Chapter 29
Fifty Years of Kriging

Jean-Paul Chilès and Nicolas Desassis

Abstract Random function models and kriging constitute the core of the geostatistical methods created by Georges Matheron in the 1960s and further developed at the research center he created in 1968 at Ecole des Mines de Paris, Fontainebleau. Initially developed to avoid bias in the estimation of the average grade of mining panels delimited for their exploitation, kriging received progressively applications in all domains of natural resources evaluation and earth sciences, and more recently in completely new domains, for example, the design and analysis of computer experiments (DACE). While the basic theory of kriging is rather straightforward, its application to a large diversity of situations requires extensions of the random function models considered and sound solutions to practical problems. This chapter presents the origins of kriging as well as the development of its theory and its applications along the last fifty years. More details are given for methods presently in development to efficiently handle kriging in situations with a large number of data and a nonstationary behavior, notably the Gaussian Markov random field (GMRF) approximation and the stochastic partial differential (SPDE) approach, with a synthetic case study concerning the latter.

29.1 Introduction

The creation of the IAMG is a landmark of year 1968, which motivates the present book. Another important event of this year is the foundation of a research center of Ecole des Mines de Paris dedicated to geostatistics and mathematical morphology, two disciplines created by Georges Matheron. Concerning geostatistics, this research center was about to develop the applications of kriging, invented by Matheron several years earlier. The theory of kriging seems so straightforward that

J.-P. Chilès (✉) · N. Desassis

Centre of Geosciences, Mines ParisTech, Fontainebleau, France
e-mail: jean-paul.chiles@mines-paristech.fr; jean-paul@chiles.name

N. Desassis
e-mail: nicolas.desassis@mines-paristech.fr

it was reasonable to imagine that, after some generalizations, kriging would become a classical tool requiring no further research. On the contrary, 50 years later it remains the subject of active research, with renewed points of view. Other paradox: originating from mining estimation problems, and very close to statistical regression from a theoretical standpoint, it was not obvious that kriging would be considered in other domains than mining and earth sciences. However applications now consider, for example, the design of aircrafts (Chung and Alonso 2002), the prediction of the mechanical properties of nanomaterials (Yan et al. 2012), the optimization of supply chain networks (Dixit et al. 2016), the construction of financial term-structures (Cousin et al. 2016), the modeling of social systems (Oliveira et al. 2013), and in all cases the quantification of the uncertainty.

It is therefore not surprising to see in Table 29.1 that the number of articles on kriging (word "kriging" or "cokriging" present in the title) published by the journals of the Scopus database doubles decade after decade. The situation is slightly different for the three journals published by the IAMG: *Mathematical Geosciences* (formerly *Journal of the International Association for Mathematical Geology*, then *Mathematical Geology*), *Computers & Geosciences*, and *Natural Resources Research*; indeed, IAMG journals played a major role in the dissemination of the geostatistical literature in English in the first decades, but have now to share this role with the journals of the new application domains. (Note incidentally that few articles were published before 1980: the literature relative to kriging was largely written in French or published in monographs and conference proceedings.)

At a closer look, the originality of kriging lies in its inclusion in the geostatistical approach, where the optimality provided by kriging rests on an analysis of the spatial variability of the phenomenon of interest. Indeed, if methods for characterizing that variability were lacking, the optimality of kriging would simply be virtual. As for the persistence of research works on kriging, it is widely bound to the evolution of the capacities of calculation and memory of computers, and to the increase of the volume of the data. At its origin kriging considered some samples in the vicinity of a target block, while it has now to take into account up to thousands or even millions of data (remote sensing, laser, seismic).

This chapter first presents the origins of kriging and its theory. It continues with further developments, roughly chronologically, up to current research. Kriging has a number of variants and generalizations. We focus here on linear kriging, moreover in a monovariate context. Cokriging and disjunctive kriging are therefore not

Table 29.1 Articles whose title includes the word "kriging" or "cokriging": number of articles per decade for IAMG journals and for all journals of the Scopus database

Decade	IAMG	Scopus
1970–1979	2	14
1980–1989	63	136
1990–1999	61	272
2000–2009	53	512
2010–2016	28	1076

considered; conversely, the use of kriging to condition geostatistical simulations is acknowledged. Our aim is not a thorough presentation of kriging, which can be found in many textbooks, for example, Chilès and Delfiner (2012).

29.2 The Origins of Kriging

One of the tasks of the mining engineer is to select the panels to be exploited, and even to delimit them if the exploitation method lets him this freedom. Indeed, to simplify, a panel deserves to be exploited only if the cost of its extraction and processing does not exceed the value of the metal which can be extracted from it. For given technico-economic parameters, this means that the panel grade has to exceed some cutoff grade. In practice the true grade of a panel is not known before its exploitation, so that the selection is made on the basis of an estimated grade. At the beginning of the 1950s the estimate was simply the average grade of the data belonging to the panel or situated at its border. Krige (1951, 1952), studying exploitation data of several orebodies, observed that for high cutoffs the panels selected that way were on average less rich than expected.

As Fig. 29.1 shows it, this is not really surprising. Two parallel galleries in a sub-horizontal deposit present segments AB and CD with grades above the cutoff, contrarily to the neighboring parts of the galleries. Therefore the decision is made to exploit the trapezoid ABDC, and its grade is anticipated to be equal to the weighted average of the grades of segments AB and CD. In fact, segments AC and BD do not represent the real border between rich and poor ores. The true (unknown) limits look like the dotted lines. Therefore, poor ore is exploited (and rich ore abandoned), so that the grade of the exploited ore is lower than expected.

Mathematically, this expresses a conditional bias: Denoting Z_v the panel grade and $\bar{Z}$ the average grade of the cores situated within the panel, the conditional expectation $E[Z_v|\bar{Z}]$ is not equal to Z_v.
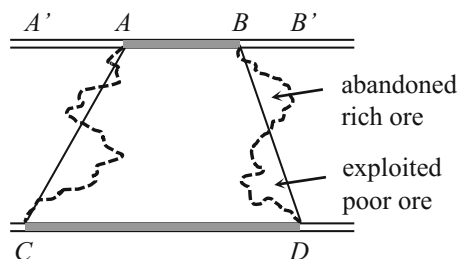

Fig. 29.1 Illustration of the estimation bias. The panel ABDC to be exploited was delimited from the rich samples observed along AB and CD. Because the true border between rich and poor ores follows a line similar to the dotted line rather than segments AC and BD, poor ore will be exploited and rich ore abandoned. (from Matheron 1961)

To avoid this bias, Krige gives a weight λ to the average grade of the data situated in the panel and the complementary weight $1 - \lambda$ to the average grade of the orebody, λ being determined by linear regression (Krige in fact considered the lognormal case and worked with grade logarithm).

Also facing problems of mining estimation, Matheron studied Krige's work and generalized his approach by assigning a proper weight to each sample, these weights being determined so as to minimize the estimation variance under the condition that the weights sum to 1 (this condition simply expresses that the estimator is a weighted average of the data).

Matheron called this method "kriging" in honor to Danie Krige. To be accurate, according to Cressie (1990), the French term "krigeage" was coined by Pierre Carlier and first used at the French Commissariat à l'énergie atomique in the late 1950s, and Matheron translated it by "kriging" in Matheron (1963b) (the first appearance of "krigeage" found by the present authors in Matheron's work is Matheron 1960, where it is mentioned as an already known concept).

29.2.1 Ordinary Kriging (OK)

Geostatistics considers natural variables distributed in space, whose behavior presents a large complexity of detail. These regionalized variables cannot be adequately represented by deterministic functions and therefore methods dedicated to random functions (RF) are considered. The theory of kriging as it is usually presented appears in Matheron (1962, 1963a). It takes place in the framework of an order-2 stationary random function (SRF) model. The regionalized variable of interest (here a grade) is considered as a realization of an SRF $Z(x)$, where x denotes a point in a two- or three-dimensional space. N data are available, at locations x_α, $\alpha = 1, 2, \ldots, N$, with values $Z_\alpha = Z(x_\alpha)$. The target Z_0 is the value $Z(x_0)$ of Z at an unobserved point x_0, or more generally the average value $Z(v)$ of Z in a given cell or block v. The kriging estimator of Z_0 is by definition of the form

$$Z^* = \sum_{\alpha=1}^{N} \lambda_\alpha Z_\alpha$$

with weights λ_α summing to 1. The weights are chosen so as to minimize the variance of the estimation error $Z^* - Z_0$ subject to the condition on their sum. This leads to a linear system of $N + 1$ equations with $N + 1$ unknowns (the N weights λ_α and a Lagrange parameter μ):

$$\begin{cases} \sum_\beta \lambda_\beta \sigma_{\alpha\beta} + \mu = \sigma_{\alpha 0} & \alpha = 1, \ldots, N \\ \sum_\beta \lambda_\beta = 1 \end{cases}$$

where $\sigma_{\alpha\beta}$ denotes the covariance of the observations Z_α and Z_β and $\sigma_{\alpha 0}$ the covariance of Z_α and the target Z_0. This is the ordinary kriging system. The ordinary kriging variance can then be expressed as:

$$\sigma^2_{OK} = E(Z^* - Z_0)^2 = \sigma_{00} - \sum_\alpha \lambda_\alpha \sigma_{\alpha 0} - \mu$$

where σ_{00} denotes the variance of Z_0.

29.2.2 Simple Kriging (SK)

Note that the kriging system and variance do not require the knowledge of the mean. If the mean m were known, we would use an estimator of the form

$$Z^* = \sum_\alpha \lambda_\alpha Z_\alpha + \left(1 - \sum_\alpha \lambda_\alpha\right) m$$

without constraint on the weights, and the minimization of the estimation variance would lead to the simple kriging system

$$\sum_\beta \lambda_\beta \sigma_{\alpha\beta} = \sigma_{\alpha 0} \quad \alpha = 1, \ldots, N$$

and to the simple kriging variance

$$\sigma^2_{SK} = E(Z^* - Z_0)^2 = \sigma_{00} - \sum_\alpha \lambda_\alpha \sigma_{\alpha 0}$$

Simple kriging receives limited applications. It is, however, important, because it has nice properties that are not shared by ordinary kriging and of course universal kriging (see Chilès and Delfiner 2012, Chap. 3). From a computational point of view, the kriging matrix being positive definite, the system can be solved by the Cholesky method.

29.2.3 Ordinary Kriging in the IRF Model

Because the mean m is not involved in ordinary kriging, it is possible to extend ordinary kriging to a more general random function model, the (order-2) intrinsic random function (IRF) model, characterized by

$$E[Z(x+h) - Z(x)] = 0$$
$$\tfrac{1}{2}E[Z(x+h) - Z(x)]^2 = \gamma(h)$$

The variogram $\gamma(h)$ summarizes the spatial variability of the random function. Geostatistics provides a set of consistent tools for choosing the variogram model adapted to a particular situation (e.g., Chilès and Delfiner 2012, Chap. 2). The above OK system and OK variance remain valid provided that $C(h)$ is formally replaced by $-\gamma(h)$ in the expressions of $\sigma_{\alpha\beta}$, $\sigma_{\alpha0}$ and σ_{00} given in the next section. This is the framework where kriging is widely used, especially in mining applications.

29.2.4 Discussion

Finally, kriging appears as nothing but (a straightforward generalization of) multiple linear regression on N data Z_α that need not to be of the form $Z(x_\alpha)$. Does it deserve a special consideration?

In fact the application of this regression requires that the covariances between the observations, and between each observation and the target, are known. They can be determined experimentally when repeated measurements are available, as is the case in meteorology, but not in usual earth sciences applications, where a unique phenomenon is considered. Applying the regression formula with a priori covariances would provide an estimator that would lose any optimality, except if by chance these covariances are perfectly suited to the data.

Kriging implies a spatial context:

- The random variables Z_α are point values of an SRF $Z(x)$ at points x_α.
- Structural analysis methods make it possible to determine the covariance function $C(h)$ of the SRF $Z(x)$.

The covariances $\sigma_{\alpha\beta}$ are then of the form $C(x_\beta - x_\alpha)$, and $\sigma_{\alpha0}$ is $C(x_0 - x_\alpha)$ if the target is $Z(x_0)$ or the average value of $C(x - x_\alpha)$ when x spans v if the target is $Z(v)$. The variance σ_{00} of Z_0 that appears in the expression of the kriging variance is $C(0)$ if the target is $Z(x_0)$ or the average value of $C(x' - x)$ when x and x' span v independently if the target is $Z(v)$.

Several authors proposed an approach similar to simple or ordinary kriging before Matheron but not in a spatial context (see Cressie 1990). The noticeable exception is Gandin (1963), who independently developed an approach similar to Matheron's one, in meteorology. SK is called *optimal interpolation*, and OK *optimal interpolation with normalization of weighting factors*. Like Matheron, Gandin was concerned by the theory and its applications; he is, for example, the first author to define and compute a variogram cloud.

29.2.5 *Analytic Calculation of Average Covariances*

In the early 1960s computers were not available, at least for mining applications. It was therefore not easy to solve linear systems of equations. Even if point (or core) data could be used to determine the variogram, kriging was applied to aggregated data. In the case of Fig. 29.1, a typical situation examined by Matheron (1961), all cores along AB are represented by their average grade Z_1, those along CD by Z_2, and those belonging to A′A and BB′ by Z_3. The target is the average grade Z_0 of the trapezoid ABDC. Kriging amounts to finding the best weights λ_1 for Z_1, λ_2 for Z_2, and $\lambda_3 = 1 - \lambda_1 - \lambda_2$ for Z_3 minimizing the variance of $\lambda_1 Z_1 + \lambda_2 Z_2 + (1 - \lambda_1 - \lambda_2) Z_3 - Z_0$. Kriging amounts to solving a system of two equations, which is straightforward, but first requires to calculate the various covariances involved. For example, if the series of contiguous cores along AB is described by a three-dimensional elongated volume s and the target block (the trapezoid ABDC in projection on the horizontal plane, with some thickness in the vertical direction) by v, σ_{10} represents $\frac{1}{|s||v|} \int_s \int_v C(x' - x) \, dx' \, dx$, which is a sextuple integral. A special variogram model, the logarithmic or de Wijsian model, was widely used because it is very tractable for analytical calculations of average covariances with Taylor expansions (see numerous technical reports of Matheron on the internet site of Mines ParisTech, Center of Geosciences, On-line geostatistical library).

29.3 Development and Maturity: Trend, Neighborhood Selection

With the availability of computers in the late 1960s, it was possible to solve linear systems with about 10–20 equations. Kriging was then carried out with about ten data in and around the target block. Usually a neighborhood of one or two rings or aureolae around the target was used. If necessary, some data were grouped whose situations with respect to the target were similar. At the first international geostatistical congress in Rome in 1975, Michel David claimed that he was able to krige a mining block for a few cents, a reasonable price for real-world applications (David 1976).

In mining applications the outputs were documents with grid cells representing the blocks; the block estimates and the associated kriging standard deviations were printed in the grid cells. Very soon applications emerged in other domains than mining, with a slightly different objective: cartography, more precisely contour mapping. See, for example, Huijbregts and Matheron (1971), Chauvet and Chilès (1975) in oceanography; Delfiner (1973), Chauvet et al. (1976) in meteorology; Delfiner and Delhomme (1975), Delhomme (1978) in hydrology. Moreover, the phenomena considered in these application domains usually present a trend: the sea floor is deeper when moving away from the coast line, aquifers have a general gradient, the top of petroleum reservoirs is usually dome shaped. This called for

developments in two directions: kriging theory, with universal kriging to account for trends, and kriging practice, with a careful design of kriging neighborhoods.

29.3.1 Universal Kriging (UK)

The assumption of a constant mean—even if unknown—became soon a limitation for the application of kriging to phenomena displaying a trend. Kriging was therefore generalized by Matheron (1969) to random functions with a polynomial drift $m(x)$ of the form

$$m(x) = \sum_{\ell=0}^{L} a_\ell f^\ell(x)$$

where the a_ℓ are unknown coefficients and the $f^\ell(x)$ are the $L+1$ monomials with degree up to a given degree k (in the one-dimensional case, $L = k$ and $f^\ell(x) = x^\ell$). For $\ell = 0, f^0(x) \equiv 1$. The kriging estimator remains of the form $Z^* = \sum_\alpha \lambda_\alpha Z_\alpha$ but, because the a_ℓ are not known, unbiasedness is ensured only under the $L+1$ constraints

$$\sum_\alpha \lambda_\alpha f_\alpha^\ell = f_0^\ell \quad \ell = 0, \ldots, L$$

where $f_\alpha^\ell = f^\ell(x_\alpha)$ and f_0^ℓ is $f^\ell(x_0)$ if the target is $Z(x_0)$ or the average value of $f^\ell(x)$ when x spans v if the target is $Z(v)$. The minimization of the estimation variance leads to a system similar to the OK system except that there are now $L+1$ constraints instead of a single one, and as many Lagrange parameters.

The UK kriging matrix is no more positive definite, so that the kriging system should be solved by Gaussian elimination, which is less efficient than the Cholesky method. However, UK can be expressed as simple kriging, followed by a drift correction. The second step appears as the solution of a linear system of $L+1$ equations with $L+1$ unknowns, whose matrix is positive definite. It is thus advantageous to exploit this result to solve the SK system and the drift correction system by the Cholesky method (an additivity property also allows the calculation of the UK variance).

The equations of UK were already presented by Goldberger (1962) but not in a spatial context and with covariances supposed to be known, whereas Matheron proposed tools for determining the underlying variogram in the presence of a drift. These tools let appear an inference problem that was adequately solved in the framework of a more general model, presented hereafter.

29.3.2 Kriging in the IRF-k Model

Like the mean for OK, the coefficients a_ℓ are not involved in universal kriging. This made it possible to extend it to a more general random function model, the model of intrinsic random functions of order k (IRF-k), where a generalized covariance function $K(h)$ is substituted to $C(h)$. The RF model was first presented by Yaglom and Pinsker (1953), and the complete theory in the n-dimensional space by Matheron (1971, 1973). It suffices to say here that the class of GCs includes ordinary covariances and covariances of the form $-\gamma(h)$ when $k = 0$, and increases with k. It includes, for example the power covariances $(-1)^{p+1} |h|^{2p+1}$, $0 \leq p \leq k$, and the "spline" covariances $(-1)^{p+1} |h|^{2p} \log |h|$, p integer, $1 \leq p \leq k$. The kriging system is the same as for UK, with K replacing C.

29.3.3 Kriging as an Interpolant

In cartography, the objective of the applications of kriging was more precisely to draw maps with isolines derived from point kriging at the nodes of a regular grid. Nowadays it is possible to locally refine the grid to precisely track an isoline. In both cases, there is a requirement that kriging is not only the optimal linear estimator for a single point or block but also has nice interpolation properties.

According to theory, when kriging is considered as an interpolant, that is, as a function $z^*(x)$ of the target point x, the kriged map inherits from the covariance or variogram model. Indeed the universal kriging estimate can be presented in its dual form

$$z^*(x) = \sum_\alpha b_\alpha C(x - x_\alpha) + \sum_\ell c_\ell f^\ell(x)$$

with the convention that C can be replaced by $-\gamma$ or by the generalized covariance K. The coefficients b_α and c_ℓ are linear functions of the data. They are obtained as solutions of a system of equations similar to the UK system (same kriging matrix). If the variogram is parabolic at the origin, then $z^*(x)$ is differentiable; if the variogram is linear at the origin (and thus with a cusp at the origin when considered as a function of vector h), $z^*(x)$ is continuous with cusps at the data points. This may not be aesthetically nice from the user's point of view, because this is not primarily the purpose of kriging. Nevertheless, a smooth map can always be obtained by applying kriging with a smooth variogram or generalized covariance model. This is the way splines were used at that time, without explicit reference to geostatistics, but Matheron (1981) showed that any spline problem is equivalent to a kriging problem in the framework of the IRF-k model. For example, in 2D, interpolating with biharmonic splines is equivalent to kriging in the framework of an IRF-1

model with the generalized covariance $|h|^2 \log |h|$. Of course if the "true" covariance model does not conform to this model, kriging loses its optimality.

29.3.4 Neighborhood Selection

The dual kriging approach is very efficient in terms of computer time but presents two limitations: (i) it does not provide the kriging variance, and (ii) like direct kriging, its above interpolation properties are valid when working globally, that is, all data points are taken into account (global neighborhood). Due to practical limitations in memory space and calculation time, there is a limit in the number N of data that can be processed (several hundreds at that time, several thousands now). Therefore, in practice kriging often continues to be used with a moving neighborhood, that is, a limited number of data points around the target point are taken into account.

Now, when kriging with a moving neighborhood, the neighborhoods of two grid nodes can differ, and this can produce spurious discontinuities, especially when an outlier data is included in the neighborhood of a grid node and not in the neighborhood of the next grid node.

The neighborhood problem is also important when building conditional simulations. The classical way at that time (and even now) for continuous variables was to work in the framework of a Gaussian RF model (if necessary after suitable transformation of the data), to generate a nonconditional simulation of the Gaussian RF, and to condition that simulation on the data with a kriging step (Journel 1974). Due to their random nature, nonconditional simulations present small-scale variations. If spurious discontinuities are added by the kriging step, it is not easy to distinguish them from natural variations, which can lead to inaccurate conclusions.

Therefore, during years, much effort was devoted by software developers to neighborhood selection (e.g., Renard and Yancey 1984). Sophisticated algorithms have been devised to reach a compromise between near and far sample points. Focusing on 2D only, neighborhoods usually include all points of the first ring and then more distant points, following a strategy that attempts to sample all directions as uniformly as possible while keeping the number of points as low as possible (octant search). Typically, 16 to 32 points are retained, from at least five octants or four noncontiguous octants. For contour mapping purposes, where continuity is important, larger neighborhoods may be considered to provide more overlap. Such an algorithm may not provide satisfactory results when data originate from profiles sampled with a short interval. The neighborhood selection then includes the requirement to have data originating from several profiles. Along years, the size of the neighborhoods increased with the improvements of computers in terms of CPU time and storage.

29.3.5 Maturity

In the 1980s kriging seemed to have reached maturity. It was widely used in mining projects to build block models of orebodies, even with a large number of sample data and a very large number of blocks. In civil engineering it enabled an accurate design of the Channel tunnel on the basis of a model of the geological layers obtained by kriging from about 100 000 data, with a sound evaluation of the uncertainty of the model (Blanchin and Chilès 1993; Chilès and Delfiner 2012, Sect. 3.8). There were further developments specific to nonlinear geostatistics (disjunctive kriging, indicator kriging) and to multivariate geostatistics (factorial kriging analysis) which are not considered here.

At the same period, Sacks et al. (1989) opened a completely new domain to kriging: the design and analysis of computer experiments (DACE). The coordinates of x are no longer geographic but represent scalar design variables, while the variable of interest Z is an objective function that depends on the design variables. A computer experiment gives the value of the objective function for chosen values of the design variables. When computer experiments are costly, kriging is used to interpolate the response surface from a limited number of data (computer experiments). Applications mainly concern engineering problems, for example, the design of aircrafts (Chung and Alonso 2002). They call for specific research works, due to the very special space considered, the sparsity of the data, the difficulty to infer the covariance. See Kleijnen (2016) for a recent review.

29.4 Iterative Use of Kriging to Handle Inequality Data

Up to the early 1980s, geostatistics provided direct solutions: kriging was obtained by solving a linear system of equations, (Gaussian) simulations were built by turning bands or other methods directly transforming a vector of independent standard normal random variables in a vector representing a discrete view of the random function. Iterative algorithms appeared to handle inequality data and more specifically to generate conditional simulations of truncated Gaussian RFs.

Inequality data were already considered in the 1980s, notably by Dubrule and Kostov (1986) and Kostov and Dubrule (1986), with a solution based on quadratic programming where inequality data are treated as constraints placed on the kriging estimate. At the end, the inequalities are classified either as inactive (they can be forgotten) or active, and in the latter case they are replaced by an equality to the upper or lower bound of the inequality. This classification is not trivial at all and is the value of the method, but the clamping effect produced by the replacement of some inequalities by their lower or upper bound is not really satisfactory.

An alternative approach proposed by Langlais (1990) is to regard inequalities as data and replace them by exact values. The procedure is to (i) simulate exact data satisfying the given inequalities while honoring the exact data and the spatial

structure, (ii) average the results over several simulations, thus generating data that will replace the inequality data, and (iii) proceed to kriging from both actual and generated data.

At the same period, truncated Gaussian RFs were considered to represent geological facies. In its simplest form, such RF is defined by a Gaussian SRF $Y(x)$ and a threshold s. The truncated Gaussian RF is simply the indicator $1_{Y(x) \geq s}$. The applications account for a threshold that varies with x (an ordinary function of x). More general models are obtained with several thresholds and possibly two or three Gaussian SRFs (plurigaussian RF). Matheron et al. (1987) proposed a method to build conditional simulations of truncated Gaussian RFs in the case of a separable exponential covariance. The method is rather simple because it fully exploits the Markov properties of that covariance model.

From that time the geostatistics community devoted a growing interest to Markov chain Monte Carlo (MCMC) methods (e.g., Tjelmeland and Holden 1993), and particularly to the Gibbs sampler (Geman and Geman 1984). Initially developed to solve optimization problems, these methods also provide useful algorithms for generating simulations of RFs at a finite number of sites (e.g., grid nodes). The Gibbs sampler gives a consistent iterative method to achieve the first step of Langlais (1990), which is the critical one: simulate exact data satisfying the inequalities. Let us consider that the inequality data are of the form $Z_\alpha \in B_\alpha$ for some values of α, where B_α denotes an interval. The procedure is initialized by generating each of these Z_α separately, by a value z_α chosen in the interval B_α. Then the following sequence is repeated:

1. Select an inequality site α.
2. Simulate Z_α conditional on $Z_\alpha \in B_\alpha$ and $Z_\beta = z_\beta$ for all $\alpha \neq \beta$ (β ranges over all sites except α), and assign the simulated value to z_α.

The procedure changes the simulated values at the inequality sites so that they progressively honor the spatial structure given by the covariance. This approach finds its theoretical justification in the ideal case of a Gaussian SRF with a known mean, where the conditional distribution of Z_α is Gaussian with mean and variance equal to the kriging estimate and the kriging variance. It is however robust and is used even in the case of an unknown mean. The same approach is used effectively to generate conditional simulations constrained by inequality data, and especially truncated Gaussian RFs (the 0 or 1 data are transformed in inequality data of the form $Y(x_\alpha) < s$ or $Y(x_\alpha) \geq s$). The algorithm should be used in global neighborhood; otherwise, care should be given to the neighborhood selection, because the algorithm may diverge.

29.5 Nonstationary Covariance

Up to now we have considered models with a stationary covariance. But reality does not care about our theoretical models. If a stationary covariance is often a reasonable assumption when a limited number of samples is available, large data sets usually show some lateral variations in the covariance or variogram, so that a global model with a stationary covariance would be a too crude approximation. This problem is obviously not new. A simple solution is to split the study domain into several subdomains, to determine a specific variogram in each subdomain, and to krige each subdomain with its own variogram. To avoid discontinuities at subdomains boundaries, the variogram parameters evolve progressively from one model to the next in a transition area. This ad hoc method was used, for example, for the study of the Channel tunnel where the 100 000 data clearly showed structural variations along the 60 km of the tunnel project. Machuca-Mory and Deutsch (2013) generalize and systematize this approach.

Global nonstationary covariance models are of course sounder than the previous approach from a theoretical point of view, and also from a practical one if they can adapt to actual situations. A simple global covariance model can be derived by generalization of the covariogram model, defined by autoconvolution of an integrable and square integrable function $w(u)$:

$$g(h) = \int w(u)w(u+h)du$$

If we replace $w(u)$ by a dilution or kernel function $w(x; u)$ also depending on x, integrable and square integrable in u whatever x, and define

$$g(x, x') = \int w(x; u)w(x'; u)du$$

then $g(x, x')$ is a nonstationary covariance model (e.g., Higdon et al. 1999). A random function with that covariance can be obtained by the dilution method (Higdon 2002).

Let us now examine the case where w, considered as a function of u for fixed x, is a Gaussian kernel with variance–covariance matrix Σ_x. The resulting correlation function can be written (e.g., Paciorek and Schervish 2006)

$$g(x, x') = |\Sigma_x|^{1/4}|\Sigma_{x'}|^{1/4}\left|\frac{\Sigma_x + \Sigma_{x'}}{2}\right|^{-1/2}\exp(-Q_{xx'})$$

with quadratic form

$$Q_{xx'} = (x' - x)^{\mathrm{T}} \left(\frac{\Sigma_x + \Sigma_{x'}}{2} \right)^{-1} (x' - x)$$

If Σ_x is constant with respect to x, then $g(x, x')$ is the standard Gaussian correlation function with global anisotropy matrix Σ_x. Otherwise, if Σ_x varies slowly, g is approximately stationary in a small neighborhood of x. This locally stationary correlation function can be generalized by replacing $\exp(-Q_{xx'})$ by $\rho(-Q_{xx'})$ where ρ is a stationary correlation function that is valid in every dimension. This class of nonstationary covariance functions can be fitted by using local variograms whose parameters are used to build local Σ_x matrices (e.g., Fouedjio et al. 2016). Emery and Arroyo (2018) describe a spectral algorithm for simulating such models.

29.6 Kriging for Large Data Sets

We have seen that kriging with moving neighborhoods provides artifacts that can be limited in their amplitude by a careful design of the neighborhood selection but not eliminated. This problem is important when putting the Gibbs algorithm into practice because the procedure might diverge. The best way to avoid artifacts is to krige in global neighborhood, that is, any target point is kriged from all the data. As the capabilities of computers in terms of memory and computational performance always increase, this becomes possible for larger and larger data sets. However, the size of most data sets is also increasing with the advent of automatic measurement stations, so that the problem remains. A direct solving of the kriging system by Gaussian elimination or the Cholesky method is possible for up to several thousand equations. Several attempts were made for processing larger systems. Before presenting two truly global approaches, let us start with a method deriving from moving neighborhoods.

29.6.1 Continuous Moving Neighborhood

Gribov and Krivoruchko (2004) developed an original method to ensure continuity with moving neighborhoods. The idea is to modify the kriging system so that data beyond a specified distance from the estimated point receive weights gradually approaching zero. This way, no discontinuity occurs when data points enter or exit the kriging neighborhood.

Rivoirard and Romary (2011) propose an equivalent approach from a different perspective: The idea is to introduce a penalty on the kriging weights in the objective function to be minimized. This penalty acts as a noise variance except that

it varies with the target point x_0. It is typically equal to 0 for data points x_α within a distance r of the estimated point x_0 (no penalty applied near the target point), and increases continuously to infinity as x_α approaches the outer boundary of the kriging neighborhood, located at a distance R. Data points at a distance larger than R thus receive a zero weight. Because this method is solely based on the addition of a noise that increases with distance, it works for all versions of kriging algorithms: OK, UK, and even IRF-k. Because it is local, this method can handle lateral changes in the covariance parameters.

29.6.2   Covariance Tapering

Large systems can be solved if the kriging matrix is sparse. This can be achieved by tapering the covariance function to zero beyond a certain range. Furrer et al. (2006), who proposed this approach, define the tapered covariance as the product of the true covariance C by a taper covariance K that has a finite range. To preserve the behavior of the true covariance C near the origin, which controls the lateral continuity of the interpolant, the taper covariance K should be more regular near the origin than C. The authors apply the method with about 6 000 data.

29.6.3   Fixed Rank Kriging

In order to reduce the complexity of the kriging system when the number of data is very large, Cressie and Johannesson (2008) represent $Z(x)$ as a linear combination of r given basis functions $S_k(x)$ with random coefficients η_k, plus a white noise $\varepsilon(x)$ (for simplicity, we omit the covariates considered by the authors as external drift functions):

$$Z(x) = \sum_{k=1}^{r} \eta_k S_k(x) + \varepsilon(x)$$

The basis functions need not be orthogonal. They are usually chosen so as to represent several scales of variation and, for each scale, to cover the whole study domain. A typical choice is wavelet functions.

Denoting by $\mathbf{S}(x)$ the vector of the basic functions $S_k(x)$, by $\mathbf{K}$ the variance–covariance matrix of the η_k, and assuming that the white-noise variance is constant and equal to σ^2, the covariance of $Z(x)$ and $Z(x')$ is

$$C(x, x') = \mathbf{S}(x)^{\mathrm{T}} \mathbf{K} \mathbf{S}(x') + \sigma^2 \delta(x' - x)$$

where δ is the Kronecker function.

Given a vector $\mathbf{Z}$ of N data $Z(x_\alpha)$, the kriging matrix is

$$\boldsymbol{\Sigma} = \mathbf{S}\,\mathbf{K}\,\mathbf{S}^{\mathrm{T}} + \sigma^2\mathbf{I}$$

where $\mathbf{S}$ is the $N \times r$ matrix whose (α, k) element is $S_k(x_\alpha)$. The authors show that the inverse of $\boldsymbol{\Sigma}$ (an $N \times N$ positive-definite matrix) in fact only requires the inversion of $\mathbf{K}$ and $\mathbf{K}^{-1} + \mathbf{S}^{\mathrm{T}}\,\mathbf{S}/\sigma^2$ (two $r \times r$ positive-definite matrices). They also show that the inference of the positive-definite matrix $\mathbf{K}$ and the variance σ^2 can be done with the classical geostatistical approach. Therefore, kriging becomes tractable even with a very large number of data. In an application to ozone satellite data, the authors use 396 basis functions, a huge reduction in comparison with the 173 000 data.

29.6.4   *Gaussian Markov Random Field Approximation*

The approach of Gaussian Markov random fields may be seen as the opposite of that of covariance tapering in the sense that it seeks to make the inverse of the covariance matrix—and not the covariance matrix itself—sparse. It was first used to generate simulations (Besag 1974, 1975) but offers a new approach to kriging (Rue and Held 2005). Let us consider a Gaussian random vector $\mathbf{Z} = \{Z_i: i = 1, \ldots, N\}$ with known mean $\mathbf{m}$ and variance–covariance matrix $\mathbf{C}$. The conditional distribution of Z_i given the other components $\{Z_j: j \neq i\}$ is Gaussian with mean and variance the kriging estimate Z^*_{-i} of Z_i (the minus sign recalls that Z_i is excluded from the data used for that kriging) and the associated kriging variance σ^2_{Ki}. Denoting by $\mathbf{B}$ the inverse of $\mathbf{C}$, the kriging weights are found to be equal to $\lambda_j(i) = -B_{ij}/B_{ii}$ so that we have

$$Z^*_{-i} = m_i - \frac{1}{B_{ii}} \sum_{j \neq i} B_{ij}\,(Z_j - m_j) \qquad \sigma^2_{Ki} = \frac{1}{B_{ii}}$$

Since B_{ii} is the inverse of the conditional variance of Z_i given $\{Z_j: j \neq i\}$ (all except the i-th), $\mathbf{B}$ is known as the precision matrix. Its off-diagonal elements are related to the conditional correlations of Z_i and Z_j given $\{Z_k: k \neq i, j\}$ by

$$\mathrm{Corr}(Z_i, Z_j | \{Z_k: k \neq i, j\}) = -\frac{B_{ij}}{\sqrt{B_{ii}\,B_{jj}}}$$

$\mathbf{B}$ is a symmetric positive-definite matrix. The pattern of zeroes of $\mathbf{B}$ can be used to define an undirected graph structure in which two nodes are connected by an edge when $B_{ij} \neq 0$. Let ne(i) denote the neighborhood of node i, that is, the set of nodes connected to i by an edge. The vector $\mathbf{Z}$ has the Markov property that Z_i is conditionally independent of $\{Z_k: k \notin \mathrm{ne}(i)\}$ given $\{Z_j: j \in \mathrm{ne}(i)\}$. The discretely indexed Gaussian $\mathbf{Z}$ is called a Gaussian Markov random field (GMRF).

If the N components Z_i are split in N_1 unknown components to be estimated and $N_2 = N - N_1$ data, it can be shown that kriging can be achieved by solving a linear system of N_1 variables and N_1 equations whose system matrix is that part of the precision matrix $\mathbf{B}$ corresponding to the N_1 unknown components. The GMRF approach is used when this matrix is sparse, so that the system can be solved even when N_1 is large.

29.6.5   The Stochastic Partial Differential Equation (SPDE) Approach

Although the GMRF approach seems particularly appealing to deal with large data sets, its use remained limited due to the fact that the link with the geostatistical models based on covariance functions was not clear, making it difficult to parameterize the precision matrix. Nevertheless, some empirical studies showed that the commonly used covariance functions could be approximated quite closely by GMRFs (e.g., Rue and Tjelmeland 2002; Hrafnkelsson and Cressie 2003). These results spurred some authors to model the data by using a Gaussian field characterized by its covariance and then to find a discretized GRMF for which the inverse of the associated precision matrix $\mathbf{B}$ provides a good approximation of the covariance matrix of the Gaussian field (Song et al. 2008; Cressie and Verzelen 2008). Although promising, these algorithms suffer from a lack of theoretical foundations, which makes their application difficult.

In their seminal paper, Lindgren et al. (2011) propose a formal link between Gaussian field and GRMFs. They use a result established by Whittle in the 1950s linking some Gaussian fields and the solutions of a class of SPDEs. More precisely, let us consider the Matérn covariance function

$$C(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{|h|}{a}\right)^\nu K_\nu\left(\frac{|h|}{a}\right)$$

where σ^2 is the sill parameter, $a > 0$ is the scale parameter, $\nu > 0$ is a regularity parameter which determines the mean-square differentiability of the Gaussian field and K_ν is the modified Bessel function of the second kind and order ν. The result of Whittle (1954) states that a Gaussian field Z with Matérn covariance function C is a solution of the linear fractional SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} Z(s) = \tau\, W(s) \quad s \in \mathbb{R}^d$$

where $\alpha = \nu + d/2, \kappa = 1/a, \tau^2 = \frac{\Gamma(\nu+d/2)(4\pi)^{d/2}\kappa^{2\nu}}{\Gamma(\nu)}$, Δ is the Laplacian operator, and W is a Gaussian white noise with unit variance. The pseudo-differential operator $(\kappa^2 - \Delta)^{\alpha/2}$ can be defined through its Fourier transform but it is simply a linear combination of iterated Laplacians when $\alpha/2$ is an integer.

Then, by using some numerical methods to solve the PDE, for example, a finite differences method (FDM) or a finite elements method (FEM), Lindgren et al. (2011) show that the resulting discretized field at the mesh points (which can include the data locations) is a discrete GRMF. The precision matrix is directly provided by the FDM or FEM implementation. It is a sparse matrix although the number of non-zero elements increases with ν. Therefore, by including the target points in the mesh generation, one can perform kriging with very large data sets by using an efficient solver for sparse matrices. Note that, when α is not an integer, the operator $(\kappa^2 - \Delta)^{\alpha/2}$ has to be approximated by $\left(\sum_{i=0}^{p} \lambda_i \Delta^i\right)^{1/2}$, where p is the smallest integer greater than α. This operator can also be discretized by a FDM or FEM.

Anisotropies can be handled with the operator $(\kappa^2 - \mathrm{div}(H.\nabla))^{\alpha/2}$ where $\mathbf{H}$ is a symmetric positive-definite matrix linked to the anisotropy matrix and div is the divergence operator.

An interesting feature of the SPDE approach is that it allows to easily incorporate varying coefficients. For instance, the matrix $\mathbf{H}$ can be replaced by $\mathbf{H}(s)$ to handle a varying anisotropy (see Fuglstad et al. 2015).

Figure 29.2 presents a synthetic vertical section that could represent a variable of interest such as porosity in a sedimentary layer. The base and top of the layers were obtained by standard geostatistical simulations. The variable of interest was built according to the model of Fuglstad et al. (2015) with $\alpha = 3/2$, the matrix $\mathbf{H}$ incorporating the anisotropy model depicted in Fig. 29.3. This anisotropy model was deduced from the model of the base and top of the layer, with a constant range along the local direction of the layer, and a shorter range, varying proportionally to layer thickness, in the orthogonal direction. Figure 29.2 shows five vertical "drill-holes" considered as the data set, and Fig. 29.4 shows the kriged section obtained with the SPDE method. The latter shows the capability of this approach to account for the anisotropy model even in areas where there are no data (provided of course that information is available concerning the anisotropy). From a computational point of view, the method is extremely efficient: in 2D a data set with about 100 000 data can be processed in about 10 s on a standard computer, with possibly a number of conditional simulations nearly in the same time.
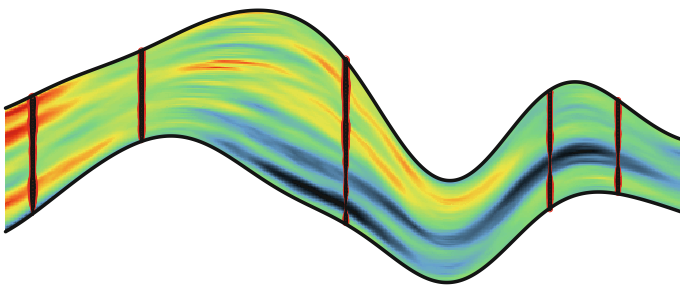

Fig. 29.2 SPDE synthetic case study: "Reality" (in fact a simulation) and sampling of five drill-holes
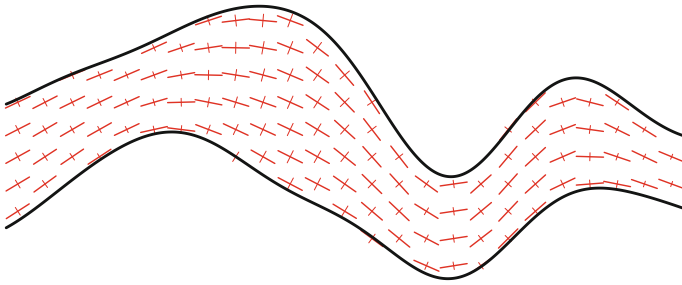
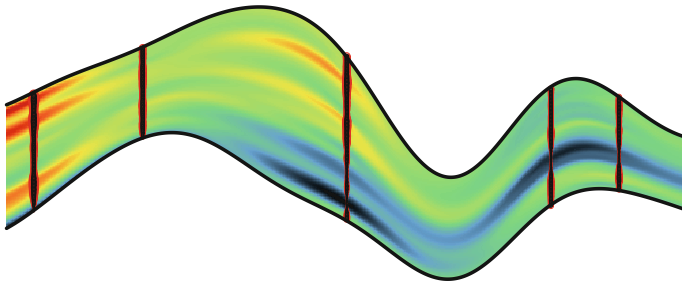**Fig. 29.3** SPDE synthetic case study: Anisotropy model

Fig. 29.4 SPDE synthetic case study: Kriging from the data of the five drill holes

29.7   Iterative Algorithms for Solving the Kriging System

Before to conclude, it is advisable to remind a presentation of two iterative kriging algorithms by Jean-François Royer in 1974, that is, in the early times of geostatistics. In meteorology, at that time, two main approaches were used to carry out the "objective analysis", that is, the interpolation of temperature and pressure at the nodes of a grid from the observations at time t, then used as input for a numerical weather forecast at time $t + 1$. One is Gandin's approach (1963), similar to simple kriging (in meteorology, the mean can be considered known thanks to a long sequence of observations). The other is an iterative approach, the method of successive corrections proposed by Cressman (1959).

Royer (1975) considers the simple monovariate situation. Rewritten with present notations, let us consider a vector **z** with $N = N_G + N_S$ components z_i, the first N_G components corresponding to grid nodes ($i \in G = \{1; \ldots; N_G\}$) and the other N_S components corresponding to observation stations ($i \in S = \{N_G + 1; \ldots; N_G + N_S\}$); z_i represents the variable of interest, at location x_i. Because the average situation for the season or month considered is known from past observations, we can subtract it and assume that **z** has mean **0**. Two iterative algorithms are proposed, depending on the set of points that drives the changes (grid nodes or observation stations). In both cases, an influence function $\rho(h)$ is used for extending a change

made at location x to location $x + h$ depending on separation h. This function satisfies $\rho(0) = 1$ and decreases to 0 when h increases. When extending to x_j a change made at x_i, the notation $\rho_{ij} = \rho(x_j - x_i)$ will be used.

Algorithm driven by grid nodes: As step $p = 0$, select a vector $\mathbf{z}^0$ with components z_i^0, for example zeroes or the values of the weather forecast for time t based on the objective analysis made at time $t - 1$. Then iterate as follows:

1. increase the step number p by 1
2. calculate the discrepancies of step $p - 1$ with regard to the data: $z_i - z_i^{p-1}$, $i \in S$
3. define model p as $z_i^p = z_i^{p-1} + \sum_{j \in S} \rho_{ij}(z_i - z_i^{p-1})$, $\quad i \in G \cup S$

Algorithm driven by the observations: As initial state, select a vector $\mathbf{z}^{\text{new}}$ with components z_i^{new}, for example zeroes or the values of the weather forecast for time t based on the objective analysis made at time $t - 1$. Then iterate as follows:

1. Set $z_j^{\text{current}} = z_j^{\text{new}}$, $\quad j \in G \cup S$
2. Select a component of S, say i, at random or by systematic scans of all the components of S
3. Calculate the discrepancy of the current value with regard to the observation: $z_i - z_i^{\text{current}}$
4. Define $z_i^{\text{new}} = z_i$
5. Update all other components so that $z_j^{\text{new}} - z_j^{\text{current}} = \rho_{ij}(z_i^{\text{new}} - z_i^{\text{current}})$, $j \neq i$

The convergence of both algorithms is ensured if and only if the matrix $\boldsymbol{\rho}$ defined by the ρ_{ij} is positive definite, which is ensured if $\rho(h)$ is a correlogram. Moreover, in that case, the iterative process converges to the solution of dual kriging. Indeed, both approaches amount to an iterative resolution of the dual kriging system (by the Jacobi method in the first approach, by the Gauss-Seidel method in the second one), followed, after each iteration, by the propagation of the changes to the point kriging estimates.

The second algorithm is very similar to the Gibbs propagation algorithm proposed nearly 40 years later by Lantuéjoul and Desassis (2012) to simulate a Gaussian vector (this algorithm is also presented in Chilès and Delfiner 2012, Sect. 7.6.3; it constitutes a further step to an algorithm proposed by Galli and Gao 1999). It is this similarity that reminded one of the present authors the paper of Royer, not exploited by geostatisticians to our knowledge, which should deserve new consideration. These iterative algorithms have the advantage that they can be used even with a very large number of data, notably when the Cholesky method cannot be used.

29.8   Conclusion

We have shown the long way from Krige's regression, which took account of two average sample grades (a local one and a global one) to avoid bias in the estimation of a panel, to present applications of kriging, which can deal with few data (e.g., a limited number of computer experiments in applications to DACE) as well as several hundred thousand data (remote sensing, seismic). We have seen the large diversity of application domains of kriging, so that is it probable that many users do not know the origin of the word: this is the price of success.

We also gave a look at current research to enable a global application of kriging to large data sets, with the requirement to also benefit from nonstationary random function models. Much work remains necessary to transform them in standard methods applicable to a large variety of situations but, in view of the large community of researchers and developers in this area, no doubt that it will be done. The future will show which approaches are the most efficient ones.

References

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). J Royal Stat Soc Ser B 36(2):192–236

Besag J (1975) Statistical analysis of non-lattice data. J Royal Stat Soc Ser D 24(3):179–195

Blanchin R, Chilès J-P (1993) The Channel Tunnel: geostatistical prediction of the geological conditions and its validation by the reality. Math Geol 25(7):963–974

Chauvet P, Chilès J-P (1975) Kriging: a method for the cartography of the sea floor. Int Hydrogr Rev 52(1):25–41

Chauvet P, Pailleux J, Chilès J-P (1976) Analyse objective des champs météorologiques par cokrigeage. La Météorologie, Sciences et Techniques 6(4):37–54

Chilès J-P, Delfiner P (2012) Geostatistics: modeling spatial uncertainty, 2nd edn. Wiley, New York

Chung H, Alonso JJ (2002) Design of a low-boom supersonic business jet using cokriging approximation models. In: 9th AIAA/ISSMO symposium on multidisciplinary analysis and optimization, p 5598

Cousin A, Maatouk H, Rullière D (2016) Kriging of financial term-structures. Eur J Oper Res 255 (2):631–648

Cressie N (1990) The origins of kriging. Math Geol 22(3):239–252

Cressie N, Johannesson G (2008) Fixed rank kriging for very large spatial data sets. J R Stat Soc Ser B 70(1):209–226

Cressie N, Verzelen N (2008) Conditional-mean least-squares fitting of Gaussian Markov random fields to Gaussian fields. Comput Stat Data Anal 52(5):2794–2807

Cressman GP (1959) An operational objective analysis system. Mon Weather Rev 87(10):367–374

David M (1976) The practice of kriging. In: Guarascio M et al (eds) Advanced geostatistics in the mining industry. Reidel, Dordrecht, pp 31–48

Delfiner P (1973) Analyse objective du géopotentiel et du vent géostrophique par krigeage universel. New print (1975): La Météorologie 5(25):1–58

Delfiner P, Delhomme J-P (1975) Optimum interpolation by kriging. In: Davis JC, McCullagh MJ (eds) Display and analysis of spatial data. Wiley, London, pp 96–114

Delhomme J-P (1978) Kriging in the hydrosciences. Adv Water Resour 1(5):251–266

Dixit V, Seshadrinath N, Tiwari MK (2016) Performance measures based optimization of supply chain network resilience: a NSGA-II+Co-Kriging approach. Comput Ind Eng 93:205–214

Dubrule O, Kostov C (1986) An interpolation method taking into account inequality constraints: I. Methodology. Math Geol 18(1):33–51

Emery X, Arroyo D (2018) On a continuous spectral algorithm for simulating non-stationary Gaussian random fields. SERRA 32(4):905–919

Fouedjio F, Desassis N, Rivoirard J (2016) A generalized convolution model and estimation for non-stationary random functions. Spat Stat 16:35–52

Fulgstad GA, Lindgren F, Simpson D, Rue H (2015) Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. Stat Sin 25:115–133

Furrer R, Genton MG, Nychka DW (2006) Covariance tapering for interpolation of large spatial datasets. J Comput Graph Stat 15(3):502–523

Galli A, Gao H (1999) Rate of convergence of the Gibbs sampler in the Gaussian case. Math Geol 33(6):653–677

Gandin LS (1963) Ob″ektivnyi analiz meteorologicheskikh polei. Gidrometeologicheskoe Izdatel'stvo, Leningrad. Translation (1965): Objective analysis of meteorological fields. Israel program for scientific translations, Jerusalem

Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 6(6):721–741

Goldberger AS (1962) Best linear unbiased prediction in the generalized regression model. J Am Stat Assoc 57:369–375

Gribov A, Krivoruchko K (2004) Geostatistical mapping with continuous moving neighbourhood. Math Geol 36(2):267–281

Higdon D (2002) Space and space-time modeling using process convolutions. In: Anderson CW et al (eds) Quantitative methods for current environmental issues. Springer, London, pp 37–54

Higdon D, Swall J, Kern J (1999) Non-stationary spatial modeling. In: Bernardo JM et al (eds) Bayesian statistics 6: proceedings of the sixth Valencia international meeting. Oxford University Press, New York, pp 761–768

Hrafnkelsson B, Cressie N (2003) Hierarchical modeling of count data with application to nuclear fall-out. Environ Ecol Stat 10(2):179–200

Huijbregts C, Matheron G (1971) Universal kriging (an optimal method for estimating and contouring in trend surface analysis). In: McGerrigle JI (ed) 9th international symposium on techniques for decision-making in the mineral industry. Canadian Institute of Mining and Metallurgy, pp 159–169

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

Kleijnen JPC (2016) Regression and Kriging metamodels with their experimental designs in simulation: A review. Eur J Oper Res 256(1):1–16

Kostov C, Dubrule O (1986) An interpolation method taking into account inequality constraints: II. Practical approach. Math Geol 18(1):53–73

Krige DG (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. J Chem Metal Min Soc S Afr, December, 119–139

Krige, DG (1952) A statistical analysis of some of the borehole values in the Orange Free State goldfield. J Chem Metal Min Soc S Afr, September, 47–64

Langlais V (1990) Estimation sous contraintes d'inégalités. Doctoral thesis, E.N.S. des Mines de Paris

Lantuéjoul C, Desassis N (2012) Simulation of a Gaussian random vector: a propagative version of the Gibbs sampler. In: The 9th international geostatistics congress, Oslo, Norway, 11–15 June 2012, pp 174–181

Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential approach. J R Stat Soc B 73(4):423–498

Machuca-Mory DF, Deutsch CV (2013) Non-stationary geostatistical modeling based on distance weighted statistics and distributions. Math Geosci 45(1):31–48

Matheron G (1960) Krigeage d'un panneau rectangulaire par sa périphérie, Note géostatistique no 28, Centre de Géostatistique, Fontainebleau, France

Matheron G (1961) Exemple de krigeage d'un panneau riche. Note géostatistique No 31. Centre de Géostatistique, Fontainebleau, France

Matheron G (1962) Traité de géostatistique appliquée, Tome I. Mémoires du Bureau de Recherches Géologiques et Minières, no. 14. Editions Technip, Paris

Matheron G (1963a) Traité de géostatistique appliquée, Tome II: Le krigeage. Mémoires du Bureau de Recherches Géologiques et Minières, no. 24, Editions B.R.G.M., Paris

Matheron G (1963b) Principles of geostatistics. Econ Geol 58(8):1246–1266

Matheron G (1969) Le krigeage universel. Cahiers du Centre de morphologie mathématique de Fontainebleau, Fasc 1, Ecole des mines de Paris

Matheron G (1971) La théorie des fonctions aléatoires intrinsèques généralisées. Note Géostatistique no 117. Technical report N-252, Centre de Géostatistique, Fontainebleau, France

Matheron G (1973) The intrinsic random functions and their applications. Adv Appl Prob 4 (3):508–541

Matheron G (1981) Splines and kriging: their formal equivalence. In: Merriam DF (ed) Down-to-earth statistics: solutions looking for geological problems. Syracuse University Geological Contributions, pp 77–95

Matheron G, Beucher H, Fouquet C de, Fouquet C de, Galli A, Guérillot D, Ravenne C (1987) Conditional simulation of the geometry of fluvio-deltaic reservoirs. SPE Paper 16753

de Oliveira MA, Possamai O, Dalla Valentina LVO, Flesch CA (2013) Modeling the leadership—project performance relation: radial basis function, Gaussian and Kriging methods as alternatives to linear regression. Expert Syst Appl 40(1):272–280

Paciorek CJ, Schervish MJ (2006) Spatial modeling using a new class of nonstationary covariance functions. Environmetrics 17(5):483–506

Renard D, Yancey JD (1984). Smoothing discontinuities when extrapolating using moving neighbourhoods. In: Verly G et al (eds) Geostatistics for natural resources characterization, Reidel, Dordrecht, part 2, pp 679–690

Rivoirard J, Romary T (2011) Continuity for kriging with moving neighborhood. Math Geosci 43 (4):469–481

Royer J-F (1975) Comparaison des méthodes d'analyse objective par interpolation optimale et par approximations successives. Technical report no 365, Etablissement d'Etudes et de Recherches Météorologiques, Direction de la Météorologie, Paris

Rue H, Held L (2005) Gaussian Markov random fields. Chapman & Hall/CRC, Boca Raton, Florida

Rue H, Tjelmeland H (2002) Fitting Gaussian Markov random fields to Gaussian fields. Scand J Stat 29(1):31–49

Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. Stat Sci 4(4):409–453

Song H, Fuentes M, Gosh S (2008) A comparative study of Gaussian geostatistical models and Gaussian Markov random field models. J Multivar Anal 99(8):1681–1697

Tjelmeland H, Holden L (1993) Semi-Markov random fields. In: Soares A (ed) Geostatistics Tróia '92. Kluwer, Dordrecht, vol 1, pp 479–491

Yaglom AM, Pinsker MS (1953) Random processes with stationary increments of order n. Dokl Akad Nauk SSSR 90(5):731–734

Yan JW, Liew KM, He LH (2012) A mesh-free computational framework for predicting buckling behaviors of single-walled carbon nanocones under axial compression based on the moving Kriging interpolation. Comput Methods Appl Mech Eng 247–248:103–112

Whittle P (1954) On stationary processes in the plane. Biometrika 41(3–4):434–449