

Applied Geostatistics with SGeMS: A Users' Guide

Nicolas Remy, Alexandre Boucher & Jianbing Wu

This document presents an extract of an upcoming book written by Nicolas Remy, Alexandre Boucher and Jianbing Wu. The book has 10 chapters detailing how to use the SGeMS software. SGeMS is a software for 3D geostatistical modeling. It implements many of the classical geostatistics algorithms, as well as new developments related to multiple-point geostatistics. The software is open source and free of charge, it can be downloaded at <http://sgems.sourceforge.net/>. Included is the book table of content and the more theoretical Chapter 3 *Geostatistics: A recall of concept*

Contents

Foreword	5
1 Introduction	7
2 General Overview	11
2.1 A Quick Tour of the Graphical User Interface	11
2.2 A Typical Geostatistical Analysis using SGeMS	15
2.2.1 Loading data into an SGeMS project	16
2.2.2 Elementary data analysis (EDA)	18
2.2.3 Variogram modeling	18
2.2.4 Creating a grid	20
2.2.5 Running a geostatistics algorithm	21
2.2.6 Displaying the results	24
2.2.7 Post-processing the results with Python	28
2.2.8 Saving the results	31
2.2.9 Automating tasks	32
2.3 Data File Formats	33
2.4 Parameter Files	35
2.5 Defining a 3D Ellipsoid	37
3 Geostatistics: A Recall of Concepts	39
3.1 Random Variable	40
3.2 Random Function	44
3.2.1 Simulated realizations	45
3.2.2 Estimated maps	48
3.3 Conditional Distributions and Simulations	50
3.3.1 Sequential simulation	51
3.3.2 Estimating the local conditional distributions	53

3.4	Inference and Stationarity	55
3.5	The Variogram, a 2-point Statistics	60
3.6	The Kriging Paradigm	63
3.6.1	Simple kriging	63
3.6.2	Ordinary kriging and other variants	67
3.6.3	Kriging with linear average variable	70
3.6.4	Cokriging	72
3.6.5	Indicator kriging	74
3.7	An Introduction to mp Statistics	75
3.8	Two-point Simulation Algorithms	79
3.8.1	Sequential Gaussian simulation	80
3.8.2	Direct sequential simulation	80
3.8.3	Direct error simulation	81
3.8.4	Indicator simulation	83
3.9	Multiple-point Simulation Algorithms	85
3.9.1	Single normal equation simulation (<i>SNESIM</i>)	85
3.9.2	Filter-based algorithm (<i>FILTERSIM</i>)	86
3.9.3	Hierarchical simulation	87
3.10	The nu/tau Expression	88
3.11	Inverse problem	94
4	Data Sets & SGeMS EDA Tools	95
4.1	The Data Sets	95
4.1.1	The 2D data set	96
4.1.2	The 3D data set	96
4.2	The SGeMS EDA Tools	100
4.2.1	Common parameters	100
4.2.2	Histogram	101
4.2.3	Q-Q plot and P-P plot	102
4.2.4	Scatter plot	103
5	Variogram computation and modeling	107
5.1	Variogram Computation in SGeMS	109
5.1.1	Selecting the head and tail properties	112
5.1.2	Computation parameters	113
5.1.3	Displaying the computed variograms	116
5.2	Variogram Modeling in SGeMS	117

6	Graphical parameters interface	121
6.1	Algorithm Panel	121
6.2	Selecting a Grid and Property	122
6.3	Selecting Multiple Properties	123
6.4	Search Neighborhood	124
6.5	Variogram	125
6.6	Kriging	126
6.7	Line Entry	127
6.8	Non-parametric Distribution	127
6.8.1	Tie breaking	128
6.9	Errors in Parameters	130
6.10	XML Format for Parameters	130
7	Estimation Algorithms	133
7.1	<i>KRIGING</i> : Univariate kriging	133
7.2	<i>INDICATOR KRIGING</i>	137
7.3	<i>COKRIGING</i> : Kriging with secondary data	144
7.4	<i>BKRIG</i> : Block Kriging Estimation	148
8	Stochastic Simulation Algorithms	161
8.1	Variogram-based Simulations	161
8.1.1	<i>LUSIM</i> : LU simulation	162
8.1.2	<i>SGSIM</i> : sequential Gaussian simulation	164
8.1.3	<i>COSGSIM</i> : sequential Gaussian co-simulation	169
8.1.4	<i>DSSIM</i> : direct sequential simulation	173
8.1.5	<i>SISIM</i> : sequential indicator simulation	178
8.1.6	<i>COSISIM</i> : sequential indicator co-simulation	185
8.1.7	<i>BSSIM</i> : Block Sequential Simulation	190
8.1.8	<i>BESIM</i> : Block Error Simulation	197
8.2	Multiple-point Simulation algorithms	204
8.2.1	<i>SNESIM</i> : Single Normal Equation Simulation	206
8.2.2	<i>FILTERSIM</i> : Filter-based Simulation	232
9	Utilities	257
9.1	<i>TRANS</i> : Histogram Transformation	258
9.2	<i>TRANSCAT</i> : Categorical Transformation	261
9.3	Post-processing of Kriging Estimates	265
9.4	<i>POSTSIM</i> : Post-processing of Realizations	269

9.5	<i>NU-TAU MODEL</i> : Combining Probability Fields	271
9.6	<i>BCOVAR</i> : Block Covariance Calculation	274
9.7	<i>IMAGE PROCESSING</i>	277
9.8	<i>MOVING WINDOW</i> : Moving Window Statistics	280
9.9	<i>TIGENERATOR</i> : Object-based image generator	283
9.9.1	Object interaction	286
9.9.2	Examples	290
10	Scripting, Commands and Plug-ins	293
10.1	Commands	293
10.1.1	Command lists	294
10.1.2	Execute command file	297
10.2	Python script	298
10.2.1	SGeMS Python modules	298
10.2.2	Running Python scripts	299
10.3	Plug-ins	301
	List of Programs	303
	List of Symbols	305
	Bibliography	322

Chapter 3

Geostatistics: A Recall of Concepts

This user's manual is no place for another presentation of the theory of geostatistics. Many excellent books and reference papers are available to such purpose: Journel and Huijbregts (1978); Isaaks and Srivastava (1989); Cressie (1993); Wackernagel (1995); Goovaerts (1997); Deutsch and Journel (1998); Chilès and Delfiner (1999); Olea (1999); Lantuejoul (2002); Mallet (2002). In this chapter we will only review the basic concepts and geostatistical principles underlying the algorithms offered in SGeMS. The more recent developments of multiple-point geostatistics are presented in relatively greater length because they are less known. Engineering-type presentations are preferred over more rigorous but less intuitive developments. These presentations point to programs coded into SGeMS whenever available.

A warning about reference: We have limited citations to a few most relevant and easily accessible references. We gave page number to the only three books by Goovaerts (1997); Deutsch and Journel (1998) and Chilès and Delfiner (1999). For an extensive list of references the reader may turn to the list proposed by Cressie (1993) and that of Chilès and Delfiner (1999).

The concept of random variable presented in Section 3.1 is introduced to model the uncertainty about a single variable. Section 3.2 extends that concept to a random function modeling the joint uncertainty about several interdependent variables distributed in space. The various possible outcomes of a random variable or random function are controlled by probability distribution functions made conditional to the data available; simulated outcomes can then be drawn from these conditional distributions, as discussed in Section 3.3. Conversely, a set of simulated outcomes can define a random function which is then seen as algorithm-

driven, the algorithm being that used to generate these outcomes. At the core of any random function there is a structural model which indicates how the various constitutive random variables relate to each other and to the data; inference of such model necessarily requires a prior decision of stationarity, as discussed in Section 3.4. That structural model can be limited to a variogram-type relation between any two variables as presented in Section 3.5. The structural model can also involve many more than two variables at a time. In the latter case, inference of the corresponding multiple-point statistics calls for a training image. Section 3.6 presents the kriging paradigm which is at the origin of most geostatistical algorithms whether aimed at estimation or simulation. Section 3.7 introduces the theory underlying the multiple-point geostatistical algorithms. The traditional variogram-based simulation algorithms, *SGSIM*, *DSSIM* and *SISIM* are presented in Section 3.8. The two multiple-point simulation algorithms, *SNESIM* and *FILTERSIM*, are presented in Section 3.9. The nu/tau expression for compositing probabilities conditional to different data events is given in Section 3.10; this non traditional yet exact expression of the fully conditioned probability provides a useful separation of data information content and data redundancy.

3.1 Random Variable

The conceptual model at the root of geostatistics, and for that matter of all of statistics and probability theory, is that of a random variable or random function. This is the model that allows making uncertainty assessment about an imperfectly known value.

A deterministic variable takes only one outcome; that outcome is either known or unknown leaving no flexibility for uncertainty. Conversely, a random variable (RV) can be seen as a variable that can take a series of possible outcomes, each with a certain probability or frequency of occurrence (Goovaerts (1997, p.63); Deutsch and Journel (1998, p.11); Jensen et al. (1997)). A random variable is traditionally denoted with a capital letter, say, Z . Its possible outcomes are denoted with the corresponding small case letter, say, $\{z_i, i = 1, \dots, n\}$ for a discrete variable with n outcomes, or $\{z \in [z_{min}, z_{max}]\}$ for a continuous variable valued in the interval bounded by a maximum and minimum value.

In the discrete case, to each outcome z_i is attached a probability value

$$p_i = \text{Prob}\{Z = z_i\} \in [0, 1], \quad \text{with:} \quad \sum_{i=1}^n p_i = 1 \quad (3.1)$$

In the continuous case, the distribution of probability values can take the form of

- a cumulative distribution function (cdf), pictured as a cumulative histogram, providing the probability for the RV not to exceed a given threshold value z , see Fig. 3.1(a):

$$F(z) = \text{Prob} \{Z \leq z\} \in [0, 1] \quad (3.2)$$

- a probability density function (pdf) or histogram, defined as the derivative of the previous cdf at z -values of non discontinuity: $f(z) = dF(z)/dz$.

From such pdf or cdf probability intervals can be derived, see Fig. 3.1(b):

$$\text{Prob} \{Z \in (a, b]\} = F(b) - F(a) = \int_a^b f(z)dz \quad (3.3)$$

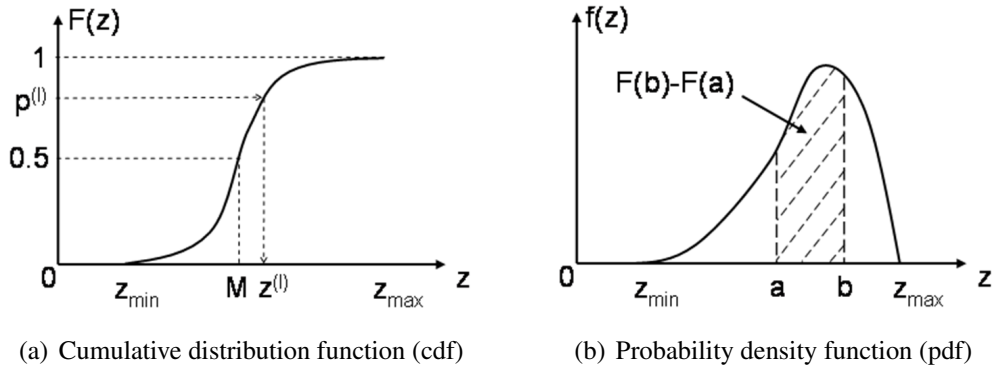


Figure 3.1: Probability distribution function

The key to a probabilistic interpretation of a variable z is the modeling of the distribution function, cdf or pdf, of the corresponding random variable Z . Note that such modeling does not mean necessarily fitting a parametric function to either the cdf or pdf; a series of classes with attached probability values is a valid model (Deutsch and Journal, 1998, p.16). That distribution function should account for all information available; it then provides all that is needed to quantify the uncertainty about the actual outcome of the variable z . For example,

- probability intervals can be derived as in Eq. 3.3;
- quantile values can be derived such as the .1 quantile or 1st decile:

$$q_{0.10} = F^{-1}(0.10) = z\text{-outcome value such that } \text{Prob}\{Z \leq q_{0.10}\} = 0.10$$

- simulated values can be drawn by reading quantile values $z^{(l)}$ corresponding to a series of random numbers $p^{(l)}, l = 1, \dots, L$ uniformly distributed in $[0, 1]$:

$$z^{(l)} = F^{-1}(p^{(l)}), \quad l = 1, \dots, L \quad (3.4)$$

This process, called Monte Carlo drawing, ensures that the cdf of the L values $z^{(l)}$ will reproduce the Z -cdf $F(z)$, see Fig. 3.1(a) and Goovaerts (1997, p.351), Deutsch and Journel (1998, p.154). Conversely, a random variable Z can be modeled by the distribution of a number L of simulated values $z^{(l)}$ generated from a process initiated with equally probable uniform random numbers $p^{(l)}$: this is the concept of algorithm-driven random variable, see hereafter and Section 3.3, Deutsch (1994a).

From the distribution of Z , specific moments or characteristic values can be derived such as,

- the mean m , or expected value of the RV Z , which can be retained as an estimate of the unknown value z , best in a least squared error sense. This mean is here written for a continuous variable Z in terms of its cdf, pdf, or as the arithmetic mean of L equiprobable realizations $z^{(l)}$ if the RV is so defined:

$$\begin{aligned} m &= E\{Z\} = \int_{z_{\min}}^{z_{\max}} z dF(z) = \int_{z_{\min}}^{z_{\max}} z f(z) dz \\ &= \frac{1}{L} \sum_{l=1}^L z^{(l)} \end{aligned} \quad (3.5)$$

- the median, or .5 quantile $q_{.50}$, a z -value which leaves 50% of the possible outcome values above it and 50% below it, see Fig. 3.1(a). The median can be used as yet another estimate of the unknown value z , best in a least absolute error sense:

$$M = q(0.50) : \text{value such that } \text{Prob}\{Z \leq M\} = 0.50 \quad (3.6)$$

If the RV Z is defined through L realizations $z^{(l)}$ with L an even number, trivial adjustment to the definition of the median realization is needed; similar adjustments may be needed for other quantile values.

- the variance which can be used as a single summary of the uncertainty around the mean estimate m :

$$\begin{aligned}\sigma^2 &= Var \{Z\} = E \{(Z - m)^2\} = \int_{z_{min}}^{z_{max}} (z - m)^2 f(z) dz \\ &= \frac{1}{L} \sum_{l=1}^L (z^{(l)} - m)^2\end{aligned}\quad (3.7)$$

Beware that the two most-used moments, mean and variance, generally do not suffice by themselves to define a distribution, hence to define probability intervals such as given in relation 3.3. Often a Gaussian-related distribution is adopted providing the missing information. The problem is that errors associated to the various data integration processes involved in spatial interpolation are almost never Gaussian-distributed as opposed to direct errors due to measurement devices.

One definite advantage of a RV modeled through a set of L realizations $z^{(l)}$ is that probability intervals can be defined without going through any variance calculation. Also these probability intervals are independent of the particular estimated value retained, as opposed to the variance (Eq. 3.7) which is specific to the mean estimate m . If one accepts that there is no unique ‘best in absolute’ estimated value for any unknown, probability intervals and uncertainty measures should indeed be independent of the particular estimated value retained (Srivastava (1987); Goovaerts (1997, p.340); Isaaks and Srivastava (1989)).

Algorithm-driven random variable

One can argue that all of predictive geostatistics amounts to the determination of a probability distribution model, a model that accounts for *all* information available about the unknown value(s) z . A distribution model cannot be reduced to its mean and variance unless some two-parameter distribution is adopted; one must then question why mean and variance should be carefully determined if the far more consequential distribution type retained is not appropriate. Instead of determining mean and variance of the possible outcomes of an unknown, modern geostatistics aims at building a process (an algorithm) mimicking the data environment of that unknown; that algorithm then allows generating many (L) alternative outcomes of

that unknown, although possibly not all of them. These L simulated realizations $z^{(l)}$ define an algorithm-driven random variable from which probability intervals for the unknown can be retrieved, as well as an estimated value which need not be the mean (Journel, 1993b; Deutsch, 1994a).

The number L of realizations can be as large as can be comfortably processed (Deutsch and Journel (1998, p.133); Chilès and Delfiner (1999, p.453)). Note that a different set of L' realizations, with possibly $L = L'$, actually defines a different random variable. The number L is part of the defining algorithm.

This book and the SGeMS software provide tools for building up these models of uncertainty. The details are in the ‘how to’ and in the presentation, mostly graphical (maps), of the results.

Beware that a model of uncertainty is just that, a model, and there could be alternative models, each delivering possibly different results such as different estimates of the unknown, yet using the same original information (data) but in a different manner. There is no unique model of uncertainty, and most troublesome, there is neither a ‘best’ model nor a fully objective model. We will return repeatedly to that point. Geostatistics, and for that matter all of probability theory, can only provide consistency with a prior model necessarily partly subjective, it cannot provide fully objective decisions (Goovaerts (1997, p.442); Chilès and Delfiner (1999, p.22); Matheron (1978); Journel (1993b); Dubrule (1994)).

3.2 Random Function

Most applications of geostatistics in the earth sciences involve mapping, which is the joint consideration of several variables in space and/or time. Some of these variables are known through sampling; most others are unknown with varying degrees of uncertainty, they should therefore be modeled as random variables. However, we are not interested in evaluating each unknown independently of the others nearby; we are interested in an assessment of the *joint* spatial distribution of all unknowns, which is an assessment of their relation and connectivity in space. The uncertainty modeling should therefore consider all unknown variables together. The concept of a random function answers that requirement (Goovaerts (1997, p.68); Chilès and Delfiner (1999, p.12)).

A Random Function (RF), denoted $Z(\mathbf{u})$, is a set of *dependent* random variables $\{Z(\mathbf{u}), \mathbf{u} \in S\}$, each marked with a coordinate vector \mathbf{u} spanning a field or study area S . That field is typically a 3D physical volume, in which case $\mathbf{u} = (x, y, z)$ is the vector of the 3 Cartesian coordinates; the common notation (z)

for the variable and the vertical coordinate does not usually pose problem. The variable could also be time in which case $\mathbf{u} = t$, or it could involve both space and time as for atmospheric pressure in which case $\mathbf{u} = (x, y, z, t)$.

Just like a single random variable Z is characterized by a distribution function, say its cdf $F(z)$ for a continuous variable, a RF $Z(\mathbf{u})$ would be characterized by its multivariate distribution function:

$$\text{Prob} \{Z(\mathbf{u}) \leq z, \mathbf{u} \in S\} \quad (3.8)$$

a function of many parameters, any number N of locations \mathbf{u} in S and the corresponding threshold values z possibly all different from one location to another.

The analytical expression of such multivariate distribution is highly impractical on sizeable grid; there can be many millions of locations in a 3D grid! Exceptions are analytical distributions defined from a very small number of parameters, e.g. Gaussian-related (Anderson (2003); Goovaerts (1997, p.265); Chilès and Delfiner (1999, p.404)). But parameter-poor distributions are very specific in their properties, hence very restrictive. Last and not least, any advantage provided by an analytical definition vanishes when data locations and values are included into the field S , unless these data honor exactly the prior RF (unlikely in practice). This process of including into the RF $Z(\mathbf{u})$ random variables that are actually sampled is known as ‘data conditioning’.

3.2.1 Simulated realizations

Just like a single RV can be defined by a finite set of simulated realizations, a RF $Z(\mathbf{u})$ is displayed and used through its realizations $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in S\}$, $l = 1, \dots, L$ (Lantuejoul (2002); Isaaks and Srivastava (1989); Goovaerts (1997, p.369); Chilès and Delfiner (1999, p.449); Deutsch and Journel (1998, p.119)). In practice these realizations take the form of a finite number L of simulated maps, each providing an alternative, equally probable, representation of the unknown ‘true’ map $z(\mathbf{u}), \mathbf{u} \in S$, see Fig. 3.2.a. Any one specific such realization is denoted $z^{(l)}(\mathbf{u}), \mathbf{u} \in S$, where the upper script l indicates the id number of that specific realization. A realization can be seen as a numerical model of the possible distribution in space of the z -values. That numerical model can be used to different purposes, including visualization and input to some transfer function representing the process under study. e.g. mining out the z -grade values.

Ideally, one would like to have access to a RF such that one of its realizations identifies the actual distribution of the true values $z(\mathbf{u}), \mathbf{u} \in S$; this is very unlikely

if S contains millions or even only thousands of unknown values $z(\mathbf{u})$. In practice, this latter limitation is not a matter of concern if the L realizations available allow a reasonable assessment of the consequent uncertainty on any processing of these unknown values. Recall the previous warning that a RF model is just a model, and asking that this model includes the unknown reality is naive.

It is the set of all such L simulated realizations, not any single realization, which provides an uncertainty assessment of the spatial distribution of the z -values over the study area. For example,

- the probability distribution of the unknown $z(\mathbf{u})$ at any particular location \mathbf{u} can be retrieved from the L simulated values $z^{(l)}(\mathbf{u})$ at that same location \mathbf{u} . This is much more than a mere estimation of $z(\mathbf{u})$, even if attached with an error variance since an error variance does not suffice to specify an error distribution. In a spatial interpolation setting, one could easily check through cross-validation that the cdf of the L simulated error values $[z^{(l)}(\mathbf{u}) - z^*(\mathbf{u})]$, $l = 1, \dots, L$, takes very different shapes depending on the data environment of location \mathbf{u} and these shapes could be quite non-Gaussian.
- the probability that two nearby unknown values $z(\mathbf{u})$ and $z(\mathbf{u}')$ be simultaneously greater than any given threshold z_0 can be evaluated by the proportion of the L realizations which display simultaneously high simulated values at these two locations.

The reader should convince himself that this result could not be obtained, in general, from sole knowledge of the two cdf's

$$F(\mathbf{u}, z_0) = \text{Prob} \{Z(\mathbf{u}) \leq z_0\} \quad \text{and} \quad F(\mathbf{u}', z_0) = \text{Prob} \{Z(\mathbf{u}') \leq z_0\};$$

indeed the uncertainties related to two nearby values $z(\mathbf{u})$ and $z(\mathbf{u}')$ are not in general independent, hence the two RVs $Z(\mathbf{u})$ and $Z(\mathbf{u}')$ are not independent, and their two cdf's cannot be combined straightforwardly.

- the probability that there exists a connected path of high z -values between two distant locations \mathbf{u} and \mathbf{u}' can be similarly evaluated by the proportion of simulated realizations (out of the L available) displaying such connected paths.

A great part of the SGeMS software deals with the generation of simulated realizations of the type shown in Fig. 3.2.a, and their utilization in assessing spatial uncertainty, that is, uncertainty involving *jointly* many different locations in space.

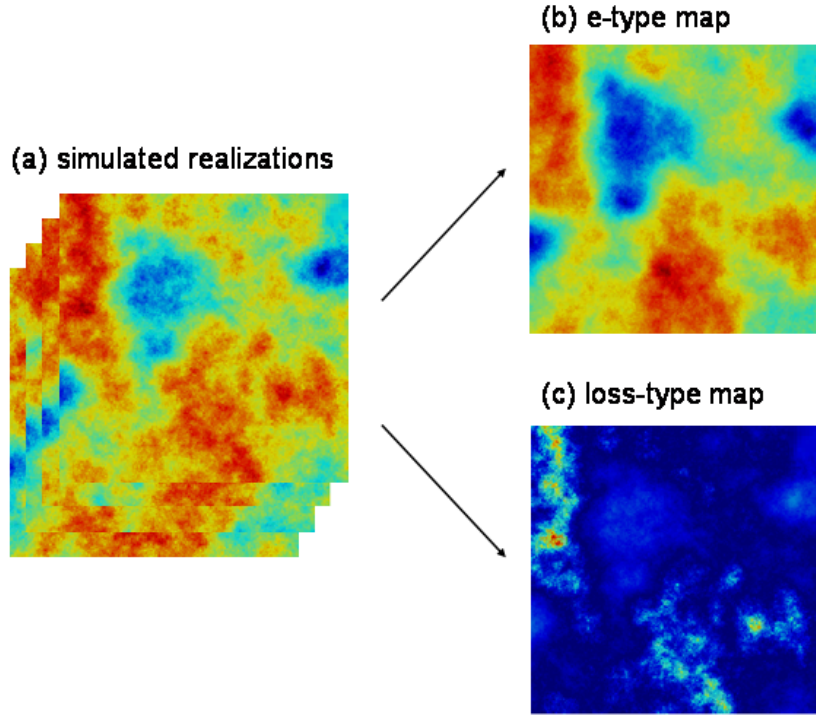


Figure 3.2: Simulated realizations and different estimation maps; a. alternative equiprobable realizations of a random function; b. E-type estimated map minimizing local squared error; c. estimated map minimizing a specific loss function of the local error

The issue of equiprobable realizations

Through the cdf $F(z) = \text{Prob}\{Z \leq z\}$ of a single analytically-defined RV, one can define the probability of any outcome value z or class thereof. The case of an algorithm-driven RF is more delicate, if only because an analytical multivariate cdf characterizing that RF is rarely available; in practice it is never available after data conditioning. Indeed real and reasonably complex data are rarely perfectly consistent with a necessarily restrictive analytically-defined RF model, hence conditioning that model to these data changes the model in an unpredictable way depending on the algorithm retained for the conditioning process.

In the vast majority of practical applications, the RF is defined through a finite number L of simulated realizations $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in S\}, l = 1, \dots, L$. If any number $n \leq L$ of realizations are *exactly* identical, these n realizations can be collated

into a single one with probability n/L . However, in most practical applications no two realizations would be exactly identical, in which case all realizations are equiprobable, each with probability $1/L$.

How many realizations?

How many realizations (L) should be drawn from a given algorithm-driven RF model (Chilès and Delfiner (1999, p.453); Deutsch and Journel (1998, p.133))? The number L is part of the algorithm defining the RF model itself, hence the question could be dismissed. However, it is a fair question for practice: one could process $L = 10$, possibly 100, not comfortably 1000 or 10,000 realizations. Since there is no reference model to approach, the number L should be chosen large enough to ensure stability of the results and small enough to allow the intended processing of the L simulated realizations.

Consider as ‘result’ a specific function $\varphi(z^{(l)}(\mathbf{u}), \mathbf{u} \in S)$ built from anyone simulated realization. The number L of realizations should be large enough such that, e.g., the variance of the L' values $\{\varphi(z^{(l)}(\mathbf{u}), \mathbf{u} \in S), l = 1, \dots, L'\}$ stabilizes as L' increases towards L .

3.2.2 Estimated maps

There can be applications where values $z(\mathbf{u})$ are dealt one at a time independently of the next one $z(\mathbf{u}')$, no matter how close are the locations \mathbf{u} and \mathbf{u}' . We will argue that those applications are few; even the most selective mining or environmental cleaning of single values $z(\mathbf{u})$ do account for close-by values $z(\mathbf{u}')$, e.g. for reason of cost and/or accessibility (Journel and Huijbregts, 1978; Isaaks and Srivastava, 1989). Notwithstanding, consider the derivation of a single-valued estimate $z^*(\mathbf{u})$ at each unsampled location \mathbf{u} .

Since the L simulated realizations of any single unsampled RV $z(\mathbf{u})$ are equiprobable, their point-wise arithmetic average provides a single estimated value, of least squared error-type, also called the E-type estimated value where E is short for ‘expected value’, more precisely ‘conditional expectation’ (Goovaerts (1997, p.341); Deutsch and Journel (1998, p.81)):

$$z_E^*(\mathbf{u}) = \frac{1}{L} \sum_{l=1}^L z^l(\mathbf{u}). \quad (3.9)$$

The E-type map corresponding to the L realizations of Fig. 3.2a is shown in Fig. 3.2b. Note how much ‘smoother’ the estimated map is compared to anyone

of the L simulated realizations. What is missing in an estimated map is the joint dependency of the many variables $Z(\mathbf{u})$, $\mathbf{u} \in S$, beyond they sharing common data. Far away from the data locations two estimated values would typically be identical; in the case of simple kriging they would be equal to the data mean value (Isaaks and Srivastava, 1989; Goovaerts, 1997, p.130, 369). Clearly geological or physical heterogeneity should not vanish just because there is no data nearby!

As a consequence, the probability that two nearby unknown values $z(\mathbf{u})$ and $z(\mathbf{u}')$ be simultaneously greater than any given threshold z_0 could not be evaluated from an estimated map. More generally any assessment or estimation involving more than one location at a time should not be made on an estimated map, particularly if that estimation is based on sparse data which is the case in many earth sciences applications.

In all rigor a set of estimated values should only be tabulated, not displayed as a map; a map entices the user to read relations between estimated z^* -values at different locations when these relations may not reflect that between the actual z -values. Only simulated values $z^{(l)}(\mathbf{u})$ should be mapped because a simulated realization is precisely a representation of the RF modeling the joint distribution in space of all random variables $Z(\mathbf{u})$, $\mathbf{u} \in S$. Beware, however, that if an estimated map is unique, there are many alternative equiprobable simulated realizations for any given data set. A simulation work ethic forces the user to face uncertainty as represented by the L simulated realizations generated.

Alternative estimation criteria

Instead of the mean value defining the E-type estimated map as in expression (3.9), one could have retained the median $z_M^*(\mathbf{u})$ of the L simulated values $z^{(l)}(\mathbf{u})$ at each location \mathbf{u} . This would define a M-type estimated map, where each estimated value $z_M^*(\mathbf{u})$ has a 50% chance to be higher (or lower) than the actual unknown value. It can be shown that such M-type estimate is ‘best’ in a least absolute error sense. In any particular application where single-location estimated values are needed, there is no a priori reason to minimize the squared error (e^2) or the absolute error ($|e|$); one may want to minimize an application-specific loss function $\text{Loss}(e)$, for example, in environmental applications using a non-linear function $\text{Loss}(\cdot)$ that penalizes more underestimation of a lethal pollutant than overestimation. Availability of the L simulated maps $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in S\}$, $l = 1, \dots, L$, allows considering such loss function-specific estimate, see Fig. 3.2.c, Srivastava (1987), Goovaerts (1997, p.340).

Kriging could provide directly and faster an estimated map, similar but not

identical to the E-type map but it does not allow the flexibility to consider other types of estimates. In addition, the kriging variance map is an incomplete, when not misleading, measure of uncertainty as opposed to the distribution provided by the L simulated maps $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in S\}$, $l = 1, \dots, L$, see later discussion in Section 3.6.1, and Goovaerts (1997, p.180), Journel (1986), Chilès and Delfiner (1999, p.178).

3.3 Conditional Distributions and Simulations

As already stated, the main task of any probabilistic assessment is to build a model for the probability distribution of the unknown(s), either taken one at a time as displayed by the histogram of Fig. 3.1(b), or altogether as displayed by the set of simulated realizations of Fig. 3.2a. The uncertainty of any unknown, or any number of unknowns taken jointly, necessarily depends on the amount and types of data available and their assumed relation to the unknown(s) considered. Take the simple example of a single unsampled continuous variable at location \mathbf{u} , and denote by $n(\mathbf{u})$ the set of data informing it. The relevant cdf providing an assessment of the uncertainty about the unsampled value $z(\mathbf{u})$ is specific to the location \mathbf{u} and the data set $n(\mathbf{u})$ and is written (Chilès and Delfiner (1999, p.380); Goovaerts (1997, p.69)):

$$F(\mathbf{u}; z|n(\mathbf{u})) = \text{Prob} \{Z(\mathbf{u}) \leq z|n(\mathbf{u})\},$$

in words, the probability that the unknown $Z(\mathbf{u})$ be valued no greater than the threshold value z conditional to (knowing) the data set $n(\mathbf{u})$.

That conditional probability is, by definition, equal to the following ratio, with as numerator the probability of the event to be assessed $Z(\mathbf{u}) \leq z$ occurring jointly with the data event, and as denominator the probability of that data event occurring:

$$\text{Prob} \{Z(\mathbf{u}) \leq z|n(\mathbf{u})\} = \frac{\text{Prob} \{Z(\mathbf{u}) \leq z, n(\mathbf{u})\}}{\text{Prob} \{n(\mathbf{u})\}}. \quad (3.10)$$

Expression (3.10) makes explicit the dependence of that cdf on the location \mathbf{u} , more precisely, the relation of that location with the $n(\mathbf{u})$ data retained. In all rigor, one should also make explicit the type, location and value of each datum constituting the set $n(\mathbf{u})$. Indeed if any aspect of that data set changes, the distribution (3.10) is changed.

The distribution (3.10) is called the conditional cumulative distribution function (ccdf) of the specific RV $Z(\mathbf{u})$ given the data set $n(\mathbf{u})$. When many unknowns

$\{z(\mathbf{u}), \mathbf{u} \in S\}$ are jointly involved, the conditional probability required becomes multivariable; it is written as (Goovaerts (1997, p.372); Anderson (2003); Johnson (1987)):

$$\text{Prob} \{Z(\mathbf{u}) \leq z, \mathbf{u} \in S | n(\mathbf{u})\} = \frac{\text{Prob} \{Z(\mathbf{u}) \leq z, \mathbf{u} \in S, n(\mathbf{u})\}}{\text{Prob} \{n(\mathbf{u})\}}. \quad (3.11)$$

Probability distributions of type (Eq. 3.1) or (Eq. 3.2) which are not made conditional to the data available are of little practical interest. Similarly, it is the moments of the conditional distributions which are of practical interest, and only those should be used for estimation. For example it is the mean of the ccdf $F(\mathbf{u}; z | n(\mathbf{u}))$ which should be used as the least squared error estimate of the unknown $z(\mathbf{u})$ at location \mathbf{u} , not the mean of the marginal distribution $F(z)$ as defined in Eq. 3.2 since that marginal distribution does not account for the specific dependence of location \mathbf{u} with the data.

Similarly, the L realizations $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in S\}$, $l = 1, \dots, L$, displayed in Fig. 3.2.a are useful only if they are outcomes of the multivariate probability distribution (Eq. 3.11) conditioned to all relevant data available over the study field S (Goovaerts, 1997, p.372). For example, the arithmetic average of the L simulated values $z^{(l)}(\mathbf{u})$ at any given location \mathbf{u} provides an estimate of the unknown $z(\mathbf{u})$ at that location; the cumulative histogram of these L simulated values provides a discrete representation of the ccdf (Eq. 3.10), itself a measure of uncertainty about $z(\mathbf{u})$ (Journel and Huijbregts (1978); Isaaks and Srivastava (1989); Goovaerts (1997, p.180)).

Consider a specific zone or block V within S ; each simulated map provides a simulated value for the average z -value over V denoted $z_V^{(l)}$, then the histogram of the L simulated values $z_V^{(l)}$, $l = 1, \dots, L$ provides a measure of uncertainty about the unknown average value z_V (Journel and Kyriakidis, 2004). Instead of a mere average taken over the zone V , any complex non-linear function of the z -values over all or part of the field S could be considered.

3.3.1 Sequential simulation

How does one go about building a complex distribution such as that in expression 3.11 involving jointly many unknowns and many data possibly of different types? This is indeed the main challenge of geostatistics. A solution to such formidable task is provided by the ‘divide and conquer’ paradigm:

1. divide the problem into a series of easier problems involving only one unknown at a time, that is address the easier problem of determining the ccdf

(Eq. 3.10) of each unknown $z(\mathbf{u})$. The task of recombining these elementary conditional probabilities accounting for spatial dependence underlies the sequential simulation algorithm (Deutsch and Journel (1998, p.125); Goovaerts (1997, p.390); Chilès and Delfiner (1999, p.462); Rosenblatt (1952));

2. divide the possibly large and complex data set $n(\mathbf{u})$ constituted of many different data types into a set of smaller more homogeneous data sets $n_k(\mathbf{u})$, $k = 1, \dots, K$, and address the easier problem of determining the K ccdf's $\text{Prob}\{Z(\mathbf{u}) < z | n_k(\mathbf{u})\}$ conditioned to each of the smaller data set $n_k(\mathbf{u})$. The nu/tau model presented in Section 3.9 addresses the problem of combining these K ccdf's into a single one of type 3.10, (Journel, 2002; Krishnan, 2004; Bordley, 1982; Polyakova and Journel, 2008).

We will develop the first divide paradigm by considering the case of only three interdependent unknowns $z(\mathbf{u}_1)$, $z(\mathbf{u}_2)$, $z(\mathbf{u}_3)$, at different locations \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 . These three interdependent variables could also relate to three different attributes, say the grades of three different metals. The generalization to more than three unknowns is immediate. The joint pdf of three random variables conditional to the same data set (n) can be decomposed as (Goovaerts, 1997, p.376):

$$\begin{aligned} \text{Prob}\{Z(\mathbf{u}_1) = z_1, Z(\mathbf{u}_2) = z_2, Z(\mathbf{u}_3) = z_3 | (n)\} = \\ \text{Prob}\{Z(\mathbf{u}_1) = z_1 | (n)\} \cdot \\ \text{Prob}\{Z(\mathbf{u}_2) = z_2 | (n), Z(\mathbf{u}_1) = z_1\} \cdot \\ \text{Prob}\{Z(\mathbf{u}_3) = z_3 | (n), Z(\mathbf{u}_1) = z_1, Z(\mathbf{u}_2) = z_2\} \end{aligned} \quad (3.12)$$

In words, the tri-variate joint pdf has been decomposed into the product of three univariate conditional pdf's, each involving only one variable, $Z(\mathbf{u}_1)$ first, then $Z(\mathbf{u}_2)$, last $Z(\mathbf{u}_3)$, but with increased data conditioning.

Provided that the problem of conditioning to the common data set (n) can be solved (see hereafter), each of these three single variable conditional pdf's can be determined. The decomposition (Eq. 3.12) then allows the process of sequential simulation (Deutsch and Journel, 1998, p.125), more precisely:

- a value for $Z(\mathbf{u}_1)$ is drawn from the first pdf $\text{Prob}\{Z(\mathbf{u}_1) = z_1 | (n)\}$, say that simulated value is $z_1^{(l)}$;
- next a value for $Z(\mathbf{u}_2)$ is drawn from the second pdf $\text{Prob}\{Z(\mathbf{u}_2) = z_2 | (n), Z(\mathbf{u}_1) = z_1^{(l)}\}$, say that value is $z_2^{(l)}$;

- last a value for $Z(\mathbf{u}_3)$ is drawn from the third pdf
 $\text{Prob} \left\{ Z(\mathbf{u}_3) = z_3 | (n), Z(\mathbf{u}_1) = z_1^{(l)}, Z(\mathbf{u}_2) = z_2^{(l)} \right\}$, say that value is $z_3^{(l)}$;

The three simulated values $z_1^{(l)}, z_2^{(l)}, z_3^{(l)}$, although drawn in sequence one after the other, stem from the joint tri-variate distribution conditional to the common data set (n) .

- If another set of three simulated values is needed, one can repeat the process using different random numbers for the drawings.

The interdependence between the three variables $z_1^{(l)}, z_2^{(l)}, z_3^{(l)}$, has been taken into account by conditioning the simulation of each single variable to values of all previously simulated variables. We have traded the problem of simulating jointly many variables for that of simulating only one variable at a time but with an increasing conditioning data set, from (n) to $(n + 1)$ then $(n + 2)$. In practice, the problem created by the increasing data set size is solved by retaining into the conditioning data set of each variable only the closest or most related previously simulated values (Gómez-Hernández and Cassiraga (1994); Goovaerts (1997, p.390,400)).

Retaining only the $n(\mathbf{u})$ closest data to inform any unknown location \mathbf{u} amounts to an approximation of Eq. 3.12 since not all the previously simulated variables are taken into consideration. On the other hand, retaining only the closest data allows tighter local conditioning. A consequence of retaining only the closest previously simulated values is that the sequence along which the nodes are visited matters. That sequence is called the simulation path, it is usually random to avoid artifacts (Daly and Verly, 1994).

The joint-pdf in Eq. 3.12 using the sequence $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and retaining only one previously simulated value in addition to the original (n) becomes, with \mathbf{u}_2 being closer to \mathbf{u}_3 than \mathbf{u}_1 :

$$\begin{aligned}
 \text{Prob} \{ Z(\mathbf{u}_1) = z_1, Z(\mathbf{u}_2) = z_2, Z(\mathbf{u}_3) = z_3 | (n) \} &\approx \\
 \text{Prob} \{ Z(\mathbf{u}_1) = z_1 | (n) \} &\cdot \\
 \text{Prob} \{ Z(\mathbf{u}_2) = z_2 | (n), Z(\mathbf{u}_1) = z_1 \} &\cdot \\
 \text{Prob} \{ Z(\mathbf{u}_3) = z_3 | (n), Z(\mathbf{u}_2) = z_2 \} &
 \end{aligned} \tag{3.13}$$

Another realization could be obtained by changing the uniform random numbers used for the drawing and/or changing the sequence in which the locations $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ are visited and simulated.

3.3.2 Estimating the local conditional distributions

The critical step in sequential simulation consists of estimating at each location \mathbf{u} along the simulation path the conditional distribution given a specific conditioning data set ($n(\mathbf{u})$). There have been essentially two avenues for approaching the determination of the single variable $Z(\mathbf{u})$ conditional pdf $\text{Prob}\{Z(\mathbf{u}) = z | n(\mathbf{u})\}$, both calling for a multiple-point (mp) RF model:

1. The traditional 2-point statistics approach consists of evaluating the relation of the single unknown $Z(\mathbf{u})$ with one datum $Z(\mathbf{u}_\alpha)$ at a time; thus no more than 2 locations or 2 variables are involved at any time. Such relation typically takes the form of a covariance/correlation or, equivalently, a variogram; these are 2-point statistics. Then calling on a prior multiple-point model requiring only 2-point stats for its calibration, the previous conditional pdf $\text{Prob}\{Z(\mathbf{u}) = z | n(\mathbf{u})\}$ is determined through some form of kriging. Examples of such simple mp models that can be calibrated from only 2-point stats are:
 - the multivariate Gaussian model underlying the sequential Gaussian simulation algorithm (SGeMS program *SGSIM*, Section 8.1.2), Goovaerts (1997, p.380); Anderson (2003); Chilès and Delfiner (1999, p.462); Gómez-Hernández and Journel (1993)
 - the truncation at order two of an expansion of the exact conditional probability in the discrete case. Such truncation underlies the indicator simulation algorithm (SGeMS program *SISIM*), see the decomposition Eq. 3.12 and Section 3.8.4, Goovaerts (1997, p.393), Journel and Alabert (1989).

In short, a 2-point statistics approach aims at dividing the data set (n) into single locations or variables. First each single data variable is related to the single unknown (1+1=2-point statistics), then these elementary results are pieced together through kriging using some simple prior multiple-point (mp) probabilistic model. The results are no better than this prior mp model, be it Gaussian-related or algorithm-driven.

2. The second approach avoids such extreme division of the conditioning data event. Instead it consists of using an explicit multiple-point (mp) model, which allows considering the (n) data altogether, or together a set $n(\mathbf{u})$ of neighbor data. The necessary $n(\mathbf{u}) + 1$ multiple-point statistics are lifted

from replicates of the $n(\mathbf{u})$ -data event found in a visually explicit training image (Ti) (Guardiano and Srivastava, 1993; Srivastava, 1994; Strebel, 2000, 2002; Zhang, et al., 2006). The results of such explicit mp geo-statistics application are no better than the prior model implicit to the Ti used. The consideration of training images, if available, allows making use of the mp structural information carried by these Ti's much beyond their variograms.

A training image is a representation of how z values are jointly distributed in space (Farmer, 1992; Strebel, 2002; Journel, 2002; Zhang, 2006). A training image (Ti) is essentially an unconditional realization of a RF model $Z(\mathbf{u})$, that is a prior conceptual depiction of the distribution in space of z -values, a depiction that need not honor at their location any of the data values included in the set (n) . The joint distribution in space of the actual unknown values $\{z(\mathbf{u}), \mathbf{u} \in S\}$ is assumed to 'look like' the Ti but would also honor the data (n) . The role of a mp simulation is strictly one of data conditioning, 'morphing' the Ti to honor the conditioning data (n) . A 2-point simulation aims at generating simulated realizations that honor the data and a variogram model. A mp simulation aims at generating simulated realizations that honor the data and the multiple-point structures present in the training image.

The necessity of an mp model

It is important to understand that there cannot be any probabilistic estimation or simulation without the necessary multiple-point (mp) statistics linking the data taken altogether to the unknown(s) (Journel, 1993b). Those mp statistics are either delivered explicitly through an analytical multivariate model or a training image or they are implicitly provided by the specific simulation algorithm retained. Traditional algorithms that call for input of only two-point statistics (variograms) adopt implicitly the higher order statistics in-built into the simulation algorithm retained, and these are most often of high entropy character. High or maximum entropy leads to maximizing disorder beyond the input variogram model(s) (Goovaerts (1997, p.335); Journel and Alabert (1989); Journel and Deutsch (1993)). An implicit model that maximizes disorder beyond specified two-point statistics is no less a model than a training image with its specific (lower entropy) structures and patterns much beyond the reach of a mere variogram. Also it can be argued that a high entropy model is often inappropriate for an earth sciences application where complex curvilinear structures involving many more than 2 locations in space are

known to exist, even if these structures are not immediately apparent from the limited local data available. The 2-point stats-based realizations are consistent with the implicit maximum entropy hypothesis beyond the input variogram model. If, however, one has any inkling of the existence of definite structures or patterns, this is precious structural information that must be accounted for in addition to the local data in the exercise of building alternative representations of the true image (Journel and Zhang, 2006). Two-point statistics, covariance or variogram, do not suffice to carry such mp information.

3.4 Inference and Stationarity

The concept of stationarity is at the basis of all probabilistic inference: you try to associate the (data) environment of any unknown value to ‘similar’ environments for which you know the outcome of the variable, this allows you to make prediction of the unknown from the known outcomes. The critical decision is that of similarity of the data environments, a decision which is never fully objective even though it defines the probabilistic model and thus impacts critically the predictions made (Goovaerts (1997, p.70); Chilès and Delfiner (1999, p.16); Deutsch and Journel (1998, p.12); Wackernagel (1995); Journel and Huijbregts (1978)).

Consider the most elementary problem of inferring the possible outcomes of a single unknown value z , which is inference of the distribution of the corresponding random variable Z . To be more specific, consider a petroleum reservoir and say that z is the unsampled porosity value at a given location \mathbf{u} , in which case the corresponding RV is denoted $Z(\mathbf{u})$. Many alternative decisions, all somewhat subjective, are possible:

- One may associate broadly the environment of the unknown $z(\mathbf{u})$ to the entire reservoir S which includes the location \mathbf{u} ; in which case the distribution of $Z(\mathbf{u})$ could be inferred from the histogram of all samples $z(\mathbf{u}_\alpha)$ available in S whether the sample location \mathbf{u}_α belongs or not to the lithofacies prevailing at location \mathbf{u} . The decision of stationarity then encompasses the whole reservoir.
- If one knows that the unsampled location \mathbf{u} is within a sand facies, it would make sense to restrict the previous histogram to only those samples known to have been collected in a sand facies. The decision of stationarity is now restricted to the sand facies. There is one caveat however: there should be

enough sand samples in S for their porosity histogram to be deemed representative; if not, one may have to pool samples from different lithofacies into the same histogram, or consider using sand porosity samples coming from deposits other than S but deemed similar. Thus, the decision of stationarity is necessarily subjective, conditioned in particular by data availability; that decision will change as the deposit S matures becoming better known and sampled. Yet, a different decision of stationarity implies a different probabilistic model, different data and different estimation results. The deposit S is the same, it is our model of it which has changed.

- Consider the favorable case of a well sampled deposit S where enough sand porosity samples are available to build a histogram deemed reliable. Should that histogram be considered as the probabilistic model for the location-specific RV $Z(\mathbf{u})$? Among those sand samples, one may want to give more weight to samples $z(\mathbf{u}_\alpha)$ at locations \mathbf{u}_α closer to the unsampled location \mathbf{u} , and also give more weight to isolated sample locations as opposed to clustered sample locations to reduce the impact of preferential over-sampling (data clustering) in certain zones. This suggestion is at the basis of the concept of kriging (Krige, 1951; Matheron, 1970; Journel and Huijbregts, 1978). In the process of kriging the Euclidean distance between any two (sand) locations $|\mathbf{u} - \mathbf{u}'|$ is replaced by a variogram distance $\gamma(\mathbf{u}, \mathbf{u}')$ read from a variogram model, itself inferred from the sand porosity samples. Inference of that variogram model calls for extending the decision of stationarity to pairs of sample values $z(\mathbf{u}_\alpha)$ and $z(\mathbf{u}_\beta)$ separated by approximately the same distance vector $\mathbf{h} = \mathbf{u}_\beta - \mathbf{u}_\alpha$ (Deutsch and Journel (1998, p.43); Goovaerts (1997, p.50)). A model of isotropy, whereby only the modulus of vector \mathbf{h} is retained, actually corresponds to yet another extension of the stationarity decision allowing the pooling of sample pairs with the same separation distance $|\mathbf{u}_\beta - \mathbf{u}_\alpha|$ irrespective of the direction of the vector $\mathbf{u}_\beta - \mathbf{u}_\alpha$. In the end, what is important is not kriging but the series of decisions of stationarity which allowed implementing that kriging, in our case stationarity of porosity in sand, then stationarity (of order 2) allowing inference of the variogram model, last and not least the local stationarity decision used to specify how far away from location \mathbf{u} should one go (within sand) to define the data event $n(\mathbf{u})$ informing that location \mathbf{u} .

In many applications involving subsurface deposits where local data are sparse, inference of a variogram is difficult particularly if the decision of stationarity re-

restricts samples to one specific facies, rock type or sub zone. Yet without a variogram, a mere 2-point statistics, there is no kriging, hence no traditional geostatistics. In presence of sparse data, the necessary variogram is often borrowed from deposits or outcrops different from the reservoir under study, or it is simply drawn to reflect the geologist's appreciation of correlation ranges. The problem is that many earth sciences structures (spatial variability) do not lend themselves well to a variogram characterization: very different patterns of variability may share the same variogram, see Fig. 3.3, Fig. 3.4 and Strebelle (2000); Caers (2005). As to borrow or draw a variogram to depict the spatial variability of a variable $z(\mathbf{u})$, $\mathbf{u} \in S$, why not borrow or draw a more relevant conceptual image of that variability? Geologists do not think in terms of variograms or covariance matrix, their expertise often takes the form of pictures, sketches, cartoons, as evident to anyone opening a structural geology book. One could consider an outcrop photograph or a geological sketch as a training image, which is a realization of a random function model. The corresponding decision of stationarity is that the actual data environment of the unsampled value $z(\mathbf{u})$ in S , has approximate replicates over that training image (Strebelle, 2002; Journel and Zhang, 2006).

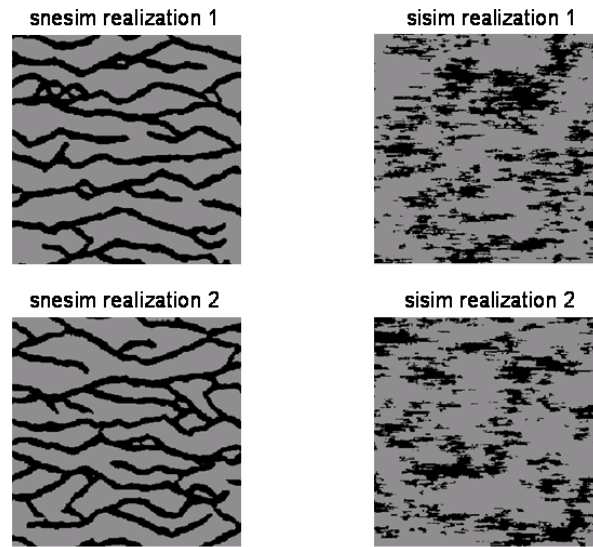


Figure 3.3: The two *SNESIM* generated images have more continuous and smoother features than the *SISIM* generated image, nevertheless they share the same variogram model

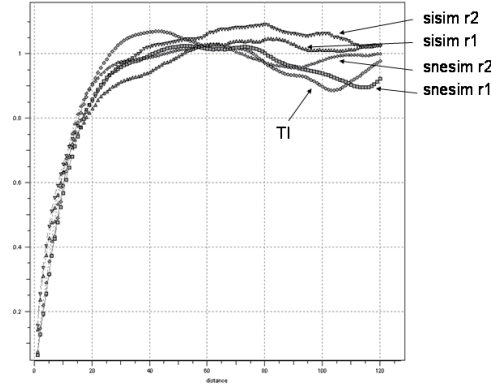


Figure 3.4: Reproduction of the E-W training image variogram by the two *SNESIM* and two *SISIM* generated images shown in Fig. 3.3.

The subjectivity of model decision

It does take some effort for accepting the fact that stationarity, which is theoretically defined as invariance by translation of spatial statistics (Goovaerts (1997, p.70); Chilès and Delfiner (1999, p.16)), is a modeling decision, not some hypothesis or property of the data which could be tested (Matheron, 1978; Journel, 1986). Such decision is necessarily subjective and can only be judged after the fact by whether or not the resulting model has helped achieving the task at hand. Subjectivity is at the source of any prediction process, it defines the repetition process that provides replicates; without such repetition there is no inference possible. Unless a prior model is given fully determined, stationarity is a necessary decision that allows building the random function model and inference of its characteristic moments.

- Accepting a fully defined model, such as the independence model, or a Gaussian model with a given covariance, or accepting a specific training image, amount to different decisions of stationarity. In the case of a training image, its selection represents a decision of stationarity, which allows its scanning to find replicates of any specific data event and retrieval of the corresponding replicates of the central value informed by that data event. The histogram of these central value replicates is then taken as the conditional distribution of the actual unknown. This is the inference process underlying the *SNESIM* mp simulation algorithm, see Section 3.7 and Guardiano and

Srivastava (1993); Strebel (2000).

- Accepting a Gaussian model allows building a kriging system to retrieve by kriging the two conditional moments (mean and variance) which suffice to specify the Gaussian conditional distribution, see Section 3.6 and Goovaerts (1997, p.266); Anderson (2003); Chilès and Delfiner (1999, p.381).

A priori, one decision is no better than another. There is no universality attached to the Gaussian model or any theoretical aura attached to the task of solving a system of equations (kriging) as opposed to the more trivial task of scanning a training image for replicates. Conversely, the patterns described by the T_i retained may not be the ones relevant to the actual deposit S . Adopting a wrong T_i may lead to severe errors, all the more dangerous that one is comforted by the final numerical representation of S fitting both the data and one's prior (and possibly erroneous) vision of the structures of variability. Hence, we insist on the absolute necessity of considering alternative different structural models, whether these are explicit training images or implicit models anchored on variogram models; these different structural models should reflect the range of possible different (geological) scenarios for the spatial variability ref.

Which model, 2-point or mp?

The better model is that which delivers the 'deemed' better result: did mimicking the training image patterns yield a 'more' satisfactory result than the less-pattern result yielded by variogram-based geostatistics? Again, the final judgment is necessarily case and application-dependent and is in part subjective.

One could leave aside some critical data, of a global nature as opposed to the local data used for conditioning, and check which of the final numerical representation of S fits best these 'check' data. Examples of such test data could be some of the production data in a producing hydrocarbon reservoir or mining deposit, they could be a subset of 'ground truth' data in an environmental application (Caers and Hoffman (2006); Goovaerts (1997, p.105); Journel and Zhang (2006)).

There is place in a geostatistical tool box for both sets of algorithms, 2-point and mp statistics are complementary: SGeMS proves it.

3.5 The Variogram, a 2-point Statistics

The main tool for traditional geostatistics, and for that matter, most statistical prediction algorithm, is the covariance or its equivalent the variogram. Consider a stationary random function $Z(\mathbf{u})$, and any two of its random variables $Z(\mathbf{u})$ and $Z(\mathbf{u} + \mathbf{h})$ separated by vector \mathbf{h} . The relation between these two RVs is characterized by any one of the following 2-point statistics, functions of the separation vector \mathbf{h} (Anderson (2003); Goovaerts (1997, p.28); Journel and Huijbregts (1978); Matheron (1970)):

- the covariance:

$$C(\mathbf{h}) = E \{ [Z(\mathbf{u}) - m] [Z(\mathbf{u} + \mathbf{h}) - m] \} \quad (3.14)$$

- the correlogram, or coefficient of correlation:

$$\rho(\mathbf{h}) = C(\mathbf{h})/C(0) \in [-1, +1]$$

- the variogram:

$$\begin{aligned} 2\gamma(\mathbf{h}) &= E \{ [Z(\mathbf{u} + \mathbf{h}) - Z(\mathbf{u})]^2 \} \\ &= 2 [C(\mathbf{h}) - C(0)], \text{ if } C(\mathbf{h}) \text{ exists,} \end{aligned}$$

where $m = E \{ Z(\mathbf{u}) \}$, $C(0) = \sigma^2 = \text{Var} \{ Z(\mathbf{u}) \}$ are the stationary marginal, or 1-point, statistics.

Any of these 2-point moments can be inferred by the corresponding experimental statistics, say from $n(\mathbf{h})$ pairs of data $z(\mathbf{u}_\alpha + \mathbf{h}), z(\mathbf{u}_\alpha)$, $\alpha = 1, \dots, n(\mathbf{h})$ approximately distant of \mathbf{h} and for the variogram (Goovaerts (1997, p.28); Wackernagel (1995); Isaaks and Srivastava (1989)):

$$2\gamma^*(\mathbf{h}) = \frac{1}{n(\mathbf{h})} \sum_{\alpha=1}^{n(\mathbf{h})} [z(\mathbf{u}_\alpha + \mathbf{h}) - z(\mathbf{u}_\alpha)]^2. \quad (3.15)$$

Modeling

In practice, available information only provides enough data pairs for a few distance $|\mathbf{h}|$ and along a few directions. However, that statistics, say $\gamma(\mathbf{h})$, is needed for all vectors $\mathbf{h} = \mathbf{u} - \mathbf{u}_\alpha$ linking an unsampled location \mathbf{u} to any nearby datum

location \mathbf{u}_α ; thus there is a need to interpolate/extrapolate experimental statistics such as $\gamma^*(\mathbf{h})$ into a model $\gamma(\mathbf{h})$ available for all \mathbf{h} . Because covariance and variogram are used to calculate variances and these variances are non-negative, not all analytical function $g(\mathbf{h})$ can be used as a covariance or a variogram model (Chilès and Delfiner (1999, p.59); Goovaerts (1997, p.87); Journel and Huijbregts (1978); Christakos (1984)). Positive linear combinations of basic acceptable (licit) models $g(\mathbf{h})$ are also acceptable, this allows defining a large family of acceptable covariance/variogram models sufficient for most practical studies. SGeMS allows consideration of any positive linear combination of the three most commonly used basic variogram models, the spherical, exponential and Gaussian models introduced in Chapter 5 (Deutsch and Journel, 1998, p.25).

The reader is referred to Goovaerts (1997, p.87) and other relevant papers and publications (Matheron, 1962, 1963; David, 1977; Journel and Huijbregts, 1978; Journel and Froidevaux, 1982; Chauvet, 1982; Cressie, 1993; Yao and Journel, 1998) for the practice of modeling experimental variograms, a sometimes delicate task in presence of directional anisotropy, sparse data and prior non-quantitative information.

Cross-variogram

In the previous expressions 3.14 and 3.15, the two RVs may relate to two difference attributes, say $Z_1(\mathbf{u})$ is porosity at location \mathbf{u} and $Z_2(\mathbf{u} + \mathbf{h})$ is seismic impedance measured at location $\mathbf{u} + \mathbf{h}$. The corresponding 2-point statistics is then, e.g. a cross-variogram defined as (Goovaerts (1997, p.46), Chilès and Delfiner (1999, p.328), Journel and Huijbregts (1978); Wackernagel (1995)):

$$2\gamma_{12}(\mathbf{h}) = E \{ [Z_1(\mathbf{u} + \mathbf{h}) - Z_1(\mathbf{u})] [Z_2(\mathbf{u} + \mathbf{h}) - Z_2(\mathbf{u})] \}. \quad (3.16)$$

In presence of only two different attributes Z_1, Z_2 , one must model a matrix of four (cross) covariance functions, $C_{11}(\mathbf{h}), C_{12}(\mathbf{h}), C_{21}(\mathbf{h}), C_{22}(\mathbf{h})$, or only three (cross) variogram functions $\gamma_{11}(\mathbf{h}), \gamma_{12}(\mathbf{h}) = \gamma_{21}(\mathbf{h}), \gamma_{22}(\mathbf{h})$, under restrictive conditions of positive definiteness, see Goovaerts (1997, p.108). In presence of N different attributes Z_1, \dots, Z_N , there would be N^2 (cross) covariance functions or $N(N - 1)/2$ (cross) variogram functions to model! Subsurface data are rarely enough to allow inference of statistics involving more than $N = 2$ attributes simultaneously, even if these statistics are only 2-point statistics involving only 2 space locations at a time, \mathbf{u} and $\mathbf{u} + \mathbf{h}$.

Multiple-point statistics

To characterize the relation between two patterns of data, say, n_1 data on attribute Z_1 : $\{z_1(\mathbf{u} + \mathbf{h}_\alpha); \alpha = 1, \dots, n_1\}$ and n_2 data on attribute Z_2 : $\{z_2(\mathbf{u}' + \mathbf{h}'_\beta); \beta = 1, \dots, n_2\}$, one would need much more than cross-covariances or cross-variograms. One needs in all rigor the **joint** distribution of the $(n_1 + n_2)$ RVs $Z_1(\mathbf{u} + \mathbf{h}_\alpha), Z_2(\mathbf{u}' + \mathbf{h}'_\beta); \alpha = 1, \dots, n_1; \beta = 1, \dots, n_2$ (Goovaerts (1997, p.72)). No experimental data would be ever enough to infer such multiple-variable, multiple-point statistics; not to mention the nightmare of their modeling. There are only two escape avenues:

1. assume a parameter-poor random function model $\{Z_1(\mathbf{u}), Z_2(\mathbf{u}')\}$ fully defined from a few low-order statistics that can be inferred from data. Most often such models are related to the multivariate Gaussian model fully characterized by the sole covariance matrix $[C_{ij}; i, j = 1, \dots, N]$, with $N = 2$ for the example above (Anderson (2003); Goovaerts (1997, p.265)).
2. build training images depicting the relation in space of the two variables $z_1(\mathbf{u})$ and $z_2(\mathbf{u}')$. These training images should reflect whatever physics or geology is known to control the joint spatial distributions of these two variables (Strebelle, 2000; Arpat, 2004; Zhang, 2006; Journel and Zhang, 2006).

In both cases, most of the structural (n-point statistics) information capitalized upon for estimation or simulation of an unsampled value (or set of values) is coming not from the data but from the model, multivariate Gaussian in case 1, the training image in case 2.

It would be a severe error, though both naive and common, to believe that one could do away with models falling in either of the two previous categories. The reason is that, whenever maps of estimated or simulated values are used, one necessarily draws from such maps much more than the 2-point or lower order statistics actually modeled from data; this ‘much more’ comes the n-point statistics of the RF model, whether Gaussian-related or training image-based (Journel, 1993b).

3.6 The Kriging Paradigm

Kriging has been historically at the source of acceptance of geostatistics (Krige, 1951; Matheron, 1970; Journel and Huijbregts, 1978); it remains a major data

integration tool and is used in most estimation and simulation algorithms. In its simplest indicator kriging form with a single (normal) equation, it identifies Bayes relation and the very definition of a conditional probability (Journel, 1983).

Kriging is in essence a generalized linear regression algorithm (Goldberger, 1962; Luenberger, 1969), extending the data-to-unknown correlation to data-to-data correlation through a non-diagonal kriging matrix. It is a regression with non-independent data: actually it can be shown that kriging consists, first of de-correlating the data by defining linear combinations of the original data which are orthogonal for a given covariance/variogram model, then of a traditional linear regression from these ‘independent’ data transforms (Journel, 1989).

The practice of geostatistics in very diverse earth sciences fields has led to a large number of variants from the theoretically rigorous yet simplest ‘simple kriging’.

3.6.1 Simple kriging

Because kriging is at the source of so many geostatistical algorithms it is worth to briefly recall here the basic simple kriging (SK) system, then its multiple variants (Goovaerts (1997, p.127); Deutsch and Journel (1998, p.77); Chilès and Delfiner (1999, p.154)).

Consider within a stationary field S the estimation of an unsampled value $z(\mathbf{u})$ from $n(\mathbf{u})$ neighboring data values $z(\mathbf{u}_\alpha)$, $\alpha = 1, \dots, n(\mathbf{u})$. If the estimate $z_{SK}^*(\mathbf{u})$ is restricted to be a linear combination of the data, it is written:

$$z_{SK}^*(\mathbf{u}) - m = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{SK}(\mathbf{u}) [z(\mathbf{u}_\alpha) - m] = \boldsymbol{\lambda}^t \cdot \mathbf{D}, \quad (3.17)$$

where $\boldsymbol{\lambda}$ is the column matrix of the $n(\mathbf{u})$ kriging weights $\lambda_\alpha^{SK}(\mathbf{u})$ and \mathbf{D} is the column matrix of the $n(\mathbf{u})$ residual data values $[z(\mathbf{u}_\alpha) - m]$ built from the stationary and assumed known mean value m . Consideration of the residual variables rather than the original Z -variables ensures unbiasedness defined as zero expected error:

$$E \{ Z_{SK}^*(\mathbf{u}) - Z(\mathbf{u}) \} = 0. \quad (3.18)$$

Remarks

- The zero expected error should be understood as a zero average error if the same geometric configuration of $n(\mathbf{u})$ data were applied elsewhere in the

stationary zone, therefore in effect averaging over all possible combinations of the $n(\mathbf{u})$ data values. Ideally, one should ensure unbiasedness conditional to both data configuration and data values (the $z(\mathbf{u}_\alpha)$'s), that is:

$$E \{Z_{SK}^*(\mathbf{u}) - Z(\mathbf{u}) | Z(\mathbf{u}_\alpha) = z(\mathbf{u}_\alpha), \alpha = 1, \dots, n(\mathbf{u})\} = 0. \quad (3.19)$$

Conditional unbiasedness for all possible combination of data values $z(\mathbf{u}_\alpha)$ entails unbiasedness as in relation Eq. 3.18, not the reverse (Journel and Huijbregts (1978); Goovaerts (1997, p.182); Deutsch and Journel (1998, p.94)). SK, as most other linear estimators such as inverse distance-based, ensures unbiasedness not conditional unbiasedness. This limitation of the unbiasedness property is at the source of many disappointments (David, 1977; Isaaks, 2004; Isaaks and Srivastava, 1989).

- A critical decision affecting the quality of the estimate (Eq. 3.17) is the choice of the $n(\mathbf{u})$ data retained to estimate any unsampled location \mathbf{u} . Consistency with the stationarity decision made to infer the covariance model would call for all locations $\mathbf{u} \in S$ to be estimated from the same data set n including all samples available over S . Such kriging with a 'global' neighborhood is rarely implemented in practice, precisely because of defiance towards the decision of stationarity (Deutsch and Journel (1998, p.32); Goovaerts (1997, p.178)). Not only one must decide the extent of the neighborhood within which the $n(\mathbf{u})$ data should be collected, but one may want to privilege certain directions, for example the direction of maximum continuity starting from \mathbf{u} (Deutsch and Journel (1998, p.33); Goovaerts (1997, p.178)).

Convexity issue

A definite advantage of kriging over traditional linear interpolators is that it is non convex: the kriging estimate need not be valued in the interval of the data values retained (Goovaerts (1997, p.177)). For example, the SK estimate $z_{SK}^*(\mathbf{u})$ may be valued greater than the largest datum value $\max \{z(\mathbf{u}_\alpha), \alpha = 1, \dots, n(\mathbf{u})\}$.

This advantage can turn into an inconvenience if the estimate $z_{SK}^*(\mathbf{u})$ is valued outside the z -physical bounds, for example a negative estimate for a positive variable such as a metal grade. One solution (not the best) to ensure convexity is to enforce the kriging weights to be all positive and sum up to 1; ordinary kriging weights do add up to 1 but are not necessarily all positive, see hereafter Section 3.6.2, and Barnes and Johnson (1984); Rao and Journel (1996).

Simple kriging system

If the estimation criterion is chosen to be ‘least squared error’, often a decision of mere convenience, the weights $\lambda_\alpha^{SK}(\mathbf{u})$ are given by a kriging system of linear equations built from a covariance model (Luenberger (1969); Matheron (1970); Journel and Huijbregts (1978); Goovaerts (1997, p.127)):

$$\mathbf{K} \cdot \boldsymbol{\lambda} = \mathbf{k} \quad (3.20)$$

where the transposed vector $\mathbf{k}^T = [C(\mathbf{u} - \mathbf{u}_\alpha), \alpha = 1, \dots, n(\mathbf{u})]$ is the data-to-unknown row covariance matrix, $\mathbf{K} = [C(\mathbf{u}_\alpha - \mathbf{u}_\beta), \alpha, \beta = 1, \dots, n(\mathbf{u})]$ is the data-to-data square covariance matrix; both matrices are built from the prior stationary covariance model:

$$C(\mathbf{h}) = \text{Cov}\{Z(\mathbf{u}), Z(\mathbf{u} + \mathbf{h})\} = C(0) - \gamma(\mathbf{h}), \quad (3.21)$$

$C(0) = \text{Var}\{Z(\mathbf{u})\}$ is the stationary variance, and

$2\gamma(\mathbf{h}) = \text{Var}\{Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})\}$ is the corresponding stationary variogram model.

The two main contributions of kriging to estimation are (Journel and Huijbregts, 1978):

1. the utilization of a variogram distance $\gamma(\mathbf{h})$ specific to the variable $Z(\mathbf{u})$ and the stationary zone S under study as opposed to a non specific Euclidean distance \mathbf{h} , as used in e.g. inverse distance interpolation. The variogram distance could be anisotropic, for example privileging data along the direction of greater continuity from the location \mathbf{u} being estimated;
2. the consideration of the data-to-data covariance matrix \mathbf{K} allows ‘data declustering’, which leads to give less weight to redundant data within a cluster of data as opposed to isolated data. This property of kriging allows correcting for bias due to preferential clustering of data, a common occurrence in earth sciences.

Kriging variance

A by-product of kriging and more generally of any least square regression is the estimation variance or kriging variance, which is the expected squared error whose minimization led to the kriging system (3.20) (Goovaerts (1997, p.179); Chilès and Delfiner (1999, p.156)):

$$\sigma_{SK}^2(\mathbf{u}) = \text{Var}\{Z(\mathbf{u}) - Z_{SK}^*(\mathbf{u})\} = C(0) - \boldsymbol{\lambda}^t \cdot \mathbf{k}. \quad (3.22)$$

This variance is often misused as a measure of accuracy of the estimator $Z_{SK}^*(\mathbf{u})$. The variance expression (3.22) is data values-independent; it depends only on the geometry of the data set $n(\mathbf{u})$ retained and the covariance model adopted; it is thus a covariance-dependent ranking index of data configuration, a valuable index for comparison but not yet a measure of estimation accuracy (Journel and Rossi, 1989). A better measure of the potential error associated with the estimator $Z_{SK}^*(\mathbf{u})$ would be the conditional error variance which is also dependent on the data values $z(\mathbf{u}_\alpha)$ (Goovaerts (1997, p.180)):

$$\sigma_{SK}^2(\mathbf{u}) = \text{Var} \{Z(\mathbf{u}) - Z_{SK}^*(\mathbf{u}) | Z(\mathbf{u}_\alpha) = z(\mathbf{u}_\alpha), \alpha = 1, \dots, n(\mathbf{u})\}.$$

It can be shown that the kriging variance (3.22) is the average of the conditional error variance over all possible joint realizations of the $n(\mathbf{u})$ data values, the data configuration being fixed (Goovaerts (1997, p.180,361)). In the general case, one cannot ignore the impact of the actual data values on estimation accuracy.

One notable exception is that provided by a Gaussian RF model where *all* constitutive RVs $Z(\mathbf{u})$, $\mathbf{u} \in S$, are assumed jointly Gaussian-distributed. In that case, the previous conditional error variance is indeed data values-independent and identifies the kriging variance (3.22), a property known as homoscedasticity (Goovaerts (1997, p.82,180); Anderson (2003)).

Distribution of error

Even if the kriging variance (3.22) could be retained as a representative error variance, one would know only two moments of the error distribution: the mean equal to zero per unbiasedness and the kriging variance. A two-parameter distribution type would have to be assumed, for example a Gaussian error distribution. There is, unfortunately, no justification for such distribution type assumption; Central Limit theorems do not apply to spatial interpolation errors essentially because data and resulting errors are not independent one from each other. If a Gaussian distribution assumption is accepted for convenience, one should be aware that this distribution has short tails and its adoption is very consequential, particularly when assessing the probability of large errors.

3.6.2 Ordinary kriging and other variants

The most restrictive aspect of any probabilistic approach is associated with the decision of stationarity (Section 3.6) which allows scanning a data set for replicates,

averaging the latter to infer the required statistics. For example, the inference of the covariance model $C(\mathbf{h})$ needed to build any kriging system calls for pooling together into a same experimental covariance pairs of data, $\{z(\mathbf{u}_\alpha), z(\mathbf{u}_\alpha + \mathbf{h})\}$, $\{z(\mathbf{u}_\beta), z(\mathbf{u}_\beta + \mathbf{h})\}$ approximately separated by the same vector \mathbf{h} but otherwise taken at different locations \mathbf{u}_α and \mathbf{u}_β . Once that covariance model is available, the harsh consequences of the stationarity decision settle in. Rigorous theory would demand that the covariance model $C(\mathbf{h})$ and the corresponding stationary mean m be frozen over the area S of the stationarity decision. Yet in many applications, local information calls for a locally variable mean and sometimes for aspects of the covariance model to be made locally variable, for example the direction of anisotropy. Variants of the previous simple kriging system were developed to allow such flexibility, all amounting to a deviation from rigorous theory.

In ordinary kriging (OK) the expected value of the random function is locally re-estimated from local data, while the covariance model is kept stationary. The OK concept has been extended to local estimation of the parameters of a functional trend (KT or kriging with a trend). A locally varying mean (LVM) can also be input directly and used in expression 3.17 in lieu of the stationary mean m .

All these variants of kriging amount to relax the decision of stationarity initially necessary to define the random function model and infer its constitutive statistics, for example a variogram model. There have been many attempts at extending the RF theory to justify such liberty with regard to the original restrictive decision of stationarity (Matheron (1971); Chilès and Delfiner (1999, p.243); Goovaerts (1997, p.143)). These attempts, besides being awkward, have resulted in other restrictions; the practice of geostatistics has essentially ignored them. Our point is that the random function theory is a non-exclusive provider of tools for assessing space/time distributions, as opposed to a binding exclusive model. The RF-originated tools can be modified to a reasonable extent with proper documentation; there is no need to build a new RF model in a vain attempt at gentrifying tools that have a proven record of practice. Possibly, the best argument for such lax consideration of RF theory is realizing that there would be no practical geostatistics without OK or without kriging with a local varying mean and, more generally without the multiple code implementations that make it ‘work’.

Ordinary kriging

The simple kriging (SK) expression (Eq. 3.17) appears as an estimate of an unknown deviation from a known stationary mean m . If that mean is considered locally variable, it can be estimated from the same $n(\mathbf{u})$ local data used in ex-

pression (Eq. 3.17); the corresponding estimate then takes its ordinary kriging (OK) expression (Goldberger (1962); Matheron (1970); Goovaerts (1997, p.132); Journel and Huijbregts (1978)):

$$z_{OK}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{OK}(\mathbf{u}) z(\mathbf{u}_{\alpha}), \quad (3.23)$$

where the kriging weights sum to 1: $\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{OK}(\mathbf{u}) = 1$.

The corresponding kriging system is similar to the SK system (Eq. 3.20) with, in addition, one Lagrange parameter and equation to account for the above constraint on the OK weights. The qualifier 'ordinary' is appropriate since OK is used more often than SK thanks to its robustness against local departures from the original decision of stationarity.

Kriging with a trend

A locally varying unknown mean $m(\mathbf{u})$ can be modeled by a function of the coordinates (\mathbf{u}) . That function, called a trend function, if of known shape or type but with unknown locally variable parameters; therefore the mean $m(\mathbf{u})$ remains unknown at any location \mathbf{u} (ref.). In the space domain where $\mathbf{u} = (x, y, z)$ coordinates, the trend function is usually a polynomial function of the coordinates; for example a trend linear in the horizontal space (x, y) but possibly quadratic in the vertical (z) would be written (Goovaerts (1997, p.141); Deutsch and Journel (1998, p.67)):

$$E \{Z(\mathbf{u})\} = m(\mathbf{u}) = a_0(\mathbf{u}) + a_1(\mathbf{u})x + a_2(\mathbf{u})y + a_3(\mathbf{u})z + a_4(\mathbf{u})z^2. \quad (3.24)$$

The 5 parameters $a_i(\mathbf{u})$ are unknown and are estimated from the $n(\mathbf{u})$ local data available. The location coordinates $\mathbf{u} = (x, y, z)$ are known.

Similarly, in the time domain, a time series $Z(t)$ may present a periodic trend modeled as a cosine function of known frequency ω but unknown varying phase and amplitude $a_0(t)$, $a_1(t)$:

$$E \{Z(t)\} = m(t) = a_0(t) + a_1(t) \cdot \cos(2\pi\omega t). \quad (3.25)$$

However such cosine trend function is not yet programmed in SGeMS.

Once the unknown parameters $a_i(\mathbf{u})$ are estimated (implicitly and by a form of kriging), a simple kriging of type (3.17) is applied by replacing at each location

\mathbf{u} the constant stationary mean m by the resulting mean estimate $m^*(\mathbf{u})$. The corresponding estimate is said to be a kriging with a trend model (KT) and is written:

$$z_{KT}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{KT}(\mathbf{u}) z(\mathbf{u}_{\alpha}). \quad (3.26)$$

The KT expression (Eq. 3.26) differs from the OK expression through its weights. These KT weights $\lambda_{\alpha}^{KT}(\mathbf{u})$ are given by a kriging system similar to the SK system (Eq. 3.20) but with additional constraints on the KT weights (Goldberger (1962); Goovaerts (1997, p.139); Journel and Huijbregts (1978)).

Actually OK is but a particular case of KT when the trend model (Eq. 3.24) reduces to the sole term $a_0(\mathbf{u})$. In cases of interpolation with the $n(\mathbf{u})$ data surrounding the location \mathbf{u} on both sides of the trend, OK would give results very close to KT. The specification of the trend functional type, say linear or quadratic, matters only in cases of extrapolation (Journel and Rossi (1989); Goovaerts (1997, p.147)).

Kriging with a local varying mean

There are applications where some ancillary information (different from the z -data) provides at *all* locations the locally varying mean (LVM), then denoted $m^*(\mathbf{u})$. The SK expression (Eq. 3.17) is then applied directly to the deviations from these locally varying mean values (Goovaerts (1997, p.190)):

$$z_{LVM}^*(\mathbf{u}) - m^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{SK}(\mathbf{u}) [z(\mathbf{u}_{\alpha}) - m^*(\mathbf{u}_{\alpha})]. \quad (3.27)$$

Non-linear kriging

The qualifier ‘non-linear’ applied to kriging is misleading if it promises some major breakaway from the fundamental limitation of considering only linear combinations of data. Most so-called non-linear kriging, including kriging of normal score transform as used in program *SGSIM* (Deutsch and Journel, 1998, p.75), disjunctive kriging (Matheron, 1973) or indicator kriging (Journel, 1983), are in fact all linear kriging applied to a non-linear transform of the variables involved. For example, lognormal kriging is but a kriging applied on the logarithms of the data (Rendu, 1979; Journel, 1980).

A non-linear transform of the original variable(s) is warranted if that transform allows defining new variables that:

- are more attuned to the problem addressed, as is the case for indicator transform (see hereafter Section 3.6.5);
- satisfy better the requirements of the algorithm being considered, e.g. the normal score transform allows meeting one requirement of the sequential Gaussian simulation algorithm (SGeMS code *SGSIM*, see Section 8.1.2, and Goovaerts (1997, p.380));
- deliver stronger correlation in space.

Remarks :

Any non-linear back transform of kriging results may lead to severe biases if not carefully attended to. Non robust unbiasedness corrections associated to back transform of kriging estimates may wipe out any benefit brought by working on the transformed variable. This is the reason for lognormal kriging having fallen into disuse: what is gained by working on the logarithm may be lost through a back transform that involves an exponentiation (Journel (1980); Chilès and Delfiner (1999, p.191); Deutsch and Journel (1998, p.76)). Note that indicator kriging estimates are used as probability estimates, without any back transform, see hereafter. Similarly, the kriging means and variances of normal score values are used directly to build the conditional distributions in sequential Gaussian simulation, these kriging results are never back transformed; it is the final simulated values that are back transformed, then indeed that back transform is sensitive to the tail extrapolation decision (Deutsch and Journel (1998, p.135); Goovaerts (1997, p.385)).

A non-linear transform of the variable does not remove the most fundamental limitation of all kriging which is that the data are related to the unknown one at a time; see the right hand side covariance matrix k in the SK system (Eq. 3.20). Kriging remains, however, the ultimate estimation paradigm if applied to functions of the data taken two by two, three by three, and ultimately taken altogether as a single multiple-point event, see Sections 3.6.5 and 3.7.

3.6.3 Kriging with linear average variable

An application of kriging with linear average variable is to estimate the average grade of a mining block from neighboring data which are both core grades and average grades of already mined-out blocks (Journel and Huijbregts, 1978; David, 1977). Another application is that related to tomographic imaging where data

are defined as linear average over diverse ray paths (1D volume data) (Gómez-Hernández et al., 2004; Hansen et al., 2006). These kriging systems call for the covariance between any two block-support z -values, which happens to be a linear average of the point covariance model $C(\mathbf{h})$, see Journel and Huijbregts (1978). Any kriging system can be used with linear average data as long as the variogram/covariance values are properly regularized (averaged).

Covariance averaging

The covariance relating a point support value $Z(\mathbf{u})$ to a linear average $B_V(\mathbf{s})$ defined over a block of support volume V centered at \mathbf{s} is derived from the point-to-point covariance model $C(\mathbf{u}, \mathbf{u} + \mathbf{h}) = C(\mathbf{h})$ as:

$$\overline{C}(\mathbf{u}, V(\mathbf{s})) = \text{Cov} \{B_V(\mathbf{s}), Z(\mathbf{u})\} = \frac{1}{|V|} \int_{\mathbf{u}' \in V(\mathbf{s})} C(\mathbf{u} - \mathbf{u}') d\mathbf{u}'$$

Similarly, an average block-to-block variogram is given as:

$$\overline{C}(V, V') = \int_{\mathbf{u} \in V} \int_{\mathbf{u}' \in V'} C(\mathbf{u} - \mathbf{u}') d\mathbf{u} d\mathbf{u}'$$

These upscaled or regularized covariances provide a valid model of point-to-block and block-to-block covariances, they are used whenever linear average data are present. Fast calculation of each block-averaged variogram/covariance is discussed in Section 7.4 and Kyriakidis et al. (2005).

Kriging with block and point data

Data taken at different scales, both on block-support and on point-support, can be considered simultaneously in the kriging system. The only condition is that all block data are linear averages of point values. For simplicity, the kriging theory is illustrated here with simple kriging.

The block data $B(\mathbf{v}_\alpha)$ is defined as the spatial linear average of point values $Z(\mathbf{u}')$ within the block volume \mathbf{v}_α (Journel and Huijbregts (1978); Hansen et al. (2006); Liu, Journel and Mukerji (2006); Goovaerts (1997, p.152)):

$$B(\mathbf{v}_\alpha) = \frac{1}{|\mathbf{v}_\alpha|} \int_{\mathbf{v}_\alpha} L_\alpha(Z(\mathbf{u}')) d\mathbf{u}' \quad \forall \alpha \quad (3.28)$$

where L_α is a known linear averaging function.

The simple kriging estimator $Z_{SK}^*(\mathbf{u})$ conditioned to both point and block data is then written:

$$Z_{SK}^*(\mathbf{u}) - m = \Lambda^t \cdot \mathbf{D} = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) \cdot [D(\mathbf{u}_{\alpha}) - m] \quad (3.29)$$

where: $\Lambda^t = [\lambda_P \ \lambda_B]$ denotes the kriging weights for point data P and block data B . $\mathbf{D}^t = [\mathbf{P} \ \mathbf{B}]$ denotes the data value vector. $D(\mathbf{u}_{\alpha})$ is a specific datum at location \mathbf{u}_{α} and $n(\mathbf{u})$ denotes the number of data. m denotes the stationary mean.

The kriging weights Λ are obtained through the kriging system:

$$\mathbf{K} \cdot \Lambda = \mathbf{k} \quad (3.30)$$

with

$$\mathbf{K} = \begin{bmatrix} \mathbf{C}_{PP'} & \bar{\mathbf{C}}_{PB} \\ \bar{\mathbf{C}}_{PB}^t & \bar{\mathbf{C}}_{BB'} \end{bmatrix} \text{ and } \mathbf{k} = \begin{bmatrix} \mathbf{C}_{PP_0} \\ \bar{\mathbf{C}}_{BP_0} \end{bmatrix}$$

where \mathbf{K} denotes the data-to-data covariance matrix, \mathbf{k} denotes the data-to-unknown covariance matrix, \mathbf{C} denotes the point covariance submatrix, $\bar{\mathbf{C}}$ denotes a covariance submatrix involving a block support and P_0 is the simulation location.

The kriging variance is then written:

$$\sigma_{SK}^2(\mathbf{u}) = \text{Var} \{Z(\mathbf{u}) - Z_{SK}^*(\mathbf{u})\} = C(0) - \Lambda^t \cdot \mathbf{k}$$

with: $C(0) = \text{Var} \{Z(\mathbf{u})\}$, being the stationary variance.

3.6.4 Cokriging

There is nothing in the theory of kriging and expressions (3.17) to (3.27) that constrains the unknown $Z(\mathbf{u})$ and the data $Z(\mathbf{u}_{\alpha})$ to relate to the same attribute. One can extend the notation Z to different attribute values $Z_k(\mathbf{u})$, $Z_{k'}(\mathbf{u}_{\alpha})$, say for estimation of a porosity value $Z_k(\mathbf{u})$ at location \mathbf{u} from porosity data $Z_k(\mathbf{u}_{\alpha})$ and seismic amplitude data $Z_{k'}(\mathbf{u}_{\alpha})$ with $k' \neq k$ at neighboring locations \mathbf{u}_{α} . Cokriging is the extension of the kriging paradigm to estimation of one attribute using a data set which contains data related to other attributes (Myers (1982); Wackernagel (1995); Goovaerts (1997, p.203); Chilès and Delfiner (1999, p.296)).

For example, the simple cokriging estimate of an unsampled porosity $z_1(\mathbf{u})$ from $n_1(\mathbf{u})$ neighboring porosity data $z_1(\mathbf{u}_{\alpha})$ and $n_2(\mathbf{u})$ seismic data $z_2(\mathbf{u}'_{\beta})$ would be written:

$$z_1^*(\mathbf{u}) - m_1 = \sum_{\alpha=1}^{n_1(\mathbf{u})} \lambda_{\alpha} [z_1(\mathbf{u}_{\alpha}) - m_1] + \sum_{\beta=1}^{n_2(\mathbf{u})} \lambda_{\beta} [z_2(\mathbf{u}'_{\beta}) - m_2], \quad (3.31)$$

where m_1 and m_2 are the two stationary means.

The only difficulty, but a serious one in practice, comes from the necessity to infer and model *jointly* many cross-covariance/variogram models, up to K^2 models in the case of cross-covariances. If it is already difficult to infer the variogram of a single variable in a 3-dimensional possibly anisotropic space, real data are rarely enough to infer a set of cross-variograms for more than $K = 3$ different attribute variables.

In order to alleviate the burden of modeling all these variograms with the linear model of coregionalization (Goovaerts (1997, p.108); Chilès and Delfiner (1999, p.339)), various shortcut models have been proposed: two of which are available in SGeMS: the Markov Model 1 and the Markov Model 2 (Almeida and Journel (1994); Journel (1999); Chilès and Delfiner (1999, p.305); Goovaerts (1997, p.237)). For text clarity, we will only consider the case of a single secondary variable ($K = 2$).

Markov Model 1 The Markov Model 1 (MM1) considers the following Markov-type screening hypothesis:

$$E\{Z_2(\mathbf{u})|Z_1(\mathbf{u}); Z_1(\mathbf{u} + \mathbf{h})\} = E\{Z_2(\mathbf{u})|Z_1(\mathbf{u})\}$$

i.e. the dependence of the secondary variable on the primary is limited to the co-located primary variable. The cross-covariance is then proportional to the auto-covariance of the primary variable:

$$C_{12}(\mathbf{h}) = \frac{C_{12}(0)}{C_{11}(0)} C_{11}(\mathbf{h}) \quad (3.32)$$

Where C_{12} is the cross-covariance between the two variables Z_1 and Z_2 and C_{11} is the covariance of the primary variable Z_1 . Solving the cokriging system under the MM1 model only calls for knowledge of C_{11} , hence the inference and modeling effort is the same as for kriging with only primary Z_1 -data. Although very congenial the MM1 model should not be used when the support of the secondary variable Z_2 is larger than the one of Z_1 , lest the variance of Z_1 would be underestimated. It is better to use the Markov Model 2 in that case.

Markov Model 2 The Markov Model 2 (MM2) was developed for the case where the volume support of the secondary variable is larger than that of the primary variable (Journel, 1999). This is often the case with remote sensing and seismic-related data. The more relevant Markov-type hypothesis is

now:

$$E\{Z_1(\mathbf{u})|Z_2(\mathbf{u}); Z_2(\mathbf{u} + \mathbf{h})\} = E\{Z_1(\mathbf{u})|Z_2(\mathbf{u})\}$$

i.e. the dependence of the primary variable on the secondary is limited to the co-located secondary variable. The cross-variogram is now proportional to the covariance of the secondary variable:

$$C_{12}(\mathbf{h}) = \frac{C_{12}(0)}{C_{11}(0)} C_{22}(\mathbf{h}) \quad (3.33)$$

In order for all three covariances C_{11} , C_{12} and C_{22} to be consistent, Journel (1999) proposed to model C_{11} as a linear combination of C_{22} and any permissible residual correlation ρ_R . Expressed in term of correlograms

$$\rho_{11}(\mathbf{h}) = \frac{C_{11}(\mathbf{h})}{C_{11}(0)}, \quad \rho_{22}(\mathbf{h}) = \frac{C_{22}(\mathbf{h})}{C_{22}(0)}$$

this is written:

$$\rho_{11}(\mathbf{h}) = \rho_{12}^2 \cdot \rho_{22}(\mathbf{h}) + (1 - \rho_{12}^2) \rho_R(\mathbf{h}) \quad (3.34)$$

where ρ_{12} is the co-located coefficient of correlation between $Z_1(\mathbf{u})$ and $Z_2(\mathbf{u})$. Independently of the cross-modeling adopted, cokriging shares all the contributions and limitations of kriging: it provides a linear, least squared error, regression combining data of diverse types accounting for their redundancy and respective variogram distances to the unknown. Cokriging considers the data one at a time and the cokriging variance being data values-independent is an incomplete measure of estimation accuracy. The linear limitation of kriging may be here more serious, since cokriging would ignore any non-linear relation between two different attributes which could be otherwise capitalized upon for cross-estimation. One possible solution is to apply cokriging on non-linear transforms of the original variables, a transform chosen to reveal such non-linear dependence.

The kriging system with block data as described in Section 3.6.3 can also be seen as a cokriging system, where the cross-dependence between points and block is given by the regularization process.

3.6.5 Indicator kriging

Indicator kriging is yet another kriging but applied to variables that are binary indicators of occurrence of an event, say:

$$I_k(\mathbf{u}) = \begin{cases} 1 & \text{if the event } k \text{ occurs at location } \mathbf{u} \\ 0 & \text{if not} \end{cases} \quad (3.35)$$

or for the continuous case :

$$I(\mathbf{u}; z_k) = \begin{cases} 1 & \text{if } Z(\mathbf{u}) \leq z_k \\ 0 & \text{if not} \end{cases}$$

The event k to be estimated could be presence of facies of type k at location \mathbf{u} , or could be that the unsampled continuous variable $z(\mathbf{u})$ is valued below the threshold z_k .

The particularity of indicator kriging (IK) is that it delivers a kriging estimate that can be interpreted directly (without any transform) as an estimate of the probability for the unsampled event to occur at location \mathbf{u} conditional to the data set $n(\mathbf{u})$ observed (Goovaerts (1997, p.293); Chilès and Delfiner (1999, p.383); Journel (1983)). The IK estimate is hereafter written under its simple kriging form:

$$\begin{aligned} I_{SK}^*(\mathbf{u}) &= \text{Prob}^* \{I(\mathbf{u}) = 1 | n(\mathbf{u})\} \\ &= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) I_k(\mathbf{u}_{\alpha}) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) \right] \cdot p_0, \end{aligned} \quad (3.36)$$

where $p_0 = E \{I(\mathbf{u})\} = \text{Prob} \{I(\mathbf{u}) = 1\}$ is the prior probability of the event occurring, $\lambda_{\alpha}(\mathbf{u})$ is the kriging weight associated to the indicator datum $I(\mathbf{u}_{\alpha})$ valued 0 or 1.

If, independently of the $n(\mathbf{u})$ indicator data, soft information is available providing a location-specific prior probability $p(\mathbf{u})$, that probability could replace p_0 in expression (3.36). Indicator kriging can be seen as an updating of that prior probability $p(\mathbf{u})$ by the indicator data $I(\mathbf{u}_{\alpha})$ ((Goovaerts, 1997, p.293)).

The fact that kriging is not a convex estimator valued between the minimum and maximum indicator data value (here 0 and 1) is particularly damaging in that IK may provide estimated probabilities outside the interval $[0, 1]$. Corrections are made to this purpose. An alternative is to consider mp statistics through products instead of linear combinations of the indicator data, see the following presentation on extended normal equations and in Section 3.10 the nu/tau model (Journel, 2002).

On the positive side, kriging being an exact estimator, if estimation is performed at any hard datum location \mathbf{u}_{α} , the resulting probability estimate is ‘hard’, that is valued 0 or 1 identifying that hard datum value; if the indicator variogram used is continuous with a small nugget effect the probability estimate would smoothly departs from that hard value (0 or 1) as the location \mathbf{u} to be estimated gets away from \mathbf{u}_{α} .

3.7 An Introduction to mp Statistics

Consider again the linear indicator kriging expression (3.36). First note that any non-linear transform of a binary (indicator) variable is non-effective in that it results in just another binary variable. To extract more from the indicator data set $\{I(\mathbf{u}_\alpha), \alpha = 1, \dots, n(\mathbf{u})\}$, one need to consider these data two by two, three by three, \dots , at the limit altogether as a single data event.

Consider then the extended IK expression (3.37) which is a linear combination:

- of the indicator data taken one at a time as in expression (3.36), there are $n(\mathbf{u})$ such indicators;
- of the indicator data taken two at a time; there are a number $(n(\mathbf{u}), 2)$ of combinations of such pairs;
- of the indicator data taken three at a time; there are $(n(\mathbf{u}), 3)$ such triplets;
- \dots
- of the indicator data taken altogether; there is only one such product:

$$\begin{aligned}
 I_{SK}^*(\mathbf{u}) = & \text{Prob}^* \{I(\mathbf{u}) = 1 | n(\mathbf{u})\} = \\
 & p_0 \quad (\text{prior probability for } I(\mathbf{u}) = 1) \\
 & + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{(1)}(\mathbf{u}) [I(\mathbf{u}_\alpha) - p_0] \quad (\text{one at a time}) \\
 & + \sum_{\alpha=1}^{(n(\mathbf{u}), 2)} \lambda_\alpha^{(2)}(\mathbf{u}) [I(\mathbf{u}_{\alpha_1})I(\mathbf{u}_{\alpha_2}) - E\{I(\mathbf{u}_{\alpha_1})I(\mathbf{u}_{\alpha_2})\}] \quad (\text{two at a time}) \\
 & + \sum_{\alpha=1}^{(n(\mathbf{u}), 3)} \lambda_\alpha^{(3)}(\mathbf{u}) [I(\mathbf{u}_{\alpha_1})I(\mathbf{u}_{\alpha_2})I(\mathbf{u}_{\alpha_3}) - E\{I(\mathbf{u}_{\alpha_1})I(\mathbf{u}_{\alpha_2})I(\mathbf{u}_{\alpha_3})\}] \\
 & \hspace{15em} (\text{three at a time}) \\
 & + \dots \\
 & + \lambda_\alpha^{(n(\mathbf{u}))}(\mathbf{u}) \left[\prod_{\alpha} I(\mathbf{u}_\alpha) - E \left\{ \prod_{\alpha} I(\mathbf{u}_\alpha) \right\} \right] \quad (\text{taken altogether})
 \end{aligned} \tag{3.37}$$

Remarks

- Expression 3.37 is a simple indicator (co)kriging estimator extended to include data taken 2, 3, up to all a time. The corresponding simple kriging system is called ‘the extended system of normal equations’ (Journel and Alabert, 1989); it has $2^{n(\mathbf{u})}$ equations yielding the $2^{n(\mathbf{u})}$ kriging weights $\lambda_\alpha^{(\cdot)}$.
- Note that a set of $n(\mathbf{u})$ binary indicator data can take $2^{n(\mathbf{u})}$ possible joint outcomes, a number precisely equal to the number of kriging weights in the extended IK expression 3.37; indeed:

$$\sum_{\alpha=1}^{n(\mathbf{u})} (n(\mathbf{u}), \alpha) = 2^{n(\mathbf{u})}.$$

It can be shown that the solution of the full extended normal system (Eq. 3.37) provides the exact conditional probability value for $I(\mathbf{u}) = 1$ for all possible data value combinations; there are $2^{n(\mathbf{u})}$ such combinations.

- The information carried by the product $I(\mathbf{u}_\alpha)I(\mathbf{u}_\beta)$ is not redundant with that carried by the two individual data $I(\mathbf{u}_\alpha)$ and $I(\mathbf{u}_\beta)$ or any of their linear combinations. By restricting the expression 3.37 to indicator data taken one at a time as in expression 3.36, one is losing precious information, that provided by observation of the data jointly at many locations. The information $I(\mathbf{u}_\alpha)I(\mathbf{u}_\beta)$ sits there ready for the taking as a covariate through a form of cokriging. As in any other cokriging, it suffices to evaluate the additional covariances required which are, in the case of data taken only up to two at a time:
 - the 3-point covariance linking any 2-point data $I(\mathbf{u}_\alpha)I(\mathbf{u}_\beta)$ to the unknown $I(\mathbf{u})$;
 - the 4-point covariance linking any two doublets of data, say $I(\mathbf{u}_1)I(\mathbf{u}_2)$ with $I(\mathbf{u}_3)I(\mathbf{u}_4)$ and measuring the redundancy between these two doublets.

The traditional IK estimator (Eq. 3.36) using the data one at a time calls for traditional 2-point covariances. An extended IK estimator using data taken two at a time would require in addition 3 and 4-point covariances. An extended IK estimator using all possible combinations of data up to only two at a time would call for a kriging system of dimension $n(\mathbf{u}) + (n(\mathbf{u}), 2)$; e.g. if $n(\mathbf{u}) = 10$, then

$n(\mathbf{u}) + (n(\mathbf{u}), 2) = 10 + 45 = 55$, a considerable increase in dimension and covariance modeling effort! Clearly this is not practical, particularly if data are to be taken many more than two at a time.

The solution is to consider all $n(\mathbf{u})$ data grouped together into a single multiple-point data event which corresponds to the last term of expression (Eq. 3.37). The corresponding IK estimate is then written:

$$I_{SK}^*(\mathbf{u}) = \lambda \cdot \prod_{\alpha=1}^{n(\mathbf{u})} I(\mathbf{u}_\alpha).$$

The corresponding kriging system reduces to one single equation, also called single normal equation, delivering the single kriging weight λ . It can be shown that this single normal equation identifies the *exact* expression of the conditional probability:

$$\begin{aligned} I_{SK}^*(\mathbf{u}) &\equiv \text{Prob} \{I(\mathbf{u}) = 1 | n(\mathbf{u})\} \\ &= \frac{\text{Prob} \{I(\mathbf{u}) = 1, n(\mathbf{u})\}}{\text{Prob} \{n(\mathbf{u})\}} \\ &= \frac{\text{Prob} \{I(\mathbf{u}) = 1, I(\mathbf{u}_\alpha) = i_\alpha, \alpha = 1, \dots, n(\mathbf{u})\}}{\text{Prob} \{I(\mathbf{u}_\alpha) = i_\alpha, \alpha = 1, \dots, n(\mathbf{u})\}} \end{aligned} \quad (3.38)$$

Note that the probability in the numerator of the exact expression (Eq. 3.38) is actually a $(n(\mathbf{u}) + 1)$ -point, non-centered, covariance while the denominator is a $n(\mathbf{u})$ -point covariance. Indeed, and as an example for the case of the specific two data values $I(\mathbf{u}_1) = 1$ and $I(\mathbf{u}_2) = 0$, that numerator is written as the 3-point non-centered covariance:

$$\text{Prob} \{I(\mathbf{u}) = 1, I(\mathbf{u}_1) = 1, I(\mathbf{u}_2) = 0\} = E \{I(\mathbf{u}) \cdot I(\mathbf{u}_1) \cdot [1 - I(\mathbf{u}_2)]\}.$$

Indicator kriging, when considering the $n(\mathbf{u})$ data altogether as a single multiple-point data event, identifies Bayes relation (Eq. 3.38). Simulating the value $I(\mathbf{u})$ from probabilities of type (3.38) is at the root of the Single Normal Equation SIMulation algorithm *SNESIM* (Strebelle, 2000). Instead of modeling a mp covariance as function of Euclidean distances \mathbf{h} as one would model a 2-point covariance $C(\mathbf{h})$, the two mp covariance values appearing as numerator and denominator of expression 3.38 are lifted directly from a training image. In even simpler terms, the conditional probability (Eq. 3.38) is identified to the experimental proportion of those training replicates of the mp data events that feature at their center location \mathbf{u}' an event $I(\mathbf{u}') = 1$ (Strebelle, 2002).

In essence, the training image provides all the necessary mp covariance values; the decision of stationarity allows scanning a specific training image for replicates (exact or approximate) of the single mp conditioning data event. This is no different from scanning a training image for replicates of pairs of values allowing the modeling of a 2-point covariance/variogram. One may argue that with a variogram one borrows less from that training image; one then forgets that the missing mp stats are then implied in a non-controllable fashion by the simulation algorithm retained; indeed there cannot be any stochastic simulation without a full mp model; recall the previous discussion on the necessity of an mp model in Section 3.3.2. The better answer is that one does trust the training image characteristic structures and mp patterns and wish to use them in the estimation/simulation exercise, and this cannot be done through 2-point statistics such as the variogram.

3.8 Two-point Simulation Algorithms

Traditional (2-point) simulation algorithms aim at reproducing a prior covariance $C(\mathbf{h})$ model, or equivalently a variogram model, that is a statistical relation between any two values $z(\mathbf{u})$ and $z(\mathbf{u} + \mathbf{h})$ in space. The missing information about what should be the relation in space of three or more values taken jointly is then *necessarily* provided by the simulation algorithm retained. Multiple-point structures, if not specified explicitly by the algorithm, are most likely of high entropy nature minimizing organization (Journel and Zhang, 2006).

If you wish the simulated realizations to reflect specific structures and patterns beyond 2-point correlation, these structures must be specified as input to a simulation algorithm that can reproduce them. Specific structures never occur by chance.

The covariance-based simulation algorithms widely used in practice stem from essentially two classes; the first class is anchored on the properties of the multivariate Gaussian RF model (Goovaerts, 1997, p.380), the second class builds on the interpretation of an indicator expected value as a conditional probability (Goovaerts, 1997, p.393), recall expression 3.36. These widely used algorithms all rely on the sequential approach (Section 3.3.1) which divides the task of joint simulation of a large number of variables into that of simulating in sequence each variable one at a time.

This initial release of SGeMS proposes the following well established covariance-based (2-point) simulation algorithms:

- *LUSIM*, or Gaussian simulation with LU decomposition, see Section 8.1.1

and Deutsch and Journel (1998, p.169)

- *SGSIM*, or sequential Gaussian simulation, see Section 8.1.2 and Deutsch and Journel (1998, p.170)
- *COSGSIM*, or sequential Gaussian co-simulation, see Section 8.1.3
- *DSSIM*, or direct sequential simulation, see Section 8.1.4
- *SISIM*, or sequential indicator simulation, see Section 8.1.5 and Deutsch and Journel (1998, p.175)
- *COSISIM*, or sequential indicator co-simulation, see Section 8.1.6
- *BSSIM*, or block sequential simulation, see Section 8.1.7
- *BESIM*, or block error simulation, see Section 8.1.8

3.8.1 Sequential Gaussian simulation

The remarkably convenient properties of the Gaussian RF model explain its success, an almost quasi monopoly of probabilistic models for continuous variable. Indeed any set of Gaussian RVs is fully characterized by its mean vector and covariance matrix; all conditional distributions are Gaussian fully characterized by only two moments, the conditional mean and variance themselves given by simple kriging (Journel and Huijbregts (1978); Anderson (2003)). Thus a Gaussian RF model would appear as the ultimate model when only two-point statistics can be inferred. Unfortunately, the qualifier is that a Gaussian RF maximizes entropy (disorder) beyond the input covariance model (Chilès and Delfiner (1999, p.412); Journel and Deutsch (1993)), hence a Gaussian-based simulation algorithm such as *SGSIM* cannot deliver any image with definite patterns or structures involving more than 2 locations at a time. The previous limitation matters little only if one is simulating a ‘homogeneously heterogeneous’ spatial distribution such as porosity or metal grade within the pre-defined geometry of a relatively homogeneous lithofacies or rock type.

In the *SGSIM* algorithm (Journel (1993a); Goovaerts (1997, p.380)) the mean and variance of the Gaussian distribution at any location along the simulation path is estimated by the kriging estimate and the kriging variance. The value drawn from that distribution is then used as conditioning data. Transform of the original data into a Gaussian distribution may be necessary and is normally performed by the normal score transform, see Section 8.1.2 for the SGeMS implementation.

3.8.2 Direct sequential simulation

It can be shown that reproduction of a covariance model does not require a Gaussian RF, but just that the mean and variance of every conditional distribution be those given by SK; the conditional distribution type need not be Gaussian, it can also vary from one simulation node to another (Journel, 1994; Bourgault, 1997). Consequently there is no need for any normal score transform and back transform. The sequential simulation can be performed *directly* with the original z -variable and data, hence the name ‘direct sequential simulation’ (program *DSSIM*).

One main advantage of *DSSIM* is that the simulation can be made conditional to local linear average z -data. Indeed kriging can accommodate data defined on volume/block support as long that these data are linear average of z -values, see Section 3.6.3. The normal score transform being non-linear would undo such linearity. The absence of a prior transformation of the data in *DSSIM* makes it an algorithm of choice for ‘downscaling’, a process whereby large scale block-support data are ‘un-averaged’ into realizations of smaller support values (Kyriakidis and Yoo, 2005; Boucher and Kyriakidis, 2006). The *DSSIM* simulated values reproduce the target covariance model and honor whatever small support data are available; in addition their block averages match the corresponding block data (Hansen et al., 2006).

The price to pay for the absence of normal score transform is absence of a back transform, hence there is no guarantee for the *DSSIM* simulated z -realizations to reproduce the z -data histogram.

Such global histogram reproduction can be obtained in two ways

- a post-processing similar to the normal score back-transform done in Gaussian simulation. Such back-transform should be such not to undo the data reproduction (Deutsch, 1994b; Journel and Xu, 1994). The utility program *TRANS* discussed in Section 9.1 allows such transform honoring the original point-support data values; this comes, however, at some loss of covariance reproduction.
- put to use the large degree of freedom represented by the ability to choose at any simulation node any distribution type (Bourgault, 1997). The procedure retained by the *DSSIM* code consists of sampling only that part of a translated z -target histogram that match the local SK mean and variance (Soares, 2001; Oz et al., 2003a).

3.8.3 Direct error simulation

In all generality one can express any unsampled value $z(\mathbf{u})$ as the sum of its estimate $z^*(\mathbf{u})$ plus the corresponding estimation error:

$$z(\mathbf{u}) = z^*(\mathbf{u}) + r(\mathbf{u}), \text{ with: } r(\mathbf{u}) = z(\mathbf{u}) - z^*(\mathbf{u})$$

The estimated value $z^*(\mathbf{u})$ is known, but the error is not. Thus simulation would consist of simulating that error $r(\mathbf{u})$ under various constraints. For example, the simulated error should have mean zero and variance equal to the known kriging variance if $z^*(\mathbf{u})$ is obtained by kriging. As for the distribution from which the simulated error should be drawn, it can be Gaussian or else. If the random variable error $R(\mathbf{u})$ is orthogonal (uncorrelated) to the random variable estimator $Z^*(\mathbf{u})$, as guaranteed by kriging (Luenberger (1969); Journel and Huijbregts (1978); Chilès and Delfiner (1999, p.465)), then the error value $r_s(\mathbf{u})$ can be drawn independently of the estimated value $z^*(\mathbf{u})$:

$$z_{cs}(\mathbf{u}) = z_K^*(\mathbf{u}) + r_s(\mathbf{u}) \quad (3.39)$$

where: $z_K^*(\mathbf{u})$ is the kriging estimate;

$r_s(\mathbf{u})$ is an error value drawn from a distribution with zero mean and variance equal to the kriging variance $\sigma_K^2(\mathbf{u}) = \text{Var} \{Z(\mathbf{u}) - Z_K^*(\mathbf{u})\}$;

$z_{cs}(\mathbf{u})$ is the simulated value.

The simulated field $\{z_{cs}(\mathbf{u}), \mathbf{u} \in \text{study area}\}$

- honors the data value $z(\mathbf{u}_\alpha)$ at data location \mathbf{u}_α since $z_K^*(\mathbf{u}_\alpha) = z(\mathbf{u}_\alpha)$ per kriging exactitude
- has the correct variance since:

$$\begin{aligned} \text{Var} \{Z_{cs}(\mathbf{u})\} &= \text{Var} \{Z_K^*(\mathbf{u})\} + \text{Var} \{R(\mathbf{u})\} \\ &= \text{Var} \{Z_K^*(\mathbf{u})\} + [\sigma_K^2(\mathbf{u}) = \text{Var} \{Z(\mathbf{u}) - Z_K^*(\mathbf{u})\}] \end{aligned}$$

per orthogonality of the error $R(\mathbf{u})$ with $Z_K^*(\mathbf{u})$.

However, there remains to ensure that the simulated field $Z_{cs}(\mathbf{u})$ features the same covariance as $Z(\mathbf{u})$. This is obtained in sequential simulation (Sections 3.8.1 and 3.8.2) by adding into the kriging data set for $z_K^*(\mathbf{u})$ all previously simulated value $z_{cs}(\mathbf{u}')$ found in its neighborhood. An alternative is to simulate the error $r_s(\mathbf{u})$ by lifting it from an error training image sharing the same (non-stationary)

covariance as the actual error $R(\mathbf{u}) = Z(\mathbf{u}) - Z_K^*(\mathbf{u})$. That error training image can be generated by repeating the estimation procedure used to generate $z_K^*(\mathbf{u})$ from data $z(\mathbf{u}_\alpha)$, $\alpha = 1, \dots, n$ on a **non-conditional** simulated realization $z_s(\mathbf{u})$ of the random function $Z(\mathbf{u})$ using the same geometric configuration of ‘simulated’ data $z_s(\mathbf{u}_\alpha)$, $\alpha = 1, \dots, n$. This process is written (Journel and Huijbregts (1978); Deutsch and Journel (1998, p.127); Chilès and Delfiner (1999, p.465)):

$$z_{cs}^{(l)}(\mathbf{u}) = z_K^*(\mathbf{u}) + \left[z_s^{(l)}(\mathbf{u}) - z_{Ks}^{*(l)}(\mathbf{u}) \right] \quad (3.40)$$

where: $z_s^{(l)}(\mathbf{u})$ is the l^{th} **non-conditional** simulated realization of the random field $Z(\mathbf{u})$ honoring its covariance model;

$z_K^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_\alpha(\mathbf{u}) \cdot z(\mathbf{u}_\alpha)$ is the kriging estimate built from the actual data value $z(\mathbf{u}_\alpha)$, $\alpha = 1, \dots, n$, the $\lambda_\alpha(\mathbf{u})$ are the kriging weights;

$z_{Ks}^{*(l)}(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_\alpha(\mathbf{u}) \cdot z_s^{(l)}(\mathbf{u}_\alpha)$ is the kriging estimate built from the simulated data values $z_s^{(l)}(\mathbf{u}_\alpha)$ taken from the non-conditionally simulated field $z_s(\mathbf{u})$ at the actual data locations \mathbf{u}_α , $\alpha = 1, \dots, n$;

$z_{cs}^{(l)}(\mathbf{u})$ is the l^{th} conditionally simulated realization.

Warning: a histogram input is required to simulate the intermediate unconditional realizations, however that histogram **may not be** reproduced in the final simulated conditional realizations.

Note that the kriging weights $\lambda_\alpha(\mathbf{u})$ used for both kriging estimates $z_K^*(\mathbf{u})$ and $z_{Ks}^{*(l)}(\mathbf{u})$ are the same, since the simulated field $Z_s(\mathbf{u})$ shares the same covariance model and the same data geometry as the actual field $Z(\mathbf{u})$. There lies the main (potential) advantage of the direct error simulation approach: one single kriging is needed per simulation node \mathbf{u} no matter the number L of conditional simulated realizations $z_{cs}^{(l)}(\mathbf{u})$, $l = 1, \dots, L$, needed. One could then utilize any fast non-conditional simulation algorithm to generate the L required fields $z_s^{(l)}(\mathbf{u})$ (Chilès and Delfiner (1999, p.494,513); Oliver (1995); Lantuejoul (2002)); the L kriginings to obtain the $z_{Ks}^{*(l)}(\mathbf{u})$ are then obtained very fast by mere matrix multiplication from the stored kriging weights $\lambda_\alpha(\mathbf{u})$; last, a mere addition (Eq. 3.39) gives the L conditional fields $z_{cs}^{(l)}(\mathbf{u})$.

The caveat though is that the error field $R(\mathbf{u}) = Z(\mathbf{u}) - Z_K^*(\mathbf{u})$ must be independent (or at least uncorrelated) to the estimated signal $Z_K^*(\mathbf{u})$. This is not a trivial requirement, only guaranteed if simple kriging is applied to a multiGaussian field $Z(\mathbf{u})$.

3.8.4 Indicator simulation

Indicator simulation was introduced to simulate categorical variables defined through a set of K binary indicator variables (Journel (1983); Goovaerts (1997, p.423); Chilès and Delfiner (1999, p.512)). The algorithm was later extended to simulation of a continuous variable made discrete over K classes. Consider the corresponding two definitions:

$$I_k(\mathbf{u}) = \begin{cases} 1 & \text{if the category } k \text{ occurs at } \mathbf{u} \\ 0 & \text{if not} \end{cases}, \quad (3.41)$$

or

$$I(\mathbf{u}; z_k) = \begin{cases} 1 & \text{if } Z(\mathbf{u}) \leq z_k \\ 0 & \text{if not} \end{cases}.$$

Indicator kriging (Section 3.6.5) would provide estimates of the K class probabilities conditional to the local data set $n(\mathbf{u})$, respectively:

$$\text{Prob} \{ \mathbf{u} \in k | n(\mathbf{u}) \} \in [0, 1], \quad k = 1, \dots, K, \quad (3.42)$$

with: $\sum_{k=1}^K \text{Prob} \{ \mathbf{u} \in k | n(\mathbf{u}) \} = 1$, or

$$\text{Prob} \{ Z(\mathbf{u}) \leq z_k | n(\mathbf{u}) \} \in [0, 1], \quad k = 1, \dots, K,$$

with $\text{Prob} \{ Z(\mathbf{u}) \leq z_k | n(\mathbf{u}) \} \leq \text{Prob} \{ Z(\mathbf{u}) \leq z_{k'} | n(\mathbf{u}) \}, \forall z_{k'} \leq z_k$.

From these IK-derived conditional probabilities a class indicator can be simulated at each simulation node \mathbf{u} , the indicator of a category or of the class to which the continuous z -value belongs.

Note that at each node $(K-1)$ indicator kriginings are needed if K classes are considered, each kriging calling for its own indicator covariance model. In the case of a continuous variable $Z(\mathbf{u})$, the modeling task is considerably reduced if one adopts a median indicator model, whereby all $(K-1)$ indicator covariance models are chosen proportional to the single covariance corresponding to the indicator defined by the median threshold value $z_k = M$ (Goovaerts (1997, p.304); Chilès and Delfiner (1999, p.384)).

Recall that the individual indicator kriging results must be corrected to honor the constraints associated to expression 3.42 (Goovaerts, 1997, p.324). These order relation corrections are made before simulated values can be drawn from the IK-estimated conditional probabilities.

It can be shown that the indicator covariance models are reproduced, except for the impact of the order relation corrections.

Remarks

The indicator formalism was originally designed for categorical variables and later extended to continuous variables. As for simulation of categorical variables, *SISIM* should not be used for more than 3 or 4 categories; beyond $K = 4$ the number and magnitude of order relation corrections become prohibitive and reproduction of the numerous indicator covariances is poor.

Hierarchy and spatial nesting of categories can be used to split the simulation of a large number K of categories into a series of independent simulations, each with a smaller number of categories (Maharaja and Journel, 2005). For example, the simulation of $K = 5$ lithofacies may be reduced to, first a simulation of the two dominant groups of facies ($K = 2$), followed by the simulation of the individual facies nested within each group, say $K = 3$ within any previously simulated first group and $K = 2$ for the second group.

3.9 Multiple-point Simulation Algorithms

The concept of multiple-point simulation was triggered by the failure of well established object-based algorithms to honor either a large amount of local data. With object-based algorithms, also called Boolean algorithms, ‘objects’ of given shape are dropped onto the simulation study area thus painting onto that area the desired shapes and patterns (Chilès and Delfiner (1999, p.545); Stoyan et al. (1987); Haldorsen and Damsleth (1990); Lantuejoul (2002); Mallet (2002)). The object shape parameters, e.g. size, anisotropy, sinuosity are made random thus making the simulation process stochastic. A cumbersome iterative process is then applied for local data conditioning: objects are displaced, transformed, removed, replaced until a reasonable match of these local data was achieved. Object-based algorithms proved ideal for building a training image with the required spatial structures and patterns, but they are notoriously difficult to condition to local data, particularly when these data are of small support volume, numerous and of diverse types. Conversely, pixel-based traditional algorithms are easy to condition because the simulation progresses one pixel (point) at a time: modifying a single point-support value to match local data does not affect a whole object area around that point. But traditional algorithms being 2-point statistics-based could only reproduce a variogram or covariance model, failing to reproduce definite shapes and patterns.

3.9.1 Single normal equation simulation (*SNESIM*)

Without losing the data conditioning flexibility of a pixel-based process, one had to find a way around the variogram limitation. That variogram comes in only when building from kriging the local conditional probability distributions (see previous Section 3.8), hence the idea of collecting directly those distributions from training images that display the required spatial patterns. By so doing one would sidestep any variogram/covariance modeling and also any kriging. The probability distributions are chosen from the training image such as to match, exactly or approximately, the local conditioning data. More precisely, the training image is scanned to retrieve replicates of the conditioning data event; these replicates define a sub training population conditioned to the data from which the previous conditional distributions could be retrieved (Guardiano and Srivastava, 1993; Strebelle, 2002). The *SNESIM* algorithm reads conditional distributions from training images that could have built using ideally suited non-conditional object-based algorithms, and progresses sequentially one pixel at a time thus capitalizing on the data conditioning ease of sequential simulation.

The main requisite, and a difficult one of the *SNESIM* algorithm is, a ‘rich’ training image where enough exact replicates can be found for any conditioning data event encountered during the sequential simulation process. At any location, if not enough such replicates are found some of the local conditioning data are dropped, allowing the possibility to find more replicates but at the cost of poorer data conditioning. That limitation becomes prohibitive if the simulation addresses too many categories ($K > 4$), or worse if the variable simulated is continuous. One must then revert to the *FILTERSIM* algorithm (Journel and Zhang, 2006; Zhang et al., 2006) which accepts approximate replicates of the conditioning data event.

The multiple-point (mp) sequential simulation algorithm whereby all conditional probabilities are read as corresponding proportions from a training image is called ‘Single Normal Equation SIMulation’. This name recalls that any such proportion is in fact the result of a single indicator kriging (normal) equation, see relation 3.38.

The original *SNESIM* code (Guardiano and Srivastava, 1993) had to re-scan the training image anew at each simulation node to collect replicates of that node conditioning data event; it gave good results but was CPU-prohibitive. The breakthrough came with the introduction of the search tree concept which allowed for a single scan of the training image and smart storage in central memory of *all* resulting training proportions (Strebelle, 2002). These proportions are then directly

read from the search tree during the course of sequential simulation. The road-block of the *SNESIM* algorithm is not anymore one of CPU, it is the demand for a large and ‘rich’ enough training image carrying enough replicates of most of the conditioning data events found in the course of simulation.

Details about the search tree-based *SNESIM* algorithm can be found in Strebelle (2000). The implementation of the *SNESIM* algorithm can be found in Section 8.2.1.

3.9.2 Filter-based algorithm (*FILTERSIM*)

A middle alternative between pixel-based and object-based algorithms is to cut into small pieces the objects or, better, a whole non conditional training image, then use those pieces to reconstruct that training image making sure the pieces fit the conditioning data (Arpat, 2004; Zhang, 2006; Journel and Zhang, 2006). The best analogy is perhaps that of building a puzzle, where each new piece patched onto the image being simulated must fit close-by previously placed pieces and original data. Similar to a smart puzzle reconstruction, the search for pieces that would fit is speeded up by looking into bins containing previously classified ‘similar’ looking pieces; say, one specific bin would contain all T_i pieces with some elements of sky in it, another bin would contain parts of trees and houses in it. As opposed to the puzzle game, any piece taken out of a bin is immediately replaced by an identical one thus no bin ever gets exhausted. Also the fit required is only approximate and it can be reconsidered later in the sequential simulation path.

Instead of putting down on the simulation field an entire T_i piece, only the central part of that piece can be patched down. That central part or patch size can be as small as the single central pixel value.

The critical key to the success of the *FILTERSIM* algorithm is the classification of local patterns of the training image into a not too large number of bins of ‘similar’ looking patterns. That classification requires reducing any pattern to a small number of characteristic scores, say, sky with clouds or sky without clouds. In *FILTERSIM* these scores are defined through linear filters applied to the set of pixel values constituting the pattern (Schneiderman and Kanade, 2004). Next, one must define a distance between a conditioning data event and any such previous bin. This is needed to select the bin closest, that is with training patterns most similar, to the conditioning data event. Future research will undoubtedly suggest better pairs of (filters + distance) than that coded in this early version of the *FILTERSIM* code.

The *FILTERSIM* algorithm, originally designed for simulation of continuous

variables, has been extended to categorical variables. However and because the notion of linear filters does not extend naturally to categorical variables, we recommend using the categorical *FILTERSIM* approach only when it is absolutely necessary to simulate jointly a large number of categorical variables ($K > 4$). For a reasonable hence small number of categories the *SNESIM* approach is a better choice provided a corresponding large and varied (rich) training image is available.

Conversely, *SNESIM* should not be used to simulate simultaneously too many categories (> 4) or the too many classes needed to discretize a continuous variable. Indeed the *SNESIM* search for exact replicates of the conditioning data event is too strict for a continuous variable: there would never be enough such exact replicates.

3.9.3 Hierarchical simulation

Because of the difficulty in obtaining very large and rich training images, particularly in 3D, and because of the RAM demand of the corresponding search trees, it may not be feasible to apply the *SNESIM* algorithm to the joint simulation of more than $K = 4$ categories. That limitation is generally not a problem in earth sciences applications, since facies or rock types are often nested in each other, which allows decomposing the problem, see Maharaja and Journel (2005); Walker (1984).

Consider, for example, the simulation of 7 facies, with facies#5 and #6 nested within facies#3, and facies#7 nested within facies#4. A first run of *SNESIM* with a four-facies Ti would yield simulated realizations of the four principal facies (#1,2,3,4) distributions. Consider any one such realization and isolate the corresponding zones simulated as facies#3 and facies#4; within the zone#3 use *SNESIM* with the proper Ti to simulate the distribution of facies#5 and #6; within the zone#4 use *SNESIM* with yet another Ti to simulate the distribution of facies#7.

3.10 The nu/tau Expression for Compositing Conditional Probabilities

Stochastic prediction is all about proposing a model for the probability distribution of possible outcomes of an unknown given all the data available. From such

distribution model, one can simulate a set of outcomes for the unknown(s). The fundamental task is thus to determine the previous conditional distributions, a task particularly difficult when data of different types are present; data that are often redundant one with each and all others, and data whose information content goes much beyond a linear correlation with the unknown being assessed. Recent developments have uncovered a dormant general formulation of that general problem, one that lends itself remarkably to the modern mp approach to data integration (Bordley, 1982; Benediktsson and Swain, 1992; Journel, 2002; Polyakova and Journel, 2008).

At this level of absolute generality, some notations are necessary; we will endeavor, however, to back these notations with intuitive examples.

Adopt the notation A for the unsampled RV, and the notations $D_i = d_i$, $i = 1, \dots, n$ for the n data events, with capital letters denoting the RVs and the corresponding small case letters denoting any observed data value. In mp applications, the D_i 's are actually vectors involving multiple data locations, but we will keep the scalar notation D_i for simplicity.

The ultimate goal of probabilistic prediction is to evaluate the fully conditional probability:

$$\text{Prob} \{A = a | D_i = d_i, i = 1, \dots, n\}, \quad (3.43)$$

a function of the $(n + 1)$ values $(a; d_i, i = 1, \dots, n)$.

If each data event D_i relates to a single location in space, say $d_i = z(\mathbf{u}_i)$, then a traditional two-point statistics such as the covariance suffices to relate any datum D_i to any other D_j or to the unknown A .

If each data event D_i involves jointly multiple data locations (it is then a vector) all related to the *same* attribute z which is also the A -attribute, then one could hope to find or build a Z -training image depicting the joint distribution of A and any vector of Z -values. Using such training image, the mp algorithms *SNESIM* and *FILTERSIM* could be implemented, see Section 3.9.1 and 3.9.2.

However, in the general situation where each data event D_i , in addition of being multiple-point, is also related to a different attribute, the task becomes insuperable. For example, D_1 could be a mp pattern of facies indicator data as interpreted from well logs, D_2 could be a set of seismic impedance data involving many locations in space but locations different from those related to D_1 ; as for A it may relate to a third attribute, say porosity, at yet a different location (or set of locations).

The solution is again to 'divide and conquer', decomposing the global data event $\mathbf{D} = \{D_i = d_i, i = 1, \dots, n\}$ into n component data events D_i for which

each of the individual conditional probabilities $\text{Prob} \{A = a | D_i = d_i\}$, $i = 1, \dots, n$ could be evaluated by traditional two-point or by mp geostatistics or by any other means. The general problem is then that of recombining n individual probabilities into an estimate of the fully conditioned probability (Eq. 3.43); this calls for determination of the integration function φ below:

$$\text{Prob} \{A = a | D_i = d_i, i = 1, \dots, n\} = \varphi (\text{Prob} \{A = a | D_i = d_i\}, i = 1, \dots, n). \quad (3.44)$$

Returning to the previous example:

$P(A|D_1)$ could be evaluated from a training image depicting the joint distribution of porosity (the A -attribute) and facies indicators (the D_1 categorical attribute),

$P(A|D_2)$ could be evaluated independently from calibration of porosity to a set of neighboring seismic impedance data (the D_2 categorical attribute),

There remains to combine these two partially conditioned probabilities accounting for the redundancy of seismic and facies data when it comes to evaluate porosity (A).

Fortunately there exists an *exact* decomposition formula of the type (Eq. 3.44), the so-called nu or tau expression. This expression has been known for some time (Bordley, 1982; Benediktsson and Swain, 1992), but its generality or exactitude had not been established until recently, nor was its importance for data integration fully recognized.

Warning

All probabilities in expressions 3.43 and 3.44 are functions of the $(n + 1)$ values a and d_i , more if \mathbf{d}_i is a mp vector of data values. However for simplicity we will use the short notations $P \{A|D\}$ and $P \{A|D_i\}$ whenever there is no risk of confusion.

Why probabilities?

Before developing the expression of the compositing function φ , one should answer the question whether a probabilistic approach is the most appropriate for this data integration problem. The answer lies in the notation 3.44 itself:

- probabilities provide a unit-free, standardized $[0, 1]$, coding of information, across all data types, which facilitates the task of data integration;

- as opposed to a deterministic estimate of A , each elementary probability $P(A = a|D_i = d_i)$ includes both the d_i information content and the uncertainty of its contribution to evaluating $A = a$.

The nu/tau expression

Consider the probability-into-distance transform of each individual probability:

$$x_0 = \frac{1 - P(A)}{P(A)}, \quad x_1 = \frac{1 - P(A|D_1)}{P(A|D_1)}, \quad x_n = \frac{1 - P(A|D_n)}{P(A|D_n)}, \quad (3.45)$$

all valued in $[0, +\infty]$.

$P(A) = P(A = a)$ is the prior probability of event $A = a$ occurring, ‘prior’ to knowing any of the n data $D_i = d_i$,

x_0 is the prior distance to $A = a$ occurring,

equal to zero if $P(A) = 1$, equal to ∞ if $P(A) = 0$,

and similarly for each of the elementary distance x_i .

We will use the notation: $1 - P(A|D_i) = P(\tilde{A}|D_i)$, where \tilde{A} stands for non A .

The distance x to $A = a$ occurring given *jointly* all n data is given by the nu, expression:

$$\frac{x}{x_0} = \prod_{i=1}^n \nu_i \frac{x_i}{x_0} = \nu_0 \cdot \prod_{i=1}^n \frac{x_i}{x_0}, \quad \text{with: } \nu_i \geq 0$$

or equivalently the tau expression:

$$\frac{x}{x_0} = \prod_{i=1}^n \left(\frac{x_i}{x_0} \right)^{\tau_i}, \quad \text{with: } \tau_i \in [-\infty, +\infty] \quad (3.46)$$

with

$$\nu_i = \left(\frac{x_i}{x_0} \right)^{\tau_i - 1}, \quad \text{or: } \tau_i = 1 + \frac{\log \nu_i}{\log \frac{x_i}{x_0}}$$

and

$$\nu_0 = \prod_{i=1}^n \nu_i \in [0, +\infty] \quad (3.47)$$

Recall that:

$$x = \frac{P(\tilde{A}|D_1, \dots, D_n)}{P(A|D_1, \dots, D_n)}$$

thus

$$P(A|D_1, \dots, D_n) = \frac{1}{1+x} \in [0, 1] \quad (3.48)$$

The tau/nu expressions give the fully conditioned relative distance x/x_0 as function of the n elementary relative distances x_i/x_0 . Recall that these n elementary distances are assumed known, the only problem addressed by the two equivalent relations 3.46 is that of combining the elementary distances into the fully conditioned distance x . Expressions 3.46 shows the combination function to be a weighted product as opposed to an approach by indicator kriging that considers weighted linear combination, see Section 3.6.5. The relative distances carry the information content of each elementary data event D_i ; the tau or nu weights account for the additional information (beyond redundancy) carried by the various data events as to evaluating the probability for $A = a$.

The exact expression of the ν -parameters is (Polyakova and Journal, 2008):

$$\nu_i = \frac{\frac{P(D_i|\tilde{A}, \overline{D_{i-1}})}{P(D_i|A, \overline{D_{i-1}})}}{\frac{P(D_i|\tilde{A})}{P(D_i|A)}} \in [0, +\infty], \quad \nu_1 = 1, \quad (3.49)$$

and similarly for the tau parameters (Krishnan, 2004):

$$\tau_i = \frac{\log \frac{P(D_i|\tilde{A}, \overline{D_{i-1}})}{P(D_i|A, \overline{D_{i-1}})}}{\log \frac{P(D_i|\tilde{A})}{P(D_i|A)}} \in [-\infty, +\infty], \quad \tau_1 = 1, \quad (3.50)$$

where $\overline{D_{i-1}} = \{D_j = d_j, j = 1, \dots, i-1\}$ denotes the set of all data events considered before the i^{th} data event $D_i = d_i$.

$\text{Prob}(D_i|A)$ is the (likelihood) probability of observing the datum value $D_i = d_i$ given the outcome $A = a$, $\text{Prob}(D_i|\tilde{A})$ is the probability of observing the same data but given \tilde{A} , thus the ratio $\frac{\text{Prob}(D_i|\tilde{A})}{\text{Prob}(D_i|A)}$ appearing in the denominator of the ν_i or τ_i expression can be read as a measure of how datum $D_i = d_i$ discriminates A from \tilde{A} . The ratio appearing in the numerator is the same discrimination measure but in presence of all previous data considered $\overline{D_{i-1}} = \{D_1 = d_1, \dots, D_{i-1} = d_{i-1}\}$. The unit value $\nu_i = \tau_i = 1$ would correspond to full information redundancy between the data event $D_i = d_i$ and the previously considered data $\overline{D_{i-1}}$. Thus the parameter values $|1 - \nu_i|$ or $|1 - \tau_i|$ could be read as the additional information content brought by D_i above the previous data $\overline{D_{i-1}}$ as to discriminating $A = a$ from $A = \text{non } a$.

Note that the single correction parameter ν_0 is data sequence-independent; also the case $\nu_0 = 1$ is more general than $\nu_i = 1, \forall i$; it encompasses complex case of data redundancies ($\nu_i \neq 1$) that cancel each other globally into $\nu_0 = 1$.

Tau or Nu model?

Recall that all expressions above (Eq. 3.43 to 3.50) are data values-dependent notwithstanding their short notation, say, $P(A|D_i)$ should be read $P(A = a|D_i = d_i)$; similarly the elementary distance x_i is both a and d_i values-dependent.

If the ν_i, τ_i parameters are actually evaluated, e.g. from training image, and made data values-dependent, the two expressions in Eq. 3.47 are equivalent. One would then prefer the nu-formulation because it puts forward a single correction parameter $\nu_0(a, d_i; i = 1, \dots, n)$ which is independent of the data sequence D_1, D_2, \dots, D_n . Also, evaluation of the τ_i parameter associated to a non-informative datum such that $P(D_i|\tilde{A}) \approx P(D_i|A)$ would run into problem because of a division by a log ratio close to zero, see Eq. 3.50.

However, if the ν_i, τ_i parameters are assumed constant, independent of the $(a, d_i; i = 1, \dots, n)$ values, then the tau formulation should be preferred. Indeed, consider the case of only two data events with the two different sets of data values:

$$\{D_1 = d_1, D_2 = d_2\} \text{ and } \{D_1 = d'_1, D_2 = d'_2\}$$

- the nu model with constant (homoscedastic) ν_0 parameter value is written:

$$\begin{aligned} \frac{x}{x_0} &= \nu_0 \cdot \frac{x_1}{x_0} \cdot \frac{x_2}{x_0} \quad \text{for data set } \{d_1, d_2\} \\ \frac{x'}{x_0} &= \nu_0 \cdot \frac{x'_1}{x_0} \cdot \frac{x'_2}{x_0} \quad \text{for data set } \{d'_1, d'_2\} \end{aligned}$$

where x, x_1, x_2 are the distance corresponding to $\{d_1, d_2\}$ and x', x'_1, x'_2 are the distance corresponding to $\{d'_1, d'_2\}$. Conditional distances are data values-dependent, as opposed to the prior distance $x_0 = x'_0$. Therefore,

$$\frac{x'}{x} = \frac{x'_1}{x_1} \cdot \frac{x'_2}{x_2}, \quad \forall \nu_0$$

The parameter ν_0 is seen to be ineffective.

- Conversely, the tau model with constant τ_1, τ_2 parameter values is written:

$$\begin{aligned} \log \frac{x}{x_0} &= \tau_1 \cdot \log \frac{x_1}{x_0} + \tau_2 \cdot \log \frac{x_2}{x_0} \quad \text{for data set } \{d_1, d_2\} \\ \log \frac{x'}{x_0} &= \tau_1 \cdot \log \frac{x'_1}{x_0} + \tau_2 \cdot \log \frac{x'_2}{x_0} \quad \text{for data set } \{d'_1, d'_2\} \end{aligned}$$

Thus:

$$\log \frac{x'}{x} = \tau_1 \cdot \log \frac{x'_1}{x_1} + \tau_2 \cdot \log \frac{x'_2}{x_2}, \text{ or equivalently}$$

$$\frac{x'}{x} = \frac{x'_1}{x_1}^{\tau_1} \cdot \frac{x'_2}{x_2}^{\tau_2}$$

The tau parameters, although data values-independent, remain effective unless $\tau_1 = \tau_2 = \nu_0 = 1$.

This latter property of the tau expression, remaining effective even if the τ_i 's are considered data values-independent, make the tau expression 3.46 a convenient heuristic to weight more certain data events. It suffices to make $\tau_i > \tau_j > 0$ to give more importance to data event D_i as compared to data event D_j , whatever the actual data values (d_i, d_j) . That heuristic utilization of the tau model completely misses the main contribution of the tau/nu expression which is the quantification of data redundancy for any specific set of values $(a, d_i; i = 1, \dots, n)$.

SGeMSproposes a utility program, *NU-TAU MODEL* see Section 9.5, to combine prior probabilities using either the nu or the tau expression 3.46 with as input data values-dependent nu or tau parameters. However, in programs *SNESIM* and *FILTERSIM*, only the tau expression is allowed with tau parameters input as data values-independent constant values.

3.11 Inverse problem

A major topic not directly addressed by the SGeMS software is that of integration of difficult data D expressed as a non-analytical, non-linear function ψ of a large number of values $z(\mathbf{u}_\alpha)$ being simulated :

$$D = \psi(z(\mathbf{u}_\alpha), \alpha = 1, \dots, n)$$

The simulated fields $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in S\}$, $l = 1, \dots, L$, must be such that they all reproduce such data, i.e.

$$D^{(l)} = \psi(\{z^{(l)}(\mathbf{u}), \alpha = 1, \dots, n\}) \approx D \quad \forall l = 1, \dots, L$$

where the function ψ is known, although typically only through an algorithm such as a flow simulator.

SGeMS provides realizations $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in S\}$ that can be selected, combined, perturbed and checked to fit approximately the data D . This is known as the general ‘inverse problem’ (Tarantola, 2005); see Hu et al. (2001); Caers and Hoffman (2006) for a geostatistical perspective.