ORIGINAL PAPER

Xavier Emery

Testing the correctness of the sequential algorithm for simulating Gaussian random fields

Abstract The sequential algorithm is widely used to simulate Gaussian random fields. However, a rigorous application of this algorithm is impractical and some simplifications are required, in particular a moving neighborhood has to be defined. To examine the effect of such restriction on the quality of the realizations, a reference case is presented and several parameters are reviewed, mainly the histogram, variogram, indicator variograms, as well as the ergodic fluctuations in the first and second-order statistics. The study concludes that, even in a favorable case where the simulated domain is large with respect to the range of the model, the realizations may poorly reproduce the second-order statistics and be inconsistent with the stationarity and ergodicity assumptions. Practical tips such as the 'multiple-grid strategy' do not overcome these impediments. Finally, extending the original algorithm by using an ordinary kriging should be avoided, unless an intrinsic random function model is sought after.

Keywords Sequential Gaussian simulation · Multigaussian distribution · Kriging neighborhood · Screening effect · Ergodic fluctuations

1 Introduction

Geostatistical simulation is currently used in various application domains, e.g. groundwater hydrology, petroleum reservoir characterization, ore reserve evaluation, environmental and soil sciences. It constitutes a helpful tool for risk assessment and for decision-making as it allows assessing the uncertainty on the unsampled values of the regionalized variable under study, viewed as a realization of a random field with a specific spatial distribution.

The sequential algorithm is one of the most widespread techniques for simulating Gaussian random fields (Ripley, 1987, p 106; Deutsch and Journel, 1992, p 141; Gómez–Hernández and Journel, 1993). However, its practical implementation requires several simplifications, but so far their effects on the accuracy of the algorithm remain unknown to a great extent. Although not exhaustive, this work aims at highlighting and explaining some limitations of the sequential algorithm. Except for its last section, it considers a stationary framework and uses simple kriging, i.e. kriging with a known mean.

X. Emery
Department of Mining Engineering,
University of Chile, Avenida Tupper 2069,
Santiago, Chile
E-mail: xemery@cec.uchile.cl
Tel.: +56-2 678 4498,
Fax: +56-2 672 3504

2 The model and the algorithm

2.1 Gaussian random fields

A random function $\{Y(\mathbf{x}), \mathbf{x} \in \mathrm{R}^d\}$ is said to be multigaussian if every linear combination of its values has a normal distribution. In particular, the prior distribution of every value is normal, but this is not a sufficient condition. Such random function, also called *Gaussian random field*, is fully determined by its moments of first and second orders (mean and covariance or variogram). In the following, these are assumed to be stationary, i.e. shift invariant. Without loss of generality, the mean can be set to zero and the variance to one, so that the spatial distribution is characterized by the variogram or, equivalently, the correlogram.

A key property of Gaussian random fields is that posterior (conditional) distributions are still Gaussian-shaped. More precisely, the distribution of $Y(\mathbf{x})$ conditional to a set of hard data is Gaussian with mean the simple kriging of $Y(\mathbf{x})$ from the dataset and variance the simple kriging variance (Chilès and Delfiner, 1999, p 381). Such property is the foundation of the sequential Gaussian algorithm.

2.2 The sequential Gaussian algorithm

Assume that a Gaussian random field has to be simulated at locations $\{\mathbf{x}_1,...,\mathbf{x}_n\}$. The simulation is performed as follows:

1. draw a normal value U_1 with zero mean and unit variance, and put $Y(\mathbf{x}_1) = U_1$;
2. for each $i \in \{2,...,n\}$, put

$$Y(\mathbf{x}_i) = Y^{SK}(\mathbf{x}_i) + \sigma_{SK}(\mathbf{x}_i)\, U_i \quad (1)$$

where $Y^{SK}(\mathbf{x}_i)$ is the simple kriging of $Y(\mathbf{x}_i)$ from $\{Y(\mathbf{x}_1),...,Y(\mathbf{x}_{i-1})\}$, $\sigma_{SK}(\mathbf{x}_i)$ is the corresponding kriging standard deviation, and U_i is a standard normal random variable independent of $\{U_1,...,U_{i-1}\}$.

At each step, the simulated value is incorporated in the dataset used for the kriging in every subsequent step: this is the "sequential" paradigm. Such description is perfectly sound as it relies on a recursive application of Bayes' identity (Chilès and Delfiner, 1999, p 463).

2.3 Pros and cons of the sequential algorithm

The sequential Gaussian algorithm is conceptually very simple and straightforward. One of its main advantages is the direct conditioning of the simulations to a set of hard data: it only requires considering these data as if they were previously simulated values. Another advantage is the possibility to refine an existing simulation, i.e. to increase its resolution.

However, several drawbacks are inherent to the sequential paradigm. The first one deals with random fields whose variogram is smooth near the origin, like a Gaussian model: such variogram may entail a quasi–singularity in the kriging system whenever a fine-scale simulation is sought after (Lantuéjoul, 1994, p 148). This problem can be avoided by performing the simulation in two steps: first a non-conditional simulation, then a conditioning kriging. Indeed, the first step does not need to resort to a kriging and its ill-conditioned matrices: for instance, the spectral method (Shinozuka and Jan, 1972; Lantuéjoul, 2002, p 192) is suited to the simulation of a Gaussian variogram over a continuous domain. Concerning the second step, the kriging matrix only involves the original data, not the simulated nodes, so there is no singularity problem either unless the data themselves are close together.

A second drawback concerns the growing computational requirements needed to perform the sequential simulation: the kriging system increases as the simulation proceeds, since it involves not only the initial data but also all the previously simulated values (Eq. 1). If many locations are considered, simplifications are necessary to speed up the algorithm. In practice, the conditioning values retained at each step (initial data + previously simulated values) are only the nearest to the node being simulated. Further values are deemed not relevant as their influence is screened by the closest ones (Deutsch and Journel, 1992, p 124). In other words, a *moving neighborhood* is defined and a prespecified maximum number of conditioning values is searched within this neighborhood, whereas the theoretical approach requires using a *unique neighborhood*. In general, the screening effect of the closest values is partial, hence the moving neighborhood entails a loss of accuracy.

Remarks

1. The ordering of the locations $\{\mathbf{x}_1,...,\mathbf{x}_n\}$ in the formal algorithm plays no role. However, in practice, it is usually chosen at random among all the possible sequences. Such "randomization" of the visiting sequence aims at avoiding artifacts in the realizations (Deutsch and Journel, 1992, p 125): indeed, a regular (for instance, row-wise) sequence may propagate a flaw or anomaly caused by the neighborhood restriction.
2. As the sequence changes from one realization to another, the kriging configurations are not repeated and no simplification can be made, hence the algorithmic implementation is quite slow (one kriging is required per realization).

2.4 How can a simulation algorithm be validated?

We now tackle the fundamental question of validating a simulation algorithm. A first idea is to compare the regional histogram and variogram of a set of non-conditional realizations with the marginal distribution and variogram of the random function model to simulate. However, since the simulated domain is always bounded, fluctuations will be observed between the realization statistics and the model statistics (Matheron, 1989a, p 78). Accordingly, the fact that the regional histogram or variogram of a realization does not match the theoretical model does not necessarily imply that the simulation algorithm does not work properly. The comparison with the model can only be made after averaging the regional statistics over a large number of realizations.

Contrarily to what many users may believe, checking that the marginal distribution and the variogram are well reproduced on average is not enough. Indeed, there exist many random function models with the same marginal distribution and same variogram, e.g. a Gaussian random field with an exponential variogram and a Poisson tessellation with Gaussian marginal. The differences between these models can be seen by examining other parameters of the spatial distribution, such as indicator variograms or moments of greater order (multiple-point statistics).

Another fundamental aspect for controlling the quality of a simulation algorithm is the analysis of the aforementioned fluctuations of the realization statistics.

For instance, the dispersion of the regional variogram around the theoretical model (*fluctuation variance*) depends on the spatial distribution of the random function, in this case the multigaussian distribution (Matheron, 1989a, p 81). Although this feature is hardly documented and often misunderstood in the geostatistical literature, it constitutes a powerful tool to validate a simulation algorithm referred to a given spatial distribution model: one can check whether the observed fluctuations are reasonably consistent with the ones expected in the theoretical model. In this respect, Lantuéjoul (1994, p 163) warned against simulation algorithms that produce regional variograms in a bounded domain without fluctuations.

In the following sections, the properties and limitations of the sequential Gaussian algorithm are studied through an example, which is deemed to constitute a "reference" case. We resort to numerical experiments because the complexity of the problem considerably increases as soon as a few locations are simulated, so that a detailed error analysis is beyond reach. We will examine the quality of reproduction of the histogram, variogram and indicator variograms (the analysis of multiple-point statistics is left aside), then we will turn to the study of the fluctuations in the first and second-order statistics.

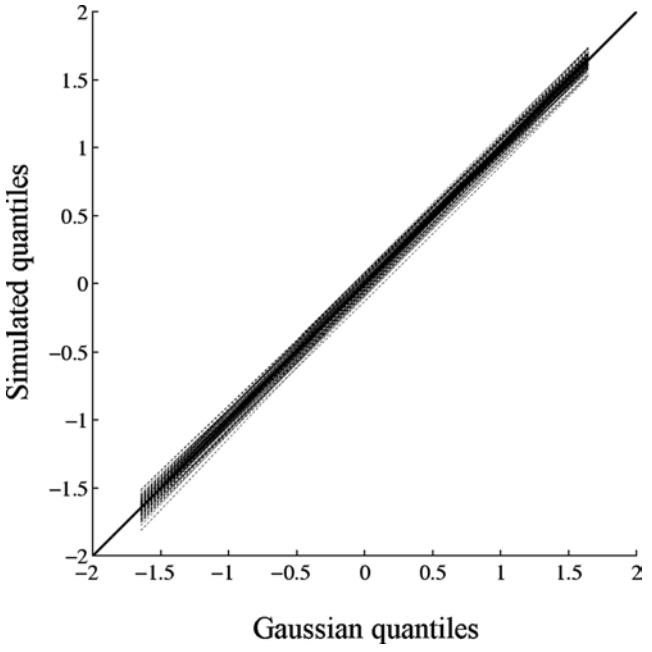


Fig. 1 Quantile-quantile plots between a standard normal distribution and the simulated distributions, the first bisector indicates a perfect coincidence between the compared distributions

3 Reproduction of marginal and bivariate distributions

In a two-dimensional domain V of size 1024×1024, a set of 200 non-conditional simulations is drawn, using a kriging neighborhood containing a maximum of 20 conditioning values. The variogram model is a nugget effect with sill 0.1 plus an isotropic spherical model with range 50 and sill 0.9; this is a very favorable situation since the range is smaller than one twentieth of the domain length. To preclude misinterpretations of the results, this example considers neither an initial dataset nor a histogram transformation.

3.1 Marginal distribution

Even if a moving neighborhood is used in the sequential algorithm, Eq. (1) ensures that each simulated value a) has a zero expectation, b) has a unit variance because of the orthogonality property of simple kriging (Chilès and Delfiner, 1999, p 162) and, c) is normally-distributed as a weighted sum of independent Gaussian random variables. Consequently, on average over all the realizations, the sequential algorithm reproduces the marginal distribution (Fig. 1).

3.2 Variogram and covariance

Figure 2 plots the simulated regional variograms along the first axis direction, together with the theoretical model. We notice that, on average, the sequential Gaussian algorithm provides a biased variogram and increases the range with respect to the underlying model. The bias is not negligible: the apparent range is close to 63 instead of 50, which means it is overvalued by 25%. Such overestimation of the range can be observed in other references (Deutsch and Journel, 1992, p 128; Tran, 1994, p 1165; Yao, 2004, p 501).

How can the bias in the variogram be explained? At first sight, using a moving neighborhood should entail a loss of correlation at large scale, since the distant conditioning data are discarded. Paradoxically the opposite situation takes place, which can be illustrated on a simple example. Consider the configuration shown in Fig. 3, where the locations are numbered according to the visiting sequence, and assume that the moving neighborhood only includes the closest node (i.e. its radius is greater than ℓ but smaller than L).

Let $\rho(\mathbf{h})$ be the theoretical covariance. The simulated values can be written as follows (Eq. 1):

$$\begin{cases} Y(\mathbf{x}_1) = U_1 \\ Y(\mathbf{x}_2) = \rho(\mathbf{x}_2 - \mathbf{x}_1)\, Y(\mathbf{x}_1) + \sqrt{1 - \rho(\mathbf{x}_2 - \mathbf{x}_1)^2}\, U_2 \\ Y(\mathbf{x}_3) = \rho(\mathbf{x}_3 - \mathbf{x}_2)\, Y(\mathbf{x}_2) + \sqrt{1 - \rho(\mathbf{x}_3 - \mathbf{x}_2)^2}\, U_3 \end{cases} \quad (2)$$

where $\{U_1, U_2, U_3\}$ are independent Gaussian variables with zero mean and unit variance. Consequently, the

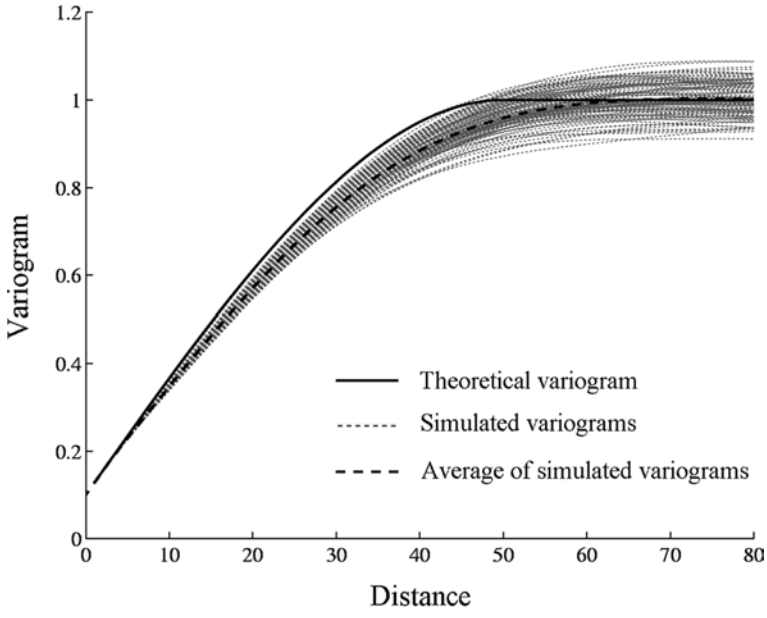


Fig. 2 Simulated variograms and theoretical model (neighborhood containing 20 conditioning values)

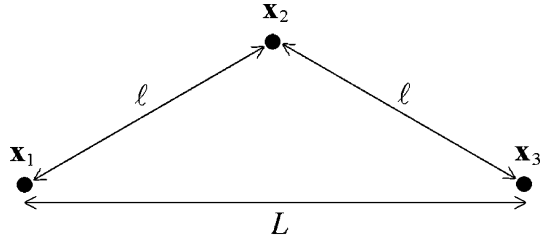


Fig. 3 The locations to simulate are the vertices of an isosceles triangle

covariance between $Y(\mathbf{x}_1)$ and $Y(\mathbf{x}_3)$ is not the expected one:

$$\begin{aligned}\mathrm{cov}[Y(\mathbf{x}_1), Y(\mathbf{x}_3)] &= \rho(\mathbf{x}_3 - \mathbf{x}_2)\rho(\mathbf{x}_2 - \mathbf{x}_1) \\ &\neq \rho(\mathbf{x}_3 - \mathbf{x}_1) \end{aligned} \quad (3)$$

For instance, in the case of a spherical model with a 10% relative nugget effect, if L is equal to the range and ℓ is half the range (aligned locations), this covariance is 0.079 instead of zero. The argument that the nearby data screen out the farthest away data is misleading: omitting the latter creates a "relay effect" (Rivoirard, 1984), and increases the range of the simulated covariance with respect to the theoretical model.

Eq. (3) shows that the theoretical covariance cannot be reproduced exactly unless this covariance is a) a pure nugget effect or, b) an exponential model in a one-dimensional space. The latter case is not surprising, since the exponential covariance produces a perfect screening effect (Chilès and Delfiner, 1999, p 202): the kriging weights are equal to zero except for the data adjacent to the location being considered. A generalization can be obtained by simulating aligned and regularly spaced locations *row-wise* and by restricting the conditioning values to the p previous nodes. This procedure provides an exact simulation of an auto-regressive model of order p over a regular 1D grid; its covariance is a mixture of exponential models and dampened sinusoids (Boulanger, 1990, p 18).

3.3 Indicator variograms and madogram

Because the distribution of a bigaussian pair $\{Y(\mathbf{x} + \mathbf{h}), Y(\mathbf{x})\}$ is symmetric, the indicator variogram associated with a threshold y is equal to:

$$\gamma_I(\mathbf{h}; y) = Prob[Y(\mathbf{x}+\mathbf{h}) \geq y, Y(\mathbf{x}) < y] \quad (4)$$

This probability can be expressed as a function of the covariance $\rho(\mathbf{h})$; for practical calculations, it can be expanded into powers of $\rho(\mathbf{h})$ (Chilès and Delfiner, 1999, p 399). A related tool is the *madogram* or *first order variogram*, which is the sum of all the indicator variograms (Matheron, 1989b, p 30):

$$\gamma_1(\mathbf{h}) = \frac{1}{2}E\{|Y(\mathbf{x}+\mathbf{h}) - Y(\mathbf{x})|\} = \int_{-\infty}^{+\infty} \gamma_I(\mathbf{h}; y)dy \quad (5)$$

In case of a Gaussian random field, the madogram is proportional to the square root of the variogram (Matheron, 1989b, p 31):

$$\gamma_1(\mathbf{h}) = \sqrt{\frac{1-\rho(\mathbf{h})}{\pi}} \quad (6)$$

Figure 4 plots the indicator variograms related to the quartiles of the marginal distribution as well as the madogram. We observe that the range is always greater than the theoretical value; the explanation is similar to the one given in the variogram analysis. Since the marginal distribution is already Gaussian, no simple

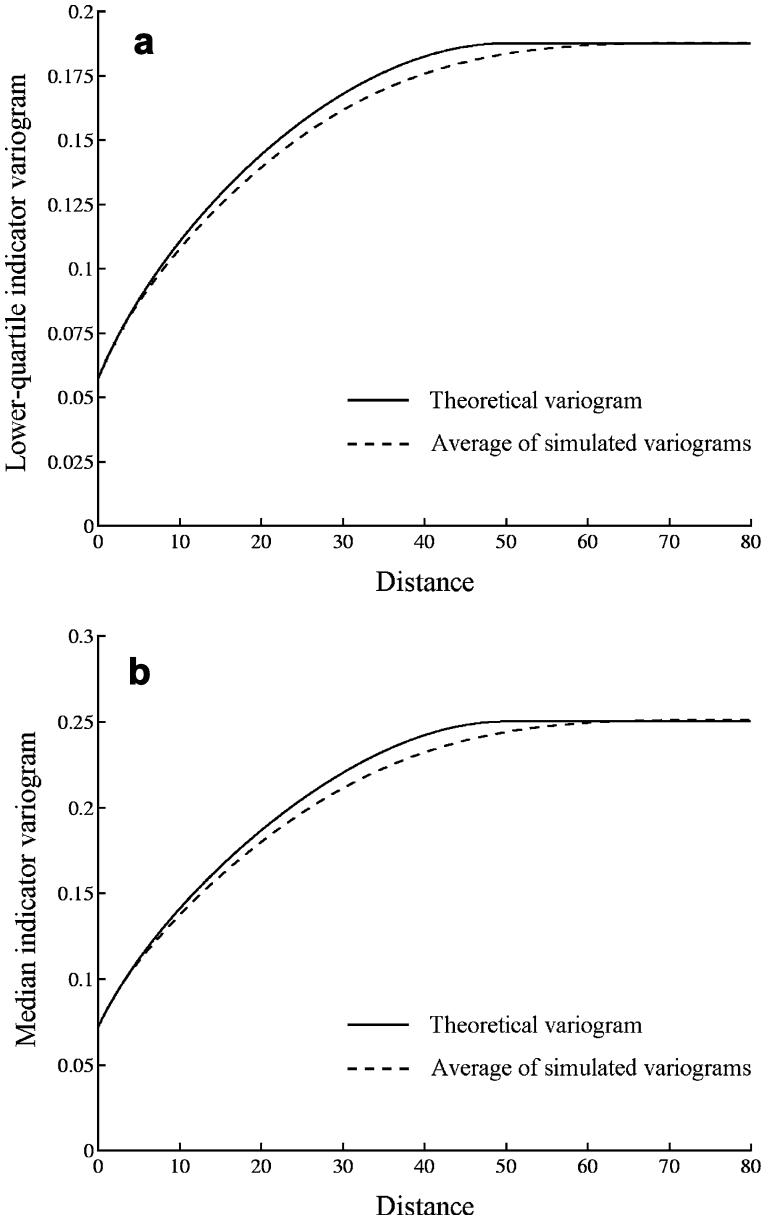


Fig. 4 Quartile indicator variograms and madogram

correction can distort the simulated bivariate distributions into the expected bigaussian distributions.

3.4 Conditions to reproduce the covariance

From a theoretical point of view, the covariance between two simulated values $\{Y(\mathbf{x}_i), Y(\mathbf{x}_j)\}$ and therefore their joint distribution are correctly reproduced if the following two conditions are fulfilled: a) the simulation of the "latest" value in the sequence, say $Y(\mathbf{x}_j)$, uses the other value $Y(\mathbf{x}_i)$ in the kriging system (Eq. 1), and b) the covariances between all the pairs of values used in this system are consistent with the theoretical model $\rho(\mathbf{h})$. Indeed, because of Eq. (1), one has:

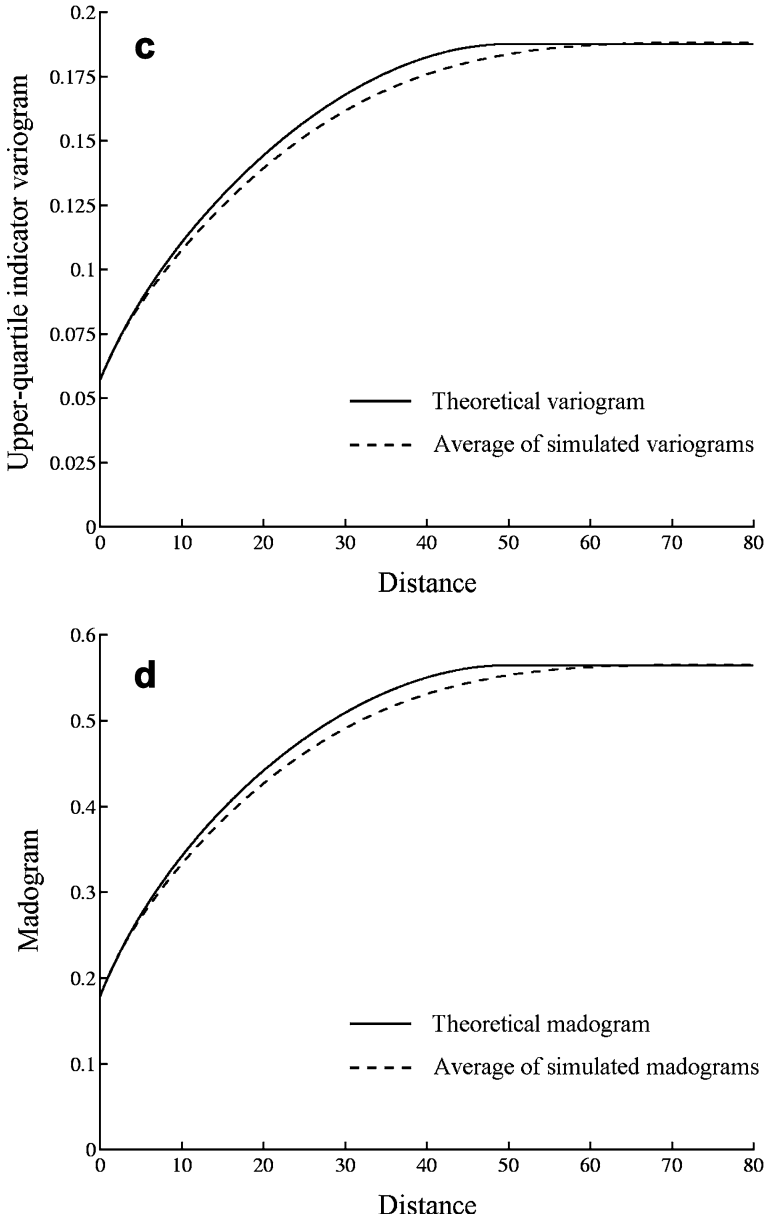
$$cov[Y(\mathbf{x}_j), Y(\mathbf{x}_i)] = \sum_{k \in K} \lambda_k^{SK} cov[Y(\mathbf{x}_k), Y(\mathbf{x}_i)] \quad (7)$$

where K stands for the indices of the values used when kriging $Y(\mathbf{x}_j)$ and $\{\lambda_k^{SK}, k \in K\}$ for the associated simple kriging weights. Under condition b), this equality becomes

$$cov[Y(\mathbf{x}_j), Y(\mathbf{x}_i)] = \sum_{k \in K} \lambda_k^{SK} \rho(\mathbf{x}_k - \mathbf{x}_i) \quad (8)$$

which matches $\rho(\mathbf{x}_j - \mathbf{x}_i)$ due to the simple kriging equations, provided that $i \in K$ (condition a). In particular, the simulated covariance between $Y(\mathbf{x}_i)$ and $Y(\mathbf{x}_j)$ will differ from the model if $Y(\mathbf{x}_i)$ is discarded due to the moving neighborhood restriction. In the example under study, $n = 1{,}048{,}576$ values are simulated, each of them from 20 conditioning values. Consequently, the number of pairs that fulfill condition a) is about $20 \times n$, whereas the total number of pairs is $n \times (n - 1) / 2$: less than 0.004% of the pairs do really reproduce the model

Fig. 4 (Contd.)



covariance! To increase this percentage to 10%, one would need to use more than 50,000 conditioning values in the search neighborhood; for practical applications, this number is clearly beyond reach.

For the pairs that do not fulfill the previous two conditions, the simulated covariance (Eq. 7) depends on the kriging configurations, hence on the visiting sequence and on the value locations, not only on their separation distance. Put another way, the simulated covariance and variogram are expected to be non-stationary. Now, one may argue that stationarity cannot be falsified on the basis of a realization and what really matters is that its "regional" variogram reproduces the variogram model. This argument is legitimate when only a single realization is available, but the principle of simulation is to draw many realizations of the random function. Therefore, the lack of stationarity is expected to provoke inconsistencies when comparing the realization properties with the model: in other words, the realizations may give an inaccurate image of the underlying spatial uncertainty. This point will be discussed in Sect. 4, when analyzing the concept of ergodic fluctuations.

3.5 Multiple-grid simulation

To minimize the effect of the moving neighborhood restriction, a "multiple-grid" strategy is often applied. It consists in simulating a coarse grid first, then in refining the simulation one or more times. Several authors

(Deutsch and Journel, 1992, p 125; Gómez–Hernández and Journel, 1993; Tran, 1994) argue that the coarse simulation ensures the reproduction of the large-scale variogram structure, whereas the successive refinements account for the short-scale variability.

However, the drawbacks above mentioned will occur when the grid mesh becomes smaller than the variogram range (i.e. when the simple kriging weights are no longer zero and a neighborhood restriction has to be applied): problems are just "delayed" until the simulation is refined to a small scale. As an illustration, let us construct a new set of 200 realizations over V, using five refinement grids. The number of nodes in the search neighborhood and the variogram model are unchanged with respect to the situation shown in Fig. 2. The sequential algorithm still leads to an important bias in the variogram reproduction (Fig. 5). Actually, even with the multiple-grid strategy, the number of pairs that reproduce the model covariance is as in the case of a random visiting sequence (see Sect. 3.4), so there is no reason to expect a great improvement neither in the variogram reproduction nor in the consistency of the realizations with the stationarity and ergodicity assumptions.

A simple example may help to understand why the bias does not disappear with the multiple-grid strategy. Let $\{\mathbf{x}_1,\ldots,\ \mathbf{x}_5\}$ be aligned and regularly spaced locations and L be the distance between $\mathbf{x}_1$ and $\mathbf{x}_5$. Suppose that a sequential simulation is performed with three successive refinement grids: the edge locations $\{\mathbf{x}_1,\ \mathbf{x}_5\}$ are simulated first, then the midpoint $\mathbf{x}_3$, finally the intermediate locations $\{\mathbf{x}_2,\ \mathbf{x}_4\}$; this procedure is known as the *random midpoint displacement method* (Chilès, 1995, p 100). To simplify the calculations, let us assume that each value is simulated from the two adjacent values only (one on each side). For short, here we will denote by $\{Y_1,\ldots,Y_5\}$ the set of simulated values. Simple but tedious calculations based on Eq. (1) lead to the following covariances:

$$\begin{cases} cov(Y_1,Y_5)=\rho(L) \\ cov(Y_1,Y_4)=cov(Y_2,Y_5)=\dfrac{\rho(\frac{L}{4})[\rho(L)+\rho(\frac{L}{2})]}{1+\rho(\frac{L}{2})} \\ cov(Y_1,Y_3)=cov(Y_3,Y_5)=\rho(\frac{L}{2}) \\ cov(Y_2,Y_4)=\dfrac{\rho^2(\frac{L}{4})[\rho(L)+2\rho(\frac{L}{2})+1]}{[1+\rho(\frac{L}{2})]^2} \\ cov(Y_1,Y_2)=cov(Y_2,Y_3)=cov(Y_3,Y_4)=cov(Y_4,Y_5)=\rho(\frac{L}{4}) \end{cases} \tag{9}$$

Therefore, in general, the simulated covariance is incorrect for two lags:

- lag $\mathbf{h}=L/2$: $\dfrac{\frac{2}{3}\rho(\frac{L}{2})+\rho^2(\frac{L}{4})\left[\rho(L)+2\rho(\frac{L}{2})+1\right]}{3[1+\rho(\frac{L}{2})]^2}$ is obtained instead of $\rho(\frac{L}{2})$
- lag $\mathbf{h}=3L/4$: $\dfrac{\rho(\frac{L}{4})[\rho(L)+\rho(\frac{L}{2})]}{[1+\rho(\frac{L}{2})]}$ is obtained instead of $\rho(\frac{3L}{4})$

The only exceptions are the pure nugget effect and the exponential variogram, cases in which the simulation produces exact results.

3.6 Sensitivity to the number of conditioning values

From the preceding statements, it appears that the only way for improving the sequential algorithm is to better design the kriging configurations or to increase the number of conditioning values. So far, the kriging neighborhood only contains 20 conditioning values and one may wonder whether the quality of the realizations would become acceptable with a larger number of values. To answer this question, a second set of realizations are drawn (same variogram model and same domain) with a kriging neighborhood containing up to 100 conditioning values. As shown in Fig. 6, the bias in the

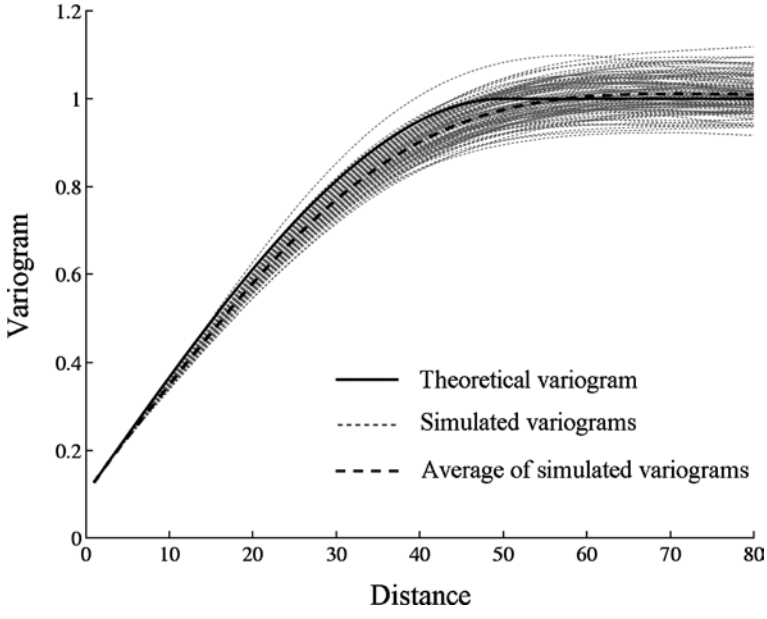


Fig. 5 Simulated variograms obtained by applying a multiple-grid strategy

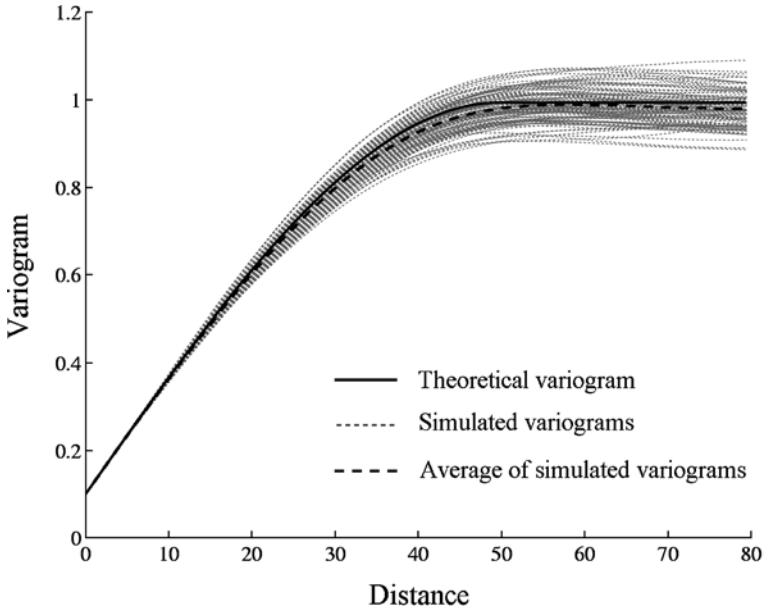


Fig. 6 Simulated variograms and theoretical model (neighborhood containing 100 conditioning values)

simulated variogram is substantially smaller, but it does not disappear (the apparent range is equal to 57, which is not fully satisfactory). It is also worth mentioning that the calculations are much more CPU-intensive than in the case of 20 conditioning values and may become prohibitive if more nodes have to be simulated or many realizations are needed.

Henceforth, we will come back to the first set of realizations, for which the kriging neighborhood contains twenty conditioning values and no multiple-grid strategy is applied, and test the correctness of the sequential algorithm via the analysis of the so-called ergodic fluctuations.

4 Ergodic fluctuations

A random function is *ergodic* in a parameter μ if the corresponding realization statistics $\hat{\mu}_W$ calculated over a subdomain W (also called "regional value over W") converges to μ when the size of W tends to infinity (Chilès and Delfiner, 1999, p 19–21). This definition makes sense only under the stationarity assumption, so that μ is shift invariant. The ergodic property makes statistical inference possible from a single realization of the random function.

In practice, one cannot make the size of W tend to infinity, since W is necessarily included in the whole simulated domain V whose area is finite. Therefore, a discrepancy or "fluctuation" between the model parameter μ and the regional value $\hat{\mu}_W$ is expected (Matheron, 1989a, p 80). Now, $\hat{\mu}_W$ itself can be randomized by substituting the random function for the regionalized variable in its definition, hence the fluctuation is converted into a random variable. Its amplitude can be characterized by a variance:

$$var(\hat{\mu}_W - \mu) = var(\hat{\mu}_W) \tag{10}$$

In the following, two fluctuations are examined, associated with the spatial average over an increasing set of subdomains and the regional variogram respectively.

4.1 Spatial average over a set of subdomains

The spatial average of the random function over a domain W is a random variable, denoted by $Y(W)$ hereafter, whose expected value is the theoretical mean and whose variance is given by

$$var[Y(W)] = \frac{1}{|W|^2}\int_W \int_W \rho(\mathbf{x}-\mathbf{x}')d\mathbf{x}d\mathbf{x}' \tag{11}$$

where $|W|$ stands for the area of W. Eq. (11) can be simplified by introducing the integral range A of the covariance model:

$$A = \frac{1}{\rho(\mathbf{0})}\int \rho(\mathbf{h})\, d\mathbf{h} \quad \text{with, here,} \quad \rho(\mathbf{0}) = 1 \tag{12}$$

Provided that $0 < A < +\infty$ and that $|W|$ is much greater than A, Eq. (11) reduces to (Matheron, 1989a, p 84; Lantuéjoul, 1991, p 393):

$$var[Y(W)] \approx \frac{A}{|W|} \tag{13}$$

In the example under study (spherical model with sill 0.9 and range 50 plus nugget effect with sill 0.1), the integral range is (Lantuéjoul, 2002, p 243)

$$A = \frac{0.9 \times \pi \times 50^2}{5} = 1413.7 \tag{14}$$

Let us now define $D^2(W \mid V)$ as the dispersion variance of W in the whole domain V:

$$D^2(W|V) = \frac{1}{P}E\left\{\sum_{p=1}^{P}[Y(W_p) - Y(V)]^2\right\} \tag{15}$$

where $\{W_p, p = 1,\ldots,P\}$ are subdomains with the same shape, size and orientation as W, that form a partition of V. In this particular case, the expectation can be estimated by an average over the 200 available realizations. By applying Krige's relationship and Eq. (13), one obtains (Matheron, 1989a, p 85; Lantuéjoul, 1991, p 394):

$$D^2(W|V) \approx A\left(\frac{1}{|W|} - \frac{1}{|V|}\right) \tag{16}$$

If V is much greater than W (say, $|V| \geq 10\ |W|$), this identity reduces to:

$$D^2(W|V) \approx \frac{A}{|W|} \tag{17}$$

In log-log coordinates, the points plotting the dispersion variance of W in V as a function of the area $|W|$ should be asymptotically aligned with slope −1. In Fig. 7, two types of subdomains are tested: squares and stripes of length 1024. In both cases, the results are slightly inconsistent since the points cross the theoretical asymptote. This may indicate that the realizations do not conform to a stationary and ergodic model (slope greater than − 1) or that the integral range is not correctly reproduced, hence the asymptote is shifted. By applying Eq. (16) to the dispersion variance of the stripes with size 1024×128 (last point in Fig. 7), the integral range is found to be 1602.8, which overestimates the true value (Eq. 14).

4.2 Regional variogram

Let $\Gamma_V(\mathbf{h})$ be the probabilistic version of the regional variogram over V and $\gamma(\mathbf{h})$ the theoretical model. For a Gaussian random field, the fluctuation variance is expressed as (Alfaro, 1979, p 29; Matheron, 1989a, p 81):

$$var[\Gamma_V(\mathbf{h})] = \frac{1}{2K_{\mathbf{h}}^2(\mathbf{0})}\int [\gamma(\mathbf{u}+\mathbf{h}) + \gamma(\mathbf{u}-\mathbf{h}) - 2\gamma(\mathbf{u})]^2 K_{\mathbf{h}}(\mathbf{u})d\mathbf{u} \tag{18}$$

where $K_{\mathbf{h}}(.)$ is the geometric covariogram of the intersection of V and V shifted by $-\mathbf{h}$. Such formula is independent of the algorithm used to construct the realizations.

Additionally, the distribution of $\Gamma_V(\mathbf{h})$ is approximately Gaussian. Indeed, in the case under study and for the distances considered (0–80 m), the regional variogram at each lag distance is the average of about 1,000,000 random variables with the same distribution (squared differences between pairs of simulated values). Although these random variables are not independent, they are located in an area greater than 700 times the integral range of the model covariance, hence a "mixing" property can be reasonably stated and the central limit theorem (Gordin, 1969) ensures that $\Gamma_V(\mathbf{h})$ is approximately normally-distributed.

In words, for each lag $\mathbf{h}$, the set of simulated variograms shown in Fig. 2 must fulfill three conditions:

1) its average must match the theoretical variogram model
2) its dispersion is ruled by Eq. (18)
3) its shape is approximately Gaussian.

The first condition has already been tested and demonstrated a bias in the variogram reproduction. Here, we focus on the other two conditions. Because of condition 3), for each lag distance, the empirical dispersion of the

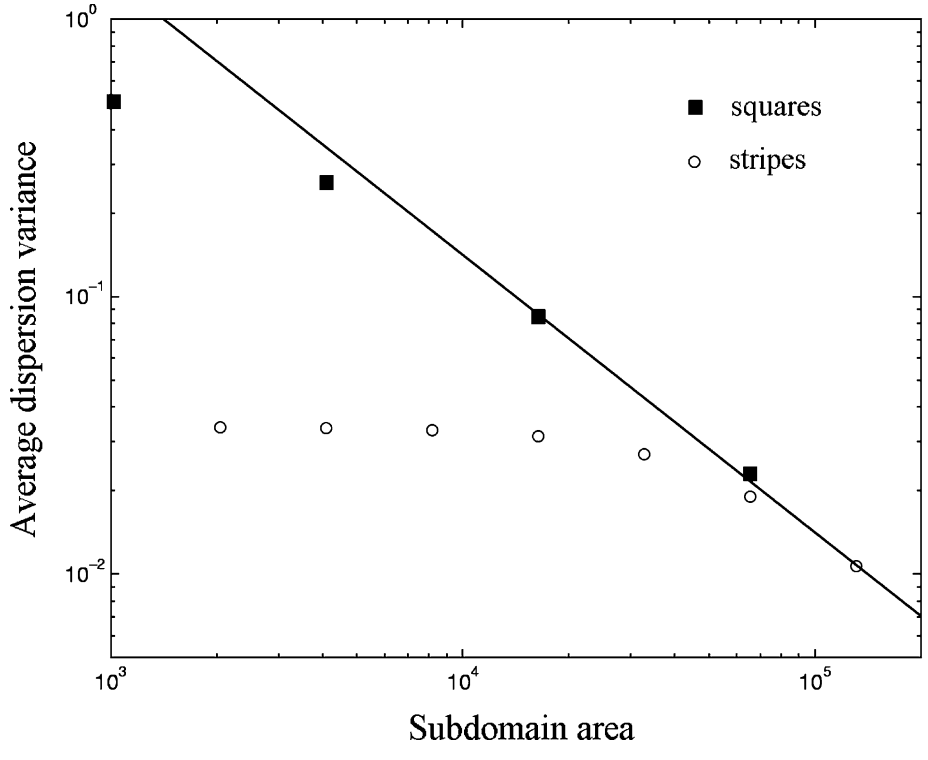


Fig. 7 Average of the empirical dispersion variances versus the subdomain area; the thick line represents the theoretical asymptote

200 simulated variograms should be proportional to a chi-squared variable with 199 degrees of freedom, the mean of which is equal to the theoretical fluctuation variance (Eq. 18). Now, to assess whether the discrepancy between the empirical dispersion of the simulated variograms and the theoretical fluctuation variance is acceptable or not, one can draw a confidence interval (e.g. with a 95% confidence) around the latter (Fig. 8).

The observed empirical dispersion matches the expected variance only when the lag is smaller than 10. For lags between 10 and 45, this dispersion lies outside the 95% confidence interval, which means that the fluctuation is understated and that the set of realizations is not consistent with a stationary and ergodic multigaussian model: in fact, the defective property is stationarity. Such underestimation of the fluctuation variance has also been observed by Ortiz and Deutsch (2002, p 179).

5 On the use of ordinary kriging

A variant of the sequential Gaussian algorithm consists in substituting an ordinary kriging for the simple kriging in Eq. (1). Let us review the implications of this approach, depending on whether a stationary or an intrinsic model is considered.

5.1 Stationary model with unknown mean

In linear geostatistics, ordinary kriging aims at estimating the values of a stationary random function whose mean is unknown and may differ from one region of space to another. Following these statements, one may use an ordinary kriging in the sequential algorithm to reproduce a locally varying mean (Deutsch and Journel, 1992, p 142). However, such approach prevents an accurate reproduction of the second-order statistics, as detailed hereafter.

Let $\{Y(\mathbf{x}), \mathbf{x} \in \mathrm{R}^{\mathrm{d}}\}$ be a Gaussian stationary random field with mean m, variance 1 and covariance $\rho(\mathbf{h})$. Let us suppose that a set of values $\{Y(\mathbf{x}_1),\ldots, Y(\mathbf{x}_{n-1})\}$ is available (say, an original dataset) and simulate a new value $Y(\mathbf{x}_n)$:

$$Y(\mathbf{x}_n) = \sum_{k=1}^{n-1} \lambda_k^{OK} Y(\mathbf{x}_k) + \sigma_{OK}(\mathbf{x}_n) U_n \tag{19}$$

In this equation, the superscript OK refers to ordinary kriging. The kriging weights and variance satisfy the following system, in which μ is a Lagrange multiplier:

$$\begin{cases} \sum_{k=1}^{n-1} \lambda_k^{OK} \rho(\mathbf{x}_k - \mathbf{x}_\ell) + \mu = \rho(\mathbf{x}_n - \mathbf{x}_\ell) \forall \ell \in \{1,\ldots,n-1\} \\ \sum_{k=1}^{n-1} \lambda_k^{OK} = 1 \\ \sigma_{OK}^2(\mathbf{x}_n) = 1 - \sum_{k=1}^{n-1} \lambda_k^{OK} \rho(\mathbf{x}_n - \mathbf{x}_k) - \mu \end{cases} \tag{20}$$

Now, the covariance between the new value and an original datum $Y(\mathbf{x}_\ell)$ (with $\ell \in \{1,\ldots,n-1\}$) is not the expected one. The difference is equal to the Lagrange multiplier introduced in Eq. (20):

$$\begin{aligned} cov[Y(\mathbf{x}_n), Y(\mathbf{x}_\ell)] &= \sum_{k=1}^{n-1} \lambda_k^{OK} cov[Y(\mathbf{x}_k), Y(\mathbf{x}_\ell)] \\ &= \rho(\mathbf{x}_n - \mathbf{x}_\ell) - \mu \neq \rho(\mathbf{x}_n - \mathbf{x}_\ell) \end{aligned} \tag{21}$$

Even if the simple kriging variance is used instead of the ordinary kriging variance (Deutsch and Journel, 1992, p 142), the previous equation is unchanged and the same

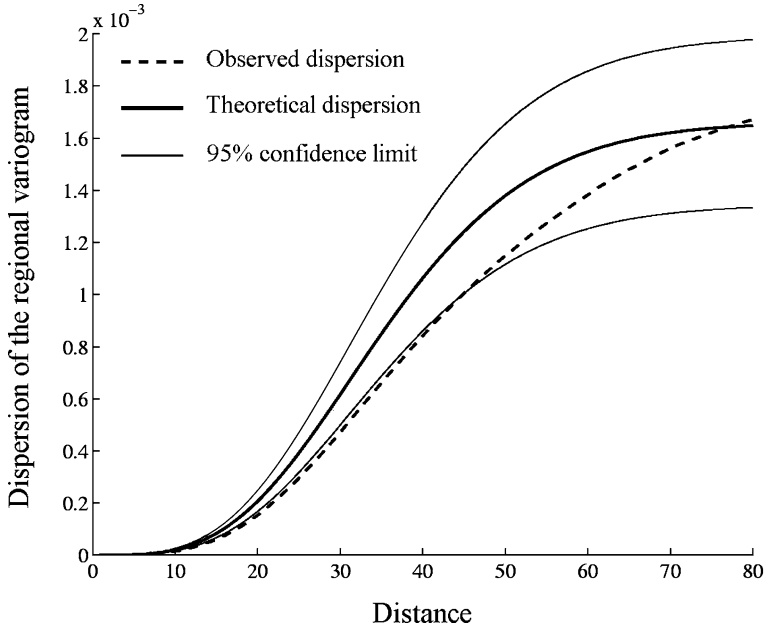


Fig. 8 Theoretical and observed dispersions of the regional variogram

conclusion prevails. By resorting to the additivity relationship (Chilès and Delfiner, 1999, p 183), one can get an alternative expression for Eq. (21) and explain what is expected to happen. If $\{\lambda_k^{SK}, k = 1, \ldots, n-1\}$ stand for the simple kriging weights of $Y(\mathbf{x}_n)$, λ_m^{SK} for the weight of the mean m and m^{OK} for the optimal estimation of this mean, then:

$$\begin{aligned} cov[Y(\mathbf{x}_n), Y(\mathbf{x}_\ell)] &= \sum_{k=1}^{n-1} \lambda_k^{SK} cov[Y(\mathbf{x}_k), Y(\mathbf{x}_\ell)] \\ &\quad + \lambda_m^{SK} cov[m^{OK}, Y(\mathbf{x}_\ell)] \\ &= \rho(\mathbf{x}_n - \mathbf{x}_\ell) + \lambda_m^{SK} \mathrm{var}(m^{OK}) \end{aligned} \tag{22}$$

hence (Eq. 21)

$$\mu = -\lambda_m^{SK} var(m^{OK}) \tag{23}$$

The last line of Eq. (22) stems from the kriging systems associated with the simple kriging of $Y(\mathbf{x}_n)$ and the ordinary kriging of m. Quite often, the weight of the mean is positive (in simple kriging, the mean compensates a lack of information when the data are not numerous or are distant, thus it is positively weighted), hence the Lagrange multiplier in Eqs. (20) and (21) is negative and the covariance between the simulated value and the former values is greater than the theoretical model:

$$\text{in general, } cov\,[Y(\mathbf{x}_n), Y(\mathbf{x}_\ell)] > \rho(\mathbf{x}_n - \mathbf{x}_\ell) \tag{24}$$

At the subsequent steps of the simulation, one will obtain an equation similar to Eq. (22) with, this time, $cov[Y(\mathbf{x}_k), Y(\mathbf{x}_\ell)] \geq \rho(\mathbf{x}_k - \mathbf{x}_\ell)$ (Eq. 24). Therefore, if the simple kriging weights are positive (which is a frequent situation), the covariance of the simulated value with the former ones will still be overestimated. We even see that the deviation between the theoretical and simulated covariances may grow up when the simulation proceeds, because the simulated values are re-used for the subsequent steps.

The variance is not reproduced either. Indeed, because of Eqs. (19) and (20), one has:

$$\begin{aligned} var[Y(\mathbf{x}_n)] &= \sum_{k,\ell=1}^{n-1} \lambda_k^{OK} \lambda_\ell^{OK} \rho(\mathbf{x}_k - \mathbf{x}_\ell) + \sigma_{OK}^2(\mathbf{x}_n) \\ &= \sum_{\ell=1}^{n-1} \lambda_\ell^{OK} [\rho(\mathbf{x}_n - \mathbf{x}_\ell) - \mu] + 1 \\ &\quad - \sum_{k=1}^{n-1} \lambda_k^{OK} \rho(\mathbf{x}_n - \mathbf{x}_k) - \mu \\ &= 1 - 2\mu \end{aligned} \tag{25}$$

and this quantity is usually greater than one (as previously mentioned, μ is often negative). In words, the simulated values are expected to overstate the prior unit variance. As an illustration, 200 realizations have been drawn in a domain of size 1024×1024 where a single datum (set to zero and located at the center of the domain) is available, with the previous variogram model (nugget + spherical with range 50) and neighborhood (20 conditioning values). Most of the simulated variograms (Fig. 9) cross the unit sill and even seem to be unbounded, as if the simulated random function had a greater (or even an infinite) variance. Again, no simple procedure can correct this situation.

5.2 Intrinsic model

The use of an ordinary kriging is relevant in case of an intrinsic random function, i.e. when the variogram $\gamma(\mathbf{h})$ does not reach a sill. Indeed, assume that a set of data $\{Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_{n-1})\}$ satisfy the intrinsic hypothesis:

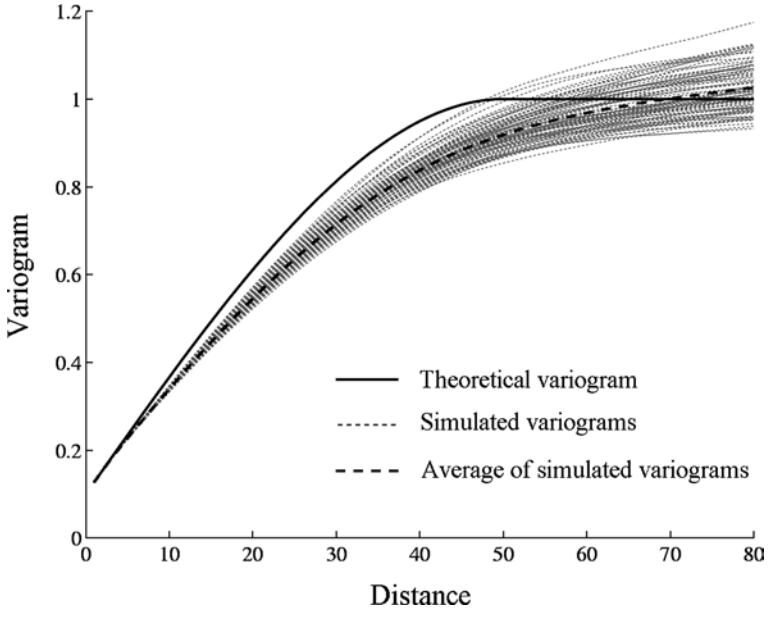


Fig. 9 Simulated variograms obtained by using an ordinary kriging

$$\forall i,j \in \{1,\dots n-1\}, \begin{cases} E[Y(\mathbf{x}_i) - Y(\mathbf{x}_j)] = 0 \\ var[Y(\mathbf{x}_i) - Y(\mathbf{x}_j)] = 2\gamma(\mathbf{x}_i - \mathbf{x}_j) \end{cases} \tag{26}$$

The increments of the random function fulfill the second-order stationarity hypothesis; their covariances can be expressed in terms of the variogram model. Indeed, let us put:

$$\begin{aligned} &\forall i,j,k,\ell \in \{1,\dots n-1\}, \\ &C_{ik,j\ell} = cov[Y(\mathbf{x}_i) - Y(\mathbf{x}_k), Y(\mathbf{x}_j) - Y(\mathbf{x}_\ell)] \end{aligned} \tag{27}$$

Using the bi-linearity of the covariance operator, it comes:

$$\begin{aligned} 2C_{ik,j\ell} &= C_{ik,ji} + C_{ik,jk} + C_{ik,i\ell} + C_{ik,k\ell} \\ &= (C_{ij,ji} + C_{jk,ji}) + C_{jk,ik} + C_{i\ell,ik} + (C_{i\ell,k\ell} + C_{\ell k,k\ell}) \\ &= -C_{ji,ji} + C_{jk,jk} + C_{i\ell,i\ell} - C_{k\ell,k\ell} \\ &= \gamma(\mathbf{x}_j - \mathbf{x}_k) + \gamma(\mathbf{x}_i - \mathbf{x}_\ell) - \gamma(\mathbf{x}_i - \mathbf{x}_j) - \gamma(\mathbf{x}_\ell - \mathbf{x}_k) \end{aligned} \tag{28}$$

In brief, the knowledge of the variogram (variance of the increments) is enough to characterize the covariance between increments. As a consequence, the variance of any linear combination whose weights add to zero can be developed formally by substituting the opposite of the variogram for the non-existent covariance (Chilès and Delfiner, 1999, p 61): if $\{\omega_i, i = 1,\dots,n-1\}$ is a set of weights that add to zero, then

$$\begin{aligned} var\left[\sum_{i=1}^{n-1} \omega_i Y(\mathbf{x}_i)\right] &= var\left\{\sum_{i=1}^{n-1} \omega_i [Y(\mathbf{x}_i) - Y(\mathbf{x}_1)]\right\} \\ &= \sum_{i,j=1}^{n-1} \omega_i \omega_j C_{i1,j1} \\ &= -\sum_{i,j=1}^{n-1} \omega_i \omega_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \end{aligned} \tag{29}$$

The kriging system of a new value $Y(\mathbf{x}_n)$ (Eq. 20) becomes:

$$\begin{cases} \sum_{k=1}^{n-1} \lambda_k^{OK} \gamma(\mathbf{x}_k - \mathbf{x}_\ell) - \mu = \gamma(\mathbf{x}_n - \mathbf{x}_\ell) \forall \ell \in \{1,\dots,n-1\} \\ \sum_{k=1}^{n-1} \lambda_k^{OK} = 1 \\ \sigma_{OK}^2(\mathbf{x}_n) = \sum_{k=1}^{n-1} \lambda_k^{OK} \gamma(\mathbf{x}_n - \mathbf{x}_k) - \mu \end{cases} \tag{30}$$

In the sequential algorithm, Eq. (19) is unchanged. Let us calculate the expectation and variance of the difference between $Y(\mathbf{x}_n)$ and the data values, using Eqs. (26), (29) and (30). For any $\ell \in \{1,\dots,n-1\}$, we have:

$$E[Y(\mathbf{x}_n) - Y(\mathbf{x}_\ell)] = \sum_{k=1}^{n-1} \lambda_k^{OK} E[Y(\mathbf{x}_k) - Y(\mathbf{x}_\ell)] = 0 \tag{31}$$

$$\begin{aligned} &var[Y(\mathbf{x}_n) - Y(\mathbf{x}_\ell)] \\ &= 2\sum_{k=1}^{n-1} \lambda_k^{OK} \gamma(\mathbf{x}_k - \mathbf{x}_\ell) - \sum_{k,p=1}^{n-1} \lambda_k^{OK} \lambda_p^{OK} \gamma(\mathbf{x}_k - \mathbf{x}_p) + \sigma_{OK}^2(\mathbf{x}_n) \\ &= 2[\gamma(\mathbf{x}_n - \mathbf{x}_\ell) + \mu] - \sum_{k=1}^{n-1} \lambda_k^{OK} [\gamma(\mathbf{x}_n - \mathbf{x}_k) + \mu] \\ &\quad + \left[\sum_{k=1}^{n-1} \lambda_k^{OK} \gamma(\mathbf{x}_n - \mathbf{x}_k) - \mu\right] \\ &= 2\gamma(\mathbf{x}_n - \mathbf{x}_\ell) \end{aligned} \tag{32}$$

Hence, the whole set $\{Y(\mathbf{x}_1),\dots, Y(\mathbf{x}_{n-1}), Y(\mathbf{x}_n)\}$ still fulfills the intrinsic hypothesis (Eqs. 26 to 28).

Remarks

1. If the random variables $\{U_1,\dots, U_n\}$ in Eq. (19) are Gaussian, the spatial distribution is fully characterized by the variogram model and is independent of the ordering of the locations $\{\mathbf{x}_1,\dots, \mathbf{x}_n\}$: the sequential algorithm simulates an intrinsic random function with Gaussian and stationary increments.
2. One original value at least is required at the beginning of the simulation to solve the ordinary kriging system. This was expectable, since an intrinsic random function is defined up to an additive constant: the initial value is needed to remove this indetermination.
3. Similarly, the sequential algorithm allows simulating a generalized intrinsic random function by using an intrinsic kriging instead of an ordinary kriging and a generalized covariance instead of a variogram (Chilès and Delfiner, 1999, p 252–265).
4. Like in the stationary case, the exactness of the sequential algorithm is true only if a unique neighborhood is used in the kriging steps. In general, the moving neighborhood restriction entails an inaccurate reproduction of the variogram. Let us come back to the configuration shown in Fig. 3, where the first value $Y(\mathbf{x}_1)$ is an initial datum; the simulated values can be written as follows:

$$\begin{cases} Y(\mathbf{x}_2) = Y(\mathbf{x}_1) + \sqrt{2\gamma(\mathbf{x}_2 - \mathbf{x}_1)} U_2 \\ Y(\mathbf{x}_3) = Y(\mathbf{x}_2) + \sqrt{2\,\gamma(\mathbf{x}_3 - \mathbf{x}_2)} U_3 \end{cases} \tag{33}$$

hence

$$\begin{aligned} var[Y(\mathbf{x}_3) - Y(\mathbf{x}_1)] &= 2\gamma(\mathbf{x}_3 - \mathbf{x}_2) + 2\gamma(\mathbf{x}_2 - \mathbf{x}_1) \\ &\neq 2\gamma(\mathbf{x}_3 - \mathbf{x}_1) \end{aligned} \tag{34}$$

The only exception is the linear variogram in a one-dimensional space. Due to the screening effect produced by this variogram, the simulation is exact as soon as the adjacent values on each side of the node being simulated are taken into account in the kriging system. The random function model is a Brownian motion (Chilès and Delfiner, 1999, p 507–510). This

assertion is no longer correct in a multidimensional space (Eq. 34), as mentioned by Chilès (1995, p 100).

6 Conclusions

Due to the moving neighborhood restriction, the sequential Gaussian algorithm fails to reproduce several basic properties of the underlying model:

- the second-order statistics such as the variogram or the indicator variograms may be biased, especially when they have a finite range;
- the realizations are not consistent with a stationary and ergodic random function model.

The bias and inconsistencies are not negligible: the reason is that each simulated value is used in the subsequent steps, so the errors tend to spread when the simulation proceeds. To overcome these impediments, the multiple-grid concept is *not* a panacea, as it only postpones the problems until the simulation is refined to a small scale. A better solution consists in increasing the number of conditioning values, but this approach quickly becomes time-consuming and unworkable. Alternatively, one may define an adequate neighborhood that minimizes the error between the simulated and theoretical statistics. Now, such task is quite difficult and impractical, since this neighborhood depends on the variogram model and the geometric configuration of the previously simulated locations (Boulanger, 1990, p 29–46).

The distinction between the model and the algorithm is sometimes confused. The model is the random function that is sought after, characterized by its spatial distribution (in particular, its histogram, variogram and indicator variograms). The algorithm is a way to obtain realizations of the model; it may be traded for another one if not suitable. In the multigaussian case, let us mention the turning bands and the discrete spectral methods as two alternatives to the sequential algorithm (Matheron, 1973, p 461; Lantuéjoul, 1994; Chilès and Delfiner, 1997). To this author's opinion, such methods present three advantages over the sequential algorithm:

1. The conditioning to a set of hard data can be performed through a kriging step that only uses the original data, not the simulated values. Hence, the kriging system is far smaller and the CPU requirements substantially decrease with respect to the sequential paradigm.
2. If the variogram is smooth near the origin and a fine-scale simulation is sought after, the quasi-singularity of the kriging matrix is avoided, unless the original data are close together. Indeed, the conditioning kriging does not involve the simulated locations, but only the data locations (recall the discussion in Sect. 2.3).
3. Since the kriging configurations are unchanged from one realization to another, the conditioning kriging needs to be performed only once for all the realizations to reproduce the original data, whereas the sequential algorithm has to achieve one kriging per realization (unless the same visiting sequence is used, which is commonly not advised). Again, this entails an important gain in terms of CPU requirements.

Acknowledgements The author would like to acknowledge the reviewers for their helpful comments and the sponsoring by Codelco-Chile for supporting this research.

References

Alfaro M (1979) Etude de la robustesse des simulations de fonctions aléatoires, Doctoral thesis, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, p 161

Boulanger F (1990) Modélisation et simulation de variables régionalisées par des fonctions aléatoires stables, Doctoral thesis, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, p 385

Chilès JP (1995) Quelques méthodes de simulation de fonctions aléatoires intrinsèques, Cahiers de Géostatistique Fascicule 5, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, pp 97–112

Chilès JP, Delfiner P (1997) Discrete exact simulation by the Fourier method. In: Baafi EY, Schofield NA (eds.) Geostatistics Wollongong'96. Kluwer Academic, Dordrecht, pp 258–269

Chilès JP, Delfiner P (1999) Geostatistics: Modeling spatial uncertainty. Wiley, New York, p 696

Deutsch CV, Journel AG (1992) GSLIB: Geostatistical software library and user's guide. Oxford University Press, New York, p 340

Gómez–Hernández J, Journel AG (1993) Joint sequential simulation of multigaussian fields. In: Soares A (ed,) Geostatistics Tróia'92. Kluwer Academic, Dordrecht, p 85–94

Gordin MI (1969) The central limit theorem for stationary processes. Soviet Mathematics Doklady 10: 1174–1176

Lantuéjoul C (1991) Ergodicity and integral range. J. of Microscopy 161(3): 387–404

Lantuéjoul C (1994) Non-conditional simulation of stationary isotropic multigaussian random functions. In: Armstrong M, Dowd PA (eds,) Geostatistical simulations. Kluwer Academic, Dordrecht, p 147–177

Lantuéjoul C (2002) Geostatistical simulation: models and algorithms. Springer-Verlag, Berlin, p 256

Matheron G (1973) The intrinsic random functions and their applications. Advances in applied probability 5: 439–468

Matheron G (1989a) Estimating and choosing: an essay on probability in practice. Springer-Verlag, Berlin, p 141

Matheron G (1989b) The internal consistency of models in geostatistics. In: Armstrong M (ed,) Geostatistics. Kluwer Academic, Dordrecht, p 21–38

Ortiz J, Deutsch CV (2002) Calculation of uncertainty in the variogram. Math Geol 34(2): 169–183

Ripley BD (1987) Stochastic simulation. Wiley, New York, p 256

Rivoirard J (1984) Le comportement des poids de krigeage, Doctoral thesis, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, p 72

Shinozuka M, Jan CM (1972) Digital simulation of random processes and its applications. J. of Sound and Vibration 25(1): 111–128

Tran T (1994) Improving variogram reproduction on dense simulation grids. Computers & Geosciences 20(7): 1161–1168

Yao T (2004) Reproduction of the mean, variance, and variogram model in spectral simulation. Math Geol 36(4): 487–506