is that kriging—in its linear version—only allows estimation of quantities that are linear in the studied variable, such as the average depth along the profile. The length L is not a linear functional of depth but essentially depends on local fluctuations of the seafloor.

7.1.2. Definition and Use of Conditional Simulations

Nonconditional Simulations and Conditional Simulations

A (nonconditional) simulation of the random function $\{Z(x) : x \in \mathbb{R}^n\}$ is simply a realization of $Z(x)$, randomly selected in the set of all possible realizations. Its construction requires the knowledge of the spatial distribution of the RF $Z(x)$, at least implicitly. In practice this poses two problems:

1. The random function $Z(x)$ is usually defined so that some regionalized variable $\{z(x) : x \in D \subset \mathbb{R}^n\}$, observed at sample points $\{x_\alpha : \alpha = 1, \dots, N\}$, can be considered as one of its realizations. Its finite-dimensional distributions can never be inferred from the data, even in a stationary framework. The situation is very different from image analysis where the complete image is available and allows, to some degree, an estimation of the multivariate distributions.¹ In geostatistics we can at best hope to obtain bivariate distributions, and possibly a few multivariate distributions, if we have data on a sufficiently regular grid.
2. Even if we knew the spatial distribution of $Z(x)$, we would in general still be unable to construct a simulation. The only truly general simulation procedure is the sequential method presented later in this chapter. However, this method has practical limitations and demands that we are able to calculate the conditional distributions, which is unusual. On the other hand, there are general methods for simulating Gaussian SRFs, for all covariance models. The Gaussian case aside, there is a vast number of random function models, each having particular spatial characteristics and corresponding to a class of possible covariance functions.

In view of this, we can consider a definition at the second order: a (nonconditional) simulation of the random function $Z(x)$ is a realization of an RF $S(x)$ chosen in the class of all RFs with the same second-order moments as $Z(x)$, namely same covariance, variogram, or generalized covariance.

The random function $S(x)$ has an infinite number of realizations. Among them, some assume at the sample points the same values as those observed and thus can be considered to better represent the regionalized variable $z(x)$. They will be called conditional simulations. A conditional simulation is therefore a realization randomly selected from the subset of realizations that match the sample points. Equivalently, it is a realization of a random function with a conditional spatial distribution.

Generating a simulation means choosing a simulation algorithm and therefore the random function $S(x)$. Whenever possible, we will attempt to increase specificity by restricting the choice of $S(x)$ to a narrower class than that defined at the second order. In the stationary case, for example, the marginal distribution of $Z(x)$ can be known, at least approximately, from the data, and we will require that $S(x)$ has the same marginal distribution. We could also ask for a reproduction of the bivariate distributions when these can be determined; un-

fortunately, in many cases we do not know how to generate a simulation with given bivariate distributions. Anyway, the choice of the simulation method must be justified and driven by the problem and the data, although a certain measure of arbitrariness in the choice is inevitable.

Use of Conditional Simulations

Conditional simulations are useful qualitatively, to obtain realistic pictures of spatial variability. Quantitatively, they are the tool of choice to evaluate the impact of spatial uncertainty on the results of complex procedures, such as numerical modeling of a dynamic system or economic optimization of the development of a natural resource. Conditional simulations fall in the scope of so-called Monte Carlo methods—they are in fact nothing but *spatially consistent* Monte Carlo simulations. Section 7.11 presents simulation case studies in mining and in petroleum.

The goal pursued here is not to reproduce the genetic mechanisms that generated the observed phenomenon. More modestly it is simply to mimic its spatial variations as realistically as possible. The results of a simulation must always be considered with a critical eye and checked against the background of the application. This is very important when studying problems such as connectivity, fluid flow, or transport, which compel us to consider, explicitly or not, a model specified well beyond its second-order moments. This is illustrated in Figures 7.2 and 7.3, showing 1D and 2D views of nonconditional simulations of various SRFs with the same exponential covariance function. They look very different (notice simulations 7.2a, 7.3a, and 7.3b: they are non-ergodic provocations!). If we built 100 simulations of each SRF instead of a single one to estimate a second-order moment numerically, each SRF would lead to the same result, up to fluctuations. But it is easy to imagine that the conclusions drawn with regards to connectivity or fluid flow would be very different, especially if we are interested in extreme situations (an illustrative example is given by Journel and Deutsch, 1993). In this respect conditioning on the actual observations is a safeguard but does not exempt us from choosing an adequate SRF model.

How Many Simulations Should We Generate?

The answer depends on the objective and on the structure of the phenomenon. When modeling a stationary field over an area much larger than the range, a single simulation can give a view of a variety of possible local situations. This is often sufficient, for example, to assess the performance of a mining scenario, which depends mainly on the local variability of ore grades or thicknesses. Conversely, when studying a nonstationary field such as a petroleum reservoir, a single simulation provides a single answer in terms of flow and production because these are global problems. It is therefore necessary to build several simulations if we want to assess the range of the possible results. Typically one hundred simulations may be needed, but this number depends largely on the distribution of the parameter of interest: if it is skewed and if we care

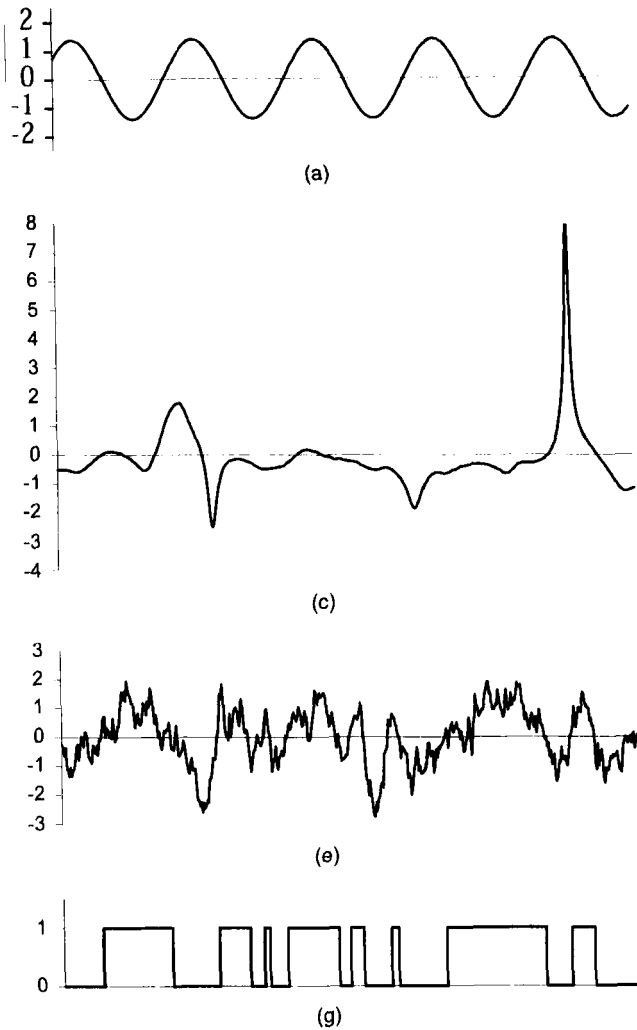


FIGURE 7.2. 1D simulations of the same exponential covariance: (a) elementary spectral process; (b) Ambartzumian process (dilution of Poisson germs); (c) dilution by a symmetric function; (d) dilution by random segments; (e) Gaussian process; (f) mosaic process with Gaussian marginal (partition into Poisson segments); (g) truncated Gaussian process; (h) mosaic random set (Poisson segments set independently to 0 or 1). Note that the processes *a*, *e*, and *f* have the same Gaussian marginal distribution and that the random sets *g* and *h* have the same mean $\frac{1}{2}$. The process *a* is not ergodic, and in practice, many elementary simulations are superimposed. Scale parameter: $1/30$ th the length.

even about extreme situations, a larger number of simulations is needed—but bear in mind that extreme value predictions usually lack robustness. It must be pointed out that conditional simulations are not equally probable; some are

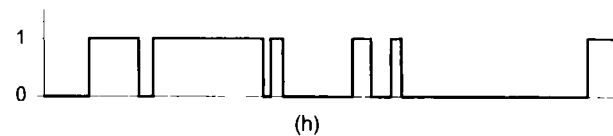
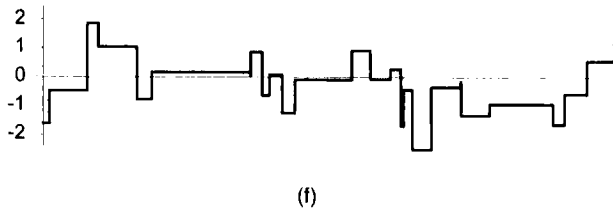
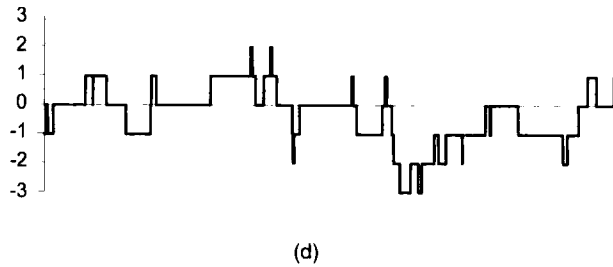
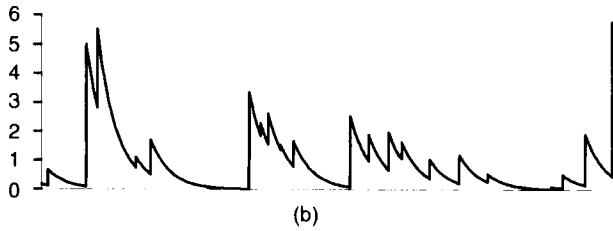


FIGURE 7.2. (Continued).

extreme and others are “average”. Rather, they represent i.i.d. drawings from (multivariate) conditional distributions that have no reason to be uniform. For some applications simulations are ranked on the basis of a regional quantity computed from the realization (e.g., the volume above a reference level).

Should We Simulate the Uncertainty on the Parameters?

When we build several simulations, a question is whether or not to include the uncertainty on the parameters (mainly the histogram and the variogram). In

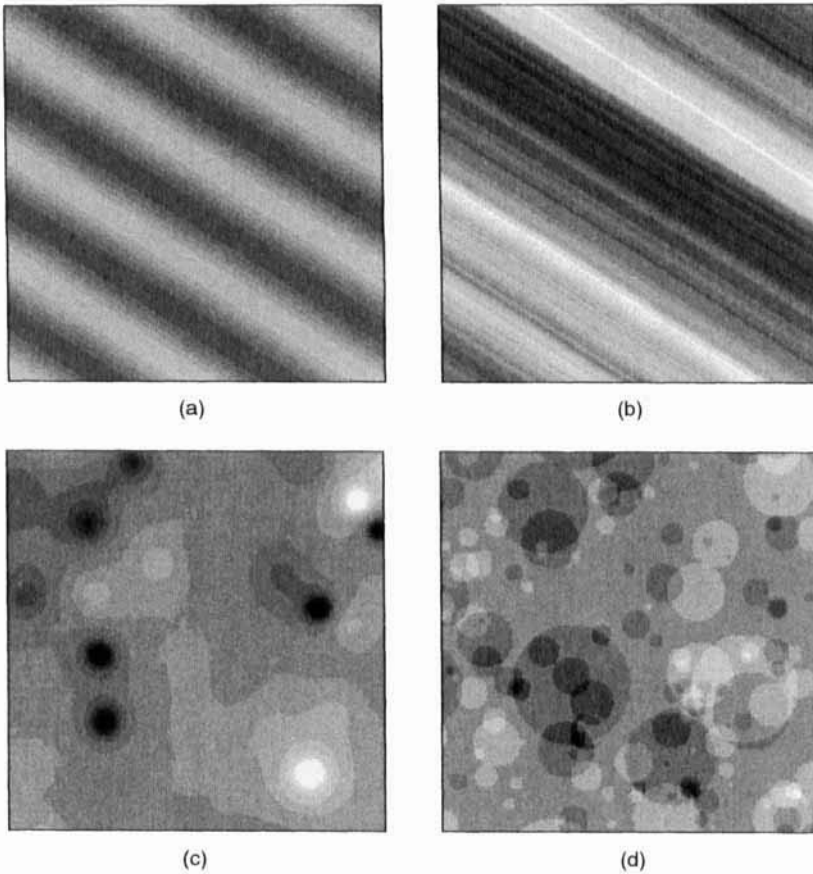
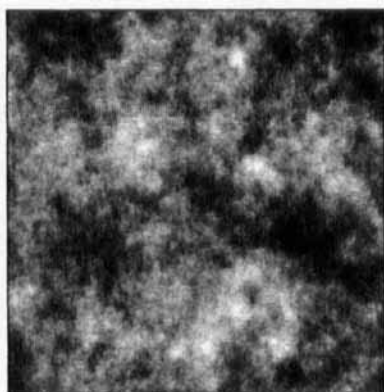
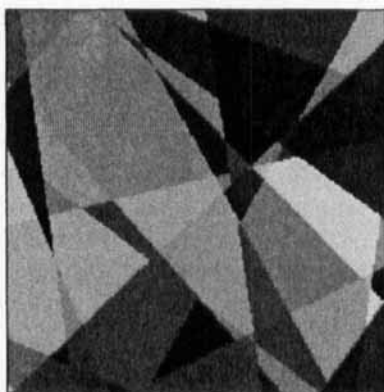


FIGURE 7.3. 2D simulations showing the same exponential covariance: (a) elementary spectral RF; (b) elementary RF by turning bands with Gaussian marginal; (c) dilution by an isotropic function; (d) dilution by random disks; (e) Gaussian RF; (f) mosaic RF with Gaussian marginal (partition into Poisson polygons); (g) truncated Gaussian RF; (h) mosaic random set (Poisson polygons set independently to 0 or 1); (i) sequential indicator simulation; (j) nonmosaic random set (Poisson polygons set alternately to 0 or 1). Note that the RFs *a*, *b*, *e*, and *f* have the same Gaussian marginal distribution, and that the random sets *g*, *h*, *i*, and *j* have the same mean $\frac{1}{2}$. The RFs *a* and *b* are not ergodic, and in practice, many elementary simulations are superimposed. Scale parameter: 1/10th the grid size.

our approach only regional parameters have an objective meaning. We know them imperfectly, and since we use models that are somehow stationary, we choose theoretical parameters close to the empirical ones. But we know that in the scope of our models, other values of the theoretical parameters are compatible with the empirical values. It may therefore be sensible to sample the histogram and variogram parameters for each simulation from a prior distribution, if this prior distribution can reasonably be assessed. It is advisable,



(e)



(f)



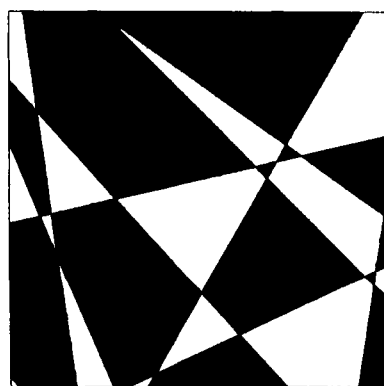
(g)



(h)



(i)



(j)

FIGURE 7.3. (Continued).

however, to study separately simulations with sensible differences in spatial variability (e.g., type of behavior of the variogram at the origin). And we must be aware that differences in the simulation algorithm can have a greater impact than variations in the parameters. Randomization of the parameters is more problematic when several variables are simulated separately and combined for modeling a complex process such as the migration of pollutants or the production of a petroleum reservoir. The definition of the possible ranges for the parameters and their possible correlations must proceed from an agreed methodology and cannot simply result from geostatistical calculations because those would imply further hypotheses about the phenomenon, for example, a full Gaussian assumption (in a similar context, see the system analysis approach adopted by Chapman et al., 1995, for the development of scenarios of evolution of a nuclear waste repository).

7.1.3. A Classification of the Methods

From the point of view of applications three broad categories of quantities are to be simulated:

1. *Continuous variables.* These are measurements of physical properties, such as a mineral grade, the thickness of a geological layer, or the velocity of sound. They are continuous variables in the sense of a continuous histogram, and also in a spatial sense, the nugget effect, if any, being treated separately. Continuous variables have a natural ranking and we can do maths on them (e.g., add, subtract, multiply, divide, take logarithms).
2. *Categorical variables.* These are discrete variables representing elements of a classification. Typical examples are lithofacies types or indicators of large genetic units. Categorical variables may or may not be rank ordered, but in general, we cannot do maths on them. The simulations define a partition of the space into cells with a constant value—within which physical variables can be simulated.
3. *Objects.* These represent entities of the application domain, whose morphology is captured as a whole and simulated in one shot. For example, a fracture is simulated as a segment or a disk with given dimensions and orientation, a meandering channel as a sinusoidal strip with given wavelength, amplitude, axis orientation, thickness, width, and optionally length. In contrast to *object-based* simulations the other types of simulations are generated point by point and often referred to as *pixel-based* simulations.

The models used to represent these quantities do not quite follow this classification, mainly because categorical variables can be simulated either directly or by truncation of a continuous variable. We will distinguish four basic model

types:

1. *Diffusive model.* The simulations are continuous functions (almost surely). The name originates from physical diffusion, the slow process by which molecules spread out evenly. The prototypical example is the Brownian process, whose transition probabilities satisfy the celebrated diffusion equation. In geology one can think of diffusive impregnation. The most used diffusive model is simply the Gaussian RF model. By analogy, in the case of an integer random function, a discrete diffusive model is one that varies by unit jumps.
2. *Jump model.* The simulations are piecewise constant and progress by random jumps (usually noninteger). The prototype is the Poisson process with random jumps and in a wider sense the dilution random functions.
3. *Mosaic model.* The simulated space is partitioned randomly into cells within which the simulation is constant, and the values taken on in the various cells are independently drawn from the same distribution. A typical example is a partition into Voronoi polygons or polyhedra based on a Poisson point process.
4. *Random set model.* The simulations assume the values 0 or 1 only, representing the indicator of a random set (binary simulation). The typical example is the Boolean model obtained by independent placement of random sets.

Table 7.1 summarizes the various methods presented in this chapter, in the order in which they are introduced, diffusive or jump models first, then mosaic and indicator models. The following characteristics are considered:

- *Covariance model.* Some methods are very general (but limited to the Gaussian case), others are limited to particular covariance models; some are limited to \mathbb{R}^1 , whereas others are usable directly in \mathbb{R}^n .
- *Model type.* For diffusive or jump models, the simulation methods are mostly proven ones, and we know how to do the conditioning, either directly or not. For mosaic or random set models, some of the simulation methods are recent, and there is no general method of conditioning. However, most methods allow the construction of conditional simulations, generally by use of iterative algorithms.
- *Ergodicity.* We generally wish each simulation to have a sample mean close to the theoretical mean m and a sample covariance close to the theoretical covariance $C(h)$, especially if the simulated domain is sufficiently large. This implies that all simulations are drawn from the realizations of an RF that is ergodic in the mean value and the covariance (second-order ergodicity). Some simulation methods do not satisfy this condition: they yield simulations whose regional covariance $C_R(h)$ (in the sense of Section 2.2.4) may bear only a remote resemblance with the theoretical covariance $C(h)$, even when the simulated domain is very large. (See the

TABLE 7.1. Main characteristics of the various simulation methods

Method	Covariance	\mathbb{R}^n	Model Type	Ergodic	Systematic			Exact Covariance
					Conditioning	Grid Required	N_s	
Sequential Gaussian	All C	All n	Gaussian	Yes	Direct	No	$< 10^3$ $> 10^3$	Yes No
Matrix decomposition	All C	All n	Gaussian	Yes	Direct	No	$< 10^3$	Yes
Turning bands	All C, γ, K	$n > 1$	*	No, if single direction	Kriging	No	All N_s	*
Autoregressive	Damped exponential and sinusoidal	$n = 1, 2$	Gaussian	Yes	Kriging	Yes	All N_s	Yes
Moving average	Most C with short range	$n = 1, 2$ $n > 2$	Gaussian	Yes	Direct Kriging	Yes	All N_s	Yes
Dilution	Most C, γ	All n	Jump	Yes	Kriging	No	All N_s	Yes
Continuous spectral	All C, γ	All n	Gaussian (Periodic)	No	Kriging	No	All N_s	Yes
Discrete spectral	Most $C, \gamma = h ^a$, polynomial K All C finite range	$n = 1$ $n > 1$	Gaussian	Yes	Kriging	Yes	All N_s	Yes

Poisson hyperplanes	Linear	All n	Jump	No	Kriging	No	All N_S	Yes
Integration	Polynomial or spline K	$n = 1$	Apparent drift	—	Kriging	No	All N_S	*
Sequential indicators	All valid C	All n	Discrete	Yes	Direct	No	All N_S	*
Truncated Gaussian	C of a particular family	All n	Discrete	*	Iterative	*	*	*
Voronoi	Specific C	All n	Mosaic	Yes	Not known	No	All N_S	Yes
Poisson polyhedra	Exponential	All n	Mosaic	Yes	Not known	No	All N_S	Yes
Substitution	Particular C	All n	—	*	Kriging	*	*	*
Boolean	Specific C	All n	Binary	Yes	Iterative	No	All N_S	Yes
Simulated annealing	—	All n	Not known	Yes	Iterative	Yes	All N_S	No

Note: C , γ and K indicate the possibility of simulating covariances, variograms and generalized covariances; n : dimension of the space; N_S : number of points simulated; for methods that are coupled with another method, * indicates a property that depends on the associated method, — means not applicable or unknown. Many variants exist, notably as regards the model type.

continuous spectral simulations of Figures 7.2a and 7.3a.) Nonergodic methods are in fact utilized by adding a large number of independent simulations.

- *Direct conditioning.* Only a few methods have this capability. In most cases a nonconditional simulation is generated first and then conditioned.
- *Conditions of use.* Considerations are points on a systematic grid or not, maximum number of simulated points, exact or approximate reproduction of the covariance, possibility of carrying out several simulations in parallel.

The methods for generating nonconditional simulations usually produce realizations of strictly stationary RFs with zero mean. It is interesting to note that the addition, properly scaled, of a large number of such independent simulations tends to a Gaussian RF. Specifically, if $S_1(\cdot), S_2(\cdot), \dots$, is a sequence of independent zero-mean, finite-variance simulations with the same spatial distribution, the spatial distribution of the random function $T_k(\cdot)$ defined by

$$T_k(x) = \frac{S_1(x) + \dots + S_k(x)}{\sqrt{k}}$$

tends to a Gaussian RF as $k \rightarrow \infty$. Indeed the central limit theorem (Feller, 1971, sec. VIII.4) ensures that the distribution of any linear combination of the form $\lambda_1 T_k(x_1) + \dots + \lambda_M T_k(x_M)$, where M , $\lambda_1, \dots, \lambda_M$, and x_1, \dots, x_M are arbitrary, tends to a Gaussian distribution as $k \rightarrow \infty$, which means that any finite-dimensional distribution of $T_k(\cdot)$ tends to a multivariate Gaussian distribution.

7.2. DIRECT CONDITIONAL SIMULATION OF A CONTINUOUS VARIABLE

In this section and the next, we consider methods that produce diffusive models, or occasionally jump models: the realizations are continuous functions (diffusive model), or piecewise continuous with fixed or random discontinuities (jump model). These methods can be used to construct realizations of Gaussian RFs (up to a transformation) either directly or by averaging a large number of independent simulations. Some methods have variants allowing the simulation of indicator or mosaic RFs (Section 7.7).

7.2.1. Sequential Simulation

Consider a vector-valued random variable $\mathbf{Z} = (Z_1, Z_2, \dots, Z_N)'$ for which a realization of the subvector $(Z_1, Z_2, \dots, Z_M)'$ is known and equal to $(z_1, z_2, \dots, z_M)'$ ($0 \leq M < N$). The distribution of the vector \mathbf{Z} conditional on $Z_i = z_i$,

$i = 1, 2, \dots, M$, can be factorized in the form

$$\begin{aligned}
 & \Pr\{z_{M+1} \leq Z_{M+1} < z_{M+1} + dz_{M+1}, \dots, z_N \leq Z_N < z_N + dz_N \mid z_1, \dots, z_M\} \\
 &= \Pr\{z_{M+1} \leq Z_{M+1} < z_{M+1} + dz_{M+1} \mid z_1, \dots, z_M\} \\
 &\quad \times \Pr\{z_{M+2} \leq Z_{M+2} < z_{M+2} + dz_{M+2} \mid z_1, \dots, z_M, z_{M+1}\} \\
 &\quad \vdots \\
 &\quad \times \Pr\{z_N \leq Z_N < z_N + dz_N \mid z_1, \dots, z_M, z_{M+1}, \dots, z_{N-1}\} \quad (7.1)
 \end{aligned}$$

We can therefore simulate the vector \mathbf{Z} sequentially by randomly selecting Z_i from the conditional distribution $\Pr\{Z_i < z_i \mid z_1, \dots, z_{i-1}\}$ for $i = M + 1, \dots, N$ and including the outcome z_i in the conditioning data set for the next step.

This procedure is absolutely general and can be used in particular for Z_i of the form $Z_i = Z(x_i)$, where $Z(x)$ is an RF and where the x_i are the sample points ($i = 1, 2, \dots, M$) and the points where we wish to simulate the RF ($i = M + 1, \dots, N$). It makes possible the construction of both a nonconditional simulation ($M = 0$) and a conditional simulation ($M > 0$). The procedure can be applied to the cosimulation of several nonindependent RFs. It produces simulations that match not only the covariance but also the spatial distribution.

The practical difficulty is that in general, we do not know how to calculate the conditional probabilities involved in (7.1), except in the ideal case of a Gaussian random vector, where the method is classically employed (e.g., see Ripley, 1987). It has been introduced in geostatistical applications by Alabert and Massonat (1990) to simulate Gaussian RFs (small-scale log-permeability variations). For a Gaussian RF *with known mean*, the conditional distribution of Z_i is Gaussian, with mean Z_i^* and variance $\sigma_{K_i}^2$, where Z_i^* is the simple kriging estimator of Z_i from $\{Z_j : j < i\}$, and $\sigma_{K_i}^2$ the associated kriging variance.

A variant of this method enables the simulation of nested indicators, either exactly or approximately (Section 7.7.1). Alabert (1987b) and Journel (1989) use it to simulate variables with a continuous marginal distribution. To do this, we define nested indicators associated with increasing thresholds of the variable under study (e.g., the deciles), and we simulate these indicators—in fact this is a cosimulation. At any given point we do not obtain a specific value but the interval between two successive thresholds to which this value belongs. This is transformed into a specific value by random selection in the interval. According to Gómez Hernández and Cassiraga (1994), this approximation does not yield a satisfactory representation of long distribution tails.

The statistical properties of the simulation are independent of the order in which the points x_i , $i > M$ are scanned. However, when i becomes large (more than a few hundred points), the kriging system cannot be solved in a satisfactory manner. Then conditioning is done using the nearest points x_j , $j < i$. In order to limit the artifacts caused by this approximation, it is recommended

to scan the points x_i in semi-random order: first simulate on a coarse grid, then on a finer grid until the final grid is reached, and at each stage select at random the grid points to be simulated (Gómez-Hernández and Cassiraga, 1994); or use a quasi-random sequence (Section 7.10.4).

Special case: Gaussian autoregressive process with an exponential covariance (1D). Among stationary processes this is the only case where the method can be applied easily without any approximation, for any set of data points and of simulated points, provided that the mean is known. Indeed (1) as we consider the Gaussian case, the conditional distribution of Z_i given $\{Z_j : j < i\}$ is normal $(Z_i^*, \sigma_{K_i}^2)$, and (2) due to the screening effect property of the exponential covariance in 1D (Section 2.3.1), the simple kriging estimator Z_i^* only involves the two adjacent data points. Constructing a nonconditional simulation is even simpler: proceeding by increasing values of x_i , Z_i^* simply involves the previously simulated value Z_{i-1} . For example, to simulate a covariance with scale parameter a and sill σ^2 at a regular spacing Δx , we select a first value Z_1 from the marginal distribution of Z , namely a Gaussian with zero mean and variance σ^2 , and then apply the recurrence relation

$$Z_i = \lambda Z_{i-1} + U_i \quad i = 2, 3, \dots \quad (7.2)$$

where $\lambda = e^{-\Delta x/a}$ and where the U_i are independent Gaussian random variables with mean zero and variance $\sigma_K^2 = \sigma^2(1 - \lambda^2)$. Autoregressive processes generalize this construction to a slightly wider class of covariances (Section 7.5.1).

7.2.2. Covariance Matrix Decomposition

In the case of a zero-mean Gaussian vector the sequential simulation in fact provides

$$Z_1 = \sigma^2 U_1 \quad Z_i = \sum_{j=1}^{i-1} \lambda_{ij} Z_j + \sigma_{K_i} U_i \quad (i > 1) \quad (7.3)$$

The λ_{ij} are the simple kriging weights of Z_i from $\{Z_j : j < i\}$, $\sigma_{K_i}^2$ is the associated kriging variance, and the successive U_i are independent standard normal random variables. Reasoning by recurrence, Z_i can also be expressed as

$$Z_i = \sum_{j=1}^i a_{ij} U_j \quad (7.4)$$

The matrix $\mathbf{A} = [a_{ij}]$ is a lower triangular matrix such that $\mathbf{A}\mathbf{A}' = \mathbf{C}$, where $\mathbf{C} = [C_{ij}]$ is the matrix of the $N \times N$ covariances.

Instead of performing a sequential simulation, we can first decompose the final matrix \mathbf{C} into the product $\mathbf{A}\mathbf{A}'$ (Cholesky decomposition; e.g., see Press

et al., 1992, sec. 5.2.9, pp. 96–98), then compute $\mathbf{Z} = \mathbf{A}\mathbf{U}$, where \mathbf{U} is a vector of N independent standard normal random variables. Since \mathbf{A} is lower triangular and \mathbf{A}' upper triangular, this method is known as that of the *LU (lower-upper) decomposition* of the covariance matrix.

To condition by M data $\{Z_i : i = 1, \dots, M\}$, it suffices to calculate first the $\{U_i : i = 1, \dots, M\}$ that allow one to generate these Z_i , namely to solve the system $\mathbf{Z}_0 = \mathbf{A}_0\mathbf{U}_0$, where \mathbf{Z}_0 , \mathbf{A}_0 , \mathbf{U}_0 are the vectors or matrices \mathbf{Z} , \mathbf{A} , \mathbf{U} reduced to their M first lines and/or columns. If we construct several simulations, this calculation needs to be done only once.

This classic method (see Ripley, 1987) has been introduced in geostatistical applications by Davis (1987a). It is applicable as long as the Cholesky decomposition is feasible, that is, in general for at most a few hundred simulated points. For larger numbers of points, numerical problems turn up and computing time can become excessive, since this increases as N^3 . To circumvent this problem, Alabert (1987a) separates the domain to be simulated into several partially overlapping zones and constructs the simulation zone by zone. On the other hand, Dowd and Saraç (1994) propose an extension of the standard algorithm that can raise the limit to about 2000 points.

Other decompositions of the covariance matrix are possible. Davis (1987b) proposes the decomposition $\mathbf{C} = \mathbf{A}\mathbf{A}'$ with a symmetric matrix \mathbf{A} that can benefit from high-performance algorithms when data are on a regular grid, due to the Toeplitz structure (or block Toeplitz in 2 or higher dimensions) of the covariance matrix in this case. For this decomposition, \mathbf{A} is called the *square root of the covariance matrix*. Dietrich and Newsam (1995) propose an improvement to Davis's algorithm. This method permits direct conditioning when there are few data points.

7.3. CONDITIONING BY KRIGING

In the rest of this chapter several simulation algorithms are presented. Most of them do not provide a direct construction of the conditional simulations. However, there are general methods that can be used to transform nonconditional simulations into conditional ones

7.3.1. Conditioning on the Data

Let us first assume that we know how to construct a nonconditional simulation, that is, a realization of a random function that has the same (generalized) covariance as the studied phenomenon but is otherwise totally unrelated to the sample data. The object of the present section is to show how to pass from a nonconditional simulation to a conditional simulation which, while retaining the structural features of the former, is calibrated on the sample data.

Consider an RF $Z(\cdot)$ known at N sample points x_α , $\alpha = 1, 2, \dots, N$. Let us assume that we have a nonconditional simulation $S(\cdot)$ independent of $Z(\cdot)$,

with the same covariance as $Z(\cdot)$. "Conditioning" is the operation by which we can pass from $S(\cdot)$ to a simulation $T(\cdot)$ that matches the sample points.

The principle, due to G. Matheron, is quite simple. Let $Z^*(x)$ denote the kriging estimator of $Z(x)$ at the point x based on the data $Z(x_\alpha)$, and let us start from the trivial decomposition

$$\begin{aligned} Z(x) &= Z^*(x) + [Z(x) - Z^*(x)] \\ \text{true value} &= \text{kriging estimator} + \text{kriging error} \end{aligned}$$

The kriging error is of course unknown since $Z(x)$ is not known. Now consider the same equality for $S(x)$, where $S^*(x)$ is the kriging estimator obtained as if the simulation were known only at the sample points x_α

$$S(x) = S^*(x) + [S(x) - S^*(x)]$$

This time the true value $S(x)$ is known and so is the error $S(x) - S^*(x)$. Hence the idea of substituting, in the decomposition of $Z(x)$, the unknown error by the simulation of this error, as shown in Figure 7.4; this gives $T(x)$ defined by

$$\begin{aligned} T(x) &= Z^*(x) + [S(x) - S^*(x)] \\ \text{conditional} &= \text{kriging} + \text{simulation of} \\ \text{simulation} &= \text{estimator} + \text{kriging error} \end{aligned} \quad (7.5)$$

Since kriging is an exact interpolator, at a sample point we have $Z^*(x_\alpha) = Z(x_\alpha)$ and $S^*(x_\alpha) = S(x_\alpha)$, so that $T(x_\alpha) = Z(x_\alpha)$.

There remains to verify that $T(\cdot)$ preserves the (generalized) covariance of $Z(\cdot)$ and $S(\cdot)$. It is obvious in the case of simple kriging: since then $Z(x) - Z^*(x)$ is uncorrelated with $Z^*(x')$ for all x' (Section 3.3.3), one has

$$\text{Cov}(Z(x), Z(x')) = \text{Cov}(Z^*(x), Z^*(x')) + \text{Cov}(Z(x) - Z^*(x), Z(x') - Z^*(x'))$$

Since S and Z are independent, the covariance of T is also the sum of the covariance of Z^* and that of $S - S^*$, which equals that of $Z - Z^*$. Hence

$$\text{Cov}(T(x), T(x')) = \text{Cov}(Z(x), Z(x'))$$

A general proof in the framework of IRF- k is presented by Delfiner (1976): any allowable linear combination has the same variance when applied to Z , S , or T .

Note, and this will be important, that in (7.5) nonconditional simulations $S(x)$ are only involved through kriging errors and therefore may be selected with mean zero.

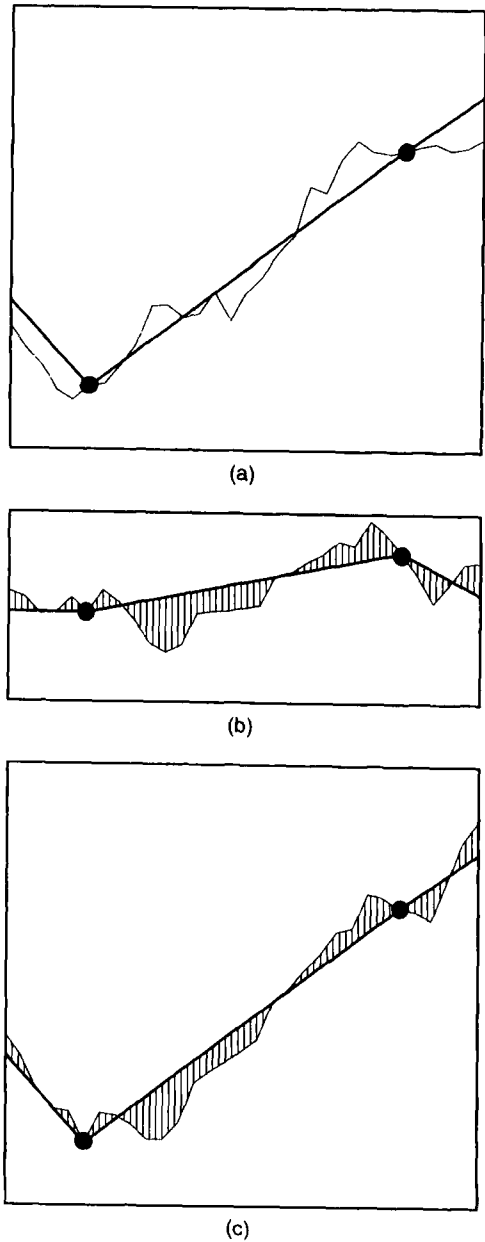


FIGURE 7.4. Conditioning a simulation: (a) real curve (unknown), sample points and kriging; (b) nonconditional simulation (known), sample points, and simulation of the kriging error; (c) kriging errors are picked from the simulation and added to the kriged curve.

When considered conditional on the $Z(x_\alpha)$, $T(x)$ is no longer stationary. The method of construction (7.5) entails that

$$E[T(x) | Z(x_\alpha) : \alpha = 1, \dots, N] = Z^*(x)$$

$$\text{Var}[T(x) | Z(x_\alpha) : \alpha = 1, \dots, N] = \text{Var}[S(x) - S^*(x)] = \sigma_K^2(x)$$

The mean of a large number of independent conditional simulations at a given point tends to the kriging estimate, and their variance tends to the kriging variance. In figurative terms, a conditional simulation “vibrates” in between the data points within an envelope defined by the kriging standard error.

A conditional simulation is meant to behave like the real field but not to estimate it. As an estimator of $Z(x)$ a simulation $T(x)$ would perform very poorly, with a variance twice the kriging variance; indeed

$$T(x) - Z(x) = [Z^*(x) - Z(x)] + [S(x) - S^*(x)]$$

so that

$$E[T(x) - Z(x)]^2 = 2\sigma_K^2(x)$$

A conditional simulation is indeed, in the words of the poet Jean Cocteau, “a lie which always tells the truth”: it lies about the values assumed by the actual realization in between the data points, but it tells the truth about what a realization should look like.

All these properties of conditional simulations hold when the data are subject to nonsystematic measurement errors provided that kriging is done in the same manner for $Z(x)$ and for the nonconditional simulation; that is, an error term with the same variance (and the same covariance if the errors are not independent) is added to the $S(x_\alpha)$ when calculating $S^*(x)$. In this case the kriging system (3.36) including error variances and covariances must of course be used. The method can obviously be extended to the multivariate case by using cokriging instead of kriging.

The proof that the substitution of errors preserves the covariance is valid for kriging (or cokriging) with a global neighborhood. If the number of data is too large, it is necessary to use local neighborhoods. A careful design of the neighborhood search algorithm is needed to avoid the introduction of spurious discontinuities due to neighborhood changes, since these discontinuities cannot be easily distinguished from the normal jitter of most simulations. It is, for example, advisable to use large overlapping neighborhoods.

By a terminology introduced by Journel (e.g., 1989), data that are specified by a single value are often called *hard data*. Conditioning by *soft data*, namely data defined by inequalities or by imprecise values with a given a priori distribution, is more difficult but possible. This problem is commonly met when simulating indicator functions, and solutions are therefore presented in Sections 7.7.1 and 7.7.3.

7.3.2. Matching the Histogram

Many nonconditional simulation methods amount to a weighted moving average of a large number of independent random numbers and produce (approximately) Gaussian simulations. But the data are not obligated to be Gaussian! If the variable has a drift, it constitutes the dominant feature of the phenomenon, and the simulation reproduces it fairly well thanks to conditioning; residuals may not have the correct distribution but this correct distribution is not known and quite difficult to infer from the data. By contrast, in the stationary case the marginal distribution of the variable is an important feature, which can be determined empirically. This distribution is not necessarily reproduced by conditioning. For example, there may be a significant amount of negative values, even though the studied variable is necessarily positive.

The usual means to restore the observed histogram, first used by Journel (1974a, 1974b), is to work with the random function $Y(x)$ obtained by Gaussian transformation of the variable of interest $Z(x)$, defined in Section 6.2.3. Let φ be the transformation expressing $Z(x)$ as a function of $Y(x)$: $Z(x) = \varphi(Y(x))$. The successive steps of the generation of a conditional simulation are the following (Fig. 7.5):

1. Transformation of the $Z(x_\alpha)$ data into $Y(x_\alpha)$ by the inverse transformation $Y(x_\alpha) = \varphi^{-1}(Z(x_\alpha))$.
2. Structural analysis of the $Y(x_\alpha)$ data, or, better, joint structural analysis of the $Y(x_\alpha)$ and $Z(x_\alpha)$ data, to obtain the variogram of $Y(x)$.
3. Nonconditional simulation of $Y(x)$, leading to $S_Y(x)$, using a Gaussian simulation method.
4. Conditioning of $S_Y(x)$ on the Gaussian data $Y(x_\alpha)$, leading to $T_Y(x)$.
5. Application of the transformation $T_Z(x) = \varphi(T_Y(x))$.

The transform φ can be defined graphically from the empirical distribution, as shown in Section 6.2.3. An expansion with Hermite polynomials can also be used (see Section 6.3.4).

If $Y(x)$ is a Gaussian RF and conditioning is done by simple kriging, $T_Y(x)$ is a Gaussian RF. The resulting simulation $T_Z(x)$ is conditioned on the $Z(x_\alpha)$ and has the specified marginal distribution. This procedure is perfectly suited to any RF that can be considered as the image of the transform of a Gaussian RF. The reproduction of the histogram is only approximate if just the marginal distribution of $Y(x)$ is Gaussian.

One notices that the nonconditional simulation and the conditioning concern the Gaussian RF $Y(x)$ and not $Z(x)$. Therefore structural analysis must be done on $Y(x)$. Another possibility is to interchange steps 4 and 5, namely to transform the nonconditional simulation $S_Y(x)$ into a nonconditional simulation $S_Z(x) = \varphi(S_Y(x))$ and condition it on the $Z(x_\alpha)$, using of course the variogram of $Z(x)$. But this is not recommended, since (1) the optimization criterion of kriging (minimum mean square error) is well adapted to Gaussian RFs and

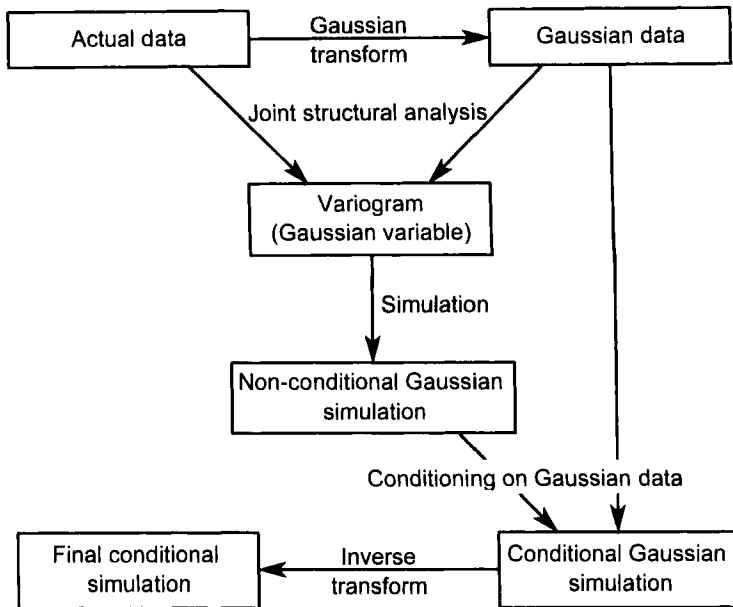


FIGURE 7.5. Flowchart of the various stages of construction of a simulation matching the histogram, the variogram, and the data.

(2) conditioning generally does not preserve the histogram when the studied random function and its nonconditional simulation are not Gaussian random functions.

We will conclude on this subject with three practical remarks:

1. Since we are using techniques that imply the framework of Gaussian SRFs, it is advisable to at least check, using the methods of Section 6.3.4, that the bivariate distributions of the transformed data $Y(x_\alpha)$ can be considered Gaussian:
 - The scatterplot of $(Y(x_\alpha), Y(x_\alpha + h))$, for fixed h , ought to be elliptical.
 - The variograms of order 1 and of order 2 (i.e., the usual variogram) of $Y(x)$ ought to satisfy $\gamma_1(h)/\gamma_1(\infty) = \sqrt{\gamma(h)/\gamma(\infty)}$.
 - The variograms of $Z(x)$ and $Y(x)$ ought to be in a relationship that depends on φ .
2. If the distribution of $Z(x)$ is not continuous, φ remains defined, but its inverse is not defined. For example, if the distribution of $Z(x)$ has a proportion $p_0 > 0$ of values equal to zero, the Gaussian value associated to a zero value of $Z(x_\alpha)$ is only known to belong to the interval $]-\infty, y_0]$, where $y_0 = G^{-1}(p_0)$. Conditioning Gaussian simulations by

such inequalities can nevertheless be achieved using the method presented in Section 7.7.3.

3. By construction, the Gaussian data $Y(x_0)$ have zero mean and unit variance. This variance is a sample variance and is thus smaller than the theoretical variance of the random function $Y(x)$ (cf. Section 2.1.2). As a consequence the variogram of $Y(x)$ is expected to have a sill greater than 1, especially when the range is large. It can result in misnormalization of the nonconditional simulation. This defect is less pronounced after conditioning. However, if it persists, this defect may be amplified by the transformation. A technique is to apply an affine correction to the simulation of the Gaussian kriging error $S_Y(x) - S_Y^*(x)$ in order to restore unit variance of $T_Y(\cdot)$. But this technique also modifies the variogram at the origin. There is no perfect solution.

7.3.3. Probability-Field Simulation

We make a digression to present a simulation technique whose statistical properties are not established but which is used for its speed and practicality. The method is based on the well known fact that to draw a random number from the distribution with c.d.f. F it suffices to draw a uniform random number p between 0 and 1 and compute $z = F^{-1}(p)$. Letting F and z depend on location x we have the relation $z(x) = F_x^{-1}[p(x)]$. When the local distributions F_x are known it is therefore equivalent to simulate $Z(x)$ or the probability field (*p-field*) $P(x)$.

Ideally these local F_x are the conditional distributions given hard data and possibly also soft data. In any case $P(x)$ is not stationary. The trick is to simulate a surrogate *stationary* *p-field*. If Z is a stationary RF the natural candidate is the uniform transform $U(x) = F[Z(x)]$ which hopefully has the same important spatial features of continuity and anisotropy as $P(x)$. The simulation algorithm comprises two steps: (1) a nonconditional simulation of the surrogate *p-field* $U(x)$, (2) the transformation $T(x) = F_x^{-1}[U(x)]$. The second step achieves automatic conditioning on hard data, since at a data point F_x is concentrated on a single value and F_x^{-1} always returns that value. By design the algorithm also reproduces the univariate distributions F_x . The spatial correlation built in $P(x)$ induces spatial correlation in the simulated $Z(x)$, which is the effect sought, but does not reproduce the nonconditional nor the conditional covariance of Z .

The covariance of the uniform transform $U(x)$ can be calculated theoretically by formula (6.23) in the case of an isofactorial model. In particular, if the bivariate distributions of Z are Gaussian the covariance of $U(x)$ is related to the correlogram $\rho(h)$ of Z by the simple formula

$$\text{Cov}[U(x), U(x+h)] = \frac{1}{2\pi} \arcsin[\tfrac{1}{2}\rho(h)]$$

In the interval of variation $[-0.5, +0.5]$ the arcsine function differs little from a straight line so that, up to a scaling factor, the covariance of U is nearly the same as $\rho(h)$, and *not* smoother as might have been expected. In fact if the distributions F_x are Gaussian with means $\mu(x)$ and standard deviations $\sigma(x)$, there is no need to go through $U(x)$ and we can simulate directly

$$T(x) = \mu(x) + \sigma(x)Y(x)$$

where $Y(x)$ is a standard Gaussian field with correlogram $\rho(h)$. Here p -field simulation amounts to a local modulation of a stationary error field. This reveals two limitations of the method. First, the covariance of T

$$\text{Cov}[T(x), T(x+h)] = \sigma(x)\sigma(x+h)\rho(h)$$

may not reflect the complexity of the nonstationarity pattern of the conditional covariance of Z . Second, when there are only hard data, the simple kriging estimate and standard deviation define $\mu(x) = z^*(x)$ and $\sigma(x) = \sigma_K(x)$; but considering the typical behavior of standard deviation maps, with cusps at data points, passing on their shape to simulations is not really desirable.

The p -field simulation technique has been proposed by Srivastava (1992) and Froidevaux (1993) and is also used in the context of sequential simulation by Deutsch and Journel (1998). We see its value as a way to introduce spatial correlation in Monte-Carlo studies in which the intervals of variation are defined from external knowledge.

7.4. TURNING BANDS

Some covariance models may be simulated directly in \mathbb{R}^n . But it is often simpler to use the turning bands method which enables the construction of simulations in space from simulations on lines. Thus methods of simulation on a line are of prime interest even if the final goal is to simulate in 2D or 3D. The turning bands method was first used by Chentsov (1957) in the special case of Brownian random functions. The general principle of the method appears as a remark in Matérn (1960, p. 16), but its development for simulations is due to Matheron (1973a).

7.4.1. Presentation of the Method in the Plane

The method is usually employed for SRFs but can be used for IRF- k as well and is presented here in this context. The turning bands method consists in adding up a large number of independent simulations defined on lines scanning the plane. Up to a scale factor, the value of the simulation at a point

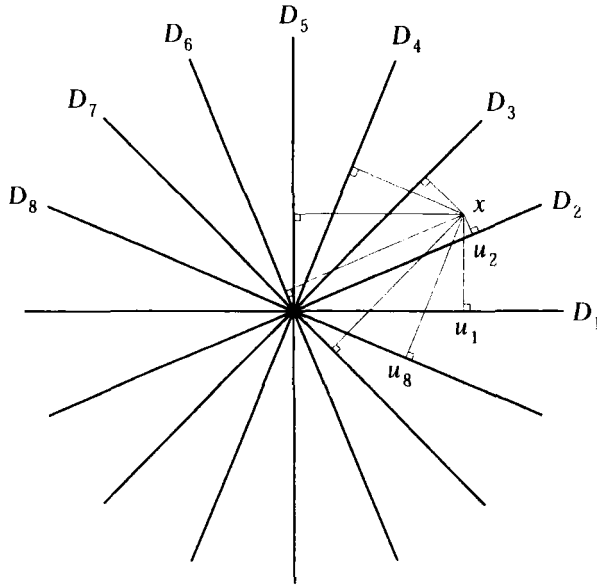


FIGURE 7.6. The principle of turning bands in 2D.

x of the plane is the sum of the values assumed at the projections of x on the different lines by the corresponding one-dimensional simulations. More specifically, consider a system of n_D lines emanating from the origin of space and scanning the plane regularly (Fig. 7.6); the angle between two adjacent lines is π/n_D . We denote by

- $\theta_i \in [0, \pi[$ the angle of the line D_i with the x -axis;
- u_i the unit vector of D_i , with components $\cos \theta_i$ and $\sin \theta_i$;
- s_i the abscissa on D_i , centered at the origin.

Independent zero-mean nonconditional simulations $S_i(s_i)$ with (generalized) covariance $K_1(h)$ are associated to the lines D_i .

Let us consider a point $x = (x, y)$ in the plane. Its projection on D_i is a point with abscissa

$$s_i = \langle x, u_i \rangle = x \cos \theta_i + y \sin \theta_i$$

The simulation at x is then defined by

$$S(x) = \frac{1}{\sqrt{n_D}} \sum_{i=1}^{n_D} S_i(s_i) \quad (7.6)$$

The elementary simulations being independent and with the same (generalized) covariance $K_1(h)$, the generalized covariance of $S(x)$ is

$$K_2(h) = \frac{1}{n_D} \sum_{i=1}^{n_D} K_1(\langle h, u_i \rangle)$$

If the number of lines is large enough, the discrete sum is an approximation of the integral

$$K_2(h) = \frac{1}{\pi} \int_0^\pi K_1(|h| \cos \theta) d\theta \quad (7.7)$$

In practice, K_2 is given, and we have to invert relation (7.7) to obtain K_1 . Brooker (1985) has shown that for $h \geq 0$ K_1 is given by

$$K_1(h) = K_2(0) + \int_0^h h(h^2 - u^2)^{-1/2} \frac{d}{du} K_2(u) du \quad (7.8)$$

In particular, the usual spherical covariance (of \mathbb{R}^3) with sill C and range a is obtained by using

$$K_1(h) = \begin{cases} C \left[1 - \frac{3\pi}{4} \left(\frac{|h|}{a} - \frac{1}{2} \frac{|h|^3}{a^3} \right) \right] & \text{if } |h| \leq a \\ C \left[1 - \frac{3}{2} \left(\frac{|h|}{a} - \frac{1}{2} \frac{|h|^3}{a^3} \right) \arcsin \left(\frac{a}{|h|} \right) - \frac{3}{4} \frac{h^2}{a^2} \sqrt{1 - \frac{a^2}{h^2}} \right] & \text{if } |h| \geq a \end{cases}$$

Note that unlike K_2 , the covariance K_1 reaches zero only asymptotically when $|h| \rightarrow \infty$. For other models see Gneiting (1998).

In principle, random directions could be used instead of regular directions. Theoretically it even suffices to take a single line with an angle drawn at random from a uniform distribution over $[0, \pi[$ to obtain the covariance (7.7) exactly. However, such procedure will only ensure the correct covariance on average over the ensemble of simulations, while none in particular would have the desired properties. For example, simulations based on a single line exhibit a zonal anisotropy, even though the parent covariance model $K_2(h)$ is isotropic (see Fig. 7.3b). For a fixed value of the number of lines, it is advisable to use a regular system of lines rather than random directions, for it ensures better second-order ergodicity.

In practice, the simulations along each line are often discretized so that the same value $S_i(s_i)$ is assigned to a whole “band” perpendicular to D_i and containing s_i . Hence the name turning “bands” given to the method.

Note that as can be seen in Figure 7.6, $S(x)$ at point x integrates the contribution of the lines at their intersections with the circle of diameter $|x|$ going

through the origin and the point x . Thus the turning bands algorithm can be seen as a back-projection. Its formalism is similar to the formalism of Radon transforms used in tomography.

7.4.2. Generalization to n -Dimensional Space

The easiest way for generalizing the method is to take a simulation $S_1(s)$ with (generalized) covariance $K_1(h)$ on a line and to define the direction of this line in \mathbb{R}^n by a random unit vector u , namely by a point on the unit sphere. The simulation at a point x of \mathbb{R}^n is then defined by

$$S(x) = S_1(\langle x, u \rangle)$$

where the inner product $\langle x, u \rangle$ stands for the projection of x onto the line. In practice, just as in the plane, several lines are used rather than a single one, and since opposite directions play the same role, the directions are selected over the unit half-sphere (see next section).

Taking the expectation with respect to the random direction u the (generalized) covariance of $S(x)$ is

$$K_n(h) = \frac{1}{S_n} \int K_1(\langle h, u \rangle) du$$

where the integration extends over the unit sphere of \mathbb{R}^n and where S_n is the surface area of this sphere. With this procedure, the covariance $K_n(h)$ is clearly isotropic though all realizations of $S(x)$ are anisotropic. Using the same notation K_n to denote this covariance as a function of $r = |h|$ we have explicitly

$$K_n(r) = \frac{2}{\sqrt{\pi}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \int_0^1 K_1(vr) (1-v^2)^{(n-3)/2} dv \quad (7.9)$$

where Γ is the gamma function (A.1). If $K_1(h)$ and $K_n(h)$ have spectral densities $f_1(u)$ in \mathbb{R}^1 and $f_n(u)$ in \mathbb{R}^n , respectively, f_1 and f_n are isotropic and can be considered as functions of $\rho = |u|$. Using the same notation f_n to represent the spectral density as a function of ρ , (7.9) can be translated in spectral terms

$$f_n(\rho) = \pi^{-n/2} \Gamma\left(\frac{n}{2}\right) \rho^{1-n} f_1(\rho)$$

Similar relations are obtained for the spectral measure of isotropic generalized covariances of IRF- k . Relation (7.9) has a very simple form when $n = 3$

$$K_3(r) = \int_0^1 K_1(vr) dv = \frac{1}{r} \int_0^r K_1(u) du \quad (7.10)$$

Formula (7.9) associates an isotropic covariance K_n in \mathbb{R}^n to every covariance K_1 in \mathbb{R}^1 . In fact this formula can be identified with the general form of an isotropic covariance in \mathbb{R}^n (Matérn, 1960; Matheron, 1973a). Thus the correspondence between K_1 and K_n is one to one, and (7.9) can be inverted. As seen in Section 7.4.1, the inversion of (7.9) is not easy for $n = 2$ (and similarly for n even). When $n = 3$, the inverse formula of (7.10) is particularly simple (and generally so when n is odd):

$$K_1(r) = \frac{d}{dr}[rK_3(r)] \quad (7.11)$$

If K_3 has a finite range a (i.e., $K_3(r) = 0$ when $r > a$), K_1 has the same property. This is generally not the case for the covariance K_1 associated with a covariance K_2 with finite range, as can be seen in (7.8). Since it is often easier to simulate a covariance when its range is finite, it may be better to generate a 2D simulation as the planar section of a 3D simulation than to simulate directly in 2D (this is of course limited to the case where the covariance K_2 is also a valid model in \mathbb{R}^3).

The correspondence between K_1 and K_n is also particularly simple, for all n , in the case of a generalized covariance of type r^α : the $K_1(vr)$ term in the integral (7.9) is factorized as $v^\alpha r^\alpha$ and K_n is of the form $B_{n\alpha} r^\alpha$ with

$$B_{n\alpha} = \frac{\Gamma(n/2)\Gamma((\alpha+1)/2)}{\sqrt{\pi}\Gamma((n+\alpha)/2)} \quad (7.12)$$

Likewise, if K_1 is a sum of $A_\alpha r^\alpha$ terms, K_n is a sum of $A_\alpha B_{n\alpha} r^\alpha$ terms. In particular, a polynomial generalized covariance in \mathbb{R}^1 generates by turning bands a polynomial generalized covariance in \mathbb{R}^n .

Note a straightforward double generalization: if u is selected from a given distribution $\mu(du)$ rather than with a uniform distribution over the unit half-sphere, and if the one-dimensional simulation S_1 has a covariance K_u that depends on u , the resulting simulation has an anisotropic covariance of type (2.67):

$$K(h) = \int K_u(\langle h, u \rangle) \mu(du)$$

7.4.3. Efficient Selection of the Line Directions

As we have just seen, a single line of random direction produces, on the average, the desired covariance. However, such a simulation is constant in all hyperplanes orthogonal to the line so that its sample covariance is not that of the model: the simulation is not ergodic in the covariance. In practice, as in (7.6) we take the normalized sum of n_D elementary simulations corresponding to n_D lines of regularly distributed or random directions. Regarding ergodicity,

for fixed n_D it is preferable to use lines with directions distributed as regularly as possible rather than randomly.

In 2D, the choice of lines is simple: discretize the directions according to $\theta_j = (j-1)\pi/n_D$, $j = 1, \dots, n_D$; one may prefer a stratified random grid and select θ_j randomly in the interval $[(j-1)\pi/n_D, j\pi/n_D]$. One should take n_D large enough (from one to a few hundred).

In 3D, the maximum number of regularly spaced directions is 15. These directions are defined by the lines joining opposite edges of a icosahedron (see Note 1 in Section 2.2.2). The first 3D simulations by turning bands were constructed from 1D simulations on these 15 lines. But these 15 special directions show up on the realizations obtained, for example, when we observe a 2D section of these simulations (Chilès, 1977). To go beyond 15 lines, one can take random directions, or several groups of 15 lines. But drawing independent random directions from a uniform distribution on the half-sphere does not provide a very regular discretization (Fig. 7.7a). Quasi-random sequences give better results in this respect (Fig. 7.7b). Freulon and de Fouquet (1991) recommend the following method:

1. Simulate points with a uniform distribution on $[0,1] \times [0,1]$ using a quasi-random sequence (they use the Van der Corput sequence in two dimensions; cf. Section 7.10.4).
2. To each point with components U_1 and U_2 , associate the direction of unit vector $(\cos(2\pi U_1)\sqrt{1-U_2^2}, \sin(2\pi U_1)\sqrt{1-U_2^2}, U_2)$, which will then be uniformly distributed over the unit half-sphere.

It is recommended to use one to several hundred lines.

Finally, when using the turning bands method, the problem of simulating $S(x)$ in \mathbb{R}^n boils down to the following two problems:

1. Find the covariance $K_1(h)$ defined in \mathbb{R}^1 associated with the isotropic covariance $K_n(r)$ to be obtained in \mathbb{R}^n ; Table A.2 gives the covariances $K_1(h)$ associated with the most common models $K_3(r)$.
2. Simulate random functions on lines with the given covariance $K_1(h)$.

For example, to simulate the exponential covariance $K_3(r) = \sigma^2 \exp(-r/a)$ in \mathbb{R}^3 , we have to simulate

$$K_1(h) = \sigma^2 \left(1 - \frac{|h|}{a}\right) \exp\left(-\frac{|h|}{a}\right)$$

on the lines. We recognize, up to a factor 2 for the scale parameter, the covariance (2.74) of the migration process presented in Section 2.5.3, which can be simulated very easily. This method of construction is specific to this particular model.

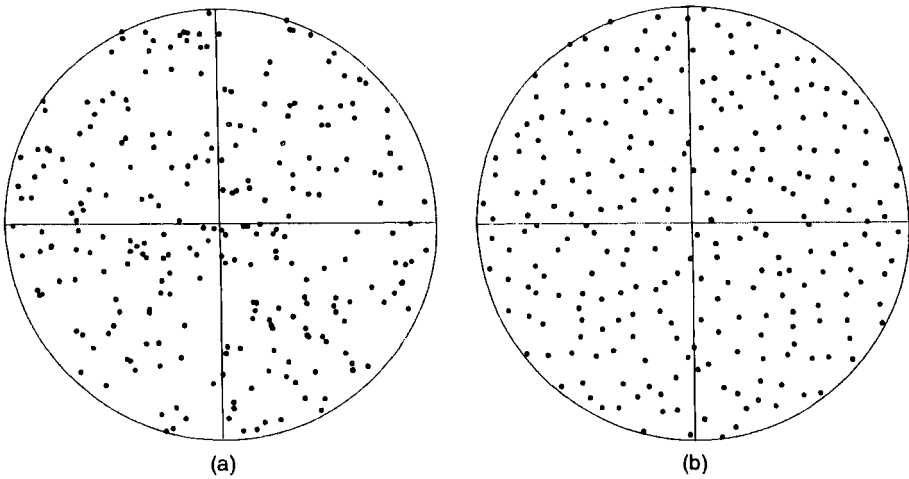


FIGURE 7.7. Schmidt equiareal projection of 256 points of the unit half-sphere, representing 256 directions in space: (a) random points; (b) quasi-random points derived from an equidistributed Van der Corput sequence.

7.5. NONCONDITIONAL SIMULATION OF A CONTINUOUS VARIABLE

We now review the main simulation methods. Some may be used directly in \mathbb{R}^n , others only in \mathbb{R}^1 , but they may be extended to \mathbb{R}^n by turning bands. We present them in the univariate case. The linear model of coregionalization of Sections 5.6.5 and 5.6.6 is well suited for the simulation of a p -variate RF: to simulate its components $\{Z_i(x) : i = 1, \dots, p\}$ it suffices to simulate the factors $Y_{jk}(x)$ independently and combine them linearly according to (5.37).

7.5.1. Autoregressive and Moving-Average Models

The two methods allowing a direct construction of conditional simulations (sequential simulation and decomposition of the covariance matrix), when applied to the simulation of Gaussian RFs, differ essentially in presentation. The first one amounts to the application of (7.3) and is of autoregressive type, whereas the second, corresponding to (7.4), is based on moving averages. These methods are not applicable when the number of points is large. But in the case of stationary discrete processes, they can be adapted to summations over a finite number of terms. Box and Jenkins (1976) have developed ARMA models with a view to fitting their parameters directly from time series data, the covariance function being derived. We present these models from our slightly different perspective: the covariance of the phenomenon is known, from a structural analysis, and the parameters of the ARMA model are derived.

1D Autoregressive Process

Consider points at a regular interval Δx in 1D space, and a “discrete-time” process $Z_t = Z(t \Delta x)$, t negative or positive integer. This definition recalls that we consider Z_t as a discrete version of a random process $Z(x)$ defined over the whole line. We will only consider processes that are stationary and have mean zero. Due to the orthogonality property of simple kriging, Z_t may be considered as the sum of the simple kriging estimator from all past values $Z_{t'}$, $t' < t$, and an uncorrelated random variable a_t with mean zero and variance σ_K^2 (the kriging variance)

$$Z_t = \sum_{i=1}^{\infty} \alpha_i Z_{t-i} + a_t \quad (7.13)$$

The process is said to be an autoregressive process of order p , or an $AR(p)$ process, when only the first p weights α_j of (7.13) are nonzero

$$Z_t = \alpha_1 Z_{t-1} + \cdots + \alpha_p Z_{t-p} + a_t \quad (7.14)$$

This process has a memory of finite length p and evolves due to random innovations a_t (the a_t are uncorrelated). Z_t has the form of a regression on its own previous values, hence the name autoregressive. This model generalizes the algorithm (7.2) presented for simulating a stochastic process with an exponential covariance.

The coefficients $\alpha_1, \dots, \alpha_p$ of the $AR(p)$ process are estimated very easily using the covariance $C(h)$ of the random function $Z(x)$ for $h = i \Delta x$ and $|i| = 0, 1, \dots, p$. Indeed, multiplying both sides of (7.14) by Z_{t-i} , and denoting $C_i = C(i \Delta x)$ the covariance of Z_{t-i} and Z_t , gives

$$C_i = \alpha_1 C_{i-1} + \alpha_2 C_{i-2} + \cdots + \alpha_p C_{i-p} \quad (7.15)$$

Doing the same for all $i = 1, 2, \dots, p$ yields the p linear equations of the simple kriging system

$$\sum_{j=1}^p \alpha_j C_{i-j} = C_i \quad (i = 1, 2, \dots, p) \quad (7.16)$$

These equations are also known as the *Yule-Walker equations*. The matrix of the system has a special pattern: the elements on symmetric diagonals are identical. Such a matrix is called a symmetric Toeplitz matrix, and an especially efficient algorithm, the Levinson algorithm, may be used to solve system (7.16) (e.g., Press et al., 1992).

The problem, when the covariance function is given, is whether it can be considered as that of an $AR(p)$ process. In practice, the choice of p is made by considering the partial correlation ϕ_k , defined as the k th value α_k obtained from the Yule-Walker equations with k terms. ϕ_k can be interpreted as the kriging

weight assigned to Z_{t-k} when kriging Z_t from Z_{t-1}, \dots, Z_{t-k} . The covariance terms C_i are those of an $AR(p)$ model if $\phi_k = 0$ for $k > p$. In practice, if ϕ_k does not reach zero, p is chosen such that $|\phi_k| < \varepsilon$ for $k > p$, where ε is a given threshold.

The knowledge of C_0, C_1, \dots, C_p determines the coefficients $\alpha_1, \dots, \alpha_p$. These in turn determine C_i for $i > p$ by iterative application of relation (7.15). Box and Jenkins (1976) show that the general form of the covariance is then

$$C_i = \sum_{j=1}^p A_j (G_j)^i \quad (i \geq 0)$$

where $G_1^{-1}, G_2^{-1}, \dots, G_p^{-1}$ are the complex roots of the characteristic equation

$$1 - \sum_{i=1}^p \alpha_i z^i = 0 \quad (z \in \mathbb{C})$$

and A_1, A_2, \dots, A_p are complex coefficients (which depend on the covariance).

In the case of an SRF, the G_i satisfy $|G_i| < 1$. Therefore the covariance is a mixture of damped exponentials (G_i real) and damped sine waves (pair of conjugate complex roots). For example, if $p = 1$, one has

$$\frac{C_i}{C_0} = \left(\frac{C_1}{C_0} \right)^i = \alpha_1^i$$

that is, an exponential model if $\alpha_1 > 0$ and a damped sine wave if $\alpha_1 < 0$ (in both cases $|\alpha_1| < 1$). Thus the method is especially suited for exponentially decreasing covariances. For example, the hole effect model $C(h) = \exp(-|h|/a) \cos(2\pi h/a)$ can be simulated at the spacing $\Delta x = a/10$ with an error less than 1% with $p = 7$ only. The method can also be used for models with a finite range, but it is usually less efficient than the moving-average method: for example, the spherical model with a range a can be reproduced at the spacing $\Delta x = a/10$, with an error less than 1%, with $p = 13$, whereas it can be simulated exactly as a moving-average model with 10 terms.

An $AR(p)$ model can be simulated with a sequential algorithm:

1. Start with a random value Z_0 , drawn from the marginal distribution of $Z(\cdot)$.
2. Simulate iteratively Z_1, Z_2, \dots, Z_{p-1} by applying, respectively, the $AR(1)$, $AR(2), \dots, AR(p-1)$ recursions associated with the covariance of Z .
3. Simulate iteratively Z_p, Z_{p+1} , and so on, by applying the $AR(p)$ recursion.

The random variables a_i are usually chosen Gaussian so that the simulation is Gaussian (see discussion below).

2D Autoregressive Model

The generalization of definitions (7.13) and (7.14) to 2D is not unique because the notions of “past” and “future” have no general meaning in this space. The simplest generalization of an $AR(p)$ process is defined by considering a regression of Z_{i_1, i_2} on $\{Z_{i_1 - i_1', i_2 - i_2'}, i_1' = 0, \dots, p, i_2' = 0, \dots, p, \text{ except the pair } i_1' = i_2' = 0\}$. Sharp and Aroian (1984) consider the case $p = 1$. Boulanger (1990) develops the method in the general case. Since an $AR(p)$ model has $(p + 1)^2 - 1$ parameters α_{j_1, j_2} , the partial correlation ϕ_k is now defined as the weight of maximum absolute value among the new $2k + 1$ weights that are introduced when increasing the order from $k - 1$ to k . As in 1D, p is chosen such that $\phi_k = 0$ for $k > p$, or at least $|\phi_k| < \varepsilon$ for $k > p$. For example, in the isotropic case an exponential covariance with a scale parameter equal to $10\Delta x$ is reproduced with an error less than 1% with $p = 6$. To obtain the same precision for a spherical model with the same scale parameter (here the range), $p = 12$ is required.

Mignolet and Spanos (1992) propose a slightly different method that gives a good approximation of the desired covariance: the second index is considered as time, and the regression involves only data on preceding columns ($i_1 = -p_1, \dots, 0, \dots, p_1$, and $i_2 = 1, \dots, p_2$).

The extension to 3D seems intractable except in special cases (e.g., factorized exponential covariance; see Section 7.11.2).

1D Moving-Average Process

As previously, we consider points at a regular interval Δx in 1D space, and a process $Z_t = Z(t\Delta x)$, t negative or positive integer. But Z_t is now defined as a weighted moving average of a sequence of independent identically distributed (i.i.d.) random variables a_t with mean $E[a_t] = 0$ and variance $\text{Var}[a_t] = \sigma^2$,

$$Z_t = \sum_{j=-\infty}^{+\infty} \beta_j a_{t-j} \quad (7.17)$$

where the weights β_j are such that $\sum \beta_j^2 < \infty$. Z_t is then a stationary process with covariance

$$C_k = \sigma^2 g_k \quad (7.18)$$

where g_k is the discrete covariogram of the weights

$$g_k = \sum_{j=-\infty}^{+\infty} \beta_j \beta_{j+k} \quad (7.19)$$

Applications deal with the case where the number of nonzero weights is finite. Since the resulting covariance is not altered by a translation of the β_j ,

we may consider that the first nonzero weight is β_0 and that only the next q weights β_1, \dots, β_q are nonzero. Definition (7.17) then becomes

$$Z_t = \beta_0 a_t + \beta_1 a_{t-1} + \dots + \beta_q a_{t-q} \quad (7.20)$$

and the discrete covariogram (7.19) of the weights has a range equal to $q + 1$, in the sense that $g_k = 0$ for $|k| \geq q + 1$.

Since the covariance is not modified if the β_j are multiplied by the same nonzero coefficient c , provided that the variance σ^2 is divided by c^2 , we may also fix $\beta_0 = 1$. We then obtain the moving average process of order q , or MA(q) process, as defined by Box and Jenkins (1976). For a time process the physical interpretation is that at time t "nature" produces a "shock" or "innovation" a_t unrelated to shocks at all other times. These shocks are combined through a filter characterizing the genetic mechanism of the phenomenon, and the output is the observed signal Z_t . In the sequel we will not impose $\beta_0 = 1$.

In our approach, which differs from that of Box and Jenkins, we use such a model for a covariance $C(h)$ that has been already identified. Thus we have to derive q and the coefficients β_j such that the discrete covariance C_k given by (7.18) and (7.19) is equal to $\bar{C}(k \Delta x)$. Unlike for AR models the discrete covariance C_k is not linear in $\beta_0, \beta_1, \dots, \beta_q$ but involves products of these parameters, which makes estimation more difficult. Boulanger (1990) proves two results:

1. If $C(h)$ has finite range a there exists a set of $q + 1$ coefficients β_j such that $C_k = C(k \Delta x)$, where $q = \text{Int}(a/\Delta x)$.
2. If the covariance only tends to zero when h tends to infinity but has a finite integral range, it can be approximated to any prespecified precision by the covariance of an MA(q) model. The order q is taken as $q = \text{Int}(a'/\Delta x)$ where a' is the smallest distance such that $|C(h)| < \varepsilon$ whenever $|h| > a'$, ε being a given threshold.

There remains to find the coefficients β_j . An approximate method is often used when the covariance function is of the form $C(h) = \sigma^2 g(h)$, where $g(h)$ is known to be the covariogram of some function $w(x)$ whose support has length a (see Section 7.5.2). Since a translation of w does not change its autoconvolution function, we may suppose without loss of generality that the support of w is $[0, a]$. As an approximation we can take $\beta_j = |\Delta x|^{1/2} w(j \Delta x)$. This leads to

$$C_{\Delta x; k} = \sigma^2 g_{\Delta x; k}$$

where

$$g_{\Delta x; k} = |\Delta x| \sum_j w(j \Delta x) w(j \Delta x + k \Delta x) \quad (7.21)$$

is the discrete approximation of $g(k \Delta x)$ from the values of w at interval Δx . The discrete covariogram $g_{\Delta x, k}$ differs slightly from the discrete covariogram g_k to be reproduced. In particular, its value at the origin is $|\Delta x| \sum_j w(j \Delta x)^2$ instead of $\int w(x)^2 dx$. Thus the simulation is normalized by a scale factor so as to ensure the correct variance. With this correction the simulated and theoretical covariance curves are usually superposable provided that the discretization step Δx is small in comparison with the range. A rigorous method where the coefficients β_j are limits of series is also available (Boulanger, 1990).

In practice, $MA(q)$ models are efficient when q has a low value, namely for covariances with a short range. The simulation of an $MA(q)$ model follows directly from its definition. The random variables a_t are usually chosen Gaussian so that the simulation is Gaussian (see discussion at the end of this Section). Boulanger also proposes a method for direct conditioning based on the fact that data perturb innovations only locally.

2D Moving-Average Model

The generalization of definition (7.17) to 2D or 3D is obvious. But again there is no unique generalization of definition (7.20). Sharp and Aroian (1984) consider, for example, a 2D $MA(1)$ process expressing Z_{t_1, t_2} as a function of a_{t_1, t_2} , a_{t_1-1, t_2} , and a_{t_1, t_2-1} . Boulanger (1990) shows that the 2D generalization of most of the results he obtained for the 1D case need a causal model, namely a lexicographic moving average, where the nonzero $\beta_{j_1 j_2}$ weights are those such that

$$\begin{aligned} j_1 = 0 \quad \text{and} \quad 1 \leq j_2 \leq q \\ 1 \leq j_1 \leq q \quad \text{and} \quad -q \leq j_2 \leq q \end{aligned}$$

With this definition of a 2D $MA(q)$ process, all the results of the 1D case can be carried over to 2D, except that now in 2D, a covariance with a finite range cannot necessarily be represented *exactly* by an $MA(q)$ process with q finite (see also Guyon, 1993, sec. 1.3.3), but as any covariance with a finite integral range, it can at least be *approximated* by such a process.

Spanos and Mignolet (1992) propose a slightly different algorithm where the moving average domain is the rectangle defined by

$$-q_1 \leq j_1 \leq q_1 \quad \text{and} \quad -q_2 \leq j_2 \leq q_2$$

The coefficients are determined so as to minimize an error in the frequency domain.

Mixed Autoregressive Moving-Average Model

There does not necessarily exist an $AR(p)$ or $MA(q)$ model, with p or q finite and not too high, that leads to the desired covariance. Box and Jenkins (1976)

have introduced a mixture of the $AR(p)$ and $MA(q)$ models that provides more flexibility. A mixed autoregressive moving-average process, or $ARMA(p, q)$ process, includes a finite number of both autoregressive and moving-average terms

$$Z_t = \alpha_1 Z_{t-1} + \cdots + \alpha_p Z_{t-p} + a_t + \beta_1 a_{t-1} + \cdots + \beta_q a_{t-q}$$

The importance of this mixed model is that it permits a “parsimonious” representation of a linear process; that is, it requires fewer parameters than a model based on a purely autoregressive or purely moving-average process. The point is obvious if one attempts to represent an $AR(p)$ model as an MA , or an $MA(q)$ model as an AR : this is possible, but an infinite number of terms is required. Mignolet and Spanos (1992) and Spanos and Mignolet (1992) propose a procedure to closely approximate an autoregressive model or a moving-average model, respectively, by an $ARMA$ model with fewer parameters (in 2D). Similarly Samaras et al. (1985) develop the use of $ARMA$ models with $p = q$, including in the multivariate case. Further results can be found in Guyon (1993, ch. I).

Spatial Distribution and Conclusion

The random variables a_t involved in the definition of $ARMA(p, q)$ models are usually chosen Gaussian, so that the simulation is Gaussian. If not, the finite-dimensional distributions of $AR(p)$ models nevertheless tend to be Gaussian by virtue of the central limit theorem (an $AR(p)$ model is equivalent to a moving average model with an infinite number of terms). The finite-dimensional distributions of $MA(q)$ models tend to be Gaussian for the same reason provided that the weights are spread among a large number of terms. The same is true for simulations obtained by turning bands from a large number of 1D $ARMA$ random functions (Fig. 7.8).

The method can be extended to the simulation of random functions that have an infinite variance, or even an infinite mean, notably stable distributions (Boulanger, 1990).

In conclusion, autoregressive moving-average models offer the following advantages:

- The multivariate distribution of the simulations can be exactly Gaussian.
- It is possible to simulate nested structures in a single step.
- The autoregressive method can be directly used in 2D, and the moving-average method in 2D or 3D, without the need for turning bands.
- It can be applied to large grids (for fixed p and q , and once the parameters are known, the computation time is simply proportional to the number of grid nodes).
- In the case of $MA(q)$ models, it is possible to directly produce conditional simulations.

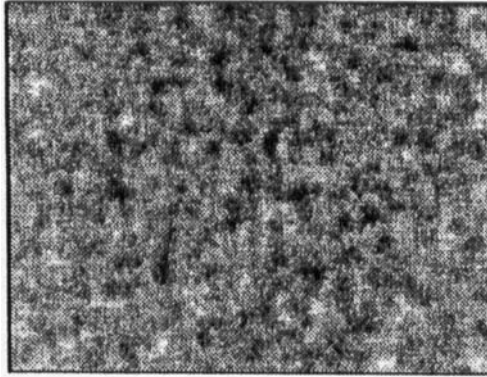


FIGURE 7.8. Simulation of a Gaussian SRF by turning bands and a 1D moving-average method. The variogram is an isotropic spherical model with range 10 pixels. The construction includes 45 turning bands (image: 256×192 pixels). From Boulanger (1990).

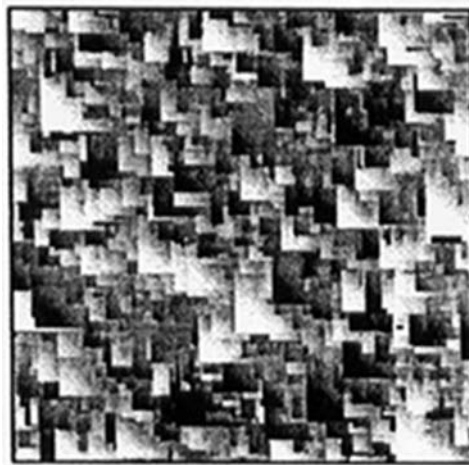


FIGURE 7.9. Simulation of a 2D autoregressive SRF with stable distribution with parameter $\alpha = 0.5$, in isoclass representation containing approximately the same number of pixels in each class. The sample autocorrelation is a factorized exponential model with 3-pixel scale parameter along the two axes (image: 256×256 pixels). From Boulanger (1990).

Their limitations are the following:

- ARMA models cannot fit all covariance models: $AR(p)$ models are for exponentially decreasing covariances, and $MA(q)$ models for covariances whose range is small compared with the simulated domain.
- Autoregressive models may present a spatial dissymmetry in 2D and cannot be used efficiently in 3D (see Fig. 7.9 for a stable random function).
- The simulation is limited to grid nodes.

A correct use of ARMA models requires some effort, but it is worth it. In particular, the moving-average method can be used efficiently in 3D to simulate

very large grids provided that the range is small in comparison with the size of the grid.

7.5.2. Dilution of Poisson Germs

The Poisson point process has long been used to define random function models (e.g., in \mathbb{R}^1 , Blanc-Lapierre and Fortet, 1953, ch. V; and in \mathbb{R}^n , Matérn, 1960, ch. 3). The method outlined here is a dilution of Poisson points considered as germs and corresponds to processes called “moving average models with constant or stochastic weight function” by Matérn and Poisson pulse processes by several authors. Let us first recall the definition and main properties of the Poisson point process.

Poisson Point Process

The Poisson point process in \mathbb{R}^n corresponds exactly to the intuitive idea of points distributed in space “at random”. The Poisson point process with *intensity*, or *density*, λ ($\lambda > 0$) is characterized by the following properties:

1. The number $N(V)$ of points inside a domain V is a Poisson random variable with parameter $\lambda|V|$, where $|V|$ represents the measure of V (length, surface, or volume)

$$\Pr\{N(V) = k\} = e^{-\lambda|V|} \frac{(\lambda|V|)^k}{k!}$$

$$E[N(V)] = \text{Var}[N(V)] = \lambda|V|$$

2. If V_i , $i = 1, 2, \dots, p$, are pairwise disjoint domains, the random variables $N(V_i)$ are mutually independent.

The Poisson point process has an important conditional property that corresponds to the notion of random points: given that the number of points $N(V)$ inside the domain V is equal to n_0 , these points are independently and uniformly distributed over V .

Thus a Poisson point process with intensity λ can be simulated within a bounded domain V as follows:

1. Draw the number of points n_0 from a Poisson distribution with mean $\lambda|V|$.
2. Draw the n_0 points independently from a uniform distribution within V .

This method can be easily applied to intervals of \mathbb{R}^n (segments in \mathbb{R}^1 , rectangles in \mathbb{R}^2 , parallelepipeds in \mathbb{R}^3). Combined with a rejection technique, it allows the simulation of a Poisson point process in a complex domain assumed to be enclosed in a union of pairwise disjoint sets.

In 1D, Poisson points delimit intervals whose lengths are independent random variables with an exponential distribution with mean $\theta = 1/\lambda$. This distribution has the following conditional property: if it is known that the interval length T is larger than a given value t , the residual length $T - t$ has the same distribution as the a priori length T . In particular, the residual length from a fixed origin to the next Poisson point has the same exponential distribution as T . Hence another method for simulating a 1D Poisson point process from an arbitrary origin, which without loss of generality we will place at $x = 0$:

1. Draw intervals T_i , $i = 1, 2, \dots$, as independent random variables from an exponential distribution with mean $\theta = 1/\lambda$.
2. Define the successive Poisson points X_i , $i = 1, 2, \dots$, associated with these intervals by $X_1 = T_1$ and the recurrence relation $X_i = X_{i-1} + T_i$, $i = 2, 3, \dots$.

The Poisson point process can be defined also on $]-\infty, 0]$ by inverting the method of construction. In all cases the origin $x = 0$ does not belong to the set of the Poisson points.

A remark on terminology: the definition of the random set of Poisson points is linked to the definition of the Poisson random process, a process that increases by unit jumps at each Poisson point (see Section 7.6.1), so that in practice the term “Poisson process” is used either for the random set or for the process itself. When we wish to emphasize the difference, we call the random set of Poisson points a “Poisson point process”.

Dilution by a Fixed Function

Let $w(x)$ be a square integrable function on \mathbb{R}^n and $g(h)$ its covariogram (2.30),

$$g(h) = (w^* \check{w})(h) = \int w(x)w(x+h)dx \quad (7.22)$$

For the sake of simplicity, we will assume that $w(x)$ is integrable. We assign a constant “dose” μ to each of the points X_i of a Poisson point process with intensity λ and dilute it by the influence function w (Fig. 7.10). At a point x the sum of the contributions of all X_i defines a random function

$$Z(x) = \mu \sum_i w(x - X_i) \quad (7.23)$$

A variant is to assign independent random doses a_i with mean μ and variance σ^2 , which corresponds to the definition

$$Z(x) = \sum_i a_i w(x - X_i) \quad (7.24)$$

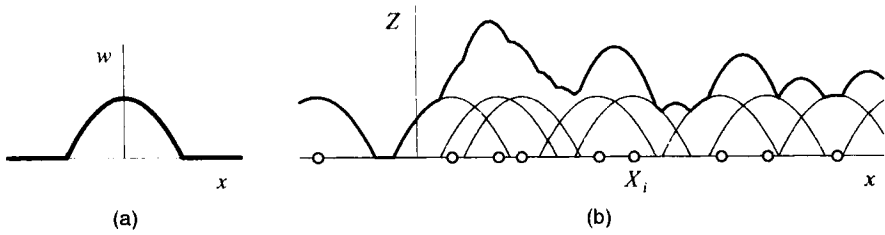


FIGURE 7.10. Construction of a simulation by dilution of Poisson germs: (a) dilution function; (b) Poisson points and the construction of the simulation.

Note that (7.23) can be considered as the special case $\sigma = 0$. With this convention $Z(x)$ is an SRF with mean m and covariance $C(h)$ given in both cases by

$$\begin{aligned} m &= \lambda \mu \int w(x) dx \\ C(h) &= \lambda(\mu^2 + \sigma^2)g(h) \end{aligned} \quad (7.25)$$

Heuristic proofs of these results are easy. A rigorous proof can be found in Blanc-Lapierre and Fortet (1953).

Some results persist with point models other than Poisson points. Consider, for example, the transposition of definition (7.24) to a regular grid of points X_i with a random origin

$$Z(x) = \sum_{i \in \mathbb{Z}^n} a_i w(x - X_0 - i \Delta x) \quad (7.26)$$

where the vector Δx represents the grid spacing and X_0 a random origin within a grid cell centered at the origin of the axes. If $\mu = 0$, it is easy to prove that formula (7.25) remains valid with $\lambda = 1/|\Delta x|$. This model constitutes the generalization of moving-average models to the continuous case. For fixed $X_0 = x_0$, the RF $Z(x)$ is not stationary, and therefore many independent simulations must be added to achieve stationarity. This model is very convenient to build the 1D simulations involved in the use of the turning bands method when the support of w is bounded: by selecting Δx equal to the length of this support, the value of $Z(x)$ at any given x involves only one dilution point (see Lantu  joul, 1994, for the 3D simulation of a spherical covariance by turning bands). The same technique can be applied directly in \mathbb{R}^n when the covariance is the autoconvolution of a function w with bounded support (e.g., the spherical covariance), by taking Δx as the smallest hypercube including the support of w .

Now the problem is the following: given a covariance function $C(h)$, does there exist an integrable function $w(x)$ such that $w^* \check{w} = C$? Since $w(x)$ has

a Fourier transform, which we denote by $\varphi(u)$, $w^*\tilde{w}$ has the Fourier transform $|\varphi(u)|^2$, and therefore the answer is limited to covariances that possess a spectral density $f(u)$ (in the sense of ordinary functions). Any function $w(x)$ whose Fourier transform satisfies $|\varphi(u)|^2 = f(u)$ has $C(h)$ as autoconvolution function, as the following diagram shows, where \mathcal{F} represents the Fourier transform and \mathcal{F}^{-1} its inverse:

$$\begin{array}{ccc} w(x) & \xrightarrow{w^*\tilde{w}} & C(h) \\ \mathcal{F} \downarrow \uparrow \mathcal{F}^{-1} & & \mathcal{F} \downarrow \uparrow \mathcal{F}^{-1} \\ \varphi(u) & \xrightarrow{|\varphi|^2} & f(u) \end{array}$$

This is the case, for example, for $w = \mathcal{F}^{-1}|f|^{1/2}$. This function is real and symmetric but does not necessarily have a bounded support when the support of $C(h)$ is bounded, which would be a desirable property. Other functions may therefore be more appropriate than $\mathcal{F}^{-1}|f|^{1/2}$. See Oliver (1995) for further discussion. It has not been proved, to our knowledge, that every covariance with bounded support can be related to a dilation function $w(x)$ with bounded support, but such solutions are known for the usual covariance models.

Example 1: Simulation of Classic Models on the Line and in Space

Most stationary covariance models can be expressed as the autoconvolution $w^*\tilde{w}$ of a simple function, which makes them very easy to simulate on a line. Table A.1 gives the function w associated with some covariance models considered in \mathbb{R}^1 (i.e., no turning bands method is applied). Table A.2 gives the function w for the covariance $C_1(h)$ in \mathbb{R}^1 associated with some covariance models $C_3(h)$ in \mathbb{R}^3 by the turning bands relation (7.11).

Note that in \mathbb{R}^1 the exponential covariance is obtained with a dilation function which is itself exponentially decreasing (but equal to zero for negative values of x). The consequence is that the process $Z(x)$ decreases exponentially at the same rate as the covariance, except at the Poisson points where it has a unit jump (Fig. 7.2b). It is an Ambartzumian process (Matheron, 1969b).

Note also that some models can be simulated directly in \mathbb{R}^n , $n > 1$, as shown by Matérn (1960, p. 30): the K -Bessel model (2.56) can be expressed as an autoconvolution function. A special case of this model, the exponential covariance (2.51), can be obtained in \mathbb{R}^n with

$$w(x) \propto \left(\frac{|x|}{a}\right)^{(n-1)/4} K_{(n-1)/4}\left(\frac{|x|}{a}\right) \quad (7.27)$$

but $w(x)$ is infinite at the origin (K_ν is the modified Bessel function of the second kind (A.4)). Figures 7.2c and 7.3c are 1D and 2D sections of a 3D

simulation built with this dilution function, which takes on the form

$$w(x) = \frac{1}{\sqrt{2\pi}a^3} \frac{\exp(-|x|/a)}{|x|/a}$$

Another procedure, based on a mixture of spherical models, will be shown later. It is more convenient than a dilution by the deterministic function (7.27), since the spherical models are the easiest ones to simulate in \mathbb{R}^n , as shown below.

The Gaussian model (2.54) can also be expressed as an autoconvolution by the function

$$w(x) \propto \exp\left(-\frac{2x^2}{a^2}\right) \quad (7.28)$$

for all n .

Example 2: Simulation of Spherical Models

The indicators of simple sets are among the most convenient dilution functions. In particular, the spherical model of \mathbb{R}^n can be obtained from the indicator of a sphere with diameter a

$$w_n(x) = \begin{cases} 1 & \text{if } |x| \leq \frac{a}{2} \\ 0 & \text{if } |x| > \frac{a}{2} \end{cases}$$

If a simulation has to be built over a domain V , the Poisson process must be simulated over the dilated² set $V \oplus S_{\tilde{w}}$, where $S_{\tilde{w}}$ is the support of the function $\tilde{w}(x) = w(-x)$. In \mathbb{R}^3 the steps are then the following:

1. Define a domain D that includes $V \oplus S_{\tilde{w}}$. If the final objective is a conditional simulation, do not forget that V must include all the conditioning data points; in practice, take the smallest parallelepiped containing all the points to be simulated and all the data points, and extend it by $a/2$ on all sides (or take the smallest sphere that contains all these points, and increase its radius by $a/2$).
2. Generate N_0 from a Poisson distribution with mean $\lambda|D|$.
3. Select N_0 independent random points X_i from a uniform distribution within D .
4. At each of these points, set up a sphere with diameter a weighted by an independent random value with mean μ and variance σ^2 .
5. Compute the value of the simulation at any given point x as the sum of the weights of the spheres containing x .

The resulting simulation has a spherical variogram with range a and sill $(\pi/6)\lambda(\mu^2 + \sigma^2)a^3$.

As seen in Section 2.3.4, the functions of \mathbb{R}^{n-2q} deduced from w_n by Radon transform of order $2q$ allow the simulation of the integrated spherical models (2.48) associated with random functions differentiable q times in \mathbb{R}^{n-2q} . For example, integrating w_5 along two coordinate axes yields

$$w_{5-2}(x) = \begin{cases} \pi \left(\frac{a^2}{4} - x^2 \right) & \text{if } |x| \leq \frac{a}{2} \\ 0 & \text{if } |x| > \frac{a}{2} \end{cases}$$

which generates the “cubic” model (2.49).

The method can be adapted to reproduce local variations in the variogram parameters. For example, a geometric anisotropy of variable amplitude can be simulated by forming ellipsoids whose shape, size, and orientation depend on location. Likewise the second moment $\mu^2 + \sigma^2$ can be adjusted if the sill varies.

Dilution by a Stochastic Function

Figure 7.11a illustrates the construction of the spherical model of \mathbb{R}^2 , namely the circular variogram (2.46). This construction is based on random disks of equal size. The circular variogram is cumbersome, and the spherical model of \mathbb{R}^3 is often used even when working in \mathbb{R}^2 . The simulation method described above amounts to picking random points in the slice delimited by the two planes parallel to the simulation plane and at distance $a/2$ from it, and forming spheres with constant diameter a . This is equivalent to picking random points in the plane (with a 2D intensity $\lambda_2 = a\lambda_3$ where λ_3 represents the 3D intensity of sphere centers) and forming disks with random diameters drawn from the c.d.f. $F_a(d)$ of the diameters of random planar sections of a sphere with diameter a

$$F_a(d) = \begin{cases} 1 - \sqrt{1 - \frac{d^2}{a^2}} & \text{if } d \leq a \\ 1 & \text{if } d \geq a \end{cases}$$

This method has been popularized by Alfaro (1980) under the name of the *random coins method* (Fig. 7.11b). However, it had been already shown by Matérn (1960, p. 36) that the spherical model of \mathbb{R}^{m+n} can be simulated in \mathbb{R}^n with

$$F_a(d) = \begin{cases} 1 - \left(1 - \frac{d^2}{a^2} \right)^{m/2} & \text{if } d \leq a \\ 1 & \text{if } d \geq a \end{cases}$$

By using other distributions F for d , the method can be extended to other covariance models with a linear behavior at the origin, and in particular to

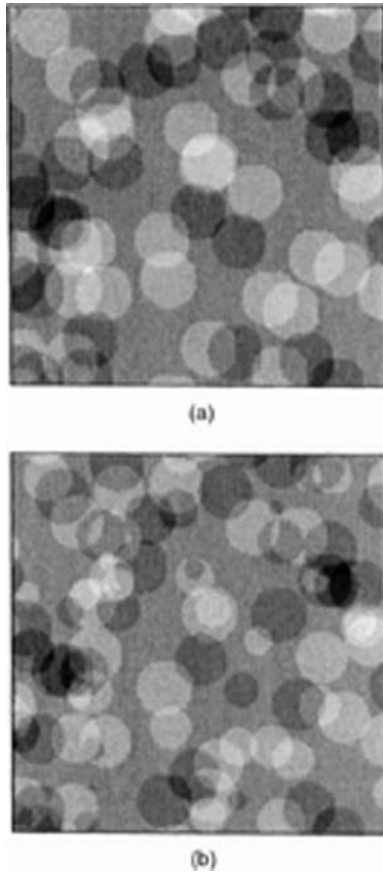


FIGURE 7.11. 2D simulation of circular and spherical covariances by dilution: (a) circular covariance (identical coins); (b) spherical covariance (random coins).

the exponential model in \mathbb{R}^n . For example, in \mathbb{R}^1 the exponential covariance $\exp(-|h|/a)$ can be simulated by random segments with lengths drawn from the c.d.f.

$$F(d) = 1 - e^{-d/a}$$

and a choice of λ , μ and σ^2 such that $\lambda(\mu^2 + \sigma^2) = 1/a$. A practical problem arises here, as for other covariances that tend only asymptotically to zero so that the distribution of d extends to infinity: even if the simulated domain is bounded, it is theoretically necessary to implement germs in the whole space \mathbb{R}^n , but most of the spheres associated with these germs will not intersect the simulated domain. Hammersley and Nelder (1955) have given a thorough general solution in \mathbb{R}^n and detailed its use for the exponential model for $n = 1, 2, 3$. The procedure is defined so that a finite number of spheres intersect

the simulated domain and is built so as to only generate these spheres. Figure 7.2d shows a simulation on the line (with random segments) and Figure 7.3d a simulation in the plane (random coins). This method allows the simulation of any completely monotone covariance in \mathbb{R}^n for all n , since such a covariance can be expressed in the form (2.28) as a mixture of exponential models.

Stochastic dilution functions other than indicators of spheres can of course be used. Since any isotropic covariance model valid in \mathbb{R}^n for all n can be considered as a mixture of Gaussian covariances according to relation (2.27), such a model can in principle be simulated with dilution functions (7.28) with a random scale parameter u (e.g., see Matérn, 1960, p. 33, for the K -Bessel and generalized Cauchy models).

Spatial Distribution and Conclusion

A Gaussian simulation is often desirable. One would think it can be obtained by using definition (7.24) with i.i.d. normal a_i . This is a double mistake:

1. The dilution of a Poisson point process cannot provide an RF whose marginal distribution is exactly Gaussian. Indeed, when the support of w is finite—which it always is in practice—the expression (7.24) reduces to a sum of N i.i.d. random variables, where N is distributed as a Poisson variable with parameter $\theta = \lambda|S_w|$, S_w being the support of w . Denoting by $\chi(u)$ the characteristic function of the random variable $a_i w(x - X_i)$, the characteristic function of $Z(x)$ is $\Phi(u) = \exp[-\theta(1 - \chi(u))]$. It cannot be of the form $\exp(-Ku^2/2)$ which is the characteristic function of a Gaussian distribution (necessarily $|\chi(u)| \leq 1$). The conclusion is the same with stochastic dilution functions.
2. On the other hand, the distribution of a_i that provides the best approximation of the marginal distribution of $Z(x)$ by a Gaussian is not the Gaussian but the discrete distribution $a_i = \pm\sigma$ with equal probability. The larger λ the better the approximation.

In fact, when $w(x)$ is an indicator function, the space is partitioned by the functions $w(x - X_i)$ into zones where the simulation is constant, thus producing a simulation of a jump model as defined in Section 7.1.3. To get simulations that look continuous rather than piecewise constant at medium scale and that are ergodic in the covariance, it is necessary to use a sufficiently large value for λ . Gaussian simulations are obtained as the limit when $\lambda \rightarrow \infty$ and $(\mu^2 + \sigma^2) \rightarrow 0$ so that $\lambda(\mu^2 + \sigma^2)$ remains constant (thus producing an infinite number of infinitely small “discontinuities” within any bounded domain).

Clearly non-Gaussian simulations are obtained for a low λ . In this case the choice of λ , of the distribution of the weights, and also of the dilution function, is an important feature of the simulation. Stable random functions can also be simulated using stable random numbers (Boulanger, 1990).

To summarize, the advantages of the dilution method are the following:

- The method can be directly used for simulating in 2D or 3D (or higher dimension) without the need for turning bands.
- The simulation can be calculated exactly at any point and not only at grid nodes.
- The method is very flexible and can be adapted to a regionalization of the covariance parameters.
- It can generate a large class of non-Gaussian random functions.

Its limitations are the following:

- The simulations can be Gaussian only asymptotically.
- Finding an appropriate dilution function may be difficult if an unusual covariance model is considered.
- Approximations are necessary when the support of the dilution function is unbounded.
- Conditioning must be done separately by kriging.

7.5.3. Continuous Spectral Method

Continuous Spectral Simulation

Since any SRF has a spectral representation (Sections 1.1.5 and 2.3.3), simulating an SRF with a given covariance can be achieved by applying (2.16), that is, by simulating an orthogonal random spectral measure satisfying (2.17). This can be obtained by simulating two uncorrelated *real* random measures satisfying (1.9) (e.g., in \mathbb{R}^1 , two uncorrelated processes with independent increments over $[0, +\infty[$). But this sets up discretization problems, and the basic simulation algorithm, first proposed by Khinchin (1934), is as follows:

Let U be a random vector of \mathbb{R}^n with probability distribution $F(du)/\sigma^2$ where $\sigma^2 = \int F(du)$ (random frequency), and let Φ be a random variable with uniform distribution over $[0, 2\pi[$ (random phase), independent of U . Then the random function in \mathbb{R}^n defined as

$$Z(x) = \sigma\sqrt{2}\cos(2\pi\langle U, x \rangle + \Phi) \quad (7.29)$$

is stationary (wide sense) with mean 0 and covariance $C(h)$ given by (2.14):

$$C(h) = \int e^{2\pi i\langle u, h \rangle} F(du) = \int \cos(2\pi\langle u, h \rangle) F(du)$$

Indeed, for fixed $U = u$, the covariance of $Z(x)$ is

$$C(h)|_{U=u} = \sigma^2 \cos(2\pi\langle u, h \rangle)$$

Randomizing U with the distribution $F(du)/\sigma^2$ yields the covariance $C(h)$. In \mathbb{R}^n , $n > 1$, choosing the random frequency vector U amounts to independently choosing its modulus and its orientation. It is then apparent that the continuous spectral method is equivalent to turning bands with a single random line and a 1D spectral simulation on this line.

Table A.1 gives the spectral density (2.18) associated with the most common covariance models in \mathbb{R}^1 , which will be used for 1D simulation of these models. Table A.2 gives the spectral density of the covariances in \mathbb{R}^1 associated by turning bands with the most common models in \mathbb{R}^3 . Mantoglou and Wilson (1982) present simulations combining turning bands and 1D spectral simulations. Early presentations of the spectral method directly in \mathbb{R}^n with applications to turbulent velocity fields and hydrology are due to Shinozuka (1971) and to Mejía and Rodríguez-Iturbe (1974), respectively (with an extension to space-time phenomena).

Spatial Distribution and Conclusion

The random function defined by (7.29) is not second-order ergodic: its realizations are sinusoidal functions in \mathbb{R}^1 , cylinders with a sinusoidal base in \mathbb{R}^2 , and so on (Fig. 7.2a and 7.3a); the covariances of these realizations are cosine functions so that none of them has the desired covariance $C(h)$ (except of course if $C(h)$ itself is a cosine covariance). Simulations of this random function will only have the correct covariance on the average. Therefore, in practice, one uses the sum, appropriately scaled, of a large number of independent basic simulations to approximate joint normality and ergodicity. Lantuéjoul (1994) shows that in the case of an exponential covariance, several thousand basic simulations must be added to correctly reproduce the bivariate Gaussian distributions (a relative error criterion, which is very stringent, was used).

The advantages of this method are the following:

- A simulation can be produced for any covariance function (provided that its Fourier transform can be computed).
- The exact value of the simulation can be computed at any location.

Its limitations are the following:

- Ergodicity is very slow to reach.
- The method may be computationally tedious if many random frequencies are superimposed in order to approximate joint normality and ergodicity (the cosine terms are all different).
- The method cannot directly produce conditional simulations.

In practice, the method is suitable when the spectrum is not widely spread, namely for differentiable SRFs, whose covariance have a parabolic behavior at the origin. In the discrete case with spacing Δx , it may be more convenient

to consider the corresponding infinite series $Z_n = Z(x_0 + n\Delta x)$, $n \in \mathbb{Z}$, with discrete covariance $C_m = C(m\Delta x)$, $m \in \mathbb{Z}$, and their spectral representations (which are periodic with period $1/\Delta x$). In 1D and for a continuous spectrum, one has, for example, the correspondence

$$f_{\Delta x}(u) = \sum_{m=-\infty}^{+\infty} C_m \exp(-2\pi i m u \Delta x)$$

$$C_m = \Delta x \int_{-1/(2\Delta x)}^{+1/(2\Delta x)} \exp(2\pi i m u \Delta x) f_{\Delta x}(u) du$$

But, in practice, the simulated sequence is finite, and discrete Fourier transforms provide a more efficient algorithm that can be used for most covariance models and produces Gaussian simulations.

7.5.4. Discrete Spectral Method

Discrete Spectral Representations in 1D

Let us first examine the 1D case. If the simulation is computed only at equally spaced points $x = n\Delta x$, $n = 0, 1, \dots, N-1$, over an interval of finite length $L = N\Delta x$, it is convenient to apply the formalism of discrete Fourier transforms (DFT). Since this formalism pertains to periodic series, the series $Z_n = Z(n\Delta x)$ is considered as a stationary process on \mathbb{Z} with period N , for which it suffices to consider the values over a single period $\{0, 1, \dots, N-1\}$.

The smallest observable frequency is $\Delta u = 1/(N\Delta x)$, and all other frequencies are a multiple of Δu , with a maximum of $1/(2\Delta x) = (N/2)\Delta u$. There is a dual correspondence between the series $Z_n = Z(n\Delta x)$ and its discrete Fourier transform $Y_k = Y(k\Delta u)$,

$$Y_k = \sum_{n=0}^{N-1} Z_n e^{-2\pi i k n / N} \quad (7.30)$$

$$Z_n = \frac{1}{N} \sum_{k=0}^{N-1} Y_k e^{2\pi i k n / N} \quad (7.31)$$

and between the covariance series $C_m = C(m\Delta x)$ and its discrete Fourier transform $F_k = F(k\Delta u)$,

$$F_k = \sum_{m=0}^{N-1} C_m e^{-2\pi i k m / N} \quad (7.32)$$

$$C_m = \frac{1}{N} \sum_{k=0}^{N-1} F_k e^{2\pi i k m / N} \quad (7.33)$$

Assuming $E(Z_n) = 0$ the Y_k satisfy $E(Y_k \bar{Y}_{k'}) = 0$ if $k \neq k'$, and $E|Y_k|^2 = NF_k$.

These two relationships can be obtained as a discrete approximation of the Fourier relationships (2.18) of the continuous case. But they also constitute an exact result that can be proved without any reference to the continuous case.

Since the covariance is real and symmetric, the F_k are also real and symmetric. Note that the periodicity and the symmetry about 0 imply a symmetry about $N/2$,

$$C_{N-m} = C_{-m} = C_m$$

$$F_{N-k} = F_{-k} = F_k$$

A key property of the F_k is to be nonnegative, which is equivalent to the assumption that the C_m constitute a valid discrete covariance model.

Since the Z_n are real random variables, the Y_k are complex random variables that satisfy the Hermitian symmetry $Y_{N-k} = \bar{Y}_k$. This means that their real and imaginary parts U_k and V_k satisfy $U_{N-k} = U_k$ and $V_{N-k} = -V_k$. In particular, $V_0 = 0$, and $V_{N/2} = 0$ if N is even. The orthogonality property is equivalent to the following two properties:

1. U_k and V_k are uncorrelated and have the same variance $\sigma_k^2 = NF_k/2$, except for $k = 0$ and for $k = N/2$ where U_k has variance $\sigma_k^2 = NF_k$ and $V_k = 0$.
2. The U_k , ($k = 0, \dots, N/2$) and V_k ($k = 1, \dots, N/2 - 1$) are mutually uncorrelated.

Hence there are in fact N independent real variables ($U_0, \dots, U_{N/2}, V_1, \dots, V_{N/2-1}$ when N is even) corresponding to N data points. Thus, by simulating independent zero-mean random variables U_k and V_k with the appropriate variances σ_k^2 and combining them through (7.31), it is possible to generate a process $Z(x)$ at discrete points $x = n\Delta x$ so that its covariance at lags $m\Delta x$ exactly reproduces a series of covariance terms C_m . Formula (7.31) should not be generalized to calculate $Z(x)$ at points other than the nodes. This is best done by interpolation between adjacent nodes.

Discrete Spectral Simulation in 1D

In practice, the calculation flow is as follows:

$$C_m \rightarrow F_k \rightarrow Y_k \rightarrow Z_n$$

1. Compute the covariance terms $C_m = C(m\Delta x)$, $m = 0, \dots, N/2$, and complete by symmetry.
2. Compute the Fourier transform of the C_m using (7.32) to obtain the Fourier coefficients F_k and the corresponding variances σ_k^2 .
3. Simulate U_0 , U_k and V_k for $0 < k < N/2$, and $U_{N/2}$ with the variances σ_k^2 ; then complete by symmetry for the U_k and by antisymmetry for the V_k .

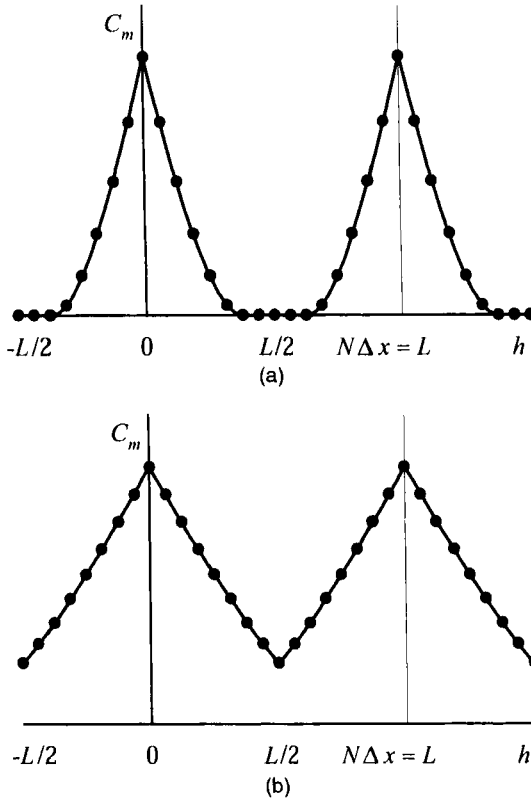


FIGURE 7.12. Periodically repeated covariance: (a) $L \geq 2a$; (b) $L < 2a$. From Chilès and Delfiner (1997), with kind permission from Kluwer Academic Publishers.

4. Invert the sequence $Y_k = U_k + iV_k$ using the inverse Fourier transform (7.31) to get the Z_n .
5. Discard the last (or the first) $N/2 - 1$ points Z_n . There are variations on this strategy, as explained below.

The reason for chopping off part of the simulated sequence lies in the circular character of the DFT. The simulated covariance is periodic with period $L = N\Delta x$, and since the interval considered for m is $[0, N-1]$ rather than $[-N/2, N/2-1]$, the graph looks like shown in Figure 7.12a. Here we have assumed that the covariance has a finite range a ($C(h) = 0$ for $|h| \geq a$) and that $L \geq 2a$. One sees that correlations vanish and then pick up again for lags $h > L - a$. To avoid this spurious effect due to the DFT formalism, it is necessary to limit the simulated sequence Z_n (also periodic with period L) to a length no greater than $L - a$.

The other case $L < 2a$ is depicted in Figure 7.12b. A section of length L is cut out from the covariance, symmetrically about the origin, and repeated pe-

riodically. Within the interval $[0, L]$ the periodic covariance coincides with the original in the interval $[0, L/2]$, and therefore one must restrict the simulated sequence to a length of $L/2$, or $N/2 + 1$ points.

To summarize, if we want to generate a simulation of Z_n over an effective interval length l (after discard), we must first consider a simulation over an interval of length $L = \min(l + a, 2l)$.

Limitations on Covariance Models and Solutions

The method is sometimes used with a variant of steps 1 and 2 where the values F_k are obtained by discretizing the spectral distribution function $F(du)$ of the continuous covariance function $C(h)$ (e.g., Shinozuka and Jan, 1972, whose method is in fact intermediate between the continuous and the discrete cases as they are presented here). Such technique, however, differs in a subtle but important manner from our recommended procedure in that the discrete spectrum F_k is not simply the value of $F(du)$ at frequency $k\Delta u$. Rather, it represents an aliased version of the spectral density in which the contributions of frequencies $(k + pN)\Delta u$ for all integers $p \in \mathbb{Z}$ are folded in the interval $[0, (N - 1)\Delta u]$ and added. Through this mechanism a correct representation of N covariance terms can be achieved with just N spectral terms, whereas a far greater number of terms may be required for an adequate representation of $F(du)$. It is only for covariances with high regularity at the origin, and thus short spectral spread, that the two approaches may coincide.

There is another important difference with the continuous case that, to our knowledge, is not mentioned by any author: the Fourier transform of $C(h)$ automatically produces a nonnegative spectral distribution $F(du)$; in the discrete case this is not always true because the periodic repetition of C_m does not necessarily lead to a valid covariance model (Chiès and Delfiner, 1996). There are, however, two cases where the F_k are necessarily nonnegative because the periodic repetition of $C(h)$ with period L is itself a valid covariance:

1. If $L \geq 2a$ (whatever the shape of $C(h)$).
2. If $C(h)$ is a convex function for $h \in [0, L/2]$.

We have already seen that there is no problem in the first case, shown in Figure 7.12a. In the second case (Fig. 7.13a), let us consider the symmetric function which is equal to $C(h) - C(L/2)$ when $|h| \leq L/2$ and to zero beyond this interval. This function decreases to zero and is convex for $h \in [0, \infty[$. It is thus a covariance function according to Pólya's theorem (Section 2.3.3), and it has a range equal to $L/2$. We are thus back to the first case. Since $C(h)$ differs from it only by the addition of the constant $C(L/2)$ when $|h| \leq L/2$, the conclusion also holds for $C(h)$ and the truncated covariance can be extended to the whole line by periodic repetition.

In other cases, such as covariances with a parabolic behavior at the origin or a hole effect, negative F_k may be found. The problem can be best illustrated

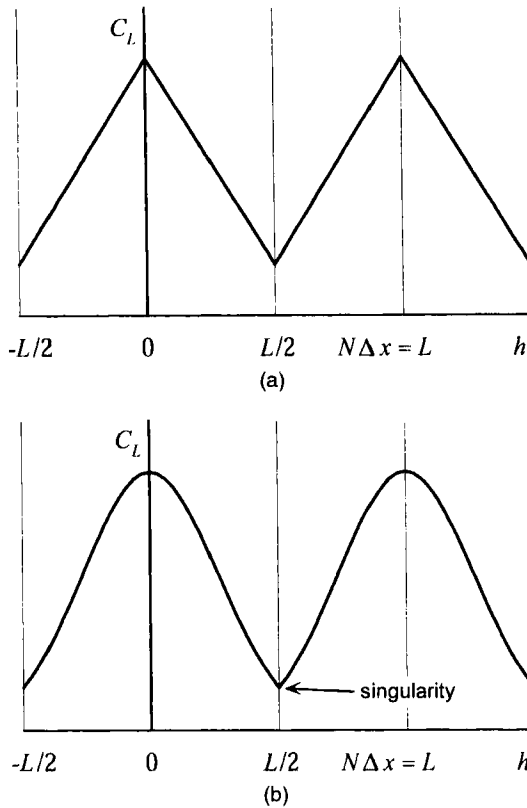


FIGURE 7.13. Periodic repetition of a truncated covariance: (a) periodic repetition of a truncated convex covariance gives a valid covariance; (b) periodic repetition of a truncated Gaussian covariance does not produce a valid covariance. From Chilès and Delfiner (1996).

by a counterexample in the continuous case. Consider a Gaussian covariance $C(h) = \exp(-h^2/a^2)$, which reaches zero only asymptotically when h tends to infinity. Consider now the restriction of $C(h)$ over the interval $[-L/2, L/2]$, and extend it to the whole line by periodic repetition, producing $C_L(h)$ as shown in Figure 7.13b. This “covariance” is twice differentiable at the origin and hence should be twice differentiable everywhere; in fact it is not differentiable for $|h| = L/2$, and so cannot be a valid model for a covariance. Some coefficients F_k of its spectrum will be negative (the Fourier transform of a periodic function is an infinite discrete spectrum).

The problem is not so severe in the discrete case, for the number of terms F_k is finite. But negative values for the spectrum may be found. There are then three solutions:

1. Extend the length L of the simulated domain. In particular, if the covariance has a finite range a and if $L \geq 2a$, all the terms F_k are nonnegative.

- If the covariance reaches the value zero only asymptotically, increasing L will limit the extent of the problem to some low negative values. The other two solutions can then be used.
2. Simply set the negative terms to zero. This will of course produce a bias, but this bias will be small if the above solution is also used, and it can be evaluated by inverting the corrected spectrum using formula (7.33).
 3. Simulate the “time-aliased” covariance $\tilde{C}(h)$ defined as the sum of shifted versions of the original covariance (e.g., on time aliasing, see Oppenheim and Shafer, 1989)

$$\tilde{C}(h) = \sum_{p=-\infty}^{+\infty} C(h + pL)$$

$\tilde{C}(h)$ and its discretized version are periodic with period $L = N \Delta x$. Using formula (7.32) the discrete spectrum \tilde{F}_k of \tilde{C}_m is found to be

$$\tilde{F}_k = \sum_{p=-\infty}^{+\infty} \sum_{m=0}^{N-1} C_{m+pN} e^{-2\pi i k m / N} = \sum_{m=-\infty}^{+\infty} C_m e^{-2\pi i k m / N}$$

\tilde{F}_k coincides with the spectrum of the complete covariance C_m (calculated at the frequencies $k \Delta u$) and is therefore nonnegative. Thus \tilde{C}_m is always a valid covariance. When $L \geq 2a$, \tilde{C}_m and C_m coincide over the interval $[-L/2, L/2]$, which explains why there is no problem. Otherwise, $\tilde{C}(h)$ and $C(h)$ are different, but it is easy to see if the difference is acceptable. This correction ought to be used only when $C(h)$ is low for $h > L/2$. If this is not the case, it is necessary to first increase L . Experimentation shows that aliasing performs very well for differentiable covariances, whereas setting negative spectrum terms to zero works better for covariances with a linear behavior at the origin.

Extension to 2D and 3D

The spectral method can also be extended to \mathbb{R}^2 and \mathbb{R}^3 (or higher dimensions): x , Δx , N and their associated indexes simply have to be replaced by vectors with the corresponding dimension, the product kn in $e^{2\pi i kn/N}$ by the scalar product $\langle k, n \rangle$, and the simple summations by multiple summations. The only problem is to correctly define the independent variables U_k and V_k due to the relations $U_{N-k} = U_k$ and $V_{N-k} = -V_k$ and the variances σ_k^2 . Borgman et al. (1984) present the correct algorithm in 2D (but unfortunately apply it by replacing the discrete spectrum by a discretization of the spectral density function). Pardo-Iguzquiza and Chica-Olmo (1993) detail the algorithm in 2D and in 3D. For example, in 2D, for $N = (N_1, N_2)$, N_1 and N_2 even, and $k = (k_1, k_2)$,

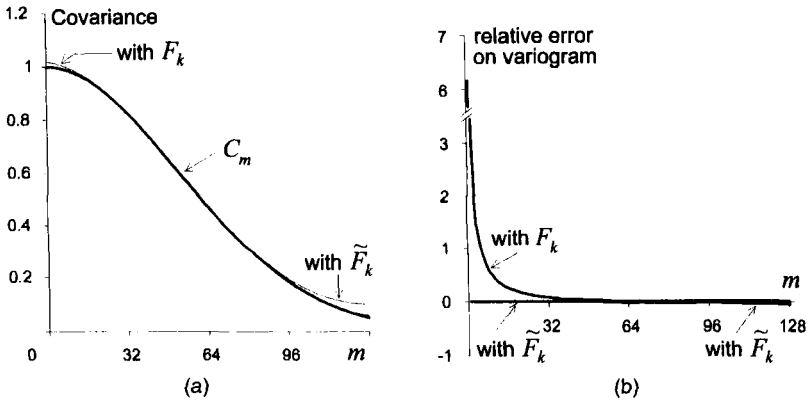


FIGURE 7.14. Approximation of a 2D isotropic Gaussian covariance as results from the discrete spectrum F_k of the truncated covariance and from the discrete spectrum \tilde{F}_k of the aliased covariance: (a) in absolute scale; (b) in terms of relative error on the variogram. The grid is $N \times N$ with $N = 256$ and is used up to $N/2$. The graphs refer to the grid directions (the approximation is better in the other directions). Note that the aliased version perfectly reproduces the behavior at the origin.

the independent variables are

- $U_{0,0}$, $U_{0,N_2/2}$, $U_{N_1/2,0}$, and $U_{N_1/2,N_2/2}$, with $\sigma_{k_1 k_2}^2 = N_1 N_2 F_{k_1 k_2}$;
- $U_{k_1 k_2}$ and $V_{k_1 k_2}$ for $\{k_1 = 0; k_2 = 1, \dots, N_2/2 - 1\}$, $\{k_1 = 1, \dots, N_1/2 - 1; k_2 = 0, \dots, N_2 - 1\}$, $\{k_1 = N_1/2; k_2 = 1, \dots, N_2/2 - 1\}$ with $\sigma_{k_1 k_2}^2 = N_1 N_2 F_{k_1 k_2}/2$.

The method allows the simulation of several covariance components in a single step. Each component may have its own geometric anisotropy, or its own zonal anisotropy, along or perpendicular to a coordinate axis. One of these components can be a nugget effect. The limitations due to negative terms for the spectrum are more serious than in the 1D case when the covariance does not reach the value zero: even convex covariances may lead to negative terms in the discrete spectrum. But as in 1D no problem can occur if the size of the simulated domain is such that the covariance reaches zero in all directions before symmetrization and periodic repetition. Figure 7.14 shows for example that the behavior of an isotropic Gaussian covariance at the origin is perfectly reproduced if $L/2$ is equal to (or greater than) the distance where the covariance is 5% of the variance and the aliased covariance is used.

FFT Algorithms

The direct and inverse Fourier transforms can be calculated with the fast Fourier transform algorithm (FFT; e.g., Press et al., 1992). It is advantageous

to select N as a power of 2. In 2D and 3D, very large grids often have to be simulated. From a practical point of view, it may become impossible to store the whole complex array in core memory, as required by the standard FFT algorithm. Fortunately other algorithms exist that use storage on an external device. An efficient one is the algorithm due to Singleton, using four files that have the same size $N_1 N_2 / 2$ and are accessed sequentially (e.g., see Press et al., 1992). Even if we can define a large array, the problem occurs if we exceed the physical memory, thus using virtual storage. Since the standard FFT algorithm accesses the array elements in a scrambled fashion, its performance drops dramatically. Singleton's algorithm can be used with four arrays stored in virtual memory but accessed sequentially.

Algorithms have been developed that do not require N to be a power of 2, but a product of prime numbers. This may be very useful; for example, if we need a 3D simulation with $N_1 = N_2 = N_3 = 72$, the standard algorithm requires to extend the size to $N_1 = N_2 = N_3 = 128$, thus multiplying the global size of the array by a factor of about 6, which can be avoided by these algorithms. In the 2D case Singleton (1979) has developed such an algorithm that also works with disk storage.

Spatial Distribution and Conclusion

The method presented here can also be regarded as a moving average method (see Ripley, 1987). In its standard application, the U_k and the V_k are selected from Gaussian distributions so that the method, by definition, produces a simulation whose multivariate distribution is Gaussian. But other choices are possible. Selecting U_k and V_k independently from a Gaussian distribution with variance $\sigma_k^2 = F_k/2$ amounts to selecting a random amplitude with mean F_k and a random phase with a uniform distribution over $[0, 2\pi[$. But we can fix the amplitude to $\sqrt{F_k}$ and only randomize the phase. This produces simulations whose sample covariance coincide (for each realization and not only on the average over a large number of realizations) with the covariance C_m . But this is true for the periodic covariance of the infinite sequence Z_n and not for the sample covariance of the finite sequence which is retained. Nevertheless this makes it possible to build simulations whose sample variograms display much lower fluctuations around the simulated model than Gaussian simulations (Chilès and Delfiner, 1997).

In conclusion, the discrete spectral method is a very powerful method of simulation of Gaussian random fields and of other types of random functions. It is seldom used, probably due to a misunderstanding of its capabilities and an often erroneous description of the algorithm. A correct implementation must use the relationships developed for the discrete case, and adequately handle the symmetries and periodicities implied by the use of discrete Fourier transforms.

The discrete spectral method offers the following advantages:

- The multivariate distribution of the simulations is exactly Gaussian (in its standard use).
- The method is completely general and allows an exact simulation of any covariance with a finite range, and either an exact simulation or an approximate one to a desired accuracy if the range is infinite.
- The method can be directly used for simulating in 2D or 3D (or higher dimension), without the need for turning bands.
- It is possible to simulate nested structures, as well as geometric anisotropies, zonal anisotropies parallel or orthogonal to the axes, and a nugget effect, in a single step.
- The FFT algorithms are computationally very efficient; variants using disk storage make it possible to process large 2D or 3D grids without losing much efficiency.

Its limitations are the following:

- Some covariance models cannot be simulated exactly: those for which the periodic repetition from the simulated domain to the whole space does not lead to a valid covariance.
- It is not possible to extend a previous simulation to a larger domain: the whole simulation must be recomputed and will not be consistent with the previous one.
- The simulation is limited to grid nodes.
- Conditioning must be done separately by kriging.

These limitations are usually mild compared to the advantages. Figure 7.15 displays 2D examples of simulations of Gaussian SRFs built with the discrete spectral method.

Simulation with Wavelets

For completeness we will briefly mention simulation methods based on wavelets (e.g., Flandrin, 1992; Zeldin and Spanos, 1995). The motivation for wavelets is to introduce spatial localization in the classic Fourier analysis. Indeed the sine waves used for Fourier analysis are perfectly localized in frequency but not at all in space, so the Fourier coefficients give no clue on the position of local features such as the contours in an image. By contrast, a wavelet decomposition uses a basis of carefully chosen functions that typically vanish outside a compact support and are obtained by scaling and shifting a basic function $\psi(x)$, the *mother wavelet*. The simplest example is the Haar function $\psi(x)$ equal to +1 for $0 \leq x < \frac{1}{2}$, to -1 for $\frac{1}{2} \leq x < 1$, and to 0 outside the interval $[0, 1]$. By dilation and translation it yields the family of step functions

$$\psi_{jn}(x) = 2^{-j/2} \psi(2^{-j}x - n) \quad j, n \in \mathbb{Z}$$

which is a discrete and orthonormal basis of $L^2(\mathbb{R})$ (square integrable functions). The support of $\psi_{jn}(x)$ is the interval of length 2^j starting at $n2^j$. Thus j is a scale index, and n is a translation index (2^j is analogous to a wavelength and 2^{-j} to a frequency). All wavelets sum to zero. The Haar

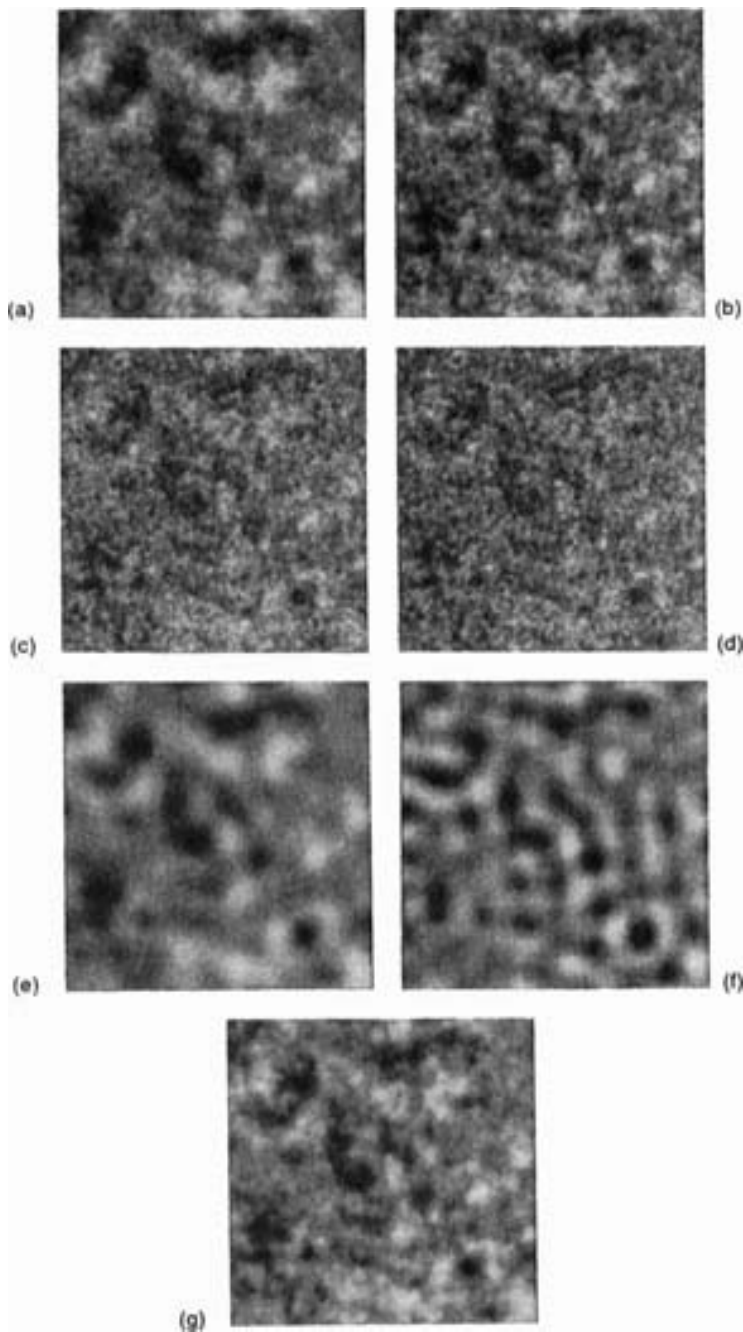


FIGURE 7.15. Examples of simulations of 2D Gaussian SRFs by the discrete spectral method: (a) spherical covariance; (b) exponential covariance; (c) stable covariance with shape parameter $\alpha = 0.5$; (d) hyperbolic covariance; (e) Gaussian covariance; (f) cardinal-sine covariance; (g) covariance $\sin((\pi/2)\exp(-|h|/a))$. Practical range: 1/8th the grid size.

function is well localized in space but not in frequency, since its Fourier transform decays only like $1/|u|$. Wavelets more regular than the Haar function are often used, notably the compactly supported Daubechies wavelets (Daubechies, 1992; Cohen, 1992; Meyer, 1993). They have higher moments that vanish as well.

Associated with the mother wavelet is the “father wavelet,” or “scaling function,” ϕ and the family

$$\phi_{jn}(x) = 2^{-j/2} \phi(2^{-j}x - n) \quad j, n \in \mathbb{Z}$$

Here ϕ is used to calculate the *approximation* of $Z(x)$ at a given resolution 2^J , while ψ is used to represent the *details*, namely the difference between two successive approximations. Equivalent terms are *trend* and *fluctuations* (see Meyer, 1993, p. 40). A scaling function sums to one. For the Haar function, for example, $\phi(x) = 1_{0 \leq x < 1}$. Specifically, for any given resolution 2^J , the following decomposition holds:

$$Z(x) = 2^{-J/2} \sum_{n=-\infty}^{+\infty} a_{Jn} \phi_{Jn}(x) + \sum_{j=-\infty}^J 2^{-j/2} \sum_{n=-\infty}^{+\infty} d_{jn} \psi_{jn}(x)$$

The coefficients a_{Jn} and d_{jn} are obtained as inner products of $Z(x)$ with $\psi_{Jn}(x)$ and $\psi_{jn}(x)$ but can also be computed recursively (recurrence in J or j). Recurrence relations also make it possible to exactly reconstruct $Z(x)$ if an initial (coarser) approximation a_{Jn} , and the different sequences of details d_{jn} at finer scales $j \leq J$ are given. This is the basis of simulations with wavelets. When $Z(x)$ is an SRF, the sequences $a_{j\cdot}$ and $d_{j\cdot}$ are stationary random sequences. The wavelet functions form an orthonormal basis, but contrary to spectral representations, the random sequences $d_{j\cdot}$ are not uncorrelated. To simulate an SRF with a given covariance function, it is therefore necessary to derive the direct and cross-covariances of the various random sequences and to simulate them accordingly. The problem seems to be more complex than it was at the beginning. Fortunately many covariance terms are negligible, and iterative algorithms can be developed if the mother wavelet is chosen adequately. Zeldin and Spanos (1995) present some 1D and 2D examples where the covariance function of the wavelet representation at a given scale approximates the desired covariance quite closely.

When $\psi(x)$ is the Haar function, the wavelet method is equivalent to the simulation via *local average subdivision* (Fenton and Vanmarcke, 1990): the average value over the whole domain (a segment in 1D) is simulated; then the region is subdivided in two subdomains, whose average values are simulated conditionally on the first average value, and so on.

To summarize, wavelet algorithms are rather complex but offer the possibility of zooming in a simulation done at a certain resolution. They are useful if this functionality is needed.

7.6. NONCONDITIONAL SIMULATION OF AN IRF- k

An IRF- k whose generalized covariance is in fact an ordinary covariance, up to an even polynomial of degree $2k$, can be regarded as the sum of an SRF and a polynomial drift of degree k . It can be simulated using the methods designed for stationary random functions. In particular, this is the case where $k = 0$ for IRF-0 whose variogram is bounded. Here we will first consider unbounded variograms, essentially the $|h|^\alpha$ variograms, then IRF- k of higher order k . We start with the unbounded variogram most popular in geostatistical applications: the linear variogram.

7.6.1. Simple 1D Example: The Linear Variogram

In one dimension, the linear variogram is a familiar figure in the theory of stochastic processes. Indeed it is involved in the two basic models of processes with independent and stationary increments: the Wiener-Lévy process and the Poisson process.

Poisson Process

A Poisson process $Z(x)$ with intensity λ on $[0, +\infty[$ can be defined as follows:

1. Consider the Poisson points $X_i, i = 1, 2, \dots$, of a 1D Poisson point process with intensity λ defined on $[0, +\infty[$.
2. Define $Z(x)$ as being equal to 0 on $[0, X_1[$ and to i on the interval $[X_i, X_{i+1}[$, $i = 1, 2, \dots$.

$Z(x)$ can be defined for negative x values by inverting the construction method. $Z(x)$ is a step function: it increases by unit jumps at discontinuity points which are the Poisson points, and it remains constant between successive discontinuity points. $Z(x+h) - Z(x)$, $h > 0$, is then equal to the number of discontinuity points $N([x, x+h])$ that occur between x and $x+h$. According to the properties of the Poisson point process, presented in Section 7.5.2, $Z(x)$ is a process with independent and strictly stationary increments (e.g., Doob, 1953). In particular, since $Z(x+h) - Z(x)$ is a Poisson random variable with parameter $\theta = \lambda|h|$, the Poisson process with intensity λ has a linear drift with slope λ and a variogram $\gamma(h) = \frac{1}{2}\lambda|h|$. Figure 7.16 (top) shows a realization of a Poisson process with intensity $\lambda = 1$ after subtraction of the drift. Note that since $Z(x)$ is a step function without accumulation points, its Hausdorff dimension is 1: even though its variogram is linear, $Z(x)$ is not fractal.

The definition can be extended to a random process with jumps of random amplitude: if the jumps are not equal to 1 but to i.i.d. random variables with mean 0 and variance σ^2 , the resulting process $Z(x)$ remains a process with independent and strictly stationary increments. It has no drift and a linear variogram $\gamma(h) = \frac{1}{2}\lambda\sigma^2|h|$. Figure 7.16 also displays such simulations, for random jumps equal to -1 or $+1$ (middle), or to a standard normal variable (bottom); both are rescaled so as to have the same variogram as the standard Poisson process. Note that $Z(x)$ is not a Gaussian IRF-0, even though the jumps are Gaussian: for example, since $Z(x)$ is piecewise constant, an increment $Z(x+h) - Z(x)$ has a nonzero probability to be equal to 0 (this probability is at least equal to $e^{-\lambda|h|}$, the probability that no Poisson point falls between x and $x+h$). It is in fact a jump RF in the sense of Section 7.1.3.

Brownian Process

The Brownian process, also named *Wiener-Lévy process*, and often *Brownian motion*,³ is usually defined as a process $Z(x)$ with independent stationary in-

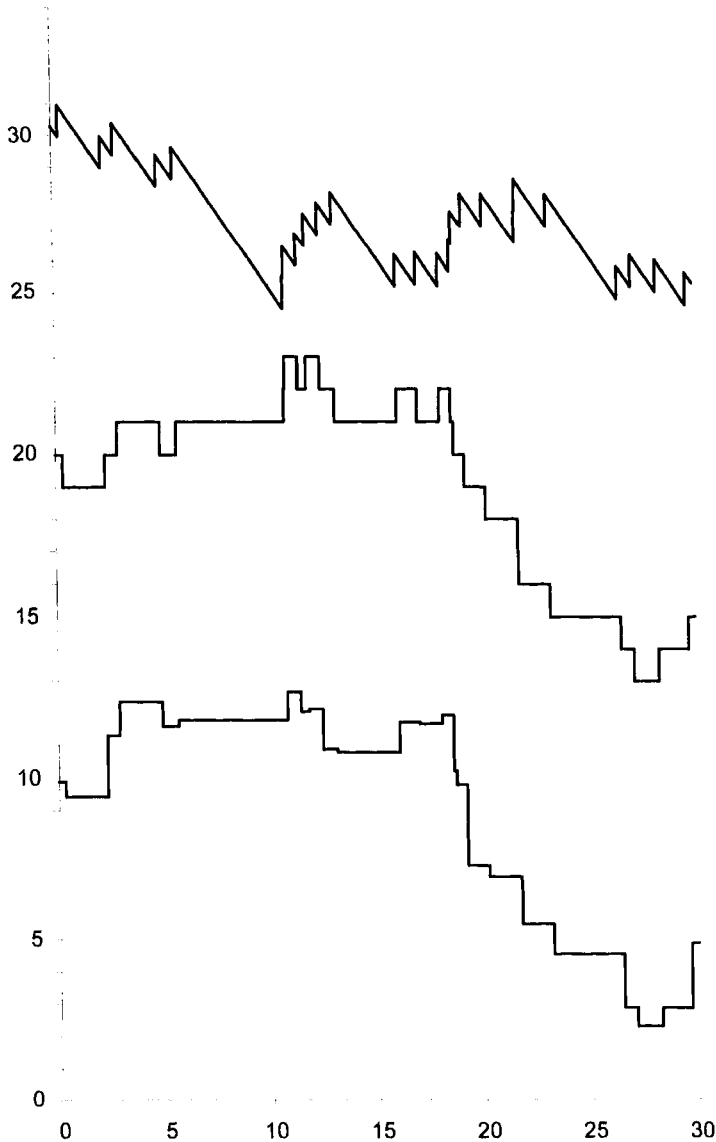


FIGURE 7.16. 1D simulations of the Poisson process. From top to bottom: pure Poisson process after subtraction of a linear drift; Poisson process with random jumps $+1$ or -1 ; Poisson process with Gaussian jumps. All are normalized so as to obtain the same linear variogram $\frac{1}{2}|h|$.

crements, such that the increment $Z(x+h) - Z(x)$ has a Gaussian distribution with mean zero and a variance proportional to $|h|$. Hence, it is a Gaussian IRF-0 with a linear variogram $\gamma(h) = b|h|$ ($b > 0$). $Z(x)$ can also be regarded as the integral over $[0, x]$ of a Gaussian white noise. Among the random processes



FIGURE 7.17. 1D simulation of the Brownian, or Wiener-Lévy, process (random walk) with variogram $\frac{1}{2}|h|$.

satisfying the definition, only separable versions are considered as Brownian processes. Almost every realization is continuous but nowhere differentiable (see, e.g., Blanc-Lapierre and Fortet, 1953, sec. IV.5; Doob, 1953, sec. VIII.2). This reflects the fractal behavior of the Brownian process: as we have seen in Section 2.5.1, its Hausdorff dimension is $3/2$. Its simulation for a set of points $x_0 < x_1 < \dots < x_N$ from an arbitrary value $z_0 = Z(x_0)$ at the origin x_0 follows immediately from its definition. It can be done iteratively using the relation

$$Z(x_i) = Z(x_{i-1}) + U_i$$

where U_i is a centered normal random variable with variance $2b(x_i - x_{i-1})$, independent of the past. Figure 7.17 shows a simulation at a regular interval Δx , known as a *random walk*. Its variogram is the same as that of the Poisson processes of Figure 7.16.

It is also possible to directly generate a conditional simulation of a Brownian process. Indeed, because it is a process with Gaussian independent increments, it is easy to prove the following two properties:

1. If we consider data points $x_0 < \dots < x_i < x_{i+1} < \dots < x_N$ and an unknown point $x \in]x_i, x_{i+1}[$, we have

$$\begin{aligned} \Pr\{Z(x) < z \mid Z(x_0) = z_0, Z(x_1) = z_1, \dots, Z(x_N) = z_N\} \\ = \Pr\{Z(x) < z \mid Z(x_i) = z_i, Z(x_{i+1}) = z_{i+1}\} \end{aligned}$$

that is, the conditioning is only dependent on the two adjacent observations which are screening the influence of the other observations.

2. This conditional distribution is Gaussian; its mean is the ordinary kriging Z^* of $Z(x)$, which amounts in this case to a linear interpolation between x_i and x_{i+1} ; its variance is the ordinary kriging variance σ_K^2 , which is given in Example 4 of Section 3.4.2.

Thanks to these properties, it is easy to implement a sequential algorithm for simulating values between data points and/or refining a simulation between previously simulated points. In the case of regularly spaced points, this is a method popularized by applications of fractals under the name of *random midpoint displacement* (Voss, 1985): if $N + 1$ points such that $N = 2^p$ are to be simulated at a regular interval Δx , the two extreme points are simulated first, by selecting the increment from a Gaussian distribution with mean zero and variance $\sigma^2 = 2\gamma(N \Delta x) = 2bN \Delta x$. The properties above are then used for simulating the midpoint so that the simulated points now define two intervals. This method is used again for simulating the midpoints of each new interval, and so on. After p such iterations, all the points are simulated.

7.6.2. Simulation of an Unbounded IRF-0

We now turn to the simulation in \mathbb{R}^n of an IRF-0 with an unbounded variogram. We consider mainly the case of the $|h|^\alpha$ variogram, $0 < \alpha < 2$. This model owes much of its popularity to the theory of fractals, since in the case of a Gaussian IRF-0 it corresponds to the fractional Brownian random function⁴ (Mandelbrot and Van Ness, 1968). This model also has the particularity of remaining invariant by turning bands: the turning bands operator transforms an $|h|^\alpha$ variogram in \mathbb{R}^1 into an $|h|^\alpha$ variogram in \mathbb{R}^n . Only the coefficient changes, in accordance with (7.12). Thus, if a method for simulating the $|h|^\alpha$ variogram were only valid in \mathbb{R}^1 , it would nevertheless make it possible to generate simulations in \mathbb{R}^n , through the turning bands operator. As for SRFs, there are various methods that can be differentiated by their degrees of rigor, speed of use, Gaussian character (or not), and special limitations. The choice of a method therefore depends on the application considered.

Shortcuts

The random midpoint displacement method simulates a Brownian process correctly. This is no longer the case for any other Gaussian random process, since the Markov property of the Brownian process does not hold. In practice, however, it gives satisfactory results for a variogram of type $|h|^\alpha$ with $\alpha > 1$. Its extension to 2D or 3D is not valid, even for a linear variogram, for the same reason. Note that with this method the zooming factor is restricted to powers of 2. Another algorithm, based on *successive random additions*, makes it possible to zoom at other rates and is also only valid in 1D for a linear variogram. Both algorithms are often used, even beyond this case, as shortcuts for quickly generating simulations that look fractal. Voss (1985) shows 2D simulations of altitude that were built with these algorithms and represent fractal landscapes.

Sequential Simulation or Covariance Matrix Decomposition

A number of methods for simulating SRFs can be used to simulate an IRF-0 or an IRF- k when they have locally stationary representations. Suppose that we must construct a simulation in a bounded domain V . As we have seen in Section 4.6.2, the usual IRF- k have locally stationary representations on V such that the covariance between two points of V is stationary. In 1D, for example, if $Z(x)$ is an IRF-0 with a linear variogram $\gamma(h) = |h|$ and if we are interested in the domain $[-R, R]$, the RF defined by

$$Y(x) = Z(x) - \frac{Z(-R) + Z(R)}{2}$$

has mean zero, and its covariance between two points of the domain is

$$E[Y(x)Y(y)] = R - |y - x| \quad x, y \in [-R, R]$$

This covariance is therefore locally stationary on $[-R, R] \times [-R, R]$.

More generally, as has been shown by Matheron (1974b), an IRF-0 with variogram $\gamma(h) = b|h|^\alpha$ possesses locally stationary representations on $[-R, R]$ with covariances of the form

$$C(h) = b(A - |h|^\alpha) \quad |h| \leq 2R \quad A \geq A_\alpha = \frac{R^\alpha}{\sqrt{\pi}} \Gamma\left(\frac{1+\alpha}{2}\right) \Gamma\left(1 - \frac{\alpha}{2}\right) \quad (7.34)$$

It follows, by application of the turning bands operator, that in \mathbb{R}^n an IRF-0 with variogram $\gamma(h) = b|h|^\alpha$ possesses within the ball with radius R locally stationary representations with covariances

$$C(h) = b(A - |h|^\alpha) \quad |h| \leq 2R \quad A \geq \frac{\Gamma(\alpha + (1+n)/2)}{\Gamma(1+\alpha)\Gamma((1+n)/2)} A_\alpha \quad (7.35)$$

In this case any method able to simulate an SRF within a ball with radius R , and only involving the covariance $C(h)$ for $|h| \leq 2R$, can be used to simulate a locally stationary representation of the IRF-0. This is the case of the sequential method and the covariance matrix decomposition method (Sections 7.2.1 and 7.2.2), but these are limited to a few hundred simulation points if applied rigorously.

Moving-Average Model or Dilution of Poisson Germs

The simulation by moving average (as presented at the end of Section 7.5.1) or by dilution of Poisson germs (Section 7.5.2) requires, in principle, square integrable functions and only permits the simulation of ordinary covariances.

These methods can, however, be extended to functions $w(x)$ such that

$$G(h) = \frac{1}{2} \int [w(x+h) - w(x)]^2 dx \quad (7.36)$$

exists. Reasoning along the same lines as before, one obtains an IRF-0 with the variogram $\gamma(k \Delta x) = \sigma^2 G_{\Delta x; k}$ (moving average) or $\gamma(h) = \lambda G(h)$ (dilution of Poisson points). In the first expression, $G_{\Delta x; k}$ is the discrete approximation of $G(h)$, defined similarly to the discrete approximation (7.21) of the covariogram (7.22) in the stationary case by

$$G_{\Delta x; k} = \frac{1}{2} |\Delta x| \sum_j [w(j \Delta x + k \Delta x) - w(j \Delta x)]^2$$

Consider now the case of an $|h|^\alpha$ model in \mathbb{R}^1 . As shown by Mandelbrot and Van Ness (1968), the function $G(h) = b|h|^\alpha$ can be written in the form (7.36) with

$$w(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \sqrt{\frac{b}{R_\alpha}} x^{(\alpha-1)/2} & \text{if } x > 0 \end{cases}$$

and

$$R_\alpha = \frac{1}{2\alpha} + \frac{1}{2} \int_0^\infty [(1+u)^{(\alpha-1)/2} - u^{(\alpha-1)/2}]^2 du$$

Mandelbrot (1975a) uses a weighting function related to the above one by antisymmetry (i.e., $w(x) = -w(-x)$ for $x < 0$; R_α is then modified). In both cases R_α is finite provided that $0 < \alpha < 2$, which is precisely the condition for $|h|^\alpha$ to be a variogram. Since the support of w is infinite, even Poisson points at a large distance from the simulated domain have an influence on the simulated values within that domain. In practice, only Poisson points up to a certain distance from the simulated domain are considered, which results in a minor approximation. As shown by Chilès (1995), a large number of Poisson points is necessary to achieve ergodicity (typically 10,000 Poisson points). An important particular case is $\alpha = 1$. Then $w(x)$ is simply a step function, so Poisson germs falling outside the simulated domain only contribute a constant value and do not modify the increments: the simulation involves no approximation and is equivalent to the standard Poisson process.

Simulations of an $|h|^\alpha$ variogram in \mathbb{R}^n can be obtained by combining independent 1D simulations by turning bands. If a 1D simulation is built along the central direction of any solid angle $d\theta$ (in practice, along a large number of regularly distributed directions) and if each 1D simulation is built from a Gaussian noise (in practice, a moving-average method with a very small

spacing), the resulting simulation is a realization of a fractional Brownian random function. This method amounts to that proposed by Chentsov (1957) for Brownian random functions and generalized by Mandelbrot (1975a) for fractional Brownian random functions.

Simulation from Poisson Hyperplanes

An alternative method is to use Poisson hyperplanes (see Section 7.7.5 for their definition and main properties). We will use the 2D terminology (with Poisson lines) for simplicity, but the generalization to \mathbb{R}^n , $n > 2$, is straightforward. Consider a disk with radius R enclosing the simulation domain and the lines of a Poisson line process that intersect this disk. Consider each of these lines as a fault subdividing the plane into two half-planes. More precisely, associate with line number i a random function $W_i(x)$ equal to an independently selected random variable a_i with mean 0 and variance σ^2 in either of the half-planes, also independently selected, and to 0 in the other half-plane. Define $Z(x)$ as the sum of all the $W_i(x)$ (the sum always contains a finite number of nonzero terms). Since the intersections of a Poisson line process with a given line is a Poisson point process with intensity λ , the values taken on by $Z(x)$ along any line are those of a Poisson process with random jumps, and hence $Z(x)$ is an IRF-0 with linear variogram $\gamma(h) = \frac{1}{2}\lambda\sigma^2|h|$. Furthermore $Z(x)$ is a locally stationary representation: if the disk with radius R is centered at the origin, the covariance of $Z(x)$ and $Z(y)$ is

$$E[Z(x)Z(y)] = \frac{1}{2}\lambda\sigma^2(\pi R - |y - x|) \quad (|x| \leq R, |y| \leq R)$$

The cross-sectional profile of the faults is a step function. If we replace it by a profile proportional to $w(s)$ (s is the abscissa along the perpendicular to the fault, the origin being on the fault), we get a method which has been proposed by Mandelbrot (1975a) for simulating an $|h|^\alpha$ variogram in \mathbb{R}^n , but now we must also consider the Poisson lines that do not intersect the disk. Figure 7.18 shows a 2D simulation of a linear variogram.

It can be seen that this method is exactly equivalent to using turning bands with a very high number of randomly oriented lines and dilution simulations with a very small Poisson point intensity on these lines (the faults are perpendicular to the radial lines). Straight faults and turning bands are two ways of looking at the same thing.

Continuous Spectral Method

As we have seen in Section 2.3.3, the spectral representation of an IRF-0 is very similar to that of an SRF. The main difference is that the spectral measure $F(du) \equiv \chi(du)/(4\pi^2|u|^2)$ associated with the variogram is no longer necessarily finite: the low frequencies can have infinite variance. In practice, if we simulate in a bounded domain V , we can define a frequency threshold u_0 such that every frequency component $u < u_0$ can be considered as constant

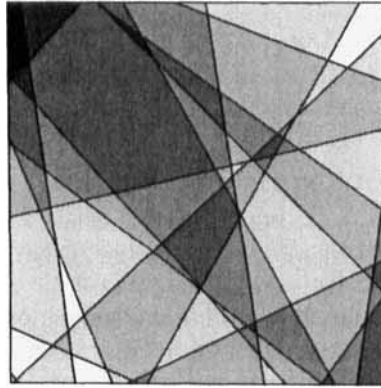


FIGURE 7.18. 2D simulation with a linear variogram, constructed from Poisson lines.

at the scale of the domain V . An IRF-0 being defined up to a constant, we can then apply the spectral simulation algorithm of the continuous case (Section 7.5.3), taking for the distribution of the random frequency U the distribution associated with the spectral measure $F(du)$ of the IRF-0 conditioned by $|U| > u_0$. Chilès (1995) examines the case of an $|h|^\alpha$ variogram in 1D: to achieve an ergodic behavior, it is necessary to take a low value for the threshold u_0 when α is large (about $1/(100L)$ when $\alpha = 1.5$, where L is the length of the simulated domain) and several thousand elementary simulations must be added.

Discrete Spectral Method

Voss (1985) presents simulations of fractional Brownian random functions (i.e., Gaussian IRF-0 with an $|h|^\alpha$ variogram) obtained by FFT. These are 2D simulations of relief and 3D simulations of clouds (by transformation of a 3D simulation of temperature). But the method uses the spectrum obtained by discretization of the spectral measure $F(du)$, which results in a bias as we saw in Section 7.5.4. This bias, which depends on α , is generally very large (Chilès, 1995).

By design the discrete spectral method is concerned with SRFs with periodic covariances. We will show the following extension: in \mathbb{R}^1 any IRF-0 whose variogram $\gamma(h)$ is concave over $[0, 2R]$ possesses a locally stationary representation with a periodic covariance, of period $4R$, and can therefore be simulated by the discrete spectral method at grid nodes discretizing the segment $[-R, R]$.

Indeed, given a variogram $\gamma(h)$ that is concave over $[0, 2R]$, consider the function $C(h)$ defined on $[-2R, 2R]$ by $C(h) = \gamma(2R) - \gamma(h)$ and extended to the whole line by periodic repetition with period $4R$. $C(h)$ is convex in the interval $[0, 2R]$, decreases from $\gamma(2R) > 0$ to zero in this interval, and therefore is a valid periodic covariance. An SRF with covariance $C(h)$ is a locally stationary representation of the IRF-0 on $[-R, R]$.

In practice, using the notations of Section 7.5.4, the simulation will be done over a segment of length $L = 4R$, but only the first half is used in view of the symmetry introduced by periodic repetition. Also note that the definition of $C(h)$ on $[-2R, 2R]$ can be replaced by $C(h) = A - \gamma(h)$ provided that $A \geq A_0$ where A_0 is the smallest value ensuring that $C(h)$ remains a covariance (A_0 is usually smaller than $\gamma(2R)$). Since the choice of A simply affects the value F_0 of the discrete spectrum, it suffices in the application of the discrete spectral method to replace $C(h)$ by $-\gamma(h)$ and set to zero the negative value that will be found for F_0 .

In view of this, it is always possible to simulate the variogram $\gamma(h) = b|h|^\alpha$ by the discrete spectral method:

- If $\alpha \leq 1$, $b|h|^\alpha$ is a concave variogram, and the discrete spectral method can be applied directly.
- If $\alpha > 1$, the *increments* of the IRF-0 have a convex covariance, and the discrete spectral method can be used to simulate these increments.

The second result is new. At first glance one might wonder why we couldn't just use the fact that $\gamma(h)$ has the locally stationary representation (7.34) and simulate that by the spectral method. The reason is that the periodic repetition of this local covariance beyond the interval $[-2R, 2R]$ yields a function that is not positive definite when $\alpha > 1$ (the Fourier series has negative terms). There remains to show that increments have a convex covariance. As in Section 7.5.4 let Z_n denote the series of the values assumed by an IRF-0 $Z(x)$ at the points of abscissa $x_n = x_0 + n \Delta x$, $n = 0, 1, 2, \dots$, and let $R_n = Z_n - Z_{n-1}$, $n = 1, 2, \dots$, denote the corresponding increments. $Z(x)$ being an IRF-0, $R(x) = Z(x) - Z(x - \Delta x)$ is an SRF, and the covariance of its increments can be expressed as a function of the variogram $\gamma(h)$ of $Z(x)$ by

$$\text{Cov}(R_n, R_{n+m}) = \gamma((m-1)\Delta x) - 2\gamma(m\Delta x) + \gamma((m+1)\Delta x)$$

When $\gamma(h) = b|h|^\alpha$, this discrete covariance is convex for $m \in \mathbb{N}$ if and only if $\alpha > 1$, the case we are considering. The proof is purely technical and will not be reproduced. It can be obtained easily if we start from the following result: if a_1, \dots, a_p are positive numbers, $[(1/p) \sum_{i=1}^p a_i^\alpha]^{1/\alpha}$ is an increasing function of α (this is the general justification of the inequality: harmonic average \leq geometric average \leq arithmetic average).

We can therefore construct a simulation of the increments R_n by the discrete spectral method and deduce a simulation of the Z_n by taking $Z_0 = 0$ and $Z_n = Z_{n-1} + R_n$, $n = 1, 2, \dots$. Figure 7.19 shows 1D simulations for $\alpha = 1/3, 2/3, 1, 4/3, 5/3$.

The discrete spectral method cannot be used to simulate $|h|^\alpha$ exactly in \mathbb{R}^n , $n > 1$, even when $\alpha \leq 1$ because the locally stationary representations cannot be periodically repeated. However, it remains possible to generate simulations in \mathbb{R}^1 and, by turning bands, to construct a simulation in \mathbb{R}^n .

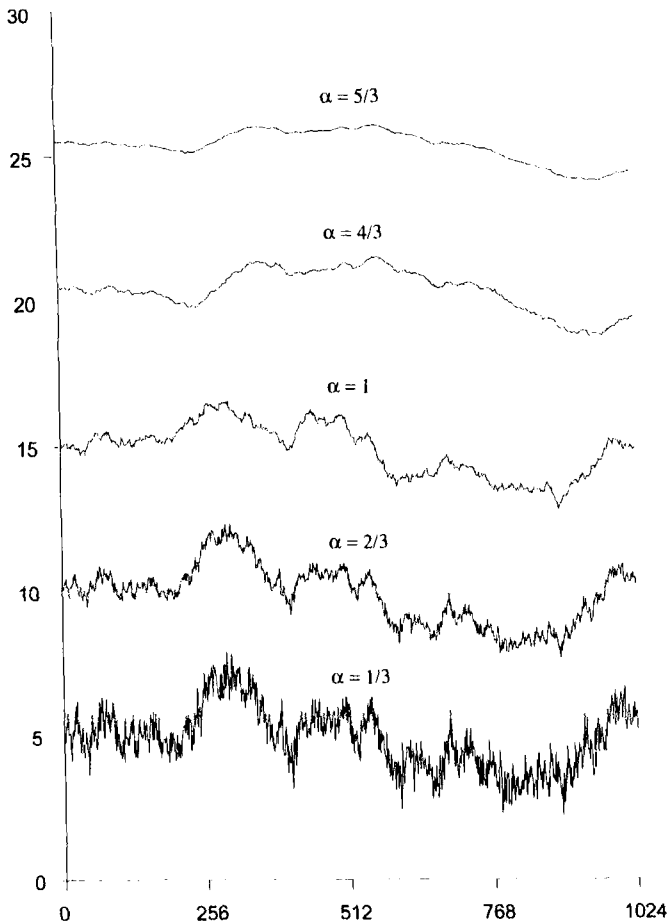


FIGURE 7.19. 1D simulations of fractional Brownian processes with variograms of type $|h|^\alpha$, for $\alpha = 1/3, 2/3, 1, 4/3, 5/3$, by the discrete spectral method. The simulations are built at 1025 discrete points. All variograms are equal to 1 for $h = 512$. The fractal dimension of these simulations is $D = 2 - \alpha/2$.

Another Shortcut: Simulation of a Weierstrass-Mandelbrot Random Function

Let us also mention a variant suggested by Voss (1985) which is to sample the frequencies in a geometric rather than arithmetic series. In 1D, $Z(x)$ is defined by

$$Z(x) = \sum_{k=-\infty}^{\infty} A_k r^{k\alpha/2} \sin(2\pi r^{-k} x + \Phi_k)$$

where the A_k are independent random variables with equal variance, and the Φ_k are independent random phases uniform on $[0, 2\pi]$. The parameter r must be chosen between 0 and 1. The amplitude of the low frequencies (negative values of k) tends to infinity when $k \rightarrow -\infty$; on the other hand, the amplitude of the high frequencies tends to zero when $k \rightarrow +\infty$. $Z(x)$ is the

Weierstrass-Mandelbrot random function. If we define lower and upper frequency limits, we can easily construct an approximate simulation of this RF, since a limited number of terms is sufficient. But this RF is neither stationary nor intrinsic. Although it gives simulations of relief that look fractal, it does not really permit the construction of realizations with an $|h|^\alpha$ variogram.

Simulation with Wavelets

The wavelet method has been extended to the simulation of the fractional Brownian random function. The success of the algorithm, namely the possibility to efficiently truncate the infinite sum to a limited number of terms, depends critically on the adequate choice of the wavelet family; for example, see Flandrin (1992), Elliott and Majda (1994), and Sellan (1995). This method achieves rigorously what is only approximated by the random midpoint displacement method. It includes as a particular case the simulation via local average subdivision, applied to the fractional Brownian random function by Fenton and Vanmarcke (1990). Let us recall that the main value of wavelet methods is the possibility to easily zoom into a simulation.

7.6.3. Simulation of an IRF- k

Simulation of an IRF- k by Integration

On the line the k th integral of an IRF-0 is an IRF- k , and its generalized covariance is derived by a $2k$ -fold integration of the generalized covariance of the IRF-0 (Section 4.5.1). This property is utilized, in combination with turning bands, to simulate IRF- k in \mathbb{R}^n . For example, in \mathbb{R}^n a generalized covariance $|h|^{\alpha+2p}$ ($0 < \alpha < 2$, $0 \leq p \leq k$) is obtained by simulating the same model on lines; to this end it suffices to integrate p times a simulation of an IRF-0 with an $|h|^\alpha$ variogram.

Simulation of a Polynomial Generalized Covariance

Polynomial generalized covariances possess locally stationary representations in \mathbb{R}^n for all n (Section 4.6.2). These representations can therefore be simulated directly in \mathbb{R}^n by the covariance matrix decomposition method. But since this method is limited to several hundred points, polynomial generalized covariances are usually simulated by turning bands (e.g., the simulations shown in Fig. 4.4). The coefficients of the polynomial covariance K_1 to be simulated on the lines are deduced from those of the polynomial covariance K_n in \mathbb{R}^n by inverting the turning bands formula (7.9) using (7.12) with $\alpha = 2p + 1$. In particular, for $k = 2$ the explicit correspondence between the covariance K_1 in 1D and the covariances K_2 and K_3 generated by turning bands in 2D and 3D, respectively, is given by

$$K_1(r) = -b_0 r^2 + b_1 r^3 - b_2 r^5$$

$$K_2(r) = -\frac{2}{\pi} b_0 r + \frac{4}{3\pi} b_1 r^3 - \frac{16}{15\pi} b_2 r^5$$

$$K_3(r) = -\frac{1}{2} b_0 r^2 + \frac{1}{4} b_1 r^3 - \frac{1}{6} b_2 r^5$$

The requested 1D IRF- k can be obtained by successive integrations of the same Poisson or Wiener-Lévy process $W(x)$, as we have seen in Section 4.5.6. Matheron (1972c, 1973a) proposes a variant that provides simulations of locally stationary representations of the IRF- k on $[-R, R]$. These simulations are constructed in a very similar manner to those of Section 4.5.6, but taking in this case the successive *centered* integrals of $W(x)$

$$Z(x) = c_0 Y_0(x) + c_1 Y_1(x) + \cdots + c_k Y_k(x)$$

where the RFs $Y_p(x)$ are defined by

$$Y_0(x) = W(x) - \frac{1}{2}[W(-R) + W(R)]$$

and the recurrence relation

$$Y_p(x) = \int_{-R}^x Y_{p-1}(u) du - \frac{1}{2} \int_{-R}^R Y_{p-1}(u) du$$

The coefficients c_k are the same as in Section 4.5.6. In particular, if we limit ourselves to $k = 2$, an IRF- k with generalized covariance

$$K(h) = -b_0|h| + b_1|h|^3 - b_2|h|^5$$

has on $[-R, R]$ a locally stationary representation with covariance

$$\begin{aligned} C(h) = & b_0(R - |h|) + b_1(2R^3 - 3Rh^2 + |h|^3) \\ & + b_2(16R^5 - 20R^3|h|^2 + 5R|h|^4 - |h|^5) \quad (h \leq 2R) \end{aligned}$$

If $W(x)$ has the generalized covariance $K_0(h) = -|h|$, the coefficients c_0, c_1, c_2 are deduced from b_0, b_1, b_2 by

$$c_0 = \sqrt{b_0} \quad c_2 = \sqrt{120b_2} \quad c_1 = \sqrt{6b_1 + 2c_0c_2}$$

Note that the terms under the square roots are positive or zero; this results from the conditions on the coefficients b_0, b_1, b_2 to ensure that $K(h)$ is a valid generalized covariance in \mathbb{R}^1 .

If $W(x)$ is simulated with the dilution method (Poisson process), $W(x)$ is piecewise constant, and therefore $Z(x)$ is continuous and piecewise linear for $k = 1$, continuous and piecewise parabolic for $k = 2$. On the other hand, a slight bias occurs when starting from a discrete simulation $W(x)$ (random walk of Section 7.6.1): due to the discretization, the integrals of $W(x)$ are not exact, and lower-degree terms $(-1)^q|h|^{2q+1}$, $q < p$, are introduced in the simulation of $(-1)^p|h|^{2p+1}$ (cf. Orfeuil, 1972).

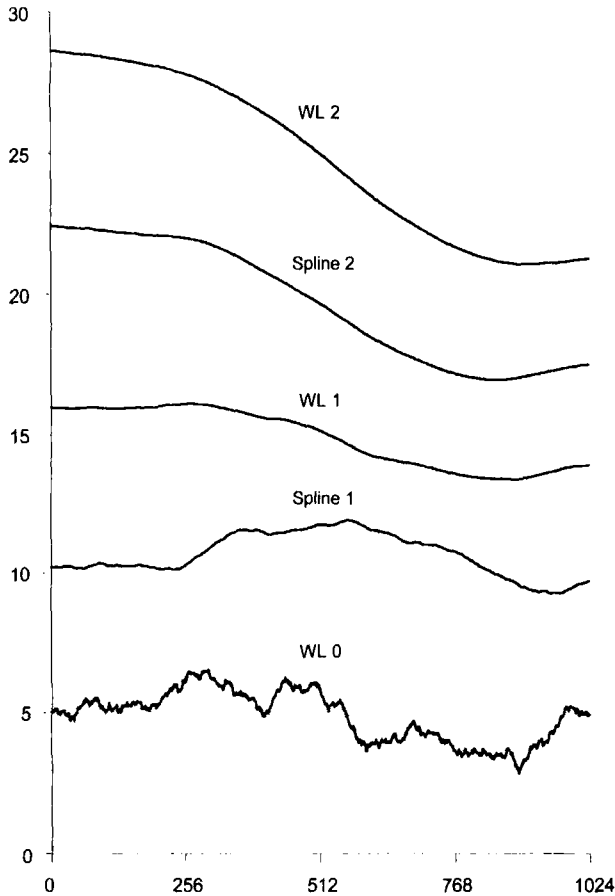


FIGURE 7.20. 1D simulations of an IRF- k with a generalized polynomial or spline covariance by the discrete spectral method: direct simulation of a locally stationary representation (covariance of type $(-1)^{k+1}|h|^{2k+1}$, labeled WL k) or simulation of finite differences and discrete integration (generalized covariance of type $(-1)^{k+1}|h|^{2k} \log|h|$, labeled Spline k).

If we wish to simulate $Z(x)$ at a regular spacing over a segment of length $2R$, we can also simply use the discrete spectral method, with $L = 4R$. Indeed Matheron (1972c) shows that the locally stationary covariance $C(h)$ is extended by periodic repetition to a function with period $4R$ which is still a covariance. Figure 7.20 shows simulations obtained in 1D by the discrete spectral method.

Simulation of a Spline Generalized Covariance

The spline generalized covariance $K(h) = |h|^2 \log(|h|)$ is valid in \mathbb{R}^n , for a drift of degree $k \geq 1$. It possesses locally stationary representations for $|h| \leq 2R$

(see Section 4.6.2), which can be simulated by the method of diagonalization of the covariance matrix if we need a few hundred points at most. In \mathbb{R}^1 Matheron (1988) proposes a very simple method of simulation at a regular spacing that does not have this limitation: let $Z_n = Z(x_n)$ be the values assumed by the IRF-1 $Z(x)$ at the points of abscissa $x_n = x_0 + n\Delta x$, $n = 0, 1, 2, \dots$, and let $R_n = Z_n - Z_{n-1}$, $n = 1, \dots, N$ be the corresponding increments. $Z(x)$ being an IRF-1, $R(x) = Z(x) - Z(x - \Delta x)$ is an IRF-0. If the generalized covariance of $Z(x)$ is $K(h) = h^2 \log(|h|)$, the variogram of the increments R_n and R_{n+m} is

$$\gamma(m\Delta x) = [(m+1)^2 \log(m+1) - 2m^2 \log(m) + (m-1)^2 \log(m-1)](\Delta x)^2 \quad (m \in \mathbb{N}_+)$$

This is a concave function for $m \in \mathbb{N}$. We can thus construct a simulation of the increments R_n by the discrete spectral method. We then deduce a simulation of the Z_n by taking $Z_0 = 0$ and $Z_n = Z_{n-1} + R_n$, $n = 1, 2, \dots$. The same procedure can be used for simulating the generalized covariance $K(h) = -h^4 \log(|h|)$ by starting from second-order increments (their variogram is concave). Figure 7.20 shows simulations obtained in this way.

To obtain a simulation in \mathbb{R}^n , we can use turning bands, since a spline covariance in \mathbb{R}^1 is transformed into a spline covariance in \mathbb{R}^n . Only the coefficient changes (it is divided by n). Similarly, turning bands transform a 1D generalized covariance of type $-h^4 \log(|h|)$ into an isotropic covariance of the same type, but with a coefficient divided by $n(n+2)/3$.

7.7. SIMULATION OF A CATEGORICAL VARIABLE

The simplest categorical variable is one that only assumes the values 0 or 1, that is, the indicator of a set, which we consider here as random. But the covariance function is an extremely poor tool for describing the geometric properties of these very special random functions. For example, the covariance does not give any information on the connectivity of the medium. In fact the covariance is the same for the random set considered and its complement (e.g., grains and pores), while their connectivities are generally very different. Richer tools have been developed in mathematical morphology, but these can be determined only if we have a continuous image of a realization of the random set. Since this book deals essentially with phenomena that are not measured continuously, nor even necessarily on a regular grid, we will not dwell long on the theory of random sets and its applications in mathematical morphology and stochastic geometry. We refer the reader to the literature on this subject, in particular, Kendall and Moran (1963), Matheron (1967, 1975a), Kendall (1974), Serra (1982), and Stoyan et al. (1987). This presentation is limited to the main random-set models that can be of use for geostatistical sim-

ulations and to simple generalizations allowing the representation of m -valued indicators ($m > 2$; e.g., facies number) or the construction of mosaic RFs: random tessellations, Boolean models, marked point processes.

Conversely, some indicator simulation methods derive from standard geostatistical methods, since they involve an underlying continuous variable. This continuous variable can have a physical meaning, for example, when we study the indicator associated with a certain cutoff grade or contamination threshold. It can also be a conventional feature of the model; for example, nested indicators (e.g., lithological facies) can be obtained by slicing a Gaussian variable at successive levels. These methods will also be presented: truncated Gaussian method, substitution random functions.

It is generally easy to simulate a random set once its type and parameters are selected. However, there is no general method for conditioning a simulation of a random set. The conditioning problem must be examined case by case.

In this section we first consider the sequential method, as adapted to indicators, which allows a direct construction of conditional simulations. Next, we outline the iterative algorithms based on Markov chains, which are often used for conditioning other simulation methods or for generating constrained simulations by simulated annealing. Finally, we review various models. Object-based methods and constrained simulations will be studied in the next sections.

7.7.1. Sequential Indicator Simulation

This is the application of the general sequential simulation method to the case of an indicator function, or more generally of several nested indicators. The method has been developed by Alabert (1987b) and Journel (1989). Let us consider for the moment the case of a single indicator, the generalization to the case of several nested indicators being straightforward. We recall that the principle of the sequential method is to sequentially draw the value at each new simulated point from the conditional distribution given the data and the values simulated previously. In the case of an indicator, the values are 0 or 1, and the conditional distribution is therefore defined by its conditional expectation, but this in general is not known. Alabert and Journel propose to replace it by the simple kriging estimate of the indicator. This approximation preserves the mean and the covariance structure of the RF when the data are randomized (provided, however, that kriging produces estimates that lie in the interval $[0, 1]$). Compared with a brute force application of the standard conditioning technique, it has the advantage of producing a binary simulation, in other words, of matching the histogram of the RF.

Basic Algorithm

More formally we denote by $I(x)$ the indicator, which is assumed to be an SRF with mean p , variance $\sigma^2 = p(1 - p)$ and covariance $C(h)$. Assume that $I(x)$ is known at the sample points $\{x_\alpha : \alpha = 1, \dots, M\}$. Let x_{M+1} be the first point

where we set out to simulate $I(x)$. The simple kriging estimator of $I(x_{M+1})$ from $I(x_1), \dots, I(x_M)$ is of the form

$$I^* = p + \sum_{\alpha=1}^M \lambda_{\alpha} (I(x_{\alpha}) - p) \quad (7.37)$$

where the λ_{α} are the solutions of the system

$$\sum_{\beta=1}^M \lambda_{\beta} C(x_{\beta} - x_{\alpha}) = C(x_{M+1} - x_{\alpha}) \quad \alpha = 1, \dots, M$$

Let us draw the simulated value T_{M+1} of $I(x_{M+1})$ from the Bernoulli distribution with mean I^* . For simplicity, we refer to the set $\{I(x_1), \dots, I(x_M)\}$ of conditioning data simply as "data." Conditionally on these data, we have

$$E[T_{M+1} \mid \text{data}] = I^*$$

$$E[T_{M+1}^2 \mid \text{data}] = I^*$$

$$E[I(x_{\alpha}) T_{M+1} \mid \text{data}] = I(x_{\alpha}) I^*$$

Thus, given expression (7.37) of I^* , we get on randomizing the data

$$E[T_{M+1}] = E[I^*] = p$$

$$E[T_{M+1}^2] = E[I^*] = p$$

$$\text{Var}[T_{M+1}] = p(1 - p)$$

$$E[I(x_{\alpha}) T_{M+1}] = p^2 + \sum_{\beta=1}^M \lambda_{\beta} E[(I(x_{\alpha}) - p)(I(x_{\beta}) - p)]$$

The (centered) covariance of $I(x_{\alpha})$ and T_{M+1} is therefore

$$\text{Cov}[I(x_{\alpha}), T_{M+1}] = \sum_{\beta=1}^M \lambda_{\beta} C(x_{\beta} - x_{\alpha}) = C(x_{M+1} - x_{\alpha})$$

This simulation method therefore does match the mean and the covariance function of $I(x)$. This property, proved for all the points x_1, \dots, x_M, x_{M+1} , remains true by iteration when we add x_{M+2} , and so on, provided that at each step the point previously simulated is included in the data set.

Note that the conditional variance of T_{M+1} is

$$\text{Var}[T_{M+1} \mid \text{data}] = I^*(1 - I^*)$$

It is not equal to σ_K^2 as for standard conditioning but takes on a value that fluctuates around σ_K^2 , its expectation being σ_K^2 . Indeed

$$\begin{aligned} E[\text{Var}(T_{M+1} \mid \text{data})] &= E[I^*(1 - I^*)] \\ &= p - p^2 - \sum_{\alpha=1}^M \sum_{\beta=1}^M \lambda_{\alpha} \lambda_{\beta} C(x_{\beta} - x_{\alpha}) \\ &= C(0) - \sum_{\alpha=1}^M \lambda_{\alpha} C(x_{M+1} - x_{\alpha}) = \sigma_K^2 \end{aligned}$$

The only theoretical problem with this method is that I^* can be less than 0 or greater than 1. We have seen in Section 3.9.1 that this can occur in one dimension with very common models (e.g., spherical covariance), and in a space of more dimensions for practically any covariance other than a pure nugget effect. When this occurs, we obviously set I^* to 0 or 1 accordingly, but then the covariance is no longer reproduced exactly. Note that nothing prevents us from using the algorithm with a covariance model that is not a genuine indicator covariance, but then the covariance of the simulated indicator cannot match the (inappropriate) input covariance. Figure 7.3i is a simulation obtained with the sequential method.

Conditioning no longer matches the covariance nor even the variogram if we condition by ordinary kriging. Indeed, going through the above steps, we obtain, for example,

$$\text{Cov}[I(x_{\alpha}), T_{M+1}] = C(x_{M+1} - x_{\alpha}) - \mu_0$$

or

$$\frac{1}{2} E[I(x_{\alpha}) - T_{M+1}]^2 = \gamma(x_{M+1} - x_{\alpha}) - \mu'_0$$

where μ_0 and μ'_0 are the Lagrange parameters of the kriging systems (3.14) and (3.16).

Generalization

The method can easily be extended to nested indicators with increasing thresholds of an RF $Z(x)$

$$I_i(x) = 1_{Z(x) < z_i} \quad \text{where} \quad z_1 < \cdots < z_{i-1} < z_i < \cdots < z_m \quad (7.38)$$

Nested indicators satisfy the characteristic property

$$I_i(x) = 1 \quad \Rightarrow \quad I_j(x) = 1 \quad \forall j < i$$

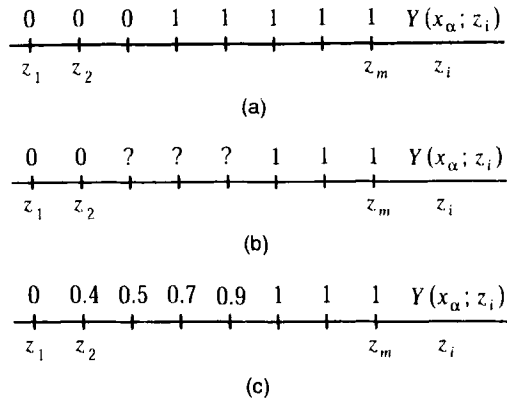


FIGURE 7.21. Coding of the information in view of sequential simulation of nested indicators associated with a continuous variable: (a) exact value (all indicators are filled in); (b) value defined by an interval (indicators in the interval are left blank); (c) value defined by an a priori distribution (c.d.f. values are used). From Journel (1989).

They can be estimated at the point x_{M+1} by cokriging, and the estimates define a cumulative distribution function provided that the usual order relationships between indicators are also satisfied by the cokriging estimates. We then draw T from this distribution. The implementation of this method raises some practical difficulties, which are discussed in Section 6.2.4 (modeling of the direct and cross-covariances, possible nonmonotonicity of the estimated c.d.f.).

Nested indicators can also be defined by reference to categorical variables, such as geologic facies types, without involving thresholds of a random function. Conversely, given nested indicators, it is possible to label the sets with values $0, 1, \dots, m$ and define a discrete RF $Y(x)$ by

$$Y(x) = 1 + \text{maximum value of } i \text{ such that } I_i(x) = 1$$

In the case where the indicators are defined by (7.38), Y is simply a discrete version of Z : $Y(x) = i$ means that $z_i \leq Z(x) < z_{i+1}$ (with the convention that $z_0 = -\infty$ and $z_{m+1} = +\infty$). If Z is continuous, Y varies by unit jumps, like a discrete diffusive random function: facies i can be in contact with facies $i - 1$ and $i + 1$ only. In the other cases Y can vary by larger jumps: facies i may be in contact with facies other than $i - 1$ and $i + 1$ (this will depend on the cross-covariances).

When the indicators derive from an RF Z , the simulation method can be applied, through an appropriate coding shown in Figure 7.21, to account for data given in the form of an interval, an inequality, or an a priori distribution function, namely “membership” data (between 0 and 1)

$$Y(x_\alpha; z_i) = \Pr\{Z(x_\alpha) < z_i \mid \text{local information at } x_\alpha\}$$

(Journel, 1989; Journel and Alabert, 1989). The local information can be the value taken on by an auxiliary variable at this point x_α such as a geological facies type, or an expert's opinion on the value $z(x_\alpha)$.

Applications to the simulation of lithological facies are presented, for example, by Murray (1993) and Dimitrakopoulos and Dagbert (1993). Figure 7.22 displays an example given by Journel and Alabert (1989). Figure 7.22a represents in eight gray levels an exhaustive sampling of a 2 ft \times 2 ft sandstone section by 40 \times 40 permeability data; Figure 7.22b represents, with the same gray levels, the kriging estimates of the permeability using only 16 data; Figure 7.22c and d shows sequential simulations of the nested indicators corresponding to these gray levels, conditionally on these 16 data in the first case, conditionally on the 16 hard data plus a regular grid of 12 \times 12 soft data in the second case (the soft information indicates that the permeability value at the data point belongs to one of three large intervals that divide the whole range of permeability values). We observe the improvement provided by the soft data.

The spatial distribution of the simulations generated by the sequential indicator algorithm are not yet known, except in special cases (e.g., Markov processes). In the case of conditional simulations, it depends on the spatial distribution of the conditioning RF and has no reason to be identical to it.

In conclusion, sequential indicator simulation is a very flexible method. Its use requires some approximations (conditioning from neighboring data only, monotonicity corrections). A shortcoming, mainly when nested indicators are simulated, is that a result is obtained even if the model defined by the indicator direct and cross-covariances is inconsistent.

7.7.2. Iterative Methods Based on Markov Chains

Simulation methods other than the sequential method do not directly provide conditional simulations of indicators. Unlike for SRFs whose realizations are continuous, there is no general conditioning algorithm as simple as adding a simulated kriging error. This algorithm can of course be used with a nonconditional indicator simulation, and will indeed preserve the covariance. However, the conditional simulation will not be an indicator function: it could take on any value between 0 and 1 (and even values outside this interval). Iterative methods make it possible, at least for some SRF models, to condition indicator simulations while preserving the binary character of the simulations. They are also used for the construction of simulations by simulated annealing. The basic methods rely on Markov chains and have been developed after the work of Metropolis et al. (1953) to simulate complex physical systems and solve optimization problems. The reader will find complementary results in Ripley (1987).

General Framework

Iterative methods are mainly used to produce or condition simulations on a grid. The general principle is to start with a simulated grid that does not fulfill

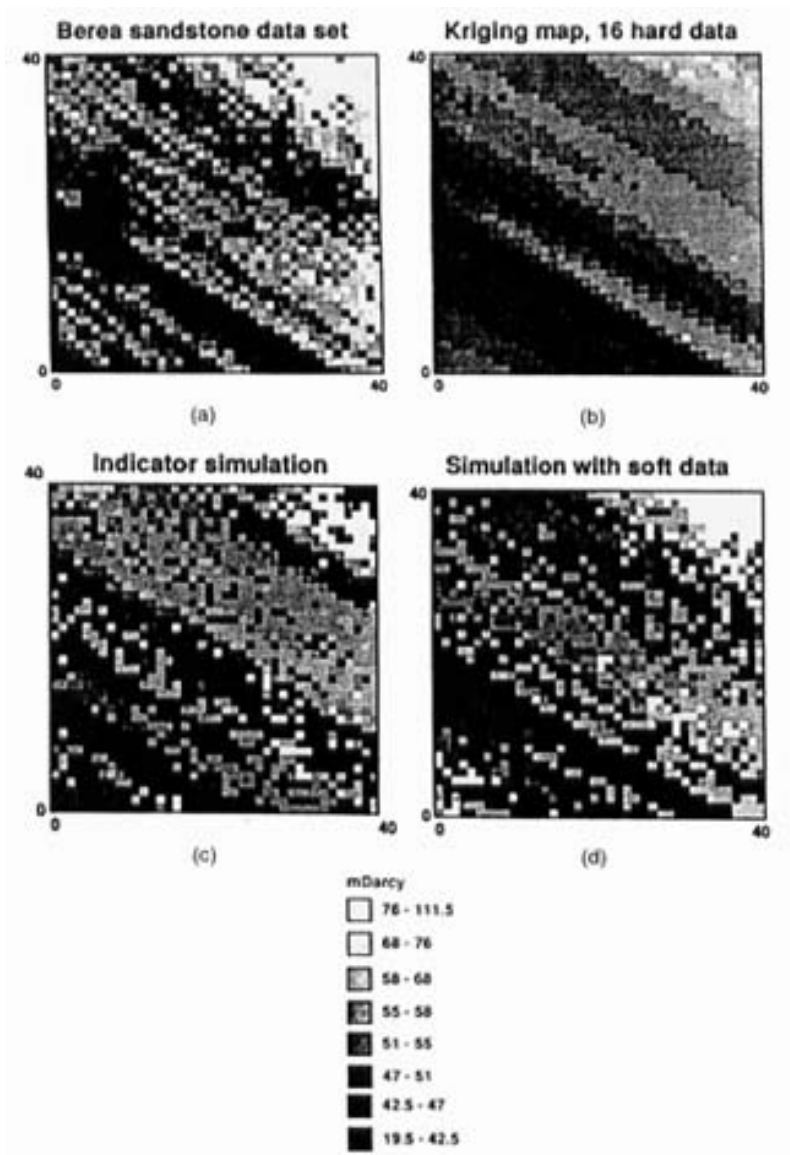


FIGURE 7.22. Conditional simulation of nested indicators: (a) exhaustive data, presented in gray levels; (b) kriging from 16 hard data, represented for the same gray levels; (c) sequential indicator simulation conditioned on the 16 data; (d) sequential simulation conditioned on the 16 hard data and a regular grid of 12×12 soft data. Notice that the use of soft data improves conditioning. From Journel and Alabert (1989).

all the requirements in terms of spatial variability or conditioning and then to update it step by step until these requirements are met. More precisely, let us consider the simulation of a random function $Z(x)$ in the following situation:

1. The points where the simulation is to be built constitute a set of N sites x_1, \dots, x_N (usually the nodes of a regular grid).
2. Z can take on p discrete values $1, 2, \dots, p$ (e.g., it is a binary variable, a categorical variable, or a class of a distribution of a continuous variable).

The realizations of $Z(x)$ satisfy $Z(x_1) = z_1, \dots, Z(x_N) = z_N$ where z_1, \dots, z_N assume one of the p discrete values. Any particular grid of values, that is, any particular value of the vector $\mathbf{z} = (z_1, \dots, z_N)'$, will be called a state and denoted by s (we do not distinguish two realizations of the random function $Z(x)$ if they coincide at the N sites). The vector $\mathbf{Z} = (Z(x_1), \dots, Z(x_N))'$ can take on p^N states s_i ($i = 1, \dots, p^N$). Its N -dimensional probability distribution is a particular finite-dimensional distribution of the random function, and it defines the monodimensional probability distribution π that assigns the probability $\pi_i > 0$ to state s_i (we consider for the sake of simplicity that any of the p^N possible states has a nonzero probability of occurrence; otherwise, the number of possible states must be decreased accordingly). Simulating $Z(x)$ over the N sites amounts to drawing a vector \mathbf{Z} from its N -dimensional distribution, or equivalently drawing a state s from the probability distribution π . The number of possible states can be very large, and the calculation of each probability π_i impossible, so particular methods are necessary. The present section describes iterative methods to sample from π : starting with an initial simulated state $S^{(0)}$ from an arbitrarily chosen distribution, we construct a sequence of random states $S^{(k)}$, $k = 1, 2, \dots$, whose probability distribution converges to π when k increases. This method derives from the properties of Markov chains (e.g., see Feller, 1968, ch. XV): the successive $S^{(k)}$ are obtained as Markov transitions from one state to the next.

The transition from $S^{(k)}$ to $S^{(k+1)}$ is defined by the transition matrix $\mathbf{P} = [p_{ij}]$, which is independent of k

$$p_{ij} = \Pr\{S^{(k+1)} = s_j \mid S^{(k)} = s_i\}$$

A Markov chain with transition matrix \mathbf{P} has the invariant distribution π if and only if

$$\sum_i \pi_i p_{ij} = \pi_j \quad \forall j \quad (7.39)$$

and this chain is reversible if and only if the reversibility conditions

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall i \neq j \quad (7.40)$$

are satisfied. Note that (7.40) implies (7.39). The invariant distribution π of a reversible Markov chain is unique, which is not necessarily the case for a more general Markov chain.

Drawing a state s randomly from the distribution π can be achieved by selecting a Markov chain with transition probabilities satisfying (7.40), building a Markov chain $S^{(k)}$ up to a sufficiently large k value for the chain to have reached the equilibrium, and retaining this final state. Details about the stability and convergence of Markov chains can be found in Meyn and Tweedie (1993). Now the problem reduces to finding a matrix \mathbf{P} satisfying (7.40). Two choices for \mathbf{P} are commonly used:

1. The *Metropolis algorithm* (a generalization of the algorithm used by Metropolis et al., 1953). We define a priori a *symmetric* transition matrix \mathbf{Q} . At each step k , denoting the current state by s_i , the matrix \mathbf{Q} is used to generate a candidate state s_j from the conditional probability \mathbf{q}_i (vector of the q_{ij} for this value of i). This candidate is accepted always if $\pi_j \geq \pi_i$, and with probability π_j/π_i if $\pi_j < \pi_i$. Otherwise, we remain in state s_i . This defines the transition probabilities

$$p_{ij} = \min\left(1, \frac{\pi_j}{\pi_i}\right) q_{ij} \quad \text{if } i \neq j$$

$$p_{ii} = q_{ii} + \sum_{j \neq i} \max\left(0, 1 - \frac{\pi_j}{\pi_i}\right) q_{ij}$$

which are seen to satisfy (7.40). The invariant distribution is unique and the Markov chain converges to it if π is not constant and \mathbf{Q} is irreducible, namely such that every state can be reached from every other state.

2. The *Barker's algorithm* (Barker, 1965). It is similar to the Metropolis algorithm, except that $\min(1, \pi_j/\pi_i)$ is replaced by $\pi_j/(\pi_i + \pi_j)$. Equation (7.40) is still satisfied. The invariant distribution is unique, and the Markov chain converges to it if \mathbf{Q} is irreducible.

More generally, Hastings (1970) proposes the following class of choices for \mathbf{P} :

$$p_{ij} = f\left(\frac{\pi_j}{\pi_i}\right) q_{ij} \quad \text{if } i \neq j$$

$$p_{ii} = q_{ii} + \sum_{j \neq i} \left(1 - f\left(\frac{\pi_j}{\pi_i}\right) q_{ij}\right)$$

where the function $f(u)$, $u \in]0, +\infty[$, is chosen so that $0 < f(u) < 1$ for $u \geq 1$ and $f(u) = uf(1/u)$ for $u < 1$. The Metropolis algorithm corresponds to the

choice $f(u) = \min(u, 1)$, whereas the Barker's algorithm corresponds to $f(u) = u/(1 + u)$. These algorithms can be used in all cases, even if π_i and π_j are not explicitly known, provided that their ratio can be calculated. An introductory exposition of the Metropolis-Hastings algorithm can be found in Chib and Greenberg (1995).

Local Replacement: Gibbs Sampler

A particular case is when each transition involves a single site: \mathbf{Q} is chosen such that $q_{ij} > 0$ only if s_i and s_j are states that differ in value at a single site. If s_i and s_j differ only by their values i_r and j_r at site x_r ,

$$\frac{\pi_j}{\pi_i} = \frac{\Pr\{Z(x_r) = j_r, \text{rest}\}}{\Pr\{Z(x_r) = i_r, \text{rest}\}} = \frac{\Pr\{Z(x_r) = j_r \mid \text{rest}\}}{\Pr\{Z(x_r) = i_r \mid \text{rest}\}}$$

The simplest case is when $Z(x)$ can take on only two different values ($p = 2$). The values i_r and j_r are then the only possible ones for $Z(x_r)$ so that

$$\Pr\{Z(x_r) = j_r \mid \text{rest}\} = \frac{\pi_j}{\pi_i + \pi_j}$$

Barker's algorithm reduces to choosing a site at random and drawing its value from the conditional distribution given the values at the other sites.

This method can be extended to the case where $Z(x)$ can take on $p > 2$ different values, or even has a continuous distribution. The following iteration is performed:

1. Select a site, say x_r , at random or by a systematic scan of all the sites.
2. Choose the new value for this site from the conditional distribution of $Z(x_r)$ given the other values $Z(x_s)$, $s = 1, \dots, r-1, r+1, \dots, N$.

Usually this conditional distribution is assumed to depend on the neighboring values only. Thus we are able to simulate the joint distribution at all sites by successive simulations of conditional univariate distributions. This method was developed by Geman and Geman (1984) who call it the "Gibbs sampler" because, when the sites constitute a square lattice, \mathbf{Z} can be considered as a Markov random field with a Gibbs distribution. Casella and George (1992) give an intuitive presentation of the Gibbs sampler, showing that it "can be thought of as a practical implementation of the fact that the knowledge of the conditional distributions is sufficient to determine a joint distribution (if it exists!)."

Spin Exchange

A particular case of Hasting's method, introduced by Flinn (1974) to simulate phase separation on a grid with two types of atoms, is usually known as the

“spin exchange” method: each transition involves only two sites that exchange their values. The algorithm is the following:

1. Select two sites at random, independently (or the second one in a neighborhood of the first one). If the values at these sites are equal, redraw the second site.
2. Pass from the present state s_i to the state s_j defined by the exchange of the values at these sites, with the probability $\min(1, \pi_j/\pi_i)$ (Metropolis algorithm) or $\pi_j/(\pi_i + \pi_j)$ (Barker’s algorithm).

The Markov chain produced is only irreducible over the set of states with the same marginal distribution: the permutations do not change the histogram of the values over the N sites. So the initial state must be selected so as to match the marginal distribution.

7.7.3. Conditioning by Indicator Values or by Inequalities

Before coming back to the simulation of random sets, let us show how iterative methods make it possible to condition a simulation of a continuous variable by indicator values or inequalities. More precisely we have

- hard data points x_α , $\alpha = 1, \dots, M$, where $Z_\alpha = Z(x_\alpha)$ is exactly known: $Z_\alpha = z_\alpha$;
- soft data points x_α , $\alpha = M + 1, \dots, N$ where $Z_\alpha = Z(x_\alpha)$ is only known to belong to an interval of \mathbb{R} : $Z(x_\alpha) \in B_\alpha$, where B_α is of the form $[a_\alpha, b_\alpha]$, $]-\infty, b_\alpha]$, $[a_\alpha, +\infty[$, or the union of a finite number of such intervals.

Section 7.7.1 presented a solution where Z is discretized so that it can be replaced by a series of nested indicator functions. We describe here another method where no discretization of Z is necessary. The construction of a simulation conditioned on the z_α and the B_α can be decomposed into two steps:

1. Select a random vector $(Z_{M+1}, \dots, Z_N)'$ from the multivariate distribution conditional on the z_α and the B_α .
2. Simulate the random function conditioned on the hard data and the above vector.

The second step is a straightforward simulation conditioned on hard data, so only the first step needs to be addressed here.

If there are only a few interval data approximate methods can be used, for example, replace each interval by its conditional expectation given B_α (e.g., Chilès, 1984; cf. Section 7.11.1) or a value that depends on the neighboring values (Isaaks, 1984). Freulon and de Fouquet (1993) propose an exact itera-

tive method for the Gaussian case. This method is a direct application of the Gibbs sampler method. The initial state $S^{(0)}$ is obtained by assigning the hard data values z_α to sites x_α , $\alpha = 1, \dots, M$, and selecting a random value within B_α for each site x_α , $\alpha = M + 1, \dots, N$, from the Gaussian p.d.f. truncated by B_α . The initial state is consistent with all the data (hard and soft) but not with the spatial structure. The Gibbs sampler permits the introduction of the spatial structure while still honoring the data. Specifically, the following sequence is iterated:

1. Select a site x_α among the soft data (randomly or by scanning these sites).
2. Ignore the value at this site, and estimate it by kriging from the values at all the other sites; also compute the corresponding kriging variance.
3. Replace the value at this site by the kriged value, plus a Gaussian residual with variance equal to the kriging variance, randomly selected so as to match the inequality B_α . An acceptance-rejection technique is preferred to the inverse Gaussian c.d.f. which has numerical problems for the large arguments associated with small kriging variances.

This algorithm converges to an ergodic conditional SRF. Freulon and de Fouquet (1993) demonstrate the use of the method for simulating a 2D graytone image conditioned on a black and white image. The problem is this: produce a continuous gray-tone image with the correct covariance such that after thresholding it coincides with the given black-and-white image. Freulon (1994) extends the method to other types of conditioning: (1) on an image corrupted by Poisson noise, where the value at grid node x_α is not Z_α but an integer random value selected from a Poisson distribution with mean Z_α , and (2) on an image corrupted by noise and deformed by linear averaging along lines. The results, shown in Figure 7.23, are impressive.

As is shown next, this method can be extended to the simulation of a series of nested indicator functions that can be considered as the result of applying thresholds to a common Gaussian SRF (a single indicator function corresponding to the case of a single threshold).

7.7.4. Truncated Gaussian Simulation

Consider an indicator or a series of indicators that originate from applying one or more thresholds to a standard Gaussian RF $Y(x)$,

$$I_i(x) = 1_{y_{i-1} \leq Y(x) < y_i} \quad \text{with} \quad -\infty = y_0 < y_1 < \dots < y_{i-1} < y_i < \dots < y_m = +\infty \quad (7.41)$$

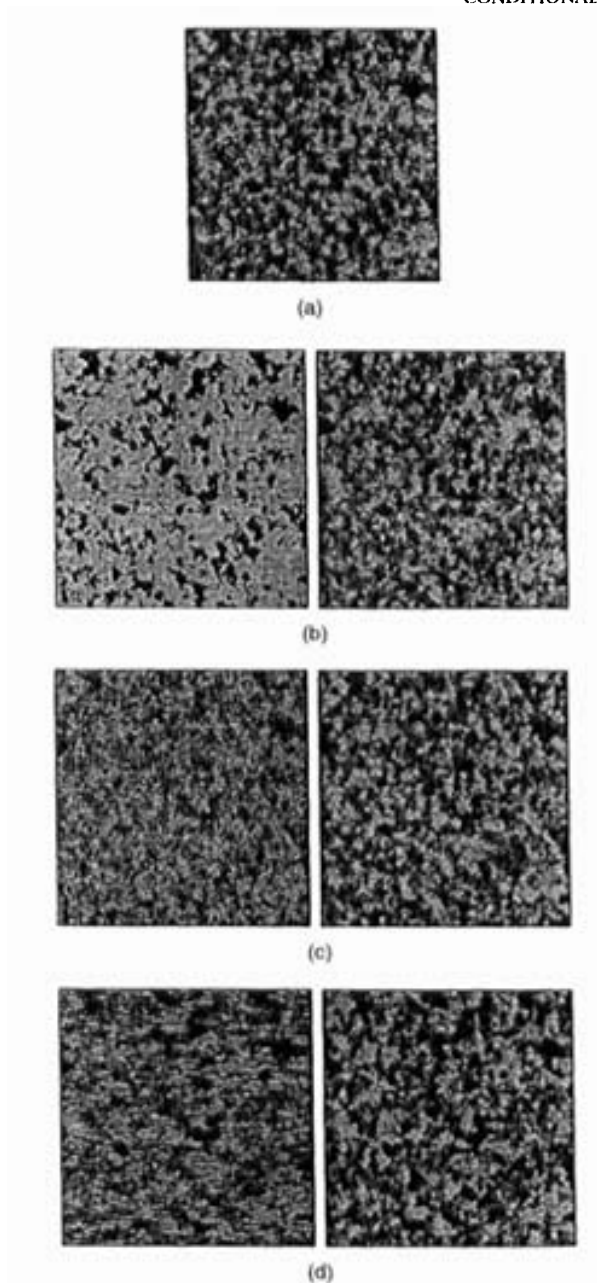


FIGURE 7.23. Simulation of a Gaussian SRF conditional on a truncated or noisy image: (a) initial image; (b) truncated image in black and white (threshold $y_1 = 0$) (left) and simulation of the (unknown) initial image conditional on the truncated image (right); (c) image corrupted by Poisson noise (left) and simulation of the (unknown) initial image conditional on the corrupted image (right); (d) image corrupted by noise and deformed by linear averaging along lines (left) and simulation of the (unknown) initial image conditional on the corrupted and deformed image (right). From Freulon (1994), with kind permission from Kluwer Academic Publishers.

The thresholds y_i are chosen so as to match the proportions p_i of the various indicators

$$y_i = G^{-1} \left(\sum_{j=1}^i p_j \right) \quad i = 1, \dots, m-1$$

Once the correlogram $\rho(h)$ of the underlying Gaussian $Y(x)$ is known, the direct and cross-covariances of the various indicators are known (see Sections 2.5.3 and 6.3.3). The simulation of the I_i reduces to that of Y . In applications we choose directly $\rho(h)$ such that the theoretical variograms of the indicators, deduced from $\rho(h)$ by the relation (2.76) or (6.27) fit the sample variograms well. Indeed these relations are not easily inverted, except if the cutoff is at $y = 0$. Furthermore, while any correlogram $\rho(h)$ gives a valid model of indicator covariance by application of (2.76), it is not certain that conversely, the inversion of this relation is a valid correlogram model, even if we start from an indicator covariance: indeed an indicator covariance is not necessarily the covariance of a truncated Gaussian SRF. Like Figure 7.3g, Figure 7.24 shows indicator simulations obtained by thresholding Gaussian simulations at $y = 0$.

Let us now examine the construction of conditional simulations. If the data are from Y directly, we are in the standard case of Section 7.3. If, on the other hand, the data are from indicators, we proceed in four steps:

1. Global structural analysis of the direct and cross-variograms of the different indicators, with a fit to a model of the form (6.26) derived from a correlogram $\rho(h)$.
2. Simulation of $Y(x)$ at the data points conditional on the indicator values, using the method described in the previous section.
3. Simulation of the whole grid conditional on these simulated hard data.
4. Transformation of the simulation of Y into a simulation of the indicators.

This method can easily be adapted to a regionalization of the proportions of the different indicators and of the corresponding thresholds. It is applied, for example, for simulating lithofacies, as shown in the example presented in Section 7.11.2 (Matheron et al., 1987, 1988). Of course it leads to simulations of diffusive type, in the sense that given (7.41), the facies i can be surrounded only by facies $i-1$ and $i+1$. The method, however, can be generalized to facies that do not follow one another in a fixed order by using two Gaussian SRFs, independent or correlated, to define the facies as shown in Figure 7.25 (truncated plurigaussian simulations: Galli et al., 1994; Le Loc'h and Galli, 1997). Finally, there are adaptations for including connectivity constraints (Allard, 1994).

Simulations of coastlines or islands carried out by Mandelbrot (1975b, 1982) are another example of application, with a single cutoff y_1 correspond-

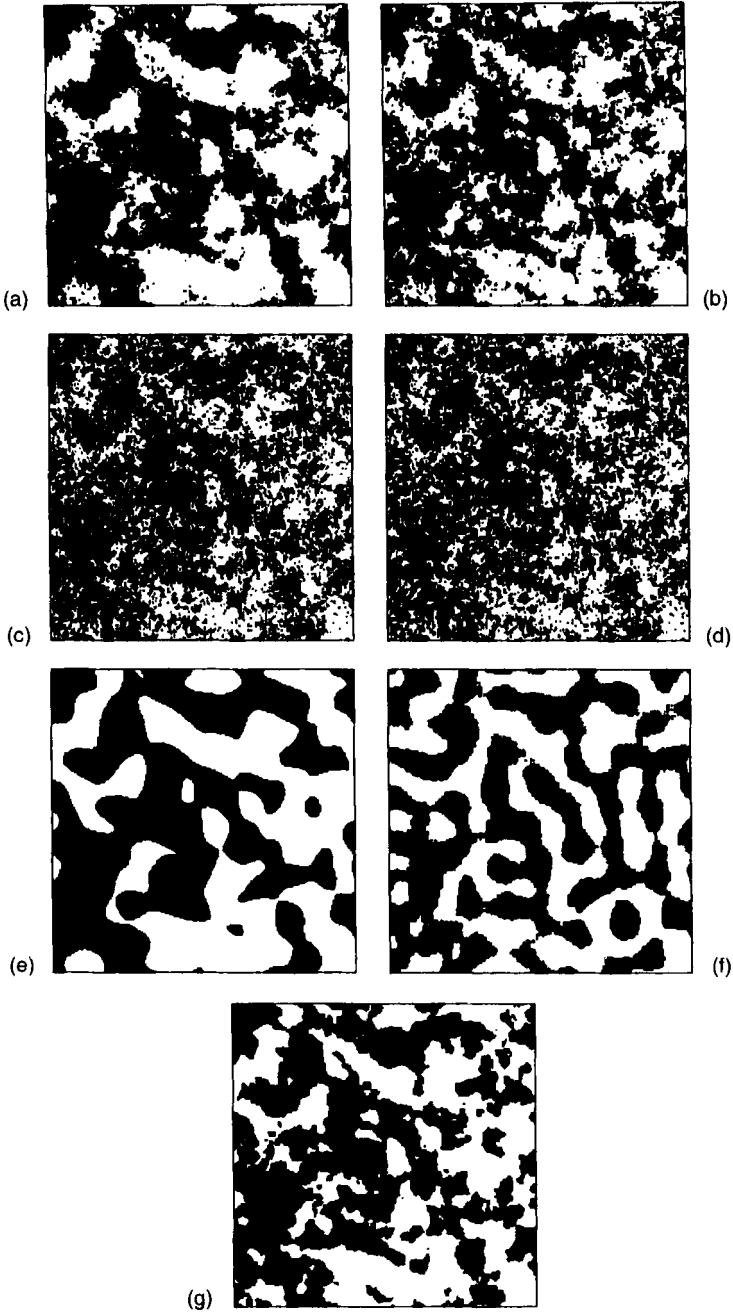


FIGURE 7.24. Examples of simulation of indicators by thresholding a Gaussian simulation. These examples correspond to the threshold $y_1 \approx 0$ of the simulations of Figure 7.15. Note that (g) has an exponential covariance.

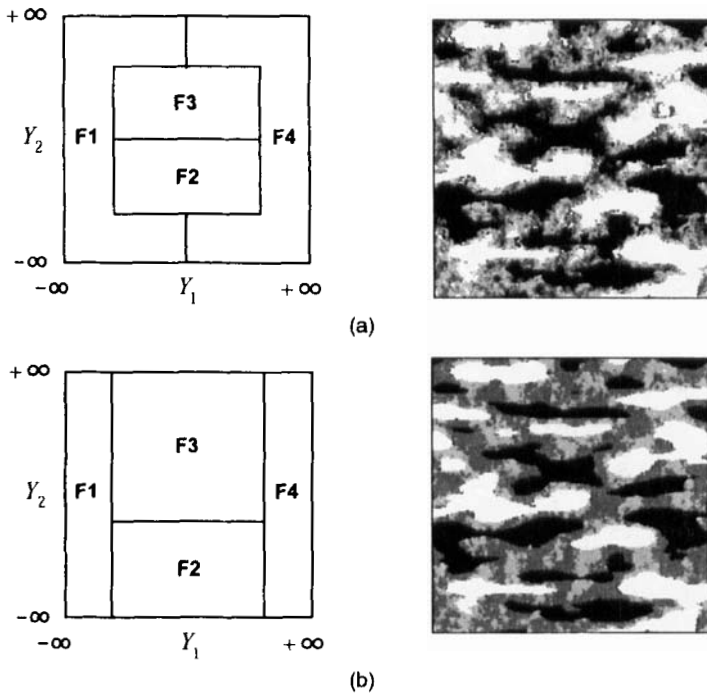


FIGURE 7.25. Truncated plurigaussian simulations. Each facies is defined by a domain in the Y_1, Y_2 plane (bounds can be infinite). In these examples Y_1 has an anisotropic cubic covariance, defining smooth well-marked contours, and Y_2 has an isotropic spherical covariance defining irregular diffuse contours: (a) facies F1 (white) and F4 (black) are only in contact for extreme values of Y_2 ; (b) facies F1 and F4 are never in contact, and the gray space between them is subdivided into facies F2 and F3 by variable Y_2 .

ing to sea level. However, Mandelbrot starts with a simulation of an IRF-0 with a variogram of type $|h|^\alpha$ and not an SRF.

7.7.5. Tessellation

A tessellation is a partition of space into nonoverlapping cells (usually convex polygons or polyhedra). By extension, it denotes a mosaic model based on that partition.

General Principle

Let us consider a stationary partition of \mathbb{R}^n in cells P_i (usually convex polyhedra). A characteristics of this partition is the function

$$g(h) = \Pr\{x \text{ and } x + h \text{ belong to the same cell}\} \quad (7.42)$$

Now assign i.i.d. random values a_i to these cells (independently of the cells), and define $Z(x) = a_i$ for $x \in P_i$. Z is by definition a stationary mosaic random function: two values $Z(x)$ and $Z(x+h)$ are equal with probability $g(h)$, or i.i.d. with probability $1 - g(h)$. If we denote by m and σ^2 the mean and variance of the random variables a_i , and suppose that they are finite, the random function $Z(x)$ has mean m and stationary covariance

$$C(h) = \sigma^2 g(h) \quad (7.43)$$

If the a_i are chosen Gaussian, $Z(x)$ has a Gaussian marginal distribution (but its joint distributions are not Gaussian). Categorical variables are obtained by using discrete distributions. Random sets correspond of course to random variables that are equal to 1 with probability p or 0 with probability $1 - p$, thus to $\sigma^2 = p(1 - p)$. Let us now review some stationary tessellation models.

Renewal Processes

In 1D a renewal process is a point process for which the lengths U_i of the intervals between successive events are independently and identically distributed with a common probability distribution $F(du)$ (e.g., see Feller, 1971, ch. XI). There is no restriction on F other than to be concentrated on $]0, \infty[$ and to have a finite mean μ . Thus

$$\mu = \int_0^\infty u F(du) = \int_0^\infty [1 - F(u)] du$$

Since, in general, the origin is not a point of the process, the distribution of the interval separating the origin from the first event to the right plays a special role. The choice of initial distribution that ensures stationarity of the process is the density

$$f_0(u) = \frac{1}{\mu} (1 - F(u))$$

This is also the distribution of the *residual waiting time* between an arbitrary point x and the next point of the process. The probability that x and $x+h$ belong to the same interval is simply the probability that the residual waiting time exceeds h and is therefore $1 - F_0(h)$. Thus

$$g(h) = \frac{1}{\mu} \int_h^\infty [1 - F(u)] du \quad (7.44)$$

The simplest case is the Poisson point process with intensity λ . Here F and F_0 coincide, $F(u) = F_0(u) = \exp(-\lambda u)$, $u > 0$, and from (7.43) $Z(x)$ has

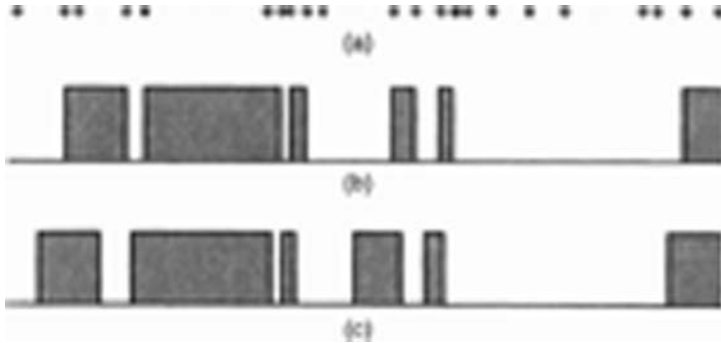


FIGURE 7.26. 1D random sets associated with a Poisson point process: (a) Poisson points; (b) 2-state Markov process (base partition defined by segments between Poisson points); (c) Voronoi process (base partition defined by segments joining interval midpoints).

the exponential covariance

$$C(h) = \sigma^2 \exp(-\lambda|h|) \quad (7.45)$$

with scale parameter $a = 1/\lambda$. If the random variables a_i of the mosaic model take on their values within a finite space of states (e.g., integers from 1 to N), the resulting process $Z(x)$ is a continuous-time homogeneous Markov chain (a jump process in the sense of Feller, 1971, sec. X.3). An example is shown in Figure 7.26a, b for $N = 2$.

Coming back to the general case, the covariance (7.44) is convex, positive, and linear at the origin. Conversely, any covariance function that satisfies these properties can be simulated with a renewal process by taking

$$F(u) = 1 - \frac{C'(u)}{C'(0)} \quad \sigma^2 = C(0) \quad F_0(u) = 1 - \frac{C(u)}{C(0)}$$

An algorithm to simulate the renewal process over the segment $[0, L]$ is the following:

1. Select the first point of the process to the right of the origin from the distribution F_0 .
2. Simulate the next segments independently from the distribution F until the end point L is reached.

Generalization to a Concave Unbounded Variogram

We note that if $C(L)$ is not equal to 0, there is a nonzero probability that the first point on the right side of the origin falls after L . In this case the simulation is constant over $[0, L]$. To avoid this, a slight modification is introduced so that the first point always belongs to the interval $[0, L]$. To obtain this without introducing a bias on the covariance, Mathéron (1988) showed that it suffices

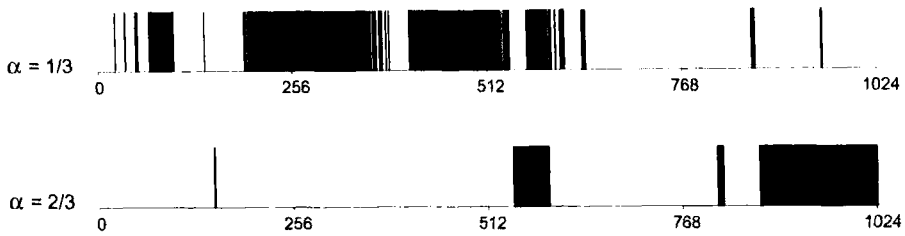


FIGURE 7.27. 1D simulations of a mosaic IRF-0 (here a random set) with a variogram locally of type $|h|^\alpha$, for $\alpha = \frac{1}{3}$ and $\alpha = \frac{2}{3}$. The simulations are built at 1025 discrete points. The values equal to 1 are represented by vertical bars. These simulations have fractal dimension $D = 2 - \alpha$. The upper graph resembles observations of fractures along a well.

to use the previous algorithm with

$$F(u) = 1 - \frac{\gamma'(u)}{\gamma'(0)} \quad \sigma^2 = \gamma(L) \quad F_0(u) = \frac{\gamma(u)}{\gamma(L)}$$

where $\gamma(h) = C(0) - C(h)$ is the variogram of $Z(x)$.

The algorithm has a broader application domain than convex covariances with a linear behavior at the origin: it can be applied to simulate any concave variogram with a linear behavior at the origin, even if this variogram is not bounded. It has been generalized by Matheron (1988) to simulate any concave variogram, even if not linear at the origin. In this case it is necessary to consider a generalization of renewal processes to point processes with accumulation points, and the simulation is built at discrete points. Let us simply give the algorithm for simulating a mosaic IRF-0 with variogram $\gamma(h) = b|h|^\alpha$, $0 < \alpha < 1$, at $x = 0, 1, \dots, L$ (L integer).

1. Select $u \in [0, L]$ from the distribution $F_0(u) = (u/L)^\alpha$. Set $u_0 = u$. From $j = 0$ to $j_0 = \text{Int}(u_0)$, set $Z(j) = a_0$, where a_0 is chosen randomly from a distribution with variance $\sigma^2 = bL^\alpha$.
2. Select independently Y_1 from a gamma distribution with shape parameter α and scale parameter b , and Y_2 from a gamma distribution with shape parameter $(1 - \alpha)$ and same scale parameter b . Set $x_0 = j_0 + 1 - u_0$ and $u_1 = x_0 Y_1 / Y_2$. From $j = j_0 + 1$ to $j_1 = j_0 + 1 + \text{Int}(u_1)$, set $Z(j) = a_1$.
3. Select new random variables Y_1 and Y_2 . Set $x_1 = j_1 + 1 - u_1$ and $u_2 = x_1 Y_1 / Y_2$. From $j = j_1 + 1$ to $j_2 = j_1 + 1 + \text{Int}(u_2)$, set $Z(j) = a_2$.
4. Repeat until the point $j = L$ is reached.

The simulations obtained with this algorithm are fractal, but with Hausdorff dimension $D = 2 - \alpha$ instead of $2 - \alpha/2$ for a Gaussian IRF-0 with the same variogram (fractional Brownian random function). They are not very ergodic. In the limiting case $\alpha = 1$, the algorithm tends to a simulation with a single discontinuity and is no longer fractal. Figure 7.27 shows simulations with $L = 1024$ for $\alpha = \frac{1}{3}$ and $\frac{2}{3}$.

The generalization of renewal processes to \mathbb{R}^n , $n > 1$ is simple only in the case of Poisson processes. We will examine now the two types of generalization based, respectively, on Poisson point processes in \mathbb{R}^n and Poisson hyperplanes.

Voronoi Tessellation

Consider points $\{x_i : i = 1, 2, \dots\}$ in \mathbb{R}^n , and associate to each x_i the set P_i of all points x in \mathbb{R}^n that are closer to x_i than to any other x_j , $j \neq i$. In 1D the P_i are the segments joining the midpoints of

the intervals $[x_i, x_{i+1}]$ (the x_i are assumed to be ordered by increasing values). In 2D they are the convex polygons delimited by the perpendicular bisectors of the segments joining neighboring points (Fig. 7.28a). These polygons are usually named Voronoi polygons (metallogeny), or in other contexts polygons of influence (mining), Thiessen polygons (hydrology), cellular networks, Dirichlet polygons (see Stoyan et al., 1987). More generally, in \mathbb{R}^n the P_i are convex polyhedra. If we randomize the x_i into a stationary random point process of points X_i , the P_i define a partition of space into a stationary random set of polyhedra, from which we can build a mosaic random function whose covariance is given by (7.43). Figure 7.28b exhibits an RF with a Gaussian marginal distribution, and Figure 7.28c a random set.

The problem is to obtain the function $g(h)$ defined by (7.42). Matérn (1960, p. 40) derives it when the X_i form a Poisson point process in \mathbb{R}^n . The result has a simple expression in the 1D case (Fig. 7.26c). Each Voronoi segment is the half of two successive segments of the Poisson process; hence the lengths of the Voronoi segments are gamma random variables with parameter $\alpha = 2$ and index $b = 2\lambda$, where λ is the Poisson point intensity. Two successive Voronoi segments contain one-half of the same Poisson segment, and therefore are not independent and do not define a renewal process. However formula (7.44) remains applicable—it can be derived directly by a geometric argument involving only stationarity—and combined with (7.43) yields the covariance

$$C(h) = \sigma^2(1 + \lambda|h|)\exp(-2\lambda|h|)$$

In \mathbb{R}^2 , Poisson-Voronoi polygons, namely Voronoi polygons associated with a planar Poisson point process, pave the plane with quasi-hexagonal polygons and are therefore of great interest in many applications. Their main properties were derived by Miles (1970). Other results have been obtained by Crain (1978) by Monte Carlo techniques. Miles (1972) presents the properties of the Poisson-Voronoi polyhedra in \mathbb{R}^3 .

Poisson Hyperplanes

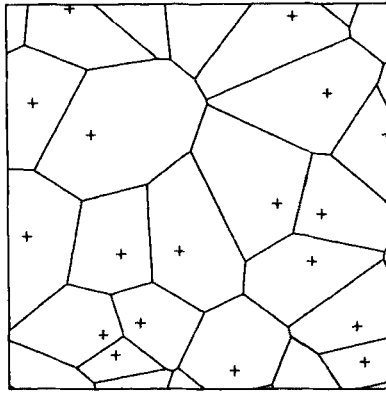
Instead of starting from random points in the plane or in space, one can start from “random lines” or “random planes.” This permits the simulation of random functions with an exponential covariance (or a linear variogram; see Section 7.6.2).

The principle for generating Poisson lines in \mathbb{R}^2 , Poisson planes in \mathbb{R}^3 , and more generally Poisson hyperplanes in \mathbb{R}^n is to start from a Poisson point process with intensity $\lambda_n d\omega$ on a line D_ω going through the origin of space and with direction the central direction of the solid angle $d\omega$. Through each of these points we erect a line, plane, or hyperplane perpendicular to D_ω , and repeat this for all directions ω of the unit half-sphere. Thus we obtain a network of lines or planes that appear random because no orientation or region of space is favored (generalizations are of course possible). Mathematical developments are beyond our present scope and can be found in Miles (1969) and Matheron (1971a, 1975a).

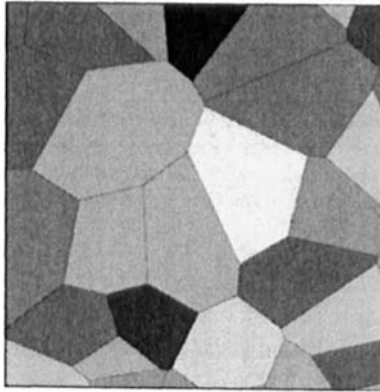
The intersection of a subspace \mathbb{R}^m with a Poisson hyperplane process of \mathbb{R}^n ($m < n$) produces a Poisson hyperplane process of \mathbb{R}^m . In particular, the intersection of a line with a Poisson hyperplane process forms a Poisson point process whose intensity is independent of the line considered. The intensity λ of the induced Poisson point processes on lines may be taken as the parameter of the Poisson hyperplane processes in \mathbb{R}^n and its subspaces.⁵

Consider now the case where we want to simulate Poisson hyperplanes within a hypersphere with diameter D centered at the origin. Each hyperplane is characterized by two parameters (Fig. 7.29):

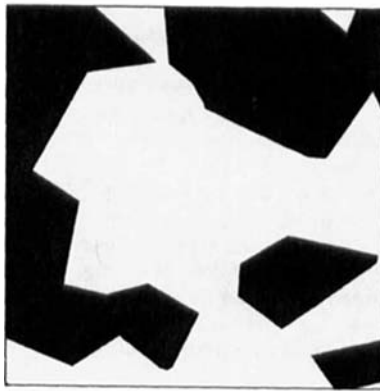
1. Its distance d to the origin.
2. The direction ω of the unit vector orthogonal to the hyperplane, directed from the origin to the hyperplane.



(a)



(b)



(c)

FIGURE 7.28. 2D mosaic simulations associated with a Voronoi partition: (a) Voronoi polygons; (b) mosaic RF with Gaussian marginal; (c) mosaic random set ($p = 0.5$). The initial point process is made of Poisson points.

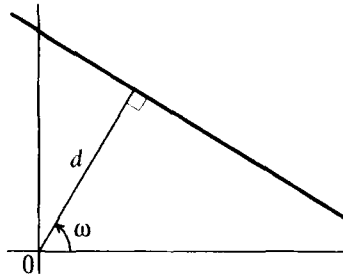


FIGURE 7.29. Parameters defining a line (2D): direction $\omega \in [0, 2\pi[$ and distance d .

Equivalently the hyperplane is characterized by the vector joining the origin to the projection of the origin onto the hyperplane. Now rotate the vectors associated with the Poisson hyperplanes so as to align them with the positive side of the first coordinate axis. The end points of the rotated vectors define a point process on the positive part of this axis. This process is also a Poisson point process. A geometric argument shows that its intensity is $\lambda S_n/V_{n-1}$, where S_n is the surface area of the unit radius sphere of \mathbb{R}^n and V_n its volume (as results from formulas (A.5), the ratio S_n/V_{n-1} is equal to 2, π , or 4 when n equals 1, 2, or 3, respectively). Hence we have the following procedure for simulating the Poisson hyperplanes that intersect a hypersphere with diameter D :

1. Generate a 1D Poisson point process with intensity $\lambda S_n/V_{n-1}$ over the segment $[0, D/2]$.
2. To each point X_i of this Poisson process, associate a random orientation ω_i drawn independently from a uniform distribution over the unit hypersphere, and erect the hyperplane characterized by the distance $d_i = |X_i|$ and the orientation ω_i .

This model can be easily extended to a nonuniform distribution of the orientations ω_i .

Poisson Polyhedra Tessellation

Poisson lines form a partition of the plane into random convex polygons (Fig. 7.30). Similarly Poisson hyperplanes divide the space \mathbb{R}^n into random convex polyhedra. Let us assign a constant value to all the points of a polyhedron and draw this constant independently from a distribution with variance σ^2 . Since the intersection of the Poisson hyperplanes with any line defines a 1D Poisson point process with intensity λ , this mosaic random function is an SRF with covariance

$$C(h) = \sigma^2 \exp(-\lambda|h|) \quad (7.46)$$

This is covariance (7.45), but this time h belongs to \mathbb{R}^n . Figure 7.3f shows a RF with a Gaussian marginal distribution, and Figure 7.3h a random set. The hyperplanes stand out because the boundaries of the random sets are aligned on them. In figurative terms, Voronoi models consider crystals that grew from germs, whereas this model is based on a network of faults. The typical polygons obtained in 2D have very different features than those of Voronoi polygons; in particular, the mean number of vertices of these polygons is equal to 4 instead of 6 for Voronoi polygons. Results can be found in Miles (1964, 1973, in \mathbb{R}^2 ; 1972 in \mathbb{R}^3 ; 1971 in \mathbb{R}^n), Matheron (1971a, 1972a, 1975a) in \mathbb{R}^2 , \mathbb{R}^3 , and \mathbb{R}^n , and also Stoyan et al. (1987).

Let us mention a variant, the *alternating random set*, which is simulated as follows: draw at random, with equal probability, the value $m - \sigma$ or $m + \sigma$ (0 or 1 for a random set, with $m = \sigma = \frac{1}{2}$) for the polygon or polyhedron that contains the origin; assign the alternate value to the polyhedra that have a face in common with this polyhedron; and iterate. $Z(x)$ and $Z(x + h)$ have therefore identical or alternate values depending on whether the number of Poisson hyperplanes intersecting the segment joining x to $x + h$ is even or odd. Simple probabilistic calculations show

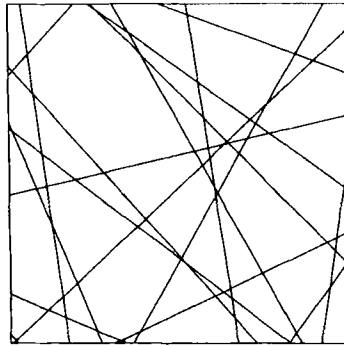


FIGURE 7.30. Simulation of Poisson lines and Poisson polygons tessellation.

(Blanc-Lapierre and Fortet, 1953, pp. 187–191) that the centered covariance is

$$C(h) = \sigma^2 \exp(-2\lambda|h|)$$

By dividing the intensity by a factor of 2, we obtain the same covariance as in (7.46) for very different simulations (Fig. 7.3j). But we have lost a degree of freedom, since the proportion associated with the value 1 is always $p = \frac{1}{2}$.

7.7.6. Substitution Random Functions

Definition and Properties

Substitution random functions (Matheron, 1989; Lantuéjoul, 1991, 1993) are the multidimensional generalization of the subordinated processes defined by Feller (1971, sec. X.7). The shift in terminology emphasizes the stereological content of the construction, which transfers certain properties of a coding process to a random function.

An RF $Z(x)$ is said to be a substitution random function if it is of the form $Z(x) = Y(T(x))$, where $T(x)$ is a random function of $x \in \mathbb{R}^n$ and $Y(t)$ a stochastic process of $t \in \mathbb{R}$. $T(x)$ is called the directing function, and $Y(t)$ the coding process. Since all kinds of models for $T(x)$ and $Y(t)$ can be used, the substitution method gives rise to a broad variety of random functions. If $T(x)$ and $Y(t)$ are continuous RFs, the isolines of $T(x)$ remain isolines of $Z(x)$: the substitution simply changes the isoline values. If $T(x)$ is constant over domains that partition \mathbb{R}^n , so is also $Z(x)$: the substitution simply changes the values assigned to the various domains. So the most interesting case is when $T(x)$ is a continuous RF and $Y(t)$ a discrete-state process. We will restrict this presentation to the case studied in detail by Lantuéjoul (1993), where $T(x)$ has strictly stationary increments, $Y(t)$ is a stationary continuous-time Markov chain with covariance $C(h)$, and $T(x)$ and $Y(t)$ are independent. This model is a generalization of the discrete diffusion random functions of Section 6.3.3,

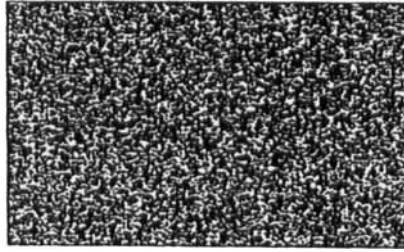


FIGURE 7.31. Simulation of a substitution RF by Markov coding of an RF whose discontinuities have a stable distribution with index $\alpha = 0.47$. From Lantuéjoul (1993), with kind permission from Kluwer Academic Publishers.

whose nice properties carry over in a straightforward manner:

- $Z(x)$ is ergodic provided that $T(x)$ is not stationary (if it were stationary this would restrict the range of the variations of T around the mean, and lead to the use of a small portion of the definition domain of $Y(t)$).
- $Z(x)$ has the same univariate distribution as $Y(t)$.
- $Z(x)$ is second-order stationary:

$$\text{Cov}[Z(x), Z(x+h)] = E[C(|T(x+h) - T(x)|)]$$

an expression which does not depend on x since T has stationary increments.

- If $Y(t)$ has an isofactorial expansion, $Z(x)$ has an isofactorial expansion with the same factors as $Y(t)$.

Examples

Lantuéjoul (1991, 1993) shows several 2D examples of such substitution random functions that demonstrate the capability of this model to represent very different morphologies. A simple model is the following:

- The directing function $T(x)$ is a 2D IRF-0 associated with random jumps on Poisson lines, defining a linear variogram, as explained in Section 7.6.2.
- The coding process $Y(t)$ is a two-state continuous-time Markov chain, whose sojourn time distributions in the states 0 and 1 have densities $b_0 \exp(-b_0 t)$ and $b_1 \exp(-b_1 t)$, respectively.

The intensity of Poisson lines is a scale factor, whereas the distribution of the random jumps has a large influence on the shape on the simulation:

- Figure 7.31 shows a simulation of $Z(x)$ when the discontinuities have a stable distribution, namely a characteristic function $\Phi(u) = \exp(-|u|^\alpha)$ with $\alpha = 0.47$. It gives a strong impression of homogeneity, like a texture.

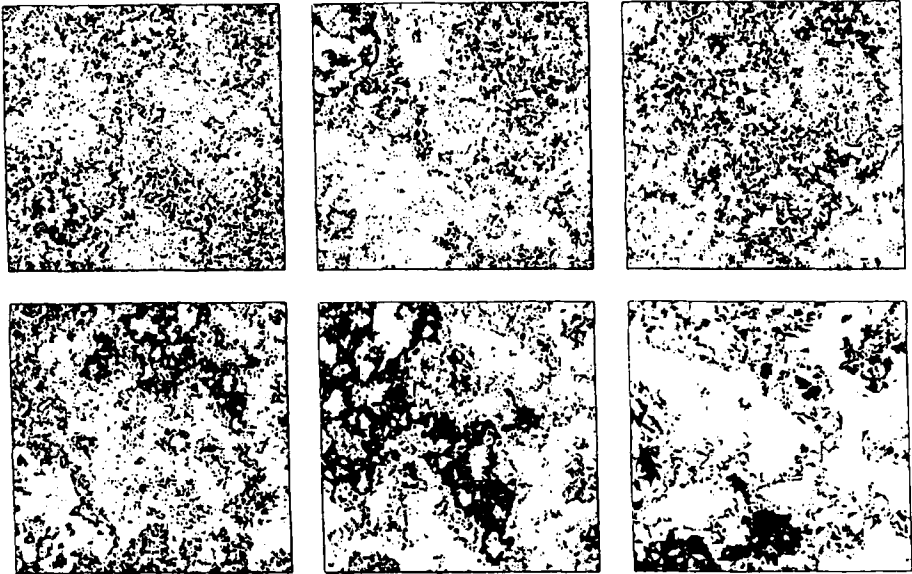


FIGURE 7.32. Simulation of a substitution RF by Markov coding of an RF whose discontinuities are ± 1 . Each image is the enlargement by a factor 2 of the upper right quarter of the previous image. From Lantuéjoul (1993), with kind permission from Kluwer Academic Publishers.

This is due to the fact that the intensity of random lines is high and the jump distribution is heavy-tailed, so very different values of T may be assigned to two neighboring points.

- Figure 7.32 displays a similar simulation when the discontinuities take on the values -1 and $+1$ only, as in the standard simulation of a linear variogram. The upper right corner of the simulation is presented at 6 different scales, $1/32$, $1/16$, $1/8$, $1/4$, $1/2$, 1 . At each scale new structures emerge, due to the slow variations of T , although, strictly speaking, this random function is not fractal given the underlying construction with Poisson lines. (These images are spectacular in color; see Lantuéjoul, 1991.)

So two variants of the same algorithm can lead to very different results. The first one gives a covariance with a finite integral range (this is true when $\alpha < 0.5$), whereas the second one leads to an infinite integral range.

Conditioning

The nonconditional simulation of a substitution random function follows directly from its definition, provided of course that we know how to simulate the independent underlying random functions $T(x)$ and $Y(t)$. But conditioning such a simulation is not so simple. If the conditioning data at each sample

point x_α were the values $t_\alpha = T(x_\alpha)$ of the directing function and $z_\alpha = Y(t_\alpha)$ of the coding process, we could meet our objective in two conditioning steps:

1. Conditional simulation of $T(x)$ given the data $T(x_\alpha) = t_\alpha$.
2. Conditional simulation of $Y(t)$ given the data $Y(t_\alpha) = z_\alpha$.

Note that from a practical point of view, $T(x)$ must be simulated at the nodes of the simulated grid, whereas $Y(t)$ must be simulated at the points defined by the values taken on by $T(x)$. But, in practice, the conditioning has to be based on the data $z_\alpha = Z(x_\alpha)$ only. Before applying the two conditioning steps, a preliminary step is to simulate the $T(x_\alpha)$. The principle is to first build a nonconditional simulation of $Y(t)$, and then simulate the $T(x_\alpha)$ under the constraints that $Y(T(x_\alpha)) = z_\alpha$. Such a constraint means that $T(x_\alpha)$ belongs to one of the intervals where $Y(t)$ is in the state z_α . This is a case of conditioning by inequalities except that we are considering unions of intervals rather than single intervals. For Gaussian RFs the solution proposed in Section 7.7.3 can be extended to the present case. Lantuéjoul (1993) presents an example of such conditioning.

7.8. OBJECT-BASED SIMULATIONS: BOOLEAN MODELS

We are considering here a class of models, the Boolean models and their variations, obtained by combining *objects* placed at random points. These models can also be considered as *marked point processes*, in the sense that they are based on a point process and marks (here the objects) attached to the points of the process (Stoyan et al., 1987, sec. 4.2). They constitute a family of very flexible models and are sometimes used on the basis of a physical or genetic interpretation which defines the objects of the particular model used. They are also used without reference to any plausible physical interpretation when they produce an acceptable fit of the observations.

7.8.1. Boolean Models

Boolean Random Sets

A Boolean random set corresponds to the intuitive idea of the union of randomly located objects. The main theoretical results are due to Matheron (1967, 1975a), though special forms have been used earlier (see Stoyan et al., 1987, p. 68).

Consider a process of Poisson points X_i in \mathbb{R}^n , $i = 1, 2, \dots$, and let A_i , $i = 1, 2, \dots$, be i.i.d. random objects (elements of volume, or more precisely compact—i.e., closed and bounded—subsets of \mathbb{R}^n). The union of the A_i shifted to points X_i constitutes by definition a Boolean random set X :

$$X = \bigcup_i \tau_{X_i} A_i \quad (7.47)$$

where τ_h denotes the operator of translation by a vector h (Fig. 7.33a). (7.47) is the set formulation of the dilution approach (7.23).

The points X_i of the Poisson process are referred to as germs and the compacts A_i as primary objects. In this model the germs grow independently until they form primary objects that can overlap. The space \mathbb{R}^n is thus divided into two phases, the objects X which is the union of the primary objects, and its complement, the background.

A very simple example consists in taking spheres with a given diameter for the objects. Variants presented in the framework of the dilution methods apply here also: for example, spheres with random and independent diameters. Figure 7.33b, c presents two 2D simulations where the objects are Voronoi polygons and Poisson polygons, respectively.

A Boolean model X depends on two parameters, the intensity λ of the Poisson point process and the probability distribution of the primary objects. In general, a random closed set X is characterized by its *Choquet capacity*, or *hitting functional*, namely the mapping T that associates to each compact subset B of \mathbb{R}^n the probability that B hits X

$$T(B) = \Pr\{X \cap B \neq \emptyset\}$$

In the case of Boolean random sets, instead of evaluating $T(B)$ directly, we evaluate the probability $Q(B) = 1 - T(B)$ that B is contained in the background. It can be shown that this is written in the form (Matheron, 1967; Serra 1982)

$$Q(B) = e^{-\lambda E|A \oplus \check{B}|} \quad \left(A \oplus \check{B} = \bigcup_{x \in A} \check{B}_x \right)$$

where \check{B} denotes the symmetric of B with respect to the origin. If B is reduced to a single point, we obtain the background probability of the Boolean model

$$q = e^{-\lambda E|A|} \quad (7.48)$$

If $B = \{x, x + h\}$ is a pair of points, then $|A \oplus \check{B}| = |A \cup A_{-h}| = 2|A| - |A \cap A_{-h}|$. Denoting by $K(h) = E|A \cap A_{-h}|$ the mean geometric covariogram of the primary objects, the bivariate distribution of the background of the Boolean model is

$$\Pr\{x \in \bar{X}, x + h \in \bar{X}\} = q^2 e^{\lambda K(h)} = q e^{-\lambda[K(0) - K(h)]} \quad (7.49)$$

Therefore the covariance of the background, which is also that of the objects, is

$$C(h) = q[e^{-\lambda[K(0) - K(h)]} - e^{-\lambda K(0)}] \quad (7.50)$$

If B is a segment or a ball, the explicit calculation of $|A \oplus \check{B}|$ is possible only if A is almost certainly convex. If so, it can be shown, for example (take

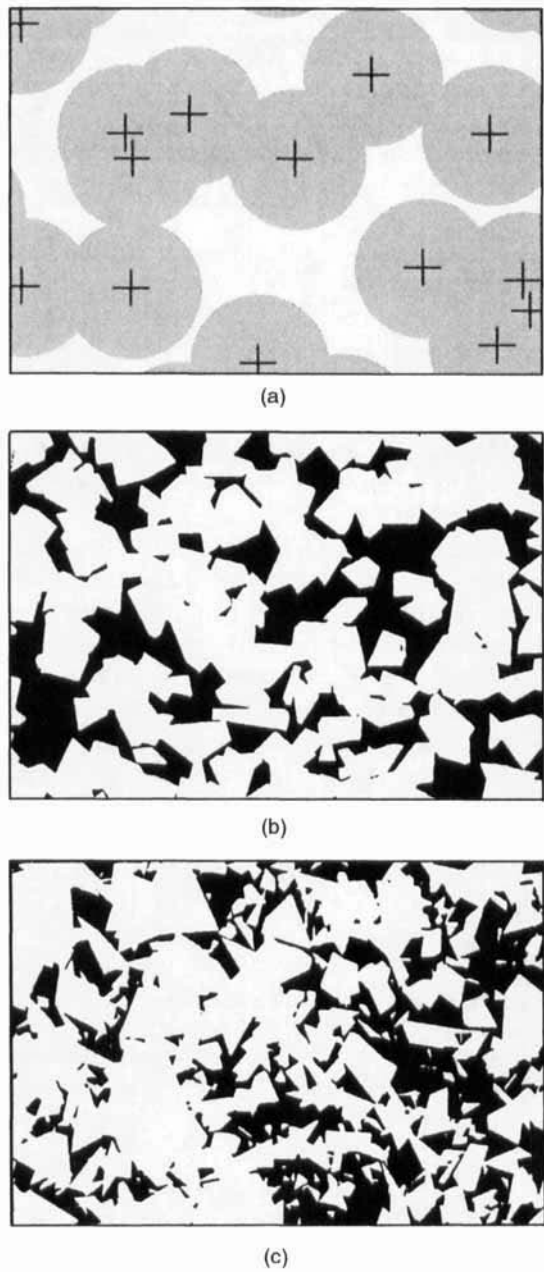


FIGURE 7.33. Realizations of Boolean random sets in 2D. The objects are: (a) disks; (b) Voronoi polygons; (c) Poisson polygons (Delfiner, 1970). From C. Lantuéjoul, personal communication.

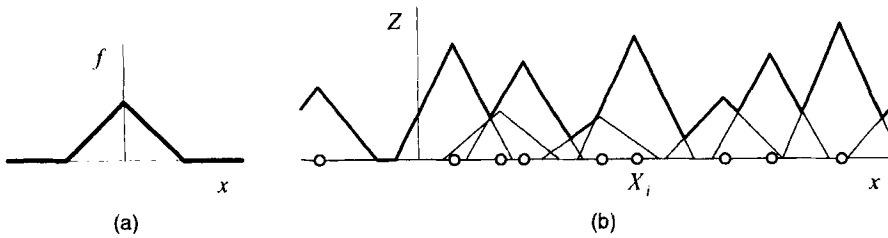


FIGURE 7.34. Construction of a simulation of a Boolean RF (1D): (a) primary function (a cone); (b) resulting simulation with cones with random height.

for B a segment), that the distribution of background intercepts is exponential. We note that the shape of the covariance function of a Boolean random set depends on λ , unlike the case of dilution SRFs. The background probability also depends on λ . The intensity λ is therefore an essential parameter of a Boolean random set.

These results are useful for testing the validity of a Boolean model with convex primary objects and for attempting the statistical inference of its parameters. When we study a random set in \mathbb{R}^n , we frequently examine the intersection of the random set with a hyperplane. The overall study is made easier by the fact that the intersection of a Boolean random set with a hyperplane is still a Boolean random set. On the other hand, in the subspace of this hyperplane this intersection can be a Boolean random set of convex objects without this being the case in \mathbb{R}^n . In practice, even if we have a complete image, statistical inference is tricky due to edge effects, and various methods have been developed to deal with this (see Serra, 1982, sec. XIII.B.8; Stoyan et al., 1987, sec. 3.3; Lantuéjoul and Schmitt, 1991, for a comparison of more recent methods).

Boolean Random Functions

A generalization of Boolean random sets to RFs is to set up at each point X_i of the Poisson process a realization of a nonnegative random function f_i , called primary function, drawn independently from one point to another. Instead of cumulating the effects of these functions as for dilution RFs, we now take the maximum and let

$$Z(x) = \sup_i f_i(x - X_i) \quad (7.51)$$

The graph of the Boolean random function $Z(x)$ is the envelope of the subgraphs of the primary functions (Fig. 7.34). Each primary function delimits a geometric object in \mathbb{R}^{n+1} defined as the space below the graph of the function. If these objects are convex, it is possible to calculate the covariance of $Z(x)$. In particular, if they are cylinders, the covariance includes a linear term at the origin. If they are cones, $Z(x)$ is much more regular, and its covariance is parabolic at the origin. More generally, consider the random primary object defined by the cutoff z on the primary function f ,

$$A_z = \{x \in \mathbb{R}^n : f(x) \geq z\}$$

The union of all such primary objects set up at Poisson points X_i forms a Boolean random set whose background probability is by (7.48),

$$F(z) = P\{Z(x) < z\} = \exp(-\lambda E|A_z|)$$

Further it can be shown that the bivariate distribution function is given by

$$F_h(z_1, z_2) = P\{Z(x) < z_1, Z(x+h) < z_2\} = F(z_1)F(z_2)\exp(\lambda E|A_{z_1} \cap \tau_h A_{z_2}|)$$

which generalizes (7.49).

The case of cylindrical objects corresponds to functions f_i of the form

$$f_i(x) = a_i 1_{A_i}(x)$$

and can already be used to describe a variety of situations. The A_i can belong to several classes of objects, that we number 1, 2, and so on. If a_i is the class number of the object A_i , we obtain a simulation where the objects of a class override the objects of the lower classes. However, if the a_i are i.i.d. random variables, there is no hierarchy between the objects. In both cases, if the a_i take on a finite number of possible values, corresponding, for example, to a facies number, two adjacent facies do not necessarily have consecutive facies numbers (Fig. 7.35). The bivariate distributions of such simulations follow an isofactorial model with orthogonal indicator residuals (see Section 6.3.3).

One of the attractions of Boolean random functions is that they are well adapted to the change of support by the maximum (dilation in the case of random sets). This type of change of support occurs, for example, if the a_i represent a pollutant level rather than a facies number (although this level may depend on the facies) and if we characterize the pollution of a zone by the highest local concentration in this zone.

Simulating such RFs is very simple (relation (7.51)). The difficulty is elsewhere, namely the design of criteria to establish that this model is compatible with the available observations and the identification of its parameters (intensity of the Poisson germs process, choice of the family of primary functions, and distribution of its parameters). Boolean RFs have been introduced by Jeulin and Jeulin (1981) to model the roughness of metallic fractures. The theory is developed by Serra (1988). The characterization and quantification of Boolean RFs from images is studied by Prêteux and Schmitt (1988). A presentation can be found in Chautru (1989). Further references and variations about this model are given by Jeulin (1991).

Dead-Leaves Models and Sequential Random Functions

If the objects of a Boolean random set are of nonzero size and if the intensity λ tends to infinity, the objects end up occupying all the space. This is the basis for the dead-leaves model (Matheron, 1968b). In this terminology the objects are replaced by leaves, which progressively cover the ground, hiding parts of dead leaves fallen previously (e.g., Serra, 1982, sec. XIII.C.2). Looked at from above at a fixed time, this defines a partition of the space whose classes are the visible portions of the leaves not entirely hidden by the subsequent ones (Fig. 7.36a). A mosaic model may then be defined from this tessellation, whose covariance is proportional to $K(h)/[2K(0) - K(h)]$, where $K(h)$ is the mean geometric covariogram of the primary leaves. A variation is the multi-dead-leaves model (Jeulin, 1979) where the leaves belong to different populations (based on type of tree, or color), which defines a multiphase tessellation (Fig. 7.36b).

A similar generalization of Boolean random functions to dead-leaves random functions is proposed by Jeulin (1989). All these models can be considered as sequential random functions if we look at their evolution as the leaves are falling. Jeulin (1991, 1997) presents additional models of sequential random functions as well as methods for their statistical inference from an image.

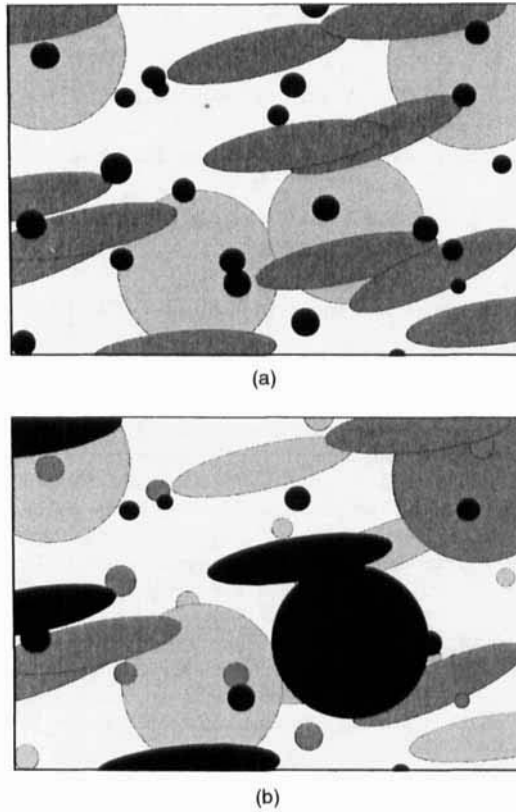


FIGURE 7.35. Realizations of Boolean RFs. The subgraphs of the primary functions are cylinders with three types of base (ellipse, large disk, small disk): (a) with hierarchy—the cylinder height (i.e., the value attached to the base) is a function of the base type (here highest for small disks, and lowest for large disks); (b) without hierarchy—the cylinder height is independent of the base type.

7.8.2. Stationary Point Processes

Procedures to construct random sets or random functions from Poisson point processes can be generalized to other stationary point processes. For example, we can define dilution random functions and Voronoi tessellations based on an arbitrary distribution of points. But this generalization is mainly applied to Boolean models and is therefore presented here. There are so many models of point processes that we will not try to be exhaustive. The interested reader is referred to the point process literature, such as Bartlett (1963, 1975), Krickeberg (1974), and Ripley (1977, 1981) and to the random set literature, such as Serra (1982) and Stoyan et al. (1987). We will only present some simple models shown in Figure 7.37 and a structural tool for identifying them: the K -function.

Cox Process

The Poisson process can be immediately generalized to the case where the intensity λ is not constant but is a measurable function $\lambda(x)$ with positive values. The number of points falling in

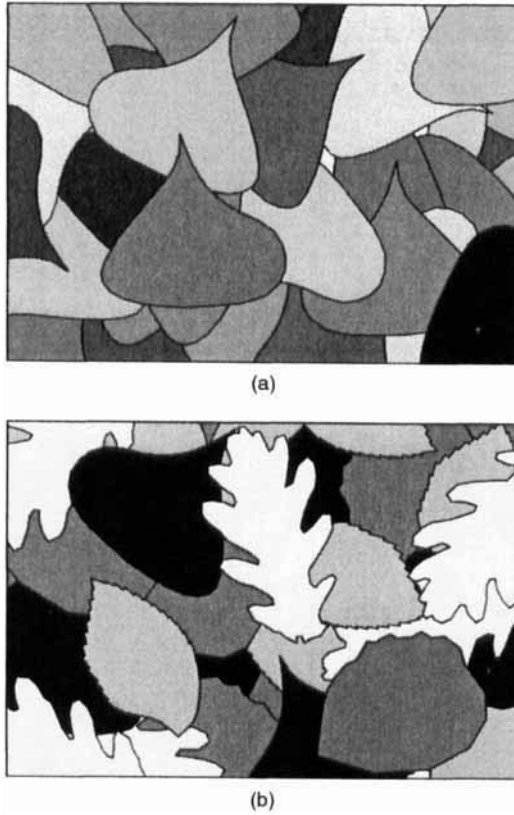


FIGURE 7.36. Realizations of mosaic random functions derived from dead-leaves models: (a) single-dead-leaves model (black poplar) with independent assignment of a value to each leaf; (b) multi-dead-leaves model with value assignment dependent on leaf species (alder, elm, oak, poplar).

V is a Poisson random variable with mean

$$\lambda(V) = \int_V \lambda(x) dx$$

These points are distributed independently according to the probability density function $\lambda(x)/\lambda(V)$, $x \in V$. The covariance between $N(V)$ and $N(V')$ is

$$\text{Cov}(N(V), N(V')) = \lambda|V \cap V'|$$

This process is not stationary. To make it stationary, it suffices to replace the regionalized intensity $\lambda(x)$ by a nonnegative stationary random function $\Lambda(x)$. We then obtain the Cox process or *doubly stochastic Poisson process* (Cox, 1955; Fig. 7.37b). $N(V)$ is distributed as a Poisson with random parameter

$$\Lambda(V) = \int_V \Lambda(x) dx$$

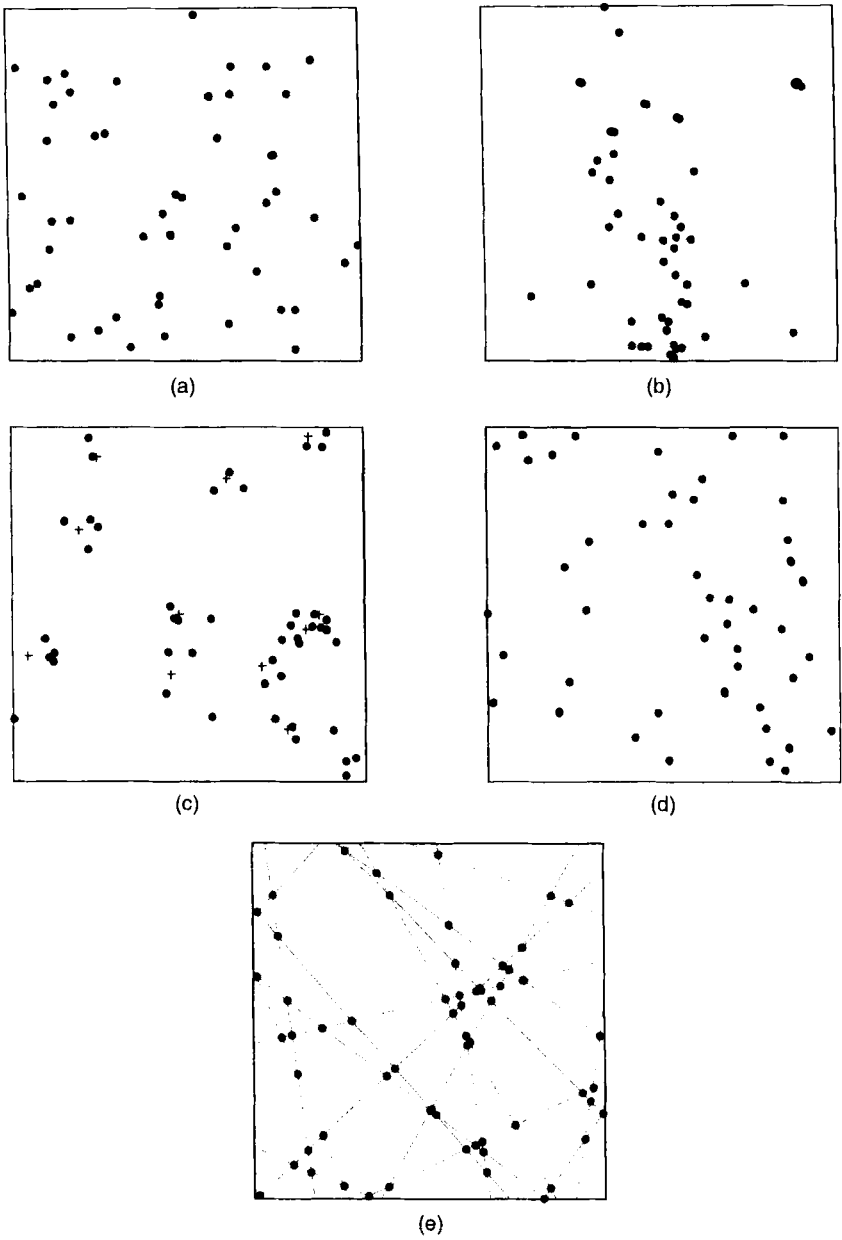


FIGURE 7.37. Simulation of point processes: (a) Poisson points; (b) Cox process (the intensity is a lognormal RF); (c) Poisson cluster process; (d) hard-core process; (e) Poisson alignments.

The probabilities and the noncentered moments are obtained by randomizing the expressions obtained for the Poisson process with regionalized intensity. In the end, and denoting respectively by m and $C(h)$ the mean and the covariance function of the intensity Λ , we get

$$E(N(V)) = m|V|$$

$$\text{Cov}(N(V), N(V')) = m|V \cap V'| + \int_V \int_{V'} C(y-x) dx dy$$

Then, even if V and V' do not overlap, $N(V)$ and $N(V')$ are generally not independent. Their covariance is the sum of two terms, the first accounting for the Poisson distribution of the points once the intensity is fixed, the second capturing the intensity fluctuations. This hierarchic character is not rare in practice. A typical example, studied by Kleingeld and Lantuéjoul (1993), is that of an alluvial deposit of precious stones where the gems are trapped in anfractuosités of the terrain. Where the terrain is regular, one expects to find few precious stones, whereas in rugged terrain one may find a cluster of stones. The Cox process has a number of remarkable properties. For example, if $N(V)$ follows a negative binomial distribution, then $\lambda(x)$ follows a gamma distribution (Feller, 1971, pp. 55–57; the negative binomial distribution is termed “the limiting form of the Pólya distribution”). Similarly, if $N(V)$ follows a Sichel distribution, then $\lambda(x)$ follows an inverse Gaussian distribution (see Sichel, 1973, for the Sichel distribution; Jørgensen, 1982, for the inverse Gaussian distribution; and also Matheron, 1981d).⁶ These properties allow Kleingeld et al. (1996, 1997) to carry out the statistical inference of the parameters of a Cox process from sample data and to simulate it.

Shooting Process or Cluster Process

The shooting process or cluster process was defined by Neyman and Scott (1958) for representing clusters of galaxies (they also considered clusters of clusters and their evolution with time, which we will not do). The properties of this model were studied by Neyman and Scott (1972). Each point X_i of a primary Poisson process with intensity λ_0 is considered as a target at which ν shots are fired. The number of shots is a random variable, and the impact of each shot is a secondary point, randomly and independently located around the target. The shooting process is formed of all the secondary points (Fig. 7.37c). Another terminology, often used when modeling natural phenomena, is *parent-daughter model*: the primary points represent the location of parents (e.g., plants) and the secondary points are their daughters (e.g., seed fall locations). Note that this process can also be seen as a particular case of the Boolean model where each random object associated with a Poisson point is in fact a cluster of objects.

It is of course possible to generalize this type of process, for example, by considering cluster processes with random intensity (Deverly, 1984b). But this only makes sense if we specify the scale of the phenomena taken into account by the two aspects (clusters and regionalization), for example, by giving the clusters a local extension and considering that the intensity changes slowly in space. This problem is already there when we choose to use the Cox process or the cluster process, since we can consider the presence of clusters as zones of high intensity, and vice versa. It can be shown that a cluster process where the number of points per cluster has a Poisson distribution is a Cox process.

Hard-Core Models

Hard-core models represent some kind of inhibition between the points. There are many models. A simple one is introduced by Matérn (1960, p. 47). Consider a realization of a Poisson point process with intensity λ_0 as a primary process. As a realization of the hard-core process, keep every point of the primary process whose nearest neighbor (in the primary process) is at a distance at least equal to a given minimum distance R (Fig. 7.37d).

Poisson Alignments

Serra (1982) proposes this model to represent aligned points. A realization of this process in 2D is composed of all the intersections of the lines of a realization of Poisson lines (Fig. 7.37e). This model can be extended to \mathbb{R}^n by considering the intersections of hyperplanes of dimension $n - 1$.

A Structural Tool: The K -Function

For all the above point processes, we know how to calculate the mean and the variance of $N(V)$ and also the covariance of $N(V)$ and $N(V_h)$. This covariance can be used as a structural tool to differentiate these models and fit their parameters, but it is not very powerful. Fortunately there are other tools, which derive from the fact that the spatial distribution of a point process can also be determined by its *Palm distribution*, where the point process is observed from a moving point of the process (for an exact definition of the Palm distribution, see Stoyan et al., 1987, sec. 4.4, or Cressie, 1991, sec. 8.3.4). An example of tool is the distribution of the distance from a point of the process to its nearest or k th nearest neighbor, particularly useful when the point process is isotropic. But the main tool in this case is the K -function.

The point processes presented above are isotropic provided that the random intensity function (Cox process) or the clusters (shooting process) are isotropic. In this case Krickeberg (1974) and Ripley (1976) propose to replace the covariance by the K -function, which has a more geometric interpretation and permits an easier discrimination of the various models. The knowledge of this function, which is also called the “second reduced moment function,” is equivalent to that of the covariance. First, consider the points of the point process that belong to the sphere of radius r centered at an arbitrary point x (this point almost surely does not coincide with a point of the process); the number of these points has for mathematical expectation $\lambda K_0(r)$ with $K_0(r) = V_n r^n$, where V_n is the volume of the unit-radius sphere of \mathbb{R}^n . Consider now a typical point x_i of the point process, and let $\lambda K(r)$ be the expected number of points of the process within the sphere of radius r centered at the point x_i , which is itself not counted. $\lambda K(r)$ usually differs from $\lambda K_0(r)$, a noticeable exception being the Poisson point process. Hence $\lambda[K(r) - K_0(r)]$ is the average excess of points belonging to the sphere of radius r centered at a point of the process (this point excluded) over and above the average number.

Figure 7.38 shows the behavior of $K_1(r) = K(r) - K_0(r)$ for the various models presented above. They are very different, excepted again for the Cox process and the cluster process which cannot be easily differentiated: for these models, K_1 is positive, whereas it is equal to zero for a Poisson point process and is negative for a hard-core process. Since the function $K_1(r)$ can be estimated directly (see Ripley, 1976; Cressie, 1991) and its theoretical expression for the various models is known, it is a good structural tool. Note that like the covariance, the K —or K_1 —function summarizes the variances and covariances of point counts, namely second-order moments, and thus it does not represent all the information about a stationary isotropic point process: two different processes can have the same K -function. Baddeley and Silverman (1984) show a 2D example of a point process that is very different from a Poisson point process and yet has the same K -function.

When studying a random set from discrete sampling, the boundaries of the individual objects are usually unknown, which makes the inference of the underlying point process difficult.

7.8.3. Conditional Simulation of a Boolean Model

Conditioning a Boolean Random Set: The Example of Fracture Networks

The simulation of a Boolean random set is straightforward once its parameters are known: intensity of the Poisson point process and probability distribution of the primary objects (type, size, orientation). But conditioning a simulation by observations is less straightforward. It can be done easily in simple cases, as we now show for fracture networks.

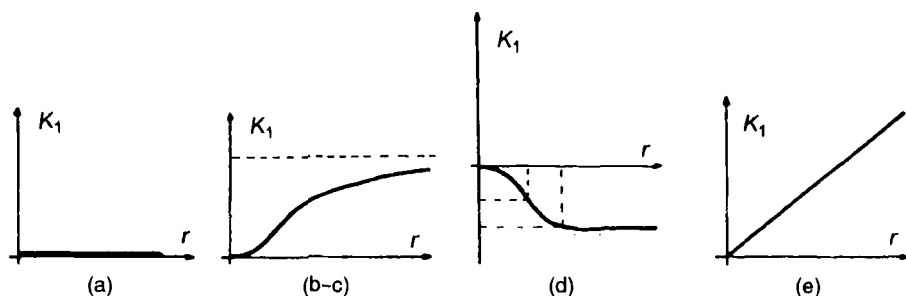


FIGURE 7.38. Shape of the function $K_1(r)$ for different models of point processes: (a) Poisson points; (b–c) Cox process and cluster process; (d) hard-core process; (e) Poisson alignments. From Serra (1982), with permission of Academic Press Limited, London.

When modeling 3D fracture networks, fractures are usually considered as planar objects. Fracture networks are often represented by Poisson models: Poisson planes for fractures that are infinite at the scale of the study, random disks for finite fractures (i.e., a 3D Boolean model of 2D disks; Baecher et al., 1977). The latter model has been generalized to a parent-daughter model with a regionalized intensity (i.e., a 3D Boolean model of disk clusters built on a Cox process; Chilès, 1989a, 1989b). The fractures are usually divided into several sets corresponding to different average fracture orientations. Each set is characterized by a distribution of the fracture orientation (the unit vector normal to the fracture) and of the fracture size (disk diameter). The data are fracture intersections with boreholes or fracture traces on drift walls or outcrops. The orientation of each fracture is measured, but its extension is not always known. These data are affected by several types of bias (censoring, truncation, size). Specific methods have been developed to handle this kind of data. 3D fracture network simulations are often used for fluid flow studies. Since actual fractures are not pairs of parallel plates, flow is often seen to be channelized. 2D Boolean models of random segments are therefore simulated within each fracture, and flow is studied in the resulting 3D channel network, as shown in Figure 7.39 (Billiaux, 1989). The reader is referred to Chilès and de Marsily (1993) for an overview of these methods with references to other authors, and to Long and Billiaux (1987), Massoud (1987), Loiseau (1987), Hestir et al. (1989), and Gervais (1993) for some modeling applications. More general marked point processes are also used (e.g., Wen and Sinding-Larsen, 1997), as well as pixel-based methods that do not require to decompose the network into individual fractures (see the next section). We will simply illustrate the conditional simulation of Poisson fracture networks.

Conditioning a model of planes or disks based on Poisson processes on observed fractures is easy because all points, lines, or planes of a Poisson process are mutually independent. The procedure consists in independently simulating the fractures that intersect the surveyed lines or surfaces (step 3) and those that do not (steps 1 and 2):

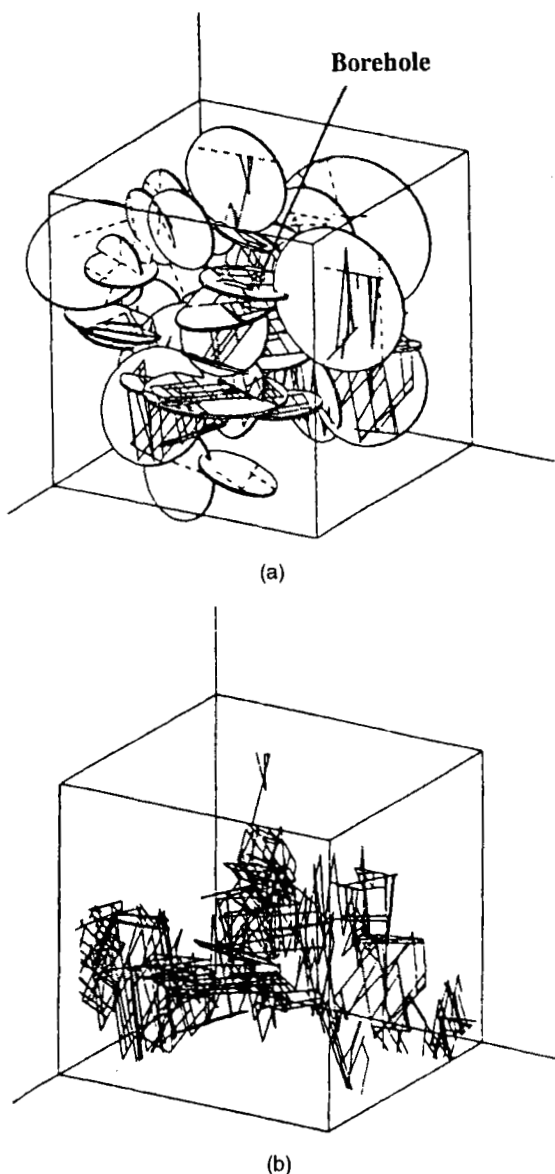


FIGURE 7.39. 3D simulation of a Boolean set of random disks and channels: (a) network of random disks (for reasons of legibility, not all the disks encountering the simulated domain are shown), and random segments simulated on each disk; (b) network of random channels (only the channels connected to a small borehole located at the center of the block are shown). Reprinted from Billaux (1989), in V. Maury and D. Fourmaintraux, eds., *Rock at great depth—Rock mechanics and rock physics at great depth/Mécanique des roches et physique des roches en condition de grande profondeur/Felsmechanik und Felsphysik in großer Tiefe/Proceedings of an international symposium, Pau, 28-31.08.89. 1989, 1620 pp., 3 vols.*

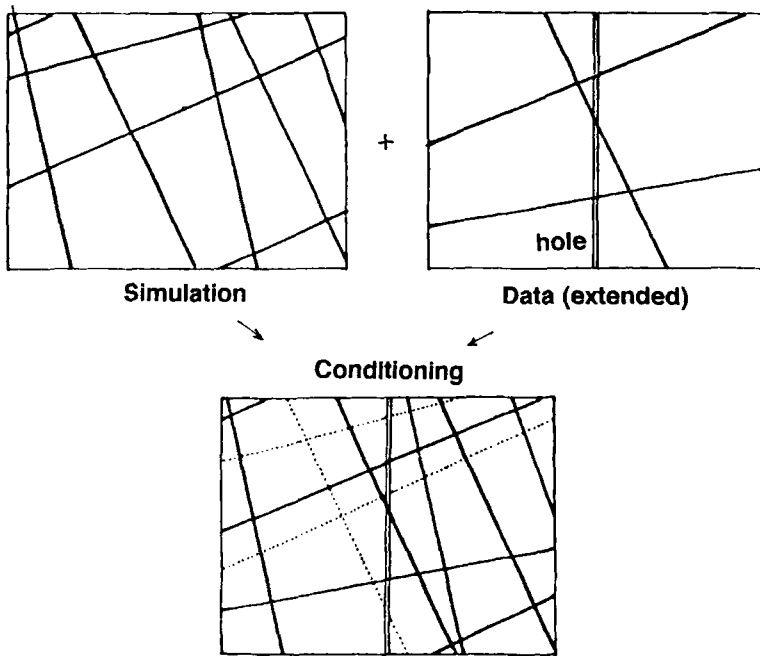


FIGURE 7.40. Conditioning of a 3D simulation of Poisson lines by drill-hole data. From Chilès (1989a, b), with kind permission from Battelle Press and Kluwer Academic Publishers.

1. Construct a nonconditional simulation.
2. Reject all the fractures that intersect the surveyed lines or surfaces, and retain all those that do not intersect.
3. Add the actual fractures that have been observed.

The simplest case is the Poisson plane model because fractures are infinite. The conditioning procedure is represented in Figure 7.40, in 2D for simplicity, the data being the intersections with a borehole (location of the intersections and fracture orientation). This procedure has been suggested by Andersson et al. (1984). In the case of the random disk model, step 3 introduces a slight complication because the extension of a fracture must be determined from an observed intersection or trace. In other words, we have to simulate the whole fracture from the joint distribution of disk diameter and disk center location conditional on the observation (point or trace). This conditional distribution results from geometric probability calculations. Rejection techniques make it possible to condition a fracture on additional observations, for example, the intersection by another borehole or a trace on another drift wall, but the user must be able to “correlate” the two intersections of the fracture. An example, based on actual data surveyed on drift walls, is presented by Chilès et al. (1992).

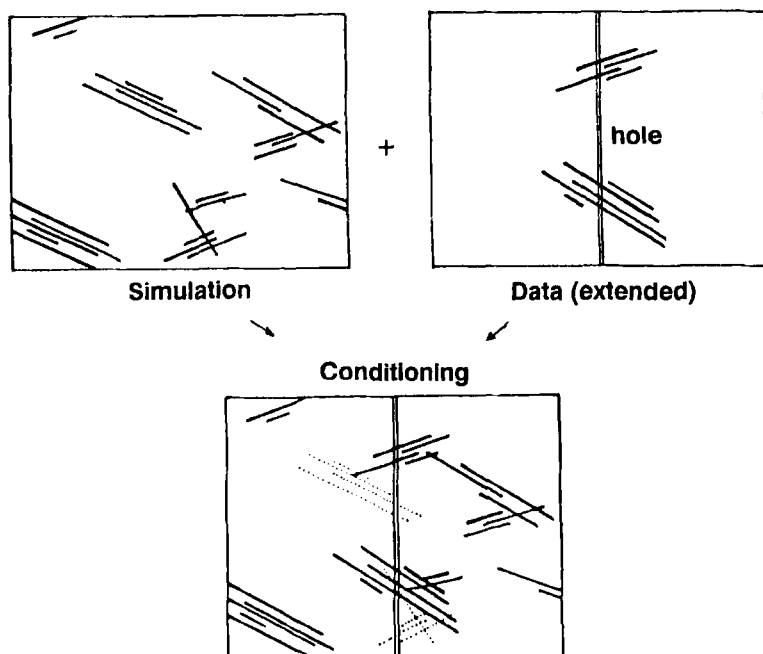


FIGURE 7.41. Conditioning of a 3D simulation of random disks clusters by drill-hole data. From Chilès (1989a, b), with kind permission from Battelle Press and Kluwer Academic Publishers.

The same approach can be used with a Boolean model of disk clusters, but since the random objects associated with the Poisson points are the clusters, the rejection defined in step 2 and the completion of the observed intersections or traces in step 3 must be applied at this level (see Fig. 7.41). This is also feasible (Chilès, 1989a, 1989b). The main constraint is practical: the user must be able to define from the data which fractures belong to the same cluster.

Conditional Simulation of a Boolean Random Set in the General Case

Conditioning a simulation of fracture network in \mathbb{R}^n is possible for two reasons: (1) the primary objects are elements of a subspace of \mathbb{R}^n so that any point belonging to the fracture network belongs to a single primary object (we neglect fracture intersections); (2) we can determine that two observations are from the same object (fracture or cluster, according to the model). In the general case, one point of the random set may belong to several objects. Moreover several data points may belong to a common object, but we have no simple means of knowing this. Therefore the method used for fracture networks cannot be transposed to the general case (see, however, Haldorsen, 1983, and Chessa, 1995, for a similar approach for simulating the geometry of a petroleum reservoir conditional on the observations in wells).

An iterative Markov procedure, suggested by G. Matheron in 1990 and presented by Lantuéjoul (1997a, b), gives a general solution provided that the objects have a strictly positive measure and the number of conditioning points, known to belong to the background or the objects, is finite. This algorithm is valid for Boolean sets based on Poisson point processes with a regionalized intensity $\lambda(x)$ that is integrable: $\int \lambda(x)dx = \lambda_1 < \infty$. The standard Boolean random set ($\lambda(x)$ constant) does not fulfill this condition, but as we will simulate it in a finite domain V , we can limit the support of $\lambda(\cdot)$ to $V \oplus S$ where S is a sphere large enough to include any elementary object.

For a simple presentation of the algorithm, it is convenient to define a λ -object as a primary object with a random location according to the density $\lambda(x)/\lambda_1$. Clearly a Boolean random set is the union of a Poisson number (with mean value λ_1) of independent λ -objects. The principle of the algorithm is the following: at time $t = 0$, start with an initial simulation, composed of primary objects, that is simply compatible with the data (but does not claim to represent a Boolean random set). Then, denoting by $\varphi(t)$ the pattern of objects at time t and by $|\varphi(t)|$ the number of objects of this pattern, in each elementary time interval $[t, t + dt]$, add a random object with probability $\lambda_1 dt$, remove an object with probability $|\varphi(t)|dt$, and do nothing with probability $1 - (\lambda_1 + |\varphi(t)|)dt$. The objects are removed and added randomly but under the constraint that the simulation always remains consistent with the data. The implementation of the algorithm is the following (Lantuéjoul, 1997a, b):

1. *Initialization.* Set $t = 0$, and simulate a pattern φ of λ -objects that respects the conditions on the objects and on the background.
2. *Moving forward in time.* Simulate an exponential holding time T with parameter $\lambda_1 + |\varphi|$, and add T to t (where $|\varphi|$ is the number of points of the pattern φ).
3. *Transition to the next state.* Simulate a random variable U equal to $+1$ with probability $\lambda_1/(\lambda_1 + |\varphi|)$, or -1 with the complementary probability $|\varphi|/(\lambda_1 + |\varphi|)$, and
 - if $U = +1$, simulate a λ -object and insert it in φ provided this does not violate the conditioning on the background;
 - if $U = -1$, randomly select one of the $|\varphi|$ present λ -objects and remove it provided this does not violate the conditioning on the objects.
4. Go to (2).

Note that this algorithm tends to equalize $|\varphi|$ with λ_1 : when $|\varphi| < \lambda_1$, objects are added more often than removed, and the other way around when $|\varphi| > \lambda_1$. Time plays no role here but will be useful for generalizations.

The following procedure can be used to initialize the process:

1. Set $\varphi = \emptyset$.

2. Simulate a λ -object A .
3. If A does not hit the conditioning points that belong to the background and hits at least one of the conditioning points belonging to the objects not yet covered by φ , insert A in φ ; else go to (2).
4. If φ does not cover all the conditioning points belonging to the objects, go to (2).

The initialization step requires a finite though possibly large number of operations, unless the data are not compatible with the model (the number of iterations needed gives an indication on the adequacy of the model to the data). The main algorithm never ends but, in practice, must be stopped after a finite time, which raises problems of convergence. First note that the general algorithm can be used even in the nonconditional case (no conditioning point), starting from an empty set. Then $|\varphi|$ evolves according to a birth-and-death process that is known to converge to the desired Poisson variable (Feller, 1968, sec. XVII.7). The proof of the convergence of the conditional algorithm is established in Lantuéjoul (1997b). In both cases the convergence rate is an exponentially decreasing function of time. Figure 7.42 shows four conditional simulations of a Boolean set of channels.⁷

Conditional Simulation of other Boolean Models

The Markov procedure of the above algorithm is a very simple one: we pass from one state to the next by adding or removing one object, and the objects are independent. This lends itself to several generalizations while remaining in the convenient setting of Markov models (Lantuéjoul, 1997b).

A first generalization is to model not only the number of objects $|\varphi|$ as a birth-and-death process with a Poisson distribution as limiting distribution, but also the pattern of the object centers as a *spatial birth-and-death process of points* (Preston, 1975; see also Stoyan et al., 1987, sec. 5.5.5; Møller, 1989; Guyon, 1993, section 6.5). This is a continuous-time Markov process where the states belong to the family of point patterns: (1) at any time t it is a point process in space, (2) it changes at isolated instants of time either by the birth of a new point, which is added to the pattern, or by the death of a current point, which is deleted, and (3) the probability of a change depends only on the current configuration of the point process. The distribution of a spatial birth-and-death process is characterized by two functions: a birth rate b and a death rate d . The meaning of these functions is as follows: let φ be the current state of the point pattern at time t ; if $x \notin \varphi$, $b(x, \varphi) dx dt$ is the probability that a birth occurs in the infinitesimal domain dx between t and $t + dt$; similarly, if $x \in \varphi$, $d(x, \varphi) dt$ is the probability that the point x dies during this time interval. The Boolean random set corresponds to the choice $b(x, \varphi) = \lambda(x)$, $d(x, \varphi) = 1$ (their sums over x being λ_1 and $|\varphi|$, respectively). Interactions between points (e.g., repulsion) can be obtained with other choices. The functions b and d should of course be chosen according to the limiting distribution that is required for the point pattern.

A second generalization consists in using a *spatial birth-and-death process of objects*. This is a continuous-time Markov process where the states are patterns of objects rather than points. The definition of a spatial birth-and-death process of objects is similar to that of a spatial birth-and-death process of points, but now φ is a pattern of objects, and birth rate and death rate are functions $b(A, \varphi)$ and $d(A, \varphi)$ where A is an object. This kind of process makes it possible to introduce interactions between the objects (e.g., the objects cannot overlap).

A third generalization consists in using a *Markov jump process*, where the transition from one pattern to the next may involve several objects rather than a single one.

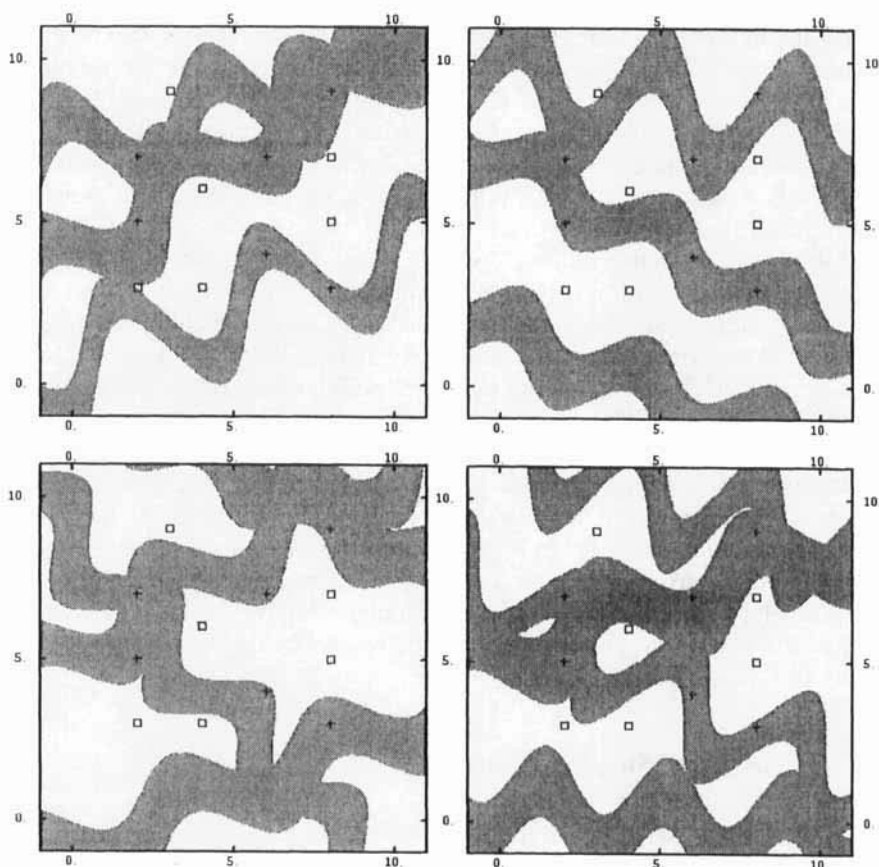


FIGURE 7.42. Conditional simulations of Boolean channels modeled as sinuous strips, constrained by six points in the objects and six points in the background (generated by SIMBOOL software from Centre de Géostatistique, Ecole des Mines).

A fourth generalization consists in combining several random set simulations using some combination rule, for example, a hierarchical rule (the random sets are ordered, and the objects of a random set override the objects of the random sets of lower order) or a dead-leaves rule (each object or leave is associated with its falling date). This generalization is very convenient to simulate multifacies models, for example, Boolean random functions of cylinders and dead-leaves models, with various dependencies between the facies. Examples are given by Lantuéjoul (1997b).

7.9. CONSTRAINED SIMULATIONS

Reproducing the covariance of a phenomenon is not sufficient to provide realistic simulations. In fact it is necessary to reproduce all finite-dimensional distributions of the parent RF. But this is not identifiable empirically, and even

if it were, we would not know how to construct a simulation of it, unless of course it had the good taste to coincide with one of the models we know how to simulate. Nevertheless, we would like to be able to constrain the simulation by a little more than its covariance, its histogram, and a few sample points. Another important challenge of geostatistics, and especially of conditional simulations, is to integrate various types of data, such as in the petroleum industry, borehole data, seismic, pumping tests, production history, as well as geological knowledge.

If the geostatistician's toolbox does not contain a simple multivariate model whose spatial distribution meets the requirements, an adaptive method might provide a solution: starting from a standard simulation, matching only the histogram and the covariance, for example, we adapt it iteratively until it matches the imposed constraints. One increasingly popular method to achieve this is simulated annealing. This is an optimization method rather than a simulation method, and its use for the construction of geostatistical simulations is based on arguments that are more heuristic than theoretical. Another approach, often combined with simulated annealing, uses a Bayesian framework to incorporate general geological knowledge when the probabilistic model cannot be completely inferred from the commonly available data. Before describing these methods, let us give two examples of simpler adaptive methods. These generalize the sequential simulation method, which, being iterative, lends itself easily to the addition of an adaptive step.

7.9.1. Examples of Simple Adaptive Methods

Reproducing Empirical Conditional Distributions

Guardiano and Srivastava (1993) present an interesting example involving finite-dimensional distributions beyond the plain covariance (Fig. 7.43). The example is about a binary image and the method assumes that we have a training image. In this case this is a sandstone formed by alternating fine, relatively impervious sediments (in black in Figure 7.43a) and coarser, more permeable sediments (in white). The simulation is constructed in a sequential manner as with the sequential indicator simulation method. The value of the simulation at the current node is drawn from a Bernoulli distribution, namely 0 with probability $1 - p$, and 1 with probability p (this depends on the node considered). A standard indicator simulation uses for p the kriging estimate I^* of the indicator calculated from points known or already simulated in the neighborhood; it reproduces the sample covariance but not the morphology of the deposit (Fig. 7.43b). For this reason Guardiano and Srivastava take for the Bernoulli distribution at the current node the conditional distribution given the points in the neighborhood (16 at most), or at least the distribution established empirically from the training image for identical configurations of the neighborhood (point geometry and values of the indicators). This simulation reproduces the covariance and the morphology of the training image much

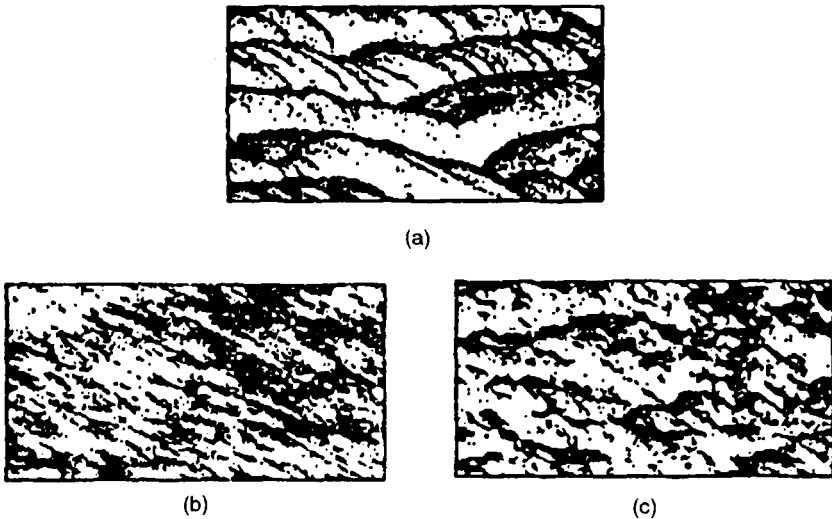


FIGURE 7.43. Sequential indicator simulation of a sandstone formed by two types of sediments: (a) training image; (b) standard sequential indicator simulation, correctly reproducing the sample covariance; (c) sequential indicator simulation making use of the sample multivariate distributions of the training image. From Guardiano and Srivastava (1993), with kind permission from Kluwer Academic Publishers.

better (Fig. 7.43c). Another similar application is the simulation of a fracture network (in *ibid.*).

Stochastic Seismic Inversion

Bortoli et al. (1993) and Haas (1993) construct a 3D simulation of acoustic impedance with the requirement that it be consistent with borehole data and with 3D seismic. The simulation is constructed on the same grid as the 3D seismic. The objective is that to the vertical of each node (x, y) of the 2D space, the synthetic seismic trace derived from the vertical profile of simulated impedances be as close as possible to the real seismic trace. To achieve this, the grid is constructed trace by trace using a sequential algorithm:

1. Select a node (x, y) at random among those where the impedance profile is unknown.
2. Construct n_s simulations of acoustic impedance along the profile to the vertical of the node (x, y) conditionally on the traces already known.
3. Transform the n_s impedance profiles into synthetic seismic traces.
4. Compare each of the n_s simulated seismic traces with the measured trace (the criterion used is the correlation coefficient or the variance of discrepancies); the impedance profile that provides the simulated trace closest to the measured trace is selected.

5. Add the selected profile to the data, and if all the profiles have not yet been simulated, return to step 1 for another profile.

The acoustic impedance is finally transformed into a porosity image by means of an acoustics-petrophysics calibration established from borehole data. In all, this method of simulation performs an inversion of the seismic data. For quality results the various steps must be carefully calibrated and validated. In particular, one must care about the choice of the required degree of similarity between the simulated synthetic trace and the real trace, given the uncertainties on the latter. The degree of similarity obtained is a function of n_s . The requirement of too high a value for n_s would lead to overconstraining the simulation. In the present case a value between 10 and 100 seems sufficient, as we can see in Figure 7.44 showing a simulated seismic section for three different values of n_s (1, 10, 100), and also the real seismic section. The real seismic is poorly reproduced by the simulated seismic when $n_s = 1$, in other words, for an ordinary simulation (the correlation coefficient is less than 0.4), whereas it is well reproduced, except for a few local anomalies, with $n_s = 100$ (the correlation coefficient reaches 0.8). Finally we note that the inclusion of the constraints tends to modify the variogram of the simulations a little, by extending its range. Since this is a case where the ranges are not known with precision, this problem is considered to be a minor one, in view of the advantage gained by taking into account the 3D seismic data.

7.9.2. Simulated Annealing

Simulated annealing is a powerful, but heuristic, means of building conditional simulations subject to complex constraints (Hegstad et al., 1994). Simulated annealing was introduced by Kirkpatrick et al. (1983) as a device to obtain improved solutions of a variety of combinatorial optimization problems. Annealing is a manufacturing process by which a molten metal is cooled very slowly to produce a stress-free solid. At high temperature the molecules of the molten metal move relatively freely and reorder themselves into a very low energy crystal structure. Nature is able to find this minimum energy state for slowly cooled systems. If the liquid metal is cooled quickly, it does not reach this state but ends up in a polycrystalline or amorphous state with a higher energy. So the essence of the process is slow cooling, allowing ample time for redistribution of the atoms as they lose mobility. In statistical physics the p.d.f. of a state s with potential energy $U(s)$ at absolute temperature $T > 0$ is of the form

$$f_T(s) \propto \exp\left(-\frac{U(s)}{kT}\right) \quad (7.52)$$

where Boltzmann's constant k relates temperature to energy. Such a distribution is known as a Gibbs, or Boltzmann, distribution. A system in thermal equilibrium at temperature T has its energy probabilistically distributed among

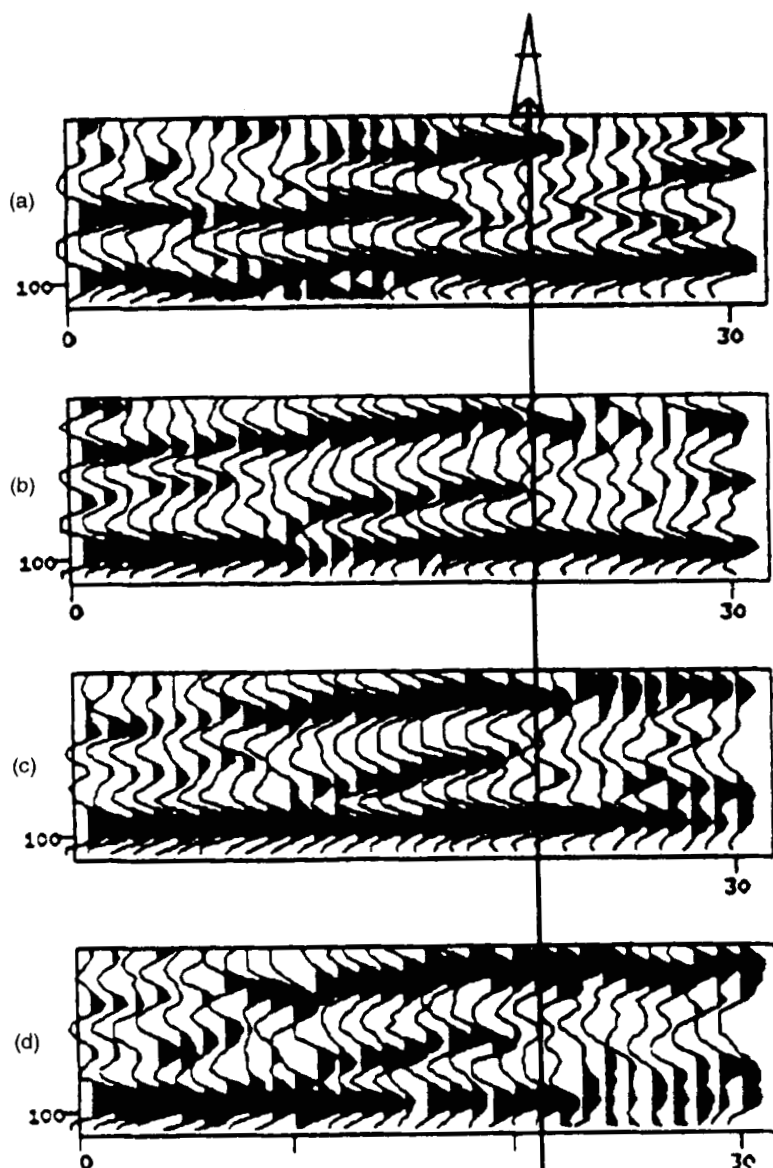


FIGURE 7.44. Geostatistical inversion of seismic data and comparison with reality on a seismic section: (a) simulation with $n_s = 1$; (b) simulation with $n_s = 10$; (c) simulation with $n_s = 100$; (d) real section. From Haas (1993).

all different states s . Even at a low temperature, there is a positive probability for the system to be in a high energy state. It is this probability that gives the system a chance to get out of a local energy minimum and to find a more global one later. However, the probability for the system to be in a high energy state

decreases when T tends to 0, since the p.d.f. f_T concentrates progressively on the set S_{\min} of the states s that have the minimum energy U_{\min} . The annealing process achieves a configuration of atoms with energy near or at the minimum of U when T returns to a low value.

Simulated annealing mimics the metallurgical process to optimize a grid of values. Starting with a first image of the grid, the image is iteratively relaxed by swapping two pixels. The selection of the successive pairs is random and combined with a rejection technique: as in the metallurgical process, the swap of two pixels will more often than not reduce the energy. In order to develop the analogy with annealing, the energy function $U(s)$ associated with any configuration or state s of the image is defined as some measure of the difference between the desired spatial features and those of the candidate realization s . The analogy with (7.52) suggests defining a p.d.f. on all images by

$$f_\lambda(s) \propto e^{-\lambda U(s)}$$

where λ plays the role of $1/T$ and so increases very slowly to infinity.

When λ remains fixed, an iterative use of this algorithm samples the states s from f_λ . This Glauber-Metropolis algorithm amounts to using the spin exchange method of Section 7.7.2 with $\pi(s_i) \propto f_\lambda(s_i)$. From a practical point of view, for a given pair of pixels and a fixed value of λ , the swap from state s_i to state s_j is made if $U(s_j) < U(s_i)$, and is randomized with probability $f_\lambda(s_j)/f_\lambda(s_i) = \exp(-\lambda[U(s_j) - U(s_i)])$ in the opposite case. As we have seen in Section 7.7.2, this enables the procedure to sample all possible states from their probability $f_\lambda(s)$ and not remain attracted by a local minimum.

By “cooling” progressively, namely by increasing λ sufficiently slowly, one approaches a uniform sampling of the set S_{\min} . In practice, λ is changed at each step. Convergence to a uniform sampling of S_{\min} is ensured if $\lambda_k = (1/c) \log(k+1)$ is taken for the value of λ at step k . In this formula c represents the minimum increase in energy required to get out of any local minimum of $U(s)$ that is not a global minimum and enter the neighborhood of a state with a lower energy (Hajek, 1988). Such convergence to infinity for λ_k is very slow. In practice, $\lambda_k = \lambda_0/\alpha^k$ is commonly used (e.g., $\alpha = 0.90$ or 0.95), but convergence is no longer assured.

It is easy to imagine variants of this algorithm. Geman and Geman (1984) use the theory of simulated annealing combined with the Gibbs sampler rather than the spin-exchange method and apply it to the restoration of images that have been degraded by an additive or multiplicative noise, a nonlinear transformation, and/or blurring. Most geostatistical applications use the Metropolis algorithm.

Millions of iterations are often required to approach a low energy state. Therefore simulated annealing will be efficient only if $U(s_j)$ can be easily obtained by updating $U(s_i)$. The spatial features commonly used in geostatistical applications, such as a regional histogram, a variogram or a bivariate distribution, can be updated locally, thus avoiding a global computation after each

iteration. A typical example of energy function is

$$U(s) = w_1 \sum_{p=1}^{n_1} [\gamma^*(h_p; s) - \gamma(h_p)]^2 + w_2 \sum_{q=1}^{n_2} [\mu_q^*(s) - \mu_q]^2$$

The first part of this expression measures the discrepancy between the variogram γ^* of the simulation and the desired variogram model γ for n_1 values h_p of the lag. The second part expresses constraints by measuring the discrepancy between some geological, statistical, or engineering properties μ_q^* measured on the simulation and their desired values μ_q observed on actual data. The weights w_1 and w_2 control the trade-off between the two parts of $U(s)$. When studying an aquifer or a petroleum reservoir, the simulated variable can be porosity, and the constraints can be, for example, head values or pumping tests results (e.g., Deutsch, 1993). Another example is two-point transition probabilities defined on a training image (Farmer, 1992). This example can be generalized to multiple-point statistics or to change-of-support statistics (e.g., applications to fracture networks by Srivastava, 1994). Additional information on the practice of simulated annealing in geostatistical applications can be found in Deutsch and Cockerham (1994) and Ouenes et al. (1994). The latter authors also review the use of *genetic algorithms* and *neural networks*.

There are relatively many, though recent, geostatistical applications. Langlais and Doyle (1993) compare the efficiency of simulated annealing (SA) using a training image, sequential indicator simulation (SIS), and truncated Gaussian simulation (TGS). SA does not incorporate the covariance structure directly, but uses an energy function with two terms:

1. One to minimize the discrepancy between the observed and simulated proportions of any facies at the various levels.
2. Another to minimize the discrepancy between the rates of transition from facies i to facies j for two points at a distance h apart, for various values of h .

This study highlights the following points:

- SA reproduces the training image best but exhibits undesirable randomness in the direction orthogonal to the image (a vertical outcrop); when a training image resembling the outcrop is added in this orthogonal direction, all SA simulations look alike: the conditions imposed on the simulations (reproduce the statistics observed on the outcrops) result in a lack of fluctuations in the statistics of the simulations. This could be a drawback, since it precludes the generation of a wide spectrum of flow simulations.
- SIS and TGS produce better results, with a better stability in the reproduction of the global and vertical proportions for TGS than for SIS.

- Taking the time required by the TGS algorithm as one time unit, SIS requires 36 time units, and SA 400 time units.

Simulated annealing seems to be an interesting technique for building simulations that honor a variety of data as well as complex constraints. But it is computationally demanding. More important, when several simulations are generated, it is hard to know if a correct level of fluctuation is reached and if all possible situations have been correctly sampled. If we imagine that the procedure can be applied until the ultimate iterations are reached, the result will be a state selected uniformly in the set S_{\min} of states with minimum energy. In general, we do not have the faintest idea on the geometry of this set and therefore on the spatial distribution of the simulations. This spatial distribution depends on the energy function we have chosen, and we do not know if this choice makes sense: it may be too loose as well as too severe (e.g., S_{\min} may contain a single state so that the simulations are all the same). Moreover in the practical procedure the “cooling” is much faster than the ideal one, so convergence is not assured, which usually leads to a much broader set of possible states.

7.9.3. Bayesian Approach

We sometimes have so few data that we cannot infer the covariance or variogram with any confidence, which is already a problem if we intend to perform kriging, and even more so if we want to construct a simulation. This situation is common in the study of petroleum reservoirs. Bayesian methods provide a framework for dealing with such problem provided that some general knowledge can be input by the geologist. It is assumed that (1) from geological considerations the geologist can specify the type of reservoir, namely in geostatistical terms, the type of RF model describing the reservoir properties of interest (e.g., lithology, porosity, permeability, oil saturation), (2) this RF model is of a type that we know how to simulate, and (3) an expert, reasoning by analogy or otherwise, can specify a prior distribution for the RF parameters.

The reservoir is usually represented by a vector-valued random variable of high dimensionality which will be denoted by R . The set of parameters of the distribution of R is a vector-valued random variable of low dimensionality which will be denoted by Θ . Our model is then a set of RFs indexed by the possible values θ of Θ .

Let us give some examples:

1. R is a Boolean model of channels, namely a set of random points with their marks (length, width, direction, curvature of the channel). Θ comprises the Poisson intensity (fixed or regionalized) and the parameters of the distribution of the channel marks.
2. R comprises an indicator or categorical RF (facies number) and a Gaussian RF (logarithm of permeability), discretized at the nodes of a grid.

Θ comprises the parameters of the categorical variable (proportion of each facies, range) and those of the Gaussian RF (logarithmic mean and variance of permeability in each facies, nugget effect, ranges).

3. *Model with constraints.* A binary function $I(r)$ specifies whether or not the reservoir $R = r$ is possible. The situation is similar to the previous ones except that the p.d.f. of R is multiplied by $I(r)$, up to a normalizing factor.

Denoting by $f(\theta)$ the prior p.d.f. of Θ and by $f(r | \theta)$ the prior p.d.f. of R when $\Theta = \theta$ is fixed, the prior p.d.f. of R is

$$f(r) = \int f(r | \theta) f(\theta) d\theta$$

A nonconditional simulation of R is simply a vector r selected from this p.d.f. In practice it is usually difficult to express $f(r)$ analytically and to simulate R directly, and the simulation is generated in two steps:

1. Select θ from the p.d.f. $f(\theta)$, which is a standard problem of sampling from a vector-valued random variable of low dimensionality.
2. Select r from the conditonal p.d.f. $f(r | \theta)$; that is, generate a nonconditional simulation with known parameters θ , which is a standard problem.

This amounts to sampling both R and Θ . In applications we have a set of observations, represented by a vector-valued random variable referred to as *data*, and we are interested in simulations conditional on these data. For the same reason as above, we will sample R from the posterior p.d.f. $f(r | \text{data})$ by sampling both R and Θ from $f(r, \theta | \text{data})$. From the Bayes's rule this posterior p.d.f. is

$$f(r, \theta | \text{data}) = \frac{f(\text{data} | r, \theta) f(r, \theta)}{f(\text{data})} \quad (7.53)$$

Given the reservoir properties $R = r$, the distribution of the data does not depend on θ so that $f(\text{data} | r, \theta)$ is a function of r only. (7.53) is thus of the form

$$f(r, \theta | \text{data}) = \text{const} \times f(\text{data} | r) f(r | \theta) f(\theta) \quad (7.54)$$

where *const* denotes $1/f(\text{data})$ which cannot usually be computed (it would require the double integration of the right-hand side in θ and r). This method requires, however, that we know $f(\text{data} | r)$ for any r . Since the observations usually comprise direct data of the reservoir properties in wells and indirect observations such as seismic or production history, this amounts to postulating that we are able to solve the forward problem, that is, compute the indirect

observations knowing the reservoir properties everywhere. The results depend on unknown data acquisition parameters (e.g., the depth of investigation of a logging tool, or the exact signature of a seismic source), so that they are represented by a probability density function, namely $f(\text{data} | r)$. Actually the determination of $f(\text{data} | r)$ is not a simple matter, especially when scale effects are present, as we will see in the next chapter.

Now the problem is to sample from (7.54). A standard technique for drawing from a probability distribution known up to a normalizing constant is the acceptance-rejection method. If we know a majorization $A(\text{data})$ of $f(\text{data} | r)$ over all possible values of r , we can construct a conditional simulation of R by adding the following third step to steps 1 and 2:

3. Let $p = f(\text{data} | r)/A(\text{data})$; accept r with the probability p , or reject it with the complementary probability $1 - p$ and return to step 1.

But this method is very inefficient. When p is low, which it is when $f(\text{data} | r)$ is not flat, a large number of simulations obtained with great effort end up being rejected. For example, with standard conditional simulations where observations are exact values of the reservoir properties, data is a subset of R . There is a very low, if not zero, probability that a nonconditional simulation (r, θ) fits the observations, so that step 3 will almost surely lead to rejection. The conclusion is the same if only some of the observations are exact values of reservoir properties. Other algorithms must be used.

In simple cases the simulation can be carried out with standard techniques. For example, Doyen et al. (1994), assuming θ known, define a sequential algorithm that combines a local prior distribution of lithology (binary basis variable sand/shale) obtained by indicator kriging with a function representing the p.d.f. of a local seismic attribute (indirect variable) given the lithological facies. Such simplification is not possible in more complex situations, and another method is used, the Metropolis-Hastings algorithm. This requires to cast the problem in terms similar to those of Section 7.7.2. For example, the reservoir (a state in the terminology of Markov fields) is represented by a multivariate grid of values, or by an array of coordinates, sizes and orientations of Boolean objects. Constraints like those in Example 3 are accounted for by simulated annealing techniques. It may be a good idea to initialize the iterative process with a conventional simulation conditioned on the observations of reservoir properties. The method must be tailored to each application and is clearly computationally intensive despite numerous approximations. Details can be found in Omre and Tjelmeland (1997).

This approach has been initiated by Haldorsen and Lake (1984) in a simplified version, and developed further by the "Norwegian School," for example, Georgsen and Omre (1993) (combination of marked point processes and Gaussian RFs, including spatial interaction), Eide et al. (1997) (Gaussian random functions, uncertainty in the data acquisition parameters), Lia et al. (1997) (a variety of data types and RF models), to cite only a few. An achievement of

this method is to produce, in principle, images that are consistent with the observations and constitute a reliable basis for quantitative evaluations. However, we basically get what we put in. If there are plenty of data the results are reliable regardless of the prior distribution; if there are little data—which is the working assumption here—the results mainly reflect the prior model. This may be dangerous if the prior model is not consistent with the available data.

7.10. PRACTICAL CONSIDERATIONS

Before concluding with examples of applications, we will make a few practical remarks.

7.10.1. Simulation of Anisotropies and Space-Dependent Variograms

Until now we have considered only simulations with a stationary and, for some methods such as turning bands, isotropic covariance. Let us now examine some generalizations.

Simulation of Anisotropies

A covariance (ordinary or generalized) may have several components, and each may have its own anisotropy. Some methods allow a direct construction of simulations with the global covariance; this is the case of the sequential method, the discrete spectral method (provided any zonal anisotropy is oriented parallel to the axes). In other cases the RFs studied can be considered as the sum of independent RFs associated with the various components of the covariance, so we simulate each component separately. The simulation of the anisotropies is not a problem:

- Simulating a zonal anisotropy is the same as making a simulation in a subspace of the observation space.
- A geometric anisotropy can always be reduced to the isotropic case by means of a simple geometric transformation; however, some methods, such as the dilution of Poisson germs, can be used to directly simulate anisotropies using anisotropic dilution functions (i.e., ellipsoids will give a spherical covariance with a geometric anisotropy).
- As we have seen at the end of Section 7.4.2, an isotropy of type (2.67) can be simulated from 1D simulations by a generalization of the turning-bands method.

Simulation of a Locally Stationary Random Function

When we study a domain of large extent, it is rarely homogeneous and with the same covariance everywhere. But generally, we can consider the phe-

nomenon as locally stationary (or locally intrinsic), with a covariance that changes slowly in space. In the neighborhood of a point x_0 , we have, for example, a variogram of the form

$$\gamma_{x_0}(h) = c(x_0)\gamma_0\left(\frac{h}{a(x_0)}\right)$$

where the sill $c(x_0)$ and the scale parameter $a(x_0)$ vary slowly with x_0 . It is important to match these structural parameters.

The variations of the sill (or more generally of a multiplicative factor) pose no problem: to obtain a nonconditional simulation displaying these variations, it suffices to perform a simulation of the basic model with variogram $\gamma_0(h)$, then multiply it by $\sqrt{c(x)}$ at each point simulated. The variations of the scale parameter are more tricky, unless we can directly use a dilution method. To simulate, for example, a spherical model with variable local range $a(x_0)$, we dilute the Poisson germs by spheres of suitable diameter: if the germ i is located at the point $X_i = x$, we dilute it by a sphere with diameter $a(x)$.

7.10.2. Simulation of Blocks

In applications we often need simulations of the mean value over blocks rather than point simulations. For example, the initialization of flow models in hydrogeology sometimes require mean values of the logarithm of transmissivity over the elementary grid cells. In mining applications the simulation of mining methods is based on simulations of average grades over blocks whose size corresponds to the selection unit.

Discretization by a Point Simulation

Some simulation methods define the simulation continuously, but it is usually not possible to integrate it analytically over a block. In practice, the simplest method for simulating a block is usually to discretize it into a fine enough grid and simulate the nodes by point simulation. The simulated block value is the mean of its simulated discretization points. The resulting approximation is acceptable if the variability is small at the scale of the discretization grid. This approach offers two advantages:

- We get both the point and the block simulations. For a simulation of mining methods, we can therefore simulate at the same time an additional reconnaissance survey (e.g., pre-mining drill-holes) and mining blocks (selection units).
- It is very fast to generate a simulation for a new size of blocks, or simulate blocks of variable dimension (e.g., patchwork grid for hydrogeologic calculations).

Discrete Gaussian Model

If we have to simulate a large number of small blocks, the discretization method may become prohibitive. We then perform a direct simulation of the blocks, assuming that they are all of the same size v . From the point covariance $C(h)$ of $Z(x)$, we can deduce the regularized covariance $C_v(h)$ of the block values $Z_v(x)$. Simulating blocks with the point model $C(h)$ simply amounts to simulating points with the regularized model $C_v(h)$.

In the stationary case we can also propagate the change of support to the probability distribution if a Gaussian or Hermitian isofactorial model of change of support is applicable (cf. Section 6.5.1). In such model the SRFs Z and Z_v are considered as deriving from two SRFs Y and Y_v with Gaussian marginal distributions by two transformations φ and φ_v (we recall that $Y_v(x)$ is not the average of $Y(x+u)$ over the block centered at x). The characteristics of the SRF Z_v (transformation function φ_v , covariance functions of Z_v and of Y_v) are deduced from those of the point-support SRF Z (transformation function φ , covariance functions of Z and of Y , themselves related). We therefore have all the elements needed to simulate Y_v and then Z_v .

The most widely used method involves the discrete Gaussian model. In this model each datum is delocalized and considered as located randomly in the small block to which it belongs. The blocks play the prime role and the point support samples can be simulated from the block values with the very simple formula (6.47) (as regards the Gaussian variables). This allows conditioning of the Gaussian simulation on the Gaussian data before transforming it into a conditional simulation of Z_v . Note that this model also allows conditioning a point or block simulation on block data.

7.10.3. Cosimulation

So far we have considered the simulation of a single variable, although on occasions we have mentioned possible ways of cosimulating several variables. Of course, the simulation of several variables requires the knowledge of the multivariate RF model. While for most classic models nonconditional simulation is easy, conditioning on the data is case-dependent. Let us give two examples.

Parallel Simulations

In a multi-gaussian context, possibly after a preliminary transformation, the sequential method can be applied directly to simulate several variables. However, a two-step approach is often easier. Consider, for example, a p -variate model in which all covariances are proportional. It suffices to start from p independent simulations, which can be constructed sequentially, or better, in parallel, if the algorithm lends itself to that (turning bands, discrete spectral method), and then to combine them linearly so as to restore correlations between variables. Conditioning is done afterward by cokriging, and can be

parallelized as well. This method can be readily generalized to the case of a linear model of coregionalization.

Cascaded Simulations

A problem often encountered in geological applications is to simulate a categorical variable representing a facies type, and one or several material properties that depend on the facies type. A favorable case is when there is a strong relationship between the physical property and the facies type, and when we can determine this facies type fairly reliably. A typical example is a formation comprising two lithologies with very different permeabilities (e.g., sand and shale). The simulation is then done in two steps: (1) univariate conditional simulation of the facies type (indicator simulation, or truncated Gaussian with several thresholds, or Boolean model), and (2) within each facies type conditional simulation of the material property using only the measurements in that facies type. The implementation of the first step assumes that for each measurement of the material property we know to which facies it belongs. If not so, we must first assign a facies type to each measurement, conditionally on the observed value (e.g., using a Bayesian model). If several simulations are constructed this assignment must be redone each time.

The other case is when the material property is independent of the facies type; then both variables are simulated separately. In practice, of course, the situation is rarely so clear-cut: there is some degree of dependence of the material property on the facies type, and we must find an adapted method.

7.10.4. Generation of Pseudorandom or Quasi-random Numbers

Simulations of random functions are in fact based on simulations of random variables. The construction of a simulation therefore implies that we have a generator of pseudorandom numbers, or numbers that can be considered as realizations of i.i.d. random variables. The problem is handled in practice in two steps (although there are of course variations):

- Generation of pseudorandom numbers with a uniform distribution over $[0, 1]$.
- Generation of numbers with any given distribution from the previous numbers.

We usually choose a generator that enables us to reproduce a sequence of random numbers generated earlier. It is then possible to reproduce a simulation already constructed (but not saved), or to appreciate the repercussions of a modification of the structural parameters (scale parameter, sill) by constructing several simulations corresponding to different parameters while reusing the same pseudorandom numbers.

Pseudorandom Numbers with a Uniform Distribution

Numerous methods have been proposed. The main ones are presented by Knuth (1981, ch. 3), Ripley (1987, ch. 2 and 3), Press et al. (1992, ch. 7). The method most commonly used, whose properties have been widely studied, is the *linear congruential method*. It produces pseudorandom numbers $x_i \in [0, 1]$ that are in fact rationals of the form $x_i = y_i/m$, where m is a fixed integer, and where the y_i form a sequence of integers less than m defined by the recurrence relation

$$y_{i+1} \equiv [ay_i + b] \pmod{m}$$

The sequence is initialized by y_0 . The multiplier a , shift b , and initial value y_0 are integers less than m . Clearly the sequence y_i can take on at most m distinct values. After a certain time we encounter a number already generated, and the same sequence is then reproduced over and over. The sequence therefore has a period $p \leq m$. Besides it can have a pseudoperiod q , a divisor of p , for which it is reproduced up to a constant (modulo m). To avoid pseudosymmetries, it is therefore recommended not to use the sequence y_i beyond half of its pseudoperiod.

All sets of values a and b do not necessarily lead to a sequence y_i with the required properties of noncorrelation between y_i and y_{i+1} . The structure of the sequence y_i depends essentially on a . Dieter (1972) examines some tests of the independence of y_i and y_{i+1} , particularly the uniformity of the distribution of the (y_i, y_{i+1}) pair. Yet it is not sufficient that y_i and y_{i+1} appear to be independent. It is necessary that the same be true of y_i and y_{i+s} for $s = 2, \dots, s_0$, where s_0 depends on the application. This can be examined by studying the structure of the sequence associated with $a^s \pmod{m}$ instead of a .

However, the absence of correlation of pairs of values up to step s_0 still does not imply that the p -tuples $(y_i, y_{i+1}, \dots, y_{i+p-1})$ ($p \leq s_0 + 1$) are equally distributed over $\{0, 1, \dots, m-1\}^p$, as we might require. Marsaglia (1972) examines this important issue. Ripley (1987, ch. 2) compares several congruential sequences from this standpoint.

As additional protection against the possible deficiencies of these sequences, a good procedure is to use two of them and scramble the order of one of the sequences using the random numbers of the other.

To conclude on these tests, we should emphasize that they pertain to the y_i sequence taken as a whole. In practice, we use only a part, and often an extremely small part of it. Also despite all the care taken in the a priori choice of parameters, it is by no means superfluous to carry out tests on the sequence of numbers that we actually use. In some cases it may be interesting to choose m , a , and b such that the half pseudoperiod is equal only to the number of values to be generated.

Finally, in view of the rapidly expanding digital storage capacity, these pseudorandom numbers generators may well be replaced in the near future by databases of pseudorandom numbers produced by more reliable algorithms or by physical white noise generators.

Pseudorandom Numbers with a Given Distribution

We generally start with pseudorandom numbers distributed uniformly over $[0, 1]$ as, for example, the x_i obtained by the linear congruential method. These values x_i can be transformed into pseudorandom numbers z_i from a c.d.f. F by inversion of F , that is, by taking $z_i = F^{-1}(x_i)$. This method has the practical advantage that a single uniform pseudorandom number suffices to generate a number from the desired distribution. The i th pseudorandom number from distribution F is associated with the i th number of the x_i sequence.

Of course there are other methods: acceptance/rejection, composition of several distributions, quotient of uniform variables, and the like, plus an abundance of algorithms specific to some distributions. A fairly complete presentation can be found in Zelen and Severo (1972), Knuth (1981), and Ripley (1987).

Quasi-random Numbers

Uniform random numbers permit the simulation of random points in a given domain D by drawing the various coordinates independently. Random points produce sequences of points x_1, x_2, \dots , that are *equidistributed* in D , in the sense that the proportion of points among x_1, x_2, \dots, x_i that belong to $D' \subset D$ tends to $|D'|/|D|$ when i tends to infinity, this for any D' . But random points are not homogeneously distributed: clusters can be observed as well as large zones without any point. When selecting the directions of the lines in the turning-bands method, for example, we do not really need random points but simply equidistributed points on the unit half-sphere; these can be generated from equidistributed points over a square (Section 7.4.3).

There exist sequences of equidistributed points that are more homogeneously distributed than random points. The construction usually relies on the intuitive principle that the i th point of the sequence must be located within D as far as possible from the $(i - 1)$ previous ones. Press et al. (1992, sec. 7.7, pp. 309–315) present several sequences that are equidistributed over a square, and more generally over a n -dimensional hypercube.

A simple example is the *Halton sequence*, also named *Van der Corput sequence* (Bouveau, 1986), which is defined as follows: The space dimensionality being n consider the first n prime numbers p_j ($p_1 = 2$, $p_2 = 3$, $p_3 = 5$, etc.). The number $i \in \mathbb{N}$ is written in base p_j

$$i = a_{j0} + a_{j1}p_j + \dots + a_{jk}p_j^k + \dots$$

We set

$$u_j(i) = \frac{a_{j0}}{p_j} + \frac{a_{j1}}{p_j^2} + \dots + \frac{a_{jk}}{p_j^{k+1}} + \dots$$

The sequence of n -tuples $x_i = (u_1(i), \dots, u_n(i))$, $i > 0$, is equidistributed over $]0, 1[^n$.

7.10.5. Check of the Simulations

Just like the real phenomenon, a simulation is not a random function but a regionalized variable that we regard as a realization of a random function. Its actual structural characteristics (histogram, variogram, etc.), evaluated from its values at the nodes of the discretization grid, differ more or less from the characteristics of the theoretical model or those of the sample data. Therefore it is useful to check that the characteristics of the simulation are not too far from those sought.

Checking the simulations is all the more useful that their construction relies in practice on pseudorandom numbers whose behavior cannot be totally guaranteed. Furthermore conditioning can introduce anomalies if the spatial distribution of the simulated RF is not really compatible with the data. We present below some general-purpose tests designed to verify that the simulations are technically correct. A validation of the physical correctness is naturally required as well, but it is application-specific.

Check of a Single Simulation

Statistical tests designed for independent random variables are not applicable here because the values taken on by a simulation are obviously correlated. But we can at least perform simple checks at various levels: for example, Gaussian nonconditional simulation, Gaussian conditional simulation, and final simulation. The checks concern the histogram of simulated values, their variogram, and even the scatterplots between the values assumed at x and $x + h$ for a few values of h . These graphs have a regional character at the scale of the simulated domain, since they are calculated from a fine discretization of this domain. We are therefore able to calculate the fluctuation variance of the spatial mean of the simulation (according to (2.33), it depends only on the covariance), and in some cases, at least the Gaussian one, of the spatial variance of the simulation

and of its variogram (cf. Section 2.9.2). This enables us to see if the deviations of these characteristics from their theoretical values are acceptable.

Check of a Set of Simulations

Several simulations of the same phenomenon are often generated in order to exhibit several images of what reality might be. In addition to the above checks on individual simulations, two methods can be applied to make sure that the different simulations are really (conditionally) uncorrelated:

1. *Check of the normed sum.* Consider N independent simulations (conditional or not) $S_i(x)$ with the same variogram $\gamma(h)$. Their normed sum

$$S(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^N S_i(x)$$

still has the variogram $\gamma(h)$ (but the histogram becomes more Gaussian—if it was not already Gaussian). We can check this using the methods mentioned above.

2. *Check of simulated errors.* This test is concerned with the conditional simulations obtained by adding to the kriging estimate a simulation of the kriging error. Using the notations of Section 7.3.1 (Z : real field; S_i : nonconditional simulation number i ; T_i : conditional simulation number i ; Z^* : kriging of Z ; S_i^* : kriging of S_i ; σ_K^2 : kriging variance), the field

$$\varepsilon_i(x) = \frac{T_i(x) - Z^*(x)}{\sigma_K(x)} = \frac{S_i(x) - S_i^*(x)}{\sigma_K(x)}$$

is a standard error. If the nonconditional simulations $S_i(x)$ are Gaussian, this must also be true of $\varepsilon_i(x)$. This often remains true to a first approximation in the non-Gaussian case, since the $\varepsilon_i(x)$ are mixtures of random variables with the same distribution. For a fixed i , the errors $\varepsilon_i(x)$ are obviously spatially correlated. On the other hand, for a fixed x , the errors $\varepsilon_i(x)$, $i = 1, \dots, N$, must be independent, centered, standardized, and possibly Gaussian. We can easily verify this for a few points of the simulation using classic tests (e.g., the χ^2 test). An example is given by Chilès (1977).

7.11. CASE STUDIES

7.11.1. Simulation of a Nickel Deposit

Conditional simulations were in fact initially motivated by the study of mineral deposits (Journel, 1973, 1974a, 1974b). Kriging enables estimation of the resources in place but not the evaluation of the recoverable reserves, which

is yet of prime importance in feasibility studies. The difficulty of evaluating recoverable reserves is due to the fact that this concept involves the mining method, its degree of selectivity, its flexibility, none of which can be reduced to linear operations, and also to the fact that the result depends heavily on the local fluctuations of the mineralization. To determine the optimal mining method and the corresponding recoverable reserves, we would need to have knowledge of the deposit at every point, which of course is never the case. A conditional simulation is then very useful. It provides a numerical model that can be known at every point. This model is used to simulate various mining methods or ore processes, perform sensitivity studies, and optimize choices. To the extent that the numerical model provided by the conditional simulation reproduces not only the global aspect of the actual deposit but also its local variability, we can reasonably consider that the method that proves optimal on the numerical model remains optimal, by and large, for the real mineral deposit.

The simulation of mining methods based on a numerical model produced by a conditional simulation constitutes a subject that was developed around 1980 and is beyond the scope of this book. The interested reader is referred to Deraisme and Dumay (1979), Da Rold et al. (1980), Deraisme et al. (1983, 1984), Kim et al. (1982), Luster (1986), and Vieira et al. (1993). Deraisme and van Deursen (1995) apply this approach to a standard civil engineering problem, the volumetric computation of dredging projects, and quantify the bias incurred by cutting on a kriged estimate rather than on simulations. The area estimate for blasting work above the cutting level is 273,500 m² based on the kriged model, whereas conditional simulations of the seabed and the geological layers give an interval of 268,000 to 391,000 m², with an average of about 310,000 m². The following example, while focused on the description of the methodology for generating conditional simulations, gives a glimpse of the contribution of the approach to the choice of a mining method.

Orebody and Objective of the Simulation

The Tiébaghi nickel orebody (New Caledonia) is sampled by several hundred drill-holes with an approximate grid spacing of 70 m × 70 m. The ore can be subdivided into six categories, from top to bottom:

- The cover: C.
- Three categories of laterites: L1, L2, L3.
- Two categories of garnierites: G1, G2.

Below we find the bedrock. Figure 7.45 shows a typical drill-hole log.

The studied zone forms a plateau. The global resources in place have been estimated by kriging and are known with good precision, namely within 8%, using a conventional confidence interval of ± 2 standard deviations. Local resources, however, are poorly estimated; a 70 m × 70 m panel can hardly be evaluated better than up to 50%. Yet the choice of the mining method

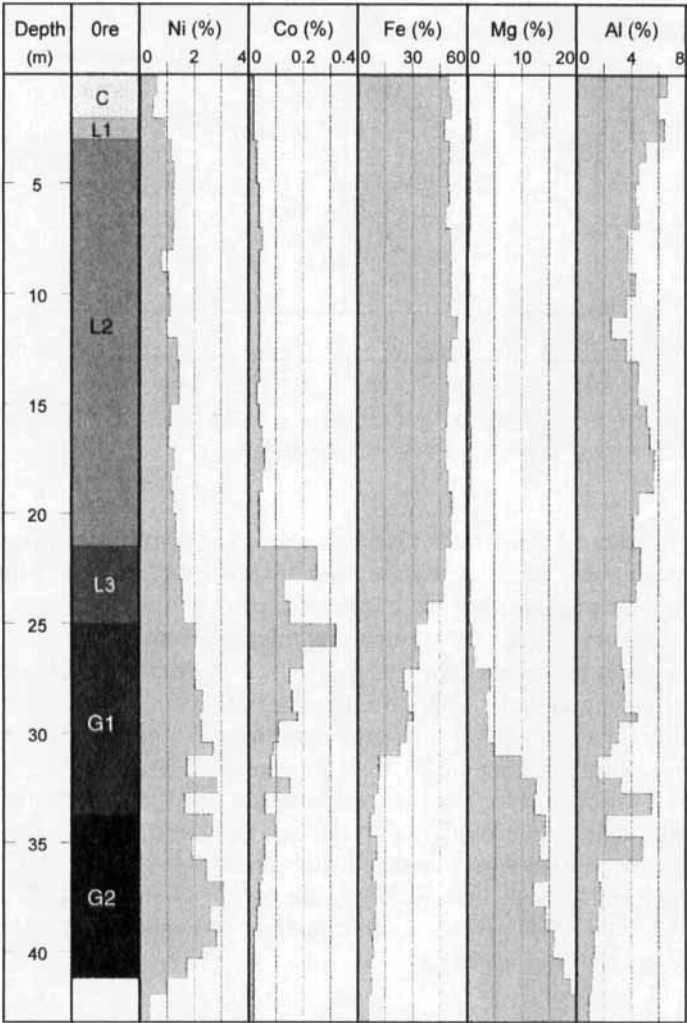


FIGURE 7.45. Typical drill-hole log in the Tiébaghi nickel orebody. From Chilès (1984), with kind permission from Kluwer Academic Publishers.

depends on even more local characteristics, such as fluctuations of the top of the garnierites or of the bedrock. It is not possible, even by halving the grid spacing (a fourfold increase in the number of drill-holes), to estimate these surfaces with reasonable precision. Hence the idea of constructing a conditional simulation. We can then test various mining methods (bulk or selective mining, homogenization stock piles, single or multiple working areas, etc.) and search for the one that best matches the ore supply requirements of the processing plant (homogeneity of the ore, regularity of supply), and also the economic criteria (cutoff grade, undesirable side product).

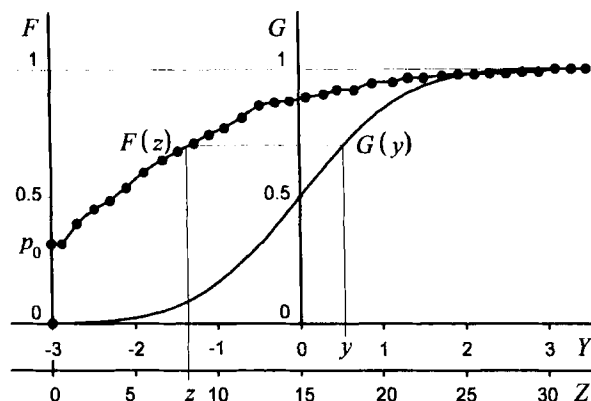


FIGURE 7.46. Graphic transformation of thickness of the L2 lateritic ore. From Chilès (1984), with kind permission from Kluwer Academic Publishers.

The simulation is constructed on a fine grid ($6.25 \text{ m} \times 6.25 \text{ m}$), which enables us to visualize local fluctuations and, by grouping of 4×4 points, to simulate with good precision the mean characteristics of the $25 \text{ m} \times 25 \text{ m}$ panels. For each category of ore, we simulate its thickness and also the mean grades of the various constituents (Ni, Co, Mg, Al). The grades are obtained in fact by simulating the associated accumulations (i.e., average grade \times thickness). Naturally thicknesses and accumulations are highly correlated and call for the use of multivariate methods described in Chapter 5. Appropriate adaptations are required to account for very special bivariate distributions of thickness and metal accumulation (not reducible to the bi-gaussian case). We focus here on the simulation of thicknesses alone, which prove to be uncorrelated between ore categories. The task then involves the construction of six different simulations. Chilès (1984) gives a more complete presentation of the simulation procedure, including for grades.

Gaussian Transform

To reproduce the sample distribution of the thickness of a given ore, we simulate the Gaussian transform associated with this thickness. Thus we start by transforming the initial data into Gaussian data, and in the end we do the reverse transformation on the Gaussian simulation.

The transformation function is deduced from the cumulated histogram. Figure 7.46 shows as an example the histogram of the thickness of the L2 ore, along with its graphic fit. We observe a discontinuity at the origin, due to the fact that 32% of the time the L2 ore is absent. This creates a problem for the transformation of the “waste” data into a Gaussian: to this zero thickness we must associate a Gaussian value less than the limit $y_0 = G^{-1}(0.32)$, but which one? (G is the c.d.f. of the standard Gaussian). We will come back to this later.

Variography

Simulations are much more sensitive than kriging to the behavior of the variogram near the origin. There is no hope of obtaining a realistic simulation at a 6.25-m grid spacing if we do not know the variogram at less than 70 m. Fortunately some complementary data are available besides the 70-m spacing data:

- A cross of drill-holes at 25-m spacing. This is located at the edge of the plateau and must be corrected for a proportional effect.
- About 30 old drill-holes, located at about 5 m from drill-holes of the 70-m grid. Since they are regularly spaced over the whole plateau, they provide for a good estimation of the variogram at 5-m lag.

The variography is simplified by the fact that the variograms of laterites turn out to be similar, as do those of garnierites. In order to compute the variograms of the Gaussians, the zero thicknesses are transformed into the conditional expectation of a Gaussian $E[Y | Y < y_0]$. The resulting bias can be calculated theoretically in the scope of a bivariate Gaussian model for Y and is taken into account in the fit of a variogram model. The main variability occurs in the first 200 meters for the laterites and in the first 100 meters for the garnierites.

The variograms of original thicknesses are of no direct use for the construction of the simulation, since this is done on the Gaussians. Note, however, that if a Hermitian transform were considered instead of a graphic transform, these variograms would be used, since the variogram of a transform is deduced from that of the Gaussian (see Section 6.3.4). The variograms of thicknesses nevertheless provide a reference for checking the quality of the final simulation (after conditioning and reverse transformation).

Construction of the Simulation

This involves three main steps:

1. *Construction of nonconditional simulations of the Gaussians.* This step is straightforward; it is carried out independently for each ore category. The random coins algorithm is used.
2. *Conditioning on the data.* Ideally the conditioning kriging should be carried out with a global neighborhood. This prevents the parasitic discontinuities caused by neighborhood changes, which make it impossible to know whether the fluctuations of the conditional simulation represent something real or are a mere artifact. Here the number of data precludes conditioning with a global neighborhood. Therefore we divide the domain into subzones, each being kriged with a global neighborhood. Some overlap of the subzones is allowed to avoid discontinuities at the boundaries. Conditioning on zero thicknesses, that is, on Gaussian

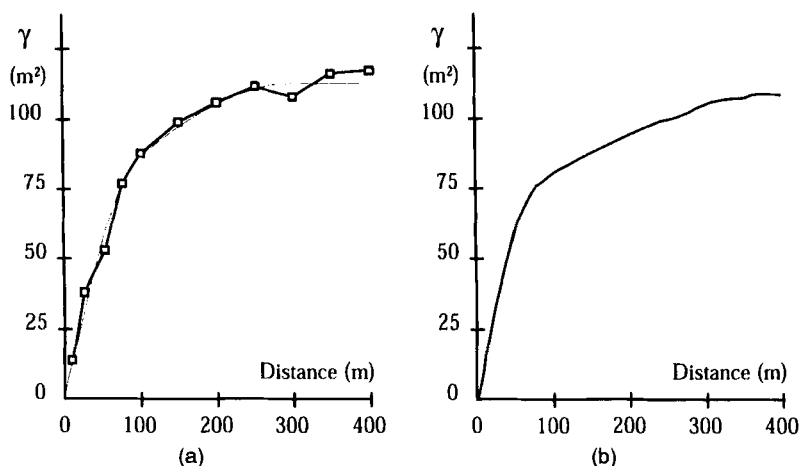


FIGURE 7.47. Sample variogram of the thickness of the laterites (a) and regional variogram of the conditional simulation (b). From Chilès (1984), with kind permission from Kluwer Academic Publishers.

data for which all we know is that they are less than the threshold u_0 , can be performed rigorously using the algorithm presented in Section 7.7.3. At the time of the study, this method was not yet known and zero thicknesses were just replaced, after Gaussian transformation, by the conditional expectation of the Gaussian.

3. *Reverse transformation to original thicknesses.* We now have conditional simulations of the Gaussians. By reverse transformation we obtain the conditional simulations of the thicknesses. In this direction the transformation is straightforward.

Check of the Simulation

Checks are made at each step. In the final stage we check that histograms of sample thicknesses are well reproduced by the simulation, and that the same is true for the variograms. Figure 7.47 displays the sample variogram of the total thickness of laterites along with that of the conditional simulation. They seem fully compatible (the slight nugget effect seen on the sample variogram, due to the uncertainty in the exact position of the interfaces, is not simulated).

A more general check is provided by a comparison between simulation and kriging. Indeed the characteristics of the real orebody lie within a range of ± 2 standard deviations about the kriged values (referring to the conventional 95% interval). The simulation must fall in the same interval. To check this, the mean thicknesses of the different ores are estimated directly by kriging. As Table 7.2 shows, the simulation is fully compatible with kriging. We note that it corresponds to an outcome relatively weak in laterites.

TABLE 7.2. Comparison of average thicknesses of the simulation and of kriging (in meters)

Ore	Simulation	Kriging	Standard Error
C	11.1	11.0	0.5
L1	4.5	5.1	0.5
L2	6.1	6.4	0.6
L3	3.7	4.3	0.6
G1	10.2	9.9	0.6
G2	2.9	3.1	0.4

Results

Figure 7.48 displays a vertical section of the orebody obtained by kriging, and the same vertical section derived from the conditional simulation. As in kriging the simulation matches the data (the two sections coincide at drill-holes locations). However, the striking difference is the character of these two sections. If we based the mining method on the kriged section, we would be making a serious mistake: the regular beds would suggest the use of an unselective bulk mining technique (e.g., a dragline or a bucket wheel excavator), while if reality exhibits variations similar to those shown by the simulation, more mobile and more selective mining techniques are to be considered, at least for the garnierites (shovels or back-hoes).

7.11.2. Simulation of an Oil Reservoir

After mining, the first applications of conditional simulations were in hydrology (Delfiner and Delhomme, 1975) and petroleum (Delfiner and Chiles, 1977). Applications to oil reservoirs grew with methods for simulating lithological variations (Alabert, 1987b; Matheron et al., 1987). A typical simulation of an oil reservoir is a three-step procedure (Dubrule, 1993):

1. Simulation of the genetic units (fluvial channels, mouth bars, turbiditic lobes).
2. Simulation of facies types within these units (sand, shale, sandstone).
3. Simulation of petrophysical variables within each facies type (porosity, permeability, saturation, etc.).

These simulations reflect the heterogeneities and the complexity of the reservoir. Their main application is realistic reservoir fluid flow and oil recovery simulation from which a range of likely economic scenarios can be derived. Examples of the complete sequence of steps for such simulation exercise are given by Hewett and Behrens (1988), and Morelon et al. (1991), and a vari-

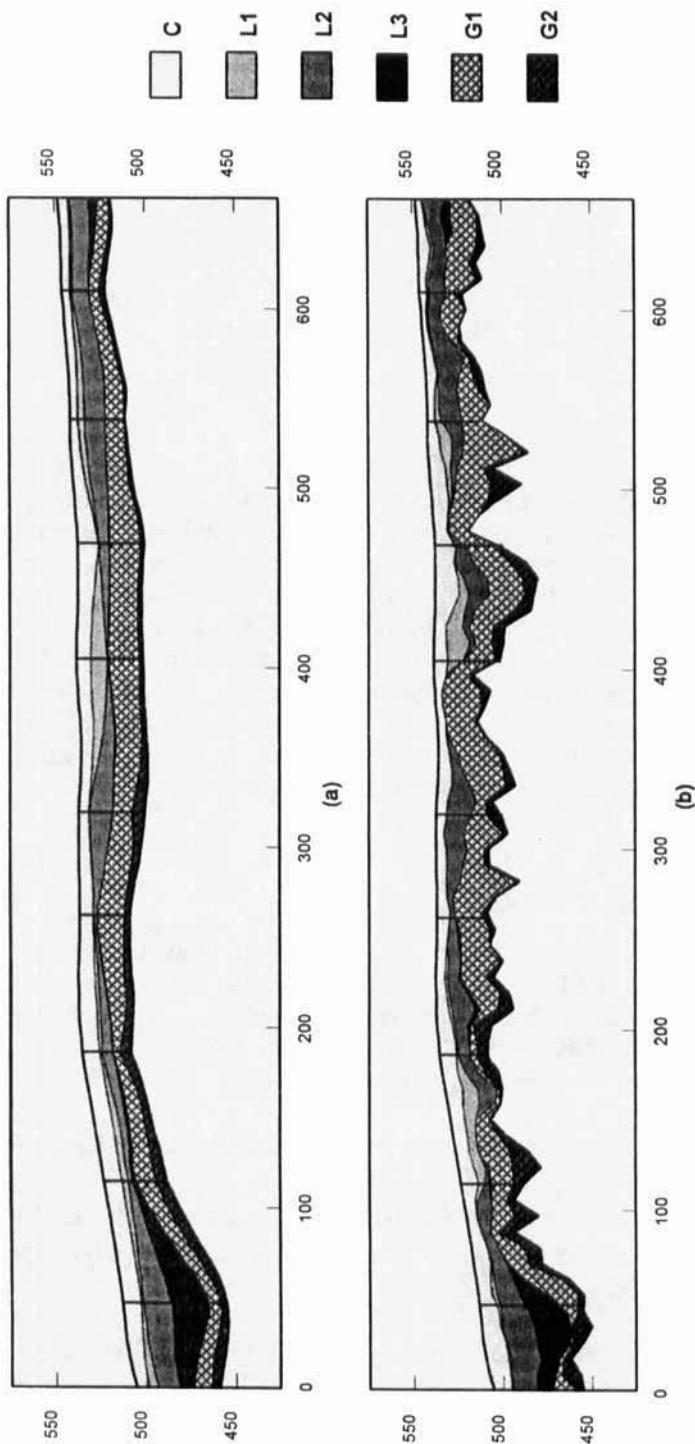


FIGURE 7.48. Example of a vertical section through the Tiébaghi nickel orebody: (a) the kriged section; (b) simulated section. From Chilès (1984), with kind permission from Kluwer Academic Publishers.

ety of case studies are described in the book edited by Yarus and Chambers (1994).

Step 1 is modeled with object-based methods: genetic units are represented as objects of random shape, size, and orientation randomly distributed in space. Early applications by Delhomme and Giannesini (1979) modeled sandstone reservoir units, drains, and shale breaks in the Hassi Messaoud field (Algeria). More recent applications abound (e.g., references cited by Dubrule, 1993). These methods offer the advantage of flexibility: the distribution of the parameters can easily be adapted to local conditions. Matheron (1969b) and Jacod and Joathon (1971) also developed *random genetic models* where the sedimentation process is governed by a differential equation relating sediment thickness to influx of sedimentary material and sea depth. A variety of situations can be simulated reflecting the choice of the parameters (rate of influx of sedimentary material, compaction, subsidence, meandering): for example, continuous layers, lenses, migration of channels. The genetic hypotheses give an explanatory value to the model and not just a descriptive one. Random genetic models can be conditioned in cases of simplification (Jacod and Joathon, 1972). The situation had been similar for Boolean models for a long time, but we have seen that a general conditioning method is now available and that simulated annealing or Bayesian techniques can be used to impose constraints or integrate other data.

Step 3 involves the techniques developed for continuous variables. The first application was the volumetric evaluation of the Lacq reservoir (Delfiner and Chilès, 1977), which generalized a method based on a shortcut to simulations proposed by Haas and Jousselin (1976). Since the reservoir is bounded by the top of a dome and truncated below by an aquifer, its estimation is a nonlinear problem, namely the estimation of the integral of $[Z_{\text{OWC}} - Z(x)]1_{Z(x) < Z_{\text{OWC}}}$ where $Z(x)$ is the depth to the reservoir top, counted positively downward, and OWC is the oil-water contact. Volume calculations were therefore based on conditional simulations. Table 7.3 shows the results obtained using the first 25 wells drilled plus two other wells selected to constrain simulations realistically ("pseudo" wells are often necessary to ensure closure of the reservoir). As there were in fact 84 wells drilled these could be used to establish "true" values.

One can see that simulations give better estimates than kriging, which systematically under-estimates volumes. It is also possible to compute volume histograms and determine the traditional P95, P50, P5 quantiles ($\Pr\{\text{Volume} > P95\} = 95\%$). The method can be extended to the calculation of oil in place by introducing porosity and fluid saturations. This type of probabilistic reserves evaluation has become common in the oil industry. Small-scale permeability variations can also be generated within the reservoir (Alabert and Massonnat, 1990).

The following example is focused on step 2, the simulation of facies types. A common approach for such applications is sequential indicator simulation: Journel and Gómez-Hernández (1989) use it to simulate the sand/shale

TABLE 7.3. Bulk rock volume estimates based on 27 wells

OWC	Bulk Rock Volume (10^6 m^3)			
	Kriging	Simulations	σ Simulations	"True" Value
620 m	302.8	324.6	44.7	331.4
630 m	362.7	394.9	54.5	397.6
640 m	430.0	473.2	65.3	470.6

Note: Volumes from kriging depths estimates; mean and standard deviation of 30 simulations; mean of 30 simulations based on 84 wells.

distribution in a clastic sequence, Alabert and Massonnat (1990) to model the spatial arrangement of sedimentary bodies (channels, turbiditic lobes, slumps, laminated facies, etc.); Doyen et al. (1994) extend the method in a Bayesian setting to incorporate seismic data in addition to well data. The approach of the Heresim Group (Matheron et al., 1987, 1988) is based on a particular instance of the truncated Gaussian model that facilitates simulation and conditioning and is very flexible. An application of this method is presented now.

Objective of the Simulation

The example considered by Matheron et al. (1987, 1988) is the Brent cliff (Great Britain). This cliff comprises three lithofacies types: shale, shaly sandstone, and sandstone exhibiting local permeabilities that differ by orders of magnitude. They have been clearly identified on a photograph of the cliff, and the objective is to construct a 3D model by conditional simulation for the purpose of studying flows. (To this effect, a permeability depending on the facies outcome is assigned to every node of the simulation.) The data are preprocessed to make the beds horizontal.

We number the three facies types (shale, shaly sandstone, sandstone), respectively, 1, 2, and 3 and denote by $Z(x)$ the number of the facies type encountered at point x . $Z(x)$ is a discrete diffusive variable, in the sense that transitions occur from facies number i to facies number $i - 1$ or $i + 1$ but not directly from facies 1 to 3. This is a typical case for the application of truncated Gaussian simulation (Section 7.7.4). Implementation details for the present case study can be found in Matheron et al. (1988). Only the main steps are outlined here.

Variography

The first hurdle when we wish to implement the truncated Gaussian method is to identify the covariance of the Gaussian. Since this cannot be determined directly, we do it indirectly by considering related variables like the facies indicators or the facies numbers. While remaining in the diffusive context, let us consider more generally the case where the space is partitioned into m ordered facies, in the proportions p_1, \dots, p_m (with of course $p_1 + \dots + p_m = 1$).

The indicators $I_1(x), \dots, I_m(x)$ of the m facies are considered as defined by the application of one or several thresholds to a standard Gaussian RF⁸ $Y(x)$,

$$I_i(x) = 1_{y_{i-1} \leq Y(x) < y_i} \quad \text{with} \quad -\infty = y_0 < y_1 < \dots < y_{i-1} < y_i < \dots < y_m = +\infty$$

The thresholds y_i are chosen so as to match the proportions p_i of the m facies

$$y_i = G^{-1} \left(\sum_{j=1}^i p_j \right) \quad i = 1, \dots, m-1$$

If the RF $Y(x)$ can be assumed stationary, we could fit its covariance by modeling the direct and cross-variograms of the cumulated indicators

$$J_i(x) = \sum_{j=i+1}^m I_j(x) = 1_{Y(x) \geq y_i} \quad i = 1, \dots, m-1$$

but this is not obvious, as mentioned in Section 6.2.4. To simplify this task, we can more simply model the variogram of the facies number $Z(x)$ from the variogram of $Y(x)$. To modulate the relative importance of the facies changes, we define $Z(x)$ more generally by

$$Z(x) = b_0 + \sum_{i=1}^{m-1} b_i J_i(x)$$

where b_i are coefficients that we fix (the standard case where $Z(x)$ is the facies number corresponds to $b_i = 1 \forall i$). Then $Z(x)$ can be expanded into normalized Hermite polynomials by applying (6.26):

$$Z = b_0 + \sum_{i=1}^{m-1} b_i [1 - G(y_i)] - \sum_{n=1}^{\infty} \left[\sum_{i=1}^{m-1} b_i g(y_i) \chi_{n-1}(y_i) \right] \frac{\chi_n(Y)}{\sqrt{n}}$$

From this and formula (6.27) we deduce the following relation between the covariance $C(h)$ of the facies number Z , which can be determined empirically, and the correlogram $\rho(h)$ of the Gaussian RF Y ,

$$C(h) = \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} b_i b_j g(y_i) g(y_j) \sum_{n=1}^{\infty} \frac{\chi_{n-1}(y_i) \chi_{n-1}(y_j)}{n} \rho(h)^n \quad (7.55)$$

The situation would be straightforward if $Z(x)$ were stationary: we would choose $\rho(h)$ so that $C(h)$ fits the sample covariance $\hat{C}(h)$. In the present case, Z

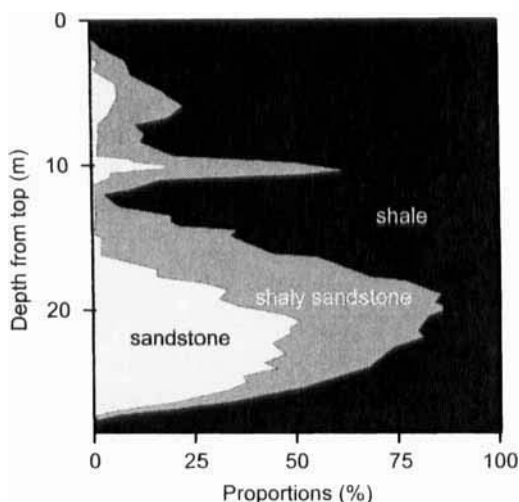


FIGURE 7.49. Proportion curves of the facies along the vertical. From Matheron et al. (1988).

cannot be considered stationary along the vertical: sandstones, for example, are located essentially in the lower half, as we see in Figure 7.49, which shows the proportion of the three facies as a function of vertical position. On the other hand, Z can be considered stationary horizontally, for the variograms of the facies indicators reach a sill at about 750 steps, as shown in Figure 7.50. We denote by (x, y, z) the coordinates of the point x , where the Ox axis is oriented horizontally and parallel to the cliff and the Oz axis is vertical, and we denote by (h_x, h_y, h_z) the components of the vector h . To account for the vertical nonstationarity, the thresholds y_i are defined as functions $y_i(z)$. The theoretical covariance of $Z(x)$ and $Z(x + h)$ is given by an expression similar to (7.55) in which y_i is replaced by $y_i(z)$ and y_j by $y_j(z + h_z)$ and is therefore a function $C(h; z)$. To fit the usual z -averaged sample covariance $\hat{C}(h)$, we postulate a correlogram $\rho(h)$ and match it with the theoretical $C(h)$ obtained by the same z -averaging of the nonstationary covariance $C(h; z)$. Figure 7.51 shows the sample covariances obtained along a horizontal line and along a vertical line (many other lines have been processed at the same time), and also their fit by the same exponential factorized model

$$\rho(h) = \exp \left(-\frac{|h_x|}{a_x} - \frac{|h_y|}{a_y} - \frac{|h_z|}{a_z} \right)$$

with $a_x = 2.4$ m and $a_z = 50$ m. a_y remains to be determined but is not involved here in the fit since $h_y = 0$. The exponential factorized model is selected for its Markov properties which greatly facilitate the construction of simulations.

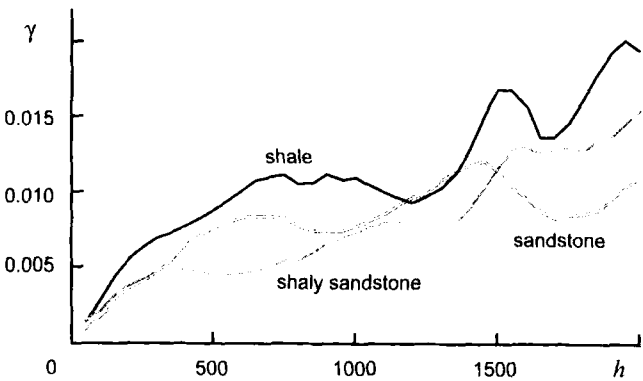


FIGURE 7.50. Horizontal variograms of the indicators of facies of the Brent cliff. From Matheron et al. (1988).

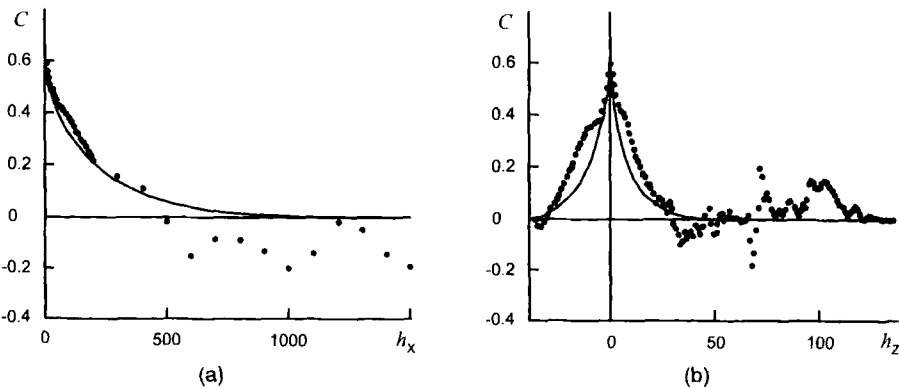


FIGURE 7.51. Sample covariance $\hat{C}(h; z)$ along the horizontal line z (a) and along the vertical lines (b), for $z = 40$, and fit by a common factorized exponential model. From Matheron et al. (1988).

Factorization of Exponential Covariances

An SRF $Y(x)$ with a factorized correlogram

$$\rho(h) = \rho_x(h_x)\rho_y(h_y)\rho_z(h_z)$$

has conditional independence properties for points at a right angle as seen in Section 3.6.1. Consider for simplicity the case of a 2D vertical regular grid with unit spacing, where x and i represent the horizontal, and z and j the vertical. If $j - 1$ verticals have already been simulated, the next vertical can be decomposed into its conditional expectation given those and a residual independent of the previous verticals

$$Y(i, j) = Y_K(i, j) + R(i, j)$$

$Y_K(i, j)$ is the simple kriging estimator of $Y(i, j)$ by the $j - 1$ columns already simulated. Due to the screening effect associated with the factorized model, only the $Y(i', j')$ located on the same line i can have a nonzero weight. Therefore $Y(i, j)$ may be written as

$$Y(i, j) = \sum_{\alpha=1}^{j-1} \lambda_{\alpha} Y(i, \alpha) + R(i, j)$$

Note that for simplicity we let α range from 1 to $j - 1$ as if these columns were the first $j - 1$ columns of the array, but in fact the columns can be simulated in any order (this property is important when we want to build conditional simulations). Each column of residuals $R(i, j)$ is independent of the others but has a vertical structure. More precisely, the vertical covariance $C_R(h)$ of the residuals $R(i, j)$ is the same as that of the Gaussian RF $Y(i, j)$, up to a multiplicative factor. This vertical covariance of Y can be written as

$$\begin{aligned} \rho_z(h) &= E[Y(i, j)Y(i + h, j)] \\ &= \sum_{\alpha=1}^{j-1} \sum_{\beta=1}^{j-1} \lambda_{\alpha} \lambda_{\beta} \rho_z(h) \rho_x(\beta - \alpha) + C_R(h) \end{aligned}$$

Hence

$$\begin{aligned} C_R(h) &= \rho_z(h) \left[1 - \sum_{\alpha=1}^{j-1} \sum_{\beta=1}^{j-1} \lambda_{\alpha} \lambda_{\beta} \rho_x(\beta - \alpha) \right] \\ &= \sigma_K^2 \rho_z(h) \end{aligned}$$

where σ_K^2 is the simple kriging variance of one point from the $j - 1$ points of the same row already simulated (note that the λ_{α} and σ_K^2 can vary from one row to the next).

Simulating $Y(i, j)$ is then very simple:

1. Simulation of Gaussian data consistent with the conditioning indicator data (using the Gibbs sampler method as explained in Section 7.7.3).
2. Reconstruction of each conditioning vertical (the conditioning data are usually vertical boreholes, but there may be missing data).
3. Nonconditional simulation of normal unit-variance residuals $R_1(i, j)$ with vertical covariance ρ_z (and no horizontal correlation).
4. Lateral conditioning (iteration over the nonconditioning verticals)

$$Y(i, j) = Y_K(i, j) + \sigma_K(j) R_1(i, j)$$

In practice, the covariances ρ_x and ρ_z are often exponential models because this simplifies the conditioning task:

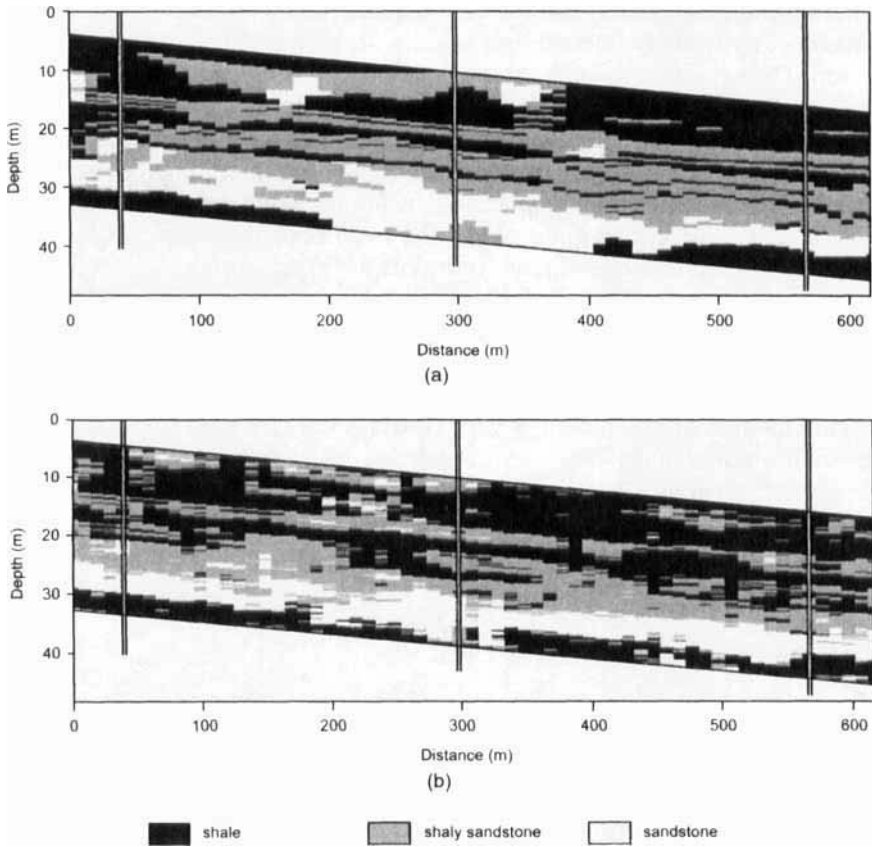


FIGURE 7.52. “Real” cliff (a) and a simulation conditioned by three “drill-holes” (b). From Matheron et al. (1988).

- Horizontally (step 4), kriging $Y(i, j)$ from previously known points of the same line involves only the nearest neighbor on either side.
- Vertically (step 2), the situation is similar, so that the 1D sequential algorithm presented at the end of Section 7.2.1 can be used.
- Step 1 can be done independently for each borehole (if the borehole interdistance is large in comparison to the range).

Finally the simulation of $Y(x)$ is transformed into a simulation of the facies numbers by truncation using the thresholds y_i .

Results

Figure 7.52 displays part of the real cliff and also a simulation conditioned on three “drill-holes” located in the section, for which a_v is not involved. The simulation reproduces the main features of the section: strong lateral continuity

of the sandstone and shaly sandstone bed at the base with local thinning of the sandstone, more shaly intermediate streaks, scattered sandstone channels near the top. The general style of the sedimentary deposits is accurately reproduced, although the simulation exhibits more irregular contours than the initial image. This results from the choice of the covariance model but does not constitute a defect, since the real cliff has been smoothed for interpretation. Matheron et al. (1988) present other simulations, including along perpendiculars to the section or in a horizontal plane, under the hypothesis that the scale parameter along v , the direction of sediment transport, is 70 m.

It is important to check if the simulations are representative of the reality as regards flow properties. A calculation of the equivalent permeability on the image of the real cliff and on simulations shows deviations of only about 15%, a reasonably good result (Matheron et al., 1987).

This method of simulation is very flexible. We can, for example, vary the thresholds horizontally. We can also define each facies not by an interval $[y_{i-1}, y_i[$ of a single Gaussian $Y(x)$ but as a domain in the 2D space defined by a pair of Gaussian SRFs, independent or correlated, which enables us to overcome the limitations of diffusive models as shown in Section 7.7.4.

NOTES

1. This is the ideal case of image analysis. In practice, images often lack spatial homogeneity, so the multivariate distributions remain unknown.
2. The dilation $V \oplus S$ is the union of all translates of V by a vector of S (see Serra, 1982).
3. This denomination is used when the random process represents the location of a particle on the line as a function of time. It is then usually denoted by $X(t)$.
4. In \mathbb{R}^n we do not use the terminology “(fractional) Brownian *motion*” in order to avoid confusion with the n -dimensional random process representing the trajectory of a particle in the space \mathbb{R}^n as a function of time.
5. Warning: Starting from different definitions, Matheron and Serra use one-half of this λ as the main parameter, and Miles uses a third parameter.
6. The Sichel distribution and the generalizations proposed by Matheron are very useful for modeling discrete distributions with a very long tail, typically the distribution of the number of precious stones in a given volume.
7. If the objects (e.g., the channels) are regarded as a permeable phase, there is a critical value of the proportion of area occupied by the objects for which the system *percolates*, in the sense that a connected component much larger than the others suddenly connects two opposite boundaries. This proportion is called the *percolation threshold* and has been studied extensively by Allard (1993) and Allard et al. (1993).
8. Here $Y(x)$ could be interpreted as a curve indicative of shaliness, such as a gamma-ray log, adequately scaled.

CHAPTER 8

Scale Effects and Inverse Problems

Felix qui potuit rerum cognoscere causas.
Virgil

8.1. INTRODUCTION

We now turn to varied and difficult physical problems for which no general solution has yet been found. Since it is not possible to be complete, we will consider the problems that arise in the discipline most explored so far, which is hydrogeology. The reader is invited to transpose the problems and approaches to other disciplines of interest, an obvious one being petroleum engineering where, for single-phase flow, it suffices to adapt the vocabulary (e.g., “reservoir” for “aquifer,” “potential” for “head,” and so on). On change of scale, this chapter owes much to the still state-of-the-art work of Matheron (1967). Geostatistical work on inverse problems has been more diversified and includes the contributions of Delhomme, de Marsily, Kitanidis, Dagan, Rubin, Gómez-Hernández, to quote only a few.

So far we have considered variables whose large-scale properties derive from small-scale properties by simple averaging. Thus the porosity of a block is simply the mean of the porosities of the thousands of cores that the block is made of, and the core values themselves are the means of porosities defined at the microscopic level (they are 0/1 functions, e.g., the indicator of the pores). Porosity is *additive*. But a physical quantity such as permeability is not additive: two permeabilities in series combine by harmonic averaging. Permeability is inherently a macroscopic quantity, defined at the scale of a core; it cannot be defined at the microscopic scale and has also no unique definition at the megascopic scale of large blocks. The change of scale does not amount to the usual change of support.

In groundwater hydrology and in petroleum engineering, permeability is an essential parameter, since it controls the geometry of flow paths and barriers. The objective of a geostatistical reservoir characterization is to provide

a numerical model of the reservoir parameters (porosity, permeability, storativity, fluid saturations) suitable as input to a flow simulator. Various scales are involved, including the macroscopic scale of core measurements in the lab, the megascopic scale of pumping tests, and the scale of block models needed by flow simulators. An important concept is that of *equivalent permeability*. In petroleum engineering terms, “the equivalent permeability of a reservoir segment is defined as the permeability of a homogeneous segment of the same dimensions that would pass the same flux under the same pressure drop” (Cardwell and Parsons, 1945). The *upscaling* problem is how to calculate equivalent permeability at a larger scale knowing permeabilities everywhere at a finer scale.

In engineering applications we never have an exhaustive sampling of the reservoir, but we may have permeability and piezometric-head data. Within a geostatistical model, this involves the *joint* spatial distribution of permeability and piezometric head. But the two quantities are related by the flow equation: when permeability and boundary conditions are given, the head is determined. Therefore consideration of the flow equation is the key for developing physically consistent stochastic models of permeability and head. This example is typical of variables whose physical behavior is governed by partial differential equations.

Permeability observations are usually scarce, for they require setting up monitoring wells to measure the water level and carry out pumping tests, from which permeabilities are deduced. Piezometry data are easier to get than permeability data. Since the water level is directly related to the permeability field, piezometry data must be used to reduce the uncertainty on permeability. This observation leads us to the inverse problem: Find a permeability field matching the permeability data and consistent with the piezometry measurements.

The present chapter comprises three main sections. The first section presents Darcy’s law and discusses the change-of-scale problem for permeabilities. The second section is concerned with general properties of stochastic differential equations, and how they differ from deterministic equations. The method of Green functions is illustrated on an important class of linear stochastic systems in time. The stochastic partial differential equation $\Delta Z = Y$ is also discussed in some detail. The third section deals with the inverse problem in the context of hydrogeology.

8.2. UPSCALING PERMEABILITY

8.2.1. Darcy’s Law: Observation and Derivation

One of the most important properties of a porous medium is its permeability, defined by Darcy’s law.¹ Let us first briefly recall the experimental context in which this law appears (Fig. 8.1). We cut out from the rock a cylindrical core with length L and cross-sectional area A . We seal it laterally and put

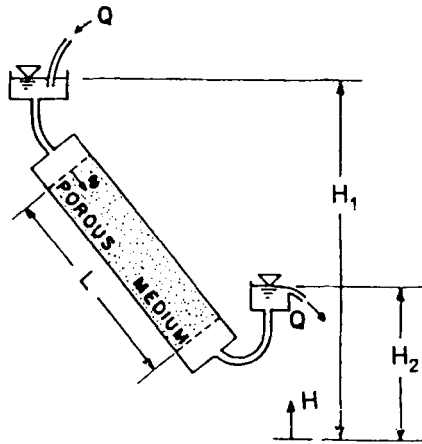


FIGURE 8.1. Sketch of Darcy's experiment. From Dagan (1989, Fig. 2.9.1), © Springer-Verlag.

the two end sections A_1 and A_2 in contact with the same fluid. We impose a hydraulic-head or piezometric-head² difference $H_1 - H_2$ between the two ends using two tanks (the head difference is equal to the difference in water level elevations between the inflow and exit tanks), wait until a steady-state flow is established, and measure the flow rate Q through the cylinder (volume of fluid passing per unit time). Alternatively, the flow rate Q can be prescribed and the head drop measured. In both cases we observe that the flow rate is proportional to the head drop. Repeating the experiment with other cylindrical cores with different lengths and cross-sectional areas, we establish empirically the following relation

$$Q = K \frac{H_1 - H_2}{L} A \quad (8.1)$$

When the fluid is water at standard temperature the proportionality coefficient K is called permeability or hydraulic conductivity and only depends on the porous medium.³

Therefore, at the macroscopic level, a steady-state flow may be described by a vector U , representing the mean velocity (or *specific discharge*), and a head gradient $\text{grad } H$, related by

$$U = -K \text{grad } H \quad (8.2)$$

This is Darcy's law. It shows that K has the dimension of a velocity [LT^{-1}]. If the medium is not macroscopically isotropic, while remaining statistically homogeneous, one only compares samples with the same in situ orientation and the coefficient K obtained is relative to this orientation. Repeating the experiment for different orientations, the same relation (8.2) is obtained in which the scalar coefficient K is replaced by a tensor. The hydraulic conductivity K ,

be it a scalar or a tensor, is a macroscopic characteristic of the porous medium. Thus $U = Q/A$ is a *filtration velocity* and not the mean fluid motion velocity, for the fluid only circulates in the pores (and in fact only in the connected pores).

Darcy's law is observed empirically at the macroscopic scale. At the microscopic scale, however, the flow is governed by the Navier-Stokes equations, which express the basic principles of mechanics for a viscous fluid. In the case of a steady-state flow, these reduce to

$$\Delta u_i = \alpha \frac{\partial h}{\partial x_i} \quad i = 1, 2, 3$$

and can be condensed into

$$\Delta u = \alpha \text{grad } h \quad (8.3)$$

where u and h represent the velocity and piezometric head at a point of the fluid, at the microscopic scale, and α is a constant that depends on the fluid.⁴ In addition the fluid, being incompressible, satisfies the continuity equation (conservation of mass)

$$\text{div } u = 0 \quad (8.4)$$

These equations only apply in the pore space and are subjected to extraordinarily complex boundary conditions, expressing that in addition to the external conditions on head or flow rate, fluid velocity cancels at the surface of contact between the grains and the pores. To effectively integrate these equations would require a detailed description of the geometry of the pore space at the microscopic level, which is clearly impossible for a real porous medium.

To approach this problem, Matheron (1967, ch. V) turns to a stochastic model. The grains are represented by a stationary and ergodic random set and the solutions of the equations by stationary vector random functions u and $\text{grad } h$ (h itself is the sum of an IRF-0 and a linear drift). The macroscopic specific discharge U is represented in this model by the expected value $U = E[u]$, and the constant macroscopic head gradient $\text{grad } H$ is equal to the (conditional) expectation of the head gradient $\text{grad } h$ at any point *in the pores*: $\text{grad } H = E[\text{grad } h(x) \mid x \text{ in the pores}]$. Matheron observes that the linearity of equations (8.3) and (8.4) entails a linear relationship between u and $\text{grad } h$, and thus between U and $\text{grad } H$, which is the linear form of Darcy's law.⁵ But this analysis does not allow a prediction of the numerical values of the proportionality coefficient K . From energetic considerations it is only proved that the permeability tensor and its inverse are positive definite.

8.2.2. Calculating Equivalent Permeabilities

Consider now passing from the macroscopic to the megascopic scale of large meshes or the gigascopic scale of the aquifer. The support over which the

medium appears as porous and homogeneous, and therefore can be characterized by its permeability, is now regarded as a point. Shifting to a higher scale, this medium can no longer be considered homogeneous, and permeability is a function of point x . The first question that arises is whether a new Darcy's law emerges, making it possible to characterize the global hydrodynamic properties at that new scale by a constant "average" permeability (to be defined).

Let us first examine the one-dimensional case, corresponding to Darcy's original experiment, but with a column made of cores with different permeabilities, which we represent as a function of the abscissa x . Darcy's law at "point" x is

$$U(x) = -K(x) \frac{dH}{dx}$$

In steady-state flow this filtration velocity is the same at all points (the fluid has only one way to go) and is therefore equal to the specific discharge $q = Q/A$. We have

$$\frac{dH}{dx} = -\frac{q}{K(x)} \quad (8.5)$$

and the head drop in the column is given by

$$H_1 - H_2 = q \int_0^L \frac{dx}{K(x)}$$

The harmonic mean K_h defined by

$$\frac{1}{K_h} = \frac{1}{L} \int_0^L \frac{dx}{K(x)}$$

enables the result to be written in the form

$$H_1 - H_2 = \frac{qL}{K_h}$$

Comparing this with Darcy's law (8.1) shows that, globally, the column behaves like a homogeneous medium with permeability K_h .

In two or three dimensions the problem is quite difficult. Many approaches have been tried, but no definitive solution has yet been found. Renard and de Marsily (1997) present an excellent review of the various methods used to calculate the equivalent permeability of a heterogeneous porous medium in the case of steady-state, uniform, and single-phase flow. We will be concerned here only with methods that fall in the scope of geostatistics. We will distinguish the gigascopic scale of the aquifer as a whole, and the megascopic scale of the meshes of the hydrodynamic models.

Effective Permeability

We now regard permeability as a regionalized variable, more precisely as a realization of a stationary and ergodic (tensor) random function $K(x)$. Darcy's law (8.2) remains valid, but U , K , and H are RFs. Our problem is to know whether at a sufficiently large scale, which we will call the gigascopic scale, the variations of detail exhibited by the RF $K(x)$ may be forgotten to be summarized by a constant tensor, equivalent from the point of view of the global flow and called *effective permeability*. If this exists it is defined by

$$E[U] = -K_{\text{ef}} E[\text{grad} H]$$

where $E[U]$ is the expected mean velocity in the given domain and $E[\text{grad} H]$ the head gradient expectation.

In the case of uniform flow, Matheron (1967, ch. VI) shows that when the domain becomes large enough with respect to the range of the permeability variogram the two expectations are proportional. In other words, there is emergence of an effective permeability K_{ef} . Such effective permeability is an intrinsic quantity independent of the conditions at the domain boundaries. When not a scalar, it is a positive-definite second-order tensor. Its value usually depends on the spatial distribution of K , but in some important special cases it actually only depends on the marginal distribution.

The most general inequality is defined by the Wiener bounds (O. Wiener, 1912, for scalar permeability; Matheron, 1967, in the general case):

$$[E(K^{-1})]^{-1} \leq K_{\text{ef}} \leq E(K) \quad (8.6)$$

where $A \leq B$ means $B - A$ positive definite. The effective permeability is always comprised between the harmonic mean $K_h = [E(K^{-1})]^{-1}$ and the arithmetic mean $K_a = E(K)$ of the "point-support" permeabilities. For example, in n -dimensional space when permeability is a product of the form

$$K(x) = K_1(x_1)K_2(x_2) \cdots K_n(x_n) \quad (8.7)$$

where x_1, x_2, \dots, x_n are the coordinates of point x and K_1, K_2, \dots, K_n independent stochastic processes with the same distribution, the effective permeability can be shown to be

$$K_{\text{ef}} = K_a^{(n-1)/n} K_h^{1/n} \quad (8.8)$$

which indeed verifies (8.6).

It is known that the geometric mean is also comprised between the arithmetic and harmonic means. Inequalities (8.6) thus justify, to some extent, the empirical geometric averaging rule often used in practice. This rule writes

$$\log K_{\text{ef}} = E(\log K) \quad (8.9)$$

In fact Matheron shows that this rule applies rigorously to *two-dimensional* flows in one circumstance: when K and its reciprocal the resistivity $W = 1/K$, scaled by their respective expectations, have the same rotation-invariant spatial distribution.⁶ These fairly restrictive conditions are met in the very interesting special case where K is a lognormal RF with isotropic covariance. Formula (8.9) is then equivalent to

$$K_{\text{ef}} = [E(K)]^{1/2} [E(K^{-1})]^{-1/2} \quad (8.10)$$

The geometric averaging rule is specific to 2D space. Indeed, for a slowly varying scalar permeability field with an isotropic covariance it is possible, following Schwydtler (1963a, b), to write K in the form $K(x) = E(K)[1 + \gamma(x)]$ with $E\gamma = 0$, which leads to the second-order approximation

$$K_{\text{ef}} = E(K) \left[1 - \frac{E\gamma^2}{n} \right] + 0(E\gamma^2) \quad (8.11)$$

where n is the dimension of space. This is compatible with geometric averaging only for $n = 2$. For $n = 1$ we have seen that K_{ef} coincides with the harmonic mean K_h . Formula (8.11) shows that K_{ef} increases with n : as the number of dimensions increases, a greater number of possible flow paths are offered to the fluid to go around low permeability zones. Note also this qualitative consequence of (8.11): when the *arithmetic* mean is fixed, effective permeability decreases as the permeability field becomes more heterogeneous.

Matheron notes that (8.11) can also be considered as an approximation to (8.8) (at order 2) and that formula (8.10) obtained in two dimensions is identical to formula (8.8) for $n = 2$. He conjectures that (8.8) is a generalization of (8.10) to 3D space in the conditions under which this was obtained, that is, when $K/E(K)$ and $W/E(W)$ have the same rotation-invariant spatial distribution. For a lognormal permeability with isotropic covariance, this conjecture takes on the form

$$\log K_{\text{ef}} = E(\log K) + \left(\frac{1}{2} - \frac{1}{n} \right) \sigma^2 \quad (8.12)$$

where σ^2 is the variance of $\log K$. A similar conjecture was already formulated by Landau and Lifshitz (1960) for electrodynamic equations. Formula (8.12) has been successfully tested by many authors (cf. Renard and de Marsily, 1997), including for values of σ^2 that are not small (up to $\sigma^2 = 7$). Qualitatively, we now see that in 3D when the *geometric* mean is fixed, effective permeability increases with heterogeneity. Physically this makes sense because greater heterogeneity creates the possibility of high permeability flow paths. (Note that there is no contradiction with the remark about formula (8.11), since the arithmetic mean and the logarithmic dispersion are not independent parameters.)

Upscaled Permeability

In engineering practice we often have to work at an intermediate scale, called the megascopic scale, where the conditions required for the emergence of an effective permeability are not satisfied. A *block permeability* is the equivalent permeability of a finite-size block. Rubin and Gómez-Hernández (1990) take as a criterion the flow through the block and define the block permeability by

$$\frac{1}{V} \int_V U(x) dx = K_b(V) \frac{1}{V} \int_V \text{grad}(H(x)) dx$$

This permeability depends on the boundary conditions and thus is not an intrinsic property of the block. In the isotropic case the Wiener bounds remain valid if mathematical expectations are replaced by means in V . There are only a few cases where an exact analytical calculation of block permeability is possible. In n -dimensional space, for example, if the permeability is a product of independent factors with the same distribution as in (8.7), the permeability of a block oriented parallel to the axes is given by

$$K_b(V) = K_a(V)^{(n-1)/n} K_h(V)^{1/n}$$

where

$$K_a(V) = \frac{1}{V} \int_V K(x) dx \quad K_h(V) = \left[\frac{1}{V} \int_V \frac{dx}{K(x)} \right]^{-1}$$

This result coincides with formula (8.8) transcribed in terms of spatial averages. By analogy it is common to plug in spatial averages in an effective permeability formula, but one must remember that it is only an approximation whose validity is not assured if the block is small. These methods, as well as heuristic and deterministic methods, are discussed by Sánchez-Vila et al. (1995) and by Renard and de Marsily (1997).

A popular method proposed by Journé et al. (1986) is *power averaging of order α*

$$K_b(V) = \left[\frac{1}{V} \int_V K(x)^\alpha dx \right]^{1/\alpha} \quad -1 \leq \alpha \leq 1$$

$\alpha = 1$ corresponds to arithmetic averaging and $\alpha = -1$ to harmonic averaging, while the limit $\alpha = 0$ corresponds to geometric averaging. In the case of a scalar isotropic permeability field, consistency with the Schwydtler formula is achieved for $\alpha = 1 - 2/n$.

As will be shown in Section 8.4, block permeability can be obtained, for specified boundary conditions, either by an approximate analytical calculation or by solving the flow equations numerically over a regular fine-mesh grid discretizing the block and its neighborhood. The first approach is used by Rubin and Gómez-Hernández (1990) in the case of lognormal permeability

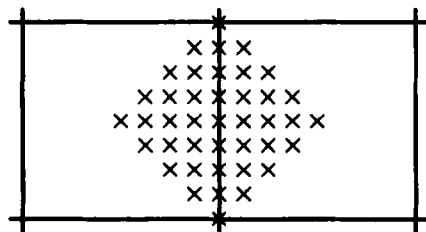


FIGURE 8.2. Support for the approximate calculation of inter-block permeability for a finite differences code (two dimensions). From Roth et al. (1996), © American Geophysical Union.

with small variance and for boundary conditions sent to infinity. The second one is used by Holden and Lia (1992) to calculate block permeability values corresponding to different boundary conditions and then best fit a permeability tensor. Such calculations can be done for all blocks of interest, or simply for some of them to serve as a basis for fitting the exponent α in the power averaging formula.

Block permeabilities are sought because they constitute the basic input to flow models at the scale of the aquifer. However, finite difference or finite element methods rely in fact on *inter-block permeabilities*, involved in the expression of the flow traversing the face common to two adjacent blocks. Assuming a constant permeability in each block and the two blocks in series, numerical methods take as inter-block permeability the harmonic mean of the permeabilities of the two blocks. This leads to potentially serious biases in the case of heterogeneous media, which is precisely the case of interest here. A better approach is to calculate directly the inter-block permeabilities without going through the step of block permeabilities. White and Horne (1987) carry out this calculation by finite differencing in a neighborhood of the interface, finely discretized (which reduces bias), and for several boundary conditions, then fit an inter-block permeability tensor. Kruehl-Romeu and Nøtinger (1995), Roth (1995), and Roth et al. (1996) propose ways to obtain scalar inter-block permeabilities that do not cause biases when used with finite differencing. In particular, Roth shows numerically that in 2D inter-block permeability can be calculated by translating the effective-permeability formula—if it exists—into a spatial average in the diamond shown in Figure 8.2. For example, if permeability is lognormal, inter-block permeability can be considered equal to the geometric mean of point-support permeabilities in this diamond (the approximation is excellent, even for a large variance of $\log K$). These methods for calculating permeability at the megascopic scale are usually implemented on simulations to determine their spatial distribution from the spatial distribution of point-support permeabilities (Rubin and Gómez-Hernández, 1990; Gómez-Hernández and Journel, 1990; Roth, 1995).

This abundance of results may give the false impression that change of scale is under control. Setting aside numerical methods, careful inspection shows,

however, that the formulas obtained only involve the distribution of point-support permeabilities, namely the marginal distribution of K for effective permeability, or the histogram of $K(x)$ in V for block permeability. Contrary to intuition, these formulas incorporate no concept of connectivity of low (resp. high) permeability zones, which may create flow barriers (resp. flow paths), nor even any consideration of the covariance of K . Yet experience shows that the connectivity of high permeability zones plays an essential role. This apparent contradiction simply underlines the fact that the formulas are only valid in specific cases of simplification (factorized or lognormal permeability) which, as important as they may be in practice, remain very special.

8.3. STOCHASTIC DIFFERENTIAL EQUATIONS

As we have seen with the Navier-Stokes equations and Darcy's law, physical laws usually take the form of differential or partial differential equations. Their transposition to the stochastic domain is fairly natural but can present specific difficulties, especially when white noise is involved. The questions asked for stochastic differential systems are the same as for deterministic ones, namely the existence and uniqueness of a solution, the analytical properties of the solutions, and how they depend on parameters and initial conditions. However, the introduction of stochastic elements brings new questions, such as the determination of the spatial distribution of the solution and the existence of stationary solutions. The study of stochastic differential equations is, by itself, a vast subject. The interested reader is referred to the work of Gikhman and Skorokhod (1972), Arnold (1973), Friedman (1975–1976), Sobczyk (1991), to quote only a few. We will give a brief presentation of general properties of stochastic differential equations and will examine in more detail the stochastic equation $\Delta Z = Y$. The next section, devoted to inverse problems in hydrogeology, will show an example of application to the earth sciences.

8.3.1. General Properties of Stochastic Differential Equations

Problem Definition

A *differential system* is a set of equations relating a variable t , often interpreted as time, a number of functions of t , and some of their derivatives. By increasing the number of unknown functions and the number of equations, one can always rewrite the system as a canonical system including only first derivatives. For example, the equation

$$f(t, z, z', z'') = 0$$

where $z' = dz/dt$ and $z'' = d^2z/dt^2$ is equivalent to the system

$$\begin{cases} z'_1 = z_2 \\ f(t, z_1, z_2, z'_2) = 0 \end{cases}$$

where the first component z_1 plays the role of z . Grouping the unknown functions in a vector $\mathbf{Z}(t)$ of dimension m and the known functions in a vector $\mathbf{Y}(t)$ of dimension p , and introducing initial conditions, we will examine systems that take on the form

$$\begin{cases} \frac{d\mathbf{Z}(t)}{dt} = \mathbf{F}(\mathbf{Z}(t), \mathbf{Y}(t)) \\ \mathbf{Z}(t_0) = \mathbf{Z}_0 \end{cases} \quad (8.13)$$

where \mathbf{F} is a vector transformation operator. The usual objective is to find the expression of $\mathbf{Z}(t)$ as a function of $\mathbf{Y}(t)$ and the initial conditions. The problem becomes stochastic if the known functions, the equation parameters, or the initial conditions are made random.

Regular Stochastic Differential Equations

First the probabilistic meaning of the differential system must be clarified. Stochastic differential equations are said to be regular if the stochastic processes and the vector transformation \mathbf{F} satisfy certain regularity conditions. There are two approaches:

- Take equation (8.13) in the mean square sense: if $\mathbf{Z}(t)$ is a solution, $\mathbf{F}(\mathbf{Z}(t), \mathbf{Y}(t))$ is the derivative of $\mathbf{Z}(t)$ in the mean square sense.
- Regard $\mathbf{Y}(t)$ as the ensemble of its realizations $\mathbf{y}_\omega(t)$ —ordinary vector functions—and equation (8.13) as a symbolic expression for an ordinary differential equation satisfied by all realizations of \mathbf{Y} , \mathbf{Z} , and \mathbf{F} (which can also be random) for every elementary event $\omega \in \Omega$. The solution $\mathbf{Z}(t)$ of the stochastic system is then defined as the ensemble of solutions $\mathbf{z}_\omega(t)$. Of course this can only make sense if the realizations of \mathbf{Y} and \mathbf{Z} are sufficiently regular.

When each of these approaches has a unique solution, the results are equivalent (cf. Sobczyk, 1991, sec. 21).⁷ As in classic differential equations, if the initial conditions are not known, the general solution of the equation is a class of stochastic processes.

Differential Equations of Generalized Stochastic Processes

More often than not differential equations involve regionalized variables that are far from being differentiable. To resolve this mathematical contradiction, one can regard them as *generalized functions*. In essence this amounts to considering regularized versions of the regionalized variables by weighting functions with the desirable differentiability properties (usually they are selected in the space \mathcal{D} of infinitely differentiable functions with bounded support). We will not go into this theory, which is developed by Schwartz (1950–1951) and by Gel'fand and Shilov (1958).

The notion of stochastic process can be extended to *generalized random process*. Just as for random processes it is possible to define stationary or

intrinsic generalized random processes. These permit a general treatment of stochastic differential equations, except that their solution can no longer be obtained by solving the deterministic system for realizations, since these are now generalized functions.⁸

Covariance of the Solution

Let us come back to regular systems (to some extent the approach can be transposed to other cases). If $\mathbf{Z}(t)$ is a solution of the stochastic system (8.13), it is a (multidimensional) stochastic process. It is interesting to know its spatial distribution (we should say, its *time distribution* if t represents time) knowing that of \mathbf{Y} , as well as the joint distribution of the \mathbf{Y} and \mathbf{Z} processes. There is no easy answer in the general case, but the moments of \mathbf{Z} and of (\mathbf{Y}, \mathbf{Z}) can be calculated either from the expression of \mathbf{Z} as a function of the other variables, or as averages over realizations. This holds for means and covariances, which completely characterize the spatial distribution in the Gaussian case.

To be specific consider a linear system such as

$$\begin{cases} \frac{d\mathbf{Z}(t)}{dt} + \mathbf{A}(t)\mathbf{Z}(t) = \mathbf{Y}(t) \\ \mathbf{Z}(t_0) = \mathbf{Z}_0 \end{cases} \quad (8.14)$$

where $\mathbf{A}(t)$ is a matrix of ordinary continuous functions, $\mathbf{Y}(t)$ a m.s. continuous stochastic process with finite second-order moments, and \mathbf{Z}_0 a deterministic or random vector with finite variance. Since \mathbf{Y} is involved linearly on the right-hand side, the solution in the m.s. sense is of the form

$$\mathbf{Z}(t) = \mathbf{G}(t, t_0)\mathbf{Z}_0 + \int_{t_0}^t \mathbf{G}(t, s)\mathbf{Y}(s)ds \quad (t \geq t_0)$$

where \mathbf{G} is a matrix. The first term is deterministic or else the product of a deterministic function by a time-independent random vector. The second term is a linear functional of \mathbf{Y} . The mean and covariance of \mathbf{Z} are easily deduced.

To simplify this, let us suppose that the system is a canonical representation of an equation of order m

$$\begin{cases} \frac{d^m Z(t)}{dt^m} + a_1(t)\frac{d^{m-1}Z(t)}{dt^{m-1}} + \cdots + a_n(t)Z(t) = Y(t) \\ Z(t_0) = \frac{dZ}{dt}(t_0) = \cdots = \frac{d^{m-1}Z}{dt^{m-1}}(t_0) = 0 \end{cases} \quad (8.15)$$

where the $a_j(t)$ are ordinary continuous functions and $Y(t)$ is a scalar m.s. continuous second-order stochastic process. The solution is of the form

$$Z(t) = \int_{t_0}^t G(t, s)Y(s)ds \quad (8.16)$$

The *Green function*—or *impulse function*— $G(t, s)$ represents the response of the system to a unit impulse at time s , as a function of time $t > s$ (for fixed s the univariate function $G(\cdot, s)$ is the solution of (8.15) when Y is the Dirac delta-function δ_s centered at s). Denoting by m_Y and σ_Y the mean and covariance functions of Y , the mean of Z is found to be

$$m_Z(t) = \int_{t_0}^t G(t, s) m_Y(s) ds$$

and its covariance

$$\sigma_Z(t_1, t_2) = \int_{t_0}^{t_1} \int_{t_0}^{t_2} G(t_1, s_1) G(t_2, s_2) \sigma_Y(s_1, s_2) ds_1 ds_2$$

If the functions a_j are constants, the impulse response is of the form $G(t, s) = p(t - s)$, and the solution (8.16) takes on the form

$$Z(t) = \int_{t_0}^t p(t - s) Y(s) ds \quad (8.17)$$

If in addition Y is a stationary process, m_Y is constant and $\sigma_Y(t_1, t_2) = C_Y(t_2 - t_1)$. Thus

$$m_Z(t) = m_Y \int_{t_0}^t p(t - s) ds$$

$$\sigma_Z(t_1, t_2) = \int_{t_0}^{t_1} \int_{t_0}^{t_2} p(t_1 - s_1) p(t_2 - s_2) C_Y(s_2 - s_1) ds_1 ds_2$$

Likewise the cross-covariance of Y and Z can be expressed in terms of the covariance σ_Y or C_Y of Y .

Existence of a Stationary Solution

If the multidimensional process $\mathbf{Y}(t)$ is strictly stationary, can equation (8.13) have a stationary solution $\mathbf{Z}(t)$? Inspection of the covariance of \mathbf{Z} provides an answer at least for second-order stationarity. In a linear system such as (8.14) with a matrix \mathbf{A} that is constant and has eigenvalues whose real part is negative, if \mathbf{Y} is second-order stationary and if the initial state is sent to $t_0 = -\infty$, \mathbf{Z} is second-order stationary.

In particular, for a linear system (8.15) where the functions a_j are constants, if Y is second-order stationary, so is the solution (8.17) with mean

$$m_Z = m_Y \int_0^\infty p(s) ds$$

and covariance (cf. (2.36))

$$C_Z = C_Y * P \quad P = p * \check{p}$$

In the frequency domain this translates into the simple formula

$$F_Z(du) = |\varphi(u)|^2 F_Y(du)$$

where $\varphi(u)$ is the Fourier transform of p and $F_Y(u)$ and $F_Z(u)$ the spectral measures of the covariances of Y and Z .

Example 1

Let us consider the regular stochastic differential equation

$$\frac{dZ(t)}{dt} + aZ(t) = Y(t)$$

where $a > 0$ and where $Y(t)$ is a stationary Gaussian process with zero mean and covariance $C_Y(h) = \sigma^2 \exp(-b|h|)$. Given the initial condition $Z(0) = z_0$, the solution is

$$Z(t) = z_0 e^{-at} + \int_0^t e^{-a(t-s)} Y(s) ds$$

Its mean is $m_Z(t) = z_0 e^{-at}$, and its covariance is

$$\sigma_Z(t_1, t_2) = \frac{\sigma^2}{a^2 - b^2} \left[e^{-b|t_2 - t_1|} - \frac{b}{a} e^{-a|t_2 - t_1|} + \left(1 + \frac{b}{a} \right) e^{-a(t_1 + t_2)} - e^{-at_1 - bt_2} - e^{-at_2 - bt_1} \right]$$

In particular, when $t \rightarrow \infty$, its behavior becomes stationary with mean zero and covariance

$$C_Z(h) = \frac{\sigma^2}{a(a+b)} \frac{ae^{-b|h|} - be^{-a|h|}}{a-b}$$

Example 2

Let us now consider the equation

$$\frac{dZ(t)}{dt} + aZ(t) = \varepsilon(t)$$

where $a > 0$ and ε is a generalized random process, not necessarily Gaussian, with covariance the generalized function $b^2 \delta(h)$, $b > 0$. Z is a generalized random process with covariance a generalized function K to be determined. Let us only search for stationary solutions whose covariance K is a generalized function. If Z were a m.s. differentiable random process with covariance $C(h)$, the covariance of $Z'(t) + aZ(t)$ would be $-C''(h) + a^2 C(h)$. This formula can be extended to a generalized random process. The starting equation thus entails

$$-K'' + a^2 K = b^2 \delta$$

whose solution is the generalized function that coincides with the ordinary function

$$K(h) = \frac{b^2}{2a} e^{-a|h|}$$

This function is not differentiable at zero. So Z can be regarded as an ordinary stochastic process but not its derivative.⁹

Example 3

Whittle (1954) generalized the preceding equation to 2D by considering the second-order equation

$$\Delta Z(x) - a^2 Z(x) = \varepsilon(x)$$

where Δ is the Laplacian operator and ε is 2D white noise with covariance $b^2 \delta(h)$. This equation has stationary solutions in the plane, with the isotropic K -Bessel covariance

$$C(r) = \frac{b^2}{2a} r K_1(ar)$$

8.3.2. The Stochastic Equation $\Delta Z = Y$

If the vectors \mathbf{Y} and \mathbf{Z} are functions of a point $x = (x_1, \dots, x_n)$ of \mathbb{R}^n , or even, for transient phenomena, of spatial coordinates x_1, \dots, x_n and time t , one can define by extension partial differential equations, complemented by initial conditions and boundary conditions. The basic properties of stochastic partial differential equations and other more general systems can be found in Sobczyk (1991, secs. 23 and 29) for random functions and in Gel'fand and Shilov (1958) for generalized random functions.

We will specialize our study to the important *Poisson* equation

$$\Delta Z = Y \tag{8.18}$$

where Y and Z are RFs of \mathbb{R}^n . In this equation Y represents a source term and is given; when Y is equal to zero, the equation is named the *Laplace* equation. The Poisson equation represents fluid flow in a porous medium with a constant permeability equal to one, where Z is the piezometric head and Y the inflow/outflow (due to rainfall, or injection, or pumping counted negatively). It can also represent the flow of heat in a homogeneous medium with thermal conductivity equal to one, where Z is the temperature and Y the heat production.¹⁰

Stochastic Model

The first question that arises is that of the existence of a stochastic model compatible with this equation. Let us first consider the 1D case. When Y is deterministic, the differential equation has for solution the second integral of Y , plus an arbitrary polynomial of degree 1. When Y is an SRF with zero mean,

we have seen in Section 4.4.2 that the second integral of Y is a representation of an IRF-1. Since the various representations of an IRF-1 differ only by an arbitrary polynomial of degree 1, the solution of the equation is an IRF-1 in the sense of the definition of Section 4.4.3. Likewise, if Y is an IRF-0, Z is an IRF-2. In both cases the solution is an IRF, that is, a class of ordinary RFs. Like in the deterministic case, boundary (or initial) conditions can restrict the generality of the solution to a particular representation. These results can be generalized to higher-dimensional spaces and to IRFs of a higher order, as shown by the following theorem from Matheron (1971b):

Theorem. If Y is a continuous IRF- k of \mathbb{R}^n , there exists a unique twice differentiable IRF- $(k+2)$ satisfying the differential equation $\Delta Z = Y$.

Uniqueness of course pertains to the IRF- $(k+2)$, regarded as a class of RFs.

Covariance Models

From this we can derive the relationship between the generalized covariances of Z and of Y , as well as the (generalized) cross-covariance of Y and Z (Matheron, 1971b; Dong, 1990, pp. 55–59). To keep things simple, let us consider only the case where Y is a zero-mean SRF. Let $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ be two points of \mathbb{R}^n . Denoting by $C(h)$ the stationary covariance of Y , we get

$$C(y-x) = E[Y(x)Y(y)] = E[\Delta Z(x)\Delta Z(y)] = E\left[\sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 Z(x)}{\partial x_i^2} \frac{\partial^2 Z(y)}{\partial y_j^2}\right]$$

With Z being twice differentiable, its nonstationary covariance $\sigma(x, y)$ is differentiable four times, and therefore

$$C(y-x) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i^2} \frac{\partial^2}{\partial y_j^2} E[Z(x)Z(y)] = \Delta_x \Delta_y \sigma(x, y)$$

where Δ_x is the Laplacian operator applied with respect to x . The ordinary covariance σ of Z and its generalized covariance K being related by (4.17), we get

$$\Delta_x \Delta_y \sigma(x, y) = \Delta_x \Delta_y K(y-x) = \Delta^2 K(h) \quad \text{for } h = y-x$$

The covariances C of Y and K of Z are thus related by the equation

$$\Delta^2 K(h) = C(h) \tag{8.19}$$

Likewise for the cross-covariance we get

$$\begin{aligned} E[Y(x)Z(y)] &= E[\Delta Z(x)Z(y)] = E\left[\sum_{i=1}^n \frac{\partial^2 Z(x)}{\partial x_i^2} Z(y)\right] \\ &= \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} E[Z(x)Z(y)] = \Delta_x \sigma(x, y) \end{aligned}$$

Given (4.17), for $k = 1$ this is of the form

$$E[Y(x)Z(y)] = -\Delta K(h) + \Delta c_0(x) + \sum_{i=1}^n \Delta c_i(x) y_i \quad h = y - x \quad y = (y_1, \dots, y_n) \quad (8.20)$$

In practice, we do not know which representation of the IRF-1 we are dealing with, and only work on linear combinations allowable at order 1. The terms involving c_0 and the c_i then disappear, and the covariance between $Y(x)$ and an ALC-1 $Z(\lambda)$ can be calculated using the generalized cross-covariance $-\Delta K(h)$.

Case Where Y Is White Noise

If Y is not a continuous IRF- k but rather a random noise with covariance $C_0 \delta$, it should be interpreted as a generalized random function. The covariance K of Z is then a generalized function satisfying

$$\Delta^2 K = C_0 \delta$$

As seen in Section 4.6.4, this is the equation satisfied by the generalized covariance associated with biharmonic spline interpolation. Consequently the function K is, up to a multiplicative factor, the isotropic generalized covariance r^3 in \mathbb{R}^1 , $r^2 \log r$ in \mathbb{R}^2 , $-r$ in \mathbb{R}^3 , $-\log r$ in \mathbb{R}^4 , and $r^{-(n-4)}$ in \mathbb{R}^n , $n > 4$. Note that the irregularity of the RF Z increases with the space dimensionality.

Existence of a Stationary Solution

As an IRF-1, can Z have a stationary representation? Its covariance $K(h)$ has a spectral representation of the form (4.19) with $k = 1$. It can coincide with a spectral representation of a stationary covariance only if $\chi(du)/(4\pi^2|u|^2)^2$ has a finite integral. Now, in view of (8.19), the spectral measure $F(du)$ associated with the covariance $C(h)$ of Y is equal to $\chi(du)$, as was shown in Section 4.5.3.

Therefore Z has a stationary representation only if $F(du)/(4\pi^2|u|^2)^2$ is a spectral measure (with a finite integral), that is, if Y has very little low frequency content. This is not true in general.

Note that (8.19) can be solved directly in simple cases, notably in 1D or when the covariance is isotropic (cf. Dong, 1990, pp. 56–57).

Solution under Boundary Conditions

Let us restrict ourselves to a domain D with boundary Γ . Let Z and Y be two ordinary functions satisfying

$$\begin{cases} \Delta Z(x) = Y(x) & \forall x \in D \\ Z(x) = z(x) & \forall x \in \Gamma \end{cases}$$

The Green function (which only depends on the shape of the boundary¹¹⁾ makes it possible to express the real solution in the integral form

$$Z(x) = \int_D G(x, x') Y(x') dx' + \int_{\Gamma} \frac{d}{d\nu} G(x, x') z(x') dx' \quad (8.21)$$

where ν is the outward unit vector at the point x' of Γ . This amounts to decomposing Z into two functions Z_1 and Z_2 such that

$$\begin{cases} \Delta Z_1 = Y & \text{over } D \\ Z_1 = 0 & \text{over } \Gamma \end{cases} \quad \text{and} \quad \begin{cases} \Delta Z_2 = 0 & \text{over } D \\ Z_2 = z & \text{over } \Gamma \end{cases}$$

where Z_1 depends on the shape of the boundary but not the boundary values. These determine Z_2 . In view of the preceding stochastic framework, the solution of (8.18) is an IRF, namely a class of RFs, but knowledge of Z on Γ characterizes a specific representation.

Cokriging of Y and Z

The solution of the above systems is generally not trivial. Moreover in usual applications Y is not known at every point of D nor Z at every point of Γ . The stochastic model then achieves its full value. Once the generalized covariance $K(h)$ of Z is known, as well as its Laplacian and its iterated Laplacian, we have all the elements to perform a cokriging of Z (or of Y) from data of both variables. Z being an IRF-1, we must of course impose conditions ensuring that in the expression of the cokriging error the terms involving Z constitute an ALC-1. This requires that we have enough Z data (at least $n+2$ in n -dimensional space). If we have no Z data at all, we can still use cokriging provided that we work on an internal representation of the IRF-1 (recall that an internal representation is an ALC-1 at every point x ; see Section 4.4.3), but the result will depend on the selected internal representation, which is arbitrary.

Given the linearity of equation (8.18), the cokriged fields Y^{**} and Z^{**} satisfy the corresponding partial differential equation

$$\Delta Z^{**} = Y^{**}$$

In particular, at points x_α where measurements of Y are available, we have $\Delta Z^{**}(x_\alpha) = Y(x_\alpha)$. In the same manner we can reconstruct conditional cosimulations of Y and Z (at least in the Gaussian case) satisfying the Poisson equation.

8.4. INVERSE PROBLEM IN HYDROGEOLOGY

Real problems are generally more complex than just the Poisson equation. They involve several variables governed by physical laws represented by several partial differential equations. They include boundary conditions of different types, which are not always well known. Also the separation between known and unknown variables is not so clear-cut: the vector \mathbf{Y} is usually known only at a limited number of points, and on the other hand, some data are available on \mathbf{Z} . When variables are treated as deterministic, the partial differential system has an infinite number of solutions that can be vastly different, many of them devoid of physical sense. A geostatistical approach makes it possible to take account of spatial variability, thereby to restrict the space of possible solutions and in the end to express the set of solutions as a family of conditional simulations. There is no general theory on this subject. We will illustrate the possibilities of this approach by modeling a heterogeneous-permeability aquifer.

8.4.1. Introductory Example of One-Dimensional Flow

Let us come back to the one-dimensional flow governed by equation (8.5),

$$\frac{dH}{dx} = -\frac{q}{K(x)}$$

where the specific discharge q is prescribed. It is simpler to take as a parameter the hydraulic resistivity $W(x) = 1/K(x)$. We then have

$$\frac{dH}{dx} = -qW(x) \quad (8.22)$$

and therefore

$$H(x) = H(0) - q \int_0^x W(s) ds$$

This solves the *direct, or forward, problem*: Knowing permeability (or resistivity) at every point as well as boundary conditions (flow, head at one end of the column), find the expression of the head at every point.

In this special case, the backward problem, namely the determination of resistivity knowing the head *everywhere* and boundary conditions, is even simpler, since the solution follows immediately from (8.22): resistivity is equal,

up to a multiplicative factor, to head gradient

$$W(x) = -\frac{1}{q} \frac{dH}{dx}$$

However, this result is entirely theoretical. The calculation of resistivity at every point would require very closely spaced and precise head data, as even very small measurement errors could have a dramatic impact on the result. Typically this is a case of an ill-posed problem. With deterministic methods it is difficult to go beyond the determination of the mean resistivity. But, if point-support resistivity and/or head data are available, and if we work in a stochastic framework, we can at least solve the classic multivariate problem: Determine the conditional expectation, or at least the cokriging estimates of W or of H from available data, or carry out a conditional simulation. In all cases a prerequisite is the joint spatial distribution of W and H , or at least their direct and cross-covariances. They derive of course from the spatial distribution—or the covariance—of W , which plays the central role from a physical point of view.

W is, up to the $-q$ factor, the stochastic derivative of H . If W is an SRF with mean m_W and covariance C_W , H is an IRF-0 with a linear drift with slope $-qm_W$ and a variogram γ_H related to C_W as we have seen in Section 2.3.1 by

$$\frac{d^2\gamma_H(h)}{dh^2} = q^2 C_W(h) \quad (8.23)$$

Denoting by $F_W(du)$ the spectral measure of the covariance C_W and by $\chi_H(du)$ the measure involved in the spectral representation (2.21) of γ_H , this relation can be written equivalently as

$$\chi_H(du) = q^2 F_W(du)$$

Can H be regarded as the sum of a linear drift and a stationary residual? The answer is yes if the spectral measure $\chi_H(du)/(4\pi^2 u^2)$ of H is integrable, in other words, since χ_H and F_W are proportional, if

$$\int \frac{F_W(du)}{u^2} < \infty$$

This condition is not satisfied in the general case. It requires the amplitude of the large wavelength components of the spectrum of W to be very small. Given (8.23), such situation is characterized by the fact that the covariance C_W is equal to minus the second derivative of a twice-differentiable stationary covariance. Such a model is obtained, for example, by taking for C_H the Radon transform of order 2 of a covariogram and for C_W its second derivative (up to the $-q^2$ coefficient). The following model, obtained by

taking the Radon transform (2.52) of an exponential covariance, is classic in hydrogeology (Gelhar, 1993, p. 98):

$$C_W(h) = \sigma_W^2 \left(1 - \frac{|h|}{a} \right) \exp \left(-\frac{|h|}{a} \right)$$

$$C_H(h) = \sigma_H^2 \left(1 + \frac{|h|}{a} \right) \exp \left(-\frac{|h|}{a} \right) \quad \text{with} \quad \sigma_H^2 = q^2 a^2 \sigma_W^2$$

The covariance C_W is linear near the origin and exhibits a hole effect (except for a scale change this is identical to the covariance (2.74) of a migration process). The covariance C_H is parabolic near the origin. Note that the scale parameter a of W is now involved in the variance of H . Since cokriging weights are sensitive to the relative levels of the different covariances, a is an important parameter of the bivariate model and must be fitted carefully. The cross-covariance of H and W is in the general case

$$\text{Cov}(H(x+h) - H(x), W(y)) = -\frac{1}{q} \left[\frac{d\gamma_H}{dh}(y-x) - \frac{d\gamma_H}{dh}(y-x-h) \right]$$

and in the special foregoing case

$$C_{HW}(h) = \sigma_H \sigma_W \frac{|h|}{a} \exp \left(-\frac{|h|}{a} \right)$$

Note the different treatment of the two boundary conditions: the flow q is considered as a known deterministic constant, whereas $H(0)$ is a point value of the RF. Figure 8.3 presents the slightly different case of a lognormal resistivity whose logarithm has an exponential covariance; ϕ then has an unbounded variogram. For consistency with the next sections, the figure is presented in terms of log-permeability (minus the log-resistivity), which changes the sign of the cross-covariance.

Whether we work in a stationary setting or not, we have available all the elements to carry out a joint structural analysis of the two variables W (or $\log W$) and H from the data at hand, and proceed to cokriging or conditional simulation.

8.4.2. Position of the Problem for a Natural Aquifer

Objective

In steady-state as well as transient flow regimes, flow and transport in an aquifer or a reservoir derive in fact from properties of geological formations (permeability) and prescribed external conditions (boundary conditions, recharge by infiltration from rainfall, pumping). The direct problem, namely the calculation of piezometric head at every point when these parameters are

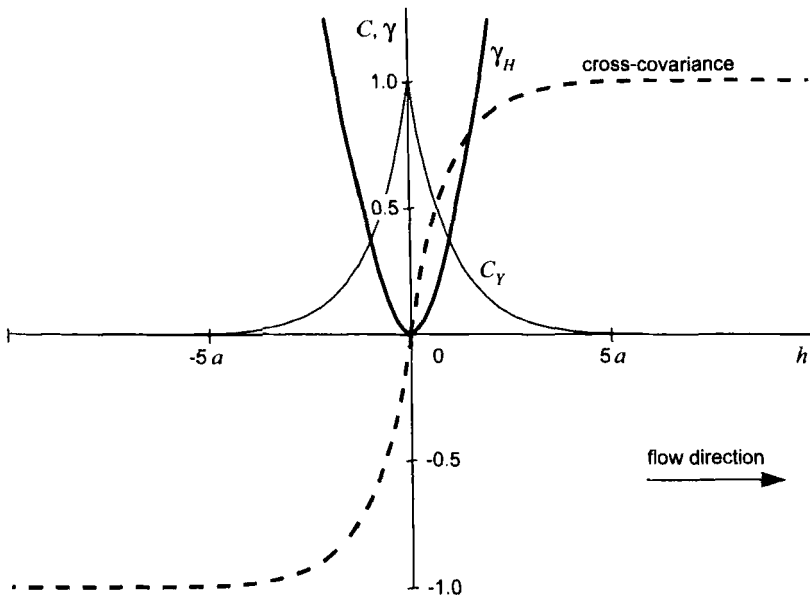


FIGURE 8.3. Covariance of log-permeability $Y \approx \log K$, variogram of head H , and cross-covariance of $Y(x)$ and $H(x+h) - H(x)$ in the one-dimensional case. The covariance of Y is assumed exponential with scale parameter a . Y is normalized by σ_Y , and H by $Ja\sigma_Y$, where J is the hydraulic gradient $-E[dH/dx]$. From Dagan (1989, Figs. 3.7.3 and 3.7.4), © Springer-Verlag.

known exhaustively, amounts to solving a partial differential equation. It is a well-posed problem in the sense that it has a solution, which is unique and stable: a perturbation of the permeability field has a very limited effect on the head.

In practice, the permeability field is known very sparsely, at those few places where pumping tests have been run. The main task of a hydrogeologic modeling exercise is therefore the reconstruction of the permeability field. If the piezometric head were known at every point, the problem would be similar to the preceding one, except that the data and the unknowns are swapped: Find the permeability field knowing piezometric head and external conditions. This time, however, the problem is ill-posed: even when existence and uniqueness of the solution is assured in a mathematical sense, the solution is unstable. To reproduce a slight perturbation of piezometric head, it is necessary to introduce a large variation of permeability; measurement errors, no matter how small, will generate completely unrealistic solutions (e.g., locally negative permeability values).

Having said this, while piezometry is easier to measure than permeability, it is also known at only a limited number of points. In real applications the problem is therefore very different: Reconstruct the permeability field on the basis of a limited number of head data, possibly complemented by permeabil-

ity data, and by more or less well-known boundary conditions. This is what we will call the *inverse problem*. It does not possess a unique solution.

A manual procedure consists in reducing the range of possible solutions by partitioning space into zones with a constant (but unknown) permeability, and fitting these few data iteratively so as to minimize the difference between calculated and observed heads. Such method obviously ignores spatial heterogeneity inside each zone. It does not permit an evaluation of the error on the piezometric head away from measurement points, and therefore also on velocities and flow paths.

Automating the foregoing procedure leads to so-called nonlinear methods, which are simply model calibration methods. They consist in finding, in discretized form, the permeability field minimizing a given penalty function. This comprises two terms, a discrepancy term measuring a squared distance between calculated and observed piezometry at observation points, and a roughness term designed to penalize solutions which are not smooth enough and to render the problem well-posed. The permeability model minimizing the penalty function is usually sought using the Gauss-Newton method or gradient-based methods. By and large this method has the same drawbacks as the manual method.

Neuman (1980), Cooley (1982, 1983), and Carrera and Neuman (1986a, b) developed a Bayesian approach similar to that of Tarantola and Valette (1982) and of Tarantola (1987) in geophysics. The region of interest being divided into blocks, the parameters to be determined include block permeabilities, inflow/outflow, boundary values, and factors allowing the expression of variances and covariances of "measurement errors" (to be understood in a wide sense). This parameter set is then estimated by a nonlinear least squares method or by maximum likelihood, on the basis of head data and of prior values for model parameters. In this approach parameters are all regarded as deterministic and the only source of randomness resides in measurement errors. The spatial structure of permeability is not used specifically, even if, following the introduction of geostatistical techniques in hydrogeology (Delhomme, 1976, 1978; de Marsily, 1986), some authors take prior values obtained by kriging (Clifton and Neuman, 1982).

We will focus here on methods where permeability is regarded as a realization of a random function, thus accounting for zoning as well as local heterogeneity. The objective of the inverse problem becomes the calculation of the conditional expectation of the permeability field—or more modestly its estimation by cokriging—or perhaps the construction of conditional simulations. Indeed the final objective is often a numerical simulation of flow and transport, both highly nonlinear processes that must be calculated from grids with the same degree of spatial heterogeneity as the real reservoir.

Working Assumptions

We will explicitly denote by (x,y,z) the coordinates of a point in 3D space. To simplify, we will consider an aquifer in which the flow is quasi-horizontal.

The piezometric head is then constant along the vertical, and by integrating vertically between the aquifer lower and upper elevations z_1 and z_2 Darcy's law takes the form

$$U(x, y) = -T(x, y) \text{grad} H(x, y) \quad (8.24)$$

where $U(x, y)$ is vertically integrated flow and $T(x, y)$ is transmissivity [$L^2 T^{-1}$] defined by

$$T(x, y) = \int_{z_1(x, y)}^{z_2(x, y)} K(x, y, z) dz$$

In the general case transmissivity is a tensor, but here we will assume it to be a scalar.

In steady state the flow satisfies the continuity (or mass-balance) equation stating that the fluid contained in an element of volume is constant

$$\text{div} U(x, y) = R(x, y) \quad (8.25)$$

where $R(x, y)$ is the inflow (recharge) rate [LT^{-1}], positive for injection, and negative for discharge.

The velocity field U being deduced directly from the transmissivity T and head H fields by application of Darcy's law, K , H , and R are taken as primary variables. Darcy's law (8.24) and the continuity equation (8.25) are then combined into the flow equation

$$\text{div}(T \text{grad} H) + R = 0 \quad (8.26)$$

The problem simplifies when either of these functions is known. In the case of a medium with a constant and known transmissivity T and a regionalized inflow R , equation (8.26) reduces to

$$T \Delta H = -R$$

namely up to a change of units, to the problem that we just treated. We will examine the more realistic case of regionalized transmissivity. Upon division by T , which is always nonzero, equation (8.26) writes

$$\frac{\partial \log T}{\partial x} \frac{\partial H}{\partial x} + \frac{\partial \log T}{\partial y} \frac{\partial H}{\partial y} + \Delta H + \frac{R}{T} = 0 \quad (8.27)$$

It is interesting to note that, in the absence of inflow, this equation involves $\log T$ —more precisely the gradient of $\log T$ —rather than T itself, and this regardless of whether or not transmissivity is lognormally distributed. The solution of the inverse problem associated with (8.27) is usually difficult because the equation is nonlinear and external conditions may be complicated.

We will first examine methods relying on a linearized approximation of the flow equation, then we will give a glimpse of iterative, nonlinear methods.

8.4.3. Perturbation Approximation to the Flow Equation

The inverse problem is greatly simplified if equation (8.27) can be replaced by an acceptable linear approximation. Indeed, in a linear setting, provided that the covariance models are consistent, there are proven methods for structural analysis, cokriging, and conditional simulations (often but not necessarily within a Gaussian model). Our main problem is thus to propose direct and cross-covariance models of transmissivity and head that are physically and mathematically consistent. As transmissivity is the basic physical property, we are back to a well-posed direct problem (which does not mean an easy problem): Find the covariance of head and the cross-covariance of transmissivity and head knowing the covariance of transmissivity and external conditions.

Linearization of the Flow Equation

The principle of the method is to decompose transmissivity (or its logarithm) and head into a deterministic mean and a random zero-mean fluctuation, called perturbation, and look for an approximation to the flow equation that is linear in the perturbations. Such attempt is only valid of course for small perturbations, a notion that will be quantified later.

The main role being played by $\log T$ rather than T , let us start from equation (8.27). We define

$$\log T(x, y) = \log T_g + Y(x, y)$$

where $\log T_g = E[\log T(x, y)]$ defines the geometric average of transmissivity and where the perturbation $Y(x, y)$ is a zero-mean random function. Likewise the head is decomposed as

$$H(x, y) = h(x, y) + \phi(x, y)$$

where the drift $h(x, y) = E[H(x, y)]$ is a deterministic function and the perturbation $\phi(x, y)$ a zero-mean random function. Equation (8.27) then becomes

$$\frac{\partial Y}{\partial x} \left(\frac{\partial h}{\partial x} + \frac{\partial \phi}{\partial x} \right) + \frac{\partial Y}{\partial y} \left(\frac{\partial h}{\partial y} + \frac{\partial \phi}{\partial y} \right) + \Delta h + \Delta \phi + \frac{R}{T_g} e^{-\gamma} = 0 \quad (8.28)$$

Analysis in the Absence of Inflow/Outflow

Let us first consider the simple case of no inflow/outflow. The equation simplifies to

$$\frac{\partial Y}{\partial x} \left(\frac{\partial h}{\partial x} + \frac{\partial \phi}{\partial x} \right) + \frac{\partial Y}{\partial y} \left(\frac{\partial h}{\partial y} + \frac{\partial \phi}{\partial y} \right) + \Delta h + \Delta \phi = 0$$

Taking mathematical expectation we get

$$E \left[\frac{\partial Y}{\partial x} \frac{\partial \phi}{\partial x} + \frac{\partial Y}{\partial y} \frac{\partial \phi}{\partial y} \right] + \Delta h = 0 \quad (8.29)$$

Subtracting this expression from the preceding one yields

$$\frac{\partial Y}{\partial x} \frac{\partial h}{\partial x} + \frac{\partial Y}{\partial y} \frac{\partial h}{\partial y} + \frac{\partial Y}{\partial x} \frac{\partial \phi}{\partial x} + \frac{\partial Y}{\partial y} \frac{\partial \phi}{\partial y} - E \left[\frac{\partial Y}{\partial x} \frac{\partial \phi}{\partial x} + \frac{\partial Y}{\partial y} \frac{\partial \phi}{\partial y} \right] + \Delta \phi = 0 \quad (8.30)$$

Head perturbations are produced by log-transmissivity perturbations and are even smaller. Therefore terms involving products of Y and ϕ perturbations will be very small and can be neglected (on the influence of this approximation refer to Loaiciga and Marino, 1990). Equations (8.29) and (8.30) simplify to

$$\Delta h = 0 \quad (8.31)$$

$$\frac{\partial Y}{\partial x} \frac{\partial h}{\partial x} + \frac{\partial Y}{\partial y} \frac{\partial h}{\partial y} + \Delta \phi = 0 \quad (8.32)$$

It is seen that T_g is not involved in these equations and therefore does not need to be known. The first equation is the ordinary (nonstochastic) Laplace equation. Its solution is a harmonic function and is uniquely defined if boundary conditions are fixed, which we assume. The second equation is a stochastic equation between perturbations. It permits the determination of ϕ as a function of Y . Note that Y is involved only through its gradient.

In some simple cases it is possible to solve these equations analytically and to derive direct and cross-covariances of Y and ϕ consistent with the flow equations. The first such attempt is due to Mizell et al. (1982), and it has been reworked in a broader context by Dong (1990). The global flow is assumed unidirectional and boundary conditions are sent to infinity (i.e., boundaries are so remote that their influence is not felt). Perhaps after a change of coordinate system, we will assume that head is a function of the x coordinate only and that it decreases when x increases. From (8.31) it follows immediately that the drift h is linear in x

$$h(x, y) = h_0 - Jx$$

where $J > 0$ is the hydraulic gradient. The stochastic equation between perturbations becomes

$$\Delta \phi = J \frac{\partial Y}{\partial x} \quad (8.33)$$

Let us introduce the RF Z solution of the partial differential equation

$$\Delta Z = Y$$

The solution of (8.33) is then the RF,

$$\phi = J \frac{\partial Z}{\partial x} \quad (8.34)$$

The first equation has been discussed in Section 8.3.2. If Y is an IRF-0 (T can then be regarded as locally stationary), this equation defines a unique twice-differentiable IRF-2 Z . The second equation then defines a unique differentiable IRF-1 ϕ . Since h is linear, $H = h + \phi$ is an IRF-1.

In applications the variogram of Y is often linear near the origin, which contradicts the differentiability assumption expressed by (8.33). We get around this difficulty by regarding Y as a generalized random function. This is all the more justified as the solution Z can be written as (8.21) (where x is a point and not the abscissa) so that ϕ is expressed as an integral of Y and does not involve its derivatives.

As for the relationships between the various covariances, let us consider for simplicity the case where Y (the log T perturbation) has a stationary covariance $C(h_x, h_y)$, h_x and h_y being the two components of the separation vector usually denoted by h (in the present section h is reserved for the head drift). Then Z is an IRF-1 whose generalized covariance $K(h_x, h_y)$ is solution to (8.19), and the variogram of its derivative, the head perturbation ϕ (an IRF-0) is

$$\gamma(h_x, h_y) = J^2 \frac{\partial^2 K}{\partial h_x^2}(h_x, h_y) \quad (8.35)$$

It depends on the direction of the vector (h_x, h_y) with respect to the general direction of flow. This anisotropy manifests itself by a more structured behavior (e.g., a longer range) in the direction orthogonal to the mean flow direction than along the mean flow direction (assuming an isotropic covariance for Y). Furthermore, if the covariance of Y behaves near the origin like r^α , where r is the modulus of the vector (h_x, h_y) , γ behaves like $r^{\alpha+2}$ in all directions and is therefore more regular. At large distance, $\gamma(h_x, h_y)$ is generally unbounded.

If the covariance C is isotropic, γ has a logarithmic behavior. There are exceptions for which H can be regarded as the sum of a linear drift and an SRF. In the isotropic case, a necessary and sufficient condition to be in this situation is that the integral range of the covariance C is equal to zero (Mizell et al., 1982). These results are particular to the 2D case; in 3D, if C is isotropic, H can always be regarded as the sum of a linear drift and an SRF (Gutjahr and Gelhar, 1981).

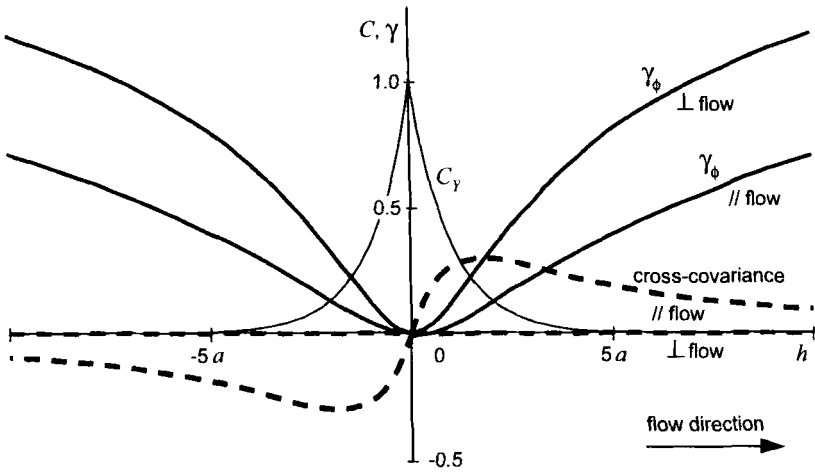


FIGURE 8.4. Covariance of $Y = \log K$, variogram of head perturbation ϕ , and cross-covariance of $Y(x)$ and $\phi(x+h) - \phi(x)$ in the two-dimensional case. The covariance of $\log K$ is assumed isotropic and exponential with scale parameter a . Y is normalized by σ_Y , and ϕ by $Ja\sigma_Y$. The global flow is considered unidirectional in an infinite aquifer. From Dagan (1989, Figs. 3.7.3 and 3.7.4), © Springer-Verlag.

Similarly the cross-covariance of Y and Z being given by (8.20), the cross-covariance between Y and a representation of ϕ is

$$E[Y(x, y)\phi(x + h_x, y + h_y)] = -J \frac{\partial}{\partial h_x}(\Delta K(h_x, h_y)) + J \Delta c_1(x, y) \quad (8.36)$$

Since ϕ is an IRF-0, this expression is used only to calculate the covariance of Y and an increment of ϕ . The second term then cancels.

One can note that the regional hydraulic gradient is involved linearly in the cross-covariance (8.36) and as a square in the head variogram (8.35). This means that in a cokriging of transmissivity, the weights on head data will be inversely proportional to J . Knowledge of this parameter is therefore essential. Figure 8.4 presents the case where Y is an SRF with an exponential isotropic covariance (the explicit formulas are given by Dagan, 1989, sec. 3.7). ϕ displays much less variability than in 1D (Fig. 8.3).

As system (8.32) relies on a first-order approximation, its solution is valid only if the variance of Y is small enough. Comparisons with the exact solution in 1D (Gutjahr et al., 1978), with a second-order approximation in 1D and 3D (Dagan, 1985b), and with results obtained from 3D simulations (Ababou et al., 1989) show that the first-order approximation produces an overestimation of the variance of ϕ of the order of 6% only for $\sigma_Y^2 = 1$, which is an order of magnitude observed in many aquifers. This overestimation remains moderate up to $\sigma_Y^2 = 5$ (it is about 50% in 1D and 30% in 3D). For more details, the

reader is referred to Gelhar (1993, ch. 4). This model seems to be very robust: Zimmerman et al. (1998) present field tests that have led to acceptable results for even higher variances, exceeding 10.

Analysis with Inflow/Outflow

We now consider the influence of an inflow/outflow term. This can be due to rainfall, to exchanges with neighboring aquifers or a stream, to pumping, and so on. It can assume different forms, and in particular the term may depend on the value of H . We will restrict the discussion to the case where R is due only to natural recharge by infiltration of rainfall. Equation (8.28) shows that in fact it is the ratio R/T_g which is involved. Therefore, if we have no information on either recharge or average transmissivity, we will be able to determine neither. Here we will assume that T_g is known, and also that the recharge function $R(x, y)$ is known (it depends mainly on meteorological parameters and soil type).

Taking mathematical expectation in (8.28) now gives

$$\Delta h = -\frac{M_0}{T_g} R \quad (8.37)$$

where $M_0 = E[e^{-Y}]$ (in particular $M_0 = \exp(\frac{1}{2}\sigma^2)$ if the marginal distribution of Y is Gaussian with variance σ^2). If boundary conditions are known this equation allows an explicit determination of h . Neglecting products of perturbations and taking (8.37) into account, we get

$$\frac{\partial Y}{\partial x} \frac{\partial h}{\partial x} + \frac{\partial Y}{\partial y} \frac{\partial h}{\partial y} + \left(1 - \frac{e^{-Y}}{M_0}\right) \Delta h + \Delta \phi = 0 \quad (8.38)$$

Equation (8.37) clearly shows that the regional flow h is no longer linear. Under conditions similar to those examined before (unidirectional regional flow, boundary conditions at infinity), and with a constant nonzero inflow, h is now quadratic. More generally, h depends both on boundary conditions and on recharge. Though the influence of recharge does not appear directly in (8.38), it exerts an influence through the form of the mean head h .

Equation (8.38) can be solved numerically but is analytically cumbersome due to the presence of an exponential term. Two approaches can be used to obtain a simplified equation, which both amount to taking a transmissivity model of the type

$$T(x, y) = T_0 [1 + Y(x, y)] \quad (8.39)$$

The first approach is to replace e^{-Y} by $1 - Y$ in (8.28). Formula (8.39) should then be regarded as a first-order approximation to T , where T_0 is the *geometric* average of T and Y the $\log T$ perturbation.

The second approach is to start from the flow equation (8.26) and take (8.39) as a definition of a decomposition of T (and not of its logarithm) into a mean value and a relative perturbation. (8.39) is then, by definition, an exact formula, where T_0 is the *arithmetic* average of transmissivity and Y a zero-mean perturbation which we do not care to relate to $\log T$.

In both cases, after linearization of the flow equation to the first order, we get for the drift term an equation similar to (8.37)

$$\Delta h = -\frac{R}{T_0} \quad (8.40)$$

and for perturbations the stochastic equation

$$\frac{\partial Y}{\partial x} \frac{\partial h}{\partial x} + \frac{\partial Y}{\partial y} \frac{\partial h}{\partial y} + Y \Delta h + \Delta \phi = 0 \quad (8.41)$$

or equivalently

$$\Delta \phi = -\text{div}[Y \text{ grad } h]$$

Model (8.39) is not compatible with a high variability of Y , which could indeed cause the occurrence of negative values of T . It assumes that the variance of $\log T$, or which is about the same, the relative variance of T , is small (to be specific, say $\sigma^2 \leq \frac{1}{2}$). The introduction of the recharge term therefore limits the robustness of the model.

8.4.4. Accounting for Boundary Conditions

Boundary conditions can be accounted for in several ways. In a first step, to simplify the presentation of the various methods, we will ignore scale effects and consider that data are obtained at the scale at which the equations are solved (points or grid cell).

Different Types of Boundary Conditions

The aquifer being limited to a domain D with boundary Γ , we will consider three classic types of boundary conditions:

- *Dirichlet conditions.* Values of the head H are prescribed at boundary points.
- *Neumann conditions.* Values of the normal derivative are prescribed on the boundary. Specifically, let ν denote the outward unit vector at a point x of Γ and denote by $\partial H / \partial \nu$ the derivative of H along ν . Neumann conditions specify values of $\partial H / \partial \nu$, which amounts to prescribing values of the flow. For example, a no-flow boundary is characterized by $\partial H / \partial \nu = 0$ at every point of Γ .

- *Fourier conditions.* These are mixed conditions where values of $H + \alpha(\partial H/\partial \nu)$ are prescribed for a given value of α .

When using a perturbation method, it is considered that boundary conditions concern the drift h and that the perturbation satisfies similar boundary conditions but with a zero value of ϕ (Dirichlet conditions) or of its gradient (Neumann conditions).

Boundary Conditions Included in the Data

A first method to take boundary conditions into account is to regard them as additional data. The RF model is that of the preceding section. Neumann conditions are treated as gradient data as was shown in Section 5.5.2. We have seen an application of this approach by J. P. Delhomme where impermeable boundaries are accounted for by zero increments of piezometric head in the direction orthogonal to the boundary (Section 5.5.4). Boundary conditions with prescribed head are handled simply by discretizing the boundary and assigning points the given head values.

This approach is very flexible. Boundary conditions may be introduced where they are known and left unspecified when they are unknown, which is a common case. The approach can be generalized to include uncertainty (introduction of variances, or even covariances, similar to those of measurement errors), or even uncertainty on the positioning of boundaries.

Boundary Conditions Included in the Model

Physically, boundary conditions are often external data. If, for example, a river or a lake is in contact with the aquifer, the piezometric head along the aquifer-river contact is equal to the water level in the river. It is more satisfactory to explicitly include this water level in the model rather than regard it merely as piezometry information. The (generalized) head covariance is no longer stationary but depends separately on the two supporting points considered.

Dagan (1985a) and Rubin and Dagan (1987) propose the following approach for a random but spatially constant recharge. If boundary conditions are known the solution of (8.41) can be written in terms of the Green function G associated with the boundaries of the domain D ,

$$\phi(x, y) = - \int_D \int \operatorname{div}[Y(x', y') \operatorname{grad} h(x', y')] G(x, y, x', y') dx' dy'$$

Since this expression is a linear functional of Y , the nonstationary covariance of ϕ as well as the nonstationary cross-covariance of Y and ϕ can be deduced from the covariance or variogram of Y . But, in practice, the explicit form of the Green function is not known, except in special cases such as boundary conditions at infinity (the case finally treated by the authors).

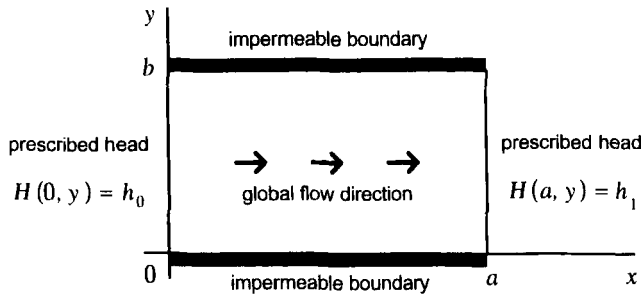


FIGURE 8.5. Rectangular domain and prescribed boundary conditions. From Roth et al. (1997), with kind permission from Kluwer Academic Publishers.

Analytic Covariance Model for a Rectangular Aquifer

Equation (8.41) can be solved directly for a rectangular aquifer and appropriate boundary conditions following a method developed by Roth (1995). The domain D is the rectangle $a \times b$ whose lower left vertex coincides with the origin of coordinates, and the boundary conditions comprise two opposite impermeable sides, namely no-flow boundaries, the other two sides being at constant prescribed head:

- Prescribed head boundaries

$$H(0, y) = h_0 \quad H(a, y) = h_1 \quad \forall y \in [0, b]$$

- No-flow boundaries: (8.42)

$$\left. \frac{\partial H(x, y)}{\partial y} \right|_{y=0} = \left. \frac{\partial H(x, y)}{\partial y} \right|_{y=b} = 0 \quad \forall x \in [0, a]$$

where h_0 and h_1 are given (Fig. 8.5). Passing to mathematical expectation, it is seen that these conditions must be satisfied by the deterministic part $h(x, y)$. Hence the head perturbation ϕ must satisfy the following boundary conditions:

- Prescribed head boundaries:

$$\phi(0, y) = \phi(a, y) = 0 \quad \forall y \in [0, b]$$

- No-flow boundaries: (8.43)

$$\left. \frac{\partial \phi(x, y)}{\partial y} \right|_{y=0} = \left. \frac{\partial \phi(x, y)}{\partial y} \right|_{y=b} = 0 \quad \forall x \in [0, a]$$

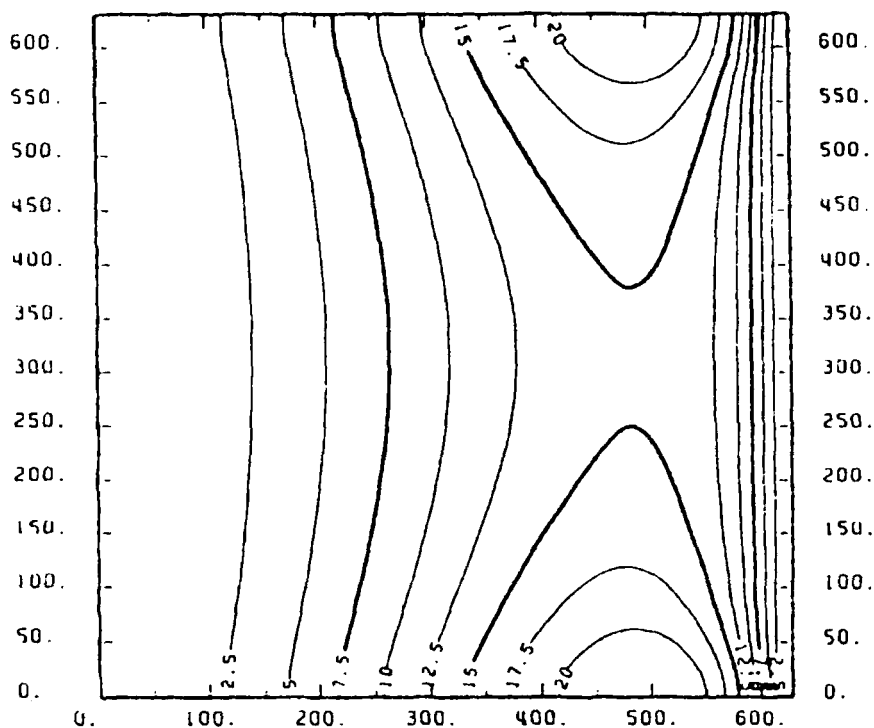


FIGURE 8.6. Nonstationary variance of head perturbation. The domain of interest and the prescribed boundary conditions are those of Figure 8.5. The log-transmissivity has a spherical covariance whose range is one tenth of the size of the domain. The recharge by rainfall infiltration is 70 mm per year. From Roth (1995).

In order to take into account the conditions at the boundaries of the rectangle, the functions Y and ϕ are periodically repeated so that the rectangle $a \times b$ corresponds to the half-period in each direction. The recharge R , assumed known is also periodically repeated so that (8.41) is valid in the whole plane. Periodic functions can be represented by Fourier series which, when considering symmetries and antisymmetries required by the continuity of Y and ϕ and the boundary conditions, amounts to a sine and cosine series. Since Y and ϕ are regarded as random functions, their Fourier coefficients are random variables. The linearized flow equation (8.41) determines the link between the coefficients of the head perturbation and those of the transmissivity perturbation, and therefore the link between their covariances. The covariance of ϕ and the cross-covariance of Y and ϕ are thus deduced analytically from the covariance of Y (Roth, 1995; Roth et al., 1997).

If we select a stationary and isotropic variogram for Y , that of ϕ is neither isotropic nor even stationary because of the influence of boundary conditions. Figure 8.6 shows the variance of $\phi(x, y)$ as a function of location in a situa-

tion inspired by a real aquifer. This variance is larger in the vicinity of the no-flow boundaries than in the center of the domain, and of course exhibits lower values near the prescribed-head boundaries. It is also larger downstream than upstream, under the effect of increasing recharge—if infiltration is uniform on the surface the aquifer transports an increasing amount of water along the direction of flow. The interpretation of the covariance of ϕ and the cross-covariance of Y and ϕ is of course more complex than in the case without boundary conditions but remains possible (to limit the effect of the nonstationarity of ϕ , variograms or correlograms are considered in strips parallel to one of the axes). This model is of course only valid when the linearization of the flow equation is justified, that is, for a moderate variance of transmissivity perturbations.

Analytic Covariance Model for an Aquifer with Arbitrary Shape

The preceding method may be extended to an arbitrary aquifer whose boundary conditions are made of two no-flow boundaries surrounding prescribed-head boundaries. Indeed, using a *conformal transformation* (e.g., Schneebeli, 1966; Polubarinova-Kochina, 1962) one can transform the natural domain D into a rectangle with boundary conditions of the type (8.42). A conformal transformation preserves angles but not areas; it preserves the steady-state flow equation provided that recharge at every point is assigned a local multiplicative factor ensuring the preservation of the quantity of water infiltrating into the aquifer. The transmissivity covariance, stationary in natural space, is transformed into a nonstationary covariance over the rectangle. From this one derives the head-perturbation covariance over the rectangle, then in natural space by application of the inverse transformation (Roth, 1995; Roth et al., 1997).

There is generally no analytic expression for the conformal transformation of a natural domain into a rectangle with similar boundary conditions. Roth carries out the transformation using a numerical method proposed by G. Matheron.

This method is in reality limited to aquifers whose shape remains relatively close to a rectangle. Otherwise, the deformation is extreme. Transmissivity, while stationary in the initial aquifer, exhibits an extremely rapid spatial variability in the transformed space, which precludes a sufficiently precise numerical calculation of its Fourier expansion over the rectangle.

Covariance Model Obtained by Discretization of the Flow Equation

When the global flow is not unidirectional or when recharge is not constant, the solution of (8.41) is more difficult since the regional piezometric head h is no longer polynomial. Kitanidis and Vomvoris (1983) in 1D then Hoeksema and Kitanidis (1984) in natural space propose a finite differencing approach similar to that of Dettinger and Wilson (1981). The global flow $h(x,y)$ is obtained analytically or numerically by solving equation (8.40), then equation

(8.41) is discretized as

$$\begin{aligned} & \frac{1}{2a} \frac{\partial h}{\partial x}(i, j)[Y(i+1, j) - Y(i-1, j)] \\ & + \frac{1}{2a} \frac{\partial h}{\partial y}(i, j)[Y(i, j+1) - Y(i, j-1)] - \Delta h(i, j)Y(i, j) \\ & + \frac{1}{a^2}[\phi(i+1, j) + \phi(i, j+1) + \phi(i-1, j) + \phi(i, j-1) - 4\phi(i, j)] = 0 \end{aligned}$$

where a is the side of the square blocks discretizing the studied domain. $Y(i, j)$ represents the perturbation of block (i, j) log-transmissivity, while partial derivatives of h and perturbation $\phi(i, j)$ are averages over block (i, j) . Considering the direct problem first, the set of $Y(i, j)$ is grouped in a vector \mathbf{Y} , the $\phi(i, j)$ of blocks located at the boundary in a vector Φ_B , and the $\phi(i, j)$ of the other blocks in a vector Φ . The grouping of the discretized equations of all blocks leads to a system of the form

$$\mathbf{A}\Phi = \mathbf{B}\mathbf{Y} + \mathbf{C}\Phi_B$$

where \mathbf{A} , \mathbf{B} , and \mathbf{C} are known constant matrices. For any given covariance of the RF Y , it is possible to derive the direct and cross covariance matrix of \mathbf{Y} and Φ conditionally on Φ_B .

Numerical Covariance Model

Another approach, initiated by Smith and Freeze (1979a, b) for variances, is numerical estimation of the covariance model, whose principle is as follows: For a given model of transmissivity covariance, M independent (non-conditional) simulations are generated, and for each one the flow equation is solved numerically using some finite difference or finite element code to obtain the corresponding head field. Then all required covariance values are computed between variable Z_i at point number α and variable Z_j at point β from M pairs of simulated values $Z_i(x_\alpha, y_\alpha)$ and $Z_j(x_\beta, y_\beta)$ ($i = 1$ for transmissivity, 2 for head). The covariances obtained are empirical ones, but since they are computed as means over several realizations, they lead, by design, to positive definite matrices and may be used directly without modeling. The cokriging estimates are consistent, and computed errors variances are non-negative.

The quality of the estimation obtained with this method depends on the convergence of the numerical covariance toward its theoretical value as N increases. The numerical covariance of transmissivity converges like $1/\sqrt{N}$, but the head covariance converges more slowly. Tests have been carried out on a lognormal transmissivity field with a range of $1/7$ of the domain size and show that several hundreds, if not thousands of simulations, are required to see

the numerical values of head covariance stabilize. However, using covariances derived from only a few hundred simulations, the cokriging of transmissivity from transmissivity and head data produces already far better results than kriging from transmissivity data alone using the theoretical variogram (Roth, 1995).

Of course this method is computationally intensive, but it is extremely versatile, since the numerical solution of the flow equation by a finite difference code does not require linearization of the flow equation and thus implies no restriction on the geometry of the aquifer, on the configuration of boundary conditions, nor even on the amplitude of transmissivity variations.

Accounting for Scale Effects

So far we have ignored the scale effect problem by assuming that all variables are relative to the same scale, either a point in the case of analytic methods, or a grid block the case of a discrete method. The problem is not critical for the head which is a very continuous variable and can be interpolated safely. Thus the mean head over a grid block does not differ much from the head value at the center of the block provided that the grid mesh is not too large.

The problem is more serious for transmissivity. In the Kitanidis and Vomvoris (1983) and Hoeksema and Kitanidis (1984) approach, transmissivities involved in the finite difference equation are block transmissivities. When considering log-transmissivities in the lognormal model, block values are arithmetic averages of point values within the block. The change of scale is thus reduced to a classic regularization problem. Assuming that the point support is that of transmissivity data, we have all the elements to calculate, from the covariance of Y , all required direct and cross-covariances and perform a cokriging estimation of transmissivity or head.

The situation is similar in the case of the numerical covariance method, where the computation is carried out at the scale of macroscopic blocks from inter-block transmissivities. In the lognormal case we have seen in Section 8.2.2 that these are expressed to an excellent approximation as geometric averages of point-support transmissivities, which translates into arithmetic averages of log-transmissivities. In the general case the link between point-support transmissivities and inter-block transmissivities is more complex, but by applying one of the approaches proposed in Section 8.2.2, we are still able to simulate Y on a fine-mesh grid, to derive inter-block transmissivities, to calculate the head numerically for the relevant boundary conditions, and therefore to calculate numerically all the covariances between point-support transmissivities, inter-block transmissivities, and piezometric head required for cokriging.

We have assumed that transmissivities are point-support data. In practice, they derive from an interpretation of pumping tests and represent an equivalent transmissivity at the scale of the zone of influence of the test. By addressing the corresponding upscaling problem like for block transmissivities all covariances involving these data can be calculated.

8.4.5. Iterative Methods for Solving the Inverse Problem

At the beginning of Section 8.4.2 we mentioned the procedure of dividing the domain of interest into zones and calibrating the permeabilities of the different zones by trial and error. This is an iterative but completely deterministic approach to the inverse problem.

One of the earliest geostatistical approach to the inverse problem is due to Delhomme (1979a). A large number of transmissivity simulations is generated conditional on transmissivity data alone, the piezometric-head field is calculated for each of them, and only simulations that are compatible with piezometry data are retained. This method works when only a few piezometric observations are available, but in general, it leads to the rejection of most simulations.

The *method of pilot points* (also called *master points*), developed first in the scope of cokriging (de Marsily et al., 1984) and then extended to conditional simulations (RamaRao et al., 1995), may be considered as the combination of the two preceding approaches. Conditional simulations of transmissivity are generated from transmissivity data plus synthetic data, the pilot points, calibrated to account for the other observations. These fictitious data are strategically located where they carry most information about the head field. They are assigned values selected so as to “explain” piezometry data, that is, to minimize an objective function (a weighted sum of squared differences, at observation points, between calculated and observed head). Coupled kriging and adjoint sensitivity analysis is employed for the optimal location of pilot points, and gradient search methods are used to derive their optimal transmissivities. The calibrated fields honor the transmissivity measurements, preserve the variogram, and match the measured piezometric heads in a least square sense. This method is computer-intensive but is not limited by the variability of transmissivity since it is not required to linearize the flow equation. It permits the incorporation of measurements performed either in transient as in steady-state flow regime. The *sequential self-calibration method* of Gómez-Hernández et al. (1997) may be considered as a simplified variant of the pilot point method.

8.4.6. Implementation

Structural Analysis

While the selection of a model suited to the aquifer type, its spatial variability, and the problem at hand is important, a correct fitting of the parameters is also essential: a comparison of seven independent geostatistical studies of a set of four aquifers (Zimmerman et al., 1998) has shown that the quality of the results depends on both the selection of the best method for the aquifer of interest and a careful variographic analysis.

Clearly the objective of structural analysis is to determine the covariance or the variogram of transmissivity or of its logarithm, since the other covariances

(generalized or not, stationary or not) are deduced from it. In principle, it suffices to model the sample transmissivity variogram. But, in practice, transmissivity data are scarce, they represent an ill-defined scale (the influence zone of a pumping test is known only approximately), and they are subjected to large measurement errors (transmissivity is computed by calibration of the drawdown curve observed during a pumping test, which calibration leaves some freedom in the choice of parameters). Conversely, piezometry data are usually numerous and reliable. It is therefore essential to make use of all data.

The ideal is to fit simultaneously the variogram of Y and that of ϕ , as well as the cross-variograms of these two variables. Since these variograms are nonstationary, except that of Y , the sample variogram is computed as usual from pairs of points separated by approximately the same distance. The theoretical variogram is then computed by averaging nonstationary variograms for the corresponding pairs.

If there are enough data, a manual fit is advisable. However, since this is performed by trial and error and since each new choice of a model for the transmissivity variogram requires one to redo the calculation for the other variograms, this approach can be very time-consuming. To simplify the problem, Kitanidis and Vomvoris (1983) start with the analytic form of the covariance of Y , specifically an exponential covariance, and fit its parameters (range and sill) by the maximum likelihood method presented in Section 2.6.2. A similar approach with other models is used by Hoeksema and Kitanidis (1985) and Dagan (1985a). Considering the preceding remarks on sensitivity, it seems desirable to check that the model obtained produces an acceptable fit of the different sample variograms.

The pilot point method incorporates piezometric-head data through a calibration procedure that does not really make use of the spatial information in this variable. The "character" of the solution thus relies heavily on the transmissivity variogram.

Cokriging or Conditional Simulation of Transmissivity

In the scope of linear methods, structural analysis provides the direct and cross-covariances of transmissivity and head. It is then possible to reconstruct the transmissivity and piezometric-head fields by cokriging from all the data (Hoeksema and Kitanidis, 1984). It is also possible to construct a conditional simulation of transmissivity (Gutjahr et al., 1994). The conditional simulation of head can be obtained simultaneously or be computed from that of transmissivity by solving the direct deterministic problem. In practice, several simulations are generated that account for the nonuniqueness of the solution to the inverse problem within the stochastic model considered. This makes it possible to derive simulations of the filtration velocity field as well as simulations of transport.

We have seen that the logarithm of transmissivity is introduced naturally in equation (8.27), independently of the assumed or not lognormal character

of transmissivity. Clearly, if the spatial distribution of transmissivity can be considered lognormal, we are in a very favorable case. On the one hand, to a good approximation, the scale effect reduces to arithmetic averaging of $\log T$, and on the other, the operational variables $\log T$ and H are Gaussian. Cokriging then coincides with conditional expectation, and the construction of conditional simulations matching the spatial distribution is easy. In the opposite case the approach remains valid, but the modeling of the scale effect and the conditional simulation method must be adapted to the type of random function.

Consistency of Simulation Results

Whichever method is used, we wish it to produce transmissivity and head fields that are consistent with the spatial structure of the aquifer, with the flow equation, and with the complete data set. The methods just presented differ by their treatment of the flow equation and the boundary conditions. All second-order methods use a coherent model of direct and cross-covariances. This guarantees, in particular, that the fields $T^{**}(x,y)$ (or $Y^{**}(x,y)$) and $H^{**}(x,y)$ obtained by cokriging from transmissivity and head data satisfy themselves the flow equation *if this is linear* (i.e., in the scope of a linear approximation to the flow equation). An iterative method based on cokriging, and ensuring that the transmissivity and head estimates satisfy the flow equation when the linear approximation is not valid, has been proposed by Yeh et al. (1995).

Concerning conditional simulations, three cases may be distinguished:

1. The linear approximation is excellent and transmissivity may be regarded as a lognormal RF. Under (multi) Gaussian theory one can generate joint simulations of T and H that match the joint spatial distribution of T and H and satisfy the flow equation.
2. The linear approximation is justified but a Gaussian model is not applicable. It is possible to generate joint conditional simulations reproducing the covariances of T and H and satisfying the flow equation. The construction of simulations matching the complete spatial distribution and the flow equation is possible only in special cases (the RF model must be one for which conditional simulation is possible, as seen in Chapter 7, and must also allow treatment of scale effects, two nontrivial conditions). A typical example, more often observed in 3D than in 2D, is a medium composed of several rock types (e.g., sand and shale), each with a constant permeability. The RF then directly derives from an indicator or categorical variable, for which upscaling rules and simulation methods are known.
3. The linear approximation is not valid. One must then resort to numerical methods, such as cokriging with numerical covariance models, or pilot points optimization. The latter ensures by design that the flow equation

is satisfied and preserves, at least to a first approximation, the structure of the transmissivity field. However, it relies on a structural analysis of transmissivity data alone. If one wishes to obtain transmissivity and head fields that satisfy the flow equation when the joint simulation method doesn't, one can simulate transmissivity alone conditional on all available data and compute numerically the head field corresponding to each simulation. The computed fields do not go exactly through the head data, but the differences are expected to be small, and if not, indicate that the model needs to be revised.

Transport Modeling

From transmissivity and head simulations one can deduce simulations of filtration velocity and simulations of transport, for example, in view of calculating the possible travel paths of a particle of pollutant from a source and the required travel time to reach a protected zone (lake, catchment, etc.). The results obtained for the different simulations can be summarized by the probability distribution of travel time (which can sometimes be obtained analytically; e.g., see Rubin and Dagan, 1992). The dispersion of this distribution reflects the fragmentary character of our knowledge of the aquifer. To make this as small as possible, it is important to make use of all available data, as demonstrated by several authors. LaVenue et al. (1995), for example, have investigated the transport of a particle from the center of a waste storage site to its periphery. Table 8.1 summarizes the location and dispersion characteristics of the travel time probability distribution according to the conditioning data: non-conditional simulation, simulation conditional on transmissivity data alone, and conditional also on piezometry data measured in steady-state and transient flows. First, one can see that conditioning by transmissivity data has the effect of drastically increasing travel time, the reason being that the "drop point" is located in a low-transmissivity zone. Second, it is seen that conditioning on piezometry data now reduces travel time because these data indicate the presence of a high-transmissivity zone downstream where few direct transmissivity data are available. The use of piezometry data also shrinks the 95% confidence interval and, in particular, lowers its upper bound considerably (note that the successive confidence intervals are not nested; there is no theoretical reason why they should be).

8.4.7. Extensions

Real aquifers, problems, and models are of course more varied than those we have presented. Hoeksema and Kitanidis (1985) take into account a drift of transmissivity; Townley and Wilson (1985) include uncertainty on boundary conditions, and the like. The monographs by Dagan (1989) and Gelhar (1993) consider other types of aquifers and of recharge than we have examined and give numerous references. They also consider the inclusion of transient-flow

TABLE 8.1. Evolution of the probability distribution of travel time (years) with conditioning data

Simulation Type	Nonconditional	Conditional on <i>T</i> Data	Conditional on <i>T</i> and Steady-State <i>H</i> Data	Conditional on <i>T</i> , Steady-State and Transient <i>H</i> Data
Median	35,000	175,000	22,000	16,000
95% Confidence Interval	10,600 to 205,000	26,500 to 1,150,000	11,100 to 43,500	9280 to 27,600

Source: LaVenue et al. (1995), © American Geophysical Union.

observations, which generally leads to significant improvements of the transmissivity model (Sun and Yeh, 1992). Fairly detailed presentations of the different inversion methods, with numerous references, can be found in Carrera et al. (1993) and in McLaughlin and Townley (1996).

We have given an overview of the various methods developed to date, including some that have not yet been tried in real field applications. The use of a geostatistical approach to the inverse problem in hydrogeology has seen a considerable development over the last two decades. Much effort is now being devoted to inverse modeling of geophysical data (gravimetry, magnetism, seismic) and integration of various types of data by multivariate inverse modeling techniques, especially for petroleum applications. Modeling spatial uncertainty while ensuring consistency with known physical equations is one of the most promising avenues of research for geostatistics.

NOTES

1. Named after Henri Darcy who originally discovered this law around 1856, in connection with a study of the fountains of Dijon, in France.
2. We only consider here aquifers in which gravity is the sole acting force and an incompressible fluid. In porous media, fluid motion velocities are very slow, and thus we can neglect the kinetic energy. The hydraulic head at point *x* then coincides with the static or piezometric head

$$H = \frac{P}{\rho g} + z$$

where *z* is the elevation of point *x* above some datum level, *P* is the (macroscopic) fluid pressure at that point [$\text{ML}^{-1}\text{T}^{-2}$], ρ the fluid (mass) density [ML^{-3}], and *g* the acceleration of gravity [LT^{-2}]. In Darcy's experiment the head at one end of the cylinder is equal to the elevation of the water level in the corresponding tank. In an aquifer with a uniform horizontal flow, the piezometric head is a constant along a vertical and equal to the elevation of the aquifer free surface. Horizontally it decreases along the direction of flow. For more details refer, for example, to de Marsily (1986).

3. The hydraulic conductivity, in general, depends on the fluid. It is expressed as

$$K = \frac{k\rho g}{\mu}$$

where μ is the dynamic viscosity of the fluid (which varies with temperature) $[\text{ML}^{-1}\text{T}^{-1}]$ and k is the intrinsic permeability $[\text{L}^2]$, which depends solely on the geometry of the void space. The latter is usually expressed in darcy units ($1 \text{ darcy} = 0.987 \cdot 10^{-12} \text{ m}^2$). Since we only consider water flows, we assimilate hydraulic conductivity to a property of the porous medium, even though the basic property is in reality the intrinsic permeability. In the petroleum industry the flow equation is written in terms of *pressure* $\rho g z$ and *mobility* k/μ (ease with which the fluid moves).

4. As we did for the macroscopic scale, we will assume that the fluid is incompressible, that the flow is slow enough to neglect kinetic energy terms, and that the only forces applied to the fluid are those of gravity. The Navier-Stokes equations are then usually written as

$$\frac{\partial p}{\partial x} - \mu \Delta u = -\rho g \text{grad } z$$

where p stands for pressure (at the microscopic scale) at point x of the fluid with elevation z . Defining the piezometric head as that at the macroscopic scale by

$$h = \frac{p}{\rho g} + z$$

we obtain equation (8.3) with $\alpha = \rho g / \mu$.

5. It is remarkable that Darcy's law can be derived from only the linearity of the Navier-Stokes equation without considering its exact form nor the boundary conditions. The reasoning behind Matheron's proof remains valid if the Navier-Stokes equations are replaced by any other linear equations. Thus the emergence of Ohm's law at the macroscopic level can be established from the relation between electrical current density and potential at the microscopic level. But there is no direct relationship between electrical and hydraulic conductivity values. (The boundary conditions are very different for electrical current and fluid flow: electrical current slides without friction along the surface of a grain, whereas fluid in contact with a grain clings to it and is at rest.)
6. The proof results from the fact that in two-dimensional space a 90° rotation transforms a gradient vector into a flux-conservative vector (zero divergence), and vice versa. It does not require permeability K to be scalar, but its mean and effective permeability are.
7. This result is proved when the first equation (8.13) is of the form

$$\frac{d\mathbf{Z}(t)}{dt} = \mathbf{F}(t, \mathbf{Z}(t))$$

where the vector transformation $\mathbf{F}(t, \mathbf{z})$ can be random—it is then a function $\mathbf{F}(t, \mathbf{z}; \omega)$ —and transforms a stochastic process of order two $\mathbf{Z}(t)$ into a stochastic process $\mathbf{F}(t, \mathbf{Z}(t))$ which also has moments of order 2. Upon simplification, the equation has one and only one solution if \mathbf{F} is continuous and satisfies a Lipschitz condition, in the mean square sense in the first case, in the ordinary sense for almost every realization \mathbf{f}_ω of \mathbf{F} in the second case. The Lipschitz condition limits the evolution with t of the difference $\mathbf{F}(t, \mathbf{Z}_2) - \mathbf{F}(t, \mathbf{Z}_1)$ with regard to the difference $\mathbf{Z}_2 - \mathbf{Z}_1$.

8. An exception concerns equations of the type

$$d\mathbf{Z}(t) = \mathbf{m}(t, \mathbf{Z}(t))dt + \sigma(t, \mathbf{Z}(t))d\mathbf{W}(t)$$

where $dW(t)$ designates Gaussian white noise deriving from a Wiener-Lévy process, which can be solved using stochastic integrals in the sense of Itô.

9. If the random noise is Gaussian, it can be interpreted as the derivative of a Wiener process $W(t)$. A reformulation of the problem in the approach of Itô, with the initial condition $Z(0) = z_0$, leads to a nonstationary solution for $Z(t)$ similar to that in Example 1 with $W(ds)$ instead of $Y(s)ds$.
10. The name *heat equation* usually refers to the equation

$$\frac{\partial Z}{\partial t} = a^2 \Delta Z$$

which defines the temperature Z of an isotropic homogeneous medium as a function of point x and time t (the Laplacian being taken with respect to space coordinates) in the absence of heat production.

11. $G(x, x')$ is zero on the boundary Γ . For fixed $x' = x_0$, $G(\cdot, x_0)$ corresponds to the solution of the Poisson equation for a point source located at x_0

$$\Delta G(x, x_0) = \delta(x - x_0)$$

If the boundary Γ is at infinity, G is an isotropic (generalized) function of $r = |x' - x|$ that depends on the space dimension n : $G(r) = r/2$ if $n = 1$, $= (\log r)/(2\pi)$ if $n = 2$, $= -1/(4\pi r)$ if $n = 3$, and in general for $n \neq 2$, $G(r) = -\Gamma(n/2)/[2(n-2)\pi^{n/2}r^{n-2}]$.

Appendix

This appendix gathers some classic definitions and results used in this book, as well as simulation formulas for a few covariance models.

A.1. MEASURE THEORY DEFINITIONS

Measures are encountered in many places of this book, especially in the context of spectral theory. For readers not familiar with measure theory, we recall a few definitions, but without going into mathematical details. Good sources are Royden (1968) and Neveu (1970).

Essentially a measure is a set function, i.e. a function which assigns a number to certain sets. The classic example is the Lebesgue measure which associates to each interval of the line its length, or to a subset in 2D or 3D its area or its volume. But we need a more general definition in order, for example, to assign a positive value to an isolated point (an atom).

σ -Algebra. A collection of subsets \mathcal{A} of a set Ω is a σ -algebra if it contains the empty set \emptyset and is closed with respect to complementation and with respect to countable unions:

$$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}; \quad A_i \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$$

From De Morgan's laws it follows that \mathcal{A} is also closed with respect to countable intersections.

Borel Sets. The collection \mathcal{B} of Borel sets of \mathbb{R}^n is the smallest σ -algebra containing all of the open sets of \mathbb{R}^n (or equivalently, all finite or infinite intervals of the line, or n -dimensional rectangles in \mathbb{R}^n).

Measurable Function-Random Variable. Given a space (Ω, \mathcal{A}) , a real-valued function $f : \Omega \rightarrow \mathbb{R}$ is said to be measurable if $f^{-1}(B) \in \mathcal{A}$ for every Borel set B of \mathbb{R} . By definition a random variable is a measurable function on (Ω, \mathcal{A}) . When the domain space Ω is \mathbb{R} or \mathbb{R}^n (and \mathcal{A} is \mathcal{B}) measurable functions include all ordinary functions of practical interest.

Measure. A measure μ on a σ -algebra \mathcal{A} of subsets of Ω is a set function mapping \mathcal{A} onto $]-\infty, +\infty\}$ such that $\mu(\emptyset) = 0$ and

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i) \quad (\sigma\text{-additivity})$$

for any sequence A_i of pairwise disjoint sets in \mathcal{A} . The measure μ is *positive* if $\mu(A) \geq 0$ ($A \in \mathcal{A}$) and *bounded* if

$$\sup\{|\mu(A)| : A \in \mathcal{A}\} < \infty$$

The measure μ is bounded if and only if $\mu(\Omega) < \infty$. A positive measure μ is called *finite* if $\mu(\Omega) < \infty$.

Some authors name “measure” a positive measure and the others “signed measures.” It is also possible to define a complex measure as a set function assigning a complex number to each $A \in \mathcal{A}$.

Examples: The Lebesgue measure $\mu(A) = |A|$ is a positive measure, but it is not bounded. A probability measure is a bounded positive measure such that $\mu(\Omega) = 1$.

Dirac Measure. The Dirac measure δ assigns the value 1 to any set B that contains the origin of coordinates and 0 to any set that doesn’t. It corresponds to a unit mass placed at the point 0. An equivalent definition is to consider δ as a pseudofunction $\delta(\cdot)$ summing to one and equal to zero everywhere, except at 0 where it is infinite (Dirac delta-function). For any continuous function φ we have

$$\delta(\varphi) = \int \delta(t)\varphi(t)dt = \varphi(0)$$

Likewise the measure δ_x corresponds to a unit mass at the point x , and we have

$$\delta_x(\varphi) = \int \delta(x-t)\varphi(t)dt = \varphi(x)$$

The above shows that δ is the identity operator for convolution, in the sense that $\delta^* \varphi = \varphi$.

Support. The support of a measure μ on \mathcal{B} is the smallest closed set F of \mathcal{B} such that μ vanishes in the complement of F . More precisely $\mu(O) = 0$ for any open set $O \subset \mathcal{B} - F$. The support of the Dirac measure is a single point, and that of the Lebesgue measure the whole space.

Absolute Continuity (Radon-Nikodym Theorem). A measure ν is absolutely continuous with respect to the positive measure μ if $\mu(A) = 0 \Rightarrow \nu(A) = 0$. Then the measure ν admits a *density function* f such that $\nu(A) = \int_A f(x)\mu(dx)$ for any $A \in \mathcal{A}$.

When μ is the Lebesgue measure, absolute continuity of ν means, in particular, that any point $\{x\}$ has measure zero $\nu(\{x\}) = 0$ (no atom).

Almost Surely, Almost Everywhere. We say that an event E occurs “almost surely” if $\Pr\{E\} = 1$, and that a property holds “almost everywhere” if the set of points where it fails to hold is of measure zero.

Orthogonal Complex Random Measure. The set function $\xi(B)$ associating a finite variance complex random variable to each $B \in \mathcal{B}$ is an orthogonal random measure if it satisfies the following properties:

$$\begin{cases} E\xi(B)\overline{\xi(B')} = 0 & \text{if } B, B' \in \mathcal{B}, \quad B \cap B' = \emptyset \\ \xi(B \cup B') = \xi(B) + \xi(B') & \text{if } B, B' \in \mathcal{B}, \quad B \cap B' = \emptyset \\ B_n \downarrow \emptyset \text{ in } \mathcal{B} & \Rightarrow E|\xi(B_n)|^2 \rightarrow 0 \end{cases}$$

Examples

1. Consider a Poisson point process with intensity λ ; the number of points $N(B)$ that fall in B minus its expected value define the random orthogonal measure $\xi(B) = N(B) - \lambda|B|$.
2. According to Chentsov (1957) Gaussian white noise is a random measure $\xi(\cdot)$ such that (a) $\xi(B)$ is normally distributed with mean 0 and variance $\lambda|B|$, and (b) when B and B' do not intersect $\xi(B)$ and $\xi(B')$ are independent and $\xi(B \cup B') = \xi(B) + \xi(B')$.
3. In 1D, let U be a random variable with a symmetric probability distribution $F(du)$, and Φ be uniform on $[0, 2\pi[$ and independent of U . The complex measure $\xi(\cdot)$ defined by

$$\sqrt{2}\xi(B) = \exp(i\Phi)\delta_U(B) + \exp(-i\Phi)\delta_{-U}(B)$$

is random and orthogonal and satisfies $E|\xi(du)|^2 = F(du)$. It is the random spectral measure of an SRF whose covariance is the Fourier transform of $F(du)$.

A.2. GAMMA FUNCTION (EULER'S INTEGRAL)

The gamma function is defined for $x > 0$ by

$$\Gamma(x) = \int_0^\infty e^{-u} u^{x-1} du \quad (\text{A.1})$$

It satisfies the recurrence relation

$$\Gamma(x+1) = x\Gamma(x)$$

Hence for a positive integer value $x = n$, one has

$$\Gamma(n+1) = n!$$

When $0 < x < 1$, it satisfies the reflection formula

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x}$$

For the particular value $x = \frac{1}{2}$, this formula implies that

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

The reflection formula defines the gamma function for $x < 0$ (but noninteger).

A.3. BESSEL FUNCTIONS

Since there are some variations about the definition of Bessel functions, we give the definition of the functions used in this book. It corresponds to the definitions used by Olver (1972).

The *Bessel function of the first kind* with index ν is defined by the development

$$J_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{(-1)^k}{k! \Gamma(\nu + k + 1)} \left(\frac{x}{2}\right)^{2k} \quad (\text{A.2})$$

Two particular cases are

$$J_{-1/2}(x) = \sqrt{\frac{2}{\pi x}} \cos x \quad J_{1/2}(x) = \sqrt{\frac{2}{\pi x}} \sin x$$

The *modified Bessel function of the first kind* is defined by the development

$$I_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(\nu + k + 1)} \left(\frac{x}{2}\right)^{2k} \quad (\text{A.3})$$

The *modified Bessel function of the second kind* is defined by

$$K_\nu(x) = \frac{\pi}{2} \frac{I_{-\nu}(x) - I_\nu(x)}{\sin \pi \nu} = K_{-\nu}(x) \quad (\text{A.4})$$

A.4. UNIT SPHERE OF \mathbb{R}^n

The surface area S_n and the volume V_n of the unit-radius sphere of \mathbb{R}^n are given by

$$S_n = \frac{2\pi^{n/2}}{\Gamma(n/2)} \quad V_n = \frac{S_n}{n} = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)} \quad (\text{A.5})$$

For example, $S_1 = 2$, $S_2 = 2\pi$, $S_3 = 4\pi$, $V_1 = 2$, $V_2 = \pi$, and $V_3 = \frac{4}{3}\pi$.

A.5. GAUSSIAN DISTRIBUTION AND HERMITE POLYNOMIALS

A.5.1. Gaussian Distribution

The standard Gaussian (synonym: standard normal) probability density function is defined by

$$g(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$$

Its dispersion indicator is

$$S = \frac{1}{\sqrt{\pi}}$$

A.5.2. Hermite Polynomials

The orthogonal polynomials associated with the standard Gaussian are the Hermite polynomials $H_n(y)$. They are defined by the Rodrigues formula

$$H_n(y)g(y) = \frac{d^n}{dy^n} g(y) \quad (\text{A.6})$$

These polynomials satisfy

$$\int H_m(y)H_n(y)g(y)dy = n! \delta_{mn}$$

An orthonormal basis for the Gaussian distribution is defined by the normalized polynomials

$$\chi_n(y) = \frac{1}{\sqrt{n!}} H_n(y)$$

For $n \geq 1$ relation (A.6) can be written as

$$H_n(y)g(y) = \frac{d}{dy} (H_{n-1}(y)g(y)) \quad (n \geq 1) \quad (\text{A.7})$$

Since $dg(y)/dy = -yg(y)$, we deduce the explicit form of Hermite polynomials:

$$H_n(y) = \sum_{i=0}^{\lfloor n/2 \rfloor} (-1)^{n-i} \frac{n!}{2^i i! (n-2i)!} y^{n-2i}$$

where $\lfloor n/2 \rfloor$ stands for the integer part of $n/2$. From this expression it follows that the derivative of $H_n(y)$ is proportional to $H_{n-1}(y)$:

$$\frac{d}{dy} H_n(y) = -nH_{n-1}(y) \quad (\text{A.8})$$

Applying definition (A.7) to H_{n+1} , we get

$$H_{n+1}(y)g(y) = \frac{d}{dy} (H_n(y)g(y)) = \left(-yH_n(y) + \frac{d}{dy} H_n(y) \right) g(y)$$

Whence

$$H_{n+1}(y) = -yH_n(y) + \frac{d}{dy} H_n(y)$$

and given (A.8),

$$H_{n+1}(y) = -yH_n(y) - nH_{n-1}(y)$$

This recurrence relation allows an easy calculation of the $H_n(y)$ from the first two polynomials $H_0(y) = 1$ and $H_1(y) = -y$. Thus, for example, $H_2(y) = y^2 - 1$, $H_3(y) = -y^3 + 3y$, and so on. Notice that the definition of H_n given here differs from the most common one by a factor $(-1)^n$.

Relation (A.7) is equivalent to

$$s_n(y) = \int_{-\infty}^y H_n(u)g(u)du = \begin{cases} G(y) & \text{if } n = 0 \\ H_{n-1}(y)g(y) & \text{if } n > 0 \end{cases} \quad (\text{A.9})$$

This formula permits an easy calculation of the coefficients of the expansion of $1_{y < y_0}$:

$$1_{y < y_0} = G(y_0) + g(y_0) \sum_{n=1}^{\infty} \frac{H_{n-1}(y_0)}{n!} H_n(y) \quad (\text{A.10})$$

In applications we also need to know the incomplete integral

$$S_{mn}(y) = \int_{-\infty}^y H_m(u)H_n(u)g(u)du$$

Using relation (A.7) and integration by parts, and then relation (A.8), it can be shown that

$$S_{mn}(y) = mS_{m-1,n-1}(y) + H_m(y)H_{n-1}(y)g(y) \quad (\text{A.11})$$

To initialize the recurrence calculation, we use the symmetry $S_{mn}(y) = S_{nm}(y)$ so as to reduce the problem to the case $n \geq m$, and we start from $S_{0,n-m} = s_{n-m}$, which is given by (A.9).

Let us also mention a useful expansion for studying lognormal distributions

$$\exp(\lambda y) = \exp\left(\frac{1}{2}\lambda^2\right) \sum_{n=0}^{\infty} (-1)^n \frac{\lambda^n}{n!} H_n(y) \quad (\text{A.12})$$

Another useful formula derived from Hochstrasser (1972, p. 786) is the integral

$$\int_{-\infty}^{+\infty} H_{2n}(tx)g(t)dt = \frac{(2n)!}{n!2^n}(x^2 - 1)^n$$

Other properties of Hermite polynomials for geostatistical applications can be found in Matheron (1974c, pp. 64–68) and in the literature on orthogonal polynomials (Szegő, 1939; Hochstrasser, 1972).

A.6. GAMMA DISTRIBUTION AND LAGUERRE POLYNOMIALS

A.6.1. Gamma Distribution

The standard gamma probability density functions are defined on $(0, \infty)$ by

$$g_{\alpha}(y) = \frac{1}{\Gamma(\alpha)} e^{-y} y^{\alpha-1} \quad (y > 0)$$

where α is a positive parameter and $\Gamma(\cdot)$ the gamma function (A.1).

The shape of the density g_{α} depends on α : if $\alpha < 1$, g_{α} is a decreasing function, unbounded at the origin; if $\alpha > 1$, g_{α} is a bell-shaped curve, which tends to a Gaussian distribution for large values of α ; in the intermediate case $\alpha = 1$, g_1 is the exponential density.

The gamma distribution has moments of all orders:

$$E[Y^n] = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)}$$

In particular,

$$E[Y] = \alpha \quad \text{Var}[Y] = \alpha$$

Its dispersion indicator is

$$S = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)} \quad (\text{A.13})$$

Its Laplace transform is

$$E[e^{-\lambda Y}] = \frac{1}{(1 + \lambda)^\alpha}$$

The standard gamma distribution corresponds to the case $b = 1$ of the general gamma distribution with parameter $\alpha > 0$ and scale $b > 0$, whose density is

$$g_{\alpha,b}(y) = \frac{b^\alpha}{\Gamma(\alpha)} e^{-by} y^{\alpha-1} \quad (y > 0)$$

If Y has the distribution $g_{\alpha,b}$, then bY has the distribution g_α .

A.6.2. Laguerre Polynomials

The orthogonal polynomials associated with the standard gamma distribution are the Laguerre polynomials $L_n^\alpha(y)$, defined on $]0, \infty[$ like g_α . They are defined by the Rodrigues formula

$$L_n^\alpha(y) g_\alpha(y) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + n)} \frac{d^n}{dy^n} [y^n g_\alpha(y)] = \frac{d^n}{dy^n} g_{\alpha+n}(y) \quad (\text{A.14})$$

These polynomials are orthogonal over the distribution g_α but are not normalized:

$$\int_0^\infty L_m^\alpha(y) L_n^\alpha(y) g_\alpha(y) dy = \frac{\Gamma(\alpha) n!}{\Gamma(\alpha + n)} \delta_{mn}$$

To get an orthonormal basis, we thus take

$$\chi_n^\alpha(y) = \sqrt{\frac{\Gamma(\alpha + n)}{\Gamma(\alpha) n!}} L_n^\alpha(y)$$

From definition (A.14) we get for $n \geq 1$,

$$L_n^\alpha(y) g_\alpha(y) = \frac{d}{dy} (L_{n-1}^{\alpha+1}(y) g_{\alpha+1}(y)) \quad (\text{A.15})$$

Using the relations

$$g_{\alpha+1}(y) = \frac{y}{\alpha} g_\alpha(y) \quad \text{and} \quad \frac{d}{dy} g_\alpha(y) = g_{\alpha-1}(y) - g_\alpha(y)$$

we deduce by recurrence that the explicit form of the n th Laguerre polynomial is

$$L_n^\alpha(y) = \sum_{i=0}^n (-1)^i \binom{n}{i} \frac{\Gamma(\alpha)}{\Gamma(\alpha+i)} y^i$$

and that its derivative is, up to a multiplicative factor, the Laguerre polynomial of degree $n-1$ associated with the distribution with parameter $\alpha+1$:

$$\frac{d}{dy} L_n^\alpha(y) = -\frac{n}{\alpha} L_{n-1}^{\alpha+1}(y) \quad (\text{A.16})$$

Laguerre polynomials with parameter α can easily be calculated from $L_0^\alpha(y) = 1$ and $L_1^\alpha(y) = 1 - (y/\alpha)$ via the recurrence relation

$$(\alpha+n)L_{n+1}^\alpha(y) = [2n+\alpha-y]L_n^\alpha(y) - nL_{n-1}^\alpha(y)$$

From (A.15) we get

$$s_n^\alpha(y) = \int_0^y L_n^\alpha(u) g_\alpha(u) du = \begin{cases} G_\alpha(y) & \text{if } n=0 \\ L_{n-1}^{\alpha+1}(y) g_{\alpha+1}(y) & \text{if } n>0 \end{cases} \quad (\text{A.17})$$

The incomplete integral

$$S_{mn}^\alpha(y) = \int_0^y L_m^\alpha(u) L_n^\alpha(u) g_\alpha(u) du$$

can be computed by recurrence, but in a somewhat more complex way than in the Gaussian case because the recurrence also involves the parameters of the gamma distribution. Setting (A.15) into the definition of $S_{mn}^\alpha(y)$, integrating by parts, and then using (A.16), we obtain

$$S_{mn}^\alpha(y) = \frac{m}{\alpha} S_{m-1,n-1}^{\alpha+1}(y) + L_m^\alpha(y) L_{n-1}^{\alpha+1}(y) g_{\alpha+1}(y) \quad (\text{A.18})$$

To initialize the recurrence calculation, we use the symmetry $S_{mn}^\alpha(y) = S_{nm}^\alpha(y)$ so as to reduce the problem to the case $n \geq m$, and we start from $S_{0,n-m}^{\alpha+m} = s_{n-m}^{\alpha+m}$.

A.7. NEGATIVE BINOMIAL DISTRIBUTION AND MEIXNER POLYNOMIALS

A.7.1. Negative Binomial Distribution

The integer random variable Y follows a negative binomial distribution if the probability that Y is equal to i is

$$p_i = (1-p)^\alpha \frac{\Gamma(\alpha+i)}{\Gamma(\alpha)} \frac{p^i}{i!} \quad i \in \mathbb{N}$$

where $\alpha > 0$ and $0 < p < 1$ are parameters and $\Gamma(\cdot)$ is the gamma function (A.1).

The negative binomial distribution may be considered as the discrete equivalent of the gamma distribution. The shape of the distribution defined by the p_i depends on α . In particular, if $\alpha < 1$, the p_i are decreasing. For $\alpha = 1$ we get the geometric, or Pascal, distribution.

The negative binomial distribution has for a generating function

$$E[s^Y] = \left(\frac{1-p}{1-ps} \right)^\alpha$$

The negative binomial distribution has moments of all orders and, in particular,

$$E[Y(Y-1)\cdots(Y-n+1)] = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \left(\frac{p}{1-p} \right)^n$$

$$E[Y] = \frac{\alpha p}{1-p} \quad \text{Var}[Y] = \frac{\alpha p}{(1-p)^2}$$

Its dispersion indicator is

$$S = \alpha \frac{p}{(1-p)^2} F\left(\alpha + 1, \frac{1}{2}; 2; -4 \frac{p}{(1-p)^2}\right) \quad (\text{A.19})$$

where F represents the hypergeometric function (Lantuéjoul, personal communication).

A.7.2. Meixner Polynomials

The orthogonal polynomials associated with the negative binomial distribution with parameters α and p are the Meixner polynomials $M_n^{\alpha p}(i)$. These can easily be calculated from $M_0^{\alpha p}(i) = 1$ and $M_1^{\alpha p}(i) = 1 - ((1-p)/p)i/\alpha$ via the recurrence

$$p(\alpha + n)M_{n+1}^{\alpha p}(i) = [n(1+p) + \alpha p - i(1-p)]M_n^{\alpha p}(i) - nM_{n-1}^{\alpha p}(i) \quad (\text{A.20})$$

Meixner polynomials satisfy the symmetry relation $M_n^{\alpha p}(i) = M_i^{\alpha p}(n)$. An orthonormal basis is obtained by taking

$$\chi_n^{\alpha p}(i) = \sqrt{\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \frac{p^n}{n!}} M_n^{\alpha p}(i)$$

A.8. LOGNORMAL DISTRIBUTION

By definition, the random variable Z is lognormal if $Y = \log Z$ is normal. Denoting by m and σ^2 the mean and variance of Y , the mean of Z is

$$M = E[\exp Y] = \exp(m + \frac{1}{2}\sigma^2) \quad (\text{A.21})$$

Likewise the vector-valued random variable $\mathbf{Z} = (Z_1, \dots, Z_n)'$ is lognormal if the vector-valued random variable $\mathbf{Y} = (Y_1, \dots, Y_n)'$, where $Y_i = \log Z_i$, is normal (i.e., has a *multivariate* Gaussian distribution).

The vector-valued random variable \mathbf{Z} being assumed lognormal, let us consider the random variable $Z = Z_1^{\lambda_1} \cdots Z_n^{\lambda_n}$. It is of the form $Z = \exp Y$ with $Y = \lambda_1 Y_1 + \cdots + \lambda_n Y_n$. The vector \mathbf{Y}

having a normal distribution, the random variable Y is Gaussian, with mean $m = \sum_{i=1}^n \lambda_i m_i$ and variance $\sigma^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \sigma_{ij}$, and thus Z is lognormal.

This property, together with formula (A.21), allows, for example, the calculation of the noncentered moments of a lognormal random variable and the noncentered covariance of two lognormal variables. In particular, we find that the variance of $Z = \exp Y$ is

$$\Sigma^2 = \text{Var}[\exp Y] = M^2(\exp(\sigma^2) - 1) \quad (\text{A.22})$$

and that if Y_1 and Y_2 are bivariate Gaussian with means m_1 and m_2 , variances σ_1^2 and σ_2^2 , and correlation coefficient ρ , the centered covariance of $\exp Y_1$ and $\exp Y_2$ is

$$\text{Cov}[\exp Y_1, \exp Y_2] = M_1 M_2 (\exp(\rho \sigma_1 \sigma_2) - 1) \quad (\text{A.23})$$

where M_1 and M_2 are the means of $\exp Y_1$ and $\exp Y_2$.

A.9. SIMULATION FORMULAS

(Tables A.1 and A.2 follow)

TABLE A.1. Simulation of common covariance models in \mathbb{R}^1 . For each covariance $C_1(h)$ in \mathbb{R}^1 the table gives a dilation function $w_1(s)$ satisfying $w_1 * \ddot{w}_1 = C_1$, and the spectral density $f_1(u)$ of $C_1(h)$. Here h stands for $|h|$, and $f_1(u)$ is expressed as a function of $\nu = 2\pi au$.

Covariance in \mathbb{R}^1	Dilation Function	Spectral Density
<i>Triangle model</i>		
$C_1(h) = \begin{cases} C \left(1 - \frac{h}{a}\right) & \text{if } h \leq a \\ 0 & \text{if } h \geq a \end{cases}$	$w_1(s) = \begin{cases} \sqrt{\frac{C}{a}} & \text{if } s \leq \frac{a}{2} \\ 0 & \text{if } s > \frac{a}{2} \end{cases}$	$f_1(u) = 2Ca \frac{1 - \cos \nu}{\nu^2}$
<i>Spherical model</i>		
$C_1(h) = \begin{cases} C \left(1 - \frac{3}{2} \frac{h}{a} + \frac{1}{2} \frac{h^3}{a^3}\right) & \text{if } h \leq a \\ 0 & \text{if } h \geq a \end{cases}$	$w_1(s) = \begin{cases} \sqrt{\frac{3C}{4a}} \left(1 - 2\frac{s}{a}\right) & \text{if } s \leq \frac{a}{2} \\ 0 & \text{if } s > \frac{a}{2} \end{cases}$	$f_1(u) = 3Ca \frac{(1 - \cos \nu)^2 + (\nu - \sin \nu)^2}{\nu^4}$

Cubic model

$$C_1(h) = \begin{cases} C \left(1 - \frac{h^2}{a^2} + \frac{35}{4} \frac{h^3}{a^3} - \frac{7}{2} \frac{h^5}{a^5} + \frac{3}{4} \frac{h^7}{a^7} \right) & \text{if } h \leq a \\ 0 & \text{if } h \geq a \end{cases}$$

$$w_1(s) = \begin{cases} \sqrt{\frac{105}{64}} \frac{C}{a} \left(1 + 2 \frac{s}{a} - 4 \frac{s^2}{a^2} - 8 \frac{s^3}{a^3} \right) & \text{if } |s| \leq \frac{a}{2} \\ 0 & \text{if } |s| > \frac{a}{2} \end{cases}$$

$$f_1(u) = 105Ca \frac{[4\nu \sin \nu + (6 - \nu^2) \cos \nu - 6]^2 + [(6 - \nu^2) \sin \nu - 4\nu \cos \nu - 2\nu]^2}{\nu^8}$$

Exponential model

$$C_1(h) = Ce^{-h/a}$$

$$w_1(s) = \begin{cases} 0 & \text{if } s < 0 \\ \sqrt{2} \frac{C}{a} e^{-s/a} & \text{if } s \geq 0 \end{cases}$$

$$f_1(u) = \frac{2Ca}{1 + \nu^2}$$

Gaussian model

$$C_1(h) = C \exp \left(-\frac{h^2}{a^2} \right)$$

$$w_1(s) = \sqrt{\frac{4}{\pi}} \frac{C}{a} \exp \left(-2 \frac{s^2}{a^2} \right)$$

$$f_1(u) = \sqrt{\pi} Ca \exp \left(-\frac{\nu^2}{4} \right)$$

TABLE A.2. Simulation of common covariance models in \mathbb{R}^3 by turning bands. For each covariance $C_3(r)$ in \mathbb{R}^3 the table gives: the associated covariance $C_1(h)$ to be simulated in \mathbb{R}^1 , a dilution function $w_1(s)$ satisfying $w_1 * w_1 = C_1$, and the spectral density $f_1(u)$ of $C_1(h)$. Here h stands for $|h|$, and $f_1(u)$ is expressed as a function of $\nu = 2\pi au$.

Covariance in \mathbb{R}^3	Covariance in \mathbb{R}^1	Dilution Function	Spectral Density
<i>Spherical model</i>			
$C_3(r) = \begin{cases} C \left(1 - \frac{3r}{2a} + \frac{1}{2} \frac{r^3}{a^3} \right) & \text{if } r \leq a \\ 0 & \text{if } r \geq a \end{cases}$	$C_1(h) = \begin{cases} C \left(1 - 3\frac{h}{a} + 2\frac{h^3}{a^3} \right) & \text{if } h \leq a \\ 0 & \text{if } h \geq a \end{cases}$	$w_1(s) = \begin{cases} \sqrt{12} \frac{C}{a} \frac{s}{a} & \text{if } s \leq \frac{a}{2} \\ 0 & \text{if } s > \frac{a}{2} \end{cases}$	$f_1(u) = \frac{12Ca}{\nu^4} \left[2 \sin \frac{\nu}{2} - \nu \cos \frac{\nu}{2} \right]^2$
<i>Cubic model</i>			
$C_3(r) = \begin{cases} C \left(1 - 7\frac{r^2}{a^2} + \frac{35}{4} \frac{r^3}{a^3} - \frac{7}{2} \frac{r^5}{a^5} + \frac{3}{4} \frac{r^7}{a^7} \right) & \text{if } r \leq a \\ 0 & \text{if } r \geq a \end{cases}$	$C_1(h) = \begin{cases} C \left(1 - 21\frac{h^2}{a^2} + 35\frac{h^3}{a^3} - 21\frac{h^5}{a^5} + 6\frac{h^7}{a^7} \right) & \text{if } h \leq a \\ 0 & \text{if } h \geq a \end{cases}$		

$$w_1(s) = \begin{cases} \sqrt{210 \frac{C}{a}} \left(\frac{1}{2} \frac{s}{a} - 2 \frac{s^3}{a^3} \right) & \text{if } |s| \leq \frac{a}{2} \\ 0 & \text{if } |s| > \frac{a}{2} \end{cases}$$

$$f_1(u) = \frac{840Ca}{\nu^8} \left[(12 - \nu^2) \sin \frac{\nu}{2} - 6\nu \cos \frac{\nu}{2} \right]^2$$

Exponential model

$$C_3(r) = C e^{-r/a}$$

$$C_1(h) = C \left(1 - \frac{h}{a} \right) e^{-h/a}$$

$$w_1(s) = \begin{cases} 0 & \text{if } s < 0 \\ 2\sqrt{\frac{C}{a}} \left(1 - \frac{s}{a} \right) e^{-s/a} & \text{if } s \geq 0 \end{cases}$$

$$f_1(u) = \frac{4Cav^2}{(1 + \nu^2)^2}$$

Gaussian model

$$C_3(r) = C \exp \left(-\frac{r^2}{a^2} \right)$$

$$C_1(h) = C \left(1 - 2 \frac{h^2}{a^2} \right) \exp \left(-\frac{h^2}{a^2} \right)$$

$$w_1(s) = \sqrt{\frac{256}{\pi} \frac{C}{a}} \frac{s}{a} \exp \left(-2 \frac{s^2}{a^2} \right)$$

$$f_1(u) = \sqrt{\pi} C a \nu^2 \exp \left(-\frac{\nu^2}{2} \right)$$

References

- Ababou, R., D. McLaughlin, L. W. Gelhar, and A. F. B. Tompson (1989). Numerical simulation of three-dimensional saturated flow in randomly heterogeneous porous media. *Transport in Porous Media*, **4**(6), 549–565.
- Abrahamsen, P. (1993). Bayesian kriging for seismic depth conversion of a multi-layer reservoir. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 385–398.
- Ahrens, L. H. (1954). The lognormal distribution of the elements (A fundamental law of geochemistry and its subsidiary). *Geochimica et Cosmochimica Acta*, **5**, 49–73.
- Aitchison, D. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.
- Alabert, F. (1987a). The practice of fast conditional simulations through the LU decomposition of the covariance matrix. *Mathematical Geology*, **19**(5), 369–386.
- Alabert, F. (1987b). *Stochastic Imaging of Spatial Distributions Using Hard and Soft Information*. M.Sc. thesis, Stanford University, California.
- Alabert, F., and G. J. Massonnat (1990). Heterogeneity in a complex turbiditic reservoir: Stochastic modeling of facies and petrophysical variability. *SPE paper* 20604, 65th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, New Orleans, Louisiana, 775–790.
- Alfaro, M. (1979). *Etude de la robustesse des simulations de fonctions aléatoires*. Doctoral thesis, E.N.S. des Mines de Paris.
- Alfaro, M. (1980). The random coin method: Solution to the problem of the simulation of a random function in the plane. *Journal of the International Association for Mathematical Geology*, **12**(1), 25–32.
- Alfaro, M. (1984). Statistical inference of the semivariogram and the quadratic model. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 45–53.
- Alfaro, M. (1993). Naturalistic simulations of deposits. Application to the problem of the kriging variance. Presented at the Conference *Geostatistics for the Next Century*, Montreal, June 1993.
- Alfsen, E. M. (1971). *Compact Convex Sets and Boundary Integrals*. Springer, Berlin.
- Allard, D. (1993). *Connexité des ensembles aléatoires: Application à la simulation des réservoirs pétroliers hétérogènes*. Doctoral thesis, E.N.S. des Mines de Paris.

- Allard, D., and HERESIM Group (1993). On the connectivity of two random set models: the truncated Gaussian and the Boolean. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 467–478.
- Allard, D. (1994). Simulating a geological lithofacies with respect to connectivity information using the truncated Gaussian model. In *Geostatistical Simulations*, M. Armstrong and P. A. Dowd, eds. Kluwer, Dordrecht, Netherlands, 197–211.
- Allègre, C. J., and E. Lewin (1995). Scaling laws and geochemical distributions. *Earth and Planetary Science Letters*, **132**, 1–13.
- Andersson, J., A. M. Shapiro, and J. Bear (1984). A stochastic model of a fractured rock conditioned by measured information. *Water Resources Research*, **20**(1), 79–88.
- Antoine, J. N., and J. P. Delhomme (1993). A method to derive dips from bedding boundaries in boreholes images. *SPE Formation Evaluation*, June, 96–102.
- Arfken, G. (1985). *Mathematical Methods for Physicists*. Academic Press.
- Armstrong, M. (1994). Is research in mining geostats as dead as a dodo? In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 303–312.
- Armstrong, M., and P. Delfiner (1980). Towards a more robust variogram: A case study on coal. Technical Report N-671, Centre de Géostatistique, Fontainebleau, France.
- Armstrong, M., and R. Jabin (1981). Variogram models must be positive-definite. *Journal of the International Association for Mathematical Geology*, **13**(5), 455–459.
- Armstrong, M., and G. Matheron (1986a). Disjunctive kriging revisited: Part I. *Mathematical Geology*, **18**(8), 711–728.
- Armstrong, M., and G. Matheron (1986b). Disjunctive kriging revisited: Part II. *Mathematical Geology*, **18**(8), 729–742.
- Armstrong, M., and H. Wackernagel (1988). The influence of the covariance function on the kriged estimator. *Sciences de la Terre, Série Informatique Géologique*, **27**(II), 245–262.
- Arnold, L. (1973). *Stochastische Differentialgleichungen*. Oldenburg, München, Germany. Translation (1974): *Stochastic Differential Equations: Theory and Applications*. Wiley, New York.
- Askey, R. (1975). *Orthogonal Polynomials and Special Functions*. Society for Industrial and Applied Mathematics, Philadelphia.
- Baddeley, A. J., and B. W. Silverman (1984). A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics*, **40**, 1089–1093.
- Baecher, G. B., N. A. Lanney, and H. H. Einstein (1977). Statistical description of rock properties and sampling. In *Proceedings of the 18th U.S. Symposium on Rock Mechanics*, AIME, New York, **5C1**, 1–8.
- Bardossy, A., I. Bogardi, and W. E. Kelly (1988). Imprecise (fuzzy) information in geostatistics. *Mathematical Geology*, **20**(4), 287–311.
- Bardossy, A., I. Bogardi, and W. E. Kelly (1990). Kriging with imprecise (fuzzy) variograms. I: Theory. II: Application. *Mathematical Geology*, **22**(1), 63–79 (I), 81–94 (II).
- Barker, A. A. (1965). Monte Carlo calculation of the radial distribution function for a proton-electron plasma. *Australian Journal of Physics*, **18**, 119–133.

- Barnes, R. J., and T. B. Johnson (1984). Positive kriging. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 231–244.
- Bartlett, M. S. (1963). The spectral analysis of point processes. *Journal of the Royal Statistical Society, Series B*, **25**(2), 264–296.
- Bartlett, M. S. (1975). *The Statistical Analysis of Spatial Pattern*. Chapman and Hall, London.
- Bastin, G., B. Lorent, C. Duque, and M. Gevers (1984). Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. *Water Resources Research*, **20**(4), 463–470.
- Beckman, P. (1973). *Orthogonal Polynomials for Engineers and Physicists*. Golem Press, Boulder, Colorado.
- Benzécri, J. P., et al. (1973). *L'analyse des données*. Tome 1, *La taxinomie*; Tome 2, *L'analyse des correspondances*. Dunod, Paris.
- Billaux, D. (1989). Influence de la connectivité d'un réseau de fractures sur sa réponse à un essai de pompage. In *Proceedings of ISRM/SPE International Symposium Rock Mechanics and Rock Physics at Great Depth*, V. Maury and D. Fourmaintraux, eds. Balkema, Rotterdam, Netherlands, 473–480.
- Blanchin, R., and J. P. Chilès (1992). Geostatistical modelling of geological layers and optimization survey design for the Channel tunnel. In *Computer Graphics in Geology: Three-dimensional Computer Graphics in Modeling Geologic Structures and Simulating Geologic Processes*, R. Pflug and J. W. Harbaugh, eds., Lecture Notes in Earth Sciences, Vol. 41, Springer-Verlag, Berlin Heidelberg, Germany, 251–256.
- Blanchin, R., and J. P. Chilès (1993a). The Channel Tunnel: Geostatistical prediction of the geological conditions and its validation by the reality. *Mathematical Geology*, **25**(7), 963–974.
- Blanchin, R., and J. P. Chilès (1993b). Channel Tunnel: Geostatistical prediction facing the ordeal of reality. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 757–766.
- Blanchin, R., J. P. Chilès, and F. Deverly (1989). Some applications of geostatistics to civil engineering. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 785–795.
- Blanc-Lapierre, A., and R. Fortet (1953). *Théorie des fonctions aléatoires*. Masson, Paris.
- Bordessoule, J. L., C. Demange, and J. Rivoirard (1988). Estimation des réserves géologiques en uranium par variables utiles à résidu autokrigeable. *Sciences de la Terre, Série Informatique Géologique*, **28**, 27–51.
- Bordessoule, J. L., C. Demange, and J. Rivoirard (1989). Using an orthogonal residual between ore and metal to estimate in-situ resources. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 923–934.
- Borgman, L., M. Taheri, and R. Hagan (1984). Three-dimensional frequency-domain simulation of geological variables. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 517–541.

- Bortoli, L. J., F. Alabert, A. Haas, and A. G. Journel (1993). Constraining stochastic images to seismic data. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 325–337.
- Boulanger, F. (1990). *Modélisation et simulation de variables régionalisées, par des fonctions aléatoires stables*. Doctoral thesis, E.N.S. des Mines de Paris.
- Bouleau, N. (1986). *Probabilités de l'ingénieur—Variables aléatoires et simulation*. Hermann, Paris.
- Bourgault, G., and D. Marcotte (1993). Spatial filtering under the linear coregionalization model. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 237–248.
- Box, G. E. P., and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*, rev. ed. Holden-Day, Oakland, California. First edition (1970).
- Bradley, R., and J. Haslett (1992). High-interaction diagnostics for geostatistical models of spatially referenced data. *Statistician*, **41**, 371–380.
- Bras, R. L., and I. Rodríguez-Iturbe (1985). *Random Functions and Hydrology*. Addison-Wesley, Reading, Massachusetts.
- Brooker, P. I. (1985). Two-dimensional simulation by turning bands. *Journal of the International Association for Mathematical Geology*, **17**(1), 81–90.
- Buxton, B. E., and A. D. Pate (1994). Joint temporal-spatial modeling of concentrations of hazardous pollutants in urban air. In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 75–87.
- Cardwell, W. T., and R. L. Parsons (1945). Average permeabilities of heterogeneous oil sands. *Trans. Am. Inst. Mining Met. Pet. Eng.*, 34–42.
- Carlier, A. (1964). *Contribution aux méthodes d'estimation des gisements d'uranium*. Doctoral thesis, Université de Paris, Série A, No. 4226. No. d'ordre 5077.
- Carr, J. R., and E. D. Deng (1987). Comparison of two techniques for applying disjunctive kriging: The Gaussian anamorphosis model versus the direct statistical inference of the bivariate distributions. *Mathematical Geology*, **19**(1), 57–68.
- Carr, J. R., D. E. Myers, and C. E. Glass (1985). Co-kriging—A computer program: *Computers in Geosciences*, **11**, 111–127.
- Carrera, J., and S. P. Neuman (1986a). Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Water Resources Research*, **22**(2), 199–210.
- Carrera, J., and S. P. Neuman (1986b). Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms. *Water Resources Research*, **22**(2), 211–227.
- Carrera, J., A. Medina, and G. Galarza (1993). Groundwater inverse problem. Discussion on geostatistical formulations and validation. *Hydrogéologie*, **(4)**, 313–324.
- Casella, G., and E. I. George (1992). Explaining the Gibbs sampler. *The American Statistician*, **46**(3), 167–174.
- Champigny, N., and M. Armstrong (1989). Estimation of fault-controlled deposits. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 971–983.
- Chapman, N. A., J. Andersson, P. Robinson, K. Skagius, C. O. Wene, M. Wiborgh, and S. Wingefors (1995). Devising scenarios for future repository evolution: A rigorous methodology. In *Materials Research Society Symposium Proceedings*, Vol. 353, Materials Research Society, 495–502.

- Chautru, J. M. (1989). The use of Boolean random functions in geostatistics. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 201–212.
- Chauvet, P. (1982). The variogram cloud. In *Proceedings of the 17th APCOM International Symposium*, Golden, Colorado, 757–764.
- Chauvet, P. (1987). *Éléments d'analyse structurale des FAI-k à 1 dimension*. Doctoral thesis, E.N.S. des Mines de Paris.
- Chauvet, P., J. Pailleux, and J. P. Chilès (1976). Analyse objective des champs météorologiques par cokrigage. *La Météorologie, Sciences et Techniques, 6ème Série*, **4**, 37–54.
- Cheng, Q. C., and F. P. Agterberg (1996). Multifractal modeling and spatial statistics. *Mathematical Geology*, **28**(1), 1–16.
- Chentsov, N. N. (1957). Lévy Brownian motion for several parameters and generalized white noise. *Theory of Probability and Its Applications*, **2**, 265–266.
- Chessa, A. (1995). *Conditional simulation of spatial stochastic models for reservoir heterogeneity*. Doctoral thesis, Technische Universiteit Delft, Holland.
- Chib, S., and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**, 327–335.
- Chilès, J. P. (1974). Analyse classique et analyse en FAI-k de données topographiques de la région de Noiretable. Technical Report N-394, Centre de Géostatistique, Fontainebleau, France.
- Chilès, J. P. (1976). How to adapt kriging to non-classical problems: three case studies. In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Huijbregts, eds. Reidel, Dordrecht, Holland, 69–89.
- Chilès, J. P. (1977). *Géostatistique des phénomènes non stationnaires (dans le plan)*. Doctoral thesis, Université de Nancy-I, France.
- Chilès, J. P. (1978). L'inférence statistique automatique des FAI-k. Technical Report N-584, Centre de Géostatistique, Fontainebleau, France.
- Chilès, J. P. (1979a). La dérive à la dérive. Technical Report N-591, Centre de Géostatistique, Fontainebleau, France.
- Chilès, J. P. (1979b). Le variogramme généralisé. Technical Report N-612, Centre de Géostatistique, Fontainebleau, France.
- Chilès, J. P. (1984). Simulation of a nickel deposit: Problems encountered and practical solutions. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 2, 1015–1030.
- Chilès, J. P. (1989a). Three-dimensional geometric modeling of a fracture network. In *Geostatistical, Sensitivity, and Uncertainty Methods for Ground-Water Flow and Radionuclide Transport Modeling*, B. E. Buxton, ed. Battelle Press, Columbus, Ohio, 361–385.
- Chilès, J. P. (1989b). Modélisation géostatistique de réseaux de fractures. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 57–76.
- Chilès, J. P. (1995). Quelques méthodes de simulation de fonctions aléatoires intrinsèques. In *Compte Rendu des Journées de Géostatistique*, C. de Fouquet, ed. *Cahiers de Géostatistique*, Fasc. 5, Ecole des Mines de Paris, 97–112.
- Chilès, J. P., and P. Delfiner (1975). Reconstitution par krigage de la surface topographique à partir de différents schémas d'échantillonnage photogrammétrique. *Bulletin de la Société Française de Photogrammétrie*, **57**, 42–50.

- Chilès, J. P., and P. Delfiner (1996). Using FFT techniques for simulating Gaussian random fields. In *Proceedings of the Conference on Mining Geostatistics, Berg-en-dal, Kruger National Park, South Africa, 19–22 September 1994*, Geostatistical Association of South Africa, 131–140.
- Chilès, J. P., and P. Delfiner (1997). Discrete exact simulation by the Fourier method. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 1, 258–269.
- Chilès, J. P., and R. Gable (1984). Three-dimensional modelling of a geothermal field. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 2, 587–598.
- Chilès, J. P., and S. Gentier (1993). Geostatistical modeling of a single fracture. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 5–108.
- Chilès, J. P., and A. Guillen (1984). Variogrammes et krigeages pour la gravimétrie et le magnétisme. *Sciences de la Terre, Série Informatique Géologique*, **20**, 455–468.
- Chilès, J. P., and H. T. Liao (1993). Estimating the recoverable reserves of gold deposits: Comparison between disjunctive kriging and indicator kriging. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 1053–1064.
- Chilès, J. P., and G. de Marsily (1993). Stochastic models of fracture systems and their use in flow and transport modeling. In *Flow and Contaminant Transport in Fractured Rock*, J. Bear, C. F. Tsang, and G. de Marsily, eds. Academic Press, San Diego, California, ch. 4, 169–236.
- Chilès, J. P., R. Gable, and R. H. Morin (1991). Contribution de la géostatistique à l'étude de la thermicité des rides océaniques du Pacifique. *Cahiers de Géostatistique*, Fasc. 1, Ecole des Mines de Paris, 51–61.
- Chilès, J. P., F. Guérin, and D. Billaux (1992). 3D stochastic simulation of fracture network and flow at Stripa conditioned on observed fractures and calibrated on measured flow rates. In *Rock Mechanics*, J. R. Tillerson and W. R. Wawersik, eds. Balkema, Rotterdam, Netherlands, 533–542.
- Chilès, J. P., B. Bourguine, and I. Niandou (1996). Designing an additional sampling pattern to determine the variogram at short distances. In *Proceedings of the Conference on Mining Geostatistics, Berg-en-dal, Kruger National Park, South Africa, 19–22 September 1994*, Geostatistical Association of South Africa, 200–212.
- Christakos, G. (1992). *Random Fields Models in Earth Sciences*. Academic Press, San Diego.
- Chung, C. F. (1984). Use of the jackknife method to estimate autocorrelation functions (or variograms). In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 55–69.
- Clark, I. (1979). Does geostatistics work? In *Proceedings of the 16th APCOM International Symposium*, T. J. O'Neil, ed. Society of Mining Engineers of the AIME, New York, 213–225.
- Clifton, P. M., and S. P. Neuman (1982). Effects of kriging and inverse modeling on conditional simulation of the Avra Valley aquifer in southern Arizona. *Water Resources Research*, **18**(4), 1215–1234.
- Cohen, A. (1992). *Ondelettes et traitement numérique du signal*. Masson, Paris.

- Cooley, R. L. (1982). Incorporation of prior information on parameters into nonlinear regression groundwater flow models. 1. Theory. *Water Resources Research*, **18**(4), 965–976.
- Cooley, R. L. (1983). Incorporation of prior information on parameters into nonlinear regression groundwater flow models. 2. Applications. *Water Resources Research*, **19**(3), 662–676.
- Cox, D. R. (1955). Some statistical models connected with series of events. *Journal of the Royal Statistical Society, Series B*, **17**(2), 129–164.
- Cox, D. R., and D. V. Hinkley (1974). *Theoretical Statistics*. Chapman and Hall, London.
- Crain, I. K. (1978). The Monte-Carlo simulation of random polygons. *Computers and Geosciences*, **4**, 131–141.
- Cramér, H. (1940). On the theory of stationary random processes. *Annals of Mathematics*, **41**(1), 215–230.
- Cramér, H. (1942). On harmonic analysis in certain functional spaces. *Arkiv för Matematik, Astronomi och Fysik (Uppsala)*, **28B**(12), 1–7.
- Cramér, H. (1945). *Mathematical Methods of Statistics*. Almqvist and Wiksells, Uppsala, Sweden. Also (1974): Princeton University Press, New Jersey.
- Craven, P., and G. Wahba (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, **31**, 377–403.
- Cressie, N. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, **17**(5), 563–586.
- Cressie, N. (1987). A nonparametric view of generalized covariances for kriging. *Mathematical Geology*, **19**(5), 425–449.
- Cressie, N. (1991). *Statistics for Spatial Data*. Wiley, New York. Reprinted (1993).
- Cressie, N., and D. M. Hawkins (1980). Robust estimation of the variogram: I. *Journal of the Mathematical Association for Mathematical Geology*, **12**(2), 115–125.
- Cressman, G. P. (1959). An operational objective analysis system. *Monthly Weather Review*, **87**(10).
- Dagan, G. (1985a). Stochastic modelling of groundwater by unconditional and conditional probabilities: the inverse problem. *Water Resources Research*, **21**(1), 65–72.
- Dagan, G. (1985b). A note on the higher-order corrections of the head covariances in steady aquifer flow. *Water Resources Research*, **21**(4), 573–578.
- Dagan, G. (1989). *Flow and Transport in Porous Formations*. Springer, Berlin Heidelberg.
- Dagbert, M., M. David, D. Crozel, and A. Desbarats (1984). Computing variograms in folded strata-controlled deposits. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 71–89.
- Daly, C., D. Jeulin, and C. Lajaunie (1989). Application of multivariate kriging to the processing of noisy images. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 749–760.
- Dantzig, G. B. (1963). *Linear Programming and Extensions*. Princeton University Press, Princeton, New Jersey.

- Da Rold, C., J. Deraisme, and R. Dumay (1980). Etude méthodologique de la simulation d'une exploitation de nickel de Nouvelle-Calédonie sur un modèle géostatistique de gisement. *Revue de l'Industrie Minérale*, **62**(12), 651–663.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia.
- David, M. (1977). *Geostatistical Ore Reserve Estimation*. Elsevier, Amsterdam, Netherlands.
- David, M. (1988). *Handbook of Applied Advanced Geostatistical Ore Reserve Evaluation*. Elsevier, Amsterdam, Netherlands.
- David, M., D. Marcotte, and M. Soulié (1984). Conditional bias in kriging and a suggested correction. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 217–230.
- Davis, M. W. (1987a). Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Mathematical Geology*, **19**(2), 91–98.
- Davis, M. W. (1987b). Generating large stochastic simulations—The matrix polynomial approximation method. *Mathematical Geology*, **19**(2), 99–107.
- Davis, M. W., and C. Grivet (1984). Kriging in a global neighborhood. *Journal of the International Association for Mathematical Geology*, **16**(3), 149–265.
- Delfiner, P. (1970). Le schéma booléen-poissonien. Technical Report N-212, Centre de Morphologie Mathématique, Fontainebleau, France.
- Delfiner, P. (1973). Analyse objective du géopotential et du vent géostrophique par krigeage universel. New print (1975): *La Météorologie*, 5^e Série, **25**, 1–57.
- Delfiner, P. (1976). Linear estimation of nonstationary spatial phenomena. In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Huijbregts, eds. Reidel, Dordrecht, Holland, 49–68.
- Delfiner, P. (1977). *Shift invariance under Linear Models*. PhD dissertation, Princeton University, New Jersey.
- Delfiner, P., and J. P. Chilès (1977). Conditional simulation: A new Monte-Carlo approach to probabilistic evaluation of hydrocarbon in place. *SPE paper* 6985.
- Delfiner, P., and J. P. Delhomme (1975). Optimum interpolation by kriging. In *Display and Analysis of Spatial Data*, J. C. Davis and M. J. McCullagh, eds. Wiley, London, 96–114.
- Delfiner, P., and R. O. Gilbert (1978). Combining two types of survey data for estimating geographical distribution of plutonium in Area 13. In *Selected Environmental Plutonium Research Reports of the NAEG*, NVO-192. Department of Energy, NTIS, Springfield, Virginia, Vol. 2, 361–404.
- Delfiner, P., J. P. Delhomme, and J. Pélissier-Combescure (1983). Application of geostatistical analysis to the evaluation of petroleum reservoirs with well logs. In *Proceedings of the SPWLA 24th Annual Logging Symposium, Calgary, June 1983*.
- Delhomme, A. E. K., and J. F. Giannesini (1979). New reservoir description technics improve simulation results in Hassi-Messaoud field—Algeria. *SPE paper* 8435, 54th Annual Fall Technical Conference and Exhibition of the Society of Petroleum Engineers, Las Vegas.

- Delhomme, J. P. (1974). La cartographie d'une grandeur physique à partir de données de différentes qualités. In *Mémoires de l'Association Internationale des Hydrogéologues, Proceedings of the Montpellier (France) Meeting*, Vol. 10, 185–194.
- Delhomme, J. P. (1976). *Applications de la théorie des variables régionalisées dans les sciences de l'eau*. Doctoral thesis, Université Pierre et Marie Curie—Paris VI. Simplified text and abstract (1978, same title): *Bulletin du B.R.G.M. (deuxième série)*, Section III, No. 4, 341–375.
- Delhomme, J. P. (1978). Kriging in the hydrosociences. *Advances in Water Resources*, 1(5), 251–266.
- Delhomme, J. P. (1979a). Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach. *Water Resources Research*, 15(2), 269–280.
- Delhomme, J. P. (1979b). Kriging under boundary conditions. Presented at the *American Geophysical Union Fall Meeting, San Francisco, December 1979*.
- Delhomme, J. P., M. Boucher, G. Meunier, and F. Jensen (1981). Apport de la géostatistique à la description des stockages de gaz en aquifère. *Revue de l'Institut Français du Pétrole*, 36(3), 309–327.
- Demange, C., C. Lajaunie, C. Lantuéjoul, and J. Rivoirard (1987). Global recoverable reserves: Testing various changes of support models on uranium data. In *Geostatistical Case Studies*, G. Matheron and M. Armstrong, eds. Reidel, Dordrecht, Holland, 135–147.
- Deraisme, J., and P. van Deursen (1995). Geostatistical modeling of multiple layers and its application to volumetric computation of dredging projects. In *Applications of Statistics and Probability—Civil Engineering Reliability and Risk Analysis*, M. Lemaire, J. L. Favre, and A. Mébarki, eds. Balkema, Rotterdam, Netherlands, Vol. 2, 1221–1227.
- Deraisme, J., and R. Dumay (1979). Geostatistics and mining processes. In *Proceedings of the 16th International APCOM Symposium*, T. J. O'Neil, ed. Society of Mining Engineers of the AIME, New York, 149–162.
- Deraisme, J., C. de Fouquet, and H. Fraisse (1983). Méthodologie de simulation d'exploitation souterraine dans les mines métalliques—Application à un gisement d'uranium en Australie. *Industrie Minérale—Les Techniques*, No. 10-83, 513–523.
- Deraisme, J., C. de Fouquet, and H. Fraisse (1984). Geostatistical orebody model for computer optimization of profits from different underground mining methods. In *Proceedings of the 18th International APCOM Symposium*, Institution of Mining and Metallurgy, London, 583–590.
- Dettinger, M. D., and J. L. Wilson (1981). First-order analysis of uncertainty in numerical models of groundwater flow. Part 1. Mathematical development. *Water Resources Research*, 17(1), 149–161.
- Deutsch, C. V. (1993). Conditioning reservoir models to well test information. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 505–518.
- Deutsch, C. V., and P. W. Cockerham (1994). Practical considerations in the application of simulated annealing to stochastic simulation. *Mathematical Geology*, 26(1), 67–82.
- Deutsch, C. V., and A. G. Journel (1998). *GSLIB: Geostatistical Software Library and User's Guide*. 2nd edition, Oxford University Press, New York.
- Deverly, F. (1984a). Geostatistical approach to mining sampling. In *Proceedings of the 18th APCOM International Symposium*, London, 379–388.

- Deverly, F. (1984b). *Echantillonnage et géostatistique*. Doctoral thesis, E.N.S. des Mines de Paris.
- Diamond, P., and M. Armstrong (1984). Robustness of variograms and conditioning of kriging matrices. *Journal of the International Association for Mathematical Geology*, **16**(8), 809–822.
- Dieter, U. (1972). Statistical interdependence of pseudo-random numbers generated by the linear congruential method. In *Applications of Number Theory to Numerical Analysis*, S. K. Zaremba, ed. Academic Press, London, 287–317.
- Dietrich, C. R., and G. N. Newsam (1995). Efficient generation of conditional simulations by Chebyshev matrix polynomial approximations to the symmetric square root of the covariance matrix. *Mathematical Geology*, **27**(2), 207–228.
- Dimitrakopoulos, R., and M. Dagbert (1993). Sequential modelling of relative indicator variables: Dealing with multiple lithology types. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 413–424.
- Dong, A. (1990). *Estimation géostatistique des phénomènes régis par des équations aux dérivées partielles*. Doctoral thesis, E.N.S. des Mines de Paris.
- Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York. Reprinted (1990).
- Dousset, P. E., and L. Sandjivy (1987). Analyse krigeante des données géochimiques multivariées prélevées sur un site stannifère en Malaisie. *Sciences de la Terre, Série Informatique Géologique*, **26**, 1–22.
- Dowd, P. A. (1982). Lognormal kriging—The general case. *Journal of the International Association for Mathematical Geology*, **14**(5), 475–499.
- Dowd, P. A. (1984). The variogram and kriging: robust and resistant estimators. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 91–106.
- Dowd, P. A. (1989). Generalized cross-covariances. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 151–162.
- Dowd, P. A., and A. G. Royle (1977). Geostatistical applications in the Athabasca tar sands. In *Proceedings of the 15th APCOM International Symposium*, Australasian Institute of Mining and Metallurgy, Parkville, Australia, 235–242.
- Dowd, P. A., and C. Saraç (1994). An extension of the LU decomposition method of simulation. In *Geostatistical Simulations*, M. Armstrong and P. A. Dowd, eds. Kluwer, Dordrecht, Netherlands, 23–36.
- Doyen, P. M., D. E. Psaila, and S. Strandenes (1994). Bayesian sequential indicator simulation of channel sands from 3-D seismic data in the Oseberg Field, Norwegian North Sea. *SPE paper 28382*, 69th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, New Orleans, Louisiana, 197–211.
- Doyen, P. M., L. D. den Boer, and W. R. Pillet (1996). Seismic porosity mapping in the Ekofisk field using a new form of collocated cokriging. *SPE paper 36498*, Annual Technical Conference and Exhibition, Denver, Colorado.
- Dubrule, O. (1981). *Krigeage et splines en cartographie automatique. Application à des exemples pétroliers*. Doctoral thesis, E.N.S. des Mines de Paris.
- Dubrule, O. (1983). Two methods with different objectives: Splines and kriging. *Journal of the International Association for Mathematical Geology*, **15**(2), 245–257.
- Dubrule, O. (1993). Introducing more geology in stochastic reservoir modelling. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 351–369.

- Dubrule, O., and C. Kostov (1986). An interpolation method taking into account inequality constraints: I. Methodology. *Mathematical Geology*, **18**(1), 33–51.
- Duchon, J. (1975). Fonctions splines de type plaque mince en dimension 2. *Séminaire d'Analyse Numérique*, N-231, Université Scientifique et Médicale de Grenoble, France.
- Duchon, J. (1976). Fonctions splines et espérances conditionnelles de champs gaussiens. *Ann. Sci. de l'Université de Clermont (France)*, **61**(14), 19–27.
- Duda, R. O. (1982). Kriging with derivative information. Technical Report No. 613, Fairchild, Palo Alto, California.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, **7**(1), 1–26.
- Eide, A. L., H. Omre, and B. Ursin (1997). Stochastic reservoir characterization conditioned on seismic data. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 1, 442–453.
- Elliott, F. W., Jr., and A. J. Majda (1994). A wavelet Monte Carlo method for turbulent diffusion with many spatial scales. *Journal of Computational Physics*, **113**, 82–111.
- Evertsz, C. J. G., and B. B. Mandelbrot (1989). Multifractal measures. In *Chaos and Fractals*, H. O. Peitgen, H. Jürgens, and D. Saupe, eds. Springer, New York, 921–953.
- Farmer, C. L. (1992). Numerical rocks. In *Mathematics of Oil Recovery*, P. R. King, ed. Oxford University Press, Oxford, 437–447.
- Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, 3d ed. Wiley, New York, Vol. 1. First edition (1950).
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, 2d ed. Wiley, New York, Vol. 2. First edition (1966).
- Fenton, G. A., and E. H. Vanmarcke (1990). Simulation of random fields via local average subdivision. *Journal of Engineering Mechanics*, **116**(8), 1733–1749.
- Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, **35**(1), 197–199.
- Flinn, P. A. (1974). Monte Carlo calculation of phase separation in a two-dimensional Ising system. *Journal of Statistical Physics*, **10**(1), 89–97.
- Formery, P. (1964). *Cours de géostatistique*. Ecole des Mines de Paris.
- François-Bongarçon, D. (1993). Geostatistical tools for the determination of fundamental sampling variances and minimum sample masses. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 989–1000.
- Freulon, X. (1994). Conditional simulation of a Gaussian random vector with non linear and/or noisy observations. In *Geostatistical Simulations*, M. Armstrong and P. A. Dowd, eds. Kluwer, Dordrecht, Netherlands, 57–71.
- Freulon, X., and C. de Fouquet (1991). Pratique des bandes tournantes à 3D. In *Compte Rendu des Journées de Géostatistique*, C. de Fouquet, ed. *Cahiers de Géostatistique*, Fasc. 1, Ecole des Mines de Paris, 101–117.
- Freulon, X., and C. de Fouquet (1993). Conditioning a Gaussian model with inequalities. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 201–212.
- Friedman, A. (1975–76). *Stochastic Differential Equations and Applications*. Academic Press, New York, Vols. 1 and 2.

- Froidevaux, R. (1993). Probability field simulation. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 73–73.
- Fry, N. (1979). Random point distributions and strain measurements in rocks. *Tectonophysics*, **60**, 89–105.
- Galli, A., F. Gerdil-Neuillet, and C. Dadou (1984a). Factorial kriging analysis: A substitute to spectral analysis of magnetic data. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 543–557.
- Galli, A., E. Murillo, and J. Thomann (1984b). Dual kriging—Its properties and its uses in direct contouring. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 2, 621–634.
- Galli, A., H. Beucher, G. Le Loc'h, B. Doligez, and Heresim Group (1994). The pros and cons of the truncated Gaussian method. In *Geostatistical Simulations*, M. Armstrong and P. A. Dowd, eds. Kluwer, Dordrecht, Netherlands, 217–233.
- Gallichand, J., D. Marcotte, and S. O. Prasher (1992). Including uncertainty of hydraulic conductivity into drainage design. *Journal of Irrigation and Drainage Engineering*, **118**(5), 744–756.
- Gandin, L. S. (1963). *Ob"ektivnyi analiz meteorologicheskikh polei*. Gidrometeorologicheskoe Izdatel'stvo, Leningrad. Translation (1965): *Objective Analysis of Meteorological Fields*. Israel Program for Scientific Translations, Jerusalem.
- Gautschi, W. (1972). Error function and Fresnel integrals. In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables—Tenth Printing*, M. Abramowitz and I. A. Stegun, eds. Wiley, New York, 295–329.
- Gel'fand, I. M., and G. E. Shilov (1958). *Obobchennye funktsii i deistviia nad nimi*. Editions Technico-Littéraire, Moscow. Translation (1964): *Generalized Functions*. Vol. 1: *Properties and Operations*. Academic Press, New York.
- Gel'fand, I. M., and N. Y. Vilenkin (1961). *Nekotorye primeneniia garmonitsheskovo analiza*. Editions de Littérature de Physique et de Mathématique, Moscow. [Translation (1964): *Generalized Functions*, Vol. 4: *Applications of Harmonic Analysis*. Academic Press, New York.]
- Gel'fand, I. M., M. I. Graev, and N. Y. Vilenkin (1962). *Obobchennye funktsii. Vypusk 5: Integral'naya geometriya svyazannye s nei voprosy teorii predstavlenii*. Moscow. [Translation (1966): *Generalized Functions*, Vol. 5: *Integral Geometry and Representation Theory*. Academic Press, New York.]
- Gelhar, L. W. (1993). *Stochastic Subsurface Hydrology*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Geman, S., and D. Geman (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**(6), 721–741.
- Georgsen, F., and H. Omre (1993). Combining fibre processes and Gaussian random functions for modelling fluvial reservoirs. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 425–439.
- Gervais, F. (1993). *Modélisation géométrique d'un réseau de fractures dans un massif rocheux stratifié. Application aux carrières marbrières de Comblanchien (Côte d'Or, France)*. Doctoral thesis, E.N.S. des Mines de Paris.

- Gikhman, I. I., and A. V. Skorokhod (1972). *Stochastic Differential Equations*. Springer, Berlin.
- Gilbert, R. O. (1987). *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York.
- Gneiting, T. (1998). Closed form solutions of the two-dimensional turning band equation. *Mathematical Geology*, **30**(4), 379–390.
- Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized regression model. *American Statistical Association Journal*, **57**, 369–375.
- Gómez-Hernández, J. J., and E. F. Cassiraga (1994). Theory and practice of sequential simulation. In *Geostatistical Simulations*, M. Armstrong and P. A. Dowd, eds. Kluwer, Dordrecht, Netherlands, 111–124.
- Gómez-Hernández, J. J., and A. G. Journel (1990). Stochastic characterization of grid-block permeabilities: From point values to block tensors. In *2nd European Conference on the Mathematics of Oil Recovery*, D. Guérillot and O. Guillon, eds. Editions Technip, Paris, 83–90.
- Gómez-Hernández, J. J., A. Sahuquillo, and J. Capilla (1997). Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. I. Theory. *Journal of Hydrology*, **203**, 162–174.
- Gonzalez-Casanova, P., and R. Alvarez (1985). Splines in geophysics. *Geophysics*, **50**(12), 2831–2848.
- Goovaerts, P. (1994). Comparative performance of indicator algorithms for modeling conditional probability distribution functions. *Mathematical Geology*, **26**(3), 389–411.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Goovaerts, P., and P. Sonnet (1993). Study of spatial and temporal variations of hydrochemical variables using Factorial Kriging Analysis. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 745–756.
- Goovaerts, P., P. Sonnet, and A. Navarre (1993). Factorial kriging analysis of sprinwater contents in the Dyle River basin, Belgium. *Water Resources Research*, **29**(7), 2115–2125.
- Goulard, M. (1989). Inference in a coregionalization model. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 397–408.
- Goulard, M., and M. Voltz (1992). Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, **24**(3), 269–286.
- Grzebyk, M. (1993). *Ajustement d'une corégionalisation stationnaire*. Doctoral thesis, E.N.S. des Mines de Paris.
- Grzebyk, M., and H. Wackernagel (1994). Multivariate analysis and spatial/temporal scales: Real and complex models. In *Proceedings of XVIIth International Conference, Hamilton, Ontario, August 1994*, Vol. 1, 19–33.
- Guarascio, M. (1976). Improving the uranium deposits estimations (the Novazza case). In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Huijbregts, eds. Reidel, Dordrecht, Holland, 351–367.
- Guardiano, F. B., and R. M. Srivastava (1993). Multivariate geostatistics: Beyond bivariate moments. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 133–144.

- Guertin, K. (1987). A correction model for conditional bias in selective mining operations. *Mathematical Geology*, **19**(5), 407–423.
- Guibal, D., and A. Remacre (1984). Local estimation of the recoverable reserves: Comparing various methods with the reality on a porphyry copper deposit. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 435–448.
- Gutjahr, A. L., and L. W. Gelhar (1981). Stochastic models for subsurface flow: Infinite versus finite domains and stationarity. *Water Resources Research*, **17**(2), 337–350.
- Gutjahr, A. L., L. W. Gelhar, A. A. Bakr, and J. R. MacMillan (1978). Stochastic analysis of spatial variability in subsurface flows. 2. Evaluation and application. *Water Resources Research*, **14**(5), 953–959.
- Gutjahr, A., B. Bullard, S. Hatch, and L. Hughson (1994). Joint conditional simulations and the spectral approach for flow modeling. *Stochastic Hydrology and Hydraulics*, **8**, 79–108.
- Guyon, X. (1993). *Champs aléatoires sur un réseau—Modélisations, statistique et applications*. Masson, Paris. Translation (1995): *Random Fields on a Network—Modeling, Statistics and Applications*. Springer, New York.
- Gy, P. (1975). *Théorie et pratique de l'échantillonnage des matières morcelées*. Editions Pierre Gy, Cannes, France.
- Gy, P. (1979). *Sampling of Particulate Materials—Theory and Practice*. Elsevier, Amsterdam, Netherlands.
- Haas, A. (1993). Simulation de réservoirs pétroliers par inversion géostatistique. *Cahiers de Géostatistique*, Fasc. 3, Ecole des Mines de Paris, 87–99.
- Haas, A., and C. Josselin (1976). Geostatistics in the petroleum industry. In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Huijbregts, eds. Reidel, Dordrecht, Holland, 333–347.
- Haas, A., G. Matheron, and J. Serra (1967). Morphologie mathématique et granulométries en place, I; II. *Annales des Mines*, I: No. 11, 735–753; II: No. 12, 767–782.
- Haas, A., P. Biver, and K. Altisen (1998). Constrained Kriging and simulation of a well layering towards a geophysical thickness map. *ECMOR VI Proceedings, 6th European Conference on Mathematics and Oil Recovery*, Peebles, Scotland, 8–11 September 1998, Paper C-31.
- Hajek, B. (1988). Cooling schedules for optimal annealing. *Mathematics of Operations Research*, **13**(2), 311–329.
- Haldorsen, H. H. (1983). *Reservoir Characterization Procedures for Numerical Simulation*. PhD dissertation, University of Texas, Austin.
- Haldorsen, H. H., and L. W. Lake (1984). A new approach to shale management in field-scale models. *SPEJ*, August 1984, 447–457.
- Halmos, P. R. (1951). *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*. Chelsea, New York.
- Hammersley, J. M., and J. A. Nelder (1955). Sampling from an isotropic Gaussian process. *Proceedings Cambridge Philosophical Society*, **51**, 652–662.
- Handcock, M. S. (1994). Measuring the uncertainty in kriging. In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 436–447.

- Handcock, M. S., and J. R. Wallis (1994). An approach to statistical spatial-temporal modeling of meteorological fields [including comments by N. Cressie, P. Guttorp, C. L. Keppenne, M. D. Dettinger, and M. Ghil, and rejoinder by M. S. Handcock]. *Journal of the American Statistical Association*, **89**(426), 368–390.
- Haslett, J. (1989). Geostatistical neighbourhoods and subset selection. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 569–577.
- Haslett, J. (1992). Spatial data analysis—Challenges. *Statistician*, **41**, 271–284.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1), 97–109.
- Hawkins, D. M., and N. Cressie (1984). Robust kriging—A proposal. *Journal of the International Association for Mathematical Geology*, **16**(1), 3–18.
- Hegstad, B. K., H. Omre, H. Tjelmeland, and K. Tyler (1994). Stochastic simulation and conditioning by annealing in reservoir description. In *Geostatistical Simulations*, M. Armstrong and P. A. Dowd, eds. Kluwer, Dordrecht, Netherlands, 43–55.
- Helterbrand, J. D., and N. Cressie (1994). Universal cokriging under intrinsic coregionalization. *Mathematical Geology*, **26**(2), 205–226.
- Hepp, V., and A. C. Dumestre (1975). CLUSTER, a method for selecting the most probable dip results from dipmeter surveys. *SPE Paper 5543*, 50th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, Dallas, Texas.
- Herzfeld, U. C. (1989a). A note on programs performing kriging with nonnegative weights. *Mathematical Geology*, **21**(3), 391–393.
- Herzfeld, U. C. (1989b). Variography of submarine morphology: Problems of deregularization, and cartographical implications. *Mathematical Geology*, **21**(7), 693–713.
- Hestir, K., J. C. S. Long, J. P. Chilès, and D. Billaux (1989). Some techniques for stochastic modeling of three-dimensional fracture networks. In *Geostatistical, Sensitivity, and Uncertainty Methods for Ground-Water Flow and Radionuclide Transport Modeling*, B. E. Buxton, ed. Battelle Press, Columbus, Ohio, 495–519.
- Hewett, T. A., and R. A. Behrens (1988). Conditional simulation of reservoir heterogeneity with fractals. *SPE paper 18326*, 63rd Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, Houston, Texas, 645–660.
- Hochstrasser, U. W. (1972). Orthogonal polynomials. In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, M. Abramowitz and I. A. Stegun, eds. Wiley, New York, 771–802.
- Hoeksema, R. J., and P. K. Kitanidis (1984). An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. *Water Resources Research*, **20**(7), 1003–1020.
- Hoeksema, R. J., and P. K. Kitanidis (1985). Comparison of Gaussian conditional mean and kriging estimation in the geostatistical solution of the inverse problem. *Water Resources Research*, **21**(6), 825–836.
- Holden, L., and O. Lia (1992). A tensor estimator for the homogenization of absolute permeability. *Transport in Porous Media*, **8**, 37–46.
- Høst, G., H. Omre, and P. Switzer (1995). Spatial interpolation errors for monitoring data. *Journal of the American Statistical Association*, **90**(431), 853–861.
- Hu, L. Y. (1988). *Mise en oeuvre du modèle gamma pour l'estimation des distributions spatiales*. Doctoral thesis, E.N.S. des Mines de Paris.

- Hu, L. Y., and C. Lantuéjoul (1988). Recherche d'une fonction d'anamorphose pour la mise en oeuvre du krigeage disjonctif isofactoriel gamma. *Sciences de la Terre, Série Informatique Géologique*, **28**, 145–173.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73–101.
- Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.
- Huijbregts, C., and G. Matheron (1971). Universal kriging (an optimal method for estimating and contouring in trend surface analysis). *9th International Symposium on Techniques for Decision-Making in the Mineral Industry*, J. I. McGerrigle, ed. Canadian Institute of Mining and Metallurgy, 159–169.
- Isaaks, E. H. (1984). Indicator simulation: Application to the simulation of a high grade uranium mineralization. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 2, 1057–1069.
- Jackson, M., A. Maréchal, et al. (1979). Recoverable reserves estimated by disjunctive kriging: A case study. In *Proceedings of the 16th APCOM International Symposium*, T. J. O'Neil, ed. Society of Mining Engineers of the AIME, New York, 240–249.
- Jacod, J., and P. Joathon (1971). Use of random-genetic models in the study of sedimentary processes. *Journal of the International Association for Mathematical Geology*, **3**(3), 219–233.
- Jacod, J., and P. Joathon (1972). Conditional simulation of sedimentary cycles in three dimensions. In *Proceedings of the International Sedimentary Congress, Heidelberg, August 1971*, D. F. Merriam, ed. Plenum Press, 139–165.
- Jaquet, O. (1989). Factorial kriging analysis applied to geological data from petroleum exploration. *Mathematical Geology*, **21**(7), 683–691.
- Jarrow, R., and A. Rudd (1983). *Options Pricing*. Irwin, Homewood, Illinois.
- Jenkins, G. W., and D. G. Watts (1968). *Spectral Analysis and Its Applications*. Holden-Day, San Francisco, California.
- Jeulin, D. (1979). *Morphologie mathématique et propriétés physiques des agglomérés de minerai de fer et de coke métallurgique*. Doctoral thesis, E.N.S. des Mines de Paris.
- Jeulin, D. (1989). Morphological modeling of images by sequential random functions. *Signal Processing*, **16**, 403–431.
- Jeulin, D. (1991). *Modèles morphologiques de structures aléatoires et de changement d'échelle*. Doctoral thesis, Université de Caen, France.
- Jeulin, D. (1992). Multivariate random image models. *Acta Stereologica*, **11**, Suppl. 1, 59–66.
- Jeulin, D. (1997). Dead leaves models: From space tessellation to random functions. In *Advances in Theory and Applications of Random Sets*, D. Jeulin, ed. World Scientific, Singapore, 137–156.
- Jeulin, D., and P. Jeulin (1981). Synthesis of rough surfaces by random morphological models. *Stereologica Jugoslavia*, **3**, Suppl. 1, 239–246.
- Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer, New York.
- Journel, A. G. (1973). Simulation de gisements miniers. *Industrie Minérale—Mine*, No. 4–73, 221–226.
- Journel, A. G. (1974a). Geostatistics for conditional simulation of orebodies. *Economic Geology*, **69**, 673–687.

- Journel, A. G. (1974b). *Simulations conditionnelles: Théorie et pratique*. Doctoral thesis, Université de Nancy-I, France.
- Journel, A. G. (1980). The lognormal approach to predicting local distributions of selective mining unit grades. *Journal of the International Association for Mathematical Geology*, **12**(4), 285–303.
- Journel, A. G. (1982). The indicator approach to estimation of spatial distributions. In *Proceedings of the 17th APCOM International Symposium*, T. B. Johnson and R. J. Barnes, eds. Society of Mining Engineers of the AIME, New York, 793–806.
- Journel, A. G. (1983). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, **15**(3), 445–468.
- Journel, A. G. (1984). The place of non-parametric geostatistics. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 307–335.
- Journel, A. G. (1989). Imaging of spatial uncertainty: A non-Gaussian approach. In *Geostatistical, Sensitivity, and Uncertainty Methods for Ground-Water Flow and Radionuclide Transport Modeling*, B. E. Buxton, ed. Battelle Press, Columbus, Ohio, 585–599.
- Journel, A. G., and F. Alabert (1989). Non-Gaussian data expansion in the Earth Sciences. *Terra Nova*, **1**(2), 123–134.
- Journel, A. G., and C. V. Deutsch (1993). Entropy and spatial disorder. *Mathematical Geology*, **25**(3), 329–355.
- Journel, A. G., and J. J. Gómez-Hernández (1989). Stochastic imaging of the Wilmington clastic sequence. *SPE paper 19857*, 64th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, San Antonio, Texas, 591–606.
- Journel, A. G., and C. J. Huijbregts (1978). *Mining Geostatistics*. Academic Press, London.
- Journel, A. G., and D. Posa (1990). Characteristic behavior and order relations for indicator variograms. *Mathematical Geology*, **22**(8), 1011–1025.
- Journel, A. G., C. Deutsch, and A. J. Desbarats (1986). Power averaging for block effective permeability. *SPE paper 15128*, 56th California Regional Meeting of the Society of Petroleum Engineers, Oakland, 329–334.
- Jowett, G. H. (1955a). The comparison of means of industrial time series. *Applied Statistics*, **4**, 32–46.
- Jowett, G. H. (1955b). Sampling properties of local statistics in stationary stochastic series. *Biometrika*, **42**, 160–169.
- Jowett, G. H. (1955c). The comparison of means of sets of observations from sections of independent stochastic series. *Journal of the Royal Statistical Society, Series B*, **17**(2), 208–227.
- Kacewicz, M. (1994). “Fuzzy” geostatistics—An integration of qualitative description into spatial analysis. In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 448–463.
- Karlin, S., and J. L. McGregor (1957). The differential equations of birth-and-death processes, and the Stieltjes moment problem. *Transactions of the American Mathematical Society*, **85**, 489–546.
- Karlin, S., and J. McGregor (1960). Classical diffusion processes and total positivity. *Journal of Mathematical Analysis and Applications*, **1**, 163–183.
- Kendall, D. G. (1974). Foundations of a theory of random sets. In *Stochastic Geometry*, E. F. Harding and D. G. Kendall, eds. Wiley, New York, 322–376.

- Kendall, M. G. (1970). *Rank Correlation Methods*, 4th ed. Griffin, London.
- Kendall, M. G., and P. A. P. Moran (1963). *Geometrical Probability*. Griffin, London.
- Khinchin, A. Y. (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, **109**(4), 604–615.
- Kim, Y. C., P. K. Medhi, and A. Arik (1982). Investigation of in-pit ore-waste selection procedures using conditionally simulated orebodies. In *Proceedings of the 17th APCOM International Symposium*, T. B. Johnson and R. J. Barnes, eds. Society of Mining Engineers of the AIME, New York, 121–142.
- Kimeldorf, G., and G. Wahba (1970). A correspondence between Bayesian estimation of stochastic processes and smoothing by splines. *Annals of Mathematical Statistics*, **41**, 495–502.
- Kirkpatrick, S., C. D. Gelatt, Jr., and M. P. Vecchi (1983). Optimization by simulated annealing. *Science*, **220**, 671–680.
- Kitanidis, P. K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research*, **19**(4), 909–921.
- Kitanidis, P. K. (1985). Minimum-variance unbiased quadratic estimation of covariances of regionalized variables. *Journal of the International Association for Mathematical Geology*, **17**(2), 195–208.
- Kitanidis, P. K., and R. W. Lane (1985). Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method. *Journal of Hydrology*, **79**, 53–71.
- Kitanidis, P. K., and E. G. Vomvoris (1983). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resources Research*, **19**(3), 677–690.
- Kleingeld, W. J. (1987). *La géostatistique pour des variables discrètes*. Doctoral thesis, E.N.S. des Mines de Paris.
- Kleingeld, W. J., and C. Lantuéjoul (1993). Sampling of orebodies with a highly dispersed mineralization. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 953–964.
- Kleingeld, W. J., C. Lantuéjoul, C. F. Prins, and M. L. Thurston (1996). The non-conditional simulation of a Cox process with application to a sampling problem. In *Proceedings of the Conference on Mining Geostatistics, Berg-en-dal, Kruger National Park, South Africa, 19–22 September 1994*, Geostatistical Association of South Africa, 26–38.
- Kleingeld, W. J., C. Lantuéjoul, M. L. Thurston, and C. F. Prins (1997). The conditional simulation of a Cox process with application to deposits with discrete particles. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 2, 683–694.
- Knuth, D. E. (1981). *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, 2d ed. Addison-Wesley, Reading, Massachusetts.
- Kolmogorov, A. N. (1940a). Kurven im Hilbertschen Raum die gegenüber eine einparametrischen Gruppe von Bewegungen invariant sind. *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS*, **26**(1), 6–9.
- Kolmogorov, A. N. (1940b). Wiener'sche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS*, **26**(2), 115–118.

- Kolmogorov, A. N. (1941a). The local structure of turbulence in incompressible viscous fluid at very large Reynolds' numbers. *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS*, **30**(4), 301–305. Reprinted (1961) in *Turbulence: Classic Papers on Statistical Theory*, S. K. Friedlander and L. Topping, eds. Interscience Publishers, New York, 151–155.
- Kolmogorov, A. N. (1941b). Dissipation of energy in the locally isotropic turbulence. *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS*, **32**(1), 16–18. Reprinted (1961) in *Turbulence: Classic Papers on Statistical Theory*, S. K. Friedlander and L. Topping, eds. Interscience Publishers, New York, 159–161.
- Kolmogorov, A. N. (1941c). Interpolation und Extrapolation von stationären zufälligen Folgen. *Izvestiia Akademii Nauk SSSR, Seriya Matematicheskaya*, **5**, 3–14.
- Koopmans, L. H. (1974). *The Spectral Analysis of Time Series*. Academic Press, New York.
- Kostov, C., and O. Dubrule (1986). An interpolation method taking into account inequality constraints: II. Practical approach. *Mathematical Geology*, **18**(1), 53–73.
- Krickeberg, K. (1974). Moments of point processes. In *Stochastic Geometry*, E. F. Harding and D. G. Kendall, eds. Wiley, New York, 89–113.
- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, December, 119–139.
- Krige, D. G. (1952). A statistical analysis of some of the borehole values in the Orange Free State goldfield. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, September, 47–64.
- Krige, D. G. (1978). *Lognormal-de Wijsian Geostatistics for Ore Evaluation*. Monograph Series, South African Institute of Mining and Metallurgy, Johannesburg.
- Krige, D. G. (1997). A practical analysis of the effects of spatial structure and of data available and accessed, on conditional biases in ordinary kriging. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 1, 799–810.
- Kruel-Romeu, R., and B. Neetinger (1995). Calculation of internodal transmissivities in finite difference models of flow in heterogeneous porous media. *Water Resources Research*, **31**(4), 943–959.
- Lajaunie, C. (1990). Comparing some approximate methods for building local confidence intervals for predicting regionalized variables. *Mathematical Geology*, **22**(1), 123–144.
- Lajaunie, C., and R. Béjaoui (1991). Sur le krigeage des fonctions complexes. Technical Report N-23/91/G, Centre de Géostatistique, Fontainebleau, France.
- Lajaunie, C., and C. Lantuéjoul (1989). Setting up the general methodology for discrete isofactorial models. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 323–334.
- Lajaunie, C., G. Courrioux, and L. Manuel (1997). Foliation fields and 3D cartography in geology: Principles of a method based on potential interpolation. *Mathematical Geology*, **29**(4), 571–584.
- Landau, L. D., and E. M. Lifshitz (1960). *Electrodynamics of Continuous Media*. Pergamon Press, Oxford.

- Langlais, V. (1989). On the neighborhood search procedure to be used when kriging with constraints. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 603–614.
- Langlais, V. (1990). *Estimation sous contraintes d'inégalités*. Doctoral thesis, E.N.S. des Mines de Paris.
- Langlais, V., and J. Doyle (1993). Comparison of several methods of lithofacies simulation on the fluvial gypsy sandstone of Oklahoma. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 299–310.
- Lantuéjoul, C. (1988). On the importance of choosing a change of support model for global reserves estimation. *Mathematical Geology*, **20**(8), 1001–1019.
- Lantuéjoul, C. (1990). *Cours de sélectivité*. Lecture Notes C-140, Centre de Géostatistique, Fontainebleau, France.
- Lantuéjoul, C. (1991). Ergodicity and integral range. *Journal of Microscopy*, **161**(3), 387–403.
- Lantuéjoul, C. (1993). Substitution random functions. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 37–48.
- Lantuéjoul, C. (1994). Non conditional simulation of stationary isotropic multigaussian random functions. In *Geostatistical Simulations*, M. Armstrong and P. A. Dowd, eds. Kluwer, Dordrecht, Netherlands, 147–177.
- Lantuéjoul, C. (1997a). Iterative algorithms for conditional simulations. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 1, 27–40.
- Lantuéjoul, C. (1997b). Conditional simulation of object-based models. In *Advances in Theory and Applications of Random Sets*, D. Jeulin, ed. World Scientific, Singapore, 271–288.
- Lantuéjoul, C., and J. Rivoirard (1984). Une méthode de détermination d'anamorphose. Technical Report N-432, Centre de Géostatistique, Fontainebleau, France.
- Lantuéjoul, C., and M. Schmitt (1991). Use of two new formulae to estimate the Poisson intensity of a Boolean model. In *Proceedings of 13rd GRETSI Symposium, Juan-les-Pins, France*, Vol. 2, 1045–1048.
- Lasky, S. G. (1950). How tonnage and grade relations help predict ore reserves. *Engineering and Mining Journal*, **151**(4), 81–85.
- Laslett, G. M. (1994). Kriging and splines: An empirical comparison of their predictive performance in some applications. *Journal of the American Statistical Association*, **89**(426), 391–409.
- LaVenue, A. M., B. S. RamaRao, G. de Marsily, and M. G. Marietta (1995). Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. 2. Application. *Water Resources Research*, **31**(3), 495–516.
- Le Loc'h, G., and A. Galli (1997). Truncated plurigaussian method: Theoretical and practical points of view. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 1, 211–222.
- Lia, O., H. Omre, H. Tjelmeland, L. Holden, and T. Egeland (1997). Uncertainties in reservoir production forecasts. *AAPG Bulletin*, **81**(5), 775–802.
- Liao, H. T. (1990). *Estimation des réserves récupérables des gisements d'or—Comparaison entre krigeage disjonctif et krigeage des indicatrices*. Doctoral thesis, Université d'Orléans; Document du BRGM, No. 202 (1991), Bureau de Recherches Géologiques et Minières, Orléans, France.

- Loader, C., and P. Switzer (1992). Spatial covariance estimation for monitoring data. In A. Walden and P. Guttorp, eds. *Statistics in Environmental and Earth Sciences*, Griffin, 52–69.
- Loaiciga, H. A., and M. A. Marino (1990). Error analysis and stochastic differentiability in subsurface flow modeling. *Water Resources Research*, **25**(12), 2897–2902.
- Loiseau, P. (1987). *Etude structurale et géostatistique des gneiss de la région du Cézallier: Modélisation tridimensionnelle de réseaux de fractures; application à l'écoulement des fluides*. Doctoral thesis, Université d'Orléans; Document du BRGM, No. 162 (1988), Bureau de Recherches Géologiques et Minières, Orléans, France.
- Long, J. C. S., and D. M. Billaux (1987). From field data to fracture network modeling: An example incorporating spatial structure. *Water Resources Research*, **23**(7), 1201–1216.
- Luster, G. R. (1986). Homogeneization and proportioning of mined materials: Applications of conditional simulation of coregionalization. In *Proceedings of the 19th APCOM International Symposium*, R. V. Ramani, ed. Society of Mining Engineers, Littleton, Colorado, 163–172.
- Ma, Y. Z., and J. J. Royer (1988). Local geostatistical filtering: application to remote sensing. *Sciences de la Terre, Série Informatique Géologique*, **27**, 17–36.
- Mandelbrot, B. B. (1967). Sporadic random functions and conditional spectral analysis: Self similar examples and limits. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. LeCam and J. Neyman, ed. University of California Press, Berkeley, California, Vol. 3, 155–179.
- Mandelbrot, B. B. (1975a). Fonctions aléatoires pluri-temporelles: approximation poissonienne du cas brownien et généralisations. *Comptes Rendus de l'Académie des Sciences de Paris, Série A*, **280**, 1075–1078.
- Mandelbrot, B. B. (1975b). *Les objets fractals: Forme, hasard, et dimension*. Flammarion, Paris. Third edition (1989).
- Mandelbrot, B. B. (1977). *Fractals: Form, Chance, and Dimension*. Freeman, San Francisco, California.
- Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. Freeman, New York.
- Mandelbrot, B. B. (1985). Self-affine fractals and fractal dimension. *Physica Scripta*, **32**, 257–260.
- Mandelbrot, B. B., and J. W. Van Ness (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, **10**(4), 422–437.
- Mantoglou, A., and J. Wilson (1982). The turning bands method for simulation of random fields using line generation by a spectral method. *Water Resources Research*, **18**(5), 1379–1394.
- Marbeau, J. H., and J. P. Marbeau (1989). Formal computation of integrals for 3-D kriging. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 773–784.
- Marbeau, J. P. (1976). *Géostatistique forestière*. Doctoral thesis, E.N.S. des Mines de Paris.
- Marcotte, D., and M. Chouteau (1993). Gravity data transformation by kriging. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 249–260.
- Marcotte, D., and M. David (1985). The bi-gaussian approach: a simple method for recovery estimation. *Journal of the International Association for Mathematical Geology*, **17**(6), 625–644.

- Maréchal, A. (1970). Cokrigage et régression en corrélation intrinsèque. Technical Report N-205, Centre de Géostatistique, Fontainebleau, France.
- Maréchal, A. (1976). The practice of transfer functions: Numerical methods and their application. In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Huijbregts, eds. Reidel, Dordrecht, Holland, 253–276.
- Maréchal, A. (1982). Local recovery estimation for co-products by disjunctive kriging. In *Proceedings of the 17th APCOM International Symposium*, T. B. Johnson and R. J. Barnes, eds. Society of Mining Engineers of the AIME, New York, 562–571.
- Maréchal, A. (1984). Kriging seismic data in presence of faults. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 385–420.
- Marsaglia, G. (1972). The structure of linear congruential sequences. In *Applications of Number Theory to Numerical Analysis*, S. K. Zaremba, ed. Academic Press, London, 249–285.
- Marshall, R. J., and K. V. Mardia (1985). Minimum norm quadratic estimation of components of spatial covariance. *Journal of the International Association for Mathematical Geology*, **17**(5), 517–525.
- Marsily, G. de (1986). *Quantitative Hydrogeology: Groundwater Hydrology for Engineers*. Academic Press, San Diego, California.
- Marsily, G. de, G. Lavedan, M. Boucher, and G. Fasanino (1984). Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 2, 831–849.
- Massoud, H. (1987). *Modélisation de la petite fracturation par les techniques de la géostatistique*. Doctoral thesis, E.N.S. des Mines de Paris; Document du BRGM, No. 155 (1988), Bureau de Recherches Géologiques et Minières, Orléans, France.
- Matérn, B. (1960). *Spatial Variation—Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*. Meddelanden från Statens Skogsforskningsinstitut, Vol. 49, No 5, Almaenna Foerlaget, Stockholm. Second edition (1986), Springer, Berlin Heidelberg.
- Matheron, G. (1955). Applications des méthodes statistiques à l'évaluation des gisements. *Annales des Mines*, No. 12, 50–75.
- Matheron, G. (1962–63). *Traité de géostatistique appliquée, Tome I; Tome II: Le krigeage*. I: Mémoires du Bureau de Recherches Géologiques et Minières, No. 14 (1962), Editions Technip, Paris; II: Mémoires du Bureau de Recherches Géologiques et Minières, No. 24 (1963), Editions B.R.G.M., Paris.
- Matheron, G. (1965). *Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux Sciences de la Nature*. Masson, Paris.
- Matheron, G. (1967). *Eléments pour une théorie des milieux poreux*. Masson, Paris.
- Matheron, G. (1968a). *Osnovy Prikladnoi Geostatistiki (Treatise of Geostatistics)*. Mir, Moscow.
- Matheron, G. (1968b). Schéma booléen séquentiel de partition aléatoire. Technical Report N-83, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1969a). *Le krigeage universel*. Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fasc. 1, Ecole des Mines de Paris.

- Matheron, G. (1969b). Les processus d'Ambarzoumian et leur application en géologie. Technical Report N-131, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1970). *La théorie des variables régionalisées et ses applications*. Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fasc. 5, Ecole des Mines de Paris. Translation (1971): *The Theory of Regionalized Variables and Its Applications*.
- Matheron, G. (1971a). Les polyèdres poissoniens isotropes. *Actes du 3^e Colloque Européen sur la Fragmentation, Cannes, France*, II-9, 509–534. *Zerkleinern 3. Vortrage Symposium, Cannes*, Dechema Monographien, Vol. 69, No. 1292–1326, Verlag Chemie GmbH, Part 2, 575–600.
- Matheron, G. (1971b). La théorie des fonctions aléatoires intrinsèques généralisées. Note Géostatistique N° 117. Technical Report N-252, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1972a). Ensembles fermés aléatoires, ensembles semi-markoviens et polyèdres poissoniens. *Advances in Applied Probability*, 4, 508–541.
- Matheron, G. (1972b). Inférence statistique pour les covariances généralisées (dans \mathbb{R}^1). Technical Report N-281, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1972c). Les covariances généralisées polynômiales. Technical Report N-299, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1973a). The intrinsic random functions and their applications. *Advances in Applied Probability*, 5, 439–468.
- Matheron, G. (1973b). Le krigeage disjonctif. Technical Report N-360, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1974a). Effet proportionnel et lognormalité ou: le retour du serpent de mer. Technical Report N-374, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1974b). Représentations stationnaires et représentations minimales pour les FAI- k . Technical Report N-377, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1974c). Les fonctions de transfert des petits panneaux. Technical Report N-395, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1975a). *Random Sets and Integral Geometry*. Wiley, New York.
- Matheron, G. (1975b). Compléments sur les modèles isofactoriels. Technical Report N-432, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1975c). Modèles isofactoriels discrets et modèle de Walsh. Technical Report N-449, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1976a). A simple substitute for conditional expectation: The disjunctive kriging. In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Huijbregts, eds. Reidel, Dordrecht, Holland, 221–236.
- Matheron, G. (1976b). Forecasting block grade distributions: the transfer functions. In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Huijbregts, eds. Reidel, Dordrecht, Holland, 237–251.
- Matheron, G. (1978). *Estimer et choisir*. Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fasc. 7, Ecole des Mines de Paris. Translation (1989): *Estimating and Choosing—An Essay on Probability in Practice*. Springer, Berlin.
- Matheron, G. (1979a). Comment traduire les catastrophes. La structure des F.A.I. générales. Technical Report N-617, Centre de Géostatistique, Fontainebleau, France.

- Matheron, G. (1979b). Recherche de simplification dans un problème de cokrigage. Technical Report N-628, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1980). Modèles isofactoriels pour l'effet zéro. Technical Report N-659, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1981a). Splines and kriging: Their formal equivalence. In *Down-to-Earth-Statistics: Solutions Looking for Geological Problems*, D. F. Merriam, ed. *Syracuse University Geological Contributions*, Syracuse, New York, 77–95.
- Matheron, G. (1981b). La sélectivité des distributions. Technical Report N-686, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1981c). Remarques sur le krigeage et son dual. Technical Report N-695, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1981d). Quatre familles discrètes. Technical Report N-703, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1982a). Pour une analyse krigeante des données régionalisées. Technical Report N-732, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1982b). La destructure des hautes teneurs et le krigeage des indicatrices. Technical Report N-761, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1984a). Modèle isofactoriel et changement de support. *Sciences de la Terre, Série Informatique Géologique*, **18**, 71–123.
- Matheron, G. (1984b). The selectivity of the distributions and the “second principle of geostatistics.” In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 421–433.
- Matheron, G. (1984c). Isofactorial models and change of support. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 449–467.
- Matheron, G. (1984d). Changement de support en modèle mosaïque. *Sciences de la Terre, Série Informatique Géologique*, **20**, 435–454.
- Matheron, G. (1984e). Une méthodologie générale pour les modèles isofactoriels discrets. *Sciences de la Terre, Série Informatique Géologique*, **21**, 1–64.
- Matheron, G. (1985). Comparaison de quelques distributions du point de vue de la sélectivité. *Sciences de la Terre, Série Informatique Géologique*, **24**, 1–21.
- Matheron, G. (1986). Sur la positivité des poids de krigeage. Technical Report N-30/86/G, Centre de Géostatistique, Fontainebleau, France.
- Matheron, G. (1987a). Suffit-il, pour une covariance, d'être de type positif? *Sciences de la Terre, Série Informatique Géologique*, **26**, 51–66.
- Matheron, G. (1987b). A simple answer to an elementary question. *Mathematical Geology*, **19**(5), 455–457.
- Matheron, G. (1988). Simulation de fonctions aléatoires admettant un variogramme concave donné. *Sciences de la Terre, Série Informatique Géologique*, **28**, 195–212.
- Matheron, G. (1989). Two classes of isofactorial models. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 309–322.
- Matheron, G. (1993). Une conjecture sur la covariance d'un ensemble aléatoire (Notes prises par E. Castelier). *Cahiers de Géostatistique*, Fasc. 3, Ecole des Mines de Paris, 107–113.

- Matheron, G., H. Beucher, C. de Fouquet, A. Galli, D. Guérillot, and C. Ravenne (1987). Conditional simulation of the geometry of fluvio-deltaic reservoirs. *SPE paper* 16753, 62nd Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, Dallas, Texas, 123–131.
- Matheron, G., H. Beucher, C. de Fouquet, A. Galli, and C. Ravenne (1988). Simulation conditionnelle à trois faciès dans une falaise de la formation du Brent. *Sciences de la Terre, Série Informatique Géologique*, **28**, 213–249.
- McLain, D. H. (1980). Interpolation methods for erroneous data. In *Mathematical Methods in Computer Graphics and Design*, K. W. Brodlie, ed. Academic Press, London, 87–104.
- McLaughlin, D., and L. R. Townley (1996). A reassessment of the groundwater inverse problem. *Water Resources Research*, **32**(5), 1131–1161.
- Mejía, J. M., and I. Rodríguez-Iturbe (1974). On the synthesis of random field sampling from the spectrum: An application to the generation of hydrologic spatial processes. *Water Resources Research*, **10**(4), 705–711.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**(6), 1087–1092.
- Meyer, Y. (1993). *Wavelets. Algorithms and Applications*. Society for Industrial and Applied Mathematics, Philadelphia.
- Meyn, S. P., and R. L. Tweedie (1993). *Markov Chains and Stochastic Stability*. Springer, London.
- Mignolet, M. P., and P. D. Spanos (1992). Simulation of homogeneous two-dimensional random fields: Part I—AR and ARMA models. *Journal of Applied Mechanics*, **59**(2), 260–269.
- Miles, R. E. (1964). Random polygons determined by random lines in a plane, I; II. *Proceedings of the National Academy of Sciences of USA*, I: **52**(4), 901–907; II: **52**(5), 1157–1160.
- Miles, R. E. (1969). Poisson flats in Euclidean spaces. Part I: A finite number of random uniform flats. *Advances in Applied Probability*, **1**, 211–237.
- Miles, R. E. (1970). On the homogeneous planar Poisson point process. *Mathematical Biosciences*, **6**(1), 85–127.
- Miles, R. E. (1971). Poisson flats in Euclidean spaces. Part II: Homogeneous Poisson flats and the complementary theorem. *Advances in Applied Probability*, **3**, 1–43.
- Miles, R. E. (1972). The random division of space. *Special Supplement to Advances in Applied Probability*, 243–266.
- Miles, R. E. (1973). The various aggregates of random polygons determined by random lines in a plane. *Advances in Mathematics*, **10**, 256–290.
- Mizell, S. A., A. L. Gutjahr, and L. W. Gelhar (1982). Stochastic analysis of spatial variability in two-dimensional steady groundwater flow assuming stationary and nonstationary heads. *Water Resources Research*, **18**(4), 1053–1067.
- Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes. *Annals of the Institute of Statistical Mathematics*, **41**(3), 565–581.
- Monestiez, P., and P. Switzer (1991) Semiparametric estimation of nonstationary spatial covariance models by metric multidimensional scaling. SIMS Technical Report No. 165, Department of Statistics and Department of Applied Earth Sciences, Stanford University, California.

- Monin, A. S., and A. M. Yaglom (1965). *Statisticheskaya gidromekhanika—Mekhanika Turbulenosti*. Nauka Press, Moscow. Translation, edition updated (1971, 1975): *Statistical Fluid Mechanics: Mechanics of Turbulence*. MIT Press, Cambridge.
- Morelon, I. F., B. Doligez, D. R. Guerillot, D. Rahon, and Y. Touffait (1991). An application of a 3D geostatistical imaging to reservoir fluid flow simulations. *SPE paper 22312*, Sixth Petroleum Computer Conference of the Society of Petroleum Engineers, Dallas, Texas, 223–231.
- Mostad, P. F., T. Egeland, N. L. Hjort, A. G. Kraggerud, and P. Y. Biver (1997). Variogram estimation in a Bayesian framework. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 1, 223–233.
- Mosteller, F., and J. W. Tukey (1977). *Data Analysis and Regression*. Addison-Wesley, Reading, Massachusetts.
- Muge, F. H., and G. Cabeçadas (1989). A geostatistical approach to eutrophication modelling. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 445–457.
- Murray, C. (1993). Indicator simulation of petrophysical rock types. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 399–411.
- Myers, D. E. (1982). Matrix formulation of co-kriging. *Journal of the International Association for Mathematical Geology*, **14**(3), 249–257.
- Myers, D. E. (1983). Estimation of linear combinations and co-kriging. *Journal of the International Association for Mathematical Geology*, **15**(5), 633–637.
- Narboni, P. (1979). Application de la méthode des variables régionalisées à deux forêts du Gabon. *Bois et Forêts des Tropiques*. No. 188, 47–68.
- Neuman, S. P. (1980). A statistical approach to the inverse problem of aquifer hydrology. 3. Improved solution method and added perspective. *Water Resources Research*, **16**(2), 331–346.
- Neumann, J. von, and I. J. Schoenberg (1941). Fourier integrals and metric geometry. *Transactions of the American Mathematical Society*, **50**, 226–251.
- Neveu, J. (1970). *Bases mathématiques du calcul des probabilités*. Masson, Paris.
- Neyman, J., and E. L. Scott. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society, Series B*, **20**(1), 1–43.
- Neyman, J., and E. L. Scott (1972). Processes of clustering and applications. In *Stochastic Point Processes, Statistical Analysis, Theory and Applications*, P. A. W. Lewis, ed. Wiley, New York, 646–681.
- Obukhov, A. M. (1941). O raspredelenii energii v spektre turbulentnogo potoka (and expanded abstract in German: Über die Energieverteilung im Spektrum des Turbulenzstromes). *Izvestiya Akademii Nauk SSSR, Seriya Geograficheskaya i Geofizicheskaya*, No. 4–5, 454–456.
- Obukhov, A. M. (1949a). Struktura temperaturnogo polya v turbulentnom potoke (Structure of the temperature field in a turbulent flow). *Izvestiya Akademii Nauk SSSR, Seriya Geograficheskaya i Geofizicheskaya*, **13**(1), 58–69.
- Obukhov, A. M. (1949b). Lokalnaya struktura atmosfernoy turbulentnosti (Local structure of atmospheric turbulence). *Doklady Akademii Nauk SSSR*, **67**(4), 643–646.
- Oliver, D. S. (1995). Moving averages for Gaussian simulation in two and three dimensions. *Mathematical Geology*, **27**(8), 939–960.

- Olver, F. W. J. (1972). Bessel functions of integer order. In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, M. Abramowitz and I. A. Stegun, eds. Wiley, New York, 355–433.
- Omre, H. (1987). Bayesian kriging—Merging observations and qualified guesses in kriging. *Mathematical Geology*, **19**(1), 25–39.
- Omre, H., and K. B. Halvorsen (1989). The Bayesian bridge between simple and universal kriging. *Mathematical Geology*, **21**(7) 767–786.
- Omre, H., and H. Tjelmeland (1997). Petroleum geostatistics. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 1, 41–52.
- Oppenheim, A. V., and R. W. Schaffer (1989). *Discrete-Time Signal Processing*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Orfeuil, J. P. (1972). Simulation du Wiener-Lévy et de ses intégrales. Technical Report N-290, Centre de Géostatistique, Fontainebleau, France.
- Ouenes, A., S. Bhagavan, P. H. Bunge, and B. J. Travis (1994). Application of simulated annealing and other global optimization methods to reservoir description: Myths and realities. *SPE paper 28415*, 69th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, New Orleans, Louisiana, 547–561.
- Pan, G., D. Gaard, K. Moss, and T. Heiner (1993). A comparison between cokriging and ordinary kriging: Case study with a polymetallic deposit. *Mathematical Geology*, **25**(3), 377–398.
- Pardo-Iguzquiza, E., and M. Chica-Olmo (1993). The Fourier integral method: An efficient spectral method for simulation of random fields. *Mathematical Geology*, **25**(2), 177–217.
- Parker, H. M. (1984). Trends in geostatistics in the mining industry. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 2, 915–934.
- Parker, H. M. (1991). Statistical treatment of outlier data in epithermal gold deposit reserve estimation. *Mathematical Geology*, **23**(2), 175–199.
- Parker, H. M., A. G. Journel, and W. C. Dixon (1979). The use of conditional lognormal probability distribution for the estimation of open-pit ore reserves in stratabound uranium deposits—A case study. In *Proceedings of the 16th APCOM International Symposium*, T. J. O'Neil, ed. Society of Mining Engineers of the AIME, New York, 133–148.
- Pawlowsky, V., R. A. Olea, and J. C. Davis (1994). Additive logratio estimation of regionalized compositional data: an application to the calculation of oil reserves. In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 371–382.
- Pilz, J. (1994). Robust Bayes linear prediction of regionalized variables. In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 464–475.
- Pilz, J., G. Spoeck, and M. G. Schimek (1997). Taking account of uncertainty in spatial covariance estimation. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 1, 302–313.
- Polubarinova-Kochina, P. Y. (1962). *Theory of Groundwater Movement* (translation). Princeton University Press, Princeton, New Jersey.

- Pólya, G. (1949). Remarks on characteristics functions. In *Proceedings of the (First) Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, ed. University of California Press, Berkeley, 115–123.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Wetterling (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2d ed. Cambridge University Press.
- Preston, C. (1975). Spatial birth-and-death processes. *Bulletin of the International Statistical Institute*, **46**(2), 371–391.
- Préteux, F., and M. Schmitt (1988). Boolean texture analysis and synthesis. In *Image Analysis and Mathematical Morphology: Theoretical Advances*, J. Serra, ed. Academic Press, London, Vol. 2, ch. 18, 377–400.
- Puente, C. E., and R. L. Bras (1986). Disjunctive kriging, universal kriging, or no kriging: Small sample results with simulated fields. *Mathematical Geology*, **18**(3), 287–305.
- Pukelsheim, F. (1994). The three sigma rule. *The American Statistician*, **48**, 88–91.
- Rabbani, M., and P. W. Jones (1991). *Digital Image Compression Techniques*. Tutorial Text Series, SPIE Press, Bellingham, Washington, Vol. TT7.
- RamaRao, B. S., A. M. LaVenue, G. de Marsily, and M. G. Marietta (1995). Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. 1. Theory and computational experiments. *Water Resources Research*, **31**(3), 475–493.
- Rao, C. R. (1970). Estimation of heteroscedastic variances in a linear model. *Journal of the American Statistical Association*, **65**, 161–172.
- Rao, C. R. (1971a). Estimation of variance and covariance components. MINQUE theory. *Journal of Multivariate Analysis*, **1**, 257–275.
- Rao, C. R. (1971b). Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis*, **1**, 445–456.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. Wiley, New York.
- Raspa, G., R. Bruno, and P. Kokkiniotis (1989). An application of disjunctive kriging: Using the negative binomial model with different change of support models. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 935–945.
- Remacre-Zaupa, A. (1984). Conditionnement uniforme. *Sciences de la Terre, Série Informatique Géologique*, **18**, 125–139.
- Remacre-Zaupa, A. (1987). Conditioning by the panel grade for recovery estimation of non-homogeneous orebodies. In *Geostatistical Case Studies*, G. Matheron and M. Armstrong, eds. Reidel, Dordrecht, Holland, 135–147.
- Renard, D. (1983). Smoothing discontinuities of kriging with moving neighborhoods. Technical Report N-837, Centre de Géostatistique, Fontainebleau, France.
- Renard, D., and P. Ruffo (1993). Depth, dip and gradient. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 167–178.
- Renard, D., and J. D. Yancey (1984). Smoothing discontinuities when “extrapolating” using moving neighborhood. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 2, 679–690.
- Renard, P., and G. de Marsily (1997). Calculating equivalent permeability: A review. *Advances in Water Resources*, **20**(5–6), 253–278.

- Rendu, J. M. (1979). Kriging, logarithmic kriging, and conditional expectation: Comparison of theory with actual results. In *Proceedings of the 16th APCOM International Symposium*, T. J. O'Neil, ed. Society of Mining Engineers of the AIME, New York, 199–212.
- Rendu, J. M. (1984). Geostatistical modelling and geological controls. In *Proceedings of the 18th APCOM International Symposium*, Institution of Mining and Metallurgy, London, 467–476.
- Rendu, J. M., and L. Readdy (1980). Geology and the semivariogram—A critical relationship. In *Proceedings of the 17th APCOM International Symposium*, T. B. Johnson and R. J. Barnes, eds. Society of Mining Engineers of the AIME, New York, 771–783.
- Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, **13**, 255–266.
- Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society, Series B*, **39**(2), 172–212.
- Ripley, B. D. (1981). *Spatial Statistics*. Wiley, New York.
- Ripley, B. D. (1987). *Stochastic Simulation*. Wiley, New York.
- Rivoirard, J. (1984). *Le comportement des poids de krigeage*. Doctoral thesis, E.N.S. des Mines de Paris.
- Rivoirard, J. (1987). Two key parameters when choosing the kriging neighborhood. *Mathematical Geology*, **19**(8), 851–856.
- Rivoirard, J. (1988). Modèles à résidus d'indicatrices autokrigeables. *Sciences de la Terre, Série Informatique Géologique*, **28**, 303–326.
- Rivoirard, J. (1989). Models with orthogonal indicator residuals. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 91–107.
- Rivoirard, J. (1990). A review of lognormal estimators for *in situ* reserves. *Mathematical Geology*, **22**(2), 213–221.
- Rivoirard, J. (1994). *Introduction to Disjunctive Kriging and Non-linear Geostatistics*. Oxford University Press, Oxford.
- Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, New Jersey.
- Rodríguez-Iturbe, I., and J. M. Mejía (1974). The design of rainfall networks in time and space. *Water Resources Research*, **10**(4), 713–728.
- Rossi, M. E., and H. M. Parker (1994). Estimating recoverable reserves: is it hopeless? In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 259–276.
- Rossi, R. E., D. J. Mulla, A. G. Journel, and E. H. Franz (1992). Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, **62**(2), 277–314.
- Roth, C. (1995). *Contribution de la géostatistique à la résolution du problème inverse en hydrogéologie*. Doctoral thesis, E.N.S. des Mines de Paris; Document du BRGM, No. 241, Bureau de Recherches Géologiques et Minières, Orléans, France.
- Roth, C., J. P. Chilès, and C. de Fouquet (1996). Adapting geostatistical transmissivity simulations to finite differences flow simulators. *Water Resources Research*, **32**(10), 3237–3242.

- Roth, C., C. de Fouquet, J. P. Chilès, and G. Matheron (1997). Geostatistics applied to hydrogeology's inverse problem: taking boundary conditions into account. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 2, 1085–1097.
- Rouhani, S., and T. J. Hall (1989). Space-time kriging of groundwater data. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 639–651.
- Royden, H. L. (1968). *Real Analysis*. Collier-Macmillan, London.
- Royer, J. F. (1975). Comparaison des méthodes d'analyse objective par interpolation optimale et par approximations successives. Technical Report No. 365, Etablissement d'Etudes et de Recherches Météorologiques, Direction de la Météorologie, Paris.
- Royer, J. J. (1988). Comparaison de quelques méthodes de déconvolution. Apport de la géostatistique. *Sciences de la Terre, Série Informatique Géologique*, **28**, 327–354.
- Rubin, Y., and G. Dagan (1987). Stochastic identification of transmissivity and effective recharge in steady groundwater flow: 1. Theory. *Water Resources Research*, **23**(7), 1185–1192.
- Rubin, Y., and G. Dagan (1992). Conditional estimation of solute travel time in heterogeneous formations: Impact of transmissivity measurements. *Water Resources Research*, **28**(4), 1033–1040.
- Rubin, Y., and J. J. Gómez-Hernández (1990). A stochastic approach to the problem of upscaling of conductivity in disordered media: Theory and unconditional numerical simulations. *Water Resources Research*, **26**(4), 691–701.
- Rutherford, I. D. (1972). Data assimilation by statistical interpolation of forecast error fields. *Journal of Atmospheric Sciences*, July 1972, 809–815.
- Sabourin, R. (1976). Application of two methods for an interpretation of the underlying variogram. In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Huijbregts, eds. Reidel, Dordrecht, Holland, 101–109.
- Samaras, E., M. Shinozuka, and A. Tsurui (1985). ARMA representation of random processes. *Journal of Engineering Mechanics*, **111**(3), 449–461.
- Sampson, P., and P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, **87**(417), 108–119.
- Sánchez-Vila, X., J. P. Girardi, and J. Carrera (1995). A synthesis of approaches to upscaling of hydraulic conductivities. *Water Resources Research*, **31**(4), 867–882.
- Sandjiv, L. (1984). The factorial kriging analysis of regionalized data: Its applications to geochemical prospecting. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 559–571.
- Sans, H., and J. R. Blaise (1987). Comparing estimated uranium grades with production figures. In *Geostatistical Case Studies*, G. Matheron and M. Armstrong, eds. Reidel, Dordrecht, Holland, 169–185.
- Santaló, L. A. (1976). *Integral Geometry and Geometric Probability*. Addison-Wesley, Reading, Massachusetts.
- Saporta, G. (1990). *Probabilités, analyse des données et statistique*. Editions Technip, Paris.
- Sarmanov, I. O. (1968). A generalized symmetric gamma correlation. *Soviet Mathematics, Doklady*, **9**(2), 547–550.

- Schlatter, T. W. (1975). Some experiments with a multivariate statistical objective analysis scheme. *Monthly Weather Review*, **103**, 246–257.
- Schneebeli, G. (1966). *Hydraulique souterraine*. Eyrolles, Paris.
- Schoenberg, I. J. (1938a). Metric spaces and completely monotone functions. *Annals of Mathematics*, **39**(4), 811–831.
- Schoenberg, I. J. (1938b). Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, **44**, 522–536.
- Schwartz, L. (1950–1951). *Théorie des distributions*. Publications de l'Institut de Mathématique de l'Université de Strasbourg, 2 Vols. New edition (1966): Hermann, Paris.
- Schwydler, M. I. (1963a). O sriednikh kharakteristikakh filtracionnikh tetchenii v sredakh so sluchajnymi neodnorodnostiami (Mean characteristics of flow in random heterogeneous media). *Izvestiia Akademii Nauk SSSR, Seriya Mekhanika i Mashinostroenie*, No. 4, 127–129.
- Schwydler, M. I. (1963b). O gidrodinamitsheskikh rastchetakh filtracionnikh tetchenii v neodnorodnikh poristyx sredakh (Hydrodynamic calculations of flow in heterogeneous porous media). *VNIi, Dobyca Nefti, Teorija i Praktika*, Annuaire 1063, 107–118.
- Séguret, S. (1989). Filtering periodic noise by using trigonometric kriging. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 481–491.
- Séguret, S. (1991). *Géostatistique des phénomènes à tendance périodique*. Doctoral thesis, E.N.S. des Mines de Paris.
- Séguret, S., and P. Huchon (1990). Trigonometric kriging: A new method for removing the diurnal variation from geomagnetic data. *Journal of Geophysical Research*, **95**(B13), 21,383–21,397.
- Sellan, F. (1995). Synthèse de mouvements browniens fractionnaires à l'aide de la transformation par ondelettes. *Comptes Rendus de l'Académie des Sciences de Paris, Série I*, **321**(3), 351–358.
- Selley, R. C. (1976). *An Introduction to Sedimentology*. Academic Press, London.
- Serra, J. (1967). Echantillonnage et estimation locale des phénomènes de transition miniers. Doctoral thesis, Université de Nancy, France.
- Serra, J. (1968). Les structures gigognes: Morphologie mathématique et interprétation métallogénique. *Mineralium Deposita*, **3**, 135–154.
- Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London.
- Serra, J. (1988). Boolean random functions. In *Image Analysis and Mathematical Morphology: Theoretical Advances*, J. Serra, ed. Academic Press, London, Vol. 2, ch. 15, 317–342.
- Serra, O. (1985). *Sedimentary Environments from Wireline Logs*. Schlumberger, Paris, New York.
- Sharp, W. E., and L. A. Aroian (1984). A rapid method for the computer simulation of geological parameters on a square net. *Sciences de la Terre, Série Informatique Géologique*, **20**(2), 495–511.
- Shinozuka, M. (1971). Simulation of multivariate and multidimensional random processes. *The Journal of the Acoustical Society of America*, **49**(1), Part 2, 357–367.

- Shinozuka, M., and C. M. Jan (1972). Digital simulation of random processes and its applications. *Journal of Sound and Vibration*, **25**(1), 111–128.
- Sichel, H. S. (1973). Statistical valuation of diamondiferous deposits. In *Proceedings of the 11th APCOM International Symposium*, M. D. G. Salamon and F. H. Lancaster, eds. The South African Institute of Mining and Metallurgy, Johannesburg, South Africa, 17–25.
- Simard, R. (1980). *Etude de la représentativité des levés gravimétriques et de son influence sur l'interprétation*. Doctoral thesis, Université de Lausanne, Switzerland.
- Sinclair, A. J., and M. Vallée (1994). Improved sampling control and data gathering for improved mineral inventories and production control. In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 323–329.
- Singleton, R. C. (1979). Two-dimensional mixed radix mass storage Fourier transform. In *Programs for Digital Signal Processing*, IEEE Press, Institute of Electrical and Electronics Engineers, New York, 1.9-1–1.9-8.
- Smith, L., and R. A. Freeze (1979a). Stochastic analysis of steady state groundwater flow in a bounded domain. 1. One-dimensional simulations. *Water Resources Research*, **15**(3), 521–528.
- Smith, L., and R. A. Freeze (1979b). Stochastic analysis of steady state groundwater flow in a bounded domain. 2. Two-dimensional simulations. *Water Resources Research*, **15**(6), 1543–1559.
- Snowden, V. (1994). Improving predictions by studying reality. In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 330–337.
- Sobczyk, K. (1991). *Stochastic Differential Equations. With Applications to Physics and Engineering*. Kluwer, Dordrecht, Netherlands.
- Solow, A. R. (1985). Bootstrapping correlated data. *Journal of the International Association for Mathematical Geology*, **17**(7), 769–775.
- Spanos, P. D., and M. P. Mignolet (1992). Simulation of homogeneous two-dimensional random fields: Part II—MA and ARMA models. *Journal of Applied Mechanics*, **59**(2), 270–277.
- Spector, A., and B. K. Bhattacharyya (1966). Energy density spectrum and autocorrelation function of anomalies due to simple magnetic models. *Geophysical Prospecting*, **14**(3), 242–272.
- Spector, A., and F. S. Grant (1970). Statistical models for interpreting aeromagnetic data. *Geophysics*, **35**(2), 293–302.
- Srivastava, R. M. (1992). Reservoir characterization with probability field simulation. *SPE paper 24753*, 67th Annual Technical Conference and Exhibition, 927–938.
- Srivastava, R. M. (1994). An annealing procedure for honouring change of support statistics in conditional simulation. In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 277–290.
- Starks, T. H., and J. M. Fang (1982). On the estimation of the generalized covariance function. *Journal of the International Association for Mathematical Geology*, **14**(1), 57–64.
- Steffens, F. E. (1993). Geostatistical estimation of animal abundance in the Kruger National Park, South Africa. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 887–897.

- Stein, M. L. (1986a). A modification of minimum norm quadratic estimation of a generalized covariance function for use with large data sets. *Mathematical Geology*, **18**(7), 625–633.
- Stein, M. L. (1986b). A simple model for spatial-temporal processes. *Water Resources Research*, **22**(13), 2107–2110.
- Stein, M. L. (1989). The loss of efficiency in kriging prediction caused by misspecification of the covariance structure. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 273–282.
- Stoyan, D., W. S. Kendall, and J. Mecke (1987). *Stochastic Geometry and Its Applications*. Wiley, New York.
- Sullivan, J. (1984). Conditional recovery estimation through probability kriging—Theory and practice. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 365–384.
- Sun, N. Z., and W. W. G. Yeh (1992). A stochastic inverse solution for transient groundwater flow: Parameter identification and reliability analysis. *Water Resources Research*, **28**(12), 3269–3280.
- Suppe, J. (1985). *Principles of Structural Geology*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Switzer, P. (1977). Estimation of spatial distributions from point sources with application to air pollution measurement. *Bulletin of the International Statistical Institute*, **47**(2), 123–137.
- Switzer, P. (1984). Inference for spatial autocorrelation functions. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 127–140.
- Switzer, P. (1989). Non-stationary spatial covariances estimated from monitoring data. In *Geostatistics*, M. Armstrong, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 127–138.
- Switzer, P. (1993). The spatial variability of prediction errors. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 1, 261–272.
- Switzer, P., and H. Parker (1976). The problem of ore versus waste discrimination. In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Huijbregts, eds. Reidel, Dordrecht, Holland, 203–218.
- Switzer, P., and K. Sølna (1996). Time-trend estimation for a geographic region. *Journal of the American Statistical Association*, **91**(434), 577–589.
- Switzer, P., and H. Xiao (1988). Approximate prediction intervals for spatial interpolation using bivariate information: An experiment. Technical Report No. 118, Department of Statistics and Department of Applied Earth Sciences, Stanford University, California.
- Szegö, G. (1939). *Orthogonal Polynomials*. Colloquium Publications, American Mathematical Society, Providence, Rhode Island. Fourth edition (1975).
- Tarantola, A. (1987). *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, Amsterdam.
- Tarantola, A., and B. Valette (1982). Generalized non-linear inverse problems solved using the least-squares criterion. *Review of Geophysics and Space Physics*, **20**(2), 219–232.

- Taylor, H. D. (1982). Modeling an orebody having lognormal grade distribution. In *Proceedings of the 17th APCOM International Symposium*, T. B. Johnson and R. J. Barnes, eds. Society of Mining Engineers of the AIME, New York, 784–788.
- Townley, L. R., and J. L. Wilson (1985). Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models of ground-water flow. *Water Resources Research*, **21**(12), 1851–1860.
- Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics*, I. Olkin, ed. Stanford University Press, Stanford, California, 448–485.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.
- Turcotte, D. L. (1986). A fractal approach to the relationship between ore grade and tonnage. *Economic Geology*, **81**, 1528–1532.
- Ver Hoef, J. M., and N. Cressie (1993). Multivariable spatial prediction. *Mathematical Geology*, **25**(2), 219–240.
- Verly, G. (1983). The multigaussian approach and its applications to the estimation of local reserves. *Journal of the International Association for Mathematical Geology*, **15**(2), 259–286.
- Verly, G. (1984). The block distribution given a point multivariate normal distribution. In *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, 495–515.
- Verly, G. (1986). Multigaussian kriging—A complete case study. In *Proceedings of the 19th APCOM International Symposium*, R. V. Ramani, ed. Society of Mining Engineers, Littleton, Colorado, 283–298.
- Vieira, J. L., J. F. Travassos, A. Gonçalves, N. Santos, and F. H. Muge (1993). Simulation of medium and short term mine production scheduling using geostatistical modelling. In *Geostatistics Tróia '92*, A. Soares, ed. Kluwer, Dordrecht, Netherlands, Vol. 2, 1013–1027.
- Vistelius, A. B. (1960). The skew frequency distributions and the fundamental law of the geochemical processes. *The Journal of Geology*, **68**(1), 1–22.
- Voss, R. F. (1985). Random fractal forgeries. In *Fundamental Algorithms for Computer Graphics*, R. A. Earnshaw, ed. Springer, Berlin Heidelberg, 13–16 and 805–835.
- Wackernagel, H. (1985). *L'inférence d'un modèle linéaire en géostatistique multivariable*. Doctoral thesis, E.N.S. des Mines de Paris.
- Wackernagel, H. (1988). Geostatistical techniques for interpreting multivariate spatial information. In C. F. Chung et al., eds. *Quantitative Analysis of Mineral and Energy Resources*, Reidel, Dordrecht, Holland, 393–409.
- Wackernagel, H. (1995). *Multivariate Geostatistics—An Introduction with Applications*. Springer, Berlin.
- Wackernagel, H., R. Webster, and M. A. Oliver (1988). A geostatistical method for segmenting multivariate sequences of soil data. In *Classification and Related Methods of Data Analysis*, H. H. Bock, ed. Elsevier, New York, 641–650.
- Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.
- Watson, G. S. (1984). Smoothing and interpolation by kriging and with splines. *Journal of the International Association for Mathematical Geology*, **16**(6), 601–615.

- Webster, R., and M. A. Oliver (1989). Optimal interpolation and isarithmic mapping of soil properties. VI. Disjunctive kriging and mapping the conditional probability. *Journal of Soil Science*, **40**, 497–512.
- Webster, R., and M. A. Oliver (1990). *Statistical Methods in Soil and Land Resource Survey*. Oxford University Press, Oxford.
- Wen, R., and R. Sinding-Larsen (1997). Stochastic modeling and simulation of small faults by marked point processes and kriging. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 1, 398–414.
- White, C. D., and R. N. Horne (1987). Computing absolute transmissibility in the presence of fine-scale heterogeneity. *SPE paper 16011*, Ninth SPE Symposium on Reservoir Simulation, San Antonio, Texas, 209–220.
- Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, **41**, 434–449.
- Wiener, N. (1942). *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series. With Engineering Applications*. Report of the Services 19, Research Project DIC-6037, MIT, Cambridge, Massachusetts. Printed in book form (1949): Wiley, New York.
- Wiener, O. (1912). Theorie des Mischkörpers für das Feld der stationären Strömung. *Abhandlungen der Mathematisch Physischen Klasse der Königlichen Sächsischen Gesellschaft der Wissenschaften*, **32**, 509–604.
- Wijs, H. J. de (1951). Statistics of ore distribution. Part I: Frequency distribution of assay values. *Geologie en Mijnbouw (Journal of the Royal Netherlands Geological and Mining Society)*, *New Series*, **13(11)**, 365–375.
- Wijs, H. J. de (1953). Statistics of ore distribution. Part II: Theory of binomial distribution applied to sampling and engineering problems. *Geologie en Mijnbouw (Journal of the Royal Netherlands Geological and Mining Society)*, *New Series*, **15(1)**, 12–24.
- Wijs, H. J. de (1976). Models for the estimation of world ore reserves. *Geologie en Mijnbouw*, **55(1–2)**, 46–50.
- Wilkinson, J. H. (1965). *The Algebraic Eigenvalue Problem*. Oxford University Press, Oxford.
- Wood, G., M. A. Oliver, and R. Webster (1990). Estimating soil salinity by disjunctive kriging. *Soil Use and Management*, **6(3)**, 97–104.
- Xu, W., T. T. Tran, R. M. Srivastava, and A. G. Journel (1992). Integrating seismic data in reservoir modeling: The collocated cokriging alternative. *SPE paper 24742*, 67th Annual Technical Conference and Exhibition.
- Yaglom, A. M. (1957). Nekotorye klassy sluchajnyh polej v n-mernom prostranstve, rodstvennye stacionarnym sluchajnym processam. *Teorija Verojatnostej i ee Prime-nenija*, **2(3)**, 292–338. Translation (1957): Some classes of random fields in n-dimensional space, related to stationary random processes. *Theory of Probability and Its Applications*, **2(3)**, 273–320.
- Yaglom, A. M. (1962). *An Introduction to the Theory of Stationary Random Functions*, rev. ed. Prentice-Hall, Englewood Cliffs, New Jersey. Reprint (1973), Dover, New York.
- Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions*, Vol. I: *Basic Results*; Vol. II: *Supplementary Notes and References*. Springer, New York.

- Yaglom, A. M., and M. S. Pinsky (1953). Random processes with stationary increments of order n . *Doklady Akademii Nauk SSSR*, **90**(5), 731–734.
- Yao, T., and T. Mukerji (1997). Application of factorial kriging to improve seismic data integration. In *Geostatistics Wollongong '96*, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, Netherlands, Vol. 1, 350–361.
- Yarus, J. M., and R. L. Chambers, eds. (1994). *Stochastic Modeling and Geostatistics. Principles, Methods, and Case Studies*. American Association of Petroleum Geologists, Tulsa, Oklahoma.
- Yates, S. R., and M. V. Yates (1988). Disjunctive kriging as an approach to management decision making. *Soil Science Society of America Journal*, **52**(6), 1554–1558.
- Yates, S. R., A. W. Warrick, and D. E. Myers (1986). Disjunctive kriging. 1. Overview of estimation and conditional probability. 2. Examples. *Water Resources Research*, **22**(5), 1: 615–621; 2: 623–630.
- Yeh, T. C. J., A. L. Gutjahr, and M. Jin (1995). An iterative cokriging-like technique for ground-water flow modeling. *Ground Water*, **33**(1), 33–41.
- Yfantis, E. A., M. Au, and F. S. Makri (1994). Image compression and kriging (including comments by P. Delfiner, and reply by E. A. Yfantis). In *Geostatistics for the Next Century*, R. Dimitrakopoulos, ed. Kluwer, Dordrecht, Netherlands, 156–170.
- Young, D. S. (1982). Development and application of disjunctive kriging model: Discrete gaussian model. In *Proceedings of the 17th APCOM International Symposium*, T. B. Johnson and R. J. Barnes, eds. Society of Mining Engineers of the AIME, New York, 544–561.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, **8**, 338–353.
- Zadeh, L. A., and J. R. Ragazzini (1950). An extension of Wiener's theory of prediction. *Journal of Applied Physics*, **21**, 645–655.
- Zeldin, B., and P. Spanos (1995). Random field simulation using wavelet bases. In *Applications of Statistics and Probability—Civil Engineering Reliability and Risk Analysis*, M. Lemaire, J. L. Favre, and A. Mébarki, eds. Balkema, Rotterdam, Netherlands, Vol. 2, 1275–1283.
- Zelen, M., and N. C. Severo (1972). Probability functions. In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, M. Abramowitz and I. A. Stegun, eds. Wiley, New York, 925–995.
- Zimmerman, D. A., G. de Marsily, C. A. Gotway, M. G. Marietta, C. L. Axness, R. Beauheim, R. Bras, J. Carrera, G. Dagan, P. B. Davies, D. P. Gallegos, A. Galli, J. Gómez-Hernández, P. Grindrod, A. L. Gutjahr, P. K. Kitanidis, A. M. Lavenue, D. McLaughlin, S. P. Neuman, B. S. Ramarao, C. Ravenne, and Y. Rubin (1998). A comparison of seven geostatistically-based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. *Water Resources Research*, **34**(6), 1373–1413.
- Zimmerman, D. L. (1989). Computationally efficient restricted maximum likelihood estimation of generalized covariance functions. *Mathematical Geology*, **21**(7), 655–672.
- Zimmermann, H. J. (1985). *Fuzzy Set Theory and Its Applications*. Kluwer-Nijhoff, Dordrecht, Netherlands.

Index

- Accumulation, 149, 200
- Additive renormalization, 252
- Additivity
 - dispersion variances, 132
 - kriging and drift, 182–183, 190, 207, 310
- Affine correction, *see* Change-of-support models
- Agriculture, 382, 419
- Aliasing, 499, 501
- Allowable linear combination, 60, 169
 - of order k (ALC- k), 236–242
 - variance and covariance, 61, 62
- Alternating
 - process, 99, 149
 - random set, 541
- Ambartsumian process, 489
- Anamorphosis, 381. *See also* Transformation function
- Anisotropy, 35, 52, 93–97, 107, 123, 571
 - geometric, 93–95, 363, 571
 - other models, 96–97, 476
 - zonal (stratified), 95–96, 156, 364, 571
- Annealing, *see* Simulated annealing
- Area of influence, 48
- ARIMA, *see* Autoregressive
- ARMA, *see* Autoregressive
- Asynchronous data, 366–368
- Atom, 637
- Autoregressive
 - integrated moving-average (ARIMA), 249
 - moving-average (ARMA), 205, 249, 483–484
 - process, 115, 479–481
- Barker's algorithm, 528
- Bathymetry, 80, 210, 217, 449
- Bayesian
 - conditional simulation, 568–571
 - covariance estimation, 110, 190
 - kriging, *see* Kriging
- Bessel
 - covariances, *see* Variogram models
 - functions, 638
- Beta distribution, 403
- Bi-gamma distribution, 404, 443, 448
- Bi-gaussian
 - distribution, 399
 - random function, 17
- Bias
 - in sample covariance, 33
 - in variogram of residuals, 121, 124
- Bilinear model of coregionalization, 350
- Bilognormal, 103, 645
- Binomial isofactorial model, 403
- Birth-and-death process, 402, 560
 - spatial, 560
- Bivariate distribution, *see* Isofactorial models
- Black-Scholes model, 229
- Block average, 195
 - distribution, *see* Change of support, Change-of-support models
 - estimation, 128–131, 197
 - kriging, 195–196, 197
 - simulation, 572–573
- BLUEPACK, 283
- Bochner's theorem, 64
- Boolean
 - objects, 545
 - random function, 401, 548–549, 561
 - random set, 545–548, 558–560
- Bootstrap, 110
- Borel set, 636

- Boundary conditions, 600, 610, 622–628, 634
- Brownian motion, 89, 172, 231, 291, 507–510, 592, 635. *See also* Brownian random function
 - fractional, 89, 259, 592
 - integral of, 249, 518
 - representation of, 257, 268
- Brownian random function, 236
 - fractional, 89, 149, 510
- Cardinal-sine covariance, *see* Variogram models
- Cartier's relation, 426
- Categorical variable, 458, 520–561
- Cauchy
 - algorithm, 63
 - covariance, *see* Variogram models
- Change of support, 75–77, 376
 - by disjunctive kriging, 445
 - by indicator kriging, 436
 - by maximum, 549
- Change-of-support models
 - affine correction, 431, 436
 - isofactorial models, 437–444. *See also* Discrete change-of-support models
 - mosaic, 433
 - multi-Gaussian, 435, 446
 - permanence of lognormality, 433, 436
- Channel tunnel, 215–224
- Charlier polynomials, 403
- Cholesky decomposition, 169, 464
- Choquet capacity, 546
- Circular covariance, *see* Variogram models
- Civil engineering, 128, 215–224, 578
- Cluster process, 553, 554
- Clustered data, 45, 378
- Codispersion, 330, 334
- Coherency spectrum, 328
- Cokriging
 - and inverse problems, 610, 630
 - and regression, 309–311, 339
 - collocated, 305–307, 360
 - disjunctive, 419
 - indicator, 386–387, 419
 - indicator-rank, 385–386
 - of errors, *see* Filtering
 - ordinary, 301, 302, 307
 - simple, 296–298
 - universal, 298–302
 - with nugget effect, 307
- Cokriging equations
 - simple, 296
 - universal, 299
 - with cross-variograms, 330
 - with derivatives, 315
- Cokriging variance, 296, 299, 305
- Cokriging weights, 300, 301, 613, 620
- Collocated cokriging, *see* Cokriging
- Completely monotone covariance, 70, 148, 225, 493
- Compositional data, 350
- Concave variograms, 68, 514, 537
- Conditional distribution, 378
 - approximation to, 379, 394
- Conditional expectation
 - as ideal estimator, 154, 191, 229, 380–382
 - defined, 14
- Conditional simulation, 452–453
- Conditioning
 - by indicator values or inequalities, 521–525, 530–531
 - by kriging, 465–468
 - histogram, 469–471
- Confidence interval, 164, 177, 191, 214, 220, 379
- Conformal transformation, 626
- Connectivity, 520, 533, 592, 602
- Consistency
 - change-of-support models, 430
 - variogram models, 97–104, 107
- Continuity
 - in the mean square, 13, 58
 - of realizations, 58, 148
- Contour mapping, 8, 207, 212
- Conventional income, *see* Selectivity curves
- Convex covariances, 66, 537
- Correlation length, 74, 148
- Correlogram, 30
- Cosimulation, 573–574, 629, 630
- Covariance function, 13, 30
 - isotropic, 31, 69
 - spectral representation, *see* Spectral representation
- Covariance function matrix, 325
- Covariance matrix decomposition, 464–465, 511
- Covariance models, *see* Variogram models
- Covariogram, 25, 70–73
 - geometric, 71
- Cox process, 550–553, 554
- Cross-covariance, 321–328
 - of indicators, 102, 387, 399
- Cross-spectral measure, 325
- Cross-validation, 111–113, 209, 215, 275, 284–285
- Cross-variogram, 328–330, 332
- Cubic
 - covariance, *see* Variogram models

- generalized covariance, *see* Generalized covariance models: polynomial
- Curvilinear distance, 45
- Darcy's law, 594–596, 616, 634
- De Wijs
 - formula, 91
 - variogram, *see* Variogram models
- Dead-leaves
 - random function, 549, 561
 - tessellation, 549
- Derivative, 58
 - covariance of, 59, 316
 - directional, 314
 - estimation of, 314–315
 - estimation with, 315–318
 - physical validity, 318–319
- Design, *see* Network design, Sampling design
- Destructuring, 401, 402, 404, 412
- Differentiability
 - in the mean square, 58, 68, 72, 73, 85, 160, 260
 - of realizations, 59
- Diffusive random function, 402, 459, 542
- Dilation, 592
- Dilation simulation method, 487–494, 511–513
- Dip, 313, 318
- Dirac measure (delta-function), 637
- Discrete change-of-support models
 - gamma, 443
 - Gaussian, 432–433, 435, 438–439, 445–446, 573
 - Hermite, 442
 - Laguerre, 448
- Disjunctive cokriging, 419
- Disjunctive kriging, 389–394
 - lognormal, 416
 - practice, 413–416
 - with change of support, *see* Change of support
 - with unknown mean, 417, 446
- Disjunctive kriging equations
 - general case, 390
 - with isofactorial model, 392
- Disjunctive kriging variance, 391, 394
- Dispersion
 - absolute, 91
 - indicator, 407, 409, 424
 - variance, 91, 131–132
- Distribution
 - empirical, 377–378, 381, 408
 - estimation, 378–394. *See also* Change of support, Change-of-support models
 - regional, 376
- Doubly stochastic Poisson process, *see* Cox process
- Drift, 57, 115–128
 - defined, 13, 165, 260, 270
 - estimation, 179–182
 - nonpolynomial, 127
 - random, 187
- Dual kriging, *see* Kriging
- Dummy points, 318
- Econometrics, 423
- Entropy, 115, 149, 412
- Environmental studies, 48, 137, 365, 419, 448, 632
- Epistemology, 3, 22–24, 137
- Equidistributed sequence, 575
- Ergodicity, 19–22, 74, 543
- Error contrast, 239
- Estimation terminology, 151
- Estimation variance, 130–131
- Exploratory data analysis, 34, 39, 41, 112
- Exponential covariance, *see* Variogram models
- Extension variance, 130
- External drift model, 354–360
- External knowledge, 351, 472
- Extreme value prediction, 214, 379, 385, 416, 454
- Facies, *see* Categorical variable
- Factor, 391, 395
 - nonpolynomial, 398, 405
 - polynomial, 397
- Factorial kriging analysis, 342–349
- Factorized covariance, 97, 202–204, 364, 589
- Fast Fourier transform (FFT), 502–503
- Fault, 54, 208, 219, 357
- Filtering, *see also* Factorial kriging analysis
 - polynomial, *see* ALC- k
 - random error, 210–212, 266, 334–336
 - systematic error, 308–309
- Filtration velocity, 596
- Finite-dimensional distributions, 29
- Flow
 - steady-state, 595
 - transient, 629, 632
- Flow equation, 616
 - perturbation approximation, 617, 618, 622
- Forestry, 138, 196–200, 308
- Fractals, 89, 91, 99, 149, 423, 449, 507, 509, 510–517, 538
- Fractures, 279, 549, 554–558
- Fuzzy, *see* Kriging, Variogram

- Gamma
 - covariance, *see* Variogram models
 - distribution, 403, 406, 641
 - function (Euler), 638
- Gaussian
 - covariance, *see* Variogram models
 - distribution, 639
 - random function, 17
 - transformation, *see* Transformation
- Generalized covariance, 252–257
 - estimation, 281–287
 - spectral representation, *see* Spectral representation
- Generalized covariance models
 - polynomial, 261–263, 517–519
 - power law, 260–261, 517
 - spline, 264, 519–520, 609
- Generalized cross-covariance, 331–332
- Generalized increment, *see* ALC- k
- Generalized least squares, *see* Least squares
- Generalized random process, 78, 603–604, 606, 609, 619
- Generalized variogram, 121, 276–279
- Genetic algorithm, 567
- Geochemistry, 423
- Geology, 8, 45, 216, 291, 318, 562. *See also*
 - Fault, Fractures
- Geophysics, *see* Gravimetry, Magnetism, Seismic
- Geostatistics, 2
- Geothermal activity, 332–336
- Germes, 486, 541, 546
- Gibbs sampler, 229, 529, 531, 566
- Global estimation, 25–26, 131, 132–137, 196, 199
 - of block distribution, 430–434
 - of point distribution, 376–378
- Global warming, 372
- Gold deposits, 52, 91, 212, 214, 412, 436
- Gradient, *see* Derivative
- Grains, *see* Boolean objects
- Gravimetry, 85, 128, 343–344
- Green function, 605, 610, 623, 635
- Guess field model, 352–354
- h -scattergram, 411
- h^a , *see* Generalized covariance models, Power law model, Variogram models
- Hankel transform, 68
- Hard data, 468, 525
- Hard-core process, 553
- Hastings algorithm, 528
- Hausdorff dimension, *see* Fractals
- Heat equation, 402, 635
- Hermite
 - isofactorial model, 404, 409
 - change of support, 437–438
 - polynomials, 399, 402, 448, 640
- Hilbert space, 161, 185, 201, 250, 297, 390
 - defined, 14
- Histogram, *see* Distribution
- Hitting functional, 546
- Hole effect, 55, 69, 92
- Homoscedasticity, 164
- Hydraulic
 - conductivity, 595, 634. *See also*
 - Permeability
 - head, 595, 633
 - resistivity, 599
- Hydrogeology, 123, 137, 319–321, 382, 594–602, 611–633
- Hyperbolic covariance, *see* Variogram models
- Icosahedron, 35, 477
- Ill-posed problem, 612, 614
- Image restoration, 531, 566
- Impulse function, *see* Green function
- Indicator, *see* Cokriging, Kriging, Sequential indicator simulation, Variogram
- Inequality constraints, *see* Conditioning, Kriging
- Infill asymptotics, *see* Micro-ergodicity
- Information effect, 376, 427, 429
- Infrared catastrophe, 67, 259
- Inhibition process, *see* Hard-core process
- Integral range, 73–74, 544, 619
- Interpolation, 150, 351. *See also* Kriging
- Interpretive models, 8
- Intrinsic correlation model, *see* Proportional covariance model
- Intrinsic properties, 251, 266
- Intrinsic random function (IRF), 17, 31
- Intrinsic random function of order k (IRF- k), 245–249
 - abstract, 249–251
- Invariance
 - under linear transformation of the f^l , 127, 183–184
 - under shifts, *see* Shift invariance
- Inverse Gaussian distribution, 553
- Inverse problem, 594, 611–633
- Isofactorial models, 391, 394–405, 543. *See also* Binomial, Hermite, Jacobi, Laguerre, Meixner, Negative binomial, Poisson, Uniform
 - barycentric, 410
 - beta, 403, 411

- change of support, 442–444. *See also*
 - Change-of-support models
 - defined, 395
 - diffusive, 401–403
 - mosaic, 400, 410
 - orthogonal indicator residuals, 401, 405, 549
 - properties, 396–397
 - pure, 403–404, 409
- Iterated exponential covariance, *see*
 - Variogram models
- Itô stochastic integral, 635
- Jackknife, 41, 284
- Jacobi
 - isofactorial model, 403
 - change of support, 444
 - polynomials, 403
- Jump model, 459, 537
- K*-function, 554
- Krawtchouk polynomials, 403
- Krige's
 - regression effect, 193–195, 200
 - relationship, 132
- Kriging
 - and regression, 163–164
 - and splines, 272–276
 - Bayesian, 187–190
 - disjunctive, *see* Disjunctive kriging
 - dual, 186–187, 226
 - fuzzy, 176
 - indicator, 382–384, 387, 419, 436
 - interpolation properties, 157–160, 171
 - intrinsic (IRF-*k*), 265–266
 - lognormal, 190–192, 382, 436
 - minimax, 176
 - of complex variable, 350
 - of spatial average, 195–196
 - ordinary, 165, 170, 185
 - positive, 225
 - probability, 385
 - random, 197
 - robust, 212–215
 - sensitivity to variogram, 175–176
 - simple, 154–164
 - trigonometric, 366–368
 - under boundary conditions, 319–321
 - under inequality constraints, 224–229
 - universal, 164–175
 - with nugget effect, 197, 204, 212
- Kriging equations
 - Bayesian, 189
 - dual, 186
 - intrinsic (IRF-*k*), 265
 - dual, 272, 275
 - ordinary, *see* universal
 - simple, 155
 - universal, 167–168
 - with variogram, 170
- Kriging neighborhood, 153
 - global, 187, 209
 - moving, 202, 234
 - selection, 207–208, 367
- Kriging variance, 156, 170, 265
 - use, 176–178
- Kriging weights, 204–206, 224
- L^1 norm, 42
- L^2 norm, 14, 42
- Lagged scatterplots, 411
- Lagrange multiplier, 167, 299
- Laguerre
 - isofactorial model, 404, 409
 - change of support, 443
 - polynomials, 403, 642
- Laplace equation, 607
- Layer cake model, 360–362
- Least squares, 271
 - generalized, 180
- Lebesgue measure, 637
- Legendre polynomials, 403
- Levinson algorithm, 479
- Linear congruential method, 574
- Linear model of coregionalization, 339–342
- Linear predictive coding, 204
- Linear variogram, *see* Variogram models
- Lipschitz condition, 634
- Local average subdivision, 506, 517
- Local estimation, *see* Block average,
 - Cokriging, Disjunctive kriging, Kriging
- Locally
 - equivalent stationary covariance, 267–270, 291
 - stationary random function, 107, 417, 446, 571
- Logarithmic variogram, *see* Variogram models
- Lognormal
 - blocks, *see* Change-of-support models
 - disjunctive kriging, *see* Disjunctive kriging
 - distribution, 644
 - grades, 56, 91, 423
 - kriging, *see* Kriging
 - random function, 104, 108, 142, 145, 382, 448, 599, 601, 631
 - variogram, 103
- LU decomposition, *see* Covariance matrix
 - decomposition

- Magnetism, 85, 363, 366–368
- Marked point process, 545
- Markov
 - chain, 402, 527, 537, 542, 560
 - process, 85, 402
 - property, 85, 91, 172, 202, 364
 - random field, 529
- Markov chain Monte Carlo, 525–530
- Markov-type model, 305, 326
- Master points, *see* Pilot points
- Maximum likelihood, 109, 630
 - restricted, 110, 281
- Measure theory, 636–638
- Measurement error, 78–79, 210, 218, 468. *See also* Filtering
- Median unbiased, 191
- Meixner
 - isofactorial model, 404, 409
 - polynomials, 403, 644
- Meteorology, 35, 90, 110, 123, 125, 316–318, 352
- Metropolis algorithm, 528, 566, 570
- Micro-ergodicity, 22, 144–145, 278
- Microstructure, 77–78
- Migration process, 100, 149, 477, 613
- Minimax, *see* Kriging
- Minimum stress surface, 291. *See also* Splines
- Mining, 56, 108, 129, 194, 214, 307, 412, 418, 435, 436, 447, 553, 577–583
- Misclassification, 388
- Monitoring data, 362, 368, 371
- Monte Carlo, 453
- Montée, *see* Radon transform
- Mosaic
 - isofactorial model, 400, 410
 - change of support, 444
 - random function, 384, 459, 535
- Moving-average process, 481–483, 511–513
- Multi-gaussian model, 381. *See also* Change-of-support models
- Multidimensional scaling, 371
- Multivariate
 - covariance model, 325
 - estimation, *see* Cokriging
 - recovery, 447
 - spectral representation, 325
- Navier-Stokes equations, 596, 634
- Nearest neighbor, 554
- Negative binomial
 - distribution, 407, 643
 - isofactorial model, 403, 405
 - change of support, 444
- Nested structures, 48, 105. *See also* Factorial kriging analysis
- Network design, 136, 366, 448
- Neural networks, 567
- Neyman-Scott process, *see* Cluster process
- Nonstationary covariance, 188
 - estimation, 368–371
- Nonstationary mean, *see* Drift
- Normality index, 424
- Nugget effect, 51, 77–78, 104
 - multivariate, 310, 327
- Objective property, 23, 144, 234
- Octant search, *see* Kriging neighborhood: moving
- Order relations, 384
- Ordinary kriging, *see* Kriging
- Ordinary least squares, *see* Least squares
- Ore tonnage, *see* Selectivity curves
- Orthogonal complex random measure, 637
- Orthogonal indicator residuals, 401, 405, 549
 - change of support, 444
- Orthogonal polynomials, 397
- Orthogonality property, 161, 184, 297, 390
- Outliers, 212, 215
- Palm distribution, 554
- Parent-daughter process, *see* Cluster process
- Pareto distribution, 423
- Percolation, 592
- Periodicities, 56
- Periodogram, 114
- Permanence of lognormality, *see* Change-of-support models
- Permeability, 525, 586, 595
 - block, 600, 628
 - effective, 598–599
 - equivalent, 594, 596–602
 - inter-block, 601, 628
 - intrinsic, 634
 - upscaled, 600–601
- Perturbation approximation, *see* Flow equation
- Petroleum, 319, 354, 563, 567, 568, 583–592
- Phase spectrum, 328
- Piezometric head, 595, 633
- Pilot points, 629
- Point process, 550–554
- Poisson
 - alignment point process, 554
 - equation, 607–611
 - isofactorial model, 403
 - lines, 268, 513, 539–541, 543
 - point process, 486–487, 536, 554, 638

- doubly stochastic, *see* Cox process
- with regionalized intensity, 551, 559
- polygons, 541, 546
- process, 89, 507
- Pólya's theorem, 66, 68
- Polygons of influence, 429, 539
- Polynomial generalized covariance, *see* Generalized covariance models
- Positioning uncertainty, 79–80, 222
- Positive definite, 59, 61, 148
 - conditionally, 61
 - k -conditionally, 258
 - strictly, 60, 64, 148, 155, 168
- Power law model, *see* Generalized covariance models, Variogram models
- Power-averaging, 600
- Prediction terminology, 151
- Principal component analysis, 345
- Prior distribution, 110, 456, 568
- Probability-field simulation, 471–472
- Projection theorem, 14
- Proportional covariance model, 336–339, 384
- Proportional effect, 45, 56, 107
- Pseudorandom numbers, 574–575
- Quadratic covariance, *see* Variogram models
- Quantity of metal, *see* Selectivity curves
- Quasi-random numbers, 477, 575–576
- Radon transform, 72–73, 82, 85, 475, 612
- Random coins method, 491
- Random error, *see* Filtering
- Random error model, 351–352
- Random function (field), 12
- Random midpoint displacement, 510
- Random noise, *see* White noise
- Random set, 99. *See also* Boolean
- Random spectral measure, 18, 65, 68, 494, 638
- Random walk, 509
- Range, 47, 106, 111, 157, 160, 179, 205
 - integral, *see* Integral range
 - practical, 84
- Rational quadratic covariance, *see* Cauchy covariance
- Reality (comparison with), 198, 220, 418, 436, 447
- Recharge, 616
- Recovery functions, 420
 - duality formulas, 423
 - isofactorial expansion, 415
- Regional
 - distribution, *see* Distribution
 - variogram, *see* Variogram
- Regionalized variable, 2
- Regression, 14, 151, 155, 179, 185, 282. *See also* Kriging
- Regularization, 75–77, 628
- Relay effect, 205
- Renewal process, 536–538
- Representation of an IRF- k , 250
 - internal, 252, 267, 610
 - locally stationary, 267, 291
 - stationary, 259, 270
- Reserves, 214, 428, 429, 431, 447, 578, 585
- Residual, 165, 229, 250
- Resistant, *see* Robust
- Resistivity, 599
- Retrodeformable, 230
- Risk assessment, 218
- Robust, 39. *See also* Kriging, Variogram
- Sample variance, 32, 38
- Sampling
 - design, 132–137
 - error, 79
 - random, 134
 - square grid, 135
 - stratified random, 135
- Scale effect, 593, 628. *See also* Upscaling
- Scale parameter, 82, 84, 85, 86
- Screening effect, 202–205, 306, 338, 364
- Seismic, 190, 207, 217, 351, 354–356, 362
- Seismic inversion, 563
- Selectivity, 420–430
 - curves, 420
 - index, 424
 - of a distribution, 425
- Self-affinity, 149
- Self-similarity, 89, 149
- Semivariogram, 32. *See also* Variogram
- Separable covariance, *see* Factorized covariance
- Separable random function, 13, 148
- Sequential Gaussian simulation, 463, 511
- Sequential indicator simulation, 521–525, 562, 567
- Sequential random function, 549
- Sequential self-calibration, 629
- Shift invariance, 239
 - of kriging, 156, 171, 184
 - of polynomials, 242–243
- Shooting process, *see* Cluster process
- Sichel distribution, 553, 592
- Sill, 47, 106, 192
- Simple kriging, *see* Kriging
- Simulated annealing, 564–568
- Singleton's algorithm, 503

- Skewed distribution, 41, 140, 192, 212, 214, 406
- Slutsky's condition, 21, 74
- Smoothing. *See also* Splines
 - filter, 207
 - relationship, 158–159, 185
- Soft data, 468, 471, 525, 530
- Soil science, *see* Agriculture
- Space-time models, 362–366
- Spatial distribution, 12
- Spatial uncertainty, 2
- Spectral
 - density, 65, 70, 113
 - filtering, 344
 - measure, 18, 21, 64, 67
 - modeling, 113–115
- Spectral representation
 - covariance, 18, 64–66
 - isotropic, 68–70
 - cross-covariance, 325
 - generalized covariance, 257–260
 - intrinsic random function, 68
 - stationary random function, 18, 64
 - variogram, 66–68
 - isotropic, 69
- Spectral simulation method
 - continuous, 494–496, 513–514
 - discrete, 496–504, 514–515, 519, 520
- Spectrum, *see* Spectral density
- Spherical covariance, *see* Variogram models
- Spin exchange, 529, 566
- Spline generalized covariance, *see* Generalized covariance models
- Splines, 272–276
- Square root method, *see* Covariance matrix decomposition
- Stability properties (covariances and variograms), 60, 61
- Stable
 - covariance, *see* Variogram models
 - distribution, 98
 - random function, 98, 484, 493
- Standard error, 156
- Stationarity
 - second-order (weak), 16–17
 - strict (strong), 16
- Statistical tests, 111
- Steady-state flow, *see* Flow
- Steiner compact set, 95
- Stochastic differential equations, 602–607
- Stochastic integral, 63
- Stochastic process, *see* Random function
- Structure function, 34. *See also* Variogram
- Subordinated process, 542
- Substitution random function, 402, 542–545
- Successive random additions, 510
- Support, 6, 153, 637. *See also* Change of support
 - random, 200, 385
- Support effect, 425, 429
- Surface area estimation, 26–27
- Systematic effect, 165, 222, 236
- Systematic error, *see* Filtering
- Tessellation, 535–542
 - Poisson polyhedra, 541–542
 - Voronoi, 538–539
- Three perpendiculars theorem, 200
- Time series, 113–115, 155, 164, 211, 479, 481, 483, 496–501
- Toeplitz matrix, 465, 479
- Topography, 125, 285
- Total uncertainty, 110, 176, 455
- Training image, 562, 567
- Transformation
 - conformal, *see* Conformal transformation
 - gamma, 406, 413
 - Gaussian, 193, 380, 469, 580
 - negative binomial, 407
 - uniform, 406, 471
- Transformation function, 381
 - modeling, 381, 405, 407, 413, 580
 - reverse, 470
- Transient flow, *see* Flow
- Transition
 - models, 72
 - phenomenon, 24, 48
- Transitive theory, 24–27
- Transmissivity, 616. *See also* Permeability
- Transport, 632
- Trend, 57, 165, 234–236, 259, 271, 372
- Triangle covariance, *see* Variogram models
- Triangular inequality, 98, 99
- Truncated Gaussian simulation, 531–535, 567, 586–592
- Truncated pluri-gaussian simulation, 533
- Turbulence, 90
- Turning bands, 472–477, 513, 517, 520
- Unbiasedness
 - conditional, 15, 164, 185, 193, 199, 428, 430
 - conditions, 166, 299, 312, 315, 355, 367
 - price for, 183
- Uniform
 - conditioning, 446, 447
 - isofactorial model, 405
- Unit sphere, 639

- Universal kriging, *see* Kriging
 Universal kriging model, 165, 233–234, 239, 259, 270, 355
 Universality conditions, *see* Unbiasedness
 Upscaling, 594

 Van der Corput sequence, 477, 576
 Variogram, 31
 anisotropic, *see* Anisotropy
 as generalized covariance, 255
 behavior at origin, 50, 104, 111
 box plot, 35
 cloud, 34–35
 concave, *see* Concave variograms
 estimation, 108–111, 137–142
 fitting, 104–111
 fluctuation, 142–145
 fuzzy, 44, 111, 176
 generalized, *see* Generalized variogram
 Huberized, 43
 indicator, 100–103, 383, 400
 map, 37
 median, 42
 misspecification, 105
 nugget effect, *see* Nugget effect
 of order 1, 98, 102, 149, 411
 of order $\frac{1}{2}$, 44
 of order α , 98–99
 of residuals, 116, 120, 123–125
 proportional, 56
 quantile, 42
 raw, 116, 120, 122–123
 regional, 30, 38, 122, 137
 regularized, 75, 90
 robust, 39–44, 100, 215
 sample, 30, 35–38, 137
 slope at origin, 33, 95, 106
 spectral representation, *see* Spectral representation
 theoretical, 30, 31, 38, 57
 underlying, 118, 120

 Variogram models, 80–97
 cardinal-sine, 93, 505
 Cauchy, 70, 85, 343, 493
 circular, 82, 490, 491
 cubic, 84, 491
 exponential, 70, 84–85, 101, 202, 203, 225, 402, 453, 464, 481, 489, 492, 505, 537, 541, 542, 606, 607
 gamma, 70
 Gaussian, 69, 85, 175, 203, 316, 490, 500, 502, 505
 hyperbolic, 70, 505
 iterated exponential, 70
 J-Bessel, 68, 93
 K-Bessel, 86, 202, 489, 493, 607
 linear, 87, 172, 182, 225, 506–510, 609
 logarithmic (de Wijsian), 90–91, 202, 609
 pentamodel, 84
 periodic, 92
 power law, 87–90, 125, 510–517, 538
 quadratic, 268
 spherical, 81–84, 205, 225, 481, 490, 491, 505
 stable, 90, 505
 triangle, 61, 66, 81
 Voronoi polygons, 538–539, 546

 Wavelets, 504–506, 517
 Weierstrass-Mandelbrot random function, 516
 Weight on the mean, 186
 White noise, 52, 78, 609, 635, 638
 Wiener bounds, 598, 600
 Wiener-Lévy process, *see* Brownian motion
 Wild horse, 229
 Winsorizing, 214

 Yule-Walker equations, 155, 479

 Zero effect, 406, 433, 434, 435