Ordinary multigaussian kriging for mapping conditional probabilities of soil properties

Xavier Emery*

Department of Mining Engineering, University of Chile, Avenida Tupper 2069, Santiago, Chile

Abstract

This paper addresses the problem of assessing the risk of deficiency or excess of a soil property at unsampled locations, and more generally of estimating a function of such a property given the information monitored at sampled sites. It focuses on a particular model that has been widely used in geostatistical applications: the multigaussian model, for which the available data can be transformed into a set of Gaussian values compatible with a multivariate Gaussian distribution.

First, the conditional expectation estimator is reviewed and its main properties and limitations are pointed out; in particular, it relies on the mean value of the normal scores data since it uses a simple kriging of these data. Then we propose a generalization of this estimator, called "ordinary multigaussian kriging" and based on ordinary kriging instead of simple kriging. Such estimator is unbiased and robust to local variations of the mean value of the Gaussian field over the domain of interest. Unlike indicator and disjunctive kriging, it does not suffer from order-relation deviations and provides consistent estimations.

An application to soil data is presented, which consists of pH measurements on a set of 165 soil samples. First, a test is proposed to check the suitability of the multigaussian distribution to the available data, accounting for the fact that the mean value is considered unknown. Then four geostatistical methods (conditional expectation, ordinary multigaussian, disjunctive and indicator kriging) are used to estimate the risk that the pH at unsampled locations is less than a critical threshold and to delineate areas where liming is needed. The case study shows that ordinary multigaussian kriging is close to the ideal conditional expectation estimator when the neighboring information is abundant, and departs from it only in under-sampled areas. In contrast, even within the sampled area, disjunctive and indicator kriging substantially differ from the conditional expectation; the discrepancy is greater in the case of indicator kriging and can be explained by the loss of information due to the binary coding of the pH data.

Keywords: Geostatistics; Conditional expectation; Gaussian random fields; Posterior distributions; Indicator kriging; Disjunctive kriging

1. Introduction

Kriging techniques are currently used to estimate soil properties, such as electrical conductivity, pH,

* Tel.: +56 2 6784498; fax: +56 2 6723504.
E-mail address: xemery@cec.uchile.cl.

nutrient or contaminant concentrations, on the basis of a limited set of samples for which these properties have been monitored. However, because of its smoothing property, linear kriging is not suitable to applications that involve the conditional distributions of the unsampled values, for instance assessing the risk of exceeding a given threshold. This problem is critical for delineating contaminated areas where a remedial treatment is needed, for determining land suitability for a specific crop, or for planning an application of fertilizer.

Determining the risk of exceeding a threshold, or more generally estimating a function of a soil property, can be dealt with either stochastic simulations or nonlinear geostatistical methods like indicator kriging or disjunctive kriging, which have found wide acceptance in soil science (Webster and Oliver, 1989, 2001; Wood et al., 1990; Oliver et al., 1996; Van Meirvenne and Goovaerts, 2001; Lark and Ferguson, 2004). An alternative to indicator and disjunctive kriging is the conditional expectation estimator. However, in practice this estimator is hardly used, except in the scope of the multigaussian model (Goovaerts, 1997, p. 271; Chilès and Delfiner, 1999, p. 381). This paper focuses on this particular model and is organized as follows. In the next section, the conditional expectation estimator is reviewed and its main properties and practical limitations are highlighted, in particular concerning the use of a simple kriging of the Gaussian data. Then, ordinary kriging is substituted for simple kriging in order to obtain an unbiased estimator that is robust to local variations of the Gaussian data mean. The proposed approach is finally illustrated and discussed through a case study in soil science.

2. On the conditional expectation estimator

2.1. The multigaussian model

A Gaussian random field, or multigaussian random function, is characterized by the fact that any weighted average of its variables follows a Gaussian distribution. Its spatial distribution is entirely determined by its first- and second-order moments (mean and covariance function or variogram), which makes the statistical inference very simple under an assumption of stationarity. In general, the available data do not have a Gaussian histogram, so that a transformation is required to turn them into Gaussian values (*normal score transform*). The mean of the transformed data is usually set to zero and their variance to one, i.e. one works with standard Gaussian distributions (Rivoirard, 1994, p. 46; Goovaerts, 1997, p. 273).

A goal of this paper is to find an estimator that is robust to departures of the data from the ideal multigaussian model, in particular concerning local variations of the mean value over the domain of interest. Therefore, the assumption of zero mean is omitted and the following hypotheses are made:

1) the transformed data have a multivariate Gaussian distribution;
2) their mean is constant (at least at the scale of the neighborhood used for local estimations) and equal to m;
3) their variance is equal to one;
4) their correlogram is known; henceforth, the correlation between the values at locations $\mathbf{x}$ and $\mathbf{x}'$ is denoted by $\rho(\mathbf{x},\mathbf{x}')$. In the stationary case, this is a function of the separation vector $\mathbf{h}=\mathbf{x}-\mathbf{x}'$ only.

2.2. Local estimation with multigaussian kriging

Let $\{Y(\mathbf{x}),\ \mathbf{x}\in D\}$ be a Gaussian random field defined on a bounded domain D and known at a set of sampling locations $\{\mathbf{x}_\alpha,\ \alpha=1\ldots\ n\}$. It can be shown (Goovaerts, 1997, p. 272) that the *conditional* distribution of $Y(\mathbf{x})$ is Gaussian-shaped, with mean equal to its simple kriging $Y(\mathbf{x})^{SK}$ from the available data and variance equal to the simple kriging variance $\sigma^2_{SK}(\mathbf{x})$. Therefore, the posterior or conditional cumulative distribution function (in short, ccdf) at location $\mathbf{x}$ is

$$\forall y\in \mathrm{R}, F(\mathbf{x};y|\text{data}) = G\left(\frac{y-Y(\mathbf{x})^{SK}}{\sigma_{SK}(\mathbf{x})}\right) \tag{1}$$

where $G(.)$ is the standard Gaussian cdf. This ccdf can be used to estimate a function of the Gaussian field, say $\varphi[Y(\mathbf{x})]$ (this is often referred to as a "transfer function"). For instance, one can calculate the expected value of the conditional distribution of $\varphi[Y(\mathbf{x})]$, which defines the "conditional expectation" estimator (Rivoirard, 1994, p. 61):

$$\{\varphi[Y(\mathbf{x})]\}^{\mathrm{CE}} = \int \varphi(y)\mathrm{d}F(\mathbf{x};y|\mathrm{data})$$
$$= \int \varphi\Big[Y(\mathbf{x})^{\mathrm{SK}} + \sigma_{\mathrm{SK}}(\mathbf{x})t\Big]g(t)\mathrm{d}t \qquad (2)$$

where $g(.)$ stands for the standard Gaussian probability distribution function. The estimator in Eq. (2) is also called "multigaussian kriging" in the geostatistical literature (Verly, 1983; David, 1988, p. 150). Its implementation relies on an assumption of strict stationarity and knowledge of the prior mean m, in order to express $Y(\mathbf{x})^{\mathrm{SK}}$.

In practice, Eq. (2) can be calculated by numerical integration (Fig. 1), by drawing a large set of values

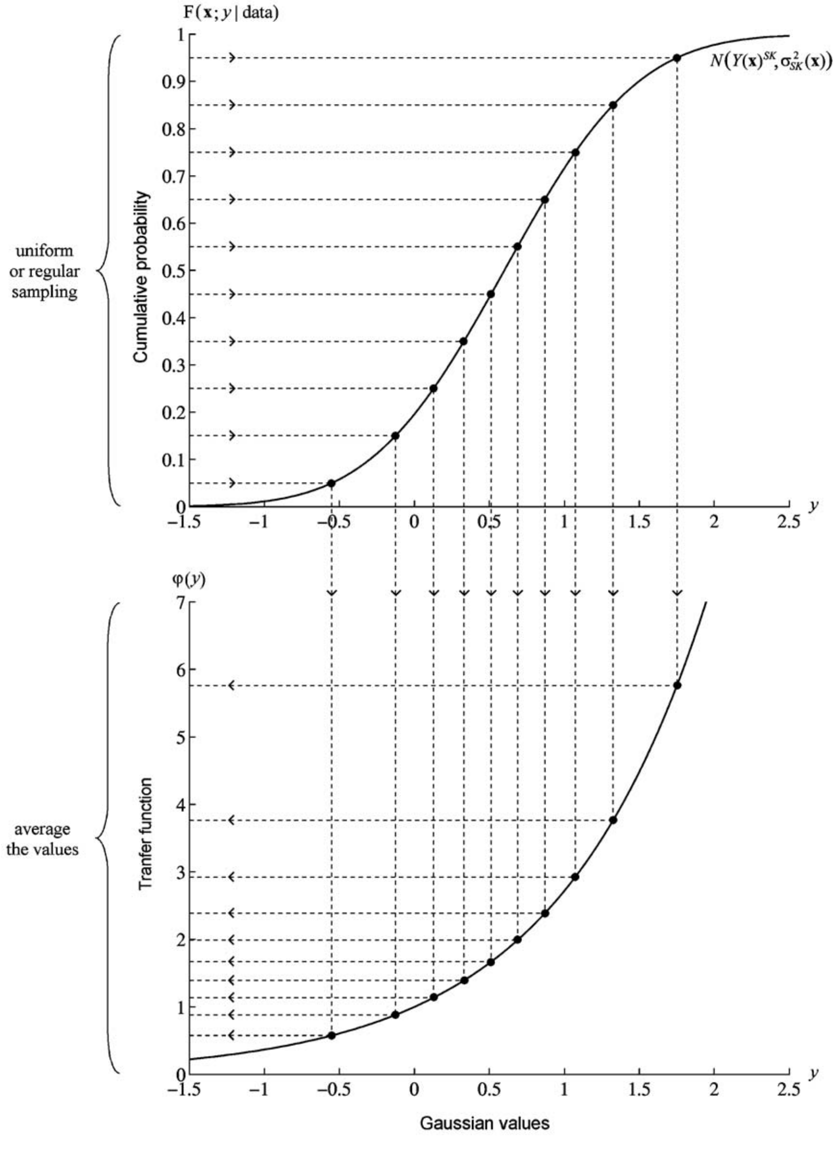

Fig. 1. Numerical integration for calculating the conditional expectation of a transfer function.

$\{u_1, \ldots u_M\}$ that sample the [0,1] interval either uniformly or regularly, and putting

$$\{\varphi[Y(\mathbf{x})]\}^{\text{CE}} \approx \frac{1}{M} \sum_{i=1}^{M} \varphi(y_i) \tag{3}$$

with $\forall i \in \{1, \ldots M\}, F(\mathbf{x}; y_i | \text{data}) = u_i$. An alternative approach is to replace the integral in Eq. (2) by a polynomial expansion, which also helps to express the estimation variance (Emery, 2005b).

Note: throughout the paper, the kriging estimators are regarded as random variables. This allows one to express their expected value and establish their unbiasedness.

2.3. Properties of the estimator

Among all the measurable functions of the data, the conditional expectation minimizes the estimation variance (variance of the error). It is unbiased and, even more, conditionally unbiased. This property is of great importance in resource assessment problems (Journel and Huijbregts, 1978, p. 458). Furthermore, since the ccdf [Eq. (1)] is an increasing function, the conditional expectation honors order relations and provides mathematically consistent results: for instance, the estimate of a positive function is always positive (Rivoirard, 1994, p. 62).

Despite these nice properties, the estimator is not often used in practice. One of the reasons is the attraction to the mean that produces the use of a simple kriging of the Gaussian data. Indeed, at locations distant from the data relatively to the range of the correlogram model, the simple kriging of $Y(\mathbf{x})$ tends to the prior mean m and the kriging variance to the prior unit variance, hence the conditional expectation in Eq. (2) tends to the prior expectation of $\varphi[Y(\mathbf{x})]$.

To avoid this effect when estimating a spatial attribute, practitioners often prefer ordinary kriging to simple kriging. This way, the mean value is considered unknown and is estimated from the data located in the kriging neighborhood. Several authors suggested a generalization of the conditional expectation estimator, by substituting in Eq. (2) an ordinary kriging for the simple kriging and leaving unchanged the simple kriging variance (Journel, 1980; Goovaerts, 1997, p. 282). However, this approach is generally ill advised, since it produces a bias when estimating a nonlinear function of the Gaussian field (Rivoirard, 1994, p. 71; Chilès and Delfiner, 1999, p. 382).

For this reason, although less precise, alternative methods are sometimes used instead of the conditional expectation, e.g. indicator or disjunctive kriging with unbiasedness constraints (Rivoirard, 1994, p. 69; Goovaerts, 1997, p. 301; Chilès and Delfiner, 1999, p. 417). The following section aims at generalizing the conditional expectation estimator, by trading simple kriging for ordinary kriging in Eq. (2) and correcting the expression of the estimator to avoid bias.

3. Ordinary multigaussian kriging

3.1. Construction of an unbiased estimator

In Eq. (2), the simple kriging estimator $Y(\mathbf{x})^{\text{SK}}$ is a Gaussian random variable (weighted average of multivariate Gaussian data) with mean m and variance

$$\text{var}\left[Y(\mathbf{x})^{\text{SK}}\right] = 1 - \sigma^2_{\text{SK}}(\mathbf{x}). \tag{4}$$

Therefore, the conditional expectation of $\varphi[Y(\mathbf{x})]$ belongs to the class of estimators that can be written in the following form:

$$\begin{aligned}\varphi[Y(\mathbf{x})]^* &= \int \varphi\left[Y^* + \sqrt{1 - \text{var}(Y^*)}t\right] g(t)\mathrm{d}t \\ &= E\left\{\varphi\left[Y^* + \sqrt{1 - \text{var}(Y^*)}T\right] | Y^*\right\}\end{aligned} \tag{5}$$

where Y^* is a Gaussian random variable with mean m and T is a standard Gaussian variable independent of Y^*.

The value of T does not need to be known, since the estimator $\varphi[Y(\mathbf{x})]^*$ is defined by an expected value with respect to T. In the case of the conditional expectation, Y^* is the simple kriging of $Y(\mathbf{x})$ and T can be interpreted as the standardized simple kriging error. In the multigaussian model, this error is independent of the kriging estimator (Chilès and Delfiner, 1999, p. 164 and 381).

A key result is that *any* estimator given by Eq. (5) constitutes an unbiased estimator of $\varphi[Y(\mathbf{x})]$, provided only that Y^* and T are independent normal variables, the first one with mean m, the second one with mean zero and unit variance. Indeed, under these conditions, $Y^* + \sqrt{1-\mathrm{var}(Y^*)}T$ and $Y(\mathbf{x})$ have the same univariate distribution (a normal distribution with mean m and unit variance), hence:

$$E\left\{\varphi[Y(\mathbf{x})]^*\right\} = E\left\{\varphi\left[Y^* + \sqrt{1-\mathrm{var}(Y^*)}T\right]\right\} = E\{\varphi[Y(\mathbf{x})]\}. \tag{6}$$

In particular, Y^* can be the ordinary kriging of $Y(\mathbf{x})$, defined by the following weighting of the Gaussian data $\{Y(\mathbf{x}_\alpha), \alpha=1\ldots n\}$:

$$Y(\mathbf{x})^{\mathrm{OK}} = \sum_{\alpha=1}^{n} \lambda_\alpha^{\mathrm{OK}} Y(\mathbf{x}_\alpha). \tag{7}$$

The ordinary kriging weights and error variance are determined by the following system, in which $\mu(\mathbf{x})$ is a Lagrange multiplier:

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_\beta^{\mathrm{OK}} \rho(\mathbf{x}_\alpha, \mathbf{x}_\beta) + \mu(\mathbf{x}) = \rho(\mathbf{x}_\alpha, \mathbf{x}) \quad \forall \alpha = 1 \ldots n \\ \sum_{\alpha=1}^{n} \lambda_\alpha^{\mathrm{OK}} = 1 \\ \sigma_{\mathrm{OK}}^2(\mathbf{x}) = 1 - \sum_{\alpha=1}^{n} \lambda_\alpha^{\mathrm{OK}} \rho(\mathbf{x}_\alpha, \mathbf{x}) - \mu(\mathbf{x}). \end{cases} \tag{8}$$

The variance of the estimator is related to the error variance and Lagrange multiplier introduced in Eq. (8):

$$\begin{aligned} \mathrm{var}\left[Y(\mathbf{x})^{\mathrm{OK}}\right] &= \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha^{\mathrm{OK}} \lambda_\beta^{\mathrm{OK}} \rho(\mathbf{x}_\alpha, \mathbf{x}_\beta) \\ &= \sum_{\alpha=1}^{n} \lambda_\alpha^{\mathrm{OK}} \rho(\mathbf{x}_\alpha, \mathbf{x}) - \mu(\mathbf{x}) \\ &= 1 - \sigma_{\mathrm{OK}}^2(\mathbf{x}) - 2\mu(\mathbf{x}). \end{aligned} \tag{9}$$

The use of an ordinary kriging in Eq. (5) defines the following estimator, named "ordinary multigaussian kriging" and associated with the superscript oMK:

$$\{\varphi[Y(\mathbf{x})]\}^{\mathrm{oMK}} = \int \varphi\left[Y(\mathbf{x})^{\mathrm{OK}} + \sqrt{\sigma_{\mathrm{OK}}^2(\mathbf{x}) + 2\mu(\mathbf{x})}\,t\right] g(t)\mathrm{d}t. \tag{10}$$

This expression does not depend on the value of the Gaussian data mean, so the latter can be considered unknown. With respect to the classical conditional expectation, the estimation of a function of $Y(\mathbf{x})$ is obtained by replacing the simple kriging by an ordinary kriging and the simple kriging variance by the ordinary kriging variance plus twice the Lagrange multiplier introduced in the kriging system. A particular case of Eq. (10) is the so-called "ordinary lognormal kriging" (Journel, 1980, p. 295; Rivoirard, 1990, p. 217), for which φ is an exponential function.

3.2. Pseudo-conditional distribution for assessing the risk of deficiency or excess of a soil property

Let us introduce the indicator function associated with threshold y:

$$\forall y \in \mathbb{R}, I_Y(\mathbf{x}; y) = \begin{vmatrix} 1 \text{ if } Y(\mathbf{x}) \le y \\ 0 \text{ otherwise.} \end{vmatrix} \tag{11}$$

The ordinary multigaussian kriging of such indicator provides a pseudo ccdf ("pseudo" because it differs from the true ccdf given in Eq. (1)), which estimates the risk that the value at location $\mathbf{x}$ is less than or equal to y conditionally to the available data at locations $\{\mathbf{x}_\alpha, \alpha=1\ldots n\}$:

$$\forall y \in \mathbb{R}, \hat{F}(\mathbf{x}; y|\mathrm{data}) = [I_Y(\mathbf{x}; y)]^{\mathrm{oMK}} = G\left(\frac{y - Y(\mathbf{x})^{\mathrm{OK}}}{\sqrt{\sigma_{\mathrm{OK}}^2(\mathbf{x}) + 2\mu(\mathbf{x})}}\right). \tag{12}$$

This corresponds to a Gaussian distribution with mean the ordinary kriging of $Y(\mathbf{x})$ and variance the ordinary kriging variance plus twice the Lagrange multiplier. As in Eq. (2), the estimator of a transfer function [Eq. (10)] can be expressed in terms of this pseudo ccdf:

$$\{\varphi[Y(\mathbf{x})]\}^{\mathrm{oMK}} = \int \varphi(y)\mathrm{d}\hat{F}(\mathbf{x}; y|\mathrm{data}) \tag{13}$$

Eq. (13) allows one to calculate the ordinary multigaussian kriging estimator by numerical integration [Eq. (3)]. The main properties of this estimator are discussed in the following subsections.

3.3. Robustness to local variations of the mean value

Contrary to the conditional expectation, ordinary multigaussian kriging [Eq. (10)] does not rely on the mean value of the Gaussian data. This mean can even vary locally in space, provided that it remains constant at the scale of the kriging neighborhood.

Ordinary multigaussian kriging is therefore suitable to a locally stationary framework and provides estimates that are robust to variations of the mean value in space. An alternative would be to use a simple kriging with a locally varying mean (in practice, estimated from the data located in the kriging neighborhood) in the conditional expectation estimator [Eq. (2)]. However, such approach does not account for any uncertainty on the mean value, which is considered known although it may be poorly estimated, and provides biased results as soon as the estimated mean departs from the true mean. Instead, ordinary multigaussian kriging appears to be simpler (there is no need to estimate the local mean) and is unbiased even if the true mean value differs from the global or local mean estimated from the data.

In practice, the empirical data never conform to the ideal model (stationary random field with known mean), so that a less efficient but more robust estimator is often preferred to the "optimal" estimator (Matheron, 1989b, p. 94; Rivoirard, 1994, p. 71; Chilès and Delfiner, 1999, p. 39). This concern for robustness is the reason why ordinary kriging is generally used instead of simple kriging when evaluating a regionalized variable in linear geostatistics. With the multigaussian approach, a nonlinear function of this variable can be estimated, which is a substantial improvement with respect to linear kriging.

3.4. Measures of local uncertainty

A limitation of ordinary multigaussian kriging concerns the calculation of local uncertainty measures, such as confidence intervals, local variance or interquartile range. Indeed, formulae (12) and (13) only give an unbiased estimator of the posterior distribution at location $\mathbf{x}$ and of any quantity that can be expressed as a linear function of the ccdf (e.g. expected value of $Y(\mathbf{x})$ or the probability that it exceeds a given threshold). But the estimated distribution [Eq. (12)] does not match the true one [Eq. (1)], in particular both distributions do not have the same variance.

In general, the pseudo ccdf understates the dispersion of the true ccdf. For instance, assume that only one datum is available in the kriging neighborhood (under-sampled area): the pseudo ccdf [Eq. (12)] is a step function (zero variance) and obviously underestimates the uncertainty prevailing at the unsampled location. This statement is true with other methods that assess the posterior distributions using ordinary kriging, such as indicator and disjunctive kriging with unbiasedness constraints: one obtains a step function if a single datum is found in the kriging neighborhood. It proves that the posterior distributions estimated with ordinary kriging are suitable for estimation purposes only, not for assessing local uncertainty or simulating the unknown values, e.g. via a sequential algorithm (Emery, 2004, p. 410).

3.5. Further properties

Ordinary multigaussian kriging [Eq. (10)] honors the value of $\varphi[Y(\mathbf{x})]$ at each data location. This exactitude property holds as soon as one uses in Eq. (5) an estimator Y^* that honors the Gaussian data, not only an ordinary kriging.

Eq. (10) does not make sense if $\sigma^2_{\mathrm{OK}}(\mathbf{x})+2\mu(\mathbf{x})<0$, or equivalently var$[Y(\mathbf{x})^{\mathrm{OK}}]>1$, because of the presence of a negative term under the square root. This situation is unlikely to occur, since the variance of the ordinary kriging estimator (weighted average of the data) is usually less than the prior unit variance. Otherwise, ordinary kriging should be traded for another weighted average of the Gaussian data with a variance less than or equal to one. A sufficient condition is that the weights are nonnegative and add to one to ensure unbiasedness; a solution to this problem has been proposed by Barnes and Johnson (1984).

Concerning the precision of ordinary multigaussian kriging, the estimation variance of $\varphi[Y(\mathbf{x})]$ is greater than the one of the conditional expectation, which is minimal among all the possible estimators of $\varphi[Y(\mathbf{x})]$. However, if the data are abundant, ordinary kriging is close to simple kriging (Goovaerts, 1997, p. 137), hence ordinary multigaussian kriging is close to the optimal conditional expectation. It therefore constitutes a worthy alternative to other nonlinear estimation methods, such as indicator or disjunctive kriging.

Table 1
Advantages and drawbacks of ordinary multigaussian kriging

Pros	Cons
Suitable for assessing conditional probabilities and transfer functions	Not suitable for assessing local uncertainty (e.g., via confidence intervals or local variance)
Unbiased estimator	
Robust to local variations of the mean value (local stationarity)	Undefined if the variance of the ordinary kriging estimator is greater than the prior unit variance
Accounts for uncertainty on the mean value	
When the data are scarce, does not yield the prior expectation	Theoretically less precise than the conditional expectation
Close to the conditional expectation when the data are abundant	Restricted to the multigaussian model
Honors the data	
The estimates do not need order-relation corrections	

Another advantage of ordinary multigaussian kriging over these two techniques is that there is no need for order-relation corrections, since the estimated ccdf [Eq. (12)] is an increasing function of the threshold.

The advantages and drawbacks of the ordinary multigaussian kriging approach are summarized in Table 1. The main restriction of this approach is certainly the multigaussian assumption, which must be suited to the available data to ensure the quality of the estimator. To validate this assumption, a test is proposed in the next section together with a case study. The other limitations indicated in Table 1 are quite mild and should not cause difficulty in practical applications, as it will be confirmed in the case study.

4. Application to soil data

4.1. Presentation of the dataset

In the following, the previous concepts and methods are applied to a dataset which consists of 165 samples quasi-regularly spaced over a 335×335 m field (Fig. 2a) located in Reunion Island (Indian Ocean) and planted to sugar cane. For each sample, several variables have been determined on three levels between 0 and 1 m in depth: available water capacity, pH, exchangeable aluminum, total acidity and cation exchange capacity. In the following, we are interested in the pH of the topsoil (0–200 mm), which has been measured in a solution of potassium chloride with a ratio soil/solution equal to 1/2.5. The pH is a critical variable for determining soil fertility, hence for predicting yield variability within the field and defining the amount of fertilizer that should be applied.

Fig. 2b and c display the histogram of the pH data and the normal scores variogram. The latter is not isotropic and has a greater range along the direction N10° W. The directional variograms have been computed for distances multiple of 25 m, with a tolerance of 20° on the azimuth and 12.5 m on the lag distance, so that each point is calculated after at least one hundred data pairs. The model is the sum of a nugget effect with sill 0.5 and a spherical structure with sill 0.5 and ranges 155 (N10° W) and 55 m (N80° E). Concerning the histogram, it has been declustered with the cell method (Goovaerts, 1997, p. 83), using a cell size of 25×25 m that roughly corresponds to the sampling mesh. It shows a bimodality, which may be symptomatic of a mixture of two populations. Actually, a look at the location map (Fig. 2a) proves that the high pH values are disseminated over the entire field of interest, hence one cannot define subdomains to study the two populations separately; this observation also explains the high relative nugget effect in the normal scores variogram (50% of the total sill).

Before performing local estimations with multigaussian kriging, it is advisable to check the compatibility of the multigaussian model with the transformed data.

4.2. Checking the multigaussian assumption

In practice, the multigaussian hypothesis cannot be fully validated because, in general, the inference of multiple-point statistics is beyond reach. Usually, only the univariate and bivariate distributions are examined (Goovaerts, 1997, p. 280). By construction of the normal scores transform, the former is Gaussian shaped and therefore consistent with the multigaussian model. Regarding the latter, there exist several ways of checking the two-point normality. One of them is the structural analysis of the indicator function

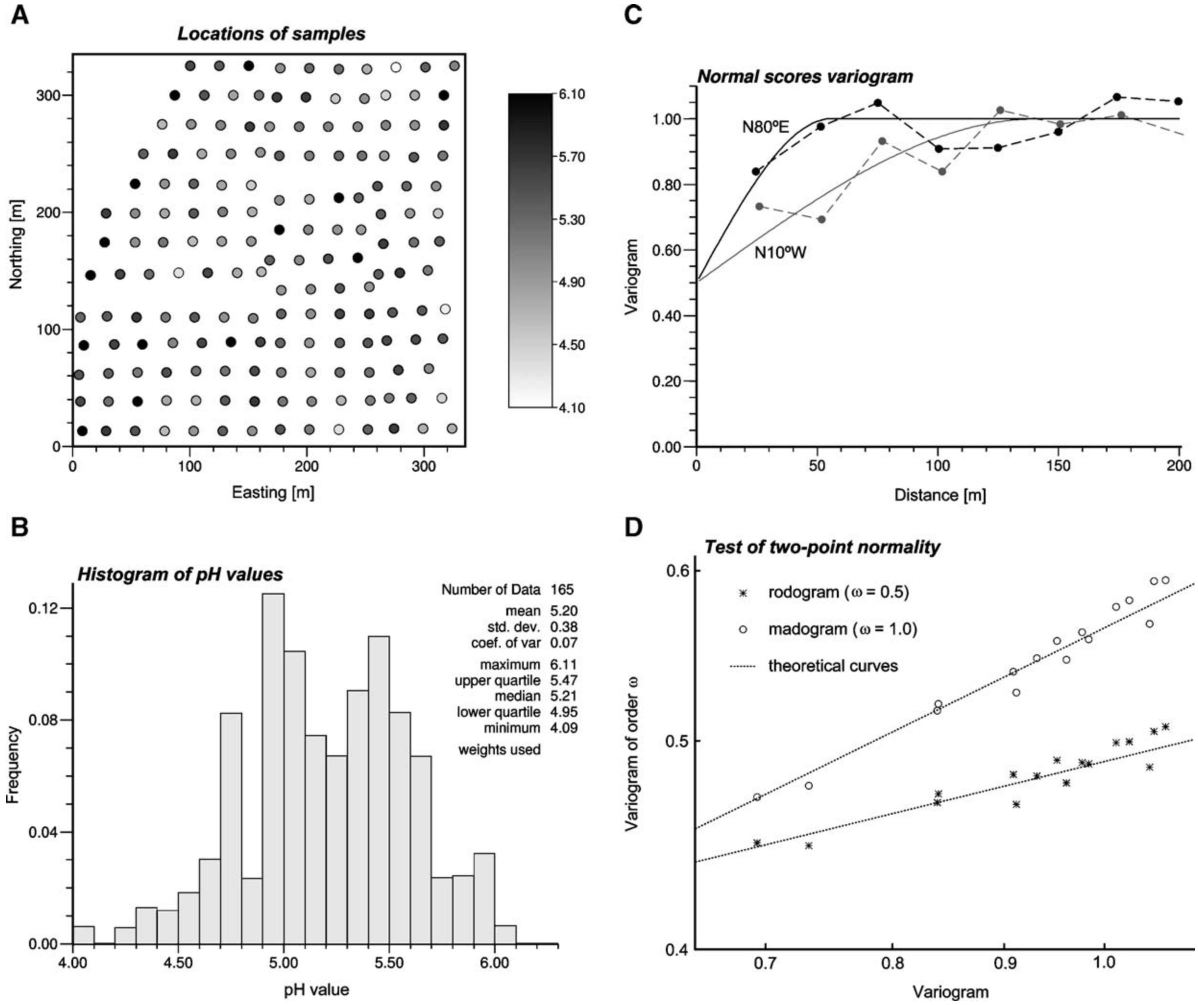

Fig. 2. Exploratory and variogram analyses of the pH data: A) location map of the samples; B) declustered pH histogram; C) normal scores variogram; D) test of the bivariate Gaussian assumption.

$I_Y(\mathbf{x},y)$ [Eq. (11)]: for several thresholds $\{y_1,\ldots y_k\}$, the empirical indicator variogram is compared to the theoretical variogram, which can be expressed as a function of the correlogram $\rho(\mathbf{h})$ of the Gaussian field (Chilès and Delfiner, 1999, p. 101) (here, local stationarity is assumed, so that the correlogram only depends on the separation vector between samples):

$$\gamma_{I,y}(\mathbf{h}) = G(y-m)[1-G(y-m)] - \frac{1}{2\pi}\int_0^{\rho(\mathbf{h})} \exp\left[-\frac{(y-m)^2}{1+u}\right]\frac{\mathrm{d}u}{\sqrt{1-u^2}} \quad (14)$$

However, one difficulty of the model is that the mean value m is considered unknown and may vary locally, so it is preferable to find a test that does not depend on this mean value. In this respect, a convenient way to check the bigaussian assumption is to study the variograms of order less than two, since these tools are defined after the *increments* of the random field. For $\omega \in]0,2]$, the variogram of order ω is defined as (Matheron, 1989a, p. 30):

$$\gamma_\omega(\mathbf{h}) = \frac{1}{2}E\{|Y(\mathbf{x}+\mathbf{h}) - Y(\mathbf{x})|^\omega\} \quad (15)$$

For $\omega=2$, one finds the classical variogram $\gamma(\mathbf{h})=1-\rho(\mathbf{h})$. Two other cases are of interest:

$\omega=1.0$ and $\omega=0.5$, which correspond to the madogram and rodogram respectively. Under the assumption that the distribution of $\{Y(\mathbf{x}+\mathbf{h}), Y(\mathbf{x})\}$ is bigaussian, one has (Emery, 2005a, p. 168):

$$\forall\omega\in]0,2], \gamma_\omega(\mathbf{h}) = \frac{2^{\omega-1}}{\sqrt{\pi}}\Gamma\left(\frac{\omega+1}{2}\right)[\gamma(\mathbf{h})]^{\omega/2} \quad (16)$$

where $\Gamma(.)$ is the gamma function.

Consequently, in log–log coordinates, the points that plot the variogram of order ω as a function of the classical variogram should be aligned with slope $\omega/2$. Once applied to the transformed data, this test is quite satisfactory (Fig. 2d), hence the multigaussian assumption is deemed acceptable for the data under study.

4.3. Mapping the risk that the pH is less than a given threshold

The pH of the soil has an effect on its fertility. When it decreases, H^+ ions become sufficiently

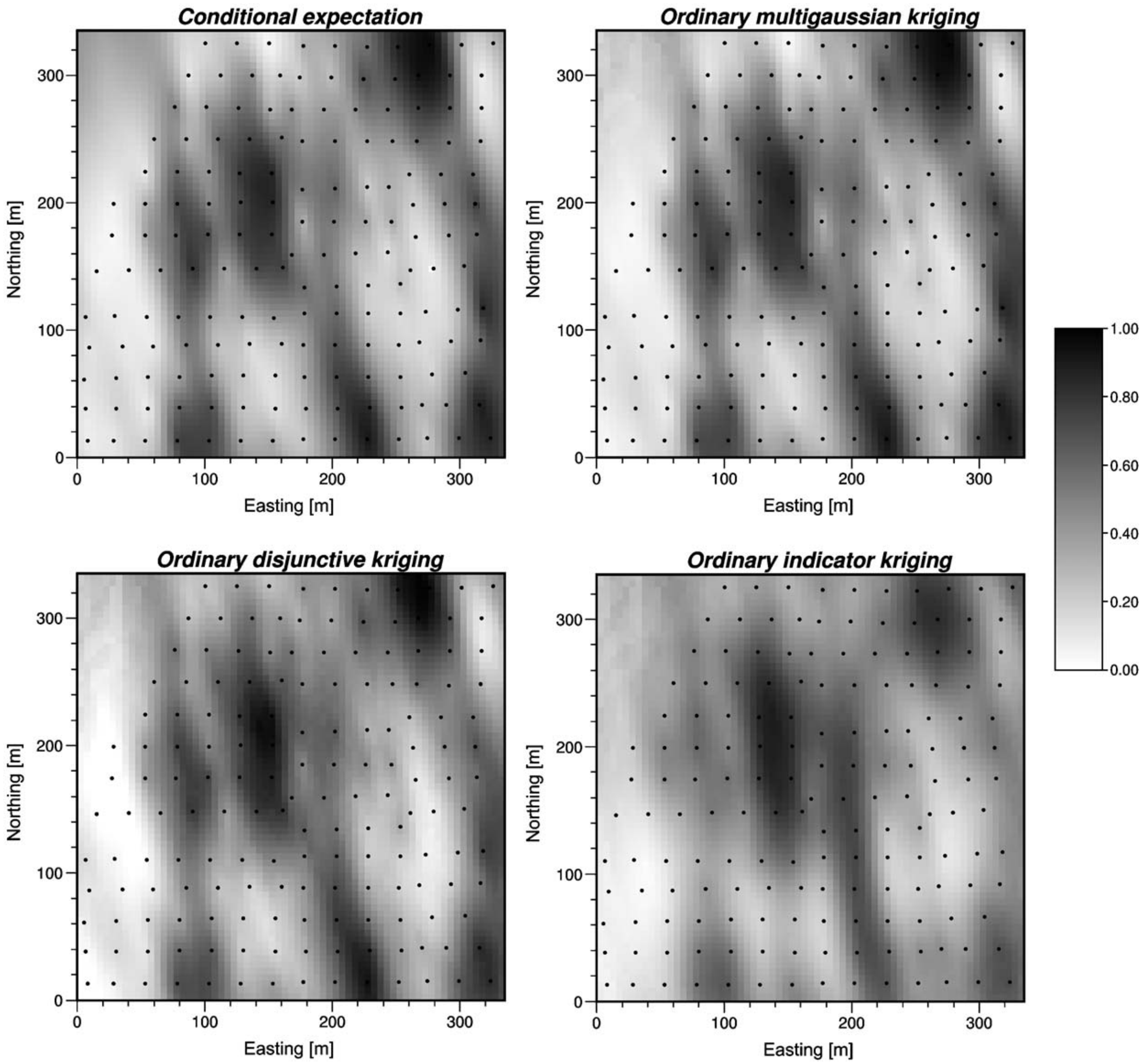

Fig. 3. Probability maps that the pH is less than 5.1. The sample locations are superimposed.

concentrated to attack the clay crystals, releasing Al^{3+} ions. At pH equal to approximately 5.1, the release becomes pronounced and is detrimental to the crops. Hence it is important to correctly assess the probability that the pH is less than a threshold of 5.1. In the following, four local estimation methods are compared:

- ***conditional expectation*** [Eq. (1)]
- ***ordinary multigaussian kriging*** [Eq. (12)]
- ***ordinary indicator kriging*** (Journel, 1983; Goovaerts, 1997, p. 301). This method relies on a coding of the pH data into indicator values associated with the target threshold (5.1) [Eq. (11)], followed by an ordinary kriging of the 0–1 values. The result is an unbiased estimation of the probability to be below the threshold;
- bigaussian disjunctive kriging with an unbiasedness constraint (Rivoirard, 1994, p. 70; Chilès and Delfiner, 1999, p. 417). For want of a better

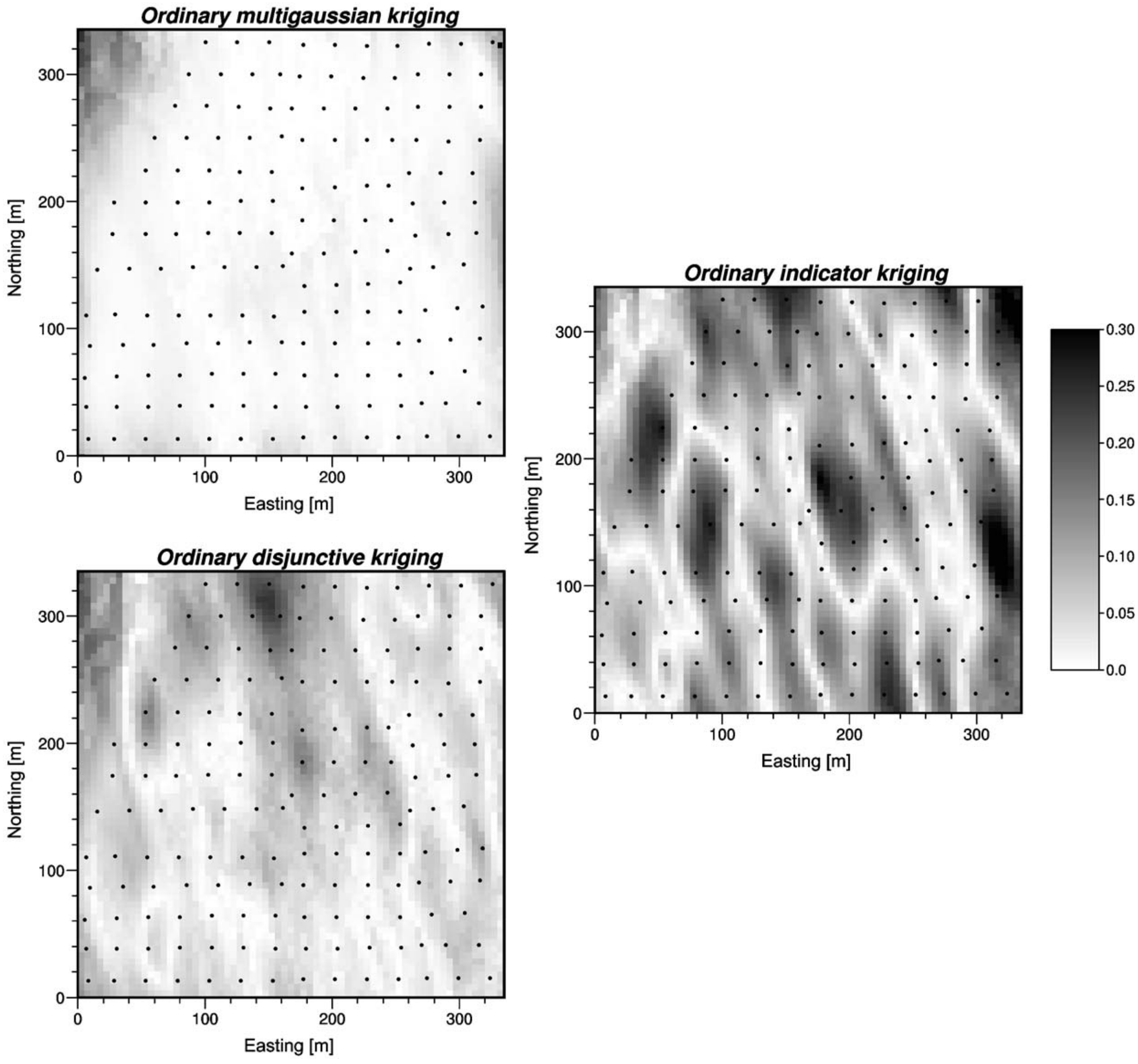

Fig. 4. Error maps plotting the absolute differences between each estimator and the conditional expectation.

name, this method will be called ***ordinary disjunctive kriging*** hereafter. It is based on an expansion of the indicator function [Eq. (11)] into Hermite polynomials and an ordinary kriging of each polynomial from its values at the data locations. The spatial covariance of the polynomial of degree p is equal to the one of the Gaussian data raised to power p. Unlike indicator kriging, ordinary disjunctive kriging avoids the loss of information due to the indicator coding of the original data (Lark and Ferguson, 2004, p. 42). It can be shown that it amounts to a full ordinary indicator cokriging, i.e. an ordinary cokriging of the indicators associated with all the possible thresholds (Liao, 1991).

In each case, kriging is performed in a moving neighborhood with a radius of 250 m along the direction of greater continuity (N10° W) and 90 m along the orthogonal direction (N80° E), looking for the 48 nearest samples. This number is deemed sufficient for estimating the local means when resorting to ordinary kriging. The results are displayed in Fig. 3. One notices that, in the northwest corner, the conditional expectation converges to the prior probability that the pH is less than 5.1 (namely, 41%), while the other three methods provide lower estimates (in general, below 30%). The reason is that the data located in the north-

Table 2
Coordinates and pH values on the first row of samples

Sample number	Easting [m]	Northing [m]	pH value
1	100.0	325.0	5.54
2	125.0	325.0	5.40
3	150.0	325.0	6.11
4	177.0	323.0	4.98
5	202.0	323.0	5.25
6	227.0	322.0	5.22
7	251.0	322.0	4.71
8	276.0	324.0	4.09
9	301.0	324.0	5.36
10	326.0	325.0	5.07

west part of the field have a higher pH than the global mean (around 5.5, whereas the average of the entire field is 5.2). In contrast, in the sampled area, little difference is observed between conditional expectation and ordinary multigaussian kriging. These two methods account for the multivariate Gaussian distribution of the data and provide consistent conditional probabilities (between 0 and 1), whereas disjunctive and indicator kriging require correcting the order-relation violations and significantly depart from conditional expectation, even inside the sampled area (Fig. 4).

The discrepancy between indicator and multigaussian kriging can be better understood on a simple example. Fig. 5 plots the estimated probabilities

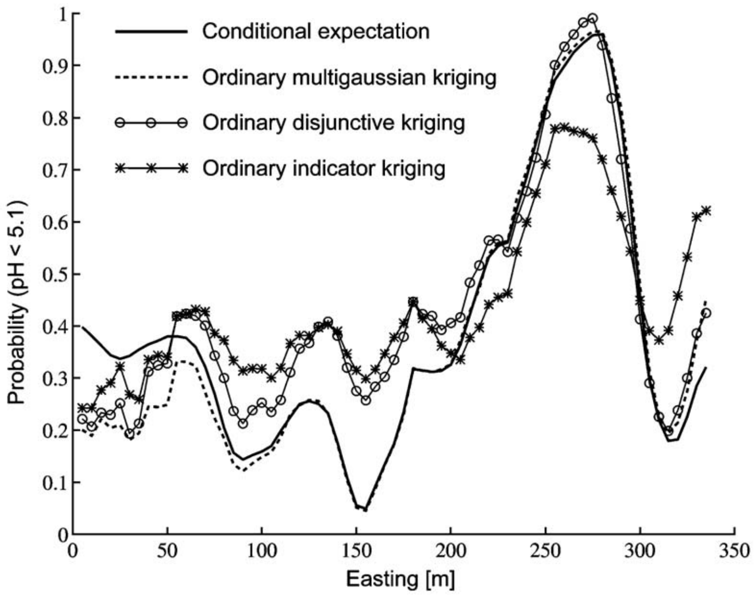

Fig. 5. Estimated probabilities along the east–west transect with north coordinate 322.5 m.

along a transect that approximately corresponds to the first row of data (north coordinate equal to 322.5 m, see Fig. 2a). The pH values of these data are presented in Table 2. As previously mentioned, in the western part of the transect, no data are available and the conditional expectation yields a higher probability (0.41) than the three other methods. Now, the discrepancy between conditional expectation and indicator kriging can be explained by the loss of information produced by the data coding into a binary variable: a datum is coded as 1 if its pH is less than 5.1, independently of whether the value is close to the threshold or not. In the example under study, the eighth datum of the transect (pH equal to 4.09) and the last datum (pH equal to 5.07) play the same role in the indicator kriging formalism, as both of them are coded as "1", which is misleading. On the contrary, the other three methods yield a higher risk that the pH is less than 5.1 in the vicinity of the eighth datum, and a lower risk in the vicinity of the last datum, which is very close to the target threshold.

4.4. Use of the probability maps for soil management

The previous methods quantify the risk that soil acidity is detrimental for crops and are useful for soil

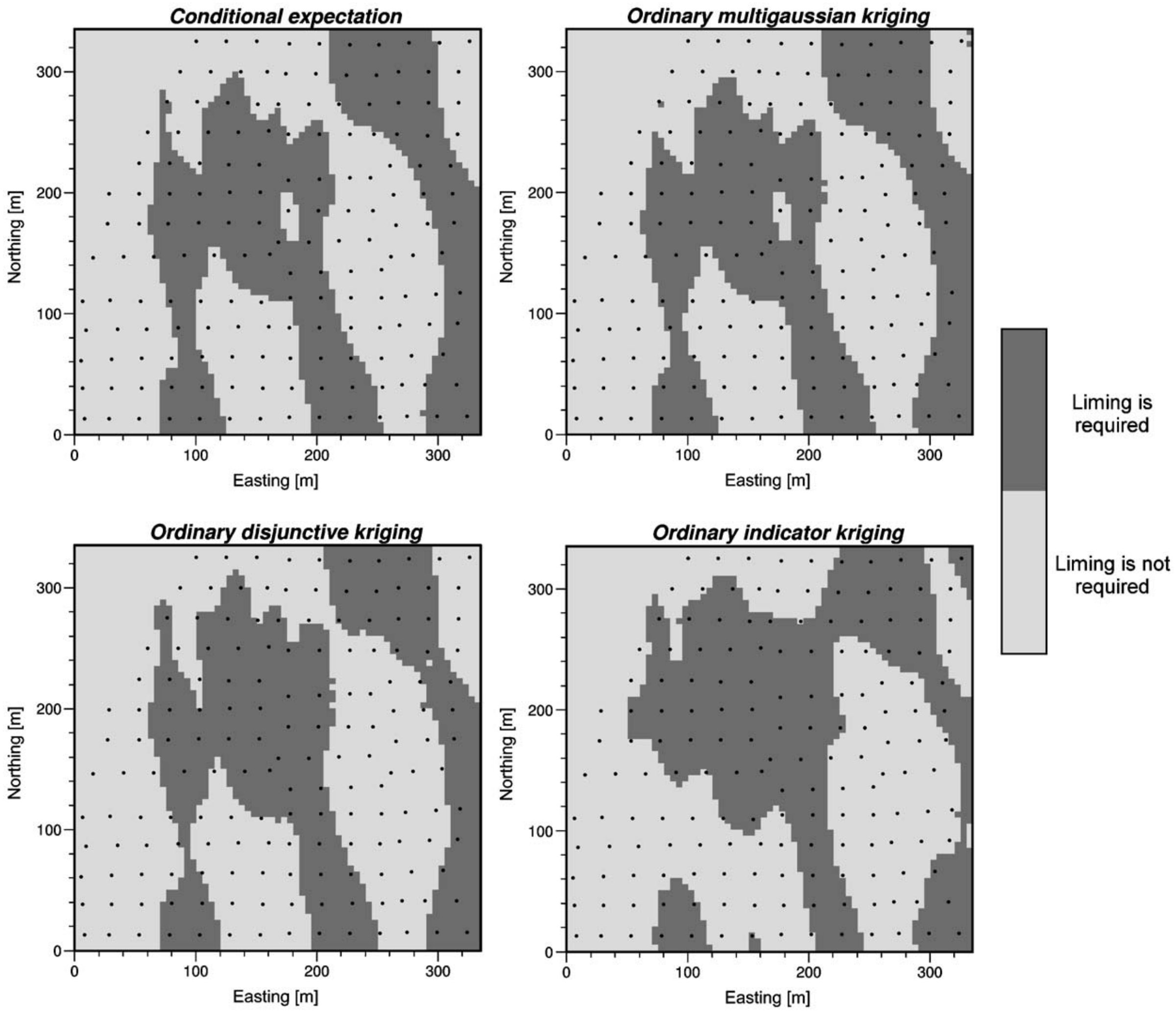

Fig. 6. Delineation of the areas where liming is needed.

management and decision-making, such as delineating the areas where a treatment is needed.

In this respect, one can calculate the proportion of the entire field whose pH is less than 5.1, either from the data histogram (Fig. 2b) or by averaging the local estimations obtained via multigaussian, disjunctive or indicator kriging. In every case, one finds a global proportion equal to 41% approximately. From this, the probability maps (Fig. 3) can be used to delineate the areas for which it is necessary to make conditions favorable for crops to tolerate soil acidity or to reduce this acidity (e.g. by liming) so as to neutralize the toxic concentrations of H^+ and Al^{3+}. A simple way is to impose that the delineated areas cover 41% of the entire field, so as to be consistent with the global estimation. The delineations are shown in Fig. 6, in which one observes that indicator kriging yields results that strongly differ from the three other (Gaussian-based) methods.

Of course, other approaches are possible for delineation. In particular, instead of using the probability maps, one may define a loss function (Goovaerts, 1997, p. 350) to determine whether liming is economically profitable or not and optimize the expected farm income by avoiding excessive liming and fertilization.

4.5. Discussion on the local estimation methods

Indicator kriging may be tedious if many thresholds are considered, while disjunctive kriging is quite convoluted for most practitioners. In contrast, ordinary multigaussian kriging is straightforward and appears to be a helpful tool for soil scientists who wish to map the conditional probabilities of a soil property and incorporate these probabilities in decision-making processes. The steps required for this approach are the following:

1) calculated declustering weights to obtain a representative histogram of the original data;
2) transform these data to normal scores, using the declustered histogram;
3) check the suitability of the two-point normality hypothesis;
4) model the variogram of the normal scores data;
5) perform an ordinary kriging of these data;
6) estimate the conditional distribution of the normal variable at any unsampled location [Eq. (12)], and back-transform it to the original unit.

In addition to its simplicity and straightforwardness, ordinary multigaussian kriging provides consistent results and no order-relation correction is needed. If the neighboring information is abundant, it is practically identical to the conditional expectation, which constitutes the "ideal" estimator of the conditional distributions. Furthermore, it is robust to local variations of the mean value over the field of interest, which is advantageous in under-sampled areas since it avoids the attraction to the prior probabilities observed when using simple kriging. For these reasons, ordinary multigaussian kriging is expected to always outperform bigaussian disjunctive kriging, as in practice the conditions to apply both methods are the same. However, the user should beware that the multigaussian assumption, or at least the bigaussian assumption, is suitable to the available data. Otherwise, one should resort to indicator kriging or to disjunctive kriging under an appropriate non-Gaussian model (Hu, 1989; Liao, 1991; Chilès and Liao, 1993; Chilès and Delfiner, 1999, p. 398–419).

5. Conclusions

The ordinary multigaussian kriging approach provides an unbiased estimation of posterior distributions and transfer functions of a spatial attribute, e.g. a soil property that can be modeled by a Gaussian random field. Although this estimator is theoretically less precise than the conditional expectation and does not allow one to measure the uncertainty prevailing at unsampled locations, it is robust to local variations of the mean value and is therefore suitable when the data do not conform to the ideal stationary model, especially in under-sampled areas. The method is simple and straightforward to apply, since a single variogram analysis and kriging are required. Furthermore, it does not suffer from order-relation problems and provides conditional probabilities that always lie in [0,1].

Like indicator and disjunctive kriging, multigaussian kriging is a helpful method in soil science applications concerned with spatial predictions. In particular, mapping the conditional probabilities of a soil property is of importance for management decisions, which are based on threshold values of this property, such as delineating safe and hazardous areas or identifying zones that are suitable for crop growth and those that must be treated.

Acknowledgements

The author would like to acknowledge the sponsoring by Codelco-Chile for supporting this research, and the French Agricultural Research Centre for International Development (CIRAD) for providing the dataset used in this work. Thanks are due to Dr. Julián Ortiz (University of Chile) and to the anonymous reviewers for their comments on an earlier version of this paper.

References

Barnes, R.J., Johnson, T.B., 1984. Positive kriging. In: Verly, G., David, M., Journel, A.G., Maréchal, A. (Eds.), Geostatistics for Natural Resources Characterization. Reidel, Dordrecht, pp. 231–244.

Chilès, J.P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York.

Chilès, J.P., Liao, H.T., 1993. Estimating the recoverable reserves of gold deposits: comparison between disjunctive kriging and indicator kriging. In: Soares, A. (Ed.), Geostatistics Tróia'92. Kluwer Academic, Dordrecht, pp. 1053–1064.

David, M., 1988. Handbook of Applied Advanced Geostatistical Ore Reserve Estimation. Elsevier, Amsterdam.

Emery, X., 2004. Testing the correctness of the sequential algorithm for simulating Gaussian random fields. Stochastic Environmental Research and Risk Assessment 18 (6), 401–413.

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

Emery, X., 2005b. Simple and ordinary multigaussian kriging for estimating recoverable reserves. Mathematical Geology 37 (3), 295–319.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Hu, L.Y., 1989. Comparing gamma isofactorial disjunctive kriging and indicator kriging for estimating local spatial distributions. In: Armstrong, M. (Ed.), Geostatistics. Kluwer Academic, Dordrecht, pp. 154–165.

Journel, A.G., 1980. The lognormal approach to predicting local distributions. Mathematical Geology 12 (4), 285–303.

Journel, A.G., 1983. Non parametric estimation of spatial distributions. Mathematical Geology 15 (3), 445–468.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London.

Lark, R.M., Ferguson, R.B., 2004. Mapping risk of soil nutrient deficiency or excess by disjunctive and indicator kriging. Geoderma 118, 39–53.

Liao, H.T., 1991. Estimation des réserves récuperables des gisements d'or – comparaison entre krigeage disjonctif et krigeage des indicatrices. Doctoral Thesis, Université d'Orléans. Report no 202, Bureau de Recherches Géologiques et Minières, Orléans, France.

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

Matheron, G., 1989b. Estimating and Choosing: An Essay on Probability in Practice. Springer-Verlag, Berlin.

Oliver, M.A., Webster, R., McGrath, S.P., 1996. Disjunctive kriging for environmental management. Environmetrics 7 (3), 333–357.

Rivoirard, J., 1990. A review of lognormal estimators for in situ reserves. Mathematical Geology 22 (2), 213–221.

Rivoirard, J., 1994. Introduction to Disjunctive Kriging and Nonlinear Geostatistics. Clarendon Press, Oxford.

Van Meirvenne, M., Goovaerts, P., 2001. Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. Geoderma 102, 75–100.

Verly, G., 1983. The multigaussian approach and its applications to the estimation of local reserves. Mathematical Geology 15 (2), 259–286.

Webster, R., Oliver, M.A., 1989. Optimal interpolation and isarithmic mapping of soil properties: VI. Disjunctive kriging and mapping the conditional probability. Journal of Soil Sciences 40, 497–512.

Webster, R., Oliver, M.A., 2001. Geostatistics for Environmental Scientists. Wiley, New York.

Wood, G., Oliver, M.A., Webster, R., 1990. Estimating soil salinity by disjunctive kriging. Soil Use and Management 6 (3), 97–104.