Stoch Environ Res Risk Assess (2010) 24:211–219
DOI 10.1007/s00477-009-0311-5

ORIGINAL PAPER

Multi-Gaussian kriging and simulation in the presence of an uncertain mean value

Xavier Emery

Published online: 18 March 2009

Abstract The multi-Gaussian model is used in geostatistical applications to predict functions of a regionalized variable and to assess uncertainty by determining local (conditional to neighboring data) distributions. The model relies on the assumption that the regionalized variable can be represented by a transform of a Gaussian random field with a known mean value, which is often a strong requirement. This article presents two variations of the model to account for an uncertain mean value. In the first one, the mean of the Gaussian random field is regarded as an unknown non-random parameter. In the second model, the mean of the Gaussian field is regarded as a random variable with a very large prior variance. The properties of the proposed models are compared in the context of non-linear spatial prediction and uncertainty assessment problems. Algorithms for the conditional simulation of Gaussian random fields with an uncertain mean are also examined, and problems associated with the selection of data in a moving neighborhood are discussed.

Keywords Spatial prediction · Conditional expectation · Spatial uncertainty · Conditional distributions · Ordinary kriging · Moving neighborhood

1 Introduction and scope of the work

Determining whether a regionalized variable at a specific location exceeds or falls short of a threshold is a problem commonly met in the geosciences. Examples of applications include the delineation of areas where a pollutant concentration is greater than a threshold, the prediction of deficiencies or excesses in soil properties, or the calculation of mineral resources that can be recovered above a given cutoff.

Because it produces smoothed predictions, linear kriging is not suitable for predicting a regionalized variable in relation to a threshold value. Instead, one can use nonlinear methods (e.g. indicator, disjunctive or multi-Gaussian kriging; Bayesian maximum entropy approaches) or conditional simulation (Journel and Huijbregts 1978; Christakos 1990; Rivoirard 1994; Goovaerts 1997; Chilès and Delfiner 1999; Chilès and Lantuéjoul 2005). The implementation of these techniques usually relies on a transformation of the original data into indicator data or into normal scores and on the knowledge of the mean value of the transformed data. This last assumption is critical when the data are scarce and their true mean value is uncertain.

This work focuses on the Gaussian random field model, which is used in mineral resource evaluation (Journel 1974), oil reservoir characterization (Delfiner and Chilès 1977), stochastic hydrology (Delhomme 1979), environmental engineering (Gotway 1994) and soil science (Chilès and Allard 2005). The purpose here is the accounting for an uncertain mean value of the normal scores data. Two applications are of particular interest and will be distinguished: (1) the prediction of functions of the regionalized variable and (2) the assessment of local uncertainty through the modeling of posterior distributions, i.e., probability distributions conditioned to the information brought by surrounding data.

One goal of this paper is to provide an overview of models useful for kriging and simulation in the presence of an uncertain mean value. These models have been

X. Emery (✉)
Department of Mining Engineering, University of Chile,
Avenida Tupper 2069, Santiago, Chile
e-mail: xemery@ing.uchile.cl

presented separately and variously in earlier papers (Emery 2006a, b, 2007, 2008), reason for which several mathematical demonstrations will be left to references rather than being reprinted in this article. Other goals—and new contributions with respect to the previous papers—are: to compare and contrast the model properties in view of both prediction and uncertainty modeling; to decide to which application each model is better suited; and to present implementation issues and guidelines for good practice, in particular regarding the use of moving neighborhoods in conditional simulation.

The outline of the paper is as follows. In the next section, we briefly recall the traditional Gaussian random field model (with a known mean) and two variations that account for an uncertain mean value. Then we discuss the weaknesses and strengths of each model in the context of spatial prediction and uncertainty assessment problems. In the last section, we examine algorithms that can be used for conditional simulation and highlight some problems related to the selection of data in a moving neighborhood.

2 Three variations of the multi-Gaussian model

2.1 Model presentation

The regionalized variable of interest is regarded as a realization of a random field $Z = \{Z_{\mathbf{x}}, \mathbf{x} \in \mathbf{D}\}$ over a given spatial domain $\mathbf{D}$. Throughout this paper, we assume that Z is the transform of a stationary Gaussian random field $Y = \{Y_{\mathbf{x}}, \mathbf{x} \in \mathbf{D}\}$, i.e., a random field whose finite-dimensional distributions are multivariate normal (or *multi-Gaussian*) and invariant under a translation in space:

$$\forall \mathbf{x} \in \mathbf{D}, \quad Z_{\mathbf{x}} = \phi(Y_{\mathbf{x}}), \tag{1}$$

where ϕ is a non-decreasing function called *anamorphosis* function. In the following, we will denote by C_Y and γ_Y the covariance function and semi-variogram of Y, respectively. Because of the stationarity assumption, these are functions of the lag separation vector.

In the following, three models will be considered:

(1) *Model with a known mean*

$$\forall \mathbf{x} \in \mathbf{D}, \quad Y_{\mathbf{x}} = m + U_{\mathbf{x}}, \tag{2}$$

where $U = \{U_{\mathbf{x}}, \mathbf{x} \in \mathbf{D}\}$ is a stationary, ergodic, standard Gaussian random field (with zero mean and unit variance), and m a known scalar parameter.

(2) *Model with an unknown mean*

Same model as in Eq. 2, except that m is unknown. Actually, this model does not define a single random field for Y, but an equivalence class of random fields, namely the class of stationary Gaussian random fields with covariance function C_Y.

(3) *Model with a random mean*

$$\forall \mathbf{x} \in \mathbf{D}, \quad Y_{\mathbf{x}} = M + U_{\mathbf{x}}, \tag{3}$$

where M is a normal random variable (constant over $\mathbf{D}$) with a variance much greater than 1, independent of U. This model can be seen as a particular case of the Bayesian models used in the scope of linear kriging prediction, in which some prior knowledge about the random variable M is assumed (Omre 1987; Omre et al. 1989; Pilz 1994). Here, the prior distribution of M is considered to have a very large variance, i.e., this distribution is non-informative and there is a complete prior uncertainty on the outcome of M (Emery 2007).

2.2 Guidelines for parameter inference

In the *model with a known mean*, the anamorphosis function ϕ and semi-variogram γ_Y can be determined through the following steps:

(1) Transform the original data (random field Z) into normal scores data (random field Y). Because the model associated with Eqs. 1, 2 admits more than one parametrization, a way to make it single is to put $m = 0$, i.e., to transform the original data into standard normal values. The anamorphosis function can then be modeled by means of a transformation table (Goovaerts 1997), an expansion into Hermite polynomials (Rivoirard 1994; Chilès and Delfiner 1999), or a Box–Cox transformation (Box and Cox 1964), to name the most common options.
(2) Calculate the sample semi-variogram of the normal scores data, and fit a model γ_Y.
(3) Validate the multi-Gaussian assumption, by examining lagged scatterplots, indicator variograms or variograms of different orders (Rivoirard 1994; Goovaerts 1997; Emery 2005).

The same steps can be followed in the *model with an unknown mean*, by temporarily assuming that the mean is zero in order to determine the anamorphosis function ϕ (step (1)), then by renouncing to the zero-mean assumption for subsequent work. Equivalently, one assumes that the original random field Z belongs to the class of random fields of the form $\phi(Y)$, where ϕ is determined at step (1) and Y is a Gaussian random field whose mean value can be any real number. Such a procedure is debatable, although it is formally identical to that realized in the context of ordinary lognormal kriging (Journel 1980), which corresponds to the case when the anamorphosis is assumed to be an exponential function. The idea behind

this procedure is that the "true" unknown mean m should be close to 0 (experimental mean of the normal scores data), so that the "true" anamorphosis function should not be so different from that determined at step (1). Furthermore, the predictors that will be constructed in the scope of this model are expected to be robust against any local fluctuation of the true mean (case of local or quasi stationarity).

Finally, in what refers to parameter inference, the *model with a random mean* (Eq. 3) is not different from the model with an unknown mean. Indeed, from a practical viewpoint, a single realization of Y is available at the data locations, hence the models with an unknown mean and with a random mean constitute two equivalent representations of the same reality. Put another way, for the realization under consideration, one has $M = m$ (unknown value), so that the above described inference procedure can be used.

In summary, the steps for inferring and validating the model parameters are the same in the three proposed models. The only difference between the models is the *decision* to finally consider the mean value of the normal scores data as known (first model) or not (last two models), while the anamorphosis function and normal scores semivariogram are assumed known in every case.

3 Comparison of models and discussion

3.1 Spatial prediction problems

We are interested in predicting a function of $Z_{\mathbf{x}}$, which is also a function of $Y_{\mathbf{x}}$, say $\varphi(Y_{\mathbf{x}})$, given a set of data at locations surrounding $\mathbf{x}$. In the following, let us denote:

- $Y_{\mathbf{x}}^{\mathrm{SK}}$ and $Y_{\mathbf{x}}^{\mathrm{OK}}$ the simple and ordinary kriging predictors of $Y_{\mathbf{x}}$
- $\mu_{\mathbf{x}}^{\mathrm{OK}}$ the Lagrange multiplier used in the ordinary kriging system
- $(\sigma_{\mathbf{x}}^{\mathrm{SK}})^2$ and $(\sigma_{\mathbf{x}}^{\mathrm{OK}})^2$ the simple and ordinary kriging variances
- g and G the standard normal probability density and cumulative distribution function.

In each of the three models, a multi-Gaussian kriging predictor of $\varphi(Y_{\mathbf{x}})$ can be designed.

(1) Model with a known mean (Rivoirard 1994)

$$[\varphi(Y_{\mathbf{x}})]_1^* = \int_{-\infty}^{+\infty} \varphi(Y_{\mathbf{x}}^{\mathrm{SK}} + \sigma_{\mathbf{x}}^{\mathrm{SK}} t)\, g(t)\, \mathrm{d}t. \quad (4)$$

(2) Model with an unknown mean (Emery 2006a, b)

$$[\varphi(Y_{\mathbf{x}})]_2^* = \int_{-\infty}^{+\infty} \varphi(Y_{\mathbf{x}}^{\mathrm{OK}} + \sqrt{(\sigma_{\mathbf{x}}^{\mathrm{OK}})^2 + 2\mu_{\mathbf{x}}^{\mathrm{OK}}}\, t)\, g(t)\, \mathrm{d}t. \quad (5)$$

Note that this predictor coincides with the well-known ordinary lognormal kriging predictor (Journel 1980) when φ is an exponential function.

(3) Model with a random mean (Emery 2008)

$$[\varphi(Y_{\mathbf{x}})]_3^* = \int_{-\infty}^{+\infty} \varphi(Y_{\mathbf{x}}^{\mathrm{OK}} + \sigma_{\mathbf{x}}^{\mathrm{OK}} t)\, g(t)\, \mathrm{d}t. \quad (6)$$

The model with a known mean is often criticized because the associated predictor (Eq. 4) is inaccurate if the assumed mean departs from the true mean. This situation is likely to occur when the mean value can be considered constant only at a local scale (quasi stationarity) and/or when it is poorly estimated because of data scarcity (Guibal and Remacre 1984; Rivoirard 1994).

In contrast, in the two models with uncertain means, the predictors (Eqs. 5, 6) do no longer rely on the mean value of the Gaussian random field. They are therefore robust to variations of this mean value in space, provided that it remains constant at the scale of the kriging neighborhood. At this stage, one may wonder which of the two models should be used for prediction. Actually, both models are theoretically sound and are therefore valid options, although there are several reasons for preferring the model with an unknown mean, as explained hereunder.

One reason is that the predictor in Eq. 5 is established without any assumption on the mean, while the predictor in Eq. 6 is based on stronger assumptions (namely, that M has a normal distribution with a very large variance and is independent of the residual random field U).

The other reason relates to the practical significance of the notion of unbiasedness. By construction, the predictor associated with the unknown mean model (Eq. 5) is unbiased, therefore (because the random field model is ergodic) the spatial average of predicted values is expected to be close to that of the unknowns. In contrast, in the random mean model (Eq. 3), ergodicity does not hold any more, so that prediction errors may not cancel in space, although the predictor in Eq. 6 (the conditional expectation of $\varphi(Y_{\mathbf{x}})$) is conditionally unbiased.

For example, suppose that the kriging neighborhood contains only one datum at location $\mathbf{x}_1$. In the model with an unknown mean, the predictor is:

$$[\varphi(Y_{\mathbf{x}})]_2^* = \varphi(Y_{\mathbf{x}_1}), \tag{7}$$

while the random mean model yields:

$$[\varphi(Y_{\mathbf{x}})]_3^* = \int_{-\infty}^{+\infty} \varphi\left(Y_{\mathbf{x}_1} + \sqrt{2\gamma_Y(\mathbf{x}_1 - \mathbf{x})}t\right) g(t)\,\mathrm{d}t. \tag{8}$$

In particular, if φ is a convex function (e.g., an exponential function), then the predictor in Eq. 8 is greater than that in Eq. 7:

$$\begin{aligned}[\varphi(Y_{\mathbf{x}})]_3^* &\geq \varphi\left(\int_{-\infty}^{+\infty} (Y_{\mathbf{x}_1} + \sqrt{2\gamma_Y(\mathbf{x}_1 - \mathbf{x})}t)\, g(t)\,\mathrm{d}t\right) \\ &= \varphi(Y_{\mathbf{x}_1}),\end{aligned} \tag{9}$$

whereas the opposite inequality holds if φ is a concave function. These inequalities lead to a curious situation: although one has designed two unbiased predictors, one of them is systematically greater than the other one. This proves that the notion of "unbiasedness" is not absolute, but relative to a model.

The only practical requirement for calculating the predictor associated with the unknown mean model (Eq. 5) is the non-negativity of the term under the square root, which is not automatically fulfilled (Table 1). The occurrence of a negative term can however be avoided by modifying the ordinary kriging system so as to constrain the weights to be nonnegative (Barnes and Johnson 1984; Herzfeld 1989).

3.2 Uncertainty modeling

Another problem of interest in applications is to quantify local uncertainty by determining probability intervals for the unknown values. To estimate the probability that the actual value of $Y_{\mathbf{x}}$ falls short of a specific threshold y, it suffices to use the previous predictors (Eqs. 4–6), with φ as the indicator function at threshold y. It comes:

(1) Model with a known mean

$$[F_{\mathbf{x}}(y)]_1^* = G\left(\frac{y - Y_{\mathbf{x}}^{\mathrm{SK}}}{\sigma_{\mathbf{x}}^{\mathrm{SK}}}\right). \tag{10}$$

(2) Model with an unknown mean

$$[F_{\mathbf{x}}(y)]_2^* = G\left(\frac{y - Y_{\mathbf{x}}^{\mathrm{OK}}}{\sqrt{(\sigma_{\mathbf{x}}^{\mathrm{OK}})^2 + 2\mu_{\mathbf{x}}^{\mathrm{OK}}}}\right). \tag{11}$$

(3) Model with a random mean

$$[F_{\mathbf{x}}(y)]_3^* = G\left(\frac{y - Y_{\mathbf{x}}^{\mathrm{OK}}}{\sigma_{\mathbf{x}}^{\mathrm{OK}}}\right). \tag{12}$$

By varying the value of y, one obtains three conditional distribution functions: $[F_{\mathbf{x}}]_1^*$, $[F_{\mathbf{x}}]_2^*$ and $[F_{\mathbf{x}}]_3^*$. Note that, in the model with a deterministic mean (Eq. 2), the *true* conditional distribution is given by $[F_{\mathbf{x}}]_1^*$ (irrespective of whether the mean is known or not), whereas in the model with a random mean, it is given by $[F_{\mathbf{x}}]_3^*$. In contrast, $[F_{\mathbf{x}}]_2^*$ is only an *estimate* of the conditional distribution of $Y_{\mathbf{x}}$ and, as such, may not yield accurate measures of uncertainty. To corroborate this assertion, consider the following three examples:

(1) If the kriging neighborhood contains only one datum, then $[F_{\mathbf{x}}]_2^*$ is a step function: the variance of the estimated conditional distribution is zero, but the variance of the true conditional distribution $[F_{\mathbf{x}}]_1^*$ (simple kriging variance) is strictly positive at locations without data.

Table 1 Properties of predictors associated with proposed models

	Model with a known mean	Model with an unknown mean	Model with a random mean
Exact interpolation at data locations?	Yes	Yes	Yes
Predictor is globally unbiased?	Yes	Yes	?
Predictor is conditionally unbiased?	Yes	No	Yes
Prediction errors are expected to cancel in space?	Yes	Yes	No
Predictor is robust to local variations of the mean value?	No	Yes	Yes
Requirements for practical implementation	Stationarity	Quasi stationarity	Quasi stationarity
	Multivariate normal distributions	Multivariate normal distributions	Multivariate normal distributions
		At least one datum in the kriging neighborhood	At least one datum in the kriging neighborhood
		Variance of ordinary kriging predictor not greater than 1	

(2) If $Y_{\mathbf{x}}$ has a pure nugget semi-variogram with sill 1, then the variance of $[F_{\mathbf{x}}]_2^*$ at a location without data is equal to $1 - 1/n$, where n is the number of data. This variance is less than the unit variance of the true conditional distribution, which here coincides with the prior distribution.

(3) If location $\mathbf{x}$ is located far from the data (beyond the range of the semi-variogram), then the variance of the true conditional distribution is equal to the prior variance (i.e., 1), while that of the estimated distribution is $1 - (\sigma_m^{\mathrm{OK}})^2$, where $(\sigma_m^{\mathrm{OK}})^2$ is the variance of the ordinary kriging estimator of m.

In conclusion, although $[F_{\mathbf{x}}]_2^*$ is an unbiased estimate of $[F_{\mathbf{x}}]_1^*$, it usually has a too small variance and does not reflect the actual uncertainty in the outcome of $Y_{\mathbf{x}}$. The deterministic mean model (Eq. 2) therefore turns out to be inappropriate for deriving conditional distributions when the mean value is unknown. In such a case, the true conditional distribution of $Y_{\mathbf{x}}$ (Eq. 10) is inaccessible. One could think of replacing the simple kriging predictor in Eq. 10 by the ordinary kriging predictor, leaving unchanged the simple kriging variance, but this approach is not theoretically sound and may still yield biased measures of uncertainty.

In contrast, the random mean model provides true conditional distributions and uncertainty measures. In particular, the conditional variance of $Y_{\mathbf{x}}$ is its ordinary kriging variance (Eq. 12) and is always greater than the simple kriging variance, a result that is intuitive insofar as not knowing the mean value should increase the uncertainty in the outcome of $Y_{\mathbf{x}}$. With respect to the traditional approach (model with a known mean, Eq. 10), one just has to replace simple kriging by ordinary kriging when determining the conditional distributions (Table 2).

4 Conditional simulation

4.1 Sequential Gaussian simulation

The previous results about uncertainty modeling can be used for simulating the Gaussian random field Y at a set of locations without data. One option is the sequential algorithm, in which the variable at each location is simulated according to its distribution conditioned to the original data and to the previously simulated variables. Now, although this approach is theoretically correct, it becomes impractical when the number of variables to simulate is large (say, more than a few thousands). To reduce the computational burden, the conditioning data must be limited to that located in a moving neighborhood.

4.1.1 Biases caused by moving neighborhood restrictions

To understand the impact caused by omitting data in the conditioning process, we will assume that there exist $n + k$ data at locations $\{\mathbf{x}_1,\ldots,\mathbf{x}_{n+k}\}$, but that the random variable $Y_{\mathbf{x}}$ is simulated conditionally to the data located at $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$.

Ideally, the simulated variable for $Y_{\mathbf{x}}$ should be

$$(Y_{\mathbf{x}})_{\mathrm{ideal}} = Y_{\mathbf{x}}^{\mathrm{OK}}\big|_{n+k\,\mathrm{data}} + \sigma_{\mathbf{x}}^{\mathrm{OK}}\big|_{n+k\,\mathrm{data}} T \tag{13}$$

with T a standard normal random variable independent of the data. However, because of neighborhood restrictions, the actually simulated variable is:

$$(Y_{\mathbf{x}})_{\mathrm{actual}} = Y_{\mathbf{x}}^{\mathrm{OK}}\big|_{n\,\mathrm{data}} + \sigma_{\mathbf{x}}^{\mathrm{OK}}\big|_{n\,\mathrm{data}} T \tag{14}$$

By denoting by $\lambda_{\alpha|n+k}(\mathbf{x})$ the ordinary kriging weight assigned to $\mathbf{x}_\alpha$ when kriging $Y_{\mathbf{x}}$ from the $n + k$ data, one has (Emery 2009):

$$Y_{\mathbf{x}}^{\mathrm{OK}}\big|_{n\,\mathrm{data}} = Y_{\mathbf{x}}^{\mathrm{OK}}\big|_{n+k\,\mathrm{data}} + \sum_{\alpha=n+1}^{n+k} \lambda_{\alpha|n+k}(\mathbf{x}) \left(Y_{\mathbf{x}_\alpha}^{\mathrm{OK}}\big|_{n\,\mathrm{data}} - Y_{\mathbf{x}_\alpha} \right) \tag{15}$$

$$\left(\sigma_{\mathbf{x}}^{\mathrm{OK}}\big|_{n\,\mathrm{data}}\right)^2 = \left(\sigma_{\mathbf{x}}^{\mathrm{OK}}\big|_{n+k\,\mathrm{data}}\right)^2 + \sum_{\alpha=n+1}^{n+k} \lambda_{\alpha|n+k}^2(\mathbf{x}) \left(\sigma_{\mathbf{x}_\alpha}^{\mathrm{OK}}\big|_{n\,\mathrm{data}}\right)^2 \tag{16}$$

Accordingly, the difference between $(Y_{\mathbf{x}})_{\mathrm{ideal}}$ and $(Y_{\mathbf{x}})_{\mathrm{actual}}$ will be greater when:

(a) the number of omitted data (k) increases (summations in Eqs. 15 and 16 contain more terms);

Table 2 Properties of conditional distributions associated with proposed models

	Model with a known mean	Model with an unknown mean	Model with a random mean
Conditional distributions account for an uncertain mean value?	No	Yes	Yes
Conditional distributions suffer from order-relation violations/inconsistencies?	No	No	No
Conditional distributions allow assessing uncertainty?	Yes	No	Yes

(b) the kriging weights that these data would receive in a unique neighborhood implementation are strongly different from 0;
(c) the omitted data cannot be predicted accurately from the remaining n data.

Point (c) is critical in the presence of a nugget effect or a short-range structure in the semi-variogram of the random field Y to be simulated, while point (a) is critical when many locations are targeted for simulation. In this respect, experience with the known mean model (using simple kriging to determine the conditional distributions) has shown that too restrictive neighborhoods may introduce biases in the reproduction of the model statistics on large grids (Gómez-Hernández and Cassiraga 1994; Tran 1994; Emery 2004).

Defining the neighborhood is even more important when using ordinary kriging, because of point (b). Indeed, the simple kriging weights assigned to the data located far from the target location are usually very small, whereas this is not necessarily the case with ordinary kriging, in particular when the mean value is assigned a large simple kriging weight (Rivoirard 1987). This difference between the behavior of simple and ordinary kriging weights is likely to be of utmost relevance in the context of sequential simulation (simulated variables are re-used as conditioning data for subsequent variables, hence errors are likely to propagate), as it will be shown hereafter on a few examples.

4.1.2 Numerical experiments

To determine the sensitivity of the neighborhood definition on the reproduction of the semi-variogram, we now present numerical results obtained by simulating a Gaussian random field with a prescribed semi-variogram over a regular grid with size 300 × 300, conditioned to one hundred data randomly located on the grid. The exercise consists in constructing 100 realizations, each one obtained via the following steps:

(1) Simulate a Gaussian random field with zero mean at the data locations by using the non-conditional version of the covariance matrix decomposition algorithm (Davis 1987). Since this simulation algorithm is perfect, no approximation is made at this stage.
(2) Simulate the Gaussian random field at the target grid nodes, conditionally to the variables obtained in the previous step, by using the sequential algorithm (with ordinary kriging to determine conditional distributions) and visiting the grid nodes according to a random path.

Because at each realization the conditioning data are generated from the non-conditional distribution (step (1)), the experimental semi-variograms of the realizations obtained at step (2) are expected to fluctuate around the prior (non-conditional) semi-variogram model.

Four models are put to the test: a pure nugget effect (Fig. 1a), a spherical (Fig. 1b), a spherical plus nugget (Fig. 1c) and a cubic model (Fig. 1d), each with a unit sill. For each model, three neighborhoods are considered, with 10, 20 and 50 conditioning data (original data and previously simulated variables) respectively. It is seen that the reproduction of the semi-variogram turns out to be poor in the presence of a nugget effect (Fig. 1). This can be explained because the nugget effect lifts the screening effect (Chilès and Delfiner 1999): the moving neighborhood omits data that would receive non-negligible weights if a unique neighborhood were used. As for the models without nugget effect, one needs at least 50 conditioning data in the kriging neighborhood to reproduce the semi-variogram fairly well.

4.1.3 Conditions for exact semi-variogram reproduction

To find out the semi-variogram models that can be reproduced exactly when using a moving neighborhood, let us go back to Eqs. 15 and 16. These equations indicate that the omission of part of the data will alter the ordinary kriging prediction and variance, unless the kriging weights associated with the omitted data are equal to zero (perfect screening effect). Such a circumstance is exceptional and occurs only with the linear semi-variogram in one-dimensional spaces (Chilès and Delfiner 1999).

To corroborate this statement, consider the case when the Gaussian random field Y is known at location $\mathbf{x}_1$ and one looks for simulating Y at locations $\mathbf{x}_2$ and $\mathbf{x}_3$. Due to neighborhood restrictions, also assume that the simulation at $\mathbf{x}_3$ only relies on the variable simulated at $\mathbf{x}_2$, i.e., the original datum at $\mathbf{x}_1$ is omitted because it is located outside the neighborhood. Under these conditions, the sequential simulation algorithm yields:

$$\begin{cases} Y_{\mathbf{x}_2} = Y_{\mathbf{x}_1} + \sqrt{2\,\gamma_Y(\mathbf{x}_2 - \mathbf{x}_1)}\,T_1 \\ Y_{\mathbf{x}_3} = Y_{\mathbf{x}_2} + \sqrt{2\,\gamma_Y(\mathbf{x}_3 - \mathbf{x}_2)}\,T_2 \end{cases} \tag{17}$$

where T_1 and T_2 are independent standard normal random variables. Accordingly, the semi-variogram between $Y_{\mathbf{x}_1}$ and $Y_{\mathbf{x}_3}$ is:

$$\frac{1}{2}\mathrm{var}\{Y_{\mathbf{x}_3} - Y_{\mathbf{x}_1}\} = \gamma_Y(\mathbf{x}_2 - \mathbf{x}_1) + \gamma_Y(\mathbf{x}_3 - \mathbf{x}_2). \tag{18}$$

This is equal to $\gamma_Y(\mathbf{x}_3 - \mathbf{x}_1)$ only if $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ are aligned in this order, i.e., simulation is performed in a one-dimensional space following a regular sequence, and γ_Y is a

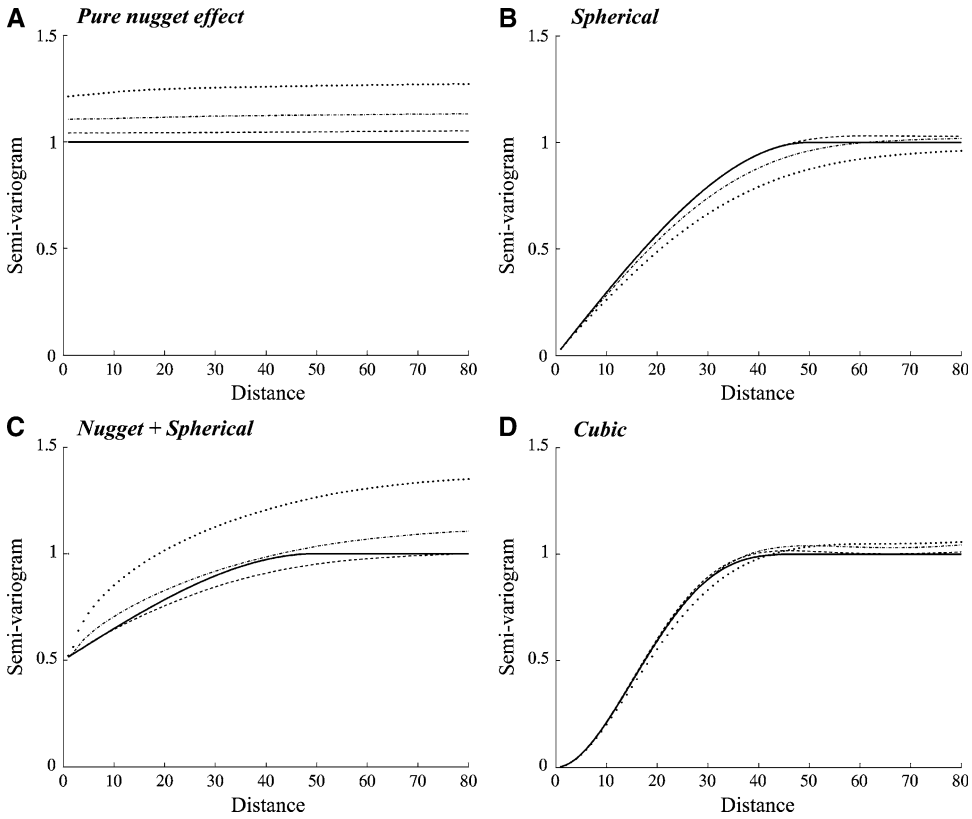

Fig. 1 Semi-variogram model (*solid line*) and average regional semi-variograms of 100 sequential Gaussian realizations obtained by using a neighborhood with 10 conditioning data (*dots*), 20 conditioning data (*dash-dots*) and 50 conditioning data (*dashed lines*). In each case, there is a total of 100 original conditioning data, and ordinary kriging is used for constructing conditional distributions

linear semi-variogram, a condition that is not compatible with the assumption of stationarity. It is concluded that no stationary semi-variogram model can be reproduced exactly with the sequential algorithm when using a moving neighborhood and ordinary kriging.

4.2 Two-step simulation

Another option for simulation is a two-step approach consisting in generating a non-conditional simulation and adding the simple kriging predictor of the "residuals" between the data and the variables simulated at the data locations (Journel and Huijbregts 1978). In the random mean model (Eq. 3), conditioning is done by ordinary kriging instead of simple kriging. Note that the non-conditional simulation is involved only through the difference with its kriging predictor and can therefore have a mean set to an arbitrary value, say 0. Any multi-Gaussian simulation algorithm can be used at this stage (discrete of continuous spectral methods, convolution methods, turning bands, or the sequential Gaussian algorithm with use of simple kriging to determine conditional distributions).

Conditioning simulations by ordinary kriging has already been suggested by Journel and Huijbregts (1978) in the scope of a stationary Gaussian field model with a non-random mean (Eq. 2). In this model however, the method is approximate since kriging errors are correlated with the conditioning data. Ordinary kriging is also used for conditioning simulations of *intrinsic* random fields with normal increments, in which case the semi-variogram is perfectly reproduced (Delfiner 1976; de Fouquet 1994). The non-ergodic model introduced in Eq. 3 provides a justification for conditioning *stationary* Gaussian random fields by ordinary kriging, and therefore offers a unified framework between stationary random fields with uncertain mean values and intrinsic random fields.

It is worthy to note that conditioning kriging only involves the original data, unlike the sequential algorithm, for which the already simulated variables have to be considered. This implies the following advantages:

- Much less calculations are required to search for nearby data and to solve the kriging system.
- Because the semi-variogram model and configuration of data locations are unchanged from one realization to another, a single kriging system has to be solved for conditioning multiple realizations.
- The number of data dropped by using a moving neighborhood instead of a unique one is significantly less than in the sequential algorithm, which decreases the biases caused by neighborhood restrictions (see Sect. 4.1.1) and facilitates the design of the kriging neighborhood, see Rivoirard (1987), Chilès and Delfiner

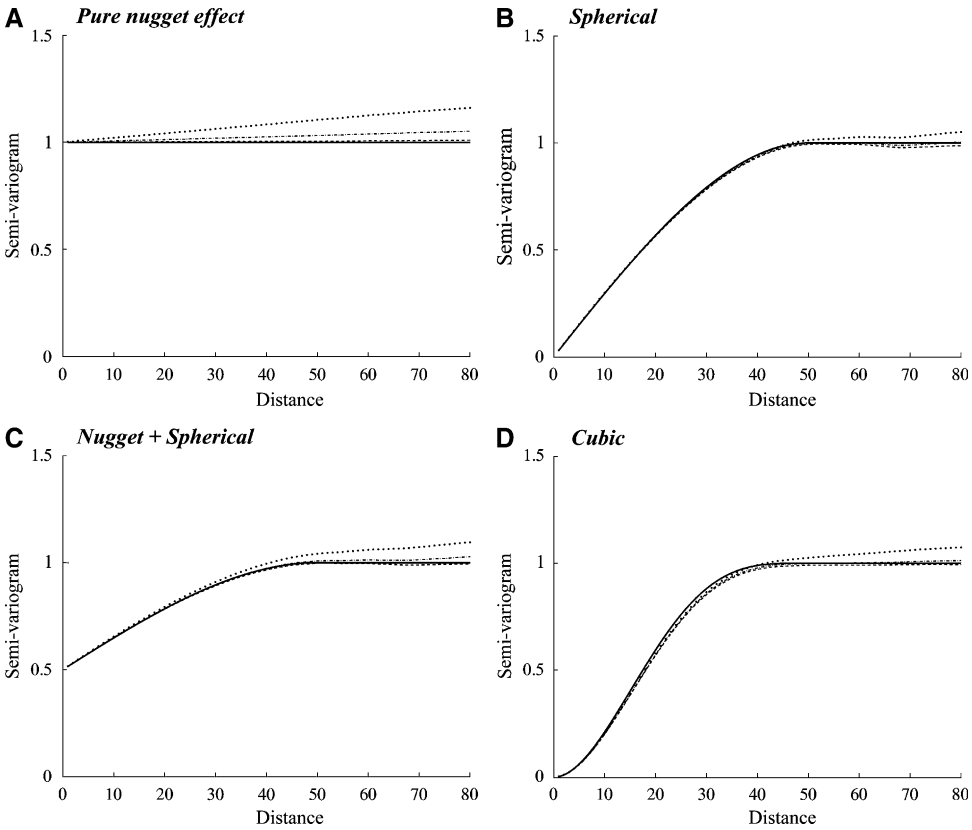

Fig. 2 Semi-variogram model (*solid line*) and average regional semi-variograms of 100 two-step realizations obtained by using a neighborhood with 10 conditioning data (*dots*), 20 conditioning data (*dash-dots*) and 50 conditioning data (*dashed lines*). In each case, there is a total of 100 original conditioning data, non-conditional realizations are constructed by recourse to the turning bands algorithm, and ordinary kriging is used for conditioning the realizations

(1999) and Emery (2009) for practical recommendations on this matter. As an illustration, the same exercise as in Sect. 4.1.2 has been made, by using the turning bands algorithms for the non-conditional simulation and considering the same numbers of neighboring data (10, 20 and 50) for conditioning kriging: the improvement in semi-variogram reproduction (Fig. 2) is notorious in comparison with the results obtained with the sequential Gaussian approach (Fig. 1), in particular for semi-variograms having a nugget effect.

5 Conclusions

So far, the multi-Gaussian model has been mainly used under the assumption of a known mean for the Gaussian random field. This paper presented two extensions of this model that account for an uncertainty in the mean value and use ordinary kriging instead of simple kriging. The first extension is suited to prediction, while the second extension is adequate for uncertainty modeling and for conditional simulation. The recourse to one or to the other extension therefore depends on the application at hand. Attention has to be paid to the definition of the kriging neighborhood, especially in the case of sequential simulation when the semi-variogram has a nugget effect.

Ordinary kriging is commonly used in linear geostatistics as it ensures robustness of the predictor to spatial variations of the mean value. It outperforms simple kriging with a locally varying mean when the local mean is poorly estimated because of data scarcity. It is hoped that its application to the multi-Gaussian model will be of interest to practitioners concerned with non-linear spatial prediction and with the characterization of spatial uncertainty.

Further extensions of the models proposed in this work can be designed. In the case of prediction or uncertainty modeling with a change of support, one can work with the joint distributions of Y at a set of locations discretizing the block under consideration. When the mean varies in space and can be represented by a polynomial drift, the models can be extended by using intrinsic kriging instead of ordinary kriging. As for multivariate problems, it suffices to substitute co-kriging for kriging.

Acknowledgments This research was funded by the Chilean Fund for Science and Technology Development (FONDECYT) and is part of project 1061103.

References

Barnes RJ, Johnson TB (1984) Positive kriging. In: Verly G, David M, Journel AG, Maréchal A (eds) Geostatistics for natural resources characterization. Reidel, Dordrecht, pp 231–244

Box GEP, Cox DR (1964) An analysis of transformations. Journal of the Royal Statistical Society B 26(2):211–252

Chilès JP, Allard D (2005) Stochastic simulation of soil variations. In: Grunwald S (ed) Environmental soil-landscape modeling: geographic information technologies and pedometrics. CRC Press, Taylor & Francis, Boca Raton, pp 289–321

Chilès JP, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York, p 695

Chilès JP, Lantuéjoul C (2005) Prediction by conditional simulation: models and algorithms. In: Bilodeau M, Meyer F, Schmitt M (eds) Space, structure and randomness. Springer, New York, pp 39–68

Christakos G (1990) A Bayesian/maximum-entropy view to the spatial estimation problem. Math Geol 22(7):763–777

Davis MW (1987) Production of conditional simulations via the LU triangular decomposition of the covariance matrix. Math Geol 19(2):91–98

de Fouquet C (1994) Reminders on the conditioning kriging. In: Armstrong M, Dowd PA (eds) Geostatistical simulations. Kluwer, Dordrecht, pp 131–145

Delfiner P (1976) Linear estimation of nonstationary spatial phenomena. In: Guarascio M, David M, Huijbregts CJ (eds) Advanced geostatistics in the mining industry. Reidel, Dordrecht, pp 49–68

Delfiner P, Chilès JP (1977) Conditional simulations: a new Monte-Carlo approach to probabilistic evaluation of hydrocarbon in place. SPE paper 6985, p 32

Delhomme JP (1979) Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach. Water Resour Res 15(2):269–280

Emery X (2004) Testing the correctness of the sequential algorithm for simulating Gaussian random fields. Stoch Environ Res Risk Assess 18(6):401–413

Emery X (2005) Variograms of order ω: a tool to validate a bivariate distribution model. Math Geol 37(2):163–181

Emery X (2006a) Multigaussian kriging for point-support estimation: incorporating constraints on the sum of the kriging weights. Stoch Environ Res Risk Assess 20(1–2):53–65

Emery X (2006b) Ordinary multigaussian kriging for mapping conditional probabilities of soil properties. Geoderma 132(1–2):75–88

Emery X (2007) Conditioning simulations of Gaussian random fields by ordinary kriging. Math Geol 39(6):607–623

Emery X (2008) Uncertainty modeling and spatial prediction by multi-Gaussian kriging: accounting for an unknown mean value. Comput Geosci 34(11):1431–1442

Emery X (2009) The kriging update equations and their application to the selection of neighboring data. Comput Geosci. doi: 10.1007/s10596-008-9116-8

Gómez-Hernández JJ, Cassiraga EF (1994) Theory and practice of sequential simulation. In: Armstrong M, Dowd PA (eds) Geostatistical simulations. Kluwer, Dordrecht, pp 111–124

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York, p 480

Gotway CA (1994) The use of conditional simulation in nuclear-waste-site performance assessment. Technometrics 36(2):129–141

Guibal D, Remacre AZ (1984) Local estimation of the recoverable reserves: comparing various methods with the reality on a porphyry copper deposit. In: Verly G, David M, Journel AG, Maréchal A (eds) Geostatistics for natural resources characterization. Reidel, Dordrecht, pp 435–448

Herzfeld UC (1989) A note on programs performing kriging with nonnegative weights. Math Geol 21(3):391–393

Journel AG (1974) Geostatistics for conditional simulation of orebodies. Econ Geol 69(5):673–687

Journel AG (1980) The lognormal approach to predicting local distributions of selective mining unit grades. Math Geol 12(4):285–303

Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London, p 600

Omre H (1987) Bayesian kriging—merging observations and qualified guesses in kriging. Math Geol 19(1):25–39

Omre H, Halvorsen KB, Berteig V (1989) A Bayesian approach to kriging. In: Armstrong M (ed) Geostatistics. Kluwer, Dordrecht, pp 109–126

Pilz J (1994) Robust Bayes linear prediction of regionalized variables. In: Dimitrakopoulos R (ed) Geostatistics for the next century. Kluwer, Dordrecht, pp 464–475

Rivoirard J (1987) Two key parameters when choosing the kriging neighborhood. Math Geol 19(8):851–856

Rivoirard J (1994) Introduction to disjunctive kriging and nonlinear geostatistics. Oxford University Press, Oxford, p 181

Tran TT (1994) Improving variogram reproduction on dense simulation grids. Comput Geosci 20(7–8):1161–1168