

Kriging and Prediction of Nonlinear Functionals

Alexander Kukush¹ and István Fazekas²

¹Kiev National University

²University of Debrecen

Abstract: The prediction of a nonlinear functional of a random field is studied. The covariance-matching constrained kriging is considered. It is proved that the optimization problem induced by it always has a solution. The proof is constructive and it provides an algorithm to find the optimal solution. Using simulation, this algorithm is compared with the method given in Aldworth and Cressie (2003).

Keywords: Kriging, Nonlinear Functional, Optimization.

1 The Nonlinear Prediction Problem

A random field $z(\mathbf{t})$, $\mathbf{t} \in D \subset \mathbb{R}^d$, of the form (1) is observed at locations $\mathbf{t}_1, \dots, \mathbf{t}_n$. Kriging means optimal predicting the unobserved values of the random field z . The method of kriging is well-known (see e.g. Krige (1951), Cressie (1991)). Several modifications of the kriging procedure are studied, e.g., in Fazekas and Kukush (2003) a version of kriging is analyzed when the locations where the field is observed are not known precisely.

In this paper, we study problems which require the prediction of nonlinear functionals of the field z . To solve this kind of problems, Aldworth and Cressie (2003) introduced a predictor, called covariance-matching constrained kriging (see also Cressie (1993)). It is an optimal linear predictor that matches not only first moments but second moments (including covariances) as well. Aldworth and Cressie (2003) proved some properties of the covariance-matching constrained kriging and applied it to real life problems. Here we prove some further properties of that predictor.

Consider the following linear geostatistical model

$$z(\mathbf{t}) = \sum_{j=0}^p \beta_j f_j(\mathbf{t}) + \delta(\mathbf{t}), \quad \mathbf{t} \in D \subseteq \mathbb{R}^d, \quad (1)$$

where d is a fixed positive integer, $f_j(\mathbf{t})$ are known functions, β_j are unknown parameters, $j = 0, 1, \dots, p$, $\delta(\mathbf{t})$ are random error terms with $\mathbb{E}\delta(\mathbf{t}) = 0$, $\mathbb{E}\delta^2(\mathbf{t}) < \infty$, $\mathbf{t} \in D$.

Our sample is $z(\mathbf{t}_1), \dots, z(\mathbf{t}_n)$, in vector form $\mathbf{z} = (z(\mathbf{t}_1), \dots, z(\mathbf{t}_n))^T \in \mathbb{R}^n$. Then $\mathbb{E}\mathbf{z} = X\boldsymbol{\beta}$, for some known matrix X of type $n \times (p+1)$. Here $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^T$. We assume that the semivariogram of our random field z is known, therefore $\text{var}(\mathbf{z}) = \Sigma$ is a known $n \times n$ matrix.

We want to predict $g(\mathbf{s})$, where g is a known scalar-valued function that is nonlinear in its vector-valued argument \mathbf{s} . Here $\mathbf{s} = (\mathbf{s}(B_1), \dots, \mathbf{s}(B_m))^T$ is a function of the underlying random field z . To predict $g(\mathbf{s})$, consider a predictor of the form $g(\hat{\mathbf{s}})$, where $\hat{\mathbf{s}} = A^T \mathbf{z}$ is a linear predictor of \mathbf{s} , and A is an $n \times m$ matrix to be calculated. To obtain an approximately unbiased predictor, Aldworth and Cressie (2003) used Taylor's expansion up to second order for $g(A^T \mathbf{z})$ and $g(\mathbf{s})$. Their approach gave the following constraints.

Assume that A satisfies the constraints $\mathbb{E}A^T \mathbf{z} = \mathbb{E}\mathbf{s}$ and $\text{var}(A^T \mathbf{z}) = \text{var}(\mathbf{s})$. The first is the standard unbiasedness condition. If \mathbf{s} and \mathbf{z} are Gaussian, then the above conditions

imply that $A^\top z$ and s have the same distribution. Therefore in the Gaussian case $g(A^\top z)$ is an unbiased predictor of $g(s)$.

As the L_2 error, i.e., $\mathbb{E}(g(A^\top z) - g(s))^2$ can not be calculated directly, Aldworth and Cressie (2003) used a first order Taylor's expansion to approximate it. They obtained the expression $\mathbf{y}^\top \text{var}(A^\top z - s)\mathbf{y}$, where $\mathbf{y} = g'(\boldsymbol{\mu}_m) = \partial g(\boldsymbol{\mu}_m)/\partial \mathbf{x}$ and $\boldsymbol{\mu}_m = \mathbb{E}s$.

To obtain a predictor that can be calculated numerically, we assume that s satisfies the following. (These assumptions are realistic if s is a linear function of z , e.g. if it is defined by an integral of z .) $\mathbb{E}s = \boldsymbol{\mu}_m = X_m^\top \boldsymbol{\beta}$, where X_m is a known $(p+1) \times m$ type matrix. Moreover, $\text{var}(s) = \Sigma_m$, $\text{cov}(z, s) = C$, where Σ_m is a known $m \times m$ type matrix, while C is a known $n \times m$ type matrix. So we obtain the following objective function and constraints.

Find the matrix A such that the objective function

$$\mathbf{y}^\top A^\top C \mathbf{y} \quad (2)$$

is maximized under constraints

$$A^\top X = X_m^\top, \quad (3)$$

$$A^\top \Sigma A = \Sigma_m, \quad (4)$$

where $\mathbf{y} = g'(\boldsymbol{\mu}_m)$.

Aldworth and Cressie (2003) gave a matrix A_0 that maximizes the Lagrangian function corresponding to the maximum problem (2)–(4) at m linearly independent directions of \mathbf{y} . However, the original problem is to find the maximum for $\mathbf{y} = g'(\boldsymbol{\mu}_m)$. Here we shall prove (see Theorem 1) that, under realistic conditions, the maximum problem (2)–(4) has a solution whatever \mathbf{y} may be. Actually, we construct the optimal solution. This construction serves the base of a numerical method what we call algorithm M. Using computer simulation, we shall compare algorithm M with the naive estimator and the method suggested by Aldworth and Cressie (2003).

2 The Solution of the Optimization Problem

Theorem 1. *Let Σ be a real symmetric positive definite matrix of type $n \times n$, Σ_m be a real symmetric positive definite matrix of type $m \times m$, C be a real matrix of type $n \times m$, X_m be a real matrix of type $(p+1) \times m$, X be a real matrix of type $n \times (p+1)$ with rank $(p+1)$. Assume that $n \geq m + p + 1$. Let \mathbf{y} be a vector of dimension m . Let*

$$P = \Sigma_m - X_m^\top (X^\top \Sigma^{-1} X)^{-1} X_m. \quad (5)$$

(a) *Then there exists a matrix A satisfying conditions (3) and (4) if and only if P is positive semidefinite.*

(b) *If there exists a matrix A satisfying constraints (3)–(4), then there exists a matrix A satisfying constraints (3)–(4) and attaining maximum of (2).*

Proof. Introduce the following notation. $E = \Sigma^{1/2} A \Sigma_m^{-1/2}$ is a matrix of type $n \times m$, $B = X_m \Sigma_m^{-1/2}$ is a matrix of type $(p+1) \times m$, $Y = \Sigma^{-1/2} X$ is a matrix of type $n \times (p+1)$. Then constraints (3)–(4) are equivalent to the following. Find the matrix E satisfying

$$E^\top Y = B^\top, \quad (6)$$

$$E^\top E = I, \quad (7)$$

where I is the $m \times m$ type unit matrix, B and Y are known matrices.

(a1) First consider the case $m = 1$. Then constraints (6)–(7) are equivalent to

$$Y^\top \mathbf{e} = \mathbf{b}, \quad \mathbf{e}^\top \mathbf{e} = 1, \quad (8)$$

where Y^\top is a given matrix of type $(p + 1) \times n$ with rank $p + 1$, \mathbf{b} is a given vector of dimension $p + 1$, while we are looking for the unknown n -dimensional vector \mathbf{e} .

By the assumptions of the theorem, $n > p + 1$. Then the general solution of $Y^\top \mathbf{e} = \mathbf{b}$ is a non-trivial linear manifold. The particular solution \mathbf{e}_0 to $Y^\top \mathbf{e} = \mathbf{b}$ having the minimal length, is the one which is orthogonal to each solution of the homogeneous linear equation $Y^\top \mathbf{e} = \mathbf{0}$. Therefore, if \mathbf{e} is orthogonal to the rows of Y^\top , then \mathbf{e}_0 is orthogonal to \mathbf{e} . So \mathbf{e}_0 is in the subspace spanned by the rows of Y^\top . That is $\mathbf{e}_0 = Y\mathbf{v}$ for some \mathbf{v} . Therefore $\mathbf{b} = Y^\top \mathbf{e}_0 = Y^\top Y\mathbf{v}$. As Y^\top is of rank $p + 1$, so $Y^\top Y$ is invertible, therefore $\mathbf{v} = (Y^\top Y)^{-1} \mathbf{b}$. So we obtain $\mathbf{e}_0 = Y(Y^\top Y)^{-1} \mathbf{b}$.

Any solution of $Y^\top \mathbf{e} = \mathbf{b}$ can be written in the form $\mathbf{e} = \mathbf{e}_0 + \mathbf{e}_1$, where \mathbf{e}_1 is a vector with $Y^\top \mathbf{e}_1 = \mathbf{0}$. As \mathbf{e}_1 is orthogonal to \mathbf{e}_0 , our condition is $1 = \mathbf{e}^\top \mathbf{e} = \|\mathbf{e}_0\|^2 + \|\mathbf{e}_1\|^2$. This can be satisfied if and only if $\|\mathbf{e}_0\|^2 \leq 1$.

So we obtained the following statement. If $m = 1$, $n > p + 1$ and Y has rank $p + 1$, then (8) has a solution if and only if

$$\mathbf{b}^\top (Y^\top Y)^{-1} \mathbf{b} \leq 1. \quad (9)$$

This is equivalent to the condition that the matrix P defined in (5) is positive semidefinite. So we obtained part (a) in the case $m = 1$.

(a2) Now let m be an arbitrary positive integer. If there exists an E satisfying (6)–(7), then for any $\mathbf{a} \in \mathbb{R}^{p+1}$, $\|\mathbf{a}\| \neq 0$, we have

$$\mathbf{a}^\top E^\top Y = \mathbf{a}^\top B^\top, \quad \mathbf{a}^\top E^\top E \mathbf{a} = \|\mathbf{a}\|^2.$$

So $\mathbf{e} = E\mathbf{a}/\|\mathbf{a}\|$ is a solution to (8) with $\mathbf{b} = B\mathbf{a}/\|\mathbf{a}\|$. By the previous part of the proof, this implies that (9) is satisfied, i.e.

$$\frac{\mathbf{a}^\top B^\top}{\|\mathbf{a}\|} (Y^\top Y)^{-1} \frac{B\mathbf{a}}{\|\mathbf{a}\|} \leq 1,$$

that is $I - B^\top (Y^\top Y)^{-1} B$ is positive semidefinite. Therefore the matrix P defined in (5) is positive semidefinite.

Now let P be positive semidefinite. Find the solution of (6)–(7) in the form

$$E = (I - Y(Y^\top Y)^{-1} Y^\top) V + Y(Y^\top Y)^{-1} B = MV + E_0, \quad (10)$$

where V is a matrix to be determined. We remark that in Aldworth and Cressie (2003) the same scheme was applied. They obtained that $V = \Sigma^{-1/2} C K \Sigma_m^{-1/2}$ is the optimal solution at m linearly independent directions of \mathbf{y} .

Now

$$E^\top = V^\top (I - Y(Y^\top Y)^{-1} Y^\top) + B^\top (Y^\top Y)^{-1} Y^\top = V^\top M + E_0^\top.$$

Y^\top . Then $\mathbf{w}^\top \mathbf{e} = \mathbf{w}_0^\top \mathbf{e}_0 + \mathbf{w}_1^\top \mathbf{e}_1$ is maximal if and only if \mathbf{e}_1 is parallel to \mathbf{w}_1 (use the Cauchy-Schwarz-Buniakovskii inequality). So we obtained part (b) of Theorem 1 in the case $m = 1$.

(b2) Now turn to part (b) in the case of arbitrary m . The proof is the basis of algorithm M.

We slightly reshape constraints (6)–(7). As the $p+1$ columns of Y are independent, we can apply Gram-Schmidt orthogonalization to it. This can be described by an invertible matrix G . Multiply from the right by this G the equation $E^\top Y = B^\top$. Then we obtain $E^\top W = V$, where $YG = W$ and $B^\top G = V$. Denote by $\mathbf{w}_1, \dots, \mathbf{w}_{p+1} \in \mathbb{R}^n$ the columns of W , by $\mathbf{e}_1, \dots, \mathbf{e}_m \in \mathbb{R}^n$ the columns of E , and by v_{ij} the elements of V . Then $E^\top W = V$ can be written into the form

$$\mathbf{e}_i^\top \mathbf{w}_j = v_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, p+1. \quad (12)$$

Here $\mathbf{w}_1, \dots, \mathbf{w}_{p+1} \in \mathbb{R}^n$ are given orthonormal vectors, v_{ij} , $i = 1, \dots, m$, $j = 1, \dots, p+1$ are given numbers. We have to find $\mathbf{e}_1, \dots, \mathbf{e}_m \in \mathbb{R}^n$.

Let \mathcal{L} denote the subspace in \mathbb{R}^n spanned by $\mathbf{w}_1, \dots, \mathbf{w}_{p+1}$ and \mathcal{L}^\perp its orthogonal complement. We search $\mathbf{e}_1, \dots, \mathbf{e}_m$ in the form

$$\mathbf{e}_i = \sum_{j=1}^{p+1} v_{ij} \mathbf{w}_j + \alpha_i \mathbf{l}_i, \quad i = 1, \dots, m, \quad (13)$$

where $\mathbf{l}_1, \dots, \mathbf{l}_m$ are unit (in some special cases zero) vectors from \mathcal{L}^\perp . (As $\mathbf{w}_1, \dots, \mathbf{w}_{p+1}$ are orthonormal vectors, the above representation is valid for vectors $\mathbf{e}_1, \dots, \mathbf{e}_m$ satisfying (12).) Because we assume that the constraints (6)–(7) can be satisfied, therefore there exist orthonormal vectors $\mathbf{e}_1, \dots, \mathbf{e}_m \in \mathbb{R}^n$ with the above representation. This implies that $\sum_{j=1}^{p+1} v_{ij}^2 \leq 1$ for $i = 1, \dots, m$, and

$$\alpha_i = \sqrt{1 - \sum_{j=1}^{p+1} v_{ij}^2}, \quad i = 1, \dots, m.$$

By orthogonality of \mathbf{e}_i and \mathbf{e}_k ,

$$0 = \mathbf{e}_i^\top \mathbf{e}_k = \sum_{j=1}^{p+1} v_{ij} v_{kj} + \alpha_i \alpha_k \mathbf{l}_i^\top \mathbf{l}_k, \quad i, k = 1, \dots, m, \quad i \neq k. \quad (14)$$

So we have to find the vectors $\mathbf{l}_1, \dots, \mathbf{l}_m$ from \mathcal{L}^\perp such that

$$\mathbf{l}_i^\top \mathbf{l}_k = d_{ik}, \quad i, k = 1, \dots, m, \quad (15)$$

where $d_{ik} = 0$ if $\alpha_i \alpha_k = 0$; $d_{ii} = 1$ if $\alpha_i \neq 0$; $d_{ik} = -\left(\sum_{j=1}^{p+1} v_{ij} v_{kj}\right) / (\alpha_i \alpha_k)$ if $i \neq k$ and $\alpha_i \alpha_k \neq 0$.

The objective function to be maximized is

$$\mathbf{y}^\top C^\top A \mathbf{y} = \mathbf{y}^\top C^\top \Sigma^{-1/2} E \Sigma_m^{1/2} \mathbf{y} = \mathbf{f}^\top \sum_{i=1}^m x_i \mathbf{e}_i, \quad (16)$$

where $\mathbf{f} = \Sigma^{-1/2} C \mathbf{y}$ and $(x_1, \dots, x_m)^\top = \mathbf{x} = \Sigma_m^{1/2} \mathbf{y}$. Using (13),

$$\sum_{i=1}^m x_i \mathbf{e}_i = \sum_{i=1}^m \sum_{j=1}^{p+1} x_i v_{ij} \mathbf{w}_j + \sum_{i=1}^m x_i \alpha_i \mathbf{l}_i. \quad (17)$$

As the first summand here is fixed, we have to find the maximum of

$$\sum_{i=1}^m \mathbf{f}^\top x_i \alpha_i \mathbf{l}_i. \quad (18)$$

Finally, introducing the notation \mathbf{f}_\perp for the orthogonal projection of \mathbf{f} to \mathcal{L}^\perp , and choosing $\gamma_i = x_i \alpha_i$, $i = 1, \dots, m$, we have to find the vectors $\mathbf{l}_1, \dots, \mathbf{l}_m$ from \mathcal{L}^\perp such that (15) is satisfied and

$$\mathbf{f}_\perp^\top \left(\sum_{i=1}^m \gamma_i \mathbf{l}_i \right) \quad (19)$$

is maximized. By the Cauchy-Schwarz-Buniakovskii inequality,

$$\mathbf{f}_\perp^\top \left(\sum_{i=1}^m \gamma_i \mathbf{l}_i \right) \leq \|\mathbf{f}_\perp\| \sqrt{\sum_{i=1}^m \sum_{j=1}^m \gamma_i \gamma_j d_{ij}}. \quad (20)$$

If $\mathbf{f}_\perp = \mathbf{0}$, or $\sum_{i=1}^m \sum_{j=1}^m \gamma_i \gamma_j d_{ij} = 0$, then $\mathbf{l}_1, \dots, \mathbf{l}_m$ can be arbitrarily chosen from \mathcal{L}^\perp such that (15) is satisfied.

Now assume that $\mathbf{f}_\perp \neq \mathbf{0}$ and $\sum_{i=1}^m \sum_{j=1}^m \gamma_i \gamma_j d_{ij} > 0$. In (20) we have equality if and only if

$$\sum_{i=1}^m \gamma_i \mathbf{l}_i = \lambda \mathbf{f}_\perp \quad (21)$$

with $\lambda > 0$. So we shall find $\mathbf{l}_1, \dots, \mathbf{l}_m$ from \mathcal{L}^\perp such that (15) and (21) are satisfied. We see that

$$\lambda = \sqrt{\sum_{i=1}^m \sum_{j=1}^m \gamma_i \gamma_j d_{ij}} / \|\mathbf{f}_\perp\|.$$

Fix an orthonormal basis $\mathbf{h}_1, \dots, \mathbf{h}_{n-(p+1)}$ in \mathcal{L}^\perp so that \mathbf{f}_\perp is in the subspace spanned by the first m basis vectors. From now on we identify the vectors in \mathcal{L}^\perp with their coordinate sequences according to the basis $\mathbf{h}_1, \dots, \mathbf{h}_{n-(p+1)}$. Find $\mathbf{l}_1, \dots, \mathbf{l}_m$ in the form

$$\mathbf{l}_i = \begin{pmatrix} D_{11} & O \\ O & D_{22} \end{pmatrix} \mathbf{h}_i, \quad i = 1, \dots, m,$$

where D_{11} is of type $m \times m$ and D_{22} is of type $(n - m - p - 1) \times (n - m - p - 1)$. Now, by (15),

$$d_{ij} = \mathbf{l}_i^\top \mathbf{l}_j = \mathbf{h}_i^\top \begin{pmatrix} D_{11}^\top D_{11} & O \\ O & D_{22}^\top D_{22} \end{pmatrix} \mathbf{h}_j = (D_{11}^\top D_{11})_{ij}, \quad i, j = 1, \dots, m.$$

Therefore $D_{11}^\top D_{11} = D_0$, where $D_0 = (d_{ij})_{i,j=1}^m$. So $D_{11} = U D_0^{1/2}$, where U is an $m \times m$ type orthogonal matrix. Then

$$\mathbf{l}_i = \begin{pmatrix} (U D_0^{1/2})_i \\ \mathbf{0} \end{pmatrix}, \quad i = 1, \dots, m,$$

where $(U D_0^{1/2})_i$ is the i th column of $U D_0^{1/2}$. We have to find U so that (21) is satisfied. This gives

$$U \left(D_0^{1/2} \sum_{i=1}^m \gamma_i \tilde{\mathbf{h}}_i \right) = \lambda \tilde{\mathbf{f}}_\perp, \quad (22)$$

where $\tilde{\mathbf{h}}_i$ is the i th unit vector in \mathbb{R}^m , while $\tilde{\mathbf{f}}_\perp$ consists of the first m coordinates of \mathbf{f}_\perp . As λ is chosen such that the vectors $\left(D_0^{1/2} \sum_{i=1}^m \gamma_i \tilde{\mathbf{h}}_i \right)$ and $\lambda \tilde{\mathbf{f}}_\perp$ have the same length, one can find an orthogonal matrix U satisfying (22).

4 Numerical Results

In this section we compare three methods of prediction of the nonlinear functional g .

The first one is algorithm M described in the constructive proof of Theorem 1 in Section 3. To obtain a feasible method, instead of \mathbf{y} we use $\widehat{\mathbf{y}}$ when we solve the optimization problem (2)–(4). Here

$$\widehat{\mathbf{y}} = g'(\widehat{\boldsymbol{\mu}}_m) = g'(X_m^\top \widehat{\boldsymbol{\beta}}), \quad (23)$$

where $\widehat{\boldsymbol{\beta}} = (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} \mathbf{z}$ is the generalized least squares estimator of $\boldsymbol{\beta}$. The resulting estimator is $\widehat{g}_{\text{opt}} = g(A_{\text{opt}}^\top \mathbf{z})$, where A_{opt}^\top is the solution given by algorithm M if \mathbf{y} in (2) is substituted by $\widehat{\mathbf{y}}$ given in (23).

The second estimator is the one given by Aldworth and Cressie (2003). This estimator is denoted by \widehat{g}_{ac} . Here

$$\widehat{g}_{\text{ac}} = g(A_0^\top \mathbf{z}),$$

where

$$A_0 = \Sigma^{-1} (I - X(X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1}) C K + \Sigma^{-1} X (X^\top \Sigma^{-1} X)^{-1} X_m, \quad (24)$$

where $K = Q^{-1/2} P^{1/2}$,

$$P = \text{var}(\mathbf{s}) - \text{var}(X_m^\top \widehat{\boldsymbol{\beta}}) = \Sigma_m - X_m^\top (X^\top \Sigma^{-1} X)^{-1} X_m \quad (25)$$

is the matrix in (5), moreover

$$\begin{aligned} Q &= \text{var}(\widehat{\mathbf{s}}_{\text{uk}}) - \text{var}(X_m^\top \widehat{\boldsymbol{\beta}}) = \\ &= C^\top (\Sigma^{-1} - \Sigma^{-1} X (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1}) C. \end{aligned} \quad (26)$$

Here

$$\widehat{\mathbf{s}}_{\text{uk}} = X_m' \widehat{\boldsymbol{\beta}} + C' \Sigma^{-1} (\mathbf{z} - X \widehat{\boldsymbol{\beta}})$$

is the universal kriging predictor.

Finally, the third estimator is the naive estimator.

Example. Consider the domain $D = [-rh, +rh] \times [-rh, +rh] \subset \mathbb{R}^2$ and a random field $z(x, y)$ on D with linear mean value function:

$$z(x, y) = \beta_0 + \beta_1 x + \beta_2 y + \delta_0(x, y) + \delta_1(x, y), \quad (27)$$

where $(x, y) \in D$. The parameters $\beta_0, \beta_1, \beta_2$ are unknown. (For simulation we chose $\beta_0 = 20, \beta_1 = 1, \beta_2 = -1$.) The field $\delta_0(x, y)$ is a zero mean Gaussian random field with Gaussian semivariogram

$$\gamma_0(\mathbf{t}) = \omega \left(1 - \exp \left(- \|\mathbf{t}\|^2 / a^2 \right) \right) \quad (28)$$

for $\mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$, where $\|\mathbf{t}\|^2 = t_1^2 + t_2^2$. (We assume that the covariance structure is completely known. Actually δ_0 is an Ornstein-Uhlenbeck type random field. For the Ornstein-Uhlenbeck random field see Terdik and Woyczynski (2005) and the references therein.) The random variables $\delta_1(x, y)$, where $(x, y) \in D$, are independent and uniformly

distributed on $[-0.5\tau, +0.5\tau]$. The random fields $\delta_0(x, y)$, $(x, y) \in D$, and $\delta_1(x, y)$, $(x, y) \in D$, are independent. We used several values of the parameters ω , a^2 and τ . By simulation we generated the field z on the h -lattice points of D . We used $r = 6$ and $h = 2$.

We predicted the function

$$g(\mathbf{s}) = \frac{1}{m} \sum_{k=-2}^1 \sum_{l=-2}^1 [z((2k+1)h, (2l+1)h)]^3, \quad (29)$$

where $m = 16$ is the number of locations, where the values of the random field are included in the above expression.

For prediction we used the observations of the random field $z(x, y)$ at locations $(x, y) = ((-r+2k)h, (-r+2l)h)$, $k, l = 0, 1, \dots, r$. I.e. our sample size was $n = 49$.

With the help of these observations we predicted $g(\mathbf{s})$.

The so called naive approach is the following. Using universal kriging, we predict the field $z(x, y)$ at the h -lattice points $(x, y) = (ih, jh)$ included in expression (29). So the naive predictor is

$$\hat{g}_{\text{naive}}(\mathbf{s}) = \frac{1}{m} \sum_{k=-2}^1 \sum_{l=-2}^1 [\hat{z}((2k+1)h, (2l+1)h)]^3, \quad (30)$$

where \hat{z} is the universal kriging predictor.

Using the same observations, we calculated \hat{g}_{ac} and \hat{g}_{opt} , too.

We made 1000 replications. (The simulations were performed with MATLAB.)

For each predictor, we calculated the ratio of the mean squared error and the average. I.e. in Table 1 the values of

$$q = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (\hat{g}_i - g_i)^2} / \frac{1}{1000} \sum_{i=1}^{1000} |g_i|$$

are shown. (We remark that in the program the value of ω is expressed in terms of π .)

The results of some other experiments were similar. The conclusions are the following. The naive approach (i.e. \hat{g}_{naive} based on universal kriging) performed poorly for non-linear predictands. The performances of the two different calculations of the covariance-matching constrained kriging, i.e. \hat{g}_{opt} and \hat{g}_{ac} are similar. The advantages of \hat{g}_{ac} are the following. It is easy to calculate \hat{g}_{ac} while \hat{g}_{opt} requires an implementation of algorithm M. Moreover, for Gaussian random fields with large variances \hat{g}_{ac} is more stable than \hat{g}_{opt} . A careful analysis of the distribution of the underlying random field z is necessary in each particular case when we want to compare \hat{g}_{opt} and \hat{g}_{ac} .

Acknowledgements

This research was partially supported by the Hungarian Foundation of Scientific Researches under Grant No. OTKA T047067/2004. The research was partially realized while A. G. Kukush was visiting the University of Debrecen.

Table 1: The relative errors of the estimators

Parameters			q values for the predictors		
ω	α^2	τ	\hat{g}_{naive}	\hat{g}_{opt}	\hat{g}_{ac}
1.571	2	3	28.17 %	0.310 %	0.356 %
3.142	4	3	25.88 %	0.844 %	0.583 %
6.283	2	3	27.38 %	0.681 %	1.306 %
12.566	4	3	27.01 %	3.008 %	2.809 %
42.412	6	3	21.32 %	5.710 %	3.304 %
1.178	6	3	26.59 %	2.737 %	2.436 %
1.571	2	4	39.55 %	0.499 %	0.411 %
6.283	2	4	40.88 %	0.850 %	1.029 %
25.133	2	4	43.21 %	2.049 %	2.636 %
50.265	4	4	41.63 %	1.393 %	1.223 %
42.412	6	0	4.78 %	5.842 %	3.346 %

References

- J. Aldworth and N. Cressie. Prediction of nonlinear spatial functionals. *J. Statistical Planning Inference*, 112:3–41, 2003.
- N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1991.
- N. A. C. Cressie. Aggregation in geostatistical problems. In A. Soares, editor, *Geostatistics Tróia 1992*, volume 1, pages 25–36. Kluwer, Dordrecht, 1993.
- I. Fazekas and A. G. Kukush. Kriging and measurement errors. In *Proc. Conf. Statistical Inference in Linear Models, Bedlewo, Poland (submitted)*. 2003.
- D. G. Krige. A statistical approach to some basic mine valuations problems on the witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52:119–139, 1951.
- Gy. Terdik and W. A. Woyczynski. Notes on fractional Ornstein-Uhlenbeck random sheets. *Publ. Math. Debrecen*, 66(1/2):153–181, 2005.

Authors' addresses:

Alexander Kukush
Faculty of Mathematics and Mechanics,
Kiev National University
Vladimirskaia st. 60
01034 Kiev
Ukraine
E-mail: alexander_kukush@univ.kiev.ua

István Fazekas
Faculty of Informatics
University of Debrecen
P.O. Box 12
4010 DEBRECEN
Hungary
E-mail: fazekasi@inf.unideb.hu