Hydrology and Earth System Sciences, 5(2), *215–223* (2001) © EGS

Hydrology & Earth System Sciences

Influence of parameter estimation uncertainty in Kriging: Part 1 – Theoretical Development

Ezio Todini

Department of Earth and Geo-Environmental Sciences, University of Bologna, Italy
Email: todini@geomin.unibo.it

Abstract

This paper deals with a theoretical approach to assessing the effects of parameter estimation uncertainty both on Kriging estimates and on their estimated error variance. Although a comprehensive treatment of parameter estimation uncertainty is covered by full Bayesian Kriging at the cost of extensive numerical integration, the proposed approach has a wide field of application, given its relative simplicity. The approach is based upon a truncated Taylor expansion approximation and, within the limits of the proposed approximation, the conventional Kriging estimates are shown to be biased for all variograms, the bias depending upon the second order derivatives with respect to the parameters times the variance-covariance matrix of the parameter estimates. A new Maximum Likelihood (ML) estimator for semi-variogram parameters in ordinary Kriging, based upon the assumption of a multi-normal distribution of the Kriging cross-validation errors, is introduced as a mean for the estimation of the parameter variance-covariance matrix.

Keywords: Kriging, maximum likelihood, parameter estimation, uncertainty

Introduction

In the study of spatially distributed hydrological phenomena, one of the most important advantages of statistical methods such as Kriging (Matheron, 1970; Delhomme, 1978; de Marsily, 1986) and objective analysis (Gandin, 1970), is their ability to quantify the uncertainty in the derived estimates. In Kriging and other techniques of spatial analysis, statistical hypotheses are made in identifying and evaluating the multidimensional spatial structure of the hydrological process of interest. In the specific case of Kriging, where linear minimum variance unbiased estimation is adopted, this is equivalent to the selection of a functional form, and the estimation of the relevant parameters for the main trend (first order moment) and for the covariance or the semi-variogram or the generalised covariance (second order moments).

Although the statistical inference of model form and of parameter values from available data obviously entails some degree of arbitrariness and uncertainty, the theoretical expressions generally used for the expected value and the error variance of the estimated variable do not take into account these sources of approximation. In particular, no account of the effects of uncertainty in parameter estimates (which can be quite large) is commonly made, which may lead to three major inconsistencies: a bias in the point estimation of the multi-dimensional variable, a different spatial distribution of the measure of uncertainty and last but not least an inappropriate model choice for the variogram.

In spite of its great concern to hydrologists and the availability of Bayesian approaches (Kitanidis, 1986), the problem of evaluating soundly the effects of model choice and parameter estimation on the value and reliability of predictions has so far received scant attention in Kriging applications. Past research has been concerned mainly with building objective and efficient algorithms for parameter estimation (see for instance Zimmerman and Zimmerman, 1991; Cressie, 1993) and for structure selection (Davis and David, 1978, Starks and Fang, 1982) while the effect of sample size on the precision of estimates has also been studied (Lettenmaier, 1981).

Unfortunately, conventional users of Kriging tend to separate the phase of the estimation of model parameters, using least squares if enough data are available, or subjective graphical methods if not, from the estimation and interpretation of results, which leads finally to the model choice. This separation of the two phases of parameter estimation and model identification leads to the selection of a model on the basis of the variance of the errors of estimate of the interpolating function (the error contrasts), once values for the previously estimated parameters are provided. No mention is generally made of the fact that, due to the interdependence of observations, the uncertainty of the parameter estimates can be quite large, and its effect on the variance of the errors of estimate will depend strongly upon the structure of the adopted model.

Bayesian approaches were introduced in Kriging applications by Kitanidis (1986) while further developments can be found in Omre and Halvorsen (1989), Le and Zidek (1992), Diggle *et al.* (1998), Woodbury and Ulrych (2001). The Bayesian approaches have the advantage of allowing the assessment and the reduction of the uncertainty on model parameters at the cost of extensive numerical integration, generally based upon Monte Carlo or Markov Chain Monte Carlo techniques.

The aim of the present paper is to present a methodology which, although approximate, can be viewed as an alternative to the Bayesian approaches and, given its relative simplicity, used on most Kriging applications. The work focuses on the assessment of the influence of parameter uncertainty of the Kriging estimates for a given model structure, while the possible source of errors induced by the wrong model choice is not dealt with. As a follow-up of the present work, the question of model choice can be addressed more appropriately within the frame of a statistical acceptance rejection test, yet to be defined, by comparing the sampling distribution of the empirical semi-variogram classes to the approximate distribution of the theoretic semi-variogram which can be obtained as a function of the parameter estimates.

Problem formulation

The Kriging problem can be formulated in two dimensions as the process of estimation of the value $\hat{z}_0$, and error variance of the estimate σ_0^2, of a spatial variable at a point (x_0, y_0) where the true (unknown) value is z_0, under the following assumptions:

i. the interpolating equation is based upon the linear interpolator given by Eqn. (1) which allows the value of z_0 in $s_0 = (x_0, y_0)$ to be expressed as a function of z_k at $s_k = (x_k, y_k)$, with *k=1, n*, as:

$$\hat{z}_0 = z^{*T} \lambda^* \tag{1}$$

where z^* is the [*n, 1*] vector of observations and λ^* the [*n, 1*] vector of interpolating weights;

ii. z^* is an intrinsic random function assumed to have a constant but unknown mean, whose system of ordinary Kriging equations, under the assumption of weak stationarity, can be written as:

$$\gamma = \Gamma \lambda \tag{2}$$

In Eqn. (2) γ is the [*n+1,1*] vector defined as:

$$\gamma = \begin{bmatrix} \gamma^* \\ \cdots \\ 1 \end{bmatrix} \tag{3}$$

with γ^* the [*n, 1*] vector of variogram values describing the spatial dependence between the estimation point (x_0, y_0) and the measurement points. Γ is the [*n+1,n+1*] matrix defined as:

$$\Gamma = \begin{bmatrix} \Gamma^* & \vdots & u \\ \cdots & \cdots & \cdots \\ u^T & \vdots & 0 \end{bmatrix} \tag{4}$$

where Γ^* is the [*n, n*] symmetric matrix of variogram values describing the spatial dependence among the *n* measurement points while *u* is an [*n, 1*] unit vector and λ is defined as:

$$\lambda = \begin{bmatrix} \lambda^* \\ \cdots \\ \mu \end{bmatrix} \tag{5}$$

with μ a Lagrange multiplier arising from the unbiasedness condition $E\{\hat{z}_0 - z_0\} = 0$;

iii. the semi-variogram, whose classical estimate $\gamma(h) = E\{[z(s) - z(s+h)]^2\}$ was proposed by Matheron (1962), is isotropic and its form can be chosen among intrinsically linear or non linear models.

Bearing in mind that $\gamma(0) = 0$ in the absence of measurement errors, the expression for a number of semi-variogram models used in this paper is given as Eqn. (6) for *h*>0, where *h* is the distance between two points and *(p, w, a)* are model parameters, *p* being generally known as the "nugget" effect,

$$\gamma(h) = p + \omega\left[1 - e^{-h/\alpha}\right] \qquad \text{Exponential}$$

$$\gamma(h) = p + \omega\left[1 - e^{-(h/\alpha)^2}\right] \qquad \text{Gaussian} \tag{6}$$

$$\begin{cases} \gamma(h) = p + \omega\left[1.875\dfrac{h}{\alpha} - 1.25\left(\dfrac{h}{\alpha}\right)^3 + 0.375\left(\dfrac{h}{\alpha}\right)^5\right] & h < \alpha \\ \gamma(h) = p + \omega & h \geq \alpha \end{cases}$$

Modified Spherical

Because Maximum Likelihood (ML) will be used to estimate the semi-variogram parameters, the spherical model was modified slightly from the original equation, as can be seen from Eqn. (6), to guarantee the continuity of the function and of its first and second order derivatives in $h=\alpha$ as required by ML to be asymptotic minimum variance (Kendall and Stuart, 1973), without modifying its convexity. The differences between the modified spherical semi-variogram and the original one are very small, as can be shown by plotting the two curves.

From the Kriging literature (see Matheron, 1970; de Marsily, 1986), under the above stated conditions, the estimation error variance of the interpolating model can be expressed as:

$$\sigma_0^2 = E\left\{(\hat{z}_0 - z_0)^2\right\} = \gamma^{*T}\lambda^* + \mu = \gamma^T\lambda \tag{7}$$

For a given set of observations, the estimate $\hat{z}_0$, from Eqn. (1), can be expressed as:

$$\hat{z}_0 = z^{*T}\begin{bmatrix} I & \vdots & 0\end{bmatrix}\begin{bmatrix}\lambda^* \\ \cdots \\ \mu\end{bmatrix} = z^T\lambda \tag{8}$$

in which the following substitution was made for computational convenience:

$$z = \begin{bmatrix} I \\ \cdots \\ 0\end{bmatrix} \quad z^* = \begin{bmatrix} z^* \\ \cdots \\ 0\end{bmatrix} \tag{9}$$

and, from Eqn. (2) and for any given model of Eqn. (6):

$$\lambda = \begin{bmatrix}\lambda^* \\ \cdots \\ \mu\end{bmatrix} = \begin{bmatrix}\Gamma^* & \vdots & u \\ \cdots & \cdots & \cdots \\ u^T & \vdots & 0\end{bmatrix}^{-1}\begin{bmatrix}\gamma^* \\ \cdots \\ 1\end{bmatrix} \tag{10}$$

The cross-validation errors are estimated by taking one value z_k at a time out of the whole set of *n* observations, by computing from Eqn. (8) at that point (x_k, y_k) the predicted value $\hat{z}_k$ and by subtracting it from the observed value z_k.

Influence of parameter estimation uncertainty

To assess the influence of parameter estimation uncertainty on the Kriging estimates, the following considerations must be made. For a given set of observations, the estimate $\hat{z}_0$ is obviously a function of the parameters and position (x_0,y_0) (see Eqn. 8) since, from Eqn. (2) and for any given model of Eqn. (6) one can see the dependence of λ on the model parameters ϑ, through Γ^* and γ^* and dependence on (x_0,y_0) through γ^*.

$$\lambda = \begin{bmatrix}\lambda^* \\ \cdots \\ \mu\end{bmatrix} = \begin{bmatrix}\Gamma^* & \vdots & u \\ \cdots & \cdots & \cdots \\ u^T & \vdots & 0\end{bmatrix}^{-1}\begin{bmatrix}\gamma^* \\ \cdots \\ 1\end{bmatrix} = \Gamma(\vartheta)^{-1}\gamma(\vartheta) = \lambda(\vartheta) \tag{11}$$

Γ(ϑ) and *γ(ϑ)* are expressed as a function of one of the models introduced in Eqn. (6) while ϑ is the *(m, 1)* parameter vector *(p,ω,α)*, with *m=1,2,3*, the number of unknown parameters to be estimated.

It is therefore possible to express directly the dependence of $\hat{z}_0$ on the set of parameters, as:

$$\hat{z}_0 = \begin{bmatrix} z^{*T} & \vdots & 0\end{bmatrix}\begin{bmatrix}\Gamma^* & \vdots & u \\ \cdots & \cdots & \cdots \\ u^T & \vdots & 0\end{bmatrix}^{-1}\begin{bmatrix}\gamma^* \\ \cdots \\ 1\end{bmatrix} = z^T\Gamma(\vartheta)^{-1}\gamma(\vartheta) = \hat{z}_0(\vartheta) \tag{12}$$

If the parameter vector ϑ is no longer considered as a set of *a priori* fixed values (the common assumption), but rather as $\hat{\vartheta}$, an estimated quantity depending both on the model structure and upon the observation vector z^*, it is easy to show that, due to the parameter estimation error, both $\hat{z}_0(\hat{\vartheta})$ and $Var\{\hat{z}_0(\hat{\vartheta})\}$ differ from $E\{\hat{z}_0(\vartheta)\}$ and $Var\{\hat{z}_0(\vartheta)\}$ respectively, even when $\hat{\vartheta}$, the parameter estimate obtained with a specific vector of observations, actually coincides with ϑ.

If the general relationship expressed by Eqn. (12) is sufficiently well behaved and if the coefficients of variation of the parameters ϑ (which depend on the degree of non-linearity of *γ(ϑ)* in the neighbourhood of the estimates of the expected values $E\{\hat{\vartheta}\}$ are small, following Todini and Ferraresi (1996), it is convenient to approximate the expected value and the variance of the Kriging estimate $\hat{z}_0$, taken as a function of the parameters, by a truncated Taylor-series expansion of $\hat{z}_0(\vartheta)$ about $E\{\hat{\vartheta}\}$ (Benjamin and Cornell, 1970). In addition, since $E\{\hat{\vartheta}\}$ is not known, it is common practice to use the estimated parameter value $\hat{\vartheta}$ instead, to give:

$$E\left\{\hat{z}_0(\vartheta) - \hat{z}_0(\hat{\vartheta})\right\} \cong \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} Cov\left\{\hat{\vartheta}_i, \hat{\vartheta}_j\right\}\left.\frac{\partial^2 \hat{z}_0(\vartheta)}{\partial\vartheta_i\partial\vartheta_j}\right|_{\vartheta=\hat{\vartheta}} \tag{13}$$

$$Var\left\{\hat{z}_0(\vartheta) - \hat{z}_0(\hat{\vartheta})\right\} \cong \sum_{i=1}^{m}\left.\frac{\partial\,\hat{z}_0(\vartheta)}{\partial\,\vartheta_i}\right|_{\vartheta=\hat{\vartheta}}\sum_{j=1}^{m} Cov\left\{\hat{\vartheta}_i, \hat{\vartheta}_j\right\}\left.\frac{\partial\,\hat{z}_0(\vartheta)}{\partial\,\vartheta_j}\right|_{\vartheta=\hat{\vartheta}} \tag{14}$$

where:

$$Cov\{\hat{\vartheta}_i, \hat{\vartheta}_j\} = \tfrac{1}{2}\,\mathrm{Tr}\left(\Gamma^{-1}\frac{\partial \Gamma}{\partial \theta_i}\Gamma^{-1}\frac{\partial \Gamma}{\partial \theta_j}\right)\Bigg|_{\vartheta=\hat{\vartheta}} \tag{15}$$

as derived in Appendix A, and where the expressions for $\frac{\partial \hat{z}_0}{\partial \vartheta_i}$, and $\frac{\partial^2 \hat{z}_0}{\partial \vartheta_i \partial \vartheta_j}$, namely:

$$\frac{\partial \hat{z}_0}{\partial \vartheta_i} = z^T \Gamma^{-1}\left(\frac{\partial \gamma}{\partial \vartheta_i} - \frac{\partial \Gamma}{\partial \vartheta_i}\Gamma^{-1}\gamma\right) \tag{16}$$

$$\frac{\partial^2 \hat{z}_0}{\partial \vartheta_i \partial \vartheta_j} = z^T \Gamma^{-1}\left[\frac{\partial^2 \gamma}{\partial \vartheta_i \partial \vartheta_j} - \frac{\partial^2 \Gamma}{\partial \vartheta_i \partial \vartheta_j}\Gamma^{-1}\gamma - \frac{\partial \Gamma}{\partial \vartheta_i}\Gamma^{-1}\left(\frac{\partial \gamma}{\partial \vartheta_j} - \frac{\partial \Gamma}{\partial \vartheta_j}\Gamma^{-1}\gamma\right) - \frac{\partial \Gamma}{\partial \vartheta_j}\Gamma^{-1}\left(\frac{\partial \gamma}{\partial \vartheta_i} - \frac{\partial \Gamma}{\partial \vartheta_i}\Gamma^{-1}\gamma\right)\right] \tag{17}$$

have been derived in Todini and Ferraresi (1996).

Equation (13) clearly shows, within the limits of the proposed approximation, that the expected value of $\hat{z}_0(\vartheta)$ does not correspond to the value estimated at $\hat{\vartheta}$, namely $\hat{z}_0(\hat{\vartheta})$; a bias correction is needed which depends on the second order derivatives of $\hat{z}_0(\vartheta)$ with respect to the parameters as well as on the covariance matrix of the parameter estimates, all computed for $\vartheta = \hat{\vartheta}$.

Equation (14) shows, within the same limitations expressed above, that the increase in the variance of the estimation error of $\hat{z}_0(\vartheta)$ depends upon the first order derivatives of the function $\hat{z}_0(\vartheta)$ as well as on the covariance matrix of the parameter estimates computed for $\vartheta = \hat{\vartheta}$.

To estimate the corrective terms of Eqns. (13) and (14) for bias correction and variance adjustment, one needs to estimate the parameter values and their covariance matrix, defined as the inverse of the Fisher information matrix. These estimates can be obtained using the ML estimator to be described in the following section.

The maximum liklihood estimator

In Kriging applications, it is usual practice to fit model semi-variogram parameters subjectively by comparing the theoretical semi-variogram graphically with mean values of its estimates obtained from the observations, as is done, for instance, in Geo-EAS (Englund and Sparks, 1988) or in GSLIB (Deutsch and Journel, 1992), which are widely available Kriging packages.

Nevertheless, more objective techniques are also available. Following the comprehensive reviews given by Zimmerman and Zimmerman (1991), and by Cressie (1993), these can be divided into least squares based techniques in the space of the semi-variogram (Journel and Huijbregts, 1978; Cressie and Hawkins, 1980; Cressie, 1985); least squares techniques in the space of observations defined in the form of generalised covariance expressed as a linear function of parameters (Delfiner, 1976; Kitanidis, 1983); Maximum Likelihood in the space of residuals from a linear trend (Mardia and Marshall, 1984); Maximum Likelihood in the space of cross-validation errors (Samper and Newman, 1989); Maximum Likelihood in the space of "error contrasts" (Kitanidis, 1983), using what is known as Restricted Maximum Likelihood (REML) (Patterson and Thompson, 1971;1974). All these techniques have pros and cons that will be addressed briefly in the sequel, but this paper will focus only on the ML and REML type estimators, since they allow for the derivation of the covariance matrix of the parameters, which can be computed as the inverse of the Fisher information matrix, and can be used for investigating the effect of parameter uncertainty over the Kriging estimates, as advocated by Kitanidis (1983) and proposed by Todini and Ferraresi (1996).

With respect to the use of ML and REML type estimators, it must be pointed out that they are asymptotically minimum variance estimators only if the Likelihood function is continuous and twice differentiable with respect to the parameters over the entire field of existence (Kendall and Stuart, 1973); this condition does not hold for the classical spherical and cubic variograms, given the discontinuity in the second derivative at $h = \alpha$. This can be overcome by a slight modification of the expressions of the spherical and of the cubic semi-variograms by imposing the continuity of the first and of the second order derivatives in $h = \alpha$ as proposed in this paper (see Eqn. 6).

The formulation presented in Todini and Pellegrini (1999) follows the development of Samper and Newman (1989), assuming a multi-normal distribution of Kriging cross-validation errors but, instead of imposing their independence, it accounts for their full covariance matrix. In addition, by realising that the cross-validation errors can be viewed as a special case of error contrasts, and by using the relationships among the Kriging variables, the problem is reduced to a computationally convenient formulation depending only on the observations, on the selected semi-variogram and on its parameters.

The *n-1* weights for each of the *n* points z_k used for cross-validation are calculated using Eqn. (10) by omitting the k^{th} row and column as appropriate. These special λ_k^* [*n*-1, 1] size distinct (k = 1, 2, … ,n) vectors of cross-validation weights relevant to each of the *n* observation points can be re-arranged to form a matrix Λ of size [*n, n*] by inserting a zero along its principal diagonal. This allows the Kriging cross-validation errors to be represented as an [*n, 1*] vector:

$$\varepsilon = \left[\Lambda^T - I\right] z^* = \left[\Lambda^T - I \;\; \vdots \;\; \mu\right] \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} z^* = \left[\Lambda^T - I \;\; \vdots \;\; \mu\right] z \qquad (18)$$

where z^* is the vector of observed data, while z is the augmented vector, its last element being set equal to zero, as in Eqn. (9), Λ is the [*n, n*] augmented matrix containing the cross-validation weights for the *n* measurements points, I is the [*n, n*] identity matrix and *m* is a vector of Lagrange multipliers which disappear. Even if *z* is not a Gaussian random field, it still seems reasonable to assume the cross-validation errors to be normally distributed, with zero mean and variance-covariance matrix V_ε (to be derived in Appendix B), i.e.:

$$\varepsilon = N(0, V_\varepsilon) \qquad (19)$$

on the grounds of the unbiasedness of Kriging for known parameters and the fact that they are obtained as a linear combination of the data. Wilks (1962) shows in fact, on the basis of earlier work due to Wald and Wolfowitz (1944), that the limiting distribution of linear functions in large samples from large finite populations is the normal distribution, even if the underlying variables are non Gaussian.

The joint probability density function of the *n* Gaussian cross validation errors (with zero mean) can thus be expressed as:

$$f(\varepsilon|\vartheta) = (2\pi)^{-n/2} \|V_\varepsilon\|^{-1/2} Exp\left(-\tfrac{1}{2}\varepsilon^T V_\varepsilon^{-1} \varepsilon\right) \qquad (20)$$

where ε is the vector of cross validation errors, ϑ denotes the vector of semi-variogram structural parameters (namely *p*, *w* and *a*), and the symbol $|\bullet|$ denotes the determinant. The vector of parameters ϑ will be determined through Maximum Likelihood estimation, namely by maximising the Likelihood:

$$L(\vartheta|\varepsilon) = f(\varepsilon|\vartheta) \qquad (21)$$

or by minimising the negative Log-likelihood function defined as:

$$\mathscr{L}(\vartheta\,|\,\varepsilon) = -ln\,L(\vartheta\,|\,\varepsilon) = \frac{n}{2} ln(2\pi) + \tfrac{1}{2} ln\|V_\varepsilon\| + \tfrac{1}{2}\varepsilon^T V_\varepsilon^{-1}\varepsilon \qquad (22)$$

Within the frame of previous research aimed at assessing the influence of parameter estimation variance in Kriging, Todini and Ferraresi (1996) (similarly to what was done by Samper and Newman (1989)) made the hypothesis of normal independent cross-validation errors by assuming the covariance matrix V_ε to be diagonal. The application of the technique to real data with this independence hypotheses showed that the assumption of independent errors may involve significant distortions in the estimation of the parameters particularly for certain types of variograms (Todini *et al*., 2001).

The new ML estimator also assumes a multi-normal distribution of the cross-validation errors but with a full covariance matrix V_ε that can be expressed either as a function of the chosen variogram and of the Kriging weights, as given for instance by de Marsily (1986) and by Samper and Newman (1989), or as the more suitable matrix expression of Eqn. (23) derived in Appendix B and used in this paper.

$$V_\varepsilon = -\Sigma\left[I \;\; \vdots \;\; 0\right] \Gamma^{-1} \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} \Sigma \qquad (23)$$

where Σ is the diagonal [*n, n*] matrix containing the n cross-validation error variances given by Eqn. (7).

Unfortunately, due to the Kriging normalisation constraint (i.e. the sum of the Kriging weights λ_k is always equal to one), the matrix V_ε is not invertible, its rank being equal to [*n-1*]. It is therefore necessary and convenient to define an ortho-normal transformation of the [*n*] dimensional vector of cross validation errors ε to project it on an [*n-1*] dimensional space by the principal component technique, disregarding the null eigen-value and the corresponding eigen-vectors. The following relationship holds:

$$P V_\eta P^T = -\Sigma\left[I \;\; \vdots \;\; 0\right] \Gamma^{-1} \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} \Sigma = V_\varepsilon \qquad (24)$$

where V_η is an [*n-1, n-1*] diagonal matrix of non-null eigen-values and P the [*n, n-1*] matrix of relevant eigen-vectors. This transformation allows one to define η as follows:

$$\eta = -P^T \Sigma\left[I \;\; \vdots \;\; 0\right] \Gamma^{-1} \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} z^* \qquad (25)$$

Consequently, the Log-likelihood becomes:

$$\mathscr{L}(\vartheta\,|\,\eta) = -ln\,L(\vartheta\,|\,\eta) = \frac{n-1}{2} ln(2\pi) + \tfrac{1}{2} ln\|V_\eta\| + \tfrac{1}{2}\eta^T V_\eta^{-1}\eta \qquad (26)$$

with:

$$V_\eta = -P^T \Sigma\left[I \;\; \vdots \;\; 0\right] \Gamma^{-1} \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} \Sigma\, P \qquad (27)$$

Equation (26) with the diagonal covariance matrix V_η given in Eqn. (27) can be viewed as a special case of Restricted Maximum Likelihood (REML) in which the cross-validation errors are a special type of independent error contrasts. The use, in Eqn. (26), of the expression provided by Eqn. (27) is not computationally convenient, since it requires a new ortho-normal transformation at each parameter modification.

Fortunately it can be shown that a more convenient and more general formulation can be found. Pre- and post-multiplying Eqn. (24) by Σ^{-1}, and substituting the resulting equation in Eqn. (25) the following relationship can be derived:

$$\eta = V_\eta P^T \Sigma^{-1} z^* \tag{28}$$

and using Eqns. (26) and (27) to derive the quadratic term appearing in the Log-likelihood, one obtains:

$$\eta^T V_\eta^{-1} \eta = z^{*T} \Sigma^{-1} P V_\eta V_\eta^{-1} V_\eta P^T \Sigma^{-1} z^* = z^{*T} \Sigma^{-1} P V_\eta P^T \Sigma^{-1} z^*$$

$$= -z^{*T} \begin{bmatrix} I & \vdots & 0 \end{bmatrix} \Gamma^{-1} \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} z^* = -z^T \Gamma^{-1} z \tag{29}$$

Thus, the new Log-likelihood can be written easily as:

$$\mathscr{L}(\vartheta \mid \eta) = -\ln L(\vartheta \mid \eta) = \frac{n-1}{2} \ln(2\pi) - \tfrac{1}{2} \ln \|\Gamma\| - \tfrac{1}{2} z^T \Gamma^{-1} z \tag{30}$$

which, although expressed in terms of the vector of observations, is defined in the multi-normal space of the Kriging cross-validation errors. Note that Eqn. (30) does not imply that $\|\Gamma\| = \|V_\eta\|$, because the change of variable in the Likelihood function requires the change of the normalising constant.

This theoretically remarkable and computationally efficient result, shows that the Log-likelihood function is independent of the Kriging weights Λ, depending only on the observations, the semi-variogram model and its parameters.

Conclusions

The theoretical approach based upon a truncated Taylor expansion approximation, introduced in this paper, aimed at assessing the effect of parameter estimation uncertainty both on Kriging estimates and on their estimated error variance, has given rise to a number of interesting considerations.

To apply the proposed methodology, a new ML estimator was also developed, that can be viewed as a special case of Restricted Maximum Likelihood with the additional advantage of a computationally efficient formulation of the Likelihood function to be minimised.

The results show that, within the limits of the proposed approximation, the conventional Kriging estimates are biased for all semi-variogram models, the bias depending upon the second order derivatives with respect to the parameters times the variance-covariance matrix of the parameter estimates. The entity and the importance of the bias and of the increase in variance will be elaborated in Part 2 by means of a case study based upon the average yearly precipitation over the Veneto Region in Italy (Todini *et al*., 2001).

Although, for the sake of simplicity, the development of the methodology in this paper does not consider errors in data, it can be expanded easily to take them into account. Extensions to co-Kriging as well as the generalisation of the proposed procedure to non-stationary fields are straightforward: in the latter case the spatial correlation structure must be expressed in terms of the generalised covariance instead of the variogram function and an adequate number of Lagrange multipliers has to be introduced, according to the degree of the polynomial drift.

References

Benjamin, J.R. and Cornell, C.A., 1970. *Probability , Statistics and Decision for Civil Engineers*. McGraw Hill, NewYork.

Cressie, N.A. and Hawkins, D.M., 1980. Robust estimation of the variogram. *Math. Geol*., **12**, 115–125.

Cressie, N.A., 1985. Fitting variogram models by weighted least squares. *Math. Geol.*, **17**, 563–586.

Cressie, N.A., 1993. *Statistics for Spatial Data*. Wiley, New York.

Davis, M.W.D. and David, M., 1978. Automatic Kriging and contouring in the presence of trends, *J. Can. Petrol. Technol.*, **17**, 1–10.

Delfiner, P., 1976. Linear estimation of non-stationary spatial phenomena. In: *Advanced Geostatistics in the Mining Industry,* M. Guarascio, M. Davis and C. Huijbregts (Eds.), Reidel, Dordrecht, The Netherlands. 49–68.

Delhomme, J.P., 1978. Kriging in the hydrosciences, *Adv. Water Resources*, **1**, 251–266.

de Marsily, G., 1986. *Quantitative hydrogeology*, Academic Press, London.

Deutsch, C.V.D. and Journel, A.G., 1992. *GSLIB Geostatistical Software Library and User's Guide*, Oxford University Press, Oxford.

Diggle, P.J., Tawn J.A. and Moyeed, R.A., 1998. Model-based geostatistics. *Appl. Roy. Stat. Soc*., **47**, 299–350.

Dubrule, O., 1983. Cross-validation of Kriging in a unique neighbourhood, *J. Math. Geol.*, **15**, 687–699.

Englund, E. and Sparks, A., 1988. *Geostatistical Environmental Assessment Software (Geo-EAS)*. U.S. Environmental Protection Agency, EPA600/4, 88/033, Las Vegas, Nevada.

Gandin, L.S., 1970. The planning of meteorological station network, *WMO Tech. Note 111*, World Meteorological Organization, Geneva.

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

Kendall, M.G. and Stuart, A., 1973. *The advanced theory of statistics*, Vol. 2. Griffin. London.

Kitanidis, P.K., 1983. Statistical estimation of polynomial generalised covariance functions and hydrologic applications, *Water Resour. Res*., **19**, 909–921.

Kitanidis, P.K., 1986. Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resour. Res.*, **22**, 499–507.

Le, N.D. and Zidek, J.V., 1992. Interpolation with uncertain spatial covariances: a Bayesian alternative to Kriging. *J. Multivariate Anal.*, **43**, 351–374.

Lettenmaier, D.P., 1981. Data requirements for Kriging: estimation and network design, *Water Resour. Res.*, **17**, 1641–1650.

Mardia, K.V. and Marshall, R.J., 1984. Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika*, **71**, 135–146.

Matheron, G. 1970. La théorie des variables regionaliséés et ses applications, *Cah. Cent. Morphol*. Math., **5**.

Omre, H. and Halvorsen, K.B., 1989. The Bayesian bridge between simple and universal Kriging. *Math. Geol*., **21**,767–786.

Patterson, H.D. and Thompson, R., 1971. Recovery of interblock information when block sizes are unequal. *Biometrika*, **58**, 545–554.

Patterson, H.D. and Thompson, R., 1974. Maximum likelihood estimation of component of variance. *Proc. 8th Int. Biometric Conf., Biometric Society*, Washington, DC. 197–207.

Samper, F.J. and Neuman, S.P., 1989. Estimation of spatial covariance structures by adjoint state maximum likelihood cross calidation: 1. Theory. *Water Resour. Res*., **25**, 351–362; 2. Synthetic experiments. *Water Resour. Res*., **25**, 363–371; 3. Application to hydrochemical and isotopic data. *Water Resour. Res*., **25**, 373–384.

Starks, T.H. and Fang, J.H., 1982. On the estimation of the generalized covariance function, *Math. Geol*., **14**, 57–64.

Todini, E. and Ferraresi, M., 1996. Influence of parameter estimation uncertainty in Kriging, *J. Hydrol*., **175**, 555–566.

Todini, E. and Pellegrini, F., 1999. A maximum likelihood estimator for semi-variogram parameters in Kriging. In: *geoENV II – Geostatistics for Environmental Applications*, J. Gòmez-Hernàndez, A. Soares and R. Froidevaux (Eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands. 187–198.

Todini E., Pellegrini, F. and Mazzetti, C., 2001. Influence of parameter estimation uncertainty in Kriging. Part 2 - Test and case study applications. *Hydrol. Earth System Sci*., **5**, xx-xx.

Wald, A. and Wolfowitz, J., 1944. Statistical tests based on permutations of the observations, *Ann. Math. Stat.*, **15**, 358–372.

Wilks, S.S., 1962. *Mathematical Statistics*. Wiley, New York.

Woodbury, A.D. and Ulrych, T.J., 2000. A full-Bayesian approach to the groundwater inverse problem for steady state flow, *Water Resour. Res.* (in press).

Zimmerman, D.L. and Zimmerman, M.B., 1991. A comparison of spatial semivariogram estimators and corresponding ordinary Kriging predictors. *Technometrics*, **33**, 77–91.

Appendix A

The variance-covariance matrix of parameter estimate

The parameter variance-covariance matrix $Cov\{\hat{\vartheta}_i, \hat{\vartheta}_j\}$ appearing in Eqns. (15) and (16) can be computed as the inverse of the Fisher information matrix, that is the inverse of the negative of the expected value of the Hessian of the Log-likelihood function defined over the parameters (Kendall and Stuart, 1973):

$$Cov\{\hat{\vartheta}_i,\hat{\vartheta}_j\} = E\left\{\left(\hat{\vartheta}_i - \vartheta_i\right)\left(\hat{\vartheta}_j - \vartheta_j\right)^T\right\} = E\{H(\mathscr{L})\}^{-1} = E\left\{\frac{\partial \mathscr{L}}{\partial \vartheta_i}\left(\frac{\partial \mathscr{L}}{\partial \vartheta_j}\right)^T\Bigg|_{\vartheta=\hat{\vartheta}}\right\} \quad \text{(A1)}$$

H being the Hessian of $\mathscr{L}$ with respect to ϑ, bearing in mind that the relationship holds asymptotically for the linear, monomial, exponential and Gaussian models, and is approximate for the spherical and cubic models, as previously mentioned.

Kitanidis (1983) , by using Gaussian moment factoring, shows that for Gaussian-distributed data a simplified expression for the parameter covariance matrix can be obtained as a function of the covariance matrix appearing in the Log-likelihood function and of its first order derivatives with respect to the parameters. By applying Kitanidis derivation to the Log-likelihood given by Eqn. (32), the following result can be obtained for the *i,j* generic element of $Cov\{\hat{\vartheta}_i, \hat{\vartheta}_j\}$:

$$Cov\{\hat{\vartheta}_i,\hat{\vartheta}_j\} = E\left\{\left[\frac{\partial \mathscr{L}}{\partial \vartheta_i}\left(\frac{\partial \mathscr{L}}{\partial \vartheta_j}\right)^T\right]_{\vartheta=\hat{\vartheta}}\right\} = \tfrac{1}{2}\,\mathrm{Tr}\left(\Gamma^{-1}\frac{\partial \Gamma}{\partial \theta_i}\Gamma^{-1}\frac{\partial \Gamma}{\partial \theta_j}\right)\Bigg|_{\vartheta=\hat{\vartheta}} \quad \text{(A2)}$$

Equation (A2) allows for the estimation of all the elements of the variance-covariance matrix of the ML estimates of Kriging parameters to be used in Eqns. (15) and (16).

Appendix B

The derivation of the covariance matrix of the cross-validation errors

As mentioned in the argument leading to Eqn. (18), for a given estimate of Kriging parameters $\hat{\vartheta}$, the cross-validation errors ε are estimated by taking one value z_k at a time out of the whole set of n observations, by computing at that point (x_k, y_k) the weights λ_k from Eqn. (10) and by subtracting the observed value z_k from the predicted value $\hat{z}_k$ obtained using Eqn. (8) in its reduced form.

Once all the λ_k have been computed and re-arranged in Λ, it is possible to evaluate the covariance matrix of the cross-validation errors to be used in the expression of the Likelihood function. This is done by taking into account that the following relationship holds:

$$\Gamma \begin{bmatrix} \Lambda - I \\ \cdots \\ \mu^T \end{bmatrix} = \begin{bmatrix} \Gamma^* & \vdots & u \\ \cdots & \cdots & \cdots \\ u^T & \vdots & 0 \end{bmatrix} \begin{bmatrix} \Lambda - I \\ \cdots \\ \mu^T \end{bmatrix} = \begin{bmatrix} \Gamma^*(\Lambda - I) + u\mu^T \\ \cdots \\ u^T(\Lambda - I) \end{bmatrix}$$

$$= \begin{bmatrix} \Gamma^*(\Lambda - I) + u\mu^T \\ \cdots \\ 0 \end{bmatrix} = \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} \Sigma \qquad \text{(B1)}$$

where Σ is the diagonal [n, n] matrix containing the n cross-validation error variances given by Eqn. (7) and where $u^T(\Lambda - I)$ equals zero, because for any given point k, the corresponding weights λ_k^* add to one.

Equation (B1) can be demonstrated by multiplying, taking into account the matrix partitioning, matrix Γ times the generic k[th] column of matrix $\begin{bmatrix} \Lambda - I \\ \cdots \\ \mu^T \end{bmatrix}$ to give:

$$\Gamma \left(\begin{bmatrix} \lambda_k^* \\ \cdots \\ 0 \\ \cdots \\ \mu_k \end{bmatrix} - \begin{bmatrix} 0 \\ \cdots \\ 1 \\ \cdots \\ 0 \end{bmatrix} \right) = \begin{bmatrix} \Gamma^* & \vdots & u \\ \cdots & \cdots & \cdots \\ u^T & \vdots & 0 \end{bmatrix} \begin{bmatrix} \lambda_k^* \\ \cdots \\ -1 \\ \cdots \\ \mu_k \end{bmatrix} = \begin{bmatrix} \Gamma_k^* & \vdots & \gamma_k^* & \vdots & u \\ \cdots & \vdots & \cdots & \vdots & \cdots \\ \gamma_k^{*T} & \vdots & 0 & \vdots & 1 \\ \cdots & \vdots & \cdots & \vdots & \cdots \\ u^T & \vdots & 1 & \vdots & 0 \end{bmatrix} \begin{bmatrix} \lambda_k^* \\ \cdots \\ -1 \\ \cdots \\ \mu_k \end{bmatrix}$$

$$= \begin{bmatrix} \Gamma_k^* \lambda_k^* - \gamma_k^* + u\mu_k \\ \cdots \\ \gamma_k^{*T} \lambda_k^* + \mu_k \\ \cdots \\ u^T \lambda_k^* - 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \cdots \\ \sigma_k^2 \\ \cdots \\ 0 \end{bmatrix} \qquad \text{(B2)}$$

The result of this product, shows that the resulting column vector has all null values except for the k[th] element, which is equal to σ_k^2 since by definition $\Gamma_k^* \lambda_k^* + u\mu_k = \gamma_k^*$; $\gamma_k^{*T} \lambda_k^* + \mu_k = \sigma_k^2$; and $u^T \lambda_k^* = 1$. Equation (B1) is then obtained by combining all the resulting columns for the different k.

Equation (B1) can also be written as:

$$\begin{bmatrix} \Lambda - I \\ \cdots \\ \mu^T \end{bmatrix} = \Gamma^{-1} \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} \Sigma \qquad \text{(B3)}$$

Using the expressions given by Eqn. (B3) the vector of cross-validation errors, defined by Eqn. (18), can be now expressed as:

$$\varepsilon = \Sigma \left[I \;\vdots\; 0 \right] \Gamma^{-1} \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} z^* \qquad \text{(B4)}$$

and its covariance matrix V_ε as:

$$V_\varepsilon = E\left\{\varepsilon \varepsilon^T\right\} = \Sigma \left[I \;\vdots\; 0 \right] \Gamma^{-1} \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} E\left\{z^* z^{*T}\right\} \left[I \;\vdots\; 0 \right] \Gamma^{-1} \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} \Sigma$$

$$= \left[\Lambda^T - I \;\vdots\; \mu \right] \begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix} V_{z^*} \left[I \;\vdots\; 0 \right] \begin{bmatrix} \Lambda - I \\ \cdots \\ \mu^T \end{bmatrix} \qquad \text{(B5)}$$

Equation (B5) shows that the covariance matrix of the cross validation errors can be expressed as a function of the variance-covariance matrix of the observations V_{z^*} which can be computed by reversing the definition of the semi-variogram:

$$V_{z^*}(i, j) = \frac{1}{2}\left(\sigma_i^2 + \sigma_j^2\right) - \gamma_{ij} \qquad \text{(B6)}$$

where σ_i^2 and σ_j^2 are the unknown variances of z_i and z_j respectively.

Writing Eqn. (B6) in matrix form one obtains:

$$V_{z^*} = \frac{1}{2}\left(\Theta + \Theta^T\right) - \Gamma^* \qquad \text{(B7)}$$

where Θ is an [n, n] matrix defined as:

$$\Theta = \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \vdots & \sigma_n^2 \\ \sigma_1^2 & \sigma_2^2 & \vdots & \sigma_n^2 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_1^2 & \sigma_2^2 & \vdots & \sigma_n^2 \end{bmatrix} \qquad \text{(B8)}$$

The covariance matrix of the cross-validation errors becomes:

$$V_\varepsilon = \left[\Lambda^T - I \;\vdots\; \mu\right]\begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix}\left(\frac{\Theta + \Theta^T}{2} - \Gamma^*\right)\left[I \;\vdots\; 0\right]\begin{bmatrix} \Lambda - I \\ \cdots \\ \mu^T \end{bmatrix} \tag{B9}$$

Matrix Θ is unknown, but given that it is constant along the columns (and conversely Θ^T is constant along the rows), the following property holds:

$$\left[\Lambda^T - I \;\vdots\; \mu\right]\left(\begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix}\frac{\Theta + \Theta^T}{2}\left[I \;\vdots\; 0\right]\right)\begin{bmatrix} \Lambda - I \\ \cdots \\ \mu^T \end{bmatrix} = 0 \tag{B10}$$

since the sum of the weights λ_k is equal to 1.

By taking into account Eqn. (B10), the covariance matrix of the cross-validation errors becomes:

$$V_\varepsilon = \left[\Lambda^T - I \;\vdots\; \mu\right]\begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix}\left(-\Gamma^*\right)\left[I \;\vdots\; 0\right]\begin{bmatrix} \Lambda - I \\ \cdots \\ \mu^T \end{bmatrix} \tag{B11}$$

which can be also written as:

$$V_\varepsilon = -\left[\Lambda^T - I \;\vdots\; \mu\right]\Gamma\begin{bmatrix} \Lambda - I \\ \cdots \\ \mu^T \end{bmatrix} \tag{B12}$$

again because the weights λ_k add to 1.

Finally, Eqn. (B12) can be further simplified by substituting for Eqn. (B3), which gives:

$$V_\varepsilon = -\Sigma\left[I \;\vdots\; 0\right]\Gamma^{-1}\Gamma\;\Gamma^{-1}\begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix}\Sigma = -\Sigma\left[I \;\vdots\; 0\right]\Gamma^{-1}\begin{bmatrix} I \\ \cdots \\ 0 \end{bmatrix}\Sigma \tag{B13}$$

Expression (B13) shows that the covariance matrix of the cross-validation errors can be expressed as a function of their variances and of the inverse of the Kriging matrix.