MULTIPLE-POINT GEOSTATISTICS

Stochastic Modeling with Training Images

Gregoire Mariethoz • Jef Caers

with website

WILEY Blackwell

Multiple-point geostatistics

Multiple-point geostatistics

Stochastic modeling with training images

Gregoire Mariethoz

Faculty of Geosciences and Environment
University of Lausanne, Switzerland

Jef Caers

Energy Resources Engineering Department
Stanford University, USA

WILEY Blackwell

This edition first published 2015 © 2015 by John Wiley & Sons, Ltd

Registered office: John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices: 9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell

Library of Congress Cataloging-in-Publication Data

Mariethoz, Gregoire, author.
Multiple-point geostatistics : stochastic modeling with training images / Gregoire Mariethoz and Jef Caers.
pages cm
Includes index.
Summary: "The topic of this book concerns an area of geostatistics that has commonly been known as multiple-point geostatistics because it uses more than two-point statistics (correlation), traditionally represented by the variogram, to model spatial phenomena"–Provided by publisher.
ISBN 978-1-118-66275-5 (hardback)
1. Geology–Statistical methods. 2. Geological modeling. I. Caers, Jef, author. II. Title.
QE33.2.S82M37 2015
551.01'5195–dc23

2014035660

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

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Cover image: Courtesy of NASA Earth Observatory.

Set in 9.5/13pt Meridien by Aptara Inc., New Delhi, India

1 2015

Contents

Preface, vii

Acknowledgments, xi

Part I Concepts

I.1 Hiking in the Sierra Nevada, 3

I.2 Spatial estimation based on random function theory, 7

I.3 Universal kriging with training images, 29

I.4 Stochastic simulations based on random function theory, 49

I.5 Stochastic simulation without random function theory, 59

I.6 Returning to the Sierra Nevada, 75

Part II Methods

II.1 Introduction, 87

II.2 The algorithmic building blocks, 91

II.3 Multiple-point geostatistics algorithms, 155

II.4 Markov random fields, 173

II.5 Nonstationary modeling with training images, 183

II.6 Multivariate modeling with training images, 199

II.7 Training image construction, 221

II.8 Validation and quality control, 239

II.9 Inverse modeling with training images, 259

II.10 Parallelization, 295

Part III Applications

III.1 Reservoir forecasting – the West Coast of Africa (WCA) reservoir, 303

III.2 Geological resources modeling in mining, 329
Coauthored by Cristian Pérez, Julian M. Ortiz, & Alexandre Boucher

III.3 Climate modeling application – the case of the Murray–Darling Basin, 345

Index, 361

Preface

Arguably, one of the important challenges in modeling, whether statistical or physical, is the presence and availability of "big data" and the advancement of "big simulations." With an increased focus on the Earth's resources, energy, and the environment comes an increased need for understanding, modeling, and simulating the processes that take place on our planet. This need is driven by a quest to forecast. Forecasting is required for decision making and for addressing engineering-type questions. How much will temperature increase? How much original oil is in place? What will be the volume and shape of the injected CO_2 plume? Where should one place a well for aquifer storage and recovery? The problems are complex; the questions and their answers are often simple.

In addressing such complex problems, uncertainty becomes an integral component. The general lack of understanding of the processes taking place and the lack of data to constrain the physical parameters of such processes make forecasting an exercise in quantifying uncertainty. As a result, forecasting methods often have two components in modeling: a stochastic and a physical component. Physical models produce deterministic outcomes or forecasts; hence, they lack the ability to produce realistic models of uncertainty in such forecasts. On the other hand, stochastic processes can only mimic physics, and although they produce models of uncertainty, these often present poor physical realism or, worse, are physically implausible. The challenge in many forecasting problems is to find the right middle ground for the intended purpose: produce physically realistic models that include the critical elements of uncertainty and are therefore able to answer the simple questions posed.

To some extent, geostatistical methods can historically be framed within this context of forecasting and within the quest for realism and truth. In the past, applications were mostly in the area of subsurface geology, in particular mineral resources, and then later oil and gas resources (as well as groundwater and hydrogeology). Perhaps a key recognition early on was that an assumption of independently and identically distributed (IID) samples taken from a spatially distributed phenomenon, such as an ore body, is a geologically ("physically") unrealistic assumption. Mineral grades show a clear spatial structure that is the direct result of the physical genesis of such deposits. The goal of geostatistics then (and still) was not to model the genesis of that deposit by means of a physical process, but to produce estimates based on a model of spatial continuity that is as realistic as possible. The predominant model was the semivariogram, which

is a statistical model, not a physical one, yet captures some elements of physical variability. Management of mineral resources constitutes a data-rich environment. Although the semivariogram is a rather limited model for describing complex physical realities, the presence of a large amount of drill holes (actual observations of physical reality) made this model of spatial continuity a plausible and successful one in the early stages of applications of geostatistics. The second major application, at least historically, is the modeling of subsurface reservoirs, where direct observations (wells) are sparse and the purpose is to forecast flow in porous media, which in itself requires physical models. In this way, two physical realities are present: the physics of deposition of clastics (sedimentation) or carbonates (growth), and the physics of fluid flow in porous media. Realism is sought in both cases. Many publications showed that geological models of the subsurface that were built based on multi-Gaussian processes (and the semivariogram as a basic parameter) lack geological realism in order to produce realistic forecasts of flow. Although any such evaluation is dependent on the nature of the flow problem considered, it appears to be the case in the large majority of practical flow-forecasting problems. A second problem in data-poor environments concerns the inference of semivariogram parameters. With data based on only a few wells, at best, one can infer some vertical semivariogram properties, but modelers were left to guess most other modeling parameters.

As a consequence, at least in reservoir modeling, Boolean (or object-based) models became fashionable because of their geological realism and flow-forecasting ability. Such models were calibrated from a richness of information available in analog outcrop models. The 1990s saw an expansion of geostatistical techniques in the traditional fields as well as application in several nongeological areas, in particular the environmental sciences. Considering the International Geostatistics Congress proceedings as a particular sample, one finds in the 1988 Avignon Congress only ~10% of applications outside traditional fields, whereas in the 2000 Cape Town Congress, environmental applications alone cover about ~20% of the papers. The 1990s therefore saw a shift in geostatistics that was twofold. Firstly, the early applications and theory that developed around semivariograms, various flavors of kriging, and multi-Gaussian simulation, including hard and soft data, were rapidly maturing. Secondly, the International Geostatistics Congress, which is held every 4 years and had long been the single platform for dissemination of novel research, saw its unique role wane because of the advent of more application-focused conferences (e.g. Petroleum Geostatistics, geoENV, and Spatial Statistics). In terms of research, and particularly in terms of the development of new methods, a drive toward non-Gaussian model development can be observed, perhaps now scattered over various areas of science and presented in various disjunctive conferences and journals. Some of the non-Gaussian methods still rely on semivariograms (or covariance functions in the statistical literature), such as the pluri-Gaussian methods or Karhune–Loeve expansions, whereas others rely on developments in the field of image analysis.

The Markov Random field (MRF), although its theory was originally developed in the 1980s, saw a proliferation of applications in both spatial and space–time modeling. The development of methods remained classical, however: data were used to fit parametric models, whether semivariograms, MRF parameters, or using traditional statistical methodologies (e.g., maximum likelihood and least squares); models were then used for estimation, or for simulation by sampling posterior distributions. Development of theoretical models is clearly based on probability theory and its extension such as Bayesian methods.

Multiple-point geostatistics, abbreviated throughout the literature as MPS, was primordially born out of a need to address the issue of lack of physical realism as well as the lack of control in the simulated fields in traditional modeling. As Matheron stated in his seminal contribution, parameters of traditional statistical models need not have a physical equivalent. Although for a theoretical probabilistic model there may be a "true" parameter, such as the Poisson intensity θ, there exists no physical property in the real world known as θ. One only has a set of true point locations within a domain when studying point processes. The data are the only physical reality. The goal of MPS is to mimic physical reality, and the vehicle to achieve this is the training image. Perhaps the name "multiple-point" suggests that this is a field of study that focuses on higher-order statistics only, but this is only partially true. The second component, namely the source of such statistics (an order of 2 or higher), is the use of a representation of the physical reality: the training image. We believe that the most important contribution in this new field, and this first book, lies in the use of training images to inform and hence include physical reality in stochastic modeling. This is a completely new contribution; it is, without exaggeration, a paradigm shift. Most of the methods covered do not follow the traditional paradigm of first parametric (or even nonparametric) modeling from data, then estimating or sampling from the given parametric model, building on probability theory only. We propose methods that skip this intermediate step (of parameterized or nonparametric models) and directly lift what is desired, whether it is the estimate or the sample or realization from the training images. The methods we propose are therefore no longer solely steeped in statistical science or probability theory (as is most of geostatistics); we borrow from computer science as well and create hybridization between these fields. For that reason, some would no longer term this "geostatistics". Labels are but labels; what matters is the content behind them.

This book is therefore a book about spatial and spatiotemporal modeling in the physical sciences (sedimentology, mineralogy, climate, environment, etc.). We do not claim any applications (yet) in other areas where spatial statistics are used (e.g., health or finance), although such applications are likely to occur in the future. This book is therefore all about practice and solving real problems, not to create more theory. The primary goal of engineering is to address engineering questions; it is not just the creation of stochastic models. However, within stochastic modeling itself, the goal is not the posterior probability distribution

function (pdf) or the model parameters; rather, it is the estimates of that reality or the simulation of that reality. All other intermediate steps, whether inferring parameters or assessing convergence of samples, are but intermediate steps to the creation of a physical reality.

This book is constructed in three major parts. In Part I, we provide by means of a virtual case, a motivation and illustration of what MPS is, the major conceptual elements of MPS, what it aims to achieve, and how training images are generated and used. Part I therefore also serves as a platform to review some assumptions that are fundamental to spatial and spatio-temporal modeling. The aim is to illustrate that a simple problem of spatial estimation and simulation can be solved with and without random function theory.

In Part II, we cover quite exhaustively the various technical details of the methodologies and algorithms currently developed in this field. Starting from basic building blocks in statistical science and computer science, the glue of algorithmic development, we provide an overview of most existing algorithms. We treat important concepts in modeling such as nonstationary and multivariate modeling, the evaluation of consistency between data and model, the construction of training images, and how such training images can be used to formulate and solve spatial inverse problems.

In Part III, we provide the application of these methods to three major application areas: reservoir modeling, mineral resources modeling, and climate science. The last part serves as an illustration of the methodology development in Part I; it should not be seen as an exhaustive list of applications but, rather, as a template for future development.

Accompanying this book is a website with a collection of training images and example test cases: http://www.trainingimages.org. In the book, we provide a reference list per chapter. For a complete and updated reference list, please visit the website http://www.trainingimages.org. PowerPoint slides of all figures in the book can also be accessed and downloaded at www.wiley.com/go/caers/multiplepointgeostatistics.

Acknowledgments

This book would not have come to existence without contributions from a number of students and colleagues who helped us with their challenging comments, provided outcomes of their own research, and reviewed parts of this book. We are particularly grateful to Celine Scheidt, Lewis Li, Pejman Tahmasebi, Kashif Mahmud, Sanjeev Jha, Siyao Xu, Satomi Suzuki, Sarah Alsaif, and Cheolkyun Jeong for contributing to some of the most innovative aspects of the research results presented in this book, and they provided us with excellent figures. Thanks are also due to Lewis Li for coding and analyzing the "universal kriging with training image" method in Part I.

We also want to thank Thomas Romary for discussions relating to Part I and for reviewing this part, to Thomas Hansen for reviewing the chapter on inverse modeling, to Odd Kolbjørnsen for reviewing the chapter on MRF, and to Philippe Renard and Julian Straubhaar for reviewing the chapters on nonstationarity and training image construction. Among the people who contributed to this book, we especially thank Alexandre Boucher, Cristian Peréz, and Julián Ortiz for coauthoring the chapters on the mining case study. Alex also provided inspiration and SGEMS code regarding the case study in Part I.

We also want to thank the anonymous reviewers of the initial proposal for this book, for unanimously endorsing the book project and making valuable suggestions. We acknowledge funding from the Stanford Center for Reservoir Forecasting and the National Centre for Groundwater Research and Training. Importantly, the support of the University of New South Wales (Australia) and of ETH Zürich are acknowledged, which both employed Gregoire Mariethoz during the time he was writing this book. Without the support of these institutions, this book would not exist.

Finally, we would like to thank Fiona Seymour and her team at Wiley, who made the publication of this book a smooth one.

PART I
Concepts

CHAPTER 1
Hiking in the Sierra Nevada

1.1 An imaginary outdoor adventure company: Buena Sierra

As is the case for any applied science, no geostatistical application is without context. This context matters; it determines modeling choices, parameter choices, and the level of detail required in such modeling. In this short first chapter, we introduce an imagined context that has elements common to many applications of geostatistics: sparse local data, indirect (secondary) or trend information, a transfer function or decision variable, as well as a specific study target. The idea of doing so is to remain general by employing a synthetic example whose elements can be linked or translated into one's own area of application.

Consider an imaginary hiking company, Buena Sierra, a start-up company interested in organizing hiking adventures in the Sierra Nevada Mountains in the area shown in Figure I.1.1(left). The company drops customers over a range of locations to hike over a famous but challenging mountain range and meets them at the other end of that range for pickup. Customers require sufficient supplies in what is considered a strenuous trip over rocky terrain, with high elevation changes on possibly hot summer days. Imagine, however, that this area lies in the vicinity of a military base; hence, no detailed topographic or digital elevation model from satellite observation is available at this point. Instead, the company must rely on sparse point information obtained from weather stations in the area, dotted over the landscape; see Figure I.1.1(right). We consider that the exact elevation of these weather stations has been determined. The company now needs to plan for the adventure trip. This would require determining the quantity of supplies needed for each customer, which would require knowing the length of the path and the cumulative elevation gain because both correlate well with effort. The hike will generally move from west to east. The starting location can be any location on the west side from grid cell (100,1) to grid cell (180,1) (see Figure I.1.2).

To make predictions about path length and cumulative elevation gain, a small routing computer program is written; although it simplifies real hiking, the program is considered adequate for this situation. More advanced routing could be applied, but this will not change the intended message of this imaginary example.

Multiple-point Geostatistics: Stochastic Modeling with Training Images, First Edition. Gregoire Mariethoz and Jef Caers.

Companion website: www.wiley.com/go/caers/multiplepointgeostatistics

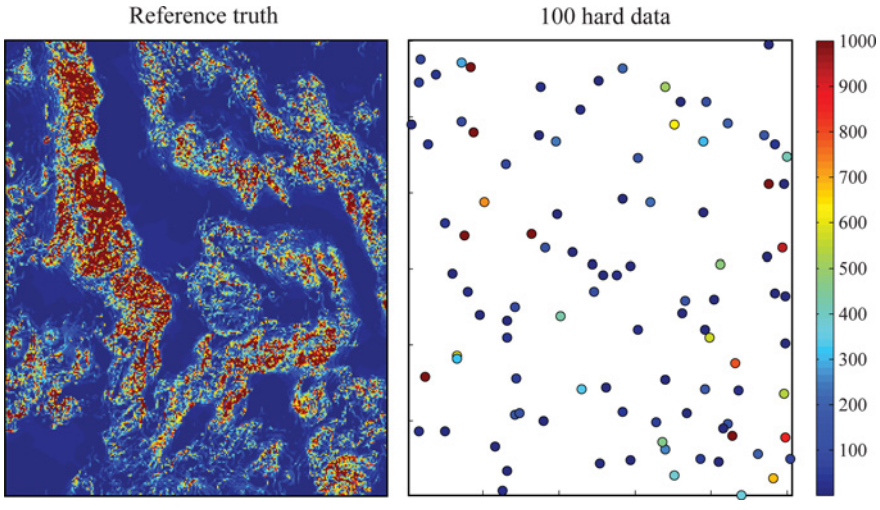

Figure I.1.1 (left) Walker Lake exhaustive digital elevation map (size: 260×300 pixels) grid; and (right) 100 extracted sample data. The colorbar represents elevation in units of ft.

The program requires as input a digital elevation map (DEM) of the area gridded on a certain grid. The program has as input a certain point on the west side, then walks by scanning for the direction that has the smallest elevation change. The program simulates two types of hikers: the minimal-effort (lazy) hiker and the maximal-effort (achiever) hiker. In both cases, the program assumes the hiker thinks only locally, namely, follows a path that is based on where they

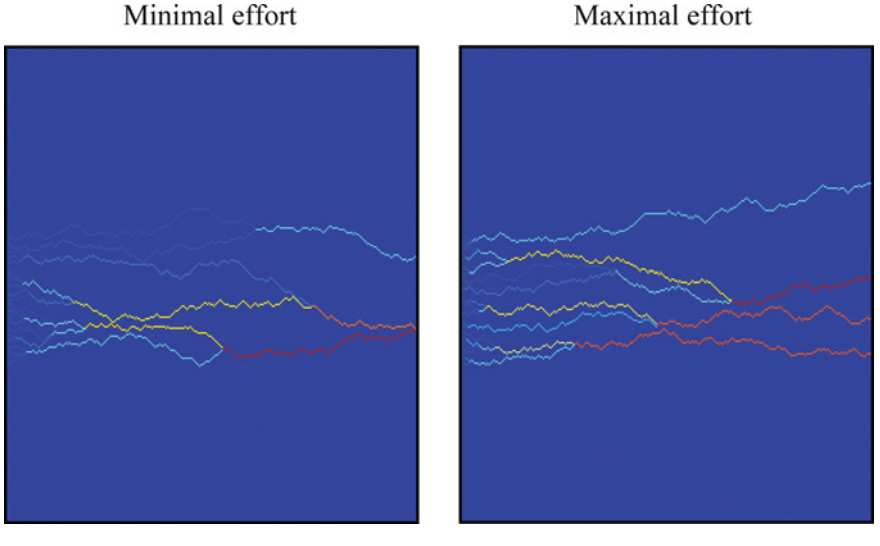

Figure I.1.2 Visualization of the 80 paths taken by hikers of two types: (left) minimal effort; and (right) maximal effort. The color indicates how frequently that portion of the path is taken, with redder color denoting higher frequency.

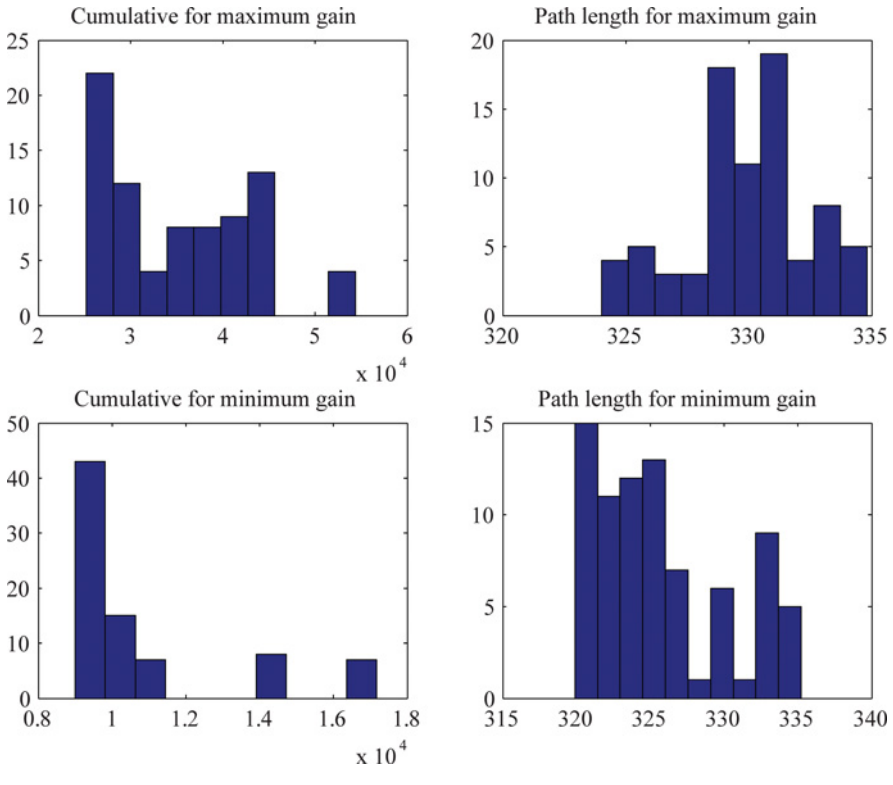

Figure I.1.3 Histograms of the cumulative elevation gain and path length for the minimal- and maximal-effort hiker. Cumulative elevation gain in units of ft, path length in units of grid cells.

are and what lies just ahead. The minimal hiker takes a path of local least resistance (steepest downhill or least uphill). The achiever hiker takes a path of maximal ascent (or minimal descent). Note that the computer program represents a deterministic transfer function: given a single DEM map, a single starting point, and a specific hiker type, it outputs a single deterministic hiking route. If the actual reference, Walker Lake, is used as input, then given starting locations from grid cell (100,1) to (180,1) on the west side, a total of 80 outcomes are generated. These 80 outcomes can be shown as a histogram; see Figure I.1.3. The resulting path statistics for both minimal effort and maximal effort are shown in Table I.1.1, which are summarized with quantiles (the eighth lowest, or P10; the 40th lowest, or P50; and the 72nd lowest, or P90).

1.2 What lies ahead

The problem evidently is that no DEM is available. How, then, would one proceed with forecasting path length and cumulative elevation change, and thereby make recommendations for Buena Sierra? We start in this Part I from very basic

Table I.1.1 Summary statistics

	Minimum effort			
	Cumulative elevation gain (ft)		Path length (cell units)	
	Median	P10–P90	Median	P10–P90
Walker Lake reference	9862	9434–14,875	324	311–327
	Maximum effort			
	Cumulative elevation gain (ft)		Path length (cell units)	
	Median	P10–P90	Median	P10–P90
Walker Lake reference	37,783	35,335–47,731	331	323–333

notions on how to formulate a theory for these kinds of problems, and then we present practical solutions based on that theory. This provides an opportunity to review important notions and assumptions that are common to most spatial prediction problems. The first such theory formulates spatial estimation, which in geostatistics is known as kriging. It is well known that kriging provides an overly smooth map, not reflecting the actual roughness of the terrain, and therefore any predictions of path length or elevation gain are biased. Such prediction would require stochastic simulation, which also allows statements of uncertainty about the calculated route statistics. Nevertheless, we will start with developing kriging because we will show how the traditional kriging (Chapter I.2) can be formulated without relying on the notions of expectation, probability, or random function theory, as long as a training image is available (Chapter I.3). The solution obtained is strikingly similar to traditional kriging, yet at no instance will we rely on random function theory.

Next, we will review stochastic simulation, which traditionally has relied on the same variogram and random function notions as kriging. In particular, we will review Gaussian theory and some popular methods that have been derived from this theory (Chapter I.4). Next, we will show, in a similar vein as for kriging, that the random function theory is not needed to perform stochastic simulation (Chapter I.5). We will present three alternative algorithms as an introduction to the many algorithms presented in Part II. These methods are compared in their ability to solve the practical problem discussed here (Chapter I.6).

CHAPTER 2

Spatial estimation based on random function theory

2.1 Assumptions of stationarity

In this chapter, we mostly review spatial estimation, a general term for estimating or guessing the outcome at unmeasured geographic locations from locations where measurements ("hard data") are available. As is the case for many statistical methods of estimation, the specification of a criterion of "best" is required. There will be only one guess or one estimate that can be given, once such a criterion has been specified. The variable being considered in the example case is the digital elevation map (DEM). One cannot directly estimate the path statistics.

Consider first a nonspatial problem, such as estimating the weight of a specific chair in a classroom. To represent this problem, we introduce the following notation. The true weight of that chair is unknown, denoted as Z, a random variable representing an unknown truth. A particular outcome, for example $z = 7\,\text{kg}$, is written with a small letter. Suppose that all other chairs in that room are similar to the chair in question and we know the weight of those chairs, denoted as $\{z_1, z_2, \ldots, z_n\}$. Based on these data, we make a histogram of the set of chairs. In doing so, an assumption is made: pooling all the weight data into a single plot, such as a histogram, entails that they are "similar" or "comparable". In probability theory, this is often referred to as "the population": a set of outcomes whose values can be grouped. They can be grouped for various reasons: similar origin, similar manufacturer, similar species, similar geological layer, similar location, and so on. However, such pooling requires a decision of what this reference population is. If one would pool tables into the set of chairs, then such pooling will lead to possibly very different results later on, and possibly very erroneous results. In many geostatistics books, this pooling and the accompanying assumption have been termed an "assumption of stationarity".

Only once a decision of stationarity has been made can estimation based on data proceed. Any guess – or, in statistical terms, any specific estimator – will be some unique function of the data, returning a single value:

$$z^* = g(z_1, z_2, \ldots, z_n) \tag{I.2.1}$$

Multiple-point Geostatistics: Stochastic Modeling with Training Images, First Edition. Gregoire Mariethoz and Jef Caers.

Companion website: www.wiley.com/go/caers/multiplepointgeostatistics

Many functions g could be considered, hence we need to specify or state some desirable properties for it. A property often stated as desirable is that of unbiasedness: namely, if a guess is made and denoted as Z^* (that guess is not yet known; it is therefore a random variable by itself), then unbiasedness can be stated based on the notion of expectation:

$$\text{Unbiasedness condition: } E[Z - Z^*] = 0 \tag{I.2.2}$$

Although this condition is common in many probability theory books, it is nontrivial for most first readers. The question is often: what is this an expectation of? What are we "averaging" over? To make such averaging feasible, one would need repeated situations, yet there are no such repeated situations: there is only one single specific chair with an actual weight that is estimated, and hence there will be only one difference between the true weight and our guess. Hence, why this "expectation"?

The unbiasedness therefore invokes a second assumption of stationary: the particular guessing procedure, if applied to (infinitely large) *similar* situations, will have the property of being, on average, equal to the truth. Suppose now that a reality exists where we would have many rooms, each such room containing a set of chairs and a specific chair for which we want to estimate the weight. Then, we need to assume that the situations presenting themselves in all these rooms form yet another population: the population of rooms. Making such a population requires, in a similar vein, an assumption of stationarity. The difficulty is that these alternative-world rooms never exist or are never truly considered; they are imaginary theoretical constructs.

A second condition often posed relates to our attitude toward making a mistake or the consequences of making errors. In the context of the chair, the particular person involved could reason as follows: overestimating the weight of the chair may be of less concern, but underestimating the weight may lead to injury upon attempting to lift it (supposing the estimating person has back problems). Clearly, making errors, whether positive or negative, may have different consequences. In the case of the chair, different attitudes may be taken. A thresholding function could be defined, where underestimating has a given consequence (a fixed hospital bill) over a certain weight value, or may gradually increase due to the increasing severity of the injury. In general, there is a function L, termed the loss function, quantifying our attitude to mistakes or to consequences of the error $z - z^*$. This leads to a second property that could be deemed desirable for any estimate or guess: minimize an expected loss, or,

$$E[L(Z - Z^*)] \text{ is minimal} \tag{I.2.3}$$

We return to the question: what are we averaging over? Averaging requires repetition of similar situations. Again, we need to consider imaginary parallel

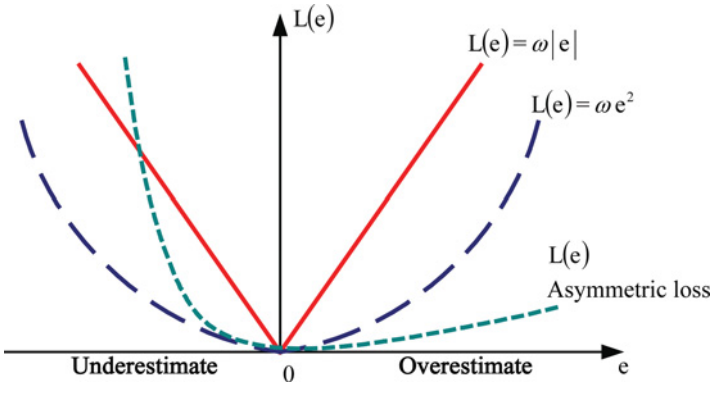

Figure I.2.1 Example loss functions; the most common choice is the parabola (least squares).

universes where the same situation occurs; and now, as a third assumption of stationarity, we assume that our attitude of loss will be similar in all those situations.

The most common loss function is to assume a parabola, as in Figure I.2.1, which is not necessarily applicable to the situation of the chair. The reason for assuming this simple squared function is not necessarily practical or aligned with reality, but rather is a mathematical convenience and driven by the elegance of the resulting solution, namely,

$$\min_{Z^*} E[(Z - Z^*)^2] \Leftrightarrow Z^* = E[Z] \tag{I.2.4}$$

In other words, the expected value minimizes a squared loss function. This basic result is the foundation of least squares theory. A simple arithmetic average is then an unbiased guess of the population mean (expected value), and, as stated in many statistical books, this occurs under the condition of independently identically distributed (IID) data. The latter entails the first hypothesis of stationarity (identically) and includes a kind of sampling that is not biased toward certain values (independently). However, as stated above, two additional assumptions of stationarity are required to get to this result.

In summary, the following stationarity hypotheses are needed to make any estimation procedure feasible, whether nonspatial or spatial:

1. Pooling of data: the creation of populations
2. Properties of estimators are defined through expectation, referring to repeated estimations in similar circumstances.
3. Loss incurred due to errors in the estimation (difference with the truth) refers to repeated situations with a similar attitude of loss.

None of these hypotheses can be tested objectively with any data; they are fundamental to the construction of the theory.

2.2 Assumption of stationarity in spatial problems

If this sounds a little bit construed from a practical viewpoint, but perfectly sound mathematically, then the situation becomes even more compelling when dealing with a spatial context. In the nonspatial context, the data are considered multiple alternative realizations or outcomes of the same truth: for example, the weight of a chair, with the caveat of an assumption of stationarity. Indeed, otherwise each chair would be "unique" and, in a way, a population on its own.

Consider analyzing the assumption of stationarity using the simple constructed example in Figure I.2.2: a single unique truth exists, and at a few locations, this unknown truth is known through its sample data. In notation, each unmeasured geographic location is associated with an unknown truth, which we denote as $Z(\mathbf{x}), \mathbf{x} = (x, y, z)$, or, if space–time is considered, $\mathbf{x} = (x, y, z, t)$. $Z(\mathbf{x})$ is considered to be a random function (as opposed to a random variable before). The term "function" refers to the fact that the outcome associated with each $\mathbf{x}$ is unique, and it also suggests a systematic variation (noting that "pure random" is a specific form of such systematic variation). If a grid is specified, then we need to deal only with a finite set of such $Z(\mathbf{x})$: $\{Z(\mathbf{x}), \mathbf{x} \in \text{Grid}\}$. At a finite set of locations, samples are recorded: $\{z(\mathbf{x}_\alpha), \alpha = 1, \ldots, n\}$.

Clearly, no repeated data exist on each $Z(\mathbf{x})$, as is the case for the nonspatial case. The only information available is measurements taken at a limited set of locations. The assumption of stationarity that is needed to make any estimation possible now requires including a geographical element, namely, an area over which pooling of data is allowed. For example, if we make a histogram of sample data over the area of study, then clearly we have made a geographical assumption of stationarity: the data at a location in that area can be pooled into a single plot, and that plot is meaningful.

We now return to the problem of estimation, in this case spatial estimation. We would like to determine at each uninformed location a "best guess" of that unknown value given some data. The problem is now spatial due to the indexing

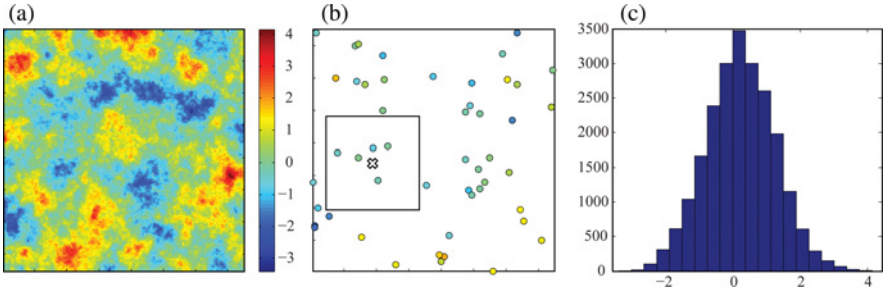

Figure I.2.2 (a) A single unique truth; (b) some sample data taken from it; and (c) its histogram. The goal is to estimate the value at the unsampled location marked with X.

with $\mathbf{x}$, hence the estimator becomes $Z^*(\mathbf{x})$. If unbiasedness is a desirable property, then

$$\text{unbiasedness condition: } E\left[Z(\mathbf{x}) - Z^*(\mathbf{x})\right] = 0 \tag{I.2.5}$$

What are we averaging over? One can imagine two types of averaging: in the first, we could average spatially, meaning over all possible $\mathbf{x}$. A second type of averaging is over all possible similar situations that could possibly occur in the universe. Again, to make this happen, we could invoke an alternate reality with many parallel universes where a similar situation as in Figure I.2.2 occurs; and, on average, over all the alternate realities, our guess would be equal to the truth.

2.3 The kriging solution

Now that some basic notions common to spatial estimation have been established, as well as basic assumptions needed to use and apply probability theory (expectation) to formulate such problems, we establish the most commonly used spatial estimation method in geostatistics: kriging. This section is mostly a simple review of basic equations of ordinary kriging, but perhaps with a more explicit statement and discussion of their underlying assumptions. Later, we will rewrite these equations without any use of expectation or random function theory.

2.3.1 Unbiasedness condition

We consider the situation in Figure I.2.2 where an estimate at only one location is required. First, we specify the function in Equation (I.2.1) as a simple linear sum:

$$z^*(\mathbf{x}) = \sum_{\alpha=1}^{n} \lambda_\alpha z(\mathbf{x}_\alpha) \tag{I.2.6}$$

or, written in terms of random variables:

$$Z^*(\mathbf{x}) = \sum_{\alpha=1}^{n} \lambda_\alpha Z(\mathbf{x}_\alpha) \tag{I.2.7}$$

Plugging this estimator into the unbiasedness condition

$$E[Z(\mathbf{x}) - Z^*(\mathbf{x})] = 0 \tag{I.2.8}$$

leads to

$$\sum_{\alpha=1}^{n} \lambda_\alpha E[Z(\mathbf{x}_\alpha)] = E[Z(\mathbf{x})] \tag{I.2.9}$$

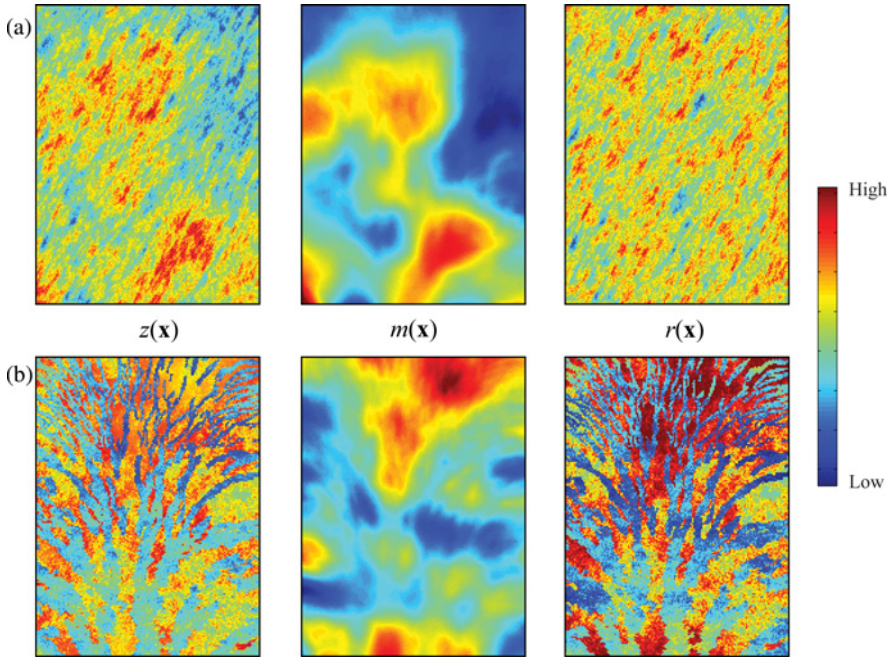

Figure I.2.3 (a) Rock density in a homogeneous layer of a carbonate reservoir; and (b) rock density in a heterogeneous deltaic reservoir.

Should the expected value be stationary, meaning constant over the domain, then the unbiasedness condition becomes

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1 \tag{I.2.10}$$

The assumption of a stationary expected value is rarely useful in practical modeling. It assumes that a phenomenon under study can be decomposed as in Figure I.2.3a, or, in mathematical terms:

$$Z(\mathbf{x}) = M + R(\mathbf{x}) \tag{I.2.11}$$

where M is some unknown but spatially constant expected value (hence, a random variable itself) and $R(\mathbf{u})$ is often termed the residual, which is spatially varying. In a Bayesian context, this expected value could have a prior itself (Omre, 1987), but most times we assume this expected value to be a constant, then

$$Z(\mathbf{x}) = m + R(\mathbf{x}) \tag{I.2.12}$$

Very few phenomena vary spatially in the same way as in Figure I.2.2, and certainly not the DEM under study, with its systematic variation of mountain and lake beds. In this context, one can propose an extension as follows:

$$Z(\mathbf{x}) = m(\mathbf{x}) + R(\mathbf{x}) \quad \text{with} \quad E[R(\mathbf{x})] = 0 \quad \forall \mathbf{x} \tag{I.2.13}$$

The phenomenon is now decomposed into two parts: (1) a slowly varying expected value, often termed "trend"; and (2) a second part, R, that varies faster than the first part and whose expected value equals zero – this is often termed the "residual". We first discuss to what extent this decomposition is useful or even realistic. The decomposition would only be useful if the estimation or modeling of the two components m and R is easier than the direct estimation or modeling of Z. If this is not the case, then the decomposition is made purely for mathematical reasons. Consider two phenomena shown in Figure I.2.3. The question is whether it is useful to write each image as

$$z(\mathbf{x}) = m(\mathbf{x}) + r(\mathbf{x}) \quad \forall \mathbf{x} \in A \tag{I.2.14}$$

In Figure I.2.3(a), we can easily make such decomposition meaningful: a slowly varying and highly varying decomposition is achieved. One can imagine that modeling each component is easier than directly modeling the z-phenomenon. This is no longer the case in Figure I.2.3(b). The trend in this image lies on certain channel properties (width and orientation), not on the image z itself. The decomposition does not achieve an easier modeling task: the residual r looks like z. From a purely mathematical point of view, all phenomena can be written as a sum of two other phenomena; the more fundamental question lies in whether this is meaningful, makes further modeling easier and, perhaps more importantly, leads to better predictions in the given modeling context.

2.3.2 Minimizing squared loss

Consider now a case where the decomposition in Equation (I.2.13) is meaningful for the phenomenon being studied. Next, we need a specification of loss, as discussed in this chapter. Consider the following specification of loss:

$$Var[Z(\mathbf{x}) - Z^*(\mathbf{x})] \text{ is minimal} \tag{I.2.15}$$

which simplifies in combination with the unbiasedness condition to

$$E[(Z(\mathbf{x}) - Z^*(\mathbf{x}))^2] \text{ is minimal} \tag{I.2.16}$$

which, using Equation (I.2.13) and Equation (I.2.7), can be rewritten as

$$E\left[\left(E[Z(\mathbf{x})] + R(\mathbf{x}) - \sum_{\alpha=1}^{n} \lambda_\alpha (E[Z(\mathbf{x}_\alpha)] + R(\mathbf{x}_\alpha))\right)^2\right] \text{ is minimal} \tag{I.2.17}$$

We can write the difference between truth and estimator as follows:

$$Z^*(\mathbf{x}) - Z(\mathbf{x}) = \sum_{\alpha=1}^{n} \lambda_\alpha Z(\mathbf{x}_\alpha) - Z(\mathbf{x}) = \sum_{\alpha=0}^{n} \lambda_\alpha Z(\mathbf{x}_\alpha) = \sum_{\alpha=0}^{n} \lambda_\alpha (E[Z(\mathbf{x}_\alpha)] + R(\mathbf{x}_\alpha)) \tag{I.2.18}$$

with

$$\lambda_0 = -1;\ \mathbf{x} = \mathbf{x}_0 \tag{I.2.19}$$

The unbiasedness condition can be written as follows:

$$\sum_{\alpha=0}^{n} \lambda_\alpha E[Z(\mathbf{x}_\alpha)] = 0 \tag{I.2.20}$$

Some simple algebra then leads to

$$\begin{aligned} &E[(Z(\mathbf{x}) - Z^*(\mathbf{x}))^2] \text{ is minimal } \Leftrightarrow \\ &E[(R(\mathbf{x}))^2] + 2\sum_{\alpha=1}^{n} \lambda_\alpha E[R(\mathbf{x})\, R(\mathbf{x}_\alpha)] + \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta E[R(\mathbf{x}_\alpha)R(\mathbf{x}_\beta)] \text{ is minimal} \end{aligned} \tag{I.2.21}$$

One notices how the expected value of Z has disappeared from the loss specification, and only residual expectations remain. This is possible only because we assume

$$E[Z(\mathbf{x})] = m(\mathbf{x}) \tag{I.2.22}$$

In other words, the expected value is not randomized (assumed to be a random variable) itself. The combination of an unbiasedness condition and a loss specification has resulted in the following minimization problem with linear constraints:

$$\begin{cases} E[(R(\mathbf{x}))^2] + \sum_{\alpha=1}^{n} \lambda_\alpha E[R(\mathbf{x})R(\mathbf{x}_\alpha)] + \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta E[R(\mathbf{x}_\alpha)R(\mathbf{x}_\beta)] \text{ is minimal} \\ \sum_{\alpha=0}^{n} \lambda_\alpha E[Z(\mathbf{x}_\alpha)] = 0 \end{cases} \tag{I.2.23}$$

Using the Lagrange formalism, the following augmented function is being minimized:

$$\begin{cases} S(\lambda_\alpha, \alpha = 1, \ldots, n; \mu) = s(\lambda_\alpha, \alpha = 1, \ldots, n) + \mu\left(\sum_{\alpha=0}^{n} \lambda_\alpha E[Z(\mathbf{x}_\alpha)]\right) \\ \text{with } s(\lambda_\alpha, \alpha = 1, \ldots, n; \mu) = E[(R(\mathbf{x}))^2] + 2\sum_{\alpha=1}^{n} \lambda_\alpha E[R(\mathbf{x})R(\mathbf{x}_\alpha)] \\ + \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta E[R(\mathbf{x}_\alpha)R(\mathbf{x}_\beta)] \\ \lambda_0 = -1; \mathbf{x}_0 = \mathbf{x} \end{cases} \tag{I.2.24}$$

Calculating and equating derivatives to zero result in the following systems of linear equations with linear constraints:

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_\beta E[R(\mathbf{x}_\alpha)R(\mathbf{x}_\beta)] + \mu = E[R(\mathbf{x})R(\mathbf{x}_\alpha)] \quad \alpha = 1, \ldots, n \\ \sum_{\beta=1}^{n} \lambda_\beta E[Z(\mathbf{x}_\beta)] = E[Z(\mathbf{x})] \end{cases} \tag{I.2.25}$$

The linear system of size $(n+1) \times (n+1)$ can be solved once the following terms are specified:

$E[R(\mathbf{x}_\alpha)R(\mathbf{x}_\beta)]$: the covariance of the residuals between different data locations

$E[R(\mathbf{x})R(\mathbf{x}_\alpha)]$: the covariance of the residual between data location and the location to be estimated

$E[Z(\mathbf{x}_\alpha)]$: the expected value at the data locations

$E[Z(\mathbf{x})]$: the expected value at the location to be estimated

Several roadblocks are still in place to obtain any kind of numerical values for these terms:

- There are no repeated data to estimate $E[Z(\mathbf{x}_\alpha)]$ or $E[Z(\mathbf{x})]$ without making additional assumptions.
- Even if we had such repeated data, it is on Z, not on R, hence we cannot estimate the above covariances of R.

It is clear that additional simplifications and assumptions need to be invoked before any numerical calculations can be carried out. First, one could assume the expected value to be the same at all geographical equations. This would take care of the bottom two terms in the above list and also simplify the linear constraint to

$$\sum_{\beta=1}^{n} \lambda_\beta = 1 \tag{I.2.26}$$

Given a stationary expected value, one can in addition assume a stationary residual. If we introduce the notation of covariance for R as follows:

$$Cov[R(\mathbf{x}'), R(\mathbf{x})] = E[R(\mathbf{x}')R(\mathbf{x})] \tag{I.2.27}$$

Then, under an assumption of stationary expected value and stationary covariance, the notation can be simplified as follows:

$$Cov[R(\mathbf{x}'), R(\mathbf{x})] = Cov_R(\mathbf{x}' - \mathbf{x}) \tag{I.2.28}$$

In other words, the covariance is only a function of the distance between geographical locations, not the exact place where geographically one is located. As a consequence, the linear system now becomes

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_\beta Cov_R(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \mu = Cov_R(\mathbf{x} - \mathbf{x}_\alpha) \quad \alpha = 1, \ldots, n \\ \sum_{\beta=1}^{n} \lambda_\beta = 1 \end{cases} \tag{I.2.29}$$

an $(n+1) \times (n+1)$ system of equations that is traditionally known as the ordinary kriging system. Once the system is solved – namely, numerical values for λ and μ are obtained – then the minimum of Equation (I.2.16) can be algebraically expressed as

$$\text{var}_{\min}(\mathbf{x}) = Var(Z) - \sum_{\alpha=1}^{n} \lambda_\alpha Cov_R(\mathbf{x}_\alpha - \mathbf{x}) - \mu \tag{I.2.30}$$

which is commonly known as the ordinary kriging variance. Next, we deal with obtaining numerical values for the covariance terms to solve Equation (I.2.29).

2.4 Estimating covariances

Solving the system of linear equations to estimate the unsampled value at the location highlighted in Figure I.2.2 requires specifying covariance values in Equation (I.2.29). In traditional approaches, this calculation is only possible through the assumption in Equation (I.2.28): "covariance is only function of the distance between geographical locations". This assumption allows pooling data pairs with similar distance (exact distance replicates rarely exist with irregular data) into a single scatterplot from which the covariance value can be calculated. This exercise can be repeated for various distances. The very existence of this single scatterplot is an explicit statement or expression of stationarity: data from different locations are pooled into one single plot. These covariance values are then grouped based on distances calculated along the same (or similar for irregular data) directions. The above linear system calls for covariances on R, not on Z. For this simple case, this poses no problem, as the mean appears fairly constant over the domain. More difficult situations are discussed in this section.

In geostatistics, semivariograms are commonly estimated. Without any assumption of stationarity, these semivariograms are defined as follows:

$$2\gamma[R(\mathbf{x}'), R(\mathbf{x})] = E[(R(\mathbf{x}') - R(\mathbf{x}))^2] \tag{I.2.31}$$