

Kriging and epistemic uncertainty : a critical discussion

Kevin Loquin and Didier Dubois

Abstract Geostatistics is a branch of statistics dealing with spatial phenomena modelled by random functions. In particular, it is assumed that, under some well-chosen simplifying hypotheses of stationarity, this probabilistic model, i.e. the random function describing spatial dependencies, can be completely assessed from the dataset by the experts. Kriging is a method for estimating or predicting the spatial phenomenon at non sampled locations from this estimated random function. In the usual kriging approach, the data is precise and the assessment of the random function is mostly made at a glance by the experts (i.e. geostatisticians) from a thorough descriptive analysis of the dataset. However, it seems more realistic to assume that spatial data is tainted with imprecision due to measurement errors and that information is lacking to properly assess a unique random function model. Thus, it would be natural to handle epistemic uncertainty appearing in both data specification and random function estimation steps of the kriging methodology. Epistemic uncertainty consists of some meta-knowledge about the lack of information on data precision or on the model variability. The aim of this paper is to discuss the pertinence of the usual random function approach to model uncertainty in geostatistics, to survey the already existing attempts to introduce epistemic uncertainty in geostatistics and to propose some perspectives for developing new tractable methods that may handle this kind of uncertainty.

Key words: geostatistics; kriging; variogram; random function; epistemic uncertainty; fuzzy subset; possibility theory

Kevin Loquin
IRIT, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex 9. e-mail: loquin@irit.fr

Didier Dubois
IRIT, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex 9. e-mail: dubois@irit.fr

1 Introduction

Geostatistics is the application of the formalism of random functions to the reconnaissance and estimation of natural phenomena.

This is how Georges Matheron [40] explains the term geostatistics in 1962 to describe a scientific approach to estimation problems in geology and mining. The development of geostatistics in the 1960s resulted from the industrial and economical need for a methodology to assess the recoverable reserves in mining deposit. Naturally, the necessity to take into account uncertainty in such methods appeared. That is the reason why statisticians were needed by geologists and mining industry to perform ore assessment consistently with the available information.

Today, geostatistics is no longer restricted to this kind of application. It is applied in disciplines such as hydrology, meteorology, oceanography, geography, forestry, environmental monitoring, landscape ecology, agriculture or for ecosystem geographical and dynamic study.

Underlying each geostatistical method is the notion of random function [12]. A random function describes a given spatial phenomenon over a domain. It consists of a set of random variables, each of which describes the phenomenon at some location of the domain. By analogy with a random process, which is a set of random variables indexed by time, a random function is a set of random variables indexed by locations. When little information is available about the spatial phenomenon, a random function is only specified by the set of means associated to its random variables over the domain and its covariance structure for all pairs of random variables induced by this random function. These parameters describe, respectively, the spatial trend and spatial dependencies of the underlying phenomenon. The dependence structural assumption underlying most of the geostatistical methods is based on the intuitive idea that, the closer are the regions of interest, the more similar is the phenomenon in these areas. In most geostatistical methods, the dependencies between the random variables are preferably described by a variogram instead of a covariance structure. The variogram depicts the variance of the increments of the quantity of interest as a function of the distance between sites.

The spatial trend and spatial dependence structure of this model are commonly supposed to be of a given form (typically, linear for the trend and spherical, power exponential, rational quadratic for the covariance or variogram structure) with a small number of unknown parameters. From the specification of these moments, many methods can be derived in geostatistics. By far, kriging is the most popular one. Suppose a spatial phenomenon is partially observed at selected sites. The aim of kriging is to predict the phenomenon at unobserved sites. This is the problem of spatial estimation, sometimes called spatial prediction. Examples of spatial phenomena estimations are soil nutrient or pollutant concentrations over a field observed on a survey grid, hydrologic variables over an aquifer observed at well locations, and air quality measurements over an air basin observed at monitoring sites.

The term kriging was coined by Matheron in honor of D.G. Krige who published an early account of this technique [37] with applications to estimation of a min-

eral ore body. In its simplest form, a kriging estimate of the field at an unobserved location is an optimized linear combination of the data at observed locations. The method has close links to Wiener optimal linear filtering in the theory of random functions, spatial splines and generalized least squares estimation in a spatial context.

A full application of a kriging method by a geostatistician involves different steps:

1. An important structural analysis is performed: usual statistical tools like histograms, empirical cumulative distributions, can be used in conjunction with an analysis of the sample variogram.
2. In place of the sample variogram, that does not respect suitable mathematical properties, a theoretical variogram is chosen. The fitting of the theoretical variogram model to the sample variogram, informed by the structural analysis, is performed.
3. Finally, from this variogram specification (which is an estimate of the dependence structure of the model), the kriging estimate is computed at the location of interest by solving a system of linear equations of the least squares type.

Kriging methods have been studied and applied extensively since 1970 and later on adapted, extended, and generalized. Georges Matheron, who founded the “Centre de Géostatistiques et de Morphologie Mathématique de l’école des Mines de Paris” in Fontainebleau, proposed the first systematic approach to kriging [40]. Many of his students or collaborators followed his steps and worked on the development and dissemination of geostatistics worldwide. We can mention here Jean-Paul Chilès, Pierre Delfiner [10] or André G. Journel [32] among others. All of them worked on extending, in many directions, the kriging methodology.

However, very few scholars discussed the nature of the uncertainty that underlies the standard Matheronian geostatistics except G. Matheron himself [43] and even fewer considered alternative theories to probability theory that could more reliably handle epistemic uncertainty in geostatistics. Epistemic uncertainty is uncertainty that stems from a lack of knowledge, from insufficient available information, about a phenomenon. It is different from uncertainty due to the variability of the phenomenon. Typically, intervals or fuzzy sets are supposed to handle epistemic uncertainty, while probability distributions are supposed to properly quantify variability.

More generally, imprecise probability theories, like possibility theory [21], belief functions [49] or imprecise previsions [53] are supposed to jointly handle those two kinds of uncertainty. Consider the didactic example of a dice toss where you have more or less information about the number of facets of the dice. When you know that a dice has 6 facets, you can easily evaluate the variability of the dice toss: 1 chance over 6 for each facet from 1 to 6; but now, suppose that you miss some information about the number of facets and that you just know that the dice has either 6 or 12 facets, you can not propose a unique model of variability of the dice toss, you can just propose two: in the first case, 1 chance over 6 for each facet from 1 to 6 and 0 chance for each facet from 7 to 12, in the second case, 1 chance over 12 for each facet from 1 to 12. This example enables the following simple

conceptual extrapolation: when you are facing a lack of knowledge or insufficient available information on the studied phenomenon, it is safer to work with a family of probability distributions, i.e. to work with sets of probability measures, to model uncertainty. Such models are generically called imprecise probability models.

Bayesian methods address the problem by attaching prior probabilities to each potential model. However, this kind of uncertainty is of purely epistemic origin and using a single subjective probability to describe it is debatable, since it represents much more information than what is actually available. In our dice toss example, choosing a probability value for the occurrence of each possible model, even if we choose a uniform distribution, i.e. a probability of $1/2$, for each possible model, is much more information than actually available about the occurrence of the possible models. Besides, it is not clear that subjective and objective probabilities can be multiplied as they represent information of a very different nature.

This paper proposes a discussion of the standard approach to kriging in relation with the presence of epistemic uncertainty pervading the data or the choice of a variogram. In the first part of the paper basics of kriging theory are recalled and the underlying assumptions discussed. Then, a survey of some existing intervallist or fuzzy extensions of kriging is offered. Finally, a preliminary discussion of the role novel uncertainty theories could play in this topic is provided.

2 Some basic concepts in probabilistic geostatistics

Geostatistics, is commonly viewed as the application of the “Theory of Regionalized Variables” to the study of spatially distributed data. This theory is not new and borrows most of its models and tools from the concept of stationary random function and from techniques of generalized least-squares prediction.

Let \mathcal{D} be a compact subset of \mathbb{R}^ℓ and $Z = \{Z(x), x \in \mathcal{D}\}$ denotes a real valued random function. A random function (or equally a random field) is made up of a set of random variables $Z(x)$, for each $x \in \mathcal{D}$. In other words, Z is a set of random variables $Z(x)$ indexed by x . Each $Z(x)$ takes its values in some real interval $\Gamma \subseteq \mathbb{R}$. In this approach, Z is the probabilistic representation of a deterministic function $z: \mathcal{D} \rightarrow \Gamma$.

The data consists of n observations $\mathcal{Z}_n = \{z(x_i), i = 1, \dots, n\}$ understood as a realization of the n random variables $\{Z(x_i), i = 1, \dots, n\}$ located at the n known distinct sampling positions $\{x_1, \dots, x_n\}$ in \mathcal{D} . \mathcal{Z}_n is the only available objective information about Z on \mathcal{D} .

2.1 Structuring assumptions

Different structuring assumptions of a random field have been proposed. They mainly aim at making the model easy to use in practice. Results of geostatistical methods highly depend on the choice of those assumptions.

2.1.1 The second-order stationary model

In geostatistics, the spatial dependence between the two random variables $Z(x)$ and $Z(x')$, located at different positions $x, x' \in \mathcal{D}$, is considered an essential aspect of the model. All geostatistical models strive to capture such spatial dependence, in order to provide information about the influence of the neighborhood of a point x on the random variable $Z(x)$.

A random function Z is said to be second-order stationary if any two random variables $Z(x)$ and $Z(x')$ have equal mean values, and their covariance function only depends on the *separation* $h = x - x'$. Formally, $\forall x, x' \in \mathcal{D}$, there exist a constant $m \in \mathbb{R}$ and a positive definite covariance function $C : \mathcal{D} \rightarrow \mathbb{R}$, such that

$$\begin{cases} \mathbb{E}[Z(x)] = m, \\ \mathbb{E}[Z(x) - m][Z(x') - m] = C(x - x') = C(h). \end{cases} \quad (1)$$

Such a model implies that the variance of the random variables $Z(x)$ is constant all over the domain \mathcal{D} . Indeed, for any $x \in \mathcal{D}$, $V(Z(x)) = C(0)$. In the simplest case, the random function is supposed to be Gaussian, and the correlation function isotropic, i.e. not depending on the direction of vector $x - x'$, so that h is a positive distance value $h = \|x - x'\|^\ell$.

A second-order stationary random function will be denoted by *SRF* in the rest of the paper.

2.1.2 The intrinsic model

This model is slightly more general than the previous one: it only assumes that the increments $Y_h(x) = Z(x+h) - Z(x)$, and not necessarily the random function Z itself, form a second-order stationary random function Y_h , for every vector h . More precisely, for each location $x \in \mathcal{D}$, $Y_h(x)$ is supposed to have a zero mean and a variance depending only on h and denoted by $2\gamma(h)$. In that case, Z is called an *intrinsic random function*, denoted by *IRF* in the rest of the paper, and characterized by:

$$\begin{cases} \mathbb{E}[Y_h(x)] = \mathbb{E}[Z(x+h) - Z(x)] = 0, \\ \mathbb{V}[Y_h(x)] = \mathbb{V}[Z(x+h) - Z(x)] = 2\gamma(h). \end{cases} \quad (2)$$

$\gamma(h)$ is the variogram. The variogram is a key concept of geostatistics. It is supposed to measure the dependence between locations, as a function of their distance.

Every *SRF* is an *IRF*, the contrary is not true in general. Indeed, from any covariance function of an *SRF*, we can derive an associated variogram as:

$$\gamma(h) = C(0) - C(h). \quad (3)$$

Indeed,

$$\begin{aligned} \gamma(h) &= \frac{1}{2}(\mathbb{V}[Z(x+h) - Z(x)]), \\ &= \frac{1}{2}(\mathbb{V}[Z(x+h)] + \mathbb{V}[Z(x)] - 2Cov(Z(x+h), Z(x))), \\ &= \frac{1}{2}(2C(0) - 2C(h)). \end{aligned}$$

In the opposite direction, the covariance function of an *IRF* is generally not of the form $C(h)$ and cannot be derived from its variogram $\gamma(h)$. Indeed, the inference from the second to the third line of the above derivation shows that equality (3) only holds if the variance of a random function is constant on the domain \mathcal{D} . This is the case for an *SRF* but not for an *IRF*. For example, unbounded variograms have no associated covariance function. It does not mean that the covariance between $Z(x)$ and $Z(x+h)$, when Z is an *IRF*, does not exist, but it is not, generally, a function of the separation h . The variogram is a more general structuring tool than the covariance function of the form $C(h)$.

2.2 Simple kriging

Kriging boils down to spatially interpolating the data set \mathcal{Z}_n by means of a linear combination of the observed values at each measurement location. The interpolation weights depend on the interpolation location and the available data over a domain of interest. In such a method, for estimating the value of the random function at an unobserved site, the dependence structure of the random function is used. Once the variogram is estimated, the kriging equations are obtained by least squares minimization.

Consider a second-order stationary random function Z , i.e. satisfying (1), informed by the data set $\mathcal{Z}_n = \{z(x_i), i = 1, \dots, n\}$. Any particular unknown value $Z(x_0)$, $x_0 \in \mathcal{D}$, is supposed to be estimated by a linear combination of the n collected data points $\{z(x_i), i = 1, \dots, n\}$. This estimation, denoted by $z^*(x_0)$, is given by:

$$z^*(x_0) = \sum_{i=1}^n \lambda_i(x_0) z(x_i). \quad (4)$$

The computation of $z^*(x_0)$ depends on the estimation of the kriging weights $\Lambda_n(x_0) = \{\lambda_i(x_0), i = 1, \dots, n\}$ at location x_0 . In the kriging paradigm, each weight $\lambda_i(x_0)$ corresponds to the *influence* of the value $z(x_i)$ in the computation of $z^*(x_0)$. More

precisely, the value $z^*(x_0)$ is the linear combination of the data set $\mathcal{Z}_n = \{z(x_i), i = 1, \dots, n\}$, weighted by the set of influence weights $\Lambda_n(x_0)$.

Kriging weights are computed by solving a system of equations induced by a least squares optimization method. It is deduced from the minimization of the estimation error variance $Z(x_0) - Z^*(x_0)$, where $Z(x_0)$ is the random variable underlying the *SRF* Z at location x_0 and $Z^*(x_0) = \sum_{i=1}^n \lambda_i(x_0)Z(x_i)$ is the “randomized” counterpart of the kriging estimate (4). The minimization of $\mathbb{V}[Z(x_0) - Z^*(x_0)]$ is carried out under the unbiasedness condition: $\mathbb{E}[Z(x_0)] = \mathbb{E}[Z^*(x_0)]$. This unbiasedness condition has a twofold consequence: first, it induces the following condition on the kriging weights:

$$\sum_{i=1}^n \lambda_i(x_0) = 1.$$

Indeed, due the stationarity of the mean (1), $\mathbb{E}[Z^*(x_0)] = \mathbb{E}[Z(x_0)] \Rightarrow \sum_{i=1}^n \lambda_i(x_0)m = m \Rightarrow \sum_{i=1}^n \lambda_i(x_0) = 1$.

Second, it implies that minimizing the variance can be rewritten in terms of mean squared error to minimize. Indeed, $\mathbb{V}[Z(x_0) - Z^*(x_0)] = \mathbb{E}[Z(x_0) - Z^*(x_0)]^2 - (\mathbb{E}[Z(x_0) - Z^*(x_0)])^2$ and the second term is zero. Thus,

$$\mathbb{V}[Z(x_0) - Z^*(x_0)] = \mathbb{E}[Z(x_0) - Z^*(x_0)]^2.$$

Thus, the kriging problem comes down to find the least squares estimate of Z at location x_0 under the constraint

$$\sum_{i=1}^n \lambda_i(x_0) = 1.$$

To obtain the kriging equations, the variance $\mathbb{V}[Z(x_0) - Z^*(x_0)]$ is rewritten as follows:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i(x_0)\lambda_j(x_0)C(x_i - x_j) - 2 \sum_{j=1}^n \lambda_j(x_0)C(x_0 - x_j) + \mathbb{V}[Z(x_0)], \quad (5)$$

where $\mathbb{V}[Z(x_0)] = C(0)$, so that kriging weights only depend on the covariance function. In order to minimize the above mean squared error, the derivative according to each kriging weight $\lambda_i(x_0)$ is computed:

$$\frac{\partial}{\partial \lambda_i(x_0)} \mathbb{V}[Z(x_0) - Z^*(x_0)] = 2 \sum_{j=1}^n \lambda_j(x_0)C(x_i - x_j) - 2C(x_0 - x_i), \quad \forall i = 1, \dots, n.$$

The equations providing the kriging weights are thus obtained by letting these partial derivatives vanish. The simple kriging equations are thus of the form:

$$C(x_0 - x_i) = \sum_{j=1}^n \lambda_j(x_0)C(x_i - x_j), \quad \forall i = 1, \dots, n. \quad (6)$$

The similarity between equations (4) and (6) is striking. The influence weights, in the simple kriging method, are the same weights as the ones that express, for all the locations $\{x_i, i = 1, \dots, n\}$, the dependence between $Z(x_0)$ and $Z(x_i)$, quantified by $C(x_0 - x_i)$, as the weighted average of the covariances $C(x_i - x_j)$ between $Z(x_i)$ and the random variables $\{Z(x_j), j = 1, \dots, n\}$. It can be summarized by this remark: the influence weights of the kriging estimate are the same as the influence weights of the dependence evaluations. It is clear that some proper dependence assessments should be the basis for any sensible interpolation of the observations. However, it does not seem to exist a direct intuitive interpretation why the observations should be combined (by means of (4)) just like the dependencies (by means of (6)).

In the case of kriging with an unknown mean based on the intrinsic model, the covariance function is replaced by the variogram in the kriging equations (6). Moreover there is an additional parameter value to be found, namely the Lagrange parameter needed to ensure the unbiasedness condition (that not always reduces to summing the kriging weights to 1); see [10], Section 3.4.

3 Variogram or covariance function estimation

In kriging, the dependence information between observations are taken into account to interpolate the set of points $\{(x_i, Z(x_i)), i = 1, \dots, n\}$. The most popular tool that models these dependencies is the variogram and not the covariance function, because the covariance function estimation is biased by the mean. Indeed, if the mean is unknown, which is generally the case, it affects the covariance function estimation. Geostatisticians proposed different functional models of variogram to comply with the observations and with the physical characteristics of a spatial domain [10]. In the first part of this section, we present the characteristics of the most popular variogram models.

Choosing one model or even combining some models to propose a new one is a subjective task requiring the geostatistician expertise and some prior descriptive analysis of the dataset \mathcal{Z}_n . The data is explicitly used only when a regression analysis is performed to fit the variogram model parameters to the empirical variogram. An empirical variogram, i.e. a variogram explicitly obtained from the dataset \mathcal{Z}_n and not by some regression on a functional model, is called a *sample variogram* in the literature. In this section, we will see that a sample variogram does not fulfil (in its general expression) the conditional negative definiteness requirement imposed on a variogram model. We will briefly discuss this point, which explains why a sample variogram is never used by geostatisticians to carry out an interpolation by kriging.

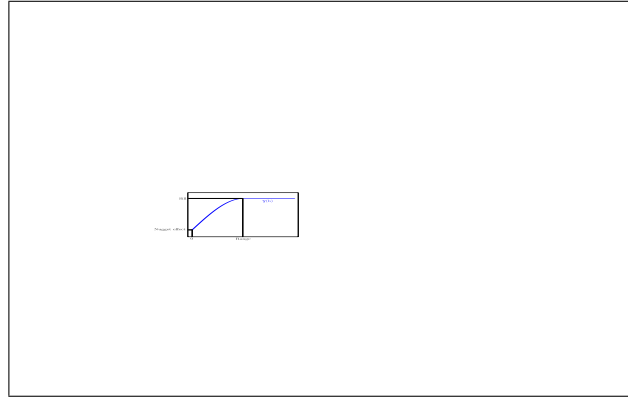


Figure 1 Qualitative properties of standard variogram models

3.1 Theoretical models of variogram or covariance functions

For the sake of clarity, we restrict this presentation of variogram models to isotropic models. An isotropic variogram is invariant to the direction of the separation $x - x'$. Thus an isotropic variogram is a function $\gamma(h)$, defined for $h \geq 0 \in \mathbb{R}$ such that $h = \|x - x'\|^\ell$.

Under the isotropy assumption, the variogram models have the following common behavior: they increase with h and, for most models, when $h \rightarrow \infty$, they stabilize at a certain level. A non-stabilized variogram models a phenomenon whose variability has no limit at large distances. If, conversely, the variogram converges to a limiting value called the **sill**, it means that there is a distance, called the **range**, beyond which $Z(x)$ and $Z(x + h)$ are uncorrelated. In some sense, the range gives some meaning to the concept of area of influence. Another parameter of a variogram that can be physically interpreted is the **nugget effect**: it is the value taken by the variogram when h tends to 0. A discontinuity at the origin is generally due to geological discontinuities, measurement noise or positioning errors. Figure 1 shows a qualitative standard variogram graph where the sill, the range and the nugget effect are represented.

Beyond this standard shape, other physical phenomena can be modelled in a variogram. For instance, the *hole effect*, understood as the tendency for high values to be surrounded by low values, is modelled by bumps on the variogram (or holes in the covariance function). *Periodicity*, which is a special case of hole effect can appear in the variogram. Explicit formulations of many popular variogram or covariance function models can be found in [10].

Usual variogram models do not perfectly match the dependence structure corresponding to the geostatistician's physical intuition and sample variogram analysis. Generally, a linear combination of variograms is used, in order to obtain a more satisfying fitting of the theoretical variogram with the sample variogram and the

geostatistician's intuition. Such a variogram is obtained by :

$$\gamma(h) = \sum_{j=1}^J \gamma_j(h).$$

The main reason is that such linear combinations preserve the negative definiteness conditions requested for variograms, as seen in the next subsection.

Moreover, when the variogram varies with the direction of the separation $x - x'$, it is said to be anisotropic. Some particular anisotropic variograms can be derived from marginal models. The most simple procedure to construct an anisotropic variogram on \mathbb{R}^ℓ is to compute the product of its marginal variograms, assuming the separability of the anisotropic variogram.

3.2 Definiteness properties of covariance and variogram functions

Mathematically, variograms, covariance functions are strongly constrained. Being extensions of the variance, some of its properties are propagated to mathematical definitions of covariance and variogram. In particular, the positive definiteness of the covariance function and similarly the conditional negative definiteness of the variogram are inherited from the positivity of variances.

The variance of linear combinations of random variables $\{Z(x_i), i = 1, \dots, p\}$, given by $\sum_{i=1}^p \mu_i Z(x_i)$, could become negative if the chosen covariance function model were not positive definite or similarly if the chosen variogram model were not conditionally negative definite [2].

When considering an *SRF*, the variance of linear combinations of random variables $\{Z(x_i), i = 1, \dots, p\}$ is expressed, in terms of the covariance function of the form $C(h)$, by

$$\mathbb{V} \left[\sum_{i=1}^p \mu_i Z(x_i) \right] = \sum_{i=1}^p \sum_{j=1}^p \mu_i \mu_j C(x_j - x_i). \quad (7)$$

Since the variance is positive, the covariance function C should be positive definite in the sense of the following definition:

Definition 1 (Positive definite function) *A real function $C(h)$, defined for any $h \in \mathbb{R}^\ell$, is positive definite if, for any natural integer p , any set of real ℓ -tuples $\{x_i, i = 1, \dots, p\}$ and any real coefficients $\{\mu_i, i = 1, \dots, p\}$,*

$$\sum_{i=1}^p \sum_{j=1}^p \mu_i \mu_j C(x_j - x_i) \geq 0.$$

Now in the case of a general *IRF*, i.e. an *IRF* with no covariance function (1) of the form $C(h)$, it can be shown [10] that the variance of any linear combination of increments of random variables $\sum_{i=1}^p \mu_i (Z(x_i) - Z(x_0))$ can be expressed, under the condition that $\sum_{i=1}^p \mu_i = 0$, by

$$\mathbb{V}\left[\sum_{i=1}^p \mu_i (Z(x_i) - Z(x_0))\right] = \mathbb{V}\left[\sum_{i=1}^p \mu_i Z(x_i)\right] = -\sum_{i=1}^p \sum_{j=1}^p \mu_i \mu_j \gamma(x_j - x_i). \quad (8)$$

Let us remark that for an *SRF*, under the condition $\sum_{i=1}^p \mu_i = 0$, expressions (7) and (8) can easily be switched by means of relation (3).

Since the variance is positive, the variogram γ should be conditionally negative definite in the sense of the following definition:

Definition 2 (Conditionally negative definite function) *A function $\gamma(h)$, defined for any $h \in \mathbb{R}^\ell$, is conditionally negative definite if, for any choice of p , $\{x_i, i = 1, \dots, p\}$ and $\{\mu_i, i = 1, \dots, p\}$, conditionally to the fact that $\sum_{i=1}^p \mu_i = 0$,*

$$\sum_{i=1}^p \sum_{j=1}^p \mu_i \mu_j \gamma(x_j - x_i) \leq 0.$$

From expression (7), the covariance function of any *SRF* is necessarily positive definite. Moreover, it can be shown that, from any positive covariance function, there exists a Gaussian random function having this covariance function. But some types of covariance functions are incompatible with some classes of random functions [1]. Note that the same problem holds for variograms and conditional negative definiteness for *IRF*. This problem, which is not solved yet, was dubbed “internal consistency of models” by Matheron [44, 45].

Since the covariance function of any *SRF* is necessarily positive definite, it means that any function that is not positive definite (resp. conditionally negative definite) cannot be the covariance of an *SRF* (resp. the variogram of an *IRF*).

3.3 Why not use the sample variogram ?

The estimation of spatial dependencies by means of the variogram or the covariance function is the key to any kriging method. The intuition underlying spatial dependencies is that points $x \sim y$ that are close together should have close values $Z(x) \sim Z(y)$ because the physical conditions are similar at those locations.

In order to make this idea more concrete, it is interesting to plot the increments $|z(x_i) - z(x_j)|$, quantifying the closeness $z(x_i) \sim z(x_j)$, as a function of the distance $r_{ij} = \|x_i - x_j\|^\ell$, that measures the closeness $x_i \sim x_j$.

The *variogram cloud* is among the most popular visualization tools used by the geostatisticians. It plots the empirical distances r_{ij} on the x-axis against the halved squared increments $v_{ij} = \frac{1}{2} (z(x_i) - z(x_j))^2$ on the y-axis. The choice of the halved squared increments is due to the definition of the variogram of an *IRF* (2).

Figure 2 shows the variogram cloud (in blue) obtained with observations taken from the Jura dataset available on the website <http://goovaerts.pierre.googlepages.com/>. This dataset is a benchmark used all along Goovaerts book [26]. This dataset presents concentrations of seven pollutants (cadmium, cobalt, chromium, copper,

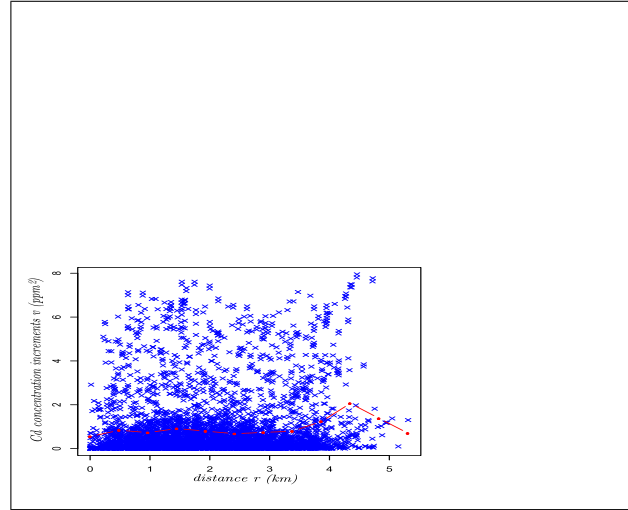


Figure 2 Variogram cloud and sample variogram

nickel, lead and zinc) measured in the French Jura region. On Figure 2, the distance is the Euclidean distance in \mathbb{R}^2 and the variogram cloud has been computed from cadmium concentrations at 100 locations.

From the variogram cloud it is possible to extract the *sample variogram*. It is obtained by computing the mean value of the halved squared increments v in classes of distance. The sample variogram can be defined by:

$$\hat{\gamma}(h) = \frac{1}{2|V_{\Delta}^h|} \sum_{i,j \in V_{\Delta}^h} (z(x_i) - z(x_j))^2,$$

where V_{Δ}^h is the set of pairs of locations such that $\|x_i - x_j\| \in [h - \Delta, h + \Delta]$. $|V_{\Delta}^h|$ is the cardinality of V_{Δ}^h , i.e. the number of pairs in V_{Δ}^h .

Figure 2 shows the sample variogram associated to the plotted cloud variogram. It has been computed for 12 sampling locations and for a class radius Δ equal to half the sampling distance.

As seen in the previous sections, geostatistics relies on sophisticated statistical models, but, in practice, geostatisticians eventually quantify these dependencies by means of a subjectively chosen theoretical variogram. Why don't they try to use the empirical variogram in order to quantify the influence of the neighborhood of a point on the value at this point? *It turns out that these empirical tools (variogram cloud or sample variogram) generally do not fulfil the conditional negative definite requirement.* In order to overcome this difficulty, two methods are generally considered: either an automated fitting (by means of a regression analysis on the parameters of a variogram model) or manual fitting made at a glance. Empirical variograms are considered by the geostatisticians only as visualization or preliminary guiding tools.

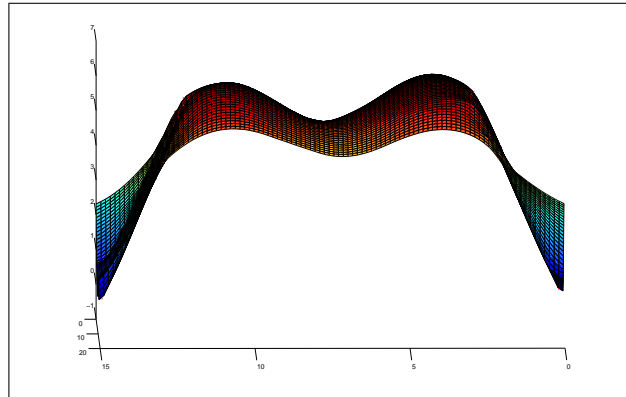


Figure 3 Kriging with a short-ranged variogram

3.4 Sensitivity of kriging to variogram parameters

The kriging parameters, i.e. range, sill and nugget effect affect the results of kriging in various ways. For one thing, while the kriging weights sum to 1, they are not necessarily all positive. In particular, the choice of the range of the variogram will affect the sign of the kriging weights.

In figures 3 and 4 we consider a set of data points that form two significantly separated clusters : there are many data-points between abscissae 0 and 5 with an increasing trend, as well as between 10 and 15 with a decreasing trend, but none between 5 and 10. In one cluster the data suggests an increasing function, and a decreasing function in the other one. Figure 3 is the result of kriging with a short-ranged variogram that only covers the area in each cluster of points. Figure 4 is the result of kriging with a long-ranged variogram covering the two clusters. In the first case, the range of the variogram does not cover the gap between the clusters. The kriged values get closer to the mean value of the data points for kriged values of locations far away from these points. This effect creates a hollow between the clusters at the center of the gap between them. The kriging weights are then all positive. On the contrary, in the second case, the general trend of the data suggests a hill, which is accounted for by the results of kriging, and can only be achieved through negative kriging weights between the clusters of data points

A positive nugget effect may prevent the kriged surface from coinciding with the data points. The effect of changing the sill is less significant. Nevertheless, it is clear that the choice of the theoretical variogram parameters has a non-negligible impact on the kriged surface.

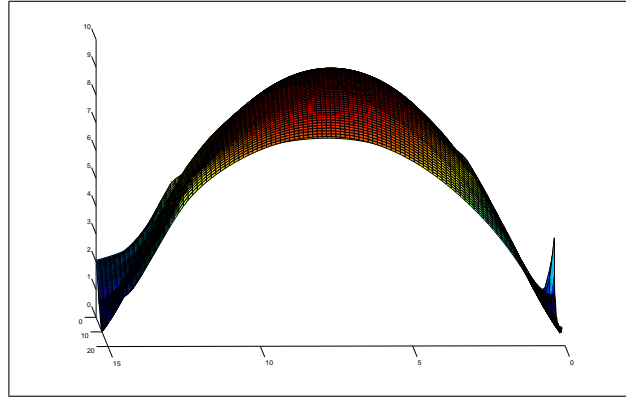


Figure 4 Kriging with a long-ranged variogram

4 Epistemic uncertainty in kriging

The traditional kriging methodology is idealized in the sense that it assumes more information than already available. The stochastic environment of the kriging approach is in some sense too heavy compared to the actual available data, which is scarce. Indeed, the actual data consists of a single realization to the presupposed random function. This issue has been addressed in critiques of the usual kriging methodology. In the kriging estimation procedure, epistemic uncertainty clearly lies in two places of the process: the knowledge of data points and the choice of the mathematical variogram. One source of global uncertainty is the lack of knowledge on the ideal variogram that is used in all the estimation locations of a kriging application. Such uncertainty is global, in the sense that it affects the random function model over the whole kriging domain. This kind of global uncertainty, to which Bayesian approaches can be applied, contrasts with some local uncertainty that may pervade the observations. In the usual approaches (Bayesian or not), these observations are supposed to be perfect, because they are modelled as precise values. However in the 1980's, some authors were concerned by the fact that epistemic uncertainty also pervades the available data, which are then modelled by means of intervals or fuzzy intervals.

Besides, the impact of epistemic uncertainty on the kriged surface should not be confused with the measure of precision obtained by the kriging variance $\mathbb{V}[Z(x_0) - Z^*(x_0)]$. This measure of precision just reflects the lack of statistical validity of kriging estimates at locations far from the data, under the assumption that the real spatial phenomenon is faithfully captured by a random function (which is not the case). The fact that the kriging variance does not depend on the measured data in a direct way makes it totally inappropriate to account for epistemic uncertainty on measurements. Moreover epistemic uncertainty on variogram parameters leads to uncertainty about the kriging variance itself.

4.1 Imprecision in the variogram

Sample variograms (see for instance Figure 2) are generally far from the ideal theoretical variogram models (see for instance Figure 1) fulfilling the conditional negative definite condition. Whether the fitting is automatic (by means of a regression analysis on the parameters of a model) or the fitting is manual and made at a glance, an important epistemic transfer can be noticed. Indeed, whatever the method, the geostatistician tries to summarize some objective information (the sample variogram) by means of a unique subjectively chosen dependence model, the theoretical variogram. As pointed out by A. G. Journel [34]:

Any serious practitioner of geostatistics would expect to spend a good half of his or her time looking at all faces of a data set, relating them to various geological interpretations, prior to any kriging.

Except in [5, 6], this fundamental step of the kriging method is never quite discussed in terms of the epistemic uncertainty it creates. Intuitively, however, there is a lack of information to properly assess a single variogram. This lack of information is a source of epistemic uncertainty, by definition [30]. As the variogram model plays a critical role in the calculation of the reliability of a kriging estimation, the epistemic uncertainty on the theoretical variogram fit should not be neglected. Forgetting about epistemic uncertainty in the variogram parameters, as propagated to the kriging estimate, may result in underestimated risks and a false confidence in the results.

4.2 Kriging in the Bayesian framework

The Bayesian kriging approach is supposed to handle this subjective uncertainty about features of the theoretical variogram, as known by experts. In practice, the structural (random function) model is not exactly known beforehand and is usually estimated from the very same data from which the predictions are made. The aim of Bayesian kriging is to incorporate epistemic uncertainty in the model estimation and thus in the associated prediction.

In Omre [46], the user has a guess on the non stationary random function Z . This guess is given by a random function Y on the domain \mathcal{D} whose moments are known and given by, $\forall x, x+h \in \mathcal{D}$,

$$\begin{cases} \mathbb{E}[Y(x)] = m_Y, \\ Cov[Y(x), Y(x+h)] = C_Y(h). \end{cases} \quad (9)$$

From the knowledge of $C_Y(h)$, the variogram can also be used thanks to the relation $\gamma_Y(h) = C_Y(0) - C_Y(h)$.

The random function Y , and more precisely functions m_Y , C_Y and γ_Y , is the available prior subjective information about the random function Z whose value must

be predicted at location x_0 . In the Bayesian updating procedure of uncertainty, how uncertainty about Y is transferred to Z , is modelled by the law that handles the uncertainty on Z conditionally to Y , i.e. the law of $Z|Y$. In our context, the covariance function or the variogram of the updating law have to be estimated. They are defined by:

$$\begin{cases} C_{Z|Y}(h) = \text{Cov}[Z(x), Z(x+h)|Y(x'); x' \in \mathcal{D}], \\ \gamma_{Z|Y}(h) = \frac{1}{2} \mathbb{V}[Z(x) - Z(x+h)|Y(x'); x' \in \mathcal{D}]. \end{cases} \quad (10)$$

From standard works on statistical Bayesian methods [7, 29], Omre extracts the Bayes updating rules for the bivariate characteristic functions of random functions, which are the variogram and the covariance function. The Bayesian updating rules that enable to compute the posterior uncertainty on Z from the prior uncertainty on Y provided by (9). The updating law (10) is equivalently obtained by:

$$\begin{cases} m_Z = a_0 + m_Y, \\ C_Z(h) = C_{Z|Y}(h) + C_Y(h), \\ \gamma_Z(h) = \gamma_{Z|Y}(h) + \gamma_Y(h), \end{cases}$$

where a_0 is an unknown constant, which is (according to Omre) introduced to make the guess less sensitive to the actual level specified, i.e. less sensitive to the assessment of m_Y .

From this updating procedure of the moments, one can retrieve the moments of Z needed for kriging. What is missing in this procedure are the covariance function or the variogram of $Z|Y$ defined by (10). Omre proposes a usual fitting procedure to estimate these functions. As a preliminary, we can observe that

$$\begin{aligned} \gamma_{Z|Y}(h) &= \gamma_Z(h) - \gamma_Y(h), \\ &= \frac{1}{2} \mathbb{V}[Z(x) - Z(x+h)] - \gamma_Y(h), \\ &= \frac{1}{2} \mathbb{E}[(Z(x) - Z(x+h))^2] - \frac{1}{2} (m_Y(x) - m_Y(x+h))^2 - \gamma_Y(h). \end{aligned}$$

A sample variogram is thus defined by

$$\hat{\gamma}_{Z|Y}(h) = \frac{1}{2|V_\Delta^h|} \sum_{i,j \in V_\Delta^h} (z(x_i) - z(x_j))^2 - (m_Y(x_i) - m_Y(x_j))^2 - 2\gamma_Y(h),$$

where V_Δ^h is the set of pairs of locations such that $\|x_i - x_j\| \in [h - \Delta, h + \Delta]$. $|V_\Delta^h|$ is the cardinality of V_Δ^h , i.e. the number of pairs in V_Δ^h .

Eventually, the Bayesian kriging system is given by

$$C_{Z|Y}(x_0 - x_i) + C_Y(x_0 - x_i) = \sum_{j=1}^n \lambda_j(x_0) C_{Z|Y}(x_i - x_j) + C_Y(x_i - x_j), \quad \forall i = 1, \dots, n.$$

Note that the rationale for this approach is not that the variogram of $Z|Y$ is easier to estimate than the one of Z . The above procedure tries to account for the epistemic uncertainty about the variogram, and the Bayesian approach does it by correcting a prior guess on the random function, then by correcting via a conditional term. Another Bayesian approach is proposed in the paper of Handcock and Stein [27]. It shows that ordinary kriging with a Gaussian stationary random function and unknown mean m can be interpreted in terms of Bayesian analysis with a prior distribution locally uniform on the mean parameter m . More generally, they propose a systematic Bayesian analysis of the kriging methodology for different mean and variogram parametric models. Other authors developed this approach [14, 25, 9]. It is supposed to take into account epistemic uncertainty in the sense that it is supposed to handle the lack of knowledge on the model parameters by assigning a prior probability distribution to these parameters.

To our view, a unique prior distribution, even if claimed to be non informative in the case of plain ignorance, is not the proper representation to capture epistemic uncertainty on the model. A unique prior models the supposedly known variability of the considered parameter, not ignorance about it. In fact it is not clear that such parameters are subject to variability. As a more consistent approach, a robust Bayesian analysis of the kriging could be performed. Robust Bayesian analysis consists of working with a family of priors in order to lay bare the sensitivity of estimators to epistemic uncertainty on the model's parameters [8, 48].

4.3 Imprecision in the data

Because available information can be of various types and qualities, ranging from measurement data to human geological experience, the treatment of uncertainty in data should reflect this diversity of origin. Moreover, there is only one observation made at each location, and this value is in essence deterministic. However one may challenge the precision or accurateness of such measurements. Especially, geological measurements are often highly imprecise.

Let us take a simple example: the measurement of permeability in an aquifer. It results from the interpretation of a pumping test: when pumping water from a well, the water level will decrease in that well and also in neighboring wells. The local permeability is obtained by fitting theoretical draw-down curves to the experimental ones. There is obviously some imprecision in such fitting that is based on approximations to the reality (e.g., homogeneous medium). Epistemic uncertainty due to measurement imperfections should pervade the measured permeability data. For the inexact (imprecise) information resulting from unique assessments of deterministic values, a nonfrequentist or subjective approach reflecting imprecision could be used.

Epistemic uncertainty about such deterministic numerical values naturally takes the form of intervals. Asserting $z(x) \in [a, b]$ comes down to claiming that the actual value of a quantity $z(x)$ lies between a and b . Note that while $z(x)$ is an objective quantity, the nature of the interval $[a, b]$ is epistemic, it represents expert knowledge

about $z(x)$ and has no existence per se. The interval $[a, b]$ is a set of mutually exclusive values one of which is the right one: the natural interpretation of the interval is that $z(x) \notin [a, b]$ is considered impossible.

A fuzzy subset F [21, 54] is a richer representation of the available knowledge in the sense that the membership degree $F(r)$ is a gradual estimation of the conformity of the value $z(x) = r$ to the expert knowledge. In most approaches, fuzzy sets are representations of knowledge about underlying precise data. The membership grade $F(r)$ is interpreted as a degree of possibility of $z(x) = r$ according to the expert [55]. In this setting, membership functions are interpreted as possibility distributions that handle epistemic uncertainty due to imprecision on the data.

Possibility distributions can often be viewed as nested sets of confidence intervals [18]. Let $F_\alpha = \{r \in \mathbb{R} : F(r) \geq \alpha\}$ be called an α -cut. F is called a fuzzy interval if and only if $\forall 0 < \alpha \leq 1, F_\alpha$ is an interval. When $\alpha = 1, F_1$ is called the *mode* of F if reduced to a singleton. If the membership function is continuous, the degree of certainty of $z(x) \in F_\alpha$ is equal to $1 - \alpha$, in the sense that any value outside F_α has possibility degree α . So it is sure that $z(x) \in S(F) = \lim_{\alpha \rightarrow 0} F_\alpha$ (this is the *support* of F), while there is no certainty that the most plausible values in F_1 contain the actual value. Note that the membership function can be retrieved from its α -cuts, by means of the relation:

$$F(r) = \sup_{r \in F_\alpha} \alpha.$$

Therefore, suppose that the available knowledge supplied by an expert comes in the form of nested confidence intervals $\{I_k, k = 1, \dots, K\}$ such that $I_1 \subset I_2 \subset \dots \subset I_K$ with increasing confidence levels $c_k > c_{k'}$ if $k > k'$, the possibility distribution defined by

$$F(r) = \min_{k=1, \dots, K} \max(1 - c_k, I_k(r)),$$

is a faithful representation of the supplied information. Viewing a possibility degree as an upper probability bound [53], F is an encoding of the probability family $\{P : P(I_k) \geq c_k\}$. If $c_K = 1$ then the support of this fuzzy interval is I_K .

If an expert only provides a mode c and a support $[a, b]$, it makes sense to represent this information as the triangular fuzzy interval with mode c and support $[a, b]$ [19]. Indeed F then encodes a family of (subjective) probability distributions containing all the unimodal ones with support included in $[a, b]$.

5 Intervallist kriging approaches

This section and the next one refers to works done in the 1980's. Even if some of them can be considered obsolete, their interest lies in their being early attempts to handle some form of epistemic uncertainty in geostatistics. While some of the proposed procedures look questionable, it is useful to understand their merits and limitations in order to avoid pitfalls and propose a well-founded methodology to that effect. Since then, it seems that virtually no new approaches have been proposed in

the recent past, even if some of the problems posed more than 20 years ago have now received more efficient solutions, for instance the solving of interval problems via Gibbs sampling [23].

5.1 The quadratic programming approach

In [22, 36], the authors propose to estimate $z^*(x_0)$, from imprecise information available as a set of constraints on the observations. Such constraints can also be seen as inequality-type data, i.e. the observation located at the position x_i is of the form $z(x_i) \geq a(x_i)$ and/or $z(x_i) \leq b(x_i)$.

This approach also assumes a global constraint which is that whatever the position $x_0 \in \mathcal{D}$, the kriging estimate $z^*(x_0)$ is bounded, which can be translated by

$$\forall x_0 \in \mathcal{D}, z^*(x_0) \in [a, b]. \quad (11)$$

For instance any ore mineral grade is necessary a value within $[0, 100\%]$.

Any kind of data, i.e. precise or inequality-type, can always be expressed in terms of an interval constraint:

$$z(x_i) \in [a(x_i), b(x_i)], \forall i = 1, \dots, n. \quad (12)$$

Indeed precise data can be modelled by constrained data (12) with equal upper and lower bound and an inequality-type data $z(x_i) \geq a(x_i)$ (resp. $z(x_i) \leq b(x_i)$) can be expressed as $[a(x_i), b]$ (resp. $[a, b(x_i)]$). Thus the data set is now given by $\mathcal{Z}_n = \{\bar{z}(x_i) = [a(x_i), b(x_i)], i = 1, \dots, n\}$.

As mentioned by A. Journel [35], this formulation of the problem allows to cope with the recurring question of the positiveness of the kriging weights, which the basic kriging approaches cannot ensure. Negative weights are generally seen as being “evil”, due to the fact that the measured spatial quantity is positive and their linear combination (4) with some negative weights could lead to a negative kriging estimate. More generally, nothing prevents the kriged values to violate range constraints induced by practical considerations on the studied quantity. Hence, one is tempted by the incorrect conclusion that all kriging weights should be positive. Actually, having some negative kriging weights is quite useful, since it allows a global kriging estimate to fall outside the range $[\min_i z(x_i), \max_i z(x_i)]$. Instead of forcing the weights to be positive, the constraint-based approach forces the estimate to be positive by adding a constraint on the estimate to the least squares optimization problem. More generally, the global constraint (11), solves the problem of getting meaningful kriging estimates.

In [39], J.L. Mallet proposes a particular solution to the problem of constrained optimization given by means of quadratic programming, i.e. to the problem of minimizing (or maximizing) a quadratic form (the error variance) under the constraint that the solution of this optimization program is inside the range $[a, b]$.

The dual expression [10] of the kriging estimate (4) is given by :

$$z^*(x_0) = \sum_{i=1}^n v_i C(x_i - x_0). \quad (13)$$

This expression is obtained by incorporating in (4), the kriging weights that are the solutions of the kriging equation (6). Thus the dual kriging weights $\{v_i, i = 1, \dots, n\}$ now reflect the dependencies between observations $\{C(x_i - x_j), i, j = 1, \dots, n\}$ and the observations $\{z(x_i), i = 1, \dots, n\}$ ¹.

Built on Mallet's approach [39], Dubrule and Kostov [22, 36] proposed a solution to this interpolation problem, that takes the form (13), where the dual kriging weights $\{v_i, i = 1, \dots, n\}$ are obtained by means of the quadratic program minimizing

$$\sum_{i=1}^n \sum_{j=1}^n v_i v_j C(x_i - x_j), \quad (14)$$

subject to n constraints

$$a(x_i) \leq \sum_{j=1}^n v_j C(x_j - x_i) \leq b(x_i)$$

induced by the dataset $\mathcal{Z}_n = \{\bar{z}(x_i) = [a(x_i), b(x_i)], i = 1, \dots, n\}$. When only precise observations (i.e. when no inequality-type constraint) are present, the system reduces to a standard simple kriging system.

However, the ensuing treatment of these constraints is ad hoc. Indeed, the authors propose to select one bound among $a(x_i), b(x_i)$ for each constraint, namely the one supposed to affect the kriging estimate. They thus select a precise data set made of the selected bounds. The choice of this data set is just influenced by the wishes of the geostatistician in front of the raw data and on the basis of some preliminary kriging steps performed from some available precise data (if any). These limitations can nowadays be tackled using Gibbs sampling methods [23].

5.2 The soft kriging approach

Methodology

In 1986, A. Journé [35] studied the same problem of adapting the kriging methodology in order to deal with what he called "soft" information. According to him,

¹ It can be noted that, in the precise framework, the dual formalism of kriging is computationally interesting. Indeed, the kriging system to be solved is obtained by minimization of (14), whatever the position of estimation x_0 . It means that the kriging system has to be solved only once to provide an interpolation over all the domain. However, this system is difficult to solve and badly conditioned. Whereas the non dual systems, where the matrices' coefficients are generally scarce, are more tractable. Therefore, it should be preferred to solve the dual kriging system in case of a high quantity of estimation points but with a small dataset and it should be preferred to solve the usual kriging system in case of a small number of estimation points but with a large dataset.

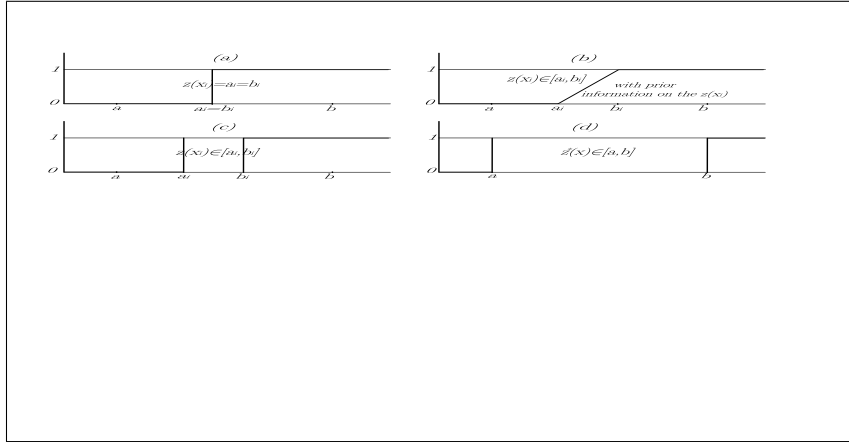


Figure 5 Prior information on the observations

“soft” information consists of imprecise data $\bar{z}(x_i)$, especially intervals, encoded by cumulative distribution functions (cdf) F_{x_i} .

The cumulative distribution function F_{x_i} , attached to a precise value $z(x_i) = a_i = b_i$ can be modelled by a step-function cdf with parameter $a_i = b_i$, i.e.:

$$F_{x_i}(s) = \begin{cases} 1, & \text{if } s \geq a(x_i) = b(x_i), \\ 0, & \text{otherwise.} \end{cases}$$

(cf. Figure 5.(a)). At each location x_i where a constraint interval $\bar{z}(x_i)$ of the form (12) is present, the associated cdf F_{x_i} is only known outside the constraint interval where it is either 0 or 1, i.e. :

$$F_{x_i}(s) = \begin{cases} 1, & \text{if } s \geq b(x_i), \\ 0, & \text{if } s \leq a(x_i), \\ ? & \text{otherwise.} \end{cases} \quad (15)$$

(cf. Figure 5.(c)). If the expert is unable to decide where, within an interval $\bar{z}(x_i) = [a(x_i), b(x_i)]$, the value $z(x_i)$ may lie, a non informative prior cdf (15) should be used, not as a uniform cdf within that interval, as the principle of maximum entropy would suggest, since it is not equivalent to a lack of information.

In addition to the constraint interval $\bar{z}(x_i)$ of Dubrule and Kostov [22, 36], some prior information allows quantifying the likelihood of value $z(x_i)$ within that interval. The corresponding cumulative distribution function F_{x_i} (cf. Figure 5.(b)) is thus completed with prior subjective probabilities.

At any other location, a minimal interval constraint exists (cf. (11) and Figure 5.(d)): $z^*(x) \in [a, b]$. This constraint, as in the quadratic programming approach of Dubrule and Kostov, enables the problem of negative weights to be addressed.

From this set of heterogeneous prior pieces of information, that we will denote by $\mathcal{Z}_n = \{\tilde{z}(x_i) = F_{x_i}, i = 1, \dots, n\}$, Journel [35] proposes to construct a “posterior” cdf at the kriging estimation location x_0 , denoted by

$$F_{x_0|\mathcal{Z}_n}(s) = P(Z(x_0) \geq s | \mathcal{Z}_n).$$

In its simplest version, the so-called “soft” kriging estimate of the “posterior” cdf $F_{x_0|\mathcal{Z}_n}$ is defined as a linear combination of the prior cdf data, for a given threshold value $s \in [a, b]$, i.e.

$$F_{x_0|\mathcal{Z}_n}(s) = \sum_{i=1}^n \lambda_i(x_0, s) F_{x_i}(s), \quad (16)$$

where the kriging weights, for a given threshold $s \in [a, b]$, are obtained by means of usual kriging based on the random function $Y(x) = F_x(s)$ at location x .

Despite its interest, there are some aspects of this approach that are debatable:

1. *The use of Bayesian semantics.* Journel proposes to use the terminology of Bayesian statistics, by means of the term *prior* for qualifying the probabilistic information attached to each piece of data and the term *posterior*, for qualifying the probabilistic information on the estimation point. However, in his approach, the computation of the posterior cdf is not made by means of the Bayesian updating procedure. He probably made this terminological choice because of the subjectivist nature of the information. However, this choice is not consistent with the Bayesian statistics.
2. The choice of a *linear combination of the cdfs* to compute the uncertain estimate. A more technical criticism of his approach concerns the definition of the kriged “posterior” cdf (16). The appropriateness of this definition supposes that the cdf of a linear combination of random variables is the linear combination of cdfs of these random variables. However, this is not correct. Propagating uncertainty on parameters of an operation is not as simple as just replacing the parameters by their cdfs in the operation. Indeed, the cdf of $Z^*(x_0)$ in (4), when $\{Z(x_i), i = 1, \dots, n\}$ are random variables with cdfs given by $\{F_{x_i}, i = 1, \dots, n\}$, is not given by (16), but via a convolution operator that could be approximated by means of a Monte Carlo method. If we assume a complete dependence between measurements of $Z(x_i)$, one may also construct the cdf of $Z^*(x_0)$ as a weighted sum of their quantile functions (inverse of cdf).

These defects make this approach theoretically unclear, with neither an interpretation in the Bayesian framework nor in the frequentist framework. Note that the author [35] already noted the strong inconsistency of his method, namely the fact that the “posterior” cdf (16) does not respect the monotonicity property, inherent to the definition of a cumulative distribution function. Indeed, when some kriging weights are negative, it is not warranted that for $s > s'$, $F_{x_0|\mathcal{Z}_n}(s) > F_{x_0|\mathcal{Z}_n}(s')$. He proposes an ad hoc correction of the kriging estimates, replacing the decreasing parts of $F_{x_0|\mathcal{Z}_n}$ by flat parts.

In spite of these criticisms of the well-foundedness of the Journel’s approach, a basic idea for handling epistemic uncertainty in the data appears in his paper. In-

deed, the way Journel proposes to encode the dataset is the first attempt by some geostatisticians, to our knowledge, to handle incomplete information (or epistemic uncertainty) in kriging. Indeed the question mark in the encoding of a uniform intervallist data (15) is the first modelling of ignorance in geostatistics. This method tends to confuse subjective, Bayesian, and epistemic uncertainty. This confusion can now be removed in the light of recent epistemic uncertainty theories. Interestingly, their emergence [21, 53] occurred when the confusion between subjectivism (de Finetti's school of probability [13]) and Bayesianism began to be clarified.

6 Fuzzy kriging

There are two main fuzzy set counterparts of statistical methods : The first one extends statistical principles like error minimisation, unbiasedness or stationarity to fuzzy set-valued realisations. Such an adaptation of prediction by kriging to triangular fuzzy data was suggested by Diamond [17]. The second one applies the extension principle to the kriging estimate [4, 5, 6] in the spirit of sensitivity analysis.

6.1 Diamond's fuzzy kriging

In the late 1980's Phil Diamond was the first to extend Matheronian statistics to the fuzzy set setting, with a view to handle imprecise data. The idea was to exploit the notion of fuzzy random variables which had emerged a few years earlier after several authors (see [11] for a bibliography). Diamond's approach relies on the Puri and Ralescu version of fuzzy random variables [47], which is influenced by the theory of random sets developed in the seventies by Matheron himself [42]. Diamond also proposed an approach to fuzzy least squares in the same spirit [15].

6.1.1 Methodology

The data used by Diamond [17] are modelled by triangular fuzzy numbers, because of both their convenience and their applicability in most practical cases. Those triangular fuzzy numbers \hat{T} are defined by their mode T^m and left and right bounds of their support T^- and T^+ . They are then denoted by $\hat{T} = (T^m; T^-, T^+)$. The set of all fuzzy triangular numbers is denoted by \mathcal{T} .

Diamond proposes to work with a distance D_2 on \mathcal{T} that makes the metric space (\mathcal{T}, D_2) complete [17] :

$$\forall \hat{A}, \hat{B} \in \mathcal{T}, D_2(\hat{A}, \hat{B}) = (A^m - B^m)^2 + (A^- - B^-)^2 + (A^+ - B^+)^2.$$

A Borel σ -algebra \mathcal{B} can be constructed on this complete metric space. This allows the definition of fuzzy random variables [47], viewed as mappings from a probability space to a specific set of functions, namely a set $(\mathcal{T}, \mathcal{B})$ of triangular fuzzy random numbers.

The expectation of a triangular fuzzy random number \hat{X} is obtained by extending the concept of Aumann integral [3], defined for random sets, to all α -cuts of \hat{X} .

Definition 3 Let \hat{X} be a triangular fuzzy random number, i.e. a \mathcal{T} -valued random variable, the α -cuts of its expectation, denoted by $\hat{\mathbb{E}}[\hat{X}]$, are given by:

$$\forall \alpha \in [0, 1], (\hat{\mathbb{E}}[\hat{X}])^\alpha = \mathbb{E}_{Aumann}[\hat{X}^\alpha]$$

It can be shown that the expected value of a triangular fuzzy random number X is a triangular fuzzy number, that will be denoted by $\hat{\mathbb{E}}[\hat{X}] = (\mathbb{E}[X]^m; \mathbb{E}[X]^-, \mathbb{E}[X]^+)$.

From those definitions, Diamond proposes to extend the concept of random function to triangular fuzzy random functions, which are \mathcal{T} -valued random functions. He proposes to work with second-order stationary triangular fuzzy random functions \hat{Z} , that verify, $\forall x, x+h \in \mathcal{D}$,

$$\begin{cases} \hat{\mathbb{E}}[\hat{Z}(x)] = (M^m; M^-, M^+) = \hat{M}, \\ \text{Cov}(\hat{Z}(x), \hat{Z}(x+h)) = (C^m(h); C^-(h), C^+(h)) = \hat{C}(h), \end{cases}$$

where the triangular fuzzy expected value is constant on \mathcal{D} and the triangular fuzzy covariance function is defined by :

$$\begin{cases} C^m(h) = \mathbb{E}[Z^m(x)Z^m(x+h)] - (M^m)^2 \\ C^-(h) = \mathbb{E}[Z^-(x)Z^-(x+h)] - (M^-)^2 \\ C^+(h) = \mathbb{E}[Z^+(x)Z^+(x+h)] - (M^+)^2 \end{cases} \quad (17)$$

Now, from this definition of fuzzy covariance function, the problem is to predict the value of the regionalized triangular fuzzy random variable $\hat{Z}(x_0)$ at x_0 . For this prediction the following linear estimator is used

$$\hat{z}^*(x_0) = \bigoplus_{i=1}^n \lambda_i(x_0) \hat{z}(x_i),$$

where $\{\hat{z}(x_i), i = 1, \dots, n\}$ are fuzzy data located on precise locations $\{x_i, i = 1, \dots, n\}$, \bigoplus is the extension of the Minkowski addition of intervals to fuzzy triangular numbers.

The set of precise kriging weights $\{\lambda_i(x_0), i = 1, \dots, n\}$ is obtained by minimization of the precise mean squared error $D = \mathbb{E}[D_2(\hat{Z}^*(x_0), \hat{Z}(x_0))^2]$. The unbiasedness condition is extended to fuzzy quantities induces the usual condition $\sum_{i=1}^n \lambda_i(x_0) = 1$ on kriging weights.

Due to the form of distance D_2 , the expression to be minimized, along the same line as simple kriging, can be expressed by :

$$D = \sum_{i=1}^n \sum_{j=1}^n \lambda_i(x_0) \lambda_j(x_0) \mathbb{C}(x_i - x_j) - 2 \sum_{j=1}^n \lambda_j(x_0) \mathbb{C}(x_0 - x_j) + \mathbb{C}(x_0 - x_0), \quad (18)$$

with $\mathbb{C}(x_i - x_j) = C^m(x_i - x_j) + C^-(x_i - x_j) + C^+(x_i - x_j)$, $\forall i, j = 0, \dots, n$. The minimization of the error (18) leads to the following kriging system:

$$\begin{cases} \sum_{j=1}^n \lambda_j(x_0) \mathbb{C}(x_i - x_j) - \mathbb{C}(x_0 - x_i) - \theta - L_i = 0, \forall i = 1, \dots, n \\ \sum_{i=1}^n \lambda_i(x_0) = 1 \\ \sum_{i=1}^n L_i \lambda_i(x_0) = 0 \\ L_i, \lambda_i(x_0) \geq 0, \forall i = 1, \dots, n. \end{cases}$$

Where L_1, L_2, \dots, L_n and θ are Lagrange multipliers which allow, under Kuhn-Tucker conditions, solving the optimization program for finding the set of kriging weights $\{\lambda_i(x_0), i = 1, \dots, n\}$ minimizing the error D . It should be noted that, in 1988, i.e. one year before the publication of his fuzzy kriging article, Philip Diamond published the same approach, restricted to interval data [16].

6.1.2 Discussion

Despite its mathematical rigor, there are several aspects of this approach that are debatable:

1. the shift from a random function to a fuzzy valued random function,
2. the choice of a scalar distance D_2 between fuzzy quantities,
3. the use of a Hukuhara difference in the computation of fuzzy covariance (17).

1. The first point presupposes a strict adherence to the Matheron school of geostatistics. However, it makes the conceptual framework (both at the conceptual and practical level) even more difficult to grasp. The metaphor of a fuzzy random field looks like an elusive artefact. The fuzzy random function is a mere substitute to a random function, and leads to a mathematical model with more parameters than the standard kriging technique. The key question is then: does it properly handle epistemic uncertainty?

2. The choice of a precise distance between fuzzy intervals is in agreement with the use of a precise variogram and it leads to a questionable way of posing the least square problem.

First, a precise distance is used to measure the variance of the difference between the triangular fuzzy random variables $\hat{Z}(x_0)$ and $\hat{Z}^*(x_0)$. This is in contradiction with using a fuzzy-valued covariance when defining the stationarity of the triangular fuzzy random function $\hat{Z}(x)$. Why not then define the covariance between the fuzzy random variables $\hat{Z}(x)$ and $\hat{Z}(x+h)$ as $\mathbb{E}[D_2(\hat{Z}(x), \hat{M})D_2(\hat{Z}(x+h), \hat{M})]$, i.e. like

the variance of $\hat{Z}(x_0) - \hat{Z}^*(x_0)$? Stationarity should then be expressed as $\mathbb{C}(h) = \mathbb{E}[D_2(\hat{Z}(x), \hat{M})D_2(\hat{Z}(x+h), \hat{M})]$.

However, insofar as fuzzy sets represent epistemic uncertainty, the fuzzy random function might represent a fuzzy set of possible standard random functions, one of which is the right one. Then, the scalar variance of a fuzzy random variable based on distance D_2 evaluates the precise variability of functions representing *the knowledge* about ill-known crisp (i.e. fuzzy) realizations. However, it does not evaluate the imprecise knowledge about the variability of the underlying precise realizations [11]. The meaning of extracting a precise variogram from fuzzy data and of the problem of minimizing the scalar variance of the membership functions (18) remains unclear. To our opinion, the approach of Diamond is not cogent for handling epistemic uncertainty. In [11] a survey of possible notions of variance of fuzzy random variables, with discussions on their significance in the scope of epistemic uncertainty is proposed. It is argued that if a fuzzy random variable represents epistemic uncertainty, its variance should be imprecise or fuzzy as well.

3. The definition of second-order stationarity for triangular fuzzy random functions is highly questionable. The fuzzy covariance function $\hat{C}(h)$ (17) proposed by Diamond is supposed to reflect the epistemic uncertainty on the covariance between $\hat{Z}(x)$ and $\hat{Z}(x+h)$, which finds its source in the epistemic uncertainty conveyed by \hat{Z} . In his definition (17) of $\hat{C}(h)$, Diamond uses the Hukuhara difference \boxminus [31] between supports of triangular fuzzy numbers

$$(\mathbb{E}[Z^m(x)Z^m(x+h)]; \mathbb{E}[Z^-(x)Z^-(x+h)], \mathbb{E}[Z^+(x)Z^+(x+h)])$$

and \hat{M} . The Hukuhara difference between two intervals is of the form $[a, b] \boxminus [c, d] = [a-c, b-d]$. Note that, the result may be such that $a-c > b-d$, i.e. not an interval. So, it is not clear that the inequalities $C^-(h) \leq C^m(h) \leq C^+(h)$ always hold when computing $\mathbb{E}[\hat{Z}(x)\hat{Z}(x+h)] \boxminus \hat{M}^2$.

The Hukuhara difference [31] \boxminus between intervals is defined such that,

$$[a, b] \boxminus [c, d] = [u, v] \iff [a, b] = [c, d] \oplus [u, v] = [c+u, d+v],$$

where \oplus is the usual Minkowski addition of intervals.

This property of the Hukuhara difference allows interpreting the epistemic transfer induced by this difference in the covariance definition of Diamond. In the standard case, the identity $\mathbb{E}[Z(x) - m][Z(x+h) - m] = \mathbb{E}[Z(x)Z(x+h)] - m^2 = C(h)$ holds. When extending it to the fuzzy case in Diamond method, it is assumed that :

- $\hat{Z}(x)\hat{Z}(x+h)$ and \hat{M}^2 are triangular fuzzy intervals when $\hat{Z}(x)$ and \hat{M} are such. This is only a coarse approximation.
- $[\mathbb{E}[Z^-(x)Z^-(x+h)], \mathbb{E}[Z^+(x)Z^+(x+h)]] = [C^-(h), C^+(h)] \oplus [M^-, M^+]$, so that the imperfect knowledge about $\hat{C}(h) \boxminus \hat{M}^2$ is identified to the imperfect knowledge about $\hat{\mathbb{E}}[\hat{Z}(x)\hat{Z}(x+h)]$. An alternative definition is to let

$$[\mathbb{E}[Z^-(x)Z^-(x+h)], \mathbb{E}[Z^+(x)Z^+(x+h)]] \ominus [M^-, M^+] = [C^-(h), C^+(h)],$$

using Minkowski difference of fuzzy intervals \ominus instead of Hukuhara difference in equation (17). It would ensure that the resulting fuzzy covariance is always a fuzzy interval, but it would be more imprecise. Choosing between both expressions require some assumption about the origin of epistemic uncertainty in this calculation.

- Besides, stating the fuzzy set equality $\hat{C}(h) = \hat{\mathbb{E}}[(\hat{Z}(x) - \hat{\mathbb{E}}(Z(x)))(\hat{Z}(x+h) - \hat{\mathbb{E}}(Z(x+h)))]$ does not enforce the equality of the underlying quantities on each side.

Finally, the Diamond approach precisely interpolates between fuzzy observations at various locations. Hence, the method does not propagate the epistemic uncertainty bearing on the variogram. Albeit fuzzy kriging provides a fuzzy interval estimate $\hat{z}^*(x_0)$, it is difficult to interpret this fuzzy estimate as picturing our knowledge about the actual $z^*(x_0)$ one would have obtained via kriging if the data had been precise. Indeed, the scalar influence coefficients in Diamond method reflect both the spatial variability of Z and the variability of *the epistemic uncertainty* of observations. This way of handling intervals or fuzzy intervals as “real” data is in fact much influenced by Matheron’s random sets where set realizations are understood as real objects (geographical areas), not as imprecise information about precise locations. The latter view of sets as epistemic constructs is more in line with Shafer’s theory of evidence [49], which also uses the formalism of random sets, albeit with the purpose of grasping incomplete information.

Overall, viewed from the point of view of epistemic uncertainty, this approach to kriging looks questionable both at the philosophical and computational levels. Nevertheless the technique has been used in practical applications [52] by Taboada et al. in a context of evaluation of reserves in an ornamental granite deposit in Galicia in Spain.

6.2 Bardossy’s fuzzy kriging

Not only may the epistemic uncertainty about the data $z(x_i)$ be modelled by intervals or fuzzy intervals, but one may argue that the variogram itself in its mathematical version should be a parametric function with interval-valued or fuzzy set-valued parameters. While Diamond was proposing a highly mathematical approach to fuzzy kriging, Bardossy et al. [4, 5, 6] between 1988 and 1990 also worked on this issue of extending kriging to epistemic uncertainty caused by fuzzy data. Beyond this adaptation of the kriging methodology to fuzzy data, they also propose in their method to handle epistemic uncertainty on the theoretical variogram model.

In their approach, the variogram is tainted with epistemic uncertainty because the parameters of the theoretical variogram model are supposed to be fuzzy subsets. The epistemic uncertainty of geostatisticians regarding these parameters is then propagated to the variogram by means of the extension principle. Introduced by Lotfi Zadeh [54], it provides a general method for extending non fuzzy models or functions in order to deal with fuzzy parameters. For instance, fuzzy set arithmetics [21],

that generalizes interval arithmetics, has been developed by applying the extension principle to the classical arithmetic operations like addition, subtraction...

Definition 4 *If U, V and W are sets, and f is a mapping from $U \times V$ to W . Let A be a fuzzy subset on U with a membership function also denoted by μ_A , likewise a fuzzy set B on V . The image of (A, B) in W by the mapping f is a fuzzy subset C on W whose membership function is obtained by:*

$$\mu_C(w) = \sup_{(u,v) \in U \times V | w=f(u,v)} \min(\mu_A(u), \mu_B(v)).$$

In terms of possibility theory, it comes down to computing the degree of possibility $\Pi(f^{-1}(w)), w \in W$. Actually, in their approach, Bardossy et al. do not directly use such a fuzzy variogram model in the kriging process. Their approach is, in a sense, more global since they propose to apply the extension principle, not only to the variogram model, but to the entire inversed kriging system and to the obtained kriging estimate $z^*(x_0)$, because it is a function of the observations $\{z(x_i), i = 1, \dots, n\}$, of the parameters of the variogram model $\{a_j, j = 1, \dots, p\}$ and of the estimation position x_0 . In other words, they express the kriging estimate as

$$z^*(x_0) = f(z(x_1), \dots, z(x_n), a_1, \dots, a_p, x_0),$$

and they apply the extension principle to propagate the epistemic uncertainty of the fuzzy observations $\{\hat{z}(x_i), i = 1, \dots, n\}$ and of the fuzzy parameters of the variogram model $\{\hat{a}_j, j = 1, \dots, p\}$ to the kriging estimate $\hat{z}^*(x_0)$. They propose to numerically solve the optimisation problem induced by their approach, without providing details.

This approach is more consistent with the epistemic uncertainty involved in the kriging methodology than the Diamond's method. However, there does not seem to be a tractable solution that can be applied to large datasets because of the costly optimisation involving fuzzy data. The question whether the epistemic uncertainty conveyed in an imprecise variogram is connected or not to the epistemic uncertainty about the data is worth considering. However, even in the presence of a precise dataset, one may argue that the chosen variogram is tainted with epistemic uncertainty that only the expert, who chooses it, could estimate.

7 Uncertainty in kriging: a prospective discussion

The extensions of kriging studied above may lead to a natural questioning about the nature of the uncertainty that pervades this interpolation method. Indeed, taking into account this kind of imperfect knowledge suggests, in the first stance, that the usual approach does not properly handle the available information. Being aware that information is partially lacking is in itself a piece of (meta-)information. Questioning the proper handling of uncertainty in kriging leads to examine two issues:

- Is the random function model proposed by Matheron and followers cogent in spatial prediction?
- How to adapt the kriging method to epistemic uncertainty without making the problem intractable?

These questions seem require a reassessment of the role of probabilistic modeling in the kriging task, supposed to be of an interpolative nature, while it heavily relies on the use of least squares methods that are more central to regression techniques than to interpolation per se.

7.1 *Spatial vs. fictitious variability*

It is commonly mentioned that probabilistic models are natural representations of phenomena displaying some form of variability. Repeatability is the central feature of the idea of probability as pointed out by Shafer and Vovk [50]. This is embodied by the use of probability trees, Markov chains and the notion of sample space. A random variable $V(\omega)$ is a mapping from a sample space Ω to the real line, and variability is captured by binding the value of V to the repeated choices of $\omega \in \Omega$, and the probability measure that equips Ω summarizes the repeatability pattern.

In the case of the random function approach to geostatistics, the role of this scenario is not quite clear. Geostatistics is supposed to handle spatial variability of a numerical quantity $z(x)$ over some geographical area \mathcal{D} . Taken at face value, spatial variability means that when the location $x \in \mathcal{D}$ changes, so does $z(x)$. However, when x is fixed $z(x)$ is a precise deterministic value. Strictly speaking, these considerations would lead us to identify the sample space with \mathcal{D} , equipped with the Lebesgue measure.

However, the classical geostatistics approach after Matheron is at odds with this simple intuition. It postulates the presence of a probability space Ω such that the quantity z depends on both x and $\omega \in \Omega$. z is taken as a random function: for each x , the actual value $z(x)$ is substituted with a random variable $Z(x)$ from a sample space Ω to the real line. The probability distribution of $Z(x)$ is thus attached to the quantity of interest $z(x)$ at a location x . It implicitly means that this quantity of interest is variable (across ω) and that you can quantify the variability of this value.

In the spatial interpolation problem solved by kriging, this kind of postulated variability at each location x of a spatial domain \mathcal{D} , corresponds to no actual phenomenon. As Chilès and Delfiner [10] (p. 24) acknowledge,

The statement “ $z(x)$ is a realization of a random function $Z(x)$ ” or even “of a stationary random function,” has no objective meaning.

Indeed, the quantity of interest at an estimation site x is deterministic and a *single* observation $z(x_i)$ for a finite set of locations x_i is available. It does not look sufficient to determine a probability distribution at each location x even if each $Z(x)$ were actually tainted with variability.

In fact, geostatisticians consider random functions not as reflecting randomness or variability actually present in natural phenomena, but as a pure mathematical model whose interest lies in the quality of predictions it can deliver. As Matheron said:

Il n'y a pas de probabilité en soi, il y a des modèles probabilistes²

The great generality of the framework, whereby a deterministic spatial phenomenon is considered as a (unique) realisation of a random function is considered to be non-constraining because it cannot be refuted by reality, and is not directly viewed as an assumption about the phenomenon under study. The spatial ergodicity assumption on the random function $Z(x)$ is instrumental to relate its fictitious variability at each location of the domain to the spatial variability of the deterministic quantity $z(x)$. While this assumption is easy to interpret in the temporal domain, it is less obvious in the spatial domain. The role of spatial ergodicity and stationarity assumptions is mainly to offer theoretical underpinnings to the least square technique used in practice. In other words, the random function approach is to be taken as a formal black-box model for data-based interpolation, and has no pretence to represent any real or epistemic phenomenon (beyond the observed data $z(x_i)$). Probability in geostatistics is neither objective nor subjective: it is mathematical.

7.2 A deterministic justification of simple kriging

One way of interpreting random functions in terms of actual (spatial) randomness is to replace pointwise locations by subareas (“blocks”) over which average estimations can be computed. Such blocks must be small enough for ensuring a meaningful spatial resolution but large enough to contain a statistically significant number of measurements. This is called the trade-off between objectivity and spatial resolution. At the limit, using a single huge block, the random function is the same at each point and reflects the variability of the whole domain. On the contrary, if the block is very small, only a single observation is available, and an ill-known deterministic function is obtained.

Some authors claim the deterministic nature of the kriging problem should be acknowledged. Journel [33] explains how to retrieve all equations of kriging without resorting to the concept of a random function. This view is close to what Matheron calls the transitive model. The first step is to define experimental mean \hat{m} , standard deviation $\hat{\sigma}$ and variogram $\hat{\gamma}$ from the set of observation points $\{(x_i, z(x_i)), i = 1, \dots, n\}$ in a block \mathcal{A} . The two first quantities are supposed to be good enough approximations of the actual mean $m_{\mathcal{A}}$ and standard deviation $\sigma_{\mathcal{A}}$ of $z(x)$ in block \mathcal{A} , viewed as a random *variable* with sample space \mathcal{A} (and no more a fictitious random variable with an elusive sample space Ω). The sample variogram value $\hat{\gamma}(h)$ approximates the quantity :

² cited by J.-P. Chilès

$$\gamma_{\mathcal{A}}(h) = \frac{\int_{\mathcal{A}_h} (z(x+h) - z(x))^2 dx}{2|\mathcal{A}_h|},$$

taken over the set \mathcal{A}_h formed by intersecting \mathcal{A} and its translated by $-h$. In fact $\gamma_{\mathcal{A}}(h) = \gamma_{\mathcal{A}}(-h)$ and the variogram value $\gamma_{\mathcal{A}}(h)$ applies to the domain $\mathcal{A}_h \cup \mathcal{A}_{-h}$. For h small enough it is representative of \mathcal{A} itself.

Journal [33] shows that there exists a *stationary random function* $Z_{\mathcal{A}}(x)$ having such empirical characteristics: $m_{\mathcal{A}}$, $\sigma_{\mathcal{A}}$ and $\gamma_{\mathcal{A}}$.

Thus, if we define $z^*(x) = \sum_{i=1}^n \lambda_i(x) z(x_i)$, the estimation variance (under the unbiasedness condition), defined by $\mathbb{V}[Z_{\mathcal{A}}(x) - Z^*(x)] = \mathbb{E}[(Z_{\mathcal{A}}(x) - Z^*(x))^2]$, where $Z^*(x)$ is the “randomized” kriging estimate of $Z_{\mathcal{A}}(x)$ coincides with the spatial integral $\frac{\int_{\mathcal{A}} (z(x) - z^*(x))^2 dx}{|\mathcal{A}|}$. Hence, ordinary kriging is basically the process of minimizing a spatially averaged squared error over the domain \mathcal{A} on the basis of available observations.

The following assumption is made:

$$\mathcal{A} \simeq \{x : x \in \mathcal{A}, x + h_i \in \mathcal{A}, i = 1, \dots, n\},$$

where $h_i = x_0 - x_i$. It means that we restrict the kriging to the vicinity of the sample points x_i and that this estimation area is well within \mathcal{A} . It leads to retrieve the kriging equations.

The unbiasedness assumption of the stochastic kriging is replaced by requiring a zero average error over \mathcal{A} that no longer depends on x_0 :

$$e_{\mathcal{A}}(x_0) = \frac{\int_{\mathcal{A}} (z(x) - z^*(x)) dx}{|\mathcal{A}|} = 0.$$

Note that $\int_{\mathcal{A}} z^*(x) dx = \sum_{i=1}^n \lambda_i(x_0) \int_{\mathcal{A}} z(x + h_i) dx$, and that due to the above assumption, $\int_{\mathcal{A}} z(x + h_i) dx = \int_{\mathcal{A}} z(x) dx$. So, $\int_{\mathcal{A}} z(x) dx = \sum_{i=1}^n \lambda_i(x_0) \int_{\mathcal{A}} z(x) dx$ and therefore, $\sum_{i=1}^n \lambda_i(x_0) = 1$.

Then the squared error can be developed as

$$(z(x) - z^*(x))^2 = z(x)^2 - 2 \sum_{i=1}^n \lambda_i(x) z(x) z(x + h_i) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i(x) \lambda_j(x) z(x + h_i) z(x + h_j).$$

The spatially averaged squared error is obtained by integrating this expression over \mathcal{A} . If we introduce the counterpart of a covariance in the form

$$C(h) = \frac{\int_{\mathcal{A}} z(x) z(x+h) dx}{|\mathcal{A}|} - m_{\mathcal{A}}^2 = \sigma_{\mathcal{A}}^2 - \gamma_{\mathcal{A}}(h),$$

it can be shown that we recognize, in the above mean squared error, the expression (5) of the simple kriging variance based on stationary random functions. Of course the obtained linear system of equations is also the same and requires positive definiteness of the covariance matrix, hence the use of a proper variogram model fitted from the sample variogram. However under the purely deterministic spatial

approach, this positiveness condition appears as a property needed to properly solve the least square equations. It is no longer related to the covariance of a random function. Failure of this condition on the sample variogram may indicate an ill-conditioning of the measured data that precludes the possibility of a sensible least square interpolation.

In summary, the whole kriging method can be explained without requiring the postulate of a random function over \mathcal{D} , which may appear as an elusive artefact. There is no random function Z a realization of which is the phenomenon under study, but rather a random variable A on each block \mathcal{A} the sample space of which is the block itself, that we can bind to a stationary random function $Z_{\mathcal{A}}$ on the block. While this remark will not affect the kriging practice (since both the deterministic and the stochastic settings lead to the same equations in the end), it becomes important when epistemic uncertainty enters the picture, as it sounds more direct to introduce it in the concrete deterministic approach than in the abstract stochastic setting. It also suggests that teaching the kriging method may obviate the need for deep, but non-refutable, stochastic concepts like ergodicity and stationarity.

7.3 Towards integrating epistemic uncertainty in spatial interpolative prediction

The above considerations lead us to a difficult task if epistemic uncertainty is to be inserted into the kriging method. Generalizing the random function framework to fuzzy random functions, whose mathematical framework is now well developed, looks hopeless. Indeed it certainly would not help providing a tractable approach, since the simplest form of kriging already requires a serious computational effort.

Adding interval uncertainty to simple kriging would also be mathematically tricky. It has been shown above that the method proposed by Diamond is not quite cogent, as it handles intervals or fuzzy intervals as objective values to which a scalar distance can be applied. The approach of Bardossy looks more convincing, even if the use of interval arithmetic is questionable. Computing an interval-valued sample variogram via optimisation is a very difficult task. Indeed, the computation of an interval-valued sample variance is a NP-hard problem [24].

The extension of the least squares method to interval-valued functions, if done properly, is also a challenging task as it comes down to inverting a matrix having interval-valued coefficients. In this respect the fuzzy least squares approach of Diamond [15], based on a scalar distance between fuzzy intervals is also problematic. It is not clear to see what the result tells us about the uncertainty concerning all least squares estimates that can be found from choosing precise original data inside the input intervals.

Diamond's kriging approach produces a scalar variogram, hence scalar influence coefficients, to be derived, which does not sound natural, as one may on the contrary expect that the more uncertain the data, the more uncertain the ensuing variogram. On the other hand, extending the least square method to fuzzy data in a meaning-

ful way, that is by letting the imprecision of the variogram impact the influence coefficients looks computationally challenging. One may think of a method dual to the Diamond's approach, that would be based on precise data plus an imprecise variogram, thus leading to imprecise interpolation between precise data. Such an imprecise variogram would be seen as a family of theoretical variograms induced by the sample variogram. Even if we could compute fuzzy influence coefficients in an efficient way from such imprecise or fuzzy variograms, it is not correct to apply interval or fuzzy interval arithmetic to the linear combination of fuzzy data when the influence coefficients are fuzzy, even if their uncertainty were independent from the uncertainty pervading the data, due to the normalisation constraint [20]. But the epistemic uncertainty of influence coefficients partially depends on the quality of the data (especially if an automatic fitting procedure is used for choosing the variogram). So it is very difficult use data uncertainty in a non-redundant way in the resulting fuzzy kriging estimates.

As far as epistemic uncertainty is concerned, there is a paradox in kriging that is also present in interpolation techniques if considered as prediction tools: the kriging result is precise. However, intuitively, the farther is x_0 from the known points x_i , the less we know about $z(x_0)$. A cogent approach to estimating the loss of information when moving away from the known locations is needed. Of course, within the kriging approach, one can resort to using the kriging variance as an uncertainty indicator, but it is known not to depend on the data values $z(x_i)$, and again relies on assumptions on the underlying fictitious random function that is the theoretical underpinning of kriging. It is acknowledged [51] that kriging variance is not estimation variance but rather some index of data configuration. Thus, it seems obvious that techniques more advanced than the usual kriging variance are required for producing a useful estimation of the kriging error or imprecision.

So, a rigorous handling of epistemic uncertainty in kriging looks like a non-trivial task. Is it worth the effort? In fact, kriging is a global interpolation method that does not take into account local specificities of terrain since the variogram relies on averages of differences of measured values at pairs of points located at a given distance from each other. Indeed parameters of the variogram are estimated globally. This critique can be found repeatedly in the literature. This point emphasizes the need to use other kinds of possibly imprecise knowledge about the terrain than the measured points.

Overall, the handling of epistemic uncertainty in spatial prediction (independently of the problem of the local validity of the kriging estimates) could be carried out using one of the following methodologies

1. Replace the kriging approach by techniques that would be mathematically simpler, more local, and where the relationship between interpolation coefficients and local dependence information would be more direct. For instance we could consider interpolation techniques taking into account local gradient estimates from neighboring points (even interpolating between locally computed slopes). This would express a more explicit impact of epistemic uncertainty, present in the measured data and in the knowledge of local variations of the ill-known spatial function, on the interpolation expression, obviating the need for reconsidering a

more genuine fuzzy least square method from scratch. This move requires further investigations of the state of the art in the interpolation area so as to find a suitable spatial prediction technique.

2. Use probabilistic methods (such as Monte-Carlo or Gibbs sampling) to propagate uncertainty taking the form of epistemic possibility distributions (intervals or fuzzy intervals) on variogram parameters and/or observed data. Such an idea is at work for instance in the transformation method of Michael Hanss [28] for mechanical engineering computations under uncertainty modelled by fuzzy sets. The idea is to sample a probability distribution so as to explore the values of a complex function over an uncertainty domain. In such a method the probability distribution is just a tool for guiding the computation process. The set of obtained results (scenarii) should not be turned into a histogram but into a range of possible outputs. The use of fuzzy sets would come down to explore a family of nested confidence domains with various confidence values, thus yielding a fuzzy set of possible outputs (e.g. a kriged value). The merit of this approach, recently developed by the authors [38] is to encapsulate already existing kriging methods within a stochastic simulation scheme, the only difference with other similar stochastic methods being the non-probabilistic exploitation of the results.

8 Conclusion

The stochastic framework of geostatistics and the ensuing kriging methodology are criticized in the literature for three reasons:

- The purely mathematical nature of the random function setting and the attached assumptions of stationarity and ergodicity, that are acknowledged to be non-refutable;
- the questionable legitimacy, for local predictions, of a global index of spatial dependence such as the variogram, that averages out local trends; of course, the use of selected neighborhoods of measured values that change with each kriged location point can address this issue, albeit at the expense of a loss of continuity of the kriged surface.
- the computational burden of the kriging interpolation method and the poor interpretability of its influence coefficients.

On the first point, it seems that the choice of modeling a deterministic quantity by a random variable does not respect the principle of parsimony. If a deterministic model yields the same equations as the stochastic one, and moreover seems to coincide with our perception of the underlying phenomenon, the simpler model should be preferred (this is the case with simple kriging, as shown above). And the practical test according to the best prediction should be mitigated by the appraisal of the complexity of the used modeling framework.

On the second point, a variogram represents global information about a domain. Here, we do face a major difficulty common to all statistical approaches. Even if

the set of observations is large over the whole domain, local predictions will have a very poor validity if the number of observations in the vicinity of the predicted value location is too small. This conflict between the requested precision of predicted values and the necessity of large observation samples is pointed out by the advocates of kriging too.

The computational burden of kriging, even if not actually so high in the simpler versions, may pose a difficulty if epistemic uncertainty must be taken into account. As shown in section 4, available methods that try to introduce epistemic uncertainty into this technique seem to make it even more complex, and sometimes mathematically debatable, while by construction, they are supposed to provide imprecise outputs. Besides, it is not so easy to relate the form of the variogram and the expressions of the kriging coefficients, and to figure out how they affect the derivatives of the interpolated function, while one may have some prior information on such derivatives from geological knowledge of a prescribed terrain.

Devising a spatial prediction method that could be simple enough to remain tractable under epistemic uncertainty, and realistic enough to provide faithful information about a given terrain where some measurements are available remains a challenging task, and an open research problem. Three lines of research have been explored so far

- Treating fuzzy observations like complex crisp observations in a suitable metric space: this approach is not really treating epistemic uncertainty, as discussed in section 6.1.
- Applying fuzzy arithmetics. This is used also by Diamond when computing the interpolation step. However, it cannot be used throughout the whole kriging method, because there is no explicit expression of the influence weights in terms of the variogram parameters. And would there be one, replacing scalar arithmetic operations by fuzzy ones would lead to a considerable loss of precision.
- Using optimisation techniques as popular in the interval analysis area. This was suggested very early by Bardossy in the fuzzy case, Dubrule and Kostov in the interval case. But it looks already computationally intractable to study the sensitivity of the kriging estimates to variogram parameters lying in intervals via optimisation.

The most promising line of research is to adapt the stochastic simulation methods to the handling of fuzzy interval analysis [38]. Indeed, it would enable existing kriging methods and stochastic exploration techniques to be exploited as such. The only difference is that the input data would be specified as representing epistemic uncertainty by nested sets of confidence intervals, and that the results of the computation would not be interpreted as a probability distribution, but exploited levelwise to form the fuzzy kriged values.

Acknowledgements

This work is supported by the French Research National Agency (ANR) through the CO2 program (project CRISCO2 ANR-06-CO2-003). The issue of handling epistemic uncertainty in geostatistics was raised by Dominique Guyonnet. The authors also wish to thank Jean-Paul Chilès and Nicolas Desassis for their comments on a first draft of this paper and their support during the project.

References

1. Armstrong M (1990) Positive Definiteness is Not Enough. *Math. Geol.* 24:135–143.
2. Armstrong M, Jabin R (1981) Variogram Models Must Be Positive Definite. *Math. Geol.* 13:455–459.
3. Aumann RJ (1965) Integrals of set-valued functions. *J. Math. Anal. Appl.* 12:1–12.
4. Bardossy A, Bogardi I, Kelly WE (1988) Imprecise (fuzzy) information in geostatistics. *Math. Geol.* 20:287–311.
5. Bardossy A, Bogardi I, Kelly WE (1990) Kriging with imprecise (fuzzy) variograms. I: Theory. *Math. Geol.* 22:63–79.
6. Bardossy A, Bogardi I, Kelly WE (1990) Kriging with imprecise (fuzzy) variograms. II: Application. *Math. Geol.* 22:81–94.
7. Berger, JO (1980) *Statistical decision theory*. Springer-Verlag, Berlin.
8. Berger J (1994) An overview of robust Bayesian analysis [with Discussion]. *Test* 3:5–124.
9. Berger JO, de Oliveira V, Sanso B (2001) Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association* 96:1361–1374.
10. Chilès JP, Delfiner P (1999) *Geostatistics, Modeling Spatial Uncertainty*. Wiley, New York, N.Y.
11. Couso I, Dubois D (2009) On the variability of the concept of variance for fuzzy random variables, *IEEE Transactions on Fuzzy Systems*, IEEE, Vol. 17 N. 5, p. 1070-1080.
12. Cressie NAC (1993) *Statistics for Spatial Data*, Revised Edition. New York: John Wiley & Sons, New York, N.Y.
13. de Finetti B (1974) *Theory of probability: a critical introductory treatment*. John Wiley & Sons, New York, N.Y.
14. de Oliveira V, Kedem B, Short DA (1997) Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association* 92:1422–1433.
15. Diamond P (1988) Fuzzy least squares. *Information Sciences*, 46:141–157.
16. Diamond P (1988) Interval-valued random functions and the kriging of intervals. *Math. Geol.* 20:145–165.
17. Diamond P (1989) Fuzzy kriging. *Fuzzy Sets and Systems* 33:315–332.
18. Dubois D (2006) Possibility theory and statistical reasoning. *Computational Statistics & Data Analysis* 51:47–69.
19. Dubois D, Foulloy L, Mauris G, Prade H (2004) Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10:273–297.
20. Dubois D, Prade H (1981) Additions of interactive fuzzy numbers”, *IEEE Trans. on Automatic Control*, 26(4), 926–936.
21. Dubois D, Prade H (1988) *Possibility Theory*. Plenum Press, New York.
22. Dubrule O, Kostov C (1986) An interpolation method taking into account inequality constraints: I. Methodology. *Math. Geol.* 18:33–51.
23. Emery X (2003) Disjunctive kriging with hard and imprecise data. *Math. Geol.* 35:699–718.
24. Ferson S, Ginzburg L, Kreinovich V, Longpr L, Aviles M (2002) Computing variance for interval data is NP-hard. *SIGACT News* 33:108–118.

25. Gaudard M, Karson M, Linder E, Sinha D (1999) Bayesian spatial prediction. *Environmental and Ecological Statistics* 6:147–171.
26. Goovaerts P (1997) *Geostatistics for Natural Resources Evaluation*. Oxford Univ. Press, New-York.
27. Handcock MS, Stein ML (1993) A Bayesian analysis of kriging. *Technometrics* 35:403–410.
28. Hanss M (2002) The transformation method for the simulation and analysis of systems with uncertain parameters. *Fuzzy Sets and Systems* 130:277–289.
29. Hartigan JA (1969) Linear Bayesian methods. *J. Royal Stat. Soc. Ser. B* 31:446–454.
30. Helton JC, Oberkampf WL (2004) Alternative representations of epistemic uncertainty. *Reliability Engineering and System Safety* 85:1–10.
31. Hukuhara M (1967) Integration des applications mesurables dont la valeur est un compact convexe. *Funkcialaj Ekvacioj*, 10:205–223.
32. Journel AG, Huijbregts CJ (1978) *Mining Geostatistics*. New York: Academic Press.
33. Journel AG (1985) The deterministic side of geostatistics *Math. Geol.* 17:1–15.
34. Journel AG (1986) *Geostatistics: Models and Tools for the Earth Sciences*. *Math. Geol.* 18:119–140.
35. Journel AG (1986) Constrained interpolation and qualitative information - The soft kriging approach. *Math. Geol.* 18:269–286.
36. Kostov C, Dubrule O (1986) An interpolation method taking into account inequality constraints: II. Practical approach. *Math. Geol.* 18:53–73.
37. Krige DG (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa* 52:119–139.
38. Loquin K, Dubois D, A fuzzy interval analysis approach to kriging with ill-known variogram and data, in preparation.
39. Mallet JL (1980) Rgression sous contraintes linaires : application au codage des variables alatoires. *Revue de Statistiques appliques* 28:57–68.
40. Matheron G, Blondel F (1962) *Traité de géostatistique appliquée*. Editions Technip, Paris.
41. Matheron G (1969) *Le krigeage universel*. Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fasc. 1, École des Mines de Paris.
42. Matheron G (1975) *Random Sets and Integral Geometry*, John Wiley & Sons, New-York, N.Y.
43. Matheron G (1978) *Estimer et choisir: essai sur la pratique des probabilités*. École des Mines de Paris.
44. Matheron G (1987) Suffit-il pour une covariance d'être de type positif? *Études Géostatistiques V, Séminaire CFSG sur la Géostatistique*, Fontainebleau, Sciences de la Terre Informatiques.
45. Matheron G (1989) The Internal Consistency of Models in Geostatistics. In M. Armstrong (Ed.), *Geostatistics, Proceedings of the Third International Geostatistics*, Avignon, Kluwer Academic Publishers, Dordrecht, 21–38.
46. Omre H (1987) Bayesian Kriging - merging observations and qualified guesses in kriging. *Math. Geol.* 19:25–39.
47. Puri ML, Ralescu DA (1986) Fuzzy random variables. *J. Math. Anal. Appl.* 114:409–422.
48. Rios Insua D, Ruggieri F (2000) *Robust Bayesian Analysis*. Springer, Berlin.
49. Shafer G (1976) *A mathematical theory of evidence*. Princeton university press Princeton, N.J.
50. Shafer G, Vovk V (2001) *Probability and Finance: It's Only a Game!* New York: Wiley.
51. Srivastava R M (1986) Philip and Watson—Quo vadunt? *Math. Geol.* 18:141–146.
52. Taboada J, Rivas T, Saavedra A, Ordóñez C, Bastante F, Giraldez E (2008) Evaluation of the reserve of a granite deposit by fuzzy kriging. *Engineering Geol.* 99:23–30.
53. Walley P (1991) *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London.
54. Zadeh LA (1965) Fuzzy sets, *Inf. Control*, 8:338–353.
55. Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems*, 1:3–28.