Proceedings of the 1. Conference on
Applied Mathematics and Computation
Dubrovnik, Croatia, September 13–18, 1999
pp. 233–243

On Simulations of Random Fields on Bounded Domains*

Luka Grubišić†, Nedžad Limić‡, and Vladivoj Valković§

Abstract. A random field on a bounded domain is usually estimated from data by using the linear least square estimation. To apply this method it is necessary to know the covariance function, while distribution of field at a point is not used. Better results of estimation are expected by applying methods for which the distribution of field must be utilized. We consider here a method for which the random field is represented by using the Loeve–Karhunen decomposition. Numerical approximations are obtained by grids discretizing the domain and by solving the corresponding finite-dimensional eigenvalue problem for finite dimensional matrices.

AMS subject classification: 60G60, 65C05

Key words: random field, Monte Carlo methods

1. Introduction

In this article the object of study is a random field on a bounded domain. The domain (an open set in $\mathbb{R}^d$) is denoted by D. It is assumed that the following property is valid: $\mathrm{int}(\bar{D}) = D$. Points of D are denoted by $\mathbf{x}$. For the considered random field $\mathbf{x} \mapsto \xi(\mathbf{x})$, it is assumed that the first two statistical moments are defined. The first order moment, i.e., the mean value field $\mathbf{E}[\xi(\cdot)]$, is assumed to be constant on D, and the covariance function $(\mathbf{x}, \mathbf{y}) \mapsto C(\mathbf{x}, \mathbf{y})$ is assumed to be continuous on $\bar{D} \times \bar{D}$. The field is measured at M points, so that the data consist of M pairs $\{(\mathbf{x}_m, \xi_m) \mid m = 1, 2, \ldots, M\}$. The problem is to estimate the considered random field at any point of D from the available data set.

The well-known method to estimate a random field from data is the linear least square estimation (LLSE) [5]. The only information needed about the field is the covariance function or variogram. An error of estimated values can be derived by using Chebyshev inequality. If additional information about the field is available,

*Supported partially by Ministry of Science and Technology, Republic of Croatia, Grants 037011 and 037012.

†Department of Mathematics, University of Zagreb, Bijenička cesta 30, 10000 Zagreb, Croatia, e–mail: luka@math.hr

‡Department of Mathematics, University of Zagreb, Bijenička cesta 30, 10000 Zagreb, Croatia, e–mail: nlimic@math.hr

§Ruđer Bošković Institute, University of Zagreb, Bijenička cesta 56, 10000 Zagreb, Croatia

one has to use other methods in order to incorporate the whole knowledge about the considered field [7]. Now we can formulate the problem by using mathematical objects.

Problem. *Let* Γ *be the following event:*

$$\Gamma = \{\xi(\mathbf{x}_m) \in I_m \mid m = 1, 2, \ldots, M\}, \tag{1}$$

where I_m *are symmetric intervals around measured values* ξ_m *with the lengths* $2\rho_m$*. The conditional statistical moments* $\mathbf{E}[\xi^r(\mathbf{x}) \,|\, \Gamma]$*,* $r \in \mathbb{N}$*, have to be estimated at each* $\mathbf{x} \in D$*, and an error of estimates must be calculated.*

In Section 2 mathematical aspects of the problem are discussed and a method for estimation is proposed. Also some properties of statistics are analyzed. In particular, criteria for statistics to be consistent estimators are clearly specified. Mathematical modelling of arsenic in soil of the Island Krk is discussed in the first part of Section 3. Numerical aspects of the problem are presented in the second part of Section 3. Discussion and concluding remarks are contained in the last section.

2. Field representation and conditional moments

We start our discussion by assuming that the random variables $\xi(\mathbf{x})$ are normal and that they have zero expectation. The eigenfunction decomposition of the covariance function

$$C(\mathbf{x}, \mathbf{y}) = \sum_{\{k|\lambda_k>0\}} \lambda_k \psi_k(\mathbf{x}) \psi_k(\mathbf{y})$$

enables the Loeve–Karhunen representation (decomposition) of the considered random field [2]:

$$\xi(\mathbf{x}) = \sum_k \sqrt{\lambda_k} \eta_k \psi_k(\mathbf{x}), \tag{2}$$

where η_k are independent random variables of the type $N(0,1)$. The functions $\mathbf{x} \mapsto \psi_k(\mathbf{x})$ are continuous on $\bar{D}$. If the eigenfunctions $\psi_k(\mathbf{x})$ are known, the expression (2) is a basis for Monte Carlo simulations. Since such situations are rarely met in applications, we have to look for other methods of simulation. A satisfactory approach to solve the problem is to use grid approximation of the considered random field. Let the unit vectors $\mathbf{e}_i$ define the coordinate system. For each $n \in \mathbb{N}$, points $\mathbf{x} = h \sum_{l=1}^{d} k_l \mathbf{e}_l$, $h = 2^{-n}$, $k_l \in \mathbb{Z}$, define a numerical grid G_n on $\mathbb{R}^d$. Elements of G_n are called grid knots. Grid knots contained in D are denoted by $G_n(D)$. Their index set is $\mathcal{J}_n$, and their number is $J_n = \operatorname{card}(\mathcal{J}_n)$. When the random field is restricted to $G_n(D)$, it becomes a J_n-dimensional normal vector:

$$\big(\xi(\mathbf{x}_k), k \in \mathcal{J}_n\big), \tag{3}$$

which is uniquely defined by $\mathbf{E}[\xi(\mathbf{x}_k)] = 0$ (or constant) and the covariances $c_{kl} = C(\mathbf{x}_k, \mathbf{x}_l)$, $k, l \in \mathcal{J}_n$. Covariance matrix is denoted by C_n and it is of the order J_n.

The analogue of (2) is the expression of the following form:

$$\xi(\mathbf{x}_k) = \sum_{l \in \mathcal{J}_n} s_{kl}\eta_l, \quad k \in \mathcal{J}_n, \tag{4}$$

where s_{kl} are matrix elements of a square root of the covariance matrix, $S_n = \sqrt{C_n}$. The representation (4) can be straightforwardly used for Monte Carlo simulation of the considered random vector $(\xi(\mathbf{x}_k), k \in \mathcal{J}_n)$.

Let f_m be the m-dimensional density of normal variables $\xi(\mathbf{x}_j)$, $j = 1, 2, \ldots, m$, and let $\phi_m(\mathbf{p})$ be the corresponding characteristic function. The product of M intervals in (1) is denoted by K, i.e., $K = I_1 \times I_2 \times \cdots \times I_M$. Let $\mathbb{1}_K$ be the indicator of K. Its Fourier transform is denoted by $\hat{\mathbb{1}}_K$. Then

$$\mathbf{E}[\xi^r(\mathbf{x}) \,|\, \Gamma]\, \mathbf{P}(\Gamma) = \left(\frac{1}{i}\,\frac{d}{dp_0}\right)^r \int_{\mathbb{R}^m} \phi_{n+1}(p_0, \mathbf{p})\, \hat{\mathbb{1}}_K(-\mathbf{p})\, d\mathbf{p} \,\Big|_{p_0=0}.$$

Explicit expression of conditional moments is not needed. If the intervals I_m from (1) shrink to points r_m, the conditional moments

$$\mathbf{E}[\xi^r(\mathbf{x}) \,|\, r_1, r_2, \ldots, r_M] = \mathbf{E}[\xi^r(\mathbf{x}) \,|\, \xi(\mathbf{x}_1) = r_1, \xi(\mathbf{x}_2) = r_2, \ldots, \xi(\mathbf{x}_M) = r_M]$$

have simple expressions.

Proposition 1. *Let $\xi(\cdot)$ be a normal random field on D. Then*

(*a*) *conditional expectation $\mathbf{E}[\xi(\mathbf{x}) \,|\, r_1, \ldots, r_M]$ is a linear combination of $C(\mathbf{x}, \mathbf{x}_j)$, $j = 1, 2, \ldots, M$,*

(*b*) *higher order conditional moments $\mathbf{E}[\xi^r(\mathbf{x}) \,|\, r_1, \ldots, r_M]$ are polynomials of the order r in $C(\mathbf{x}, \mathbf{x})$ and $C(\mathbf{x}, \mathbf{x}_j)$, $j = 1, 2, \ldots, M$.*

The well-known linear least square estimator

$$\xi_{\mathrm{LLSE}}(\mathbf{x}) = \sum_{k=1}^{M} w_k(\mathbf{x})\xi(\mathbf{x}_k), \quad \sum_{l=1}^{M} C(\mathbf{x}_k, \mathbf{x}_l) w_l(\mathbf{x}) = C(\mathbf{x}, \mathbf{x}_k), \quad k = 1, \ldots, M, \tag{5}$$

has the same conditional expectation as the field itself. Higher order conditional moments differ.

If the field $\xi(\cdot)$ is not normal, but two-dimensional densities and statistical moments of the second order exist, many of mentioned results are valid. For positive eigenvalues of the covariance function, the random variables

$$\eta_j = \frac{1}{\sqrt{\lambda_j}} \int_D \psi_j(\mathbf{x})\xi(\mathbf{x})\, d\mathbf{x}$$

exist, they are mutually orthogonal, and (2) is the desired Loeve–Karhunen representation of the considered field. The conditional moments of the field and of the

expression (5) differ. In this case, an estimate of field values conditioned to Γ cannot be obtained from linear least square estimation. However, simulation of conditional moments based on the representation (2) gives results consistent with assumptions of the field.

An estimation of statistical moments from data is a crucial step in applications. It is worthy to discuss conditions on the field ensuring that statistics can be used as consistent estimators of relevant statistical moments. Since the first and second order statistics are of common interest in applications, we consider here the statistics for expectation and covariances. In accordance with statistical models of geostatistics, it is assumed that the expectation $\mathbf{E}[\xi(\mathbf{x})]$ is $\mathbf{x}$-independent and the covariance function is homogeneous and isotropic, i.e., there exists a positive definite function $x \mapsto K(x)$ on $\mathbb{R}$ such that $C(\mathbf{x}, \mathbf{y}) = K(|\mathbf{x} - \mathbf{y}|)$. The statistics for expectation are defined by expressions:

$$s(\xi, J_n) = \frac{1}{J_n} \sum_{j \in \mathcal{J}_n} \xi(\mathbf{x}_j). \tag{6}$$

The sequence of statistical moments $\mathbf{E}[s^2(\xi, J_n)]$ converges,

$$\lim_n \mathbf{E}[s^2(\xi, J_n)] = \frac{1}{\mathrm{vol}^2(D)} \int_{D \times D} C(\mathbf{x}, \mathbf{y})\, d\mathbf{x}\, d\mathbf{y}$$

in the Riemannian sense, due to the continuity of the covariance function.

The statistics for $K(|\mathbf{h}|)$ can also be constructed. Let $\mathbf{h}$ be such that there exists $n \in \mathbb{N}$ with the following property: each $\mathbf{x} \in G_n$ implies $\mathbf{x} + \mathbf{h} \in G_n$. For each $\mathbf{h}$ of this type, the corresponding unbiased statistics are:

$$s(K(|\mathbf{h}|), I_n) = \frac{1}{I_n} \sum_{k \in \mathcal{I}_n} \xi(\mathbf{x}_k)\xi(\mathbf{x}_k + \mathbf{h}), \tag{7}$$

where $\mathcal{I}_n = \big\{k \in \mathcal{J}_n \mid \mathbf{x}_k \in G_n(D), \mathbf{x}_k + \mathbf{h} \in G_n(D)\big\}$, and $I_n = \mathrm{card}(\mathcal{I}_n)$.

It is easy to calculate variances and answer the basic question about a convergence of statistics to the corresponding moments. To write expressions as briefly as possible, we extend functions C and ψ_k by zero outside of $\bar{D}$. The $L_2(D)$–scalar product is denoted by $(\cdot|\cdot)$.

Proposition 2. *Let $\xi(\cdot)$ be a random field on D with a constant expectation and a continuous covariance function. Then:*

- (*i*) *the statistics* (6) *and* (7) *converge in the mean to certain random variables, $s(\xi)$, and $s(K(\mathbf{h}))$, respectively,*
- (*ii*) *the statistics* (6) *converge to $\mathbf{E}[\xi(\mathbf{x})]$ in the mean, if and only if $(\psi_k|1) = 0$ for all k for which $\lambda_k > 0$,*
- (*iii*) *the random variable $s(K(\mathbf{h}))$ has a positive variance for small values of $|\mathbf{h}|$.*

An undesirable conclusion follows from this analysis. The statistics (6) and (7) are consistent estimators only in some exceptional cases. In order to construct statistics which converge to the corresponding statistical moments, one is forced to introduce additional suppositions into the model of random field. It can be assumed that the random field is actually defined on the whole $\mathbb{R}^d$ and considered on D for the present purpose. Sometimes data can be interpreted as measurements from a large sample of the J_n–dimensional variable (3), etc.

3. Arsenic concentration in soil. A case study

Model of trace element concentration in soil

Mapping of a trace element filed in rock and soil is one of typical demands in environmental studies. The developed mathematical tool is applied to estimate the concentration field of arsenic in soil of the Island Krk. The chosen trace element has an increased concentration in rock and soil of Kvarner region if compared to concentration levels in surrounding regions. In order to define the arsenic random field, we follow some usual assumptions of geostatistics as well as assumptions in accordance with developed mathematical tool.

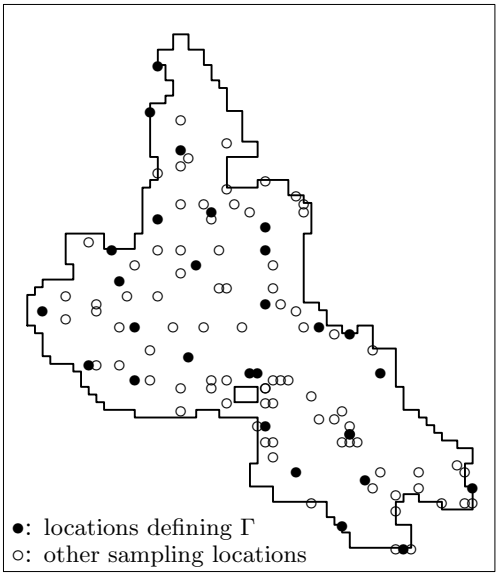

Figure 1. Sampling locations.

parameter	value
$m = \mathbf{E}[\xi_g]$	$26.42\,\mathrm{ppm}$
$A_l + A_g$	$95.01\,\mathrm{ppm}^2$
$A_g/(A_l + A_g)$	0.784
ρ_g	$0.037\,\mathrm{km}^{-2}$
$\mathbf{Var}[s(\xi)]$	$3.1\,\mathrm{ppm}$
$\mathbf{Var}[s(K_g(0))]$	$240\,\mathrm{ppm}^2$

Table 1.

Soil of the Island Krk is modeled by a domain $D \subset \mathbb{R}^2$. The concentration field of a trace element in soil can be represented by

$$\xi(\mathbf{x}) = \xi_l(\mathbf{x}) + \xi_g(\mathbf{x}), \tag{8}$$

where the components ξ_l and ξ_g are independent and represent local and global variation of concentration. They differ in the range of covariance functions. The former one has a short range covariance function, while the latter one has a long range covariance function. The former one can be considered as a noise summed with a smooth filed ξ_g. The corresponding covariance functions are assumed to have the following general structure:

$$K_l(x) = A_l \exp(-\rho_l |x|), \tag{9}$$

$$K_g(x) = A_g \exp(-\rho_g x^2). \tag{10}$$

Hence, they are defined by four parameters.

Assumption 1. *The arsenic concentration field on D has the following properties:*

(*a*) *the mean concentration fields of ξ_l and ξ_g are constant on D, $\mathbf{E}[\xi_l] = 0$,*

(*b*) *the covariance functions are homogeneous, isotropic, and have the structure* (9), (10)*,*

(*c*) *the fields ξ_l, ξ_g are Gaussian.*

All three assumptions must be verified by statistical tests.

A zone of an increased concentration of arsenic is naturally limited to Kvarner region. Therefore, a supposition about an unbounded domain of the arsenic concentration field with a constant expectation must be ruled out. Trace elements in soil are subject to permanent chemical reactions. Hence, in soil it is not possible to identify records from various time periods in order to generate a sample of trace element concentration. We are forced to conclude that, regardless of the number of data, statistics for the first two statistical moments cannot be consistent estimators. Estimated values of parameters are necessarily random, and, therefore, results of estimation must by subdued to testing.

The data set consists of measurements of arsenic in soil at 112 locations on the Island Krk [8]. Locations are evenly spread over the island and illustrated in Figure 1. This data set is used to estimate necessary parameters. By using smoothing and regression analysis, the parameters are estimated and the obtained values are presented in Table 1. The parameter ρ_l in (9) could not be estimated from available data because of too large distances between the measurement locations. We can only say that its value is larger than $5\,\mathrm{km}^{-1}$. This confirms the name "short range" component for ξ_l in the representation (8). The estimated value of ρ_g clearly indicates that the field ξ_g is "slowly" varying, since $K_g(r)/K_g(0) > 0.1$ for $r < 4.65\,\mathrm{km}$.

The event Γ of (1) is defined with $28 = 112/4$ data (Figure 1). The chosen 28 locations are evenly spread over D. The intervals I_m, $m = 1, 2, \ldots, 28$, of (1) are centred at the measured values and all have the same length equal to $4\sigma_l$, where $\sigma_l = \sqrt{A_l}$. The region D is covered with a grid having the grid step equal 0.588 km. The number of grid knots in D is $J_n = 1683$ and their index set is denoted by $\mathcal{J}_n$.

Computational aspects of eigenfunction decomposition and simulation

The basis of simulation is (4), where s_{kl} are matrix elements of $\sqrt{C_n}$, and C_n is the J_n–dimensional covariance matrix having the elements $c_{kl} = K_g(|\mathbf{x}_k - \mathbf{x}_l|)$ as defined in (10). Let (λ_k, e_k), $k = 1, 2, \ldots, J_n$, denote pairs consisting of eigenvalues λ_k and the corresponding eigenvectors e_k of C_n, such that $\|e_k\|_2^2 = \sum_j |e_{jk}|^2 = 1$. It is assumed that the eigenvalues are sorted in a nonincreasing order. The expression (4) can be written as

$$\xi(x_k) = \sum_{l \in \mathcal{J}_n} \sqrt{\lambda_l} e_{lk} \eta_l.$$

The computational problems related to $\sqrt{C_n}$ are addressed first. Since the matrix C_n is neither sparse nor too large, its square root can be efficiently calculated by using eigenvalue decomposition. An alternative approach would be the Newton's method [4]. The routine `dsyev` from LAPACK library has been used to compute all eigenvalues and eigenvectors of the matrix C_n. Let $\widetilde{\Lambda}$ be the diagonal matrix of computed eigenvalues in floating point arithmetic. The routine `dsyev` is designed to produce the matrix $\widetilde{\Lambda}$ such that there exists an orthogonal matrix $\widetilde{U}$ so that

$$\widetilde{U}\widetilde{\Lambda}\widetilde{U}^{\tau} = C_n + \delta C_n, \quad \|\delta C_n\|_2 \leq \alpha \cdot \texttt{eps}\, \|C_n\|_2,$$

where δC_n is a symmetric matrix, α is a constant depending on the dimension of the matrix, and `eps` is the machine precision. Let $\tilde{\lambda}_i$ be the computed values of λ_i. Weyl's theorem can be applied in order to discuss the accuracy of computed eigenvalues,

$$\frac{|\lambda_i - \tilde{\lambda}_i|}{\lambda_1} \leq \alpha\, 10^{-16} \kappa_2(C_n),$$

where $\kappa_2(C_n) = \|C_n\|_2 \|C_n^{-1}\|_2$ is the condition number, and $\texttt{eps} = 10^{-16}$ for double precision. A computed eigenvalue $\tilde{\lambda}$ is considered to be accurate if λ is of the same order of magnitude as λ_1. Eigenvectors that belong to those eigenvalues can also be considered as accurate. In our case, the largest computed eigenvalue is equal 77.05 and 104 eigenvalues have computed values larger than 1. We consider those eigenpairs as accurately computed.

Now we address problems related to simulations. Instead of the exact expression (4) for simulations at $\mathbf{x} \in G_n(D)$, we use an approximation

$$\xi_{app}(\mathbf{x}_k) = \sum_{l=1}^{J_{red}} \sqrt{\lambda_l} e_{lk} \eta_l, \tag{11}$$

where J_{red} is chosen so that the relative error of this approximation, measured in variance, is less than 10%,

$$\frac{\mathbf{Var}[\xi(\mathbf{x}) - \xi_{app}(\mathbf{x})]}{\mathbf{Var}[\xi(\mathbf{x})]} = \frac{\sum_{l > J_{red}} \lambda_l}{\sum_{l=1}^{J_n} \lambda_l} \leq 0.1, \quad \mathbf{x} \in G_n(D).$$

It follows that $J_{red} = 57$. Consequently, only accurately computed pairs (λ_k, e_k) of C_n form the representation (11).

Let N be the number of simulations of (11), and let $N_\Gamma < N$ be the number of simulations for which $\xi_{app} \in \Gamma$, so that the estimated probability of $\mathbf{P}(\Gamma)$ is equal $\mathbf{P}(\Gamma) = N_\Gamma/N$. Due to a small value of $\mathbf{P}(\Gamma)$, the number of simulations must be very large. Therefore, a periodicity of sequences of generated random variables must be analyzed and ruled out. Periodicity can be avoided by random mixing of random numbers from various sources [6]. In our computations `slaran` random number generator from LAPACK library is taken as the primary source. Random mixing was carried out by using streams of random bits from two sources, [6] and [3], in the usual way, by producing random numbers from sequences of 32 random bits. Estimated values of $\mathbf{P}(\Gamma)$ from 5 different mixing protocols are presented in Table 2.

$N = 2 \cdot 10^7$	$\mathbf{P}(\Gamma)$	secondary source
LAPACK	$0.211 \cdot 10^{-4}$	none
rezma1	$0.211 \cdot 10^{-4}$	[6]
rez5001	$0.207 \cdot 10^{-4}$	[6]
rez1001	$0.221 \cdot 10^{-4}$	[3]
rez5002	$0.209 \cdot 10^{-4}$	[1]

Table 2.

Results of simulation

The estimated field is mapped and illustrated in Figure 2. The random variables

$$\zeta(\mathbf{x}) = \xi(\mathbf{x}) - \mathbf{E}[\xi_g(\mathbf{x}) \,|\, \Gamma], \quad \mathbf{x} \in G_n(D),$$

have the zero conditional expectation and conditional variance equal to:

$$\sigma^2(\mathbf{x}, \Gamma) = K_l(0) + \mathbf{Var}[\xi_g(\mathbf{x}) \,|\, \Gamma].$$

Because the variance $\mathbf{Var}[\xi_g(\mathbf{x}) \,|\, \Gamma]$ is also estimated in the process of simulation, the normalized random variables

$$\eta(\mathbf{x}) = \frac{\xi(\mathbf{x}) - \mathbf{E}[\xi_g(\mathbf{x}) \,|\, \Gamma]}{\sigma(\mathbf{x}, \Gamma)}$$

can be utilized for a comparison with measurements. Their first two conditional moments are constant, $\mathbf{E}[\eta(\mathbf{x}) \,|\, \Gamma] = 0$, $\mathbf{Var}[\eta(\mathbf{x}) \,|\, \Gamma] = 1$. The measured values

$$\eta_m = \frac{\xi_m - \mathbf{E}[\xi_g(\mathbf{x}_m) \,|\, \Gamma]}{\sigma(\mathbf{x}_m, \Gamma)}$$

are mapped and random volumes

$$\Theta(\kappa) = \int_D \mathbb{1}_{|\eta|>\kappa}(\mathbf{x})\, d\mathbf{x}$$

are calculated. The random variable $\Theta(\kappa)$ can be interpreted as the volume of excessive values ξ_m, i.e., the values outside the layer in $\mathbb{R}^3$ with mid-surface $\mathbf{E}[\xi_g(\mathbf{x}) \,|\, \Gamma]$ and thickness $4\sigma_l\kappa$. For 5 estimates of the conditional moments, the following results are obtained:

$$\frac{\Theta(1)}{\Theta(0)} \approx 0.035 \sim 0.038.$$

This is a measure of confidence. Although this is not a genuine statistical measure, it has to be used, together with statistical tests, as a criterion for acceptance of the obtained results.

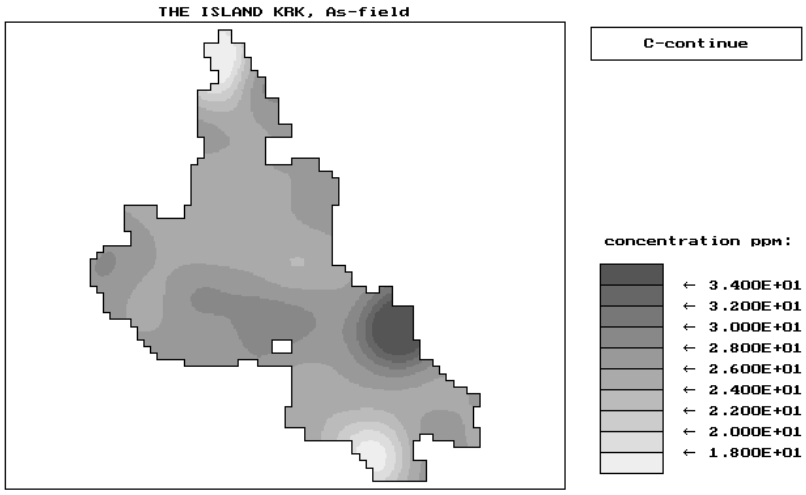

Figure 2. The global component of arsenic field.

4. Discussion and conclusion

Since the mathematical objects utilized in simulation are well-known, we have to discuss only their appropriate choice and efficiency in solving the defined problem.

Random characterization of trace element concentration in soil is analyzed in terms of the following stochastic model. A trace element concentration in soil is defined

by the local and global field components (8), corresponding covariance functions (9), (10), and Loeve–Karhunen decomposition of the global components. In order to verify whether data structure is consistent with the supposed model structure, statistical tests must be applied to data set. After a reliable estimation of the first two statistical moments, the data set is split into two spatially disjoint portions. One portion is used for simulation of the field, and the other one for verification of the supposed random structure in Loeve–Karhunen decomposition. In case of the analyzed data set (arsenic field in soil), the following conclusions can be drawn:

1. The data on arsenic field in soil were first treated without any smoothing. The result was poor. The only conclusion was that correlation could not be identified and the set of 112 data behaves as a sample of normal variables. After smoothing, the supposed structure of (8) was clear. Unfortunately, the variance of estimated covariance function is too large and the mutual portion of two fields, ξ_l, ξ_g, cannot be reliably estimated. We have to conclude from this analysis that the punctual sampling of soil cannot give us a reliable information about the structure of arsenic field in soil of the Krk island. Additional sampling must be planned with the aim to get a reliable data set for the determination of long range trace element field in soil.

2. There exists a local component of trace element field which behaves like a noise. In order to analyze the structure of noise, at few locations at least, punctual samples from a small region must be available.

3. The Assumption $1(a)$ of the constant mean value, $m = \mathbf{E}[\xi_g(\mathbf{x})]$, seems to be acceptable because of a relatively small variance of the estimated value m. The Assumption $1(b)$ is already discussed in the first item. The Assumption $1(c)$ can be efficiently verified and other nonnormal fields utilized and simulated, after a reliable statistics is obtained.

4. If the field is normal (Gaussian), the conditional expectation can be obtained in principle by LLSE of (5). To get an estimate at $\mathbf{x} \in G_n(D)$, one has to solve the linear algebraic system (5) in variables $w_k(\mathbf{x})$. In the considered case, the order of the system is 112, but the matrix of the system is ill-conditioned. The maximal and minimal eigenvalues are 7.5 and $3.0 \cdot 10^{-5}$, respectively. The matrix elements of its inverse are graphically presented in Figure 3. The inverse matrix is obtained from the spectral decomposition.

5. Other trace elements in soil [8] were also analyzed in order to check similarity with arsenic. All of them have properties mentioned in item 1, no correlation is established in the original data set. It seems that all measured trace elements have structure (8). Since the covariances of various trace element concentrations at a location have meaning only for the corresponding global components, multivariate analysis is not possible until the structure (8) is established for each trace element.

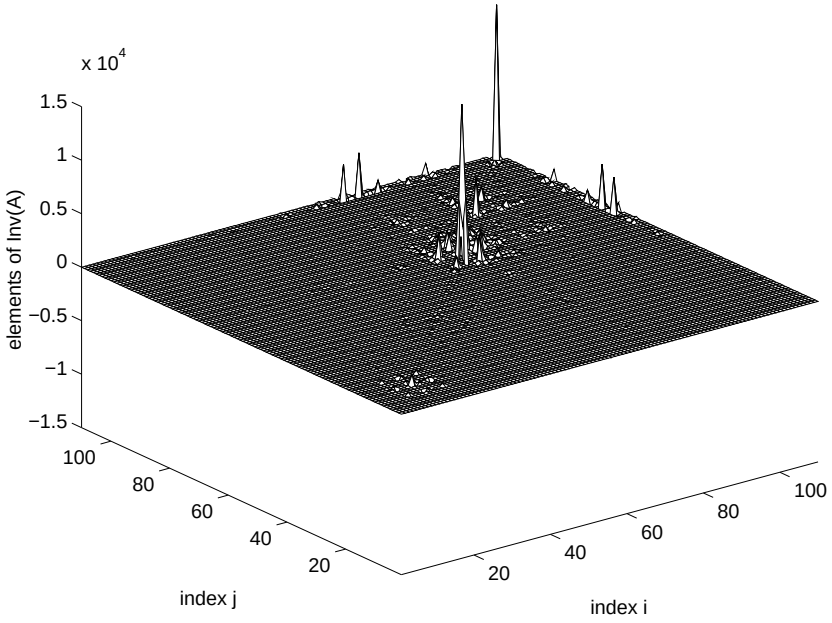

Figure 3. The inverse of the covariance matrix.

References

[1] E. Anderson, Z. Bai, C. H. Bischof, J. W. Demmel, J. J. Dongarra, J. Du Croz, A. Greenbaum, S. J. Hammarling, A. McKenney, S. Ostrouchov, and D. C. Sorensen, *LAPACK Users' Guide, Release 2.0*, SIAM, Philadelphia, 1995.

[2] V. I. Bogachev, *Gaussian Measures*, Mathematical Surveys and Monographs 62, AMS, Providence, RI, 1998.

[3] D. Eastlake, S. Crocker, and J. Schiller, *Randomness Recommendations for Security*, `http://www.cis.ohio-state.edu/htbin/rfc/rfc1750.html`

[4] N. J. Higham, *Newton's method for the matrix square root*, Math. Comp., 46 (1986), pp. 537–549.

[5] A. G. Journel and Ch. J. Hujibregts, *Mining Geostatistics*, Academic Press, New York, 1978.

[6] G. Marsaglia, *The Marsaglia random number CDROM.*

[7] M. A. Sironvalle, *The random coin method: solution of the problem of the simulation of a random function in the plane*, J. Internat. Assoc. Math. Geol., 12 (1980), pp. 25–32.

[8] V. Valković, ed., *An estimate of environmental conditions on the Island Krk*, A&CT, Zagreb, 1999.