A turning bands program for conditional co-simulation of cross-correlated Gaussian random fields ☆

Xavier Emery*

Department of Mining Engineering, University of Chile, Avenida Tupper 2069, Santiago 837 0451, Chile

Abstract

A Matlab program (TBCOSIM) is provided for co-simulating a set of stationary or intrinsic Gaussian random fields in $\mathbb{R}^3$, whose simple and cross-covariance functions are fitted by a linear model of coregionalization. It relies on the turning bands method, which performs three-dimensional simulation via a series of one-dimensional simulations along lines that span $\mathbb{R}^3$. There is no restriction on the number of random fields to simulate, on the number of basic structures used in the coregionalization model, and on the number and configuration of the locations where simulation has to be performed. Additionally, the realizations can be made conditional to data, back-transformed and averaged over a block support. TBCOSIM uses parallel simulation algorithms: at each location, the random fields are simulated simultaneously and a single co-kriging is needed for conditioning all the realizations. The capabilities of the program are illustrated with the analysis of a set of non-conditional realizations and with an application to a soil contamination dataset.

Keywords: Multivariate geostatistics; Stochastic simulation; Co-kriging; Linear model of coregionalization; Multivariate normal distribution

1. Introduction

Many disciplines in the geosciences are concerned with the characterization of co-regionalized variables monitored at sampling locations. Examples of applications include the evaluation of recoverable resources in polymetallic deposits, the modeling of the petrophysical properties of a reservoir or an aquifer, the remediation of soils polluted by heavy metals, or the prediction of soil properties for agricultural land management.

In this context, practitioners may be interested in predicting the values of the coregionalization over a given domain or in assessing the uncertainty attached to these values, either at a single location or jointly over several locations. The prediction problem can be solved by using co-kriging methods (Goovaerts, 1997; Wackernagel, 2003), while spatial uncertainty can be assessed by resorting to geostatistical co-simulation, which aims at constructing outcomes (*realizations*) that reproduce the spatial variability of each variable as well as the relationships between the variables. This approach requires defining a random field model and an algorithm to

☆ Code available from server at http://www.iamg.org/CGEditor/index.htm.

*Tel.: +56 2 978 4498; fax: +56 2 978 4985.
E-mail address: xemery@ing.uchile.cl

construct realizations of this model (Chilès and Lantuéjoul, 2005). To date, the simplest and most widespread model is that of Gaussian random fields, whose finite-dimensional distributions are multivariate normal.

A variety of algorithms have been proposed in the literature. Although some of them provide perfect simulations of Gaussian random fields, they suffer from limitations concerning the number or the spatial configuration of the locations where these random fields can be simulated. For instance, the discrete spectral algorithm proposed by Pardo-Igúzquiza and Chica-Olmo (1993, 1994) requires the locations to be evenly spaced; also, this algorithm cannot reproduce exactly some covariance models that are smooth near the origin, as pointed out by Chilès and Delfiner (1997). The multivariate extension of the matrix decomposition method (Myers, 1989) is not suitable to simulating random fields on domains containing more than a few thousands locations.

To avoid such limitations, some approximations are required. In this respect, one can distinguish algorithms that generate Gaussian random fields with covariance functions that differ from the desired model, and algorithms that do not generate Gaussian random fields, although the covariance model may be accurately reproduced. In the first category, one finds algorithms based on a discretization of continuous spectral representations (Robin et al., 1993; Gutjahr et al., 1997) and the variants of the sequential Gaussian algorithm that depart from a unique neighborhood implementation (Tran, 1994; Gómez-Hernández and Cassiraga, 1994; Emery, 2004).

The second category of approximate algorithms includes the continuous spectral and the turning bands methods. The former (Shinozuka and Jan, 1972) relies on a weighted sum of cosine functions with random frequencies and random phases. It is not general as it assumes that the spectral density of the random field is zero outside a bounded interval. Also, the convergence of the sum to random fields with multivariate normal distributions is slow if the covariance model is not smooth at the origin (Lantuéjoul, 2002). Regarding the turning bands method (Matheron, 1973; Mantoglou, 1987), it performs the simulation of a d-dimensional random field via a series of one-dimensional (1D) simulations along lines that span $\mathbb{R}^d$. A common criticism to this method is the artifact banding (stripping effect) caused when using too few lines or when discretizing the 1D simulations. However, this stripping can be considerably attenuated with a suitable choice of the line simulation algorithm and of the number and distribution of lines in $\mathbb{R}^d$ (Lantuéjoul, 1994; Gneiting, 1999; Emery and Lantuéjoul, 2006).

Among the available softwares for co-simulating Gaussian random fields, the turning bands program proposed by Carr and Myers (1985) is limited to 2D spaces and spherical covariance models; more importantly, this program fails at reproducing the cross-correlation between non-conditional random fields (Myers, 1989). Many other softwares, such as the GSLIB library (Deutsch and Journel, 1992), GCOSIM3D (Gómez-Hernández and Journel, 1993) or Gstat (Pebesma, 2004), only offer sequential Gaussian co-simulation programs; some of them are restricted to Markov-type covariance models and colocated co-kriging, while others allow using a full co-kriging (in practice, limited to a moving neighborhood) to derive conditional distributions. The objective of this work is to present a Matlab program for conditional co-simulation in $\mathbb{R}^3$ using the turning bands method.

2. Co-simulation of cross-correlated Gaussian random fields

2.1. Modeling spatial correlations in a multivariate framework

A difficulty with the simulation of co-regionalized variables is the requirement for cross-covariance functions that model the relationships between the different variables. Let us consider a stationary vector random field $\boldsymbol{Y} = (Y_1, \ldots Y_M)^t$ whose components have zero means and multivariate normal distributions. These distributions are characterized by the matrix of simple and cross-covariances

$$\forall \mathbf{x}, \mathbf{x}+\mathbf{h} \in \mathbb{R}^d, \quad \mathbf{C}(\mathbf{h}) = E\{\boldsymbol{Y}(\mathbf{x})\boldsymbol{Y}(\mathbf{x}+\mathbf{h})^t\}. \tag{1}$$

There exist several approaches for modeling the simple and cross-covariances, e.g.:

- Markov-type model (Almeida and Journel, 1994).
- Intrinsic correlation (also known as "intrinsic coregionalization" or "proportional covariance model") (Journel and Huijbregts, 1978).
- Linear model of coregionalization (Journel and Huijbregts, 1978).
- Bilinear model of coregionalization (Grzebyk and Wackernagel, 1994).

- Use of spectral representations (Pardo-Igúzquiza and Chica-Olmo, 1994; Gutjahr et al., 1997)
- Models based on the square roots of covariance functions (Oliver, 2003).
- Models induced by relationships (e.g. partial differential equations) between the variables (Chilès and Delfiner, 1999; de Fouquet, 2001).

Because of this diversity of covariance modeling approaches, it seems difficult (if not impossible) to design an all-purpose co-simulation algorithm. In this paper, we will focus on the *linear model of coregionalization*, in which the matrix of simple and cross-covariances is of the form (Wackernagel, 2003, p. 175)

$$\mathbf{C}(\mathbf{h}) = \sum_{n=1}^{N} \mathbf{B}_n \rho_n(\mathbf{h}), \tag{2}$$

where $\{\rho_n,\ n = 1 \ldots N\}$ is a set of covariance functions (positive semi-definite functions); $\{\mathbf{B}_n, n = 1 \ldots N\}$ is a set of symmetric positive semi-definite matrices.

To broaden the scope of this work, the stationarity hypothesis will be weakened by considering the more general model in which $(Y_1, \ldots Y_M)$ are jointly intrinsic Gaussian random fields (i.e. random fields for which increments, or generalized increments, are stationary and have multivariate normal distributions). In such a case, the spatial correlation at lag $\mathbf{h}$ is characterized by a matrix of *generalized* simple and cross-covariance functions (Dowd, 1989; Chilès and Delfiner, 1999, p. 331). In the sequel, we will still denote by $\mathbf{C}(\mathbf{h})$ this matrix and assume that it is of the form given in Eq. (2), except that $\{\rho_n, n = 1 \ldots N\}$ may now be generalized covariances instead of ordinary covariances. Henceforth, the term "covariance" will be used indistinctly to designate an ordinary or a generalized covariance.

2.2. Non-conditional simulation

For $n = 1 \ldots N$, denote by $\mathbf{Q}_n$ and $\mathbf{D}_n$ the matrices of eigenvectors and eigenvalues of $\mathbf{B}_n$ (an orthogonal and a diagonal matrix, respectively) and define the square root matrix of $\mathbf{B}_n$

$$\mathbf{B}_n = \mathbf{A}_n \mathbf{A}_n^t \text{ with } \mathbf{A}_n = \mathbf{Q}_n \sqrt{\mathbf{D}_n}. \tag{3}$$

Let $\boldsymbol{X}_n$ be a Gaussian vector random field with M independent components, each with covariance function ρ_n. As shown by Wackernagel (2003, p. 176), the random field $\boldsymbol{Y}$ can be written in the following fashion:

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \boldsymbol{Y}(\mathbf{x}) = \sum_{n=1}^{N} \mathbf{A}_n \boldsymbol{X}_n(\mathbf{x}). \tag{4}$$

The simulation of a Gaussian vector random field with cross-correlated components ($\boldsymbol{Y}$) therefore boils down to that of a set of independent scalar Gaussian random fields (the components of $\boldsymbol{X}_1 \ldots \boldsymbol{X}_N$). Note that the decomposition given in Eq. (4) is the basis of factorial kriging analysis (Goovaerts, 1993, 1997; Wackernagel, 2003).

2.3. Conditioning of the data

To produce realizations conditioned to a set of data, a supplementary step is required. In the univariate case, this is done by adding a non-conditional realization and the kriging of the difference between the data values and the simulated values at the data locations (Journel and Huijbregts, 1978, p. 495). This result holds in the multivariate case, except that co-kriging must be used instead of kriging (Carr and Myers, 1985; Myers, 1989; Gutjahr et al., 1997)

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \boldsymbol{Y}_{CS}(\mathbf{x}) = \boldsymbol{Y}_S(\mathbf{x}) + [\boldsymbol{Y}(\mathbf{x}) - \boldsymbol{Y}_S(\mathbf{x})]^*, \tag{5}$$

where the asterisk indicates the co-kriging operator from the values at the data locations and subscripts S and CS stand for non-conditional and conditional simulation, respectively.

Several comments are worth being made:

(1) The co-kriging weights needed in Eq. (5) only depend on the covariance model and on the configuration of data locations and locations targeted for simulation. This entails that a single co-kriging system has to be solved to condition multiple realizations.
(2) Care must be taken in the design of the co-kriging neighborhood. Over-restricted neighborhoods (with small search radii and few conditioning data) often introduce discontinuities in the simulation maps, which can be spuriously considered as part of the spatial variability of the coregionalization, and provoke a loss of precision and an increase of conditional bias in the co-kriging estimates (Rivoirard, 1987; Chilès and Delfiner, 1999). These drawbacks are likely to be more severe for multivariate simulation than for univariate simulation, as the optimal choice of co-kriging neighborhood

depends on the data configuration of the variables of interest and on their simple and cross-covariance functions (Rivoirard, 2004).

(3) When a unique neighborhood is considered, the conditioning process (Eq. (5)) can be made faster by using the dual form of co-kriging, in which a single matrix system has to be solved for all the locations where simulation is required (Chilès and Delfiner, 1999, p. 186).

(4) Co-kriging and co-simulation can be performed with heterotopic datasets, for which not all the variables are known at each data location.

(5) Simple co-kriging is used for conditioning the realizations of stationary random fields (those for which the covariance functions have a finite sill) with known mean values. Ordinary co-kriging is used for stationary random fields with uncertain mean values or for random fields with stationary increments (intrinsic random fields of order 0) (Emery, 2004, 2007b). Intrinsic co-kriging is used for random fields with polynomial drifts and stationary residuals and for intrinsic random fields (Dimitrakopoulos, 1990; de Fouquet, 1994).

3. Program description

The Matlab program TBCOSIM uses the turning bands algorithm for simulating the components of $\boldsymbol{X}_1,\ldots\ \boldsymbol{X}_N$ (Eq. (4)). It is restricted to simulation in $\mathbb{R}^3$ (1D or 2D simulations can be performed by setting the remaining coordinates to a constant, say zero) and relies on a code developed in the univariate framework by Emery and Lantuéjoul (2006). The main features of this program are the following ones:

(1) Use of parallel algorithms: at each location, the M variables are simulated all together (regardless of the values simulated at other locations) and all the realizations are made conditional by solving a single co-kriging system. This allows a considerable reduction of calculations with respect to sequential algorithms, in which the variables are simulated in turn conditionally to previously simulated values and one co-kriging is needed for each realization. Co-kriging is performed in a moving neighborhood defined by an ellipsoid that may be divided into octants. A unique neighborhood is assumed if the radius of this ellipsoid is set to infinity; in such a case, dual co-kriging is used.

(2) Use of continuous simulation algorithms along the lines, which avoids discretizing the 1D simulations and allows reproducing the simple and cross-covariances of $\boldsymbol{Y}$ without bias, even if the simulation is performed at unevenly spaced locations in $\mathbb{R}^3$ (Emery and Lantuéjoul, 2006).

(3) No restriction on the number of nested structures used in the linear model of coregionalization. Each structure can have a geometric anisotropy defined by three angles and three scale factors, following GSLIB conventions (Deutsch and Journel, 1992, p. 23). The available covariance models are the spherical, exponential, gamma, stable, cubic, Gaussian, cardinal sine, J-Bessel, K-Bessel (Matérn), generalized Cauchy, exponential sine, linear, power, mixed power and spline (Emery and Lantuéjoul, 2006) (see subroutine COVA for details).

(4) The convergence of the simulated distributions to multivariate normality is controlled by the number of lines used to simulate each nested structure. Suggestions for choosing this number have been given by Tompson et al. (1989), Freulon and de Fouquet (1991), Gneiting (1999), Lantuéjoul (1994, 2002) and Emery and Lantuéjoul (2006), among others; a general recommendation for 3D simulation is to use hundreds or thousands of lines.

(5) Post-processing options: the realizations can be back-transformed from the Gaussian unit to that of the variables of interest by specifying a set of transformation tables and parameters for tail extrapolation, following the approach proposed by Emery (2006). More precisely, let $\{(z_k^{(m)}, y_k^{(m)}), k = 1\ldots K\}$ be the transformation table for the mth variable Y_m, and $z_{\min}^{(m)}$ and $z_{\max}^{(m)}$ the extrema of the back-transformed variable. The transformation function (ϕ_m) is modeled through a piecewise interpolation of the transformation table, completed by exponential functions that depend on two positive parameters (λ_m,λ'_m):

$$\begin{cases} \forall y < y_1^{(m)}, & \phi_m(y) = z_{\min}^{(m)} + (z_1^{(m)} - z_{\min}^{(m)})\mathrm{e}^{\lambda_m(y-y_1^{(m)})}, \\ \forall y > y_K^{(m)}, & \phi_m(y) = z_{\max}^{(m)} + (z_K^{(m)} - z_{\max}^{(m)})\mathrm{e}^{\lambda'_m(y_K^{(m)}-y)}. \end{cases} \tag{6}$$

(6) Also, the realizations can be averaged over a block support, via a discretization of the block.

TBCOSIM can be run directly in the Matlab workspace with its input arguments (see the program

file for details). Alternatively, it can be used with a parameter file: in such a case, no input is needed and the user is prompted for the name of the parameter file (by default, TBCOSIM.PAR) (Table 1); the coordinates and values of the Gaussian data and, if the simulation has to be performed at scattered locations, the coordinates of these locations must be stored in ASCII files. TBCOSIM does not produce any output in the Matlab workspace, but creates an external ASCII file in which each column corresponds to one realization of one variable (the first M columns correspond to the first realization of the whole coregionalization, and so on). For grid simulation, the ordering of the grid nodes is the same as in GSLIB, i.e. point by point to the east, then row by row to the north, and level by level upward (Deutsch and Journel, 1992, p. 20). Missing data values (e.g. in heterotopic datasets) can be codified by values outside a trimming interval or by NaN (not-a-number).

TBCOSIM uses 12 sub-routines:

- BACKTR: back-transform from Gaussian to original scale.
- COKRIGE: compute co-kriging weights (moving neighborhood).
- COVA: compute covariance values.
- CREATE_PARAMFILE: create default parameter file TBCOSIM.PAR.
- DUAL: compute dual co-kriging weights (unique neighborhood).
- PICKSUPR: build templates of super-blocks (moving neighborhood).
- SEARCH: search for data in a moving neighborhood.
- SETDUAL: compute right-hand side of dual co-kriging system.
- SETROT: set up rotation matrix to transform Cartesian coordinates.
- SUPERBLK: set up super-block strategy (moving neighborhood).
- TBMAIN: perform non-conditional simulation along the lines.
- VDC: generate equidistributed lines over the unit 3D sphere.

4. Checking the reproduction of the random field model

In this section, we consider the non-conditional simulation of a pair of Gaussian random fields over a 200×200 grid. The coregionalization model consists of a nugget effect, an isotropic spherical structure with range 50 and an isotropic Matérn (K-Bessel) structure with scale factor 10 and shape parameter 2:

$$\mathbf{C}(\mathbf{h}) = \begin{pmatrix} 0.2 & -0.1 \\ -0.1 & 0.1 \end{pmatrix} \text{nugget} + \begin{pmatrix} 1 & 0.4 \\ 0.4 & 0.6 \end{pmatrix} \text{sph}_{50}(\mathbf{h}) + \begin{pmatrix} 1.8 & -1.4 \\ -1.4 & 1.5 \end{pmatrix} K - \text{Bessel}_{10;2}(\mathbf{h}). \quad (7)$$

For the simulation of each nested structure, 1000 lines are used in the turning bands algorithm. To validate the quality of the realizations, two questions can be asked:

(1) Do they conform to the above linear model of coregionalization (Eq. (7))?
(2) Do they conform to a vector random field with multivariate normal distributions?

Concerning the first question, a simple validation exercise consists in examining the sample variograms of the realizations: on average over many realizations, they should match (up to reasonable statistical fluctuations) the prior variogram model. The results obtained for 100 realizations (Fig. 1, left) show a good agreement between average sample and theoretical variograms. Note that this agreement may not hold if *conditional* realizations are used, for three reasons:

(a) the recourse to a moving neighborhood for conditioning co-kriging may introduce spurious fluctuations (recall second comment in Section 2.3);
(b) the conditioning data may not exactly conform to the coregionalization model;
(c) even so, the simple and cross-variograms of the conditional simulations usually differ from the prior variogram models, unless the simulation domain is very large with respect to the domain surrounding the conditioning data. Put another way, the data introduce a "distorsion" with respect to the prior model (Lantuéjoul, 2002, p. 204; Emery, 2007b).

As for the second question (multivariate normality), the validation is often limited to the bivariate distributions. One option is to examine the sample indicator variograms for some quantiles and compare

Table 1
Default parameter file for TBCOSIM

```
START OF PARAMETERS:
0                                  type of co-simulation: 0 = gridded locations; 1 = scattered locations
locations.prn                      if  = 1:    file with coordinates of locations for co-simulation
1 2 3                                          columns for location coordinates
5.0 5.0 5.0                        if  = 0:    x0, y0, z0
20 20 10                                       nx, ny, nz
10.0 10.0 10.0                                 dx, dy, dz
5 5 2                              block discretization (1 1 1 for point-support co-simulation)
2                                  number of variables to simulate
nscore.out                         file with conditioning data
1 2 3                                          columns for coordinates
4 5                                            columns for Gaussian data
-10 10                                         trimming limits for Gaussian data
nscore1.trn                        variable 1:          file with conversion table (raw-Gaussian)
0.0 10.0                                                minimum and maximum values for raw variable
1.0 5.0                                                 parameters for lower-tail and upper-tail extrapolation
nscore2.trn                        variable 2:          file with conversion table (raw-Gaussian)
0.0 4.0                                                 minimum and maximum values for raw variable
2.0 1.0                                                 parameters for lower-tail and upper-tail extrapolation
2                                  number of nested structures
1 170 120 100 30 0 0 1 1000                1st structure:    it a1 a2 a3 ang1 ang2 ang3 b nlines
0.45 0.30 0.30 0.52                                          variance-covariance matrix
4 100 100 50 0 0 0 0.5 1000                2nd structure:    it a1 a2 a3 ang1 ang2 ang3 b nlines
0.45 0.25 0.25 0.30                                          variance-covariance matrix
0.10 0.11 0.11 0.18                nugget effect variance-covariance matrix
30                                 number of realizations
9784498                            seed for random number generation
200 200 100                        maximum search radii in the rotated system
30 0 0                             angles for search ellipsoid
1                                  divide into octants? 1 = yes, 0 = no
4                                  optimal number of data per octant (if octant = 1) or in total (if 0)
1                                  co-kriging type: simple = 0, ordinary = 1, intrinsic (order k) = k + 1
tbcosim.out                        name of output file
3                                  number of decimals for values in the output file
1                                  create a header in the output file? 1 = yes, 0 = no
5000                               maximum number of locations to simulate simultaneously
```

Available model types:

1: spherical
2: exponential
3: gamma (parameter b > 0)
4: stable (parameter b < 2)
5: cubic
6: Gaussian
7: cardinal sine
8: J-Bessel (parameter b > 0.5)
9: K-Bessel (parameter b > 0)
10: generalized Cauchy (parameter b > 0)
11: exponential sine
12: linear
13: power (exponent b > 0)
14: mixed power (exponent b <= 2)
15: spline (exponent b = even integer)

their averages over the realizations with the corresponding prior models (Goovaerts, 1997, p. 277). Fig. 1 (right) presents the results associated with the median indicator, which corroborate the assumption of bivariate normal distributions. This validation could be refuted if an insufficient number of lines were used to simulate each nested structure in the turning bands algorithm, see comment 4 in Section 3.

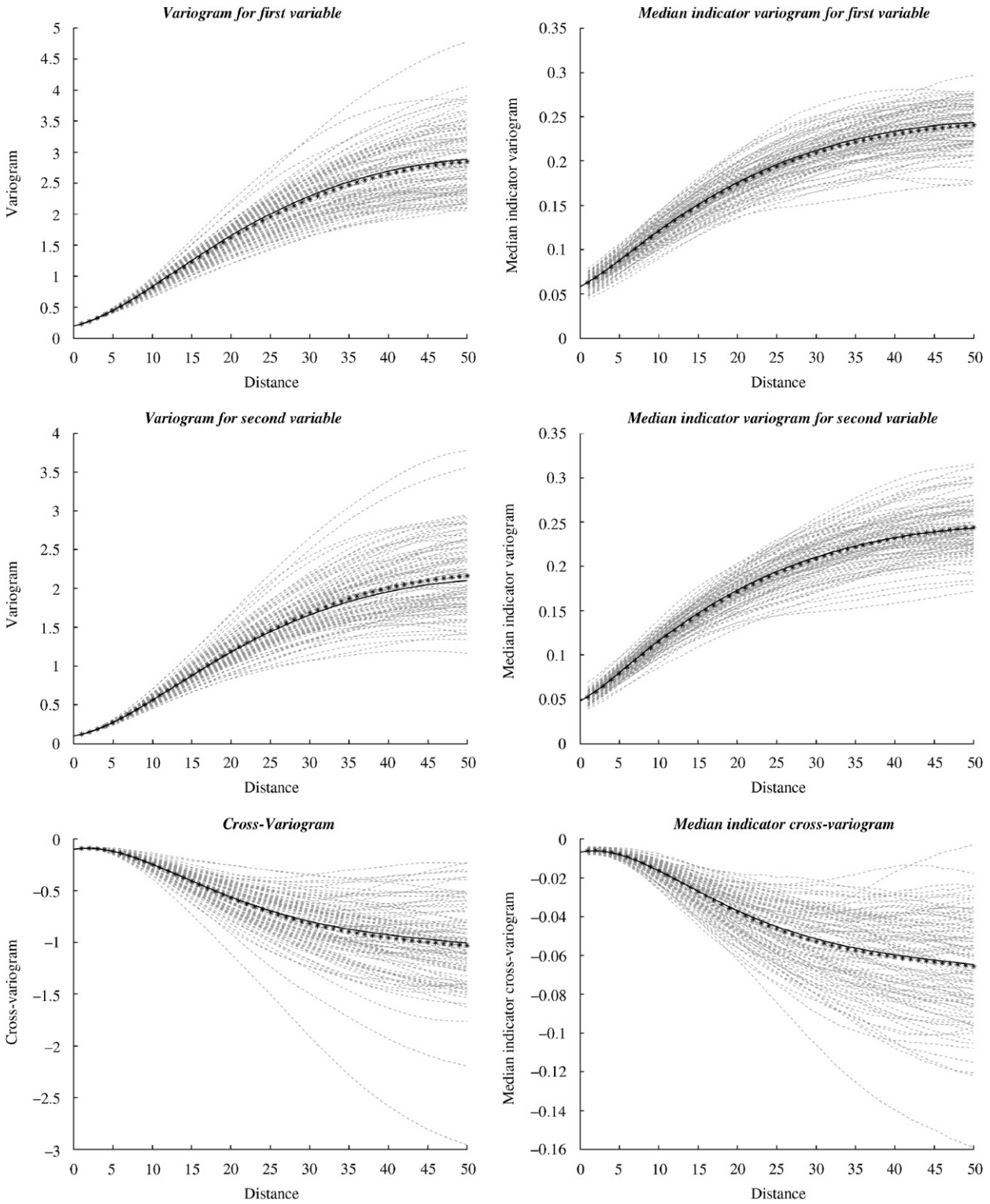

Fig. 1. Variograms of Gaussian random fields (left) and of their median indicators (right) for a set of 100 non-conditional realizations. Dashed lines correspond to sample variograms calculated along abscissa axis; asterisks indicate their average over 100 realizations; solid lines represent theoretical models.

5. Case study: Swiss Jura dataset

To illustrate the importance of the co-simulation approach in multivariate problems and the capabilities of program TBCOSIM, a case study is now presented. We will consider a dataset from the Swiss Federal Institute of Technology that consists of 359 measurements of heavy metal concentrations (cadmium, cobalt, chromium, copper, nickel, lead and zinc) in the topsoil of a contaminated site in the Swiss Jura. The site covers approximately 1450 ha and the sampling has been performed on a regular grid with mesh $250 \times 250\,m^2$ completed by several clusters (Fig. 2). Details on the site and laboratory procedures are given by Atteia et al. (1994). In the following, the analysis is focused on three contaminants: copper, lead and zinc (Table 2).

The goal of the study is to assess the risk that the contaminant concentrations exceed regulatory thresholds, in order to delineate the areas where cleanup is needed. We will assume that the remediation units are blocks with size $20 \times 20\,m^2$ and the thresholds for toxicity are 40 ppm for copper, 65 ppm for lead and 100 ppm for zinc.

The study consists of the following steps.

(1) Transform the original data (metal concentrations) into normal scores, accounting for declustering weights to correct for the irregular sampling design. In the present case, the weights have been obtained by the cell declustering technique (Goovaerts, 1997, p. 83). This technique accounts for the spatial configuration of the data locations, but ignores the correlation between data. To account for this correlation and improve the consistency of the declustering scheme, the weights could be defined via cokriging or via redundancy coefficients, by generalizing the approaches proposed by Switzer (1977), Bourgault (1997) or Bogaert (1999) to the multivariate context, based on the simple and cross-variograms of the variables or of their indicators.

(2) Calculate the sample (simple and cross) variograms for the normal scores data and fit a linear model of coregionalization. A nugget effect and two nested isotropic spherical structures, the first one with a range of 0.4 km and the second one with a range of 1.5 km, are used (Fig. 3)

$$\mathbf{C}(\mathbf{h}) = \begin{pmatrix} 0.18 & 0.20 & 0.06 \\ 0.20 & 0.38 & 0.14 \\ 0.06 & 0.14 & 0.15 \end{pmatrix} \text{nugget} + \begin{pmatrix} 0.84 & 0.50 & 0.57 \\ 0.50 & 0.36 & 0.31 \\ 0.57 & 0.31 & 0.58 \end{pmatrix} \text{sph}_{0.4}(\mathbf{h}) + \begin{pmatrix} 0.043 & 0.023 & 0.11 \\ 0.023 & 0.28 & 0.24 \\ 0.11 & 0.24 & 0.41 \end{pmatrix} \text{sph}_{1.5}(\mathbf{h}). \quad (8)$$

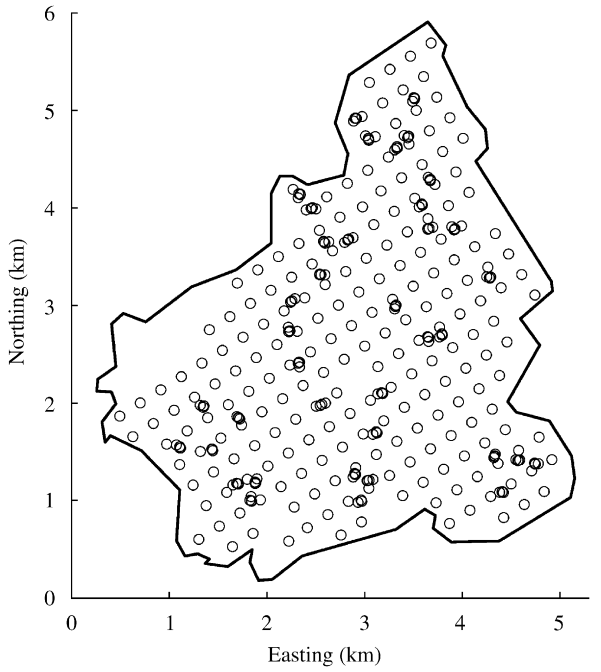

Fig. 2. Map of data locations (circles) and delineation of area under study (Jura dataset).

Table 2
Basic statistics on contaminant concentrations, expressed in ppm (Jura dataset)

	Mean	Standard deviation	Minimum	Lower quartile	Median	Upper quartile	Maximum
Cu	23.59	22.27	3.55	10.46	17.20	27.02	166.40
Pb	54.63	33.10	18.68	36.30	46.80	60.30	300.00
Zn	75.88	30.82	25.00	54.50	73.56	90.00	259.84

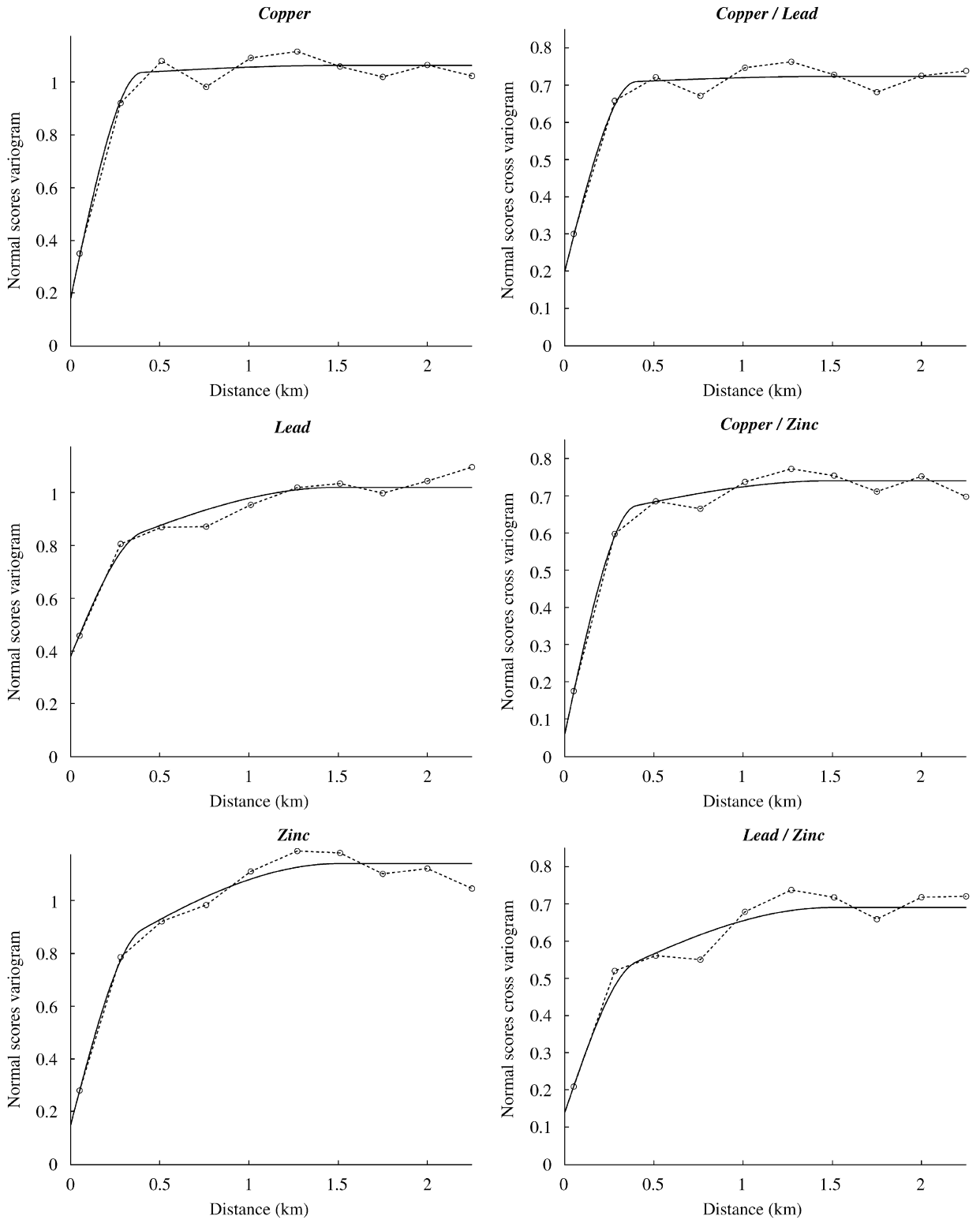

Fig. 3. Sample (dashed lines and circles) and modeled (solid lines) variograms for normal scores of copper, lead and zinc data (Jura dataset).

(3) Check for bivariate normality by examining the indicator variograms at some specific thresholds, see Section 4 and files INSTRUCTIONS_JURA, GAMV_IND (sample indicator variogram calculation) and VMODEL_IND (theoretical indicator variogram calculation) for computer programs and results.

(4) Perform conditional co-simulation of the metal concentrations over the remediation units, with the normal scores data and the fitted coregionalization model as input in TBCOSIM. The following parameters are considered:

- discretization of remediation units set to 6×6 points;
- simulation of each spherical structure based on 1000 lines;
- conditioning by ordinary co-kriging;
- definition of a moving neighborhood with a radius of 1.5 km and a maximum of 8 data in each quadrant;
- generation of 100 realizations.

As an illustration, the maps of contaminant concentrations corresponding to the first realization are shown in Fig. 4.

(5) For each remediation unit, evaluate the probability that the three metal concentrations are below the regulatory thresholds:

$$\begin{aligned} &1 - \text{Prob}\{\text{hazard}\}_{\text{mutivariate analysis}} \\ &= \text{Prob}\{\text{Cu} < 40, \text{Pb} < 65, \text{Zn} < 100\} \\ &= \text{Prob}\{\text{Cu} < 40 | \text{Pb} < 65, \text{Zn} < 100\} \\ &\times \text{Prob}\{\text{Pb} < 65 | \text{Zn} < 100\} \times \text{Prob}\{\text{Zn} < 100\} \end{aligned} \quad (9)$$

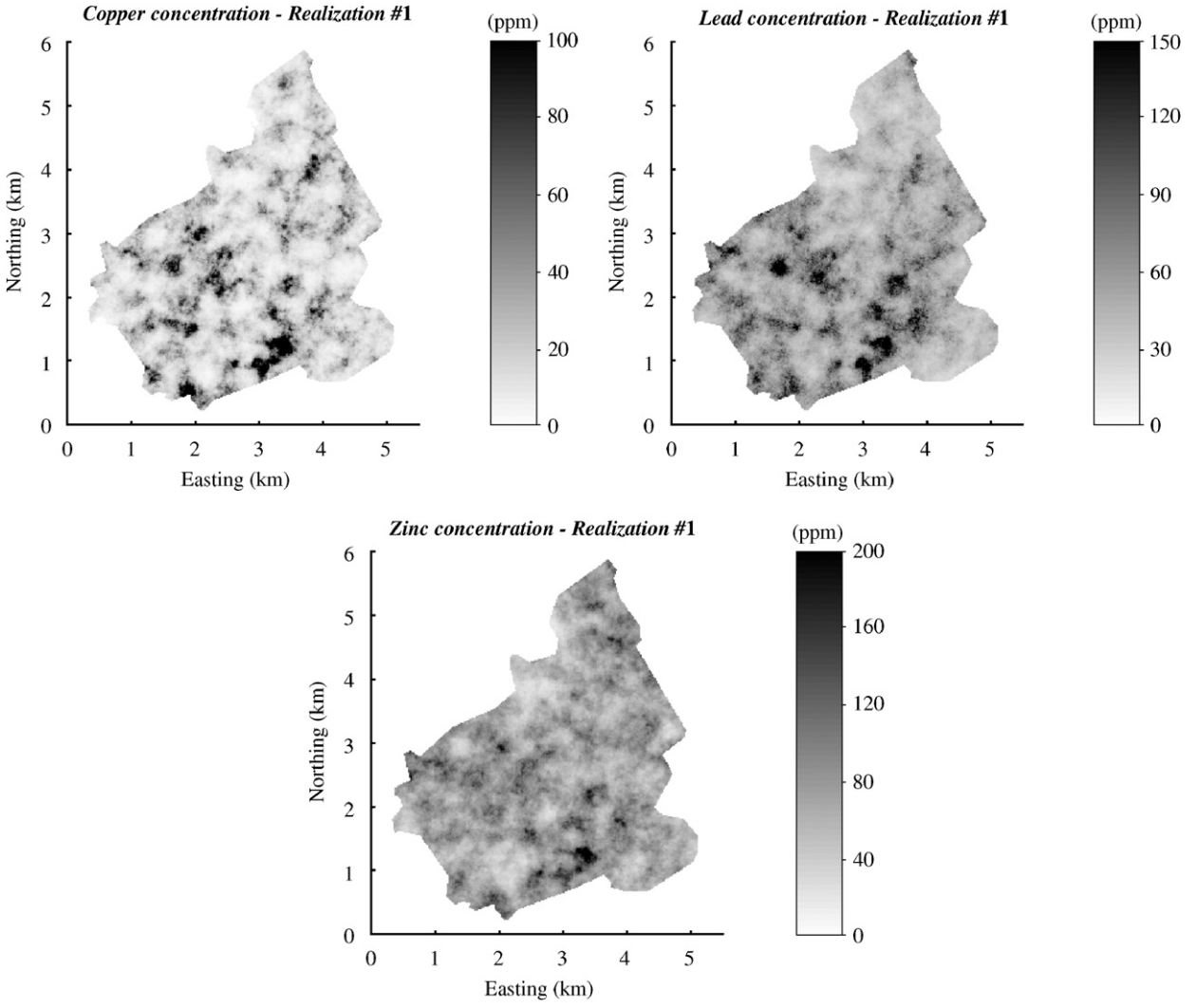

Fig. 4. Maps of simulated contaminant concentrations averaged over remediation units, corresponding to first conditional realization.

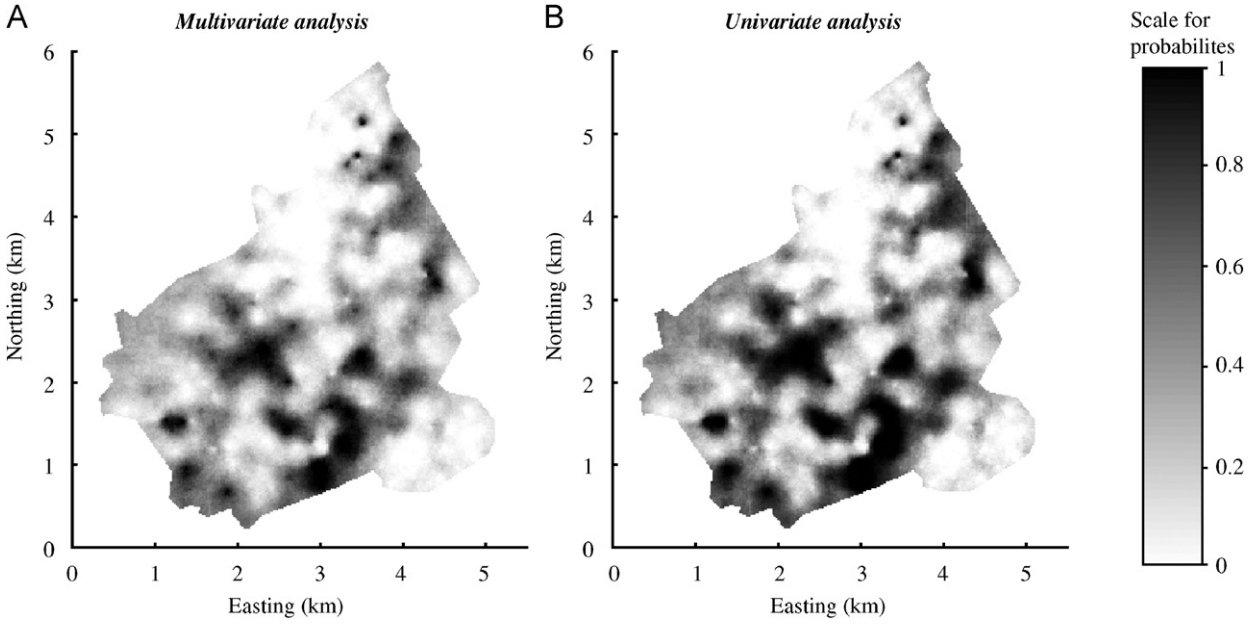

Fig. 5. Maps of hazard probability (probability that at least one contaminant concentration exceeds a regulatory threshold), calculated after 100 realizations averaged over remediation unit support.

(concentrations and thresholds in Eq. (9) are expressed in ppm). The probability of finding hazardous contaminant concentrations is mapped in Fig. 5A. For comparison, Fig. 5B shows the map that would be obtained if the metal concentrations were assumed independent and treated separately:

$$1 - \text{Prob}\{\text{hazard}\}_{\text{univariate analysis}} = \text{Prob}\{\text{Cu} < 40\} \times \text{Prob}\{\text{Pb} < 65\} \times \text{Prob}\{\text{Zn} < 100\}. \quad (10)$$

The hazard probability calculated by considering the variables independently tends to overstate the probability obtained by accounting for the cross-correlations between these variables, as corroborated in Fig. 6. This bias is explained because the copper, lead and zinc concentrations are positively correlated, hence, in general

$$\begin{cases} \text{Prob}\{\text{Cu} < 40 | \text{Pb} < 65, \text{Zn} < 100\} \geqslant \text{Prob}\{\text{Cu} < 40\}, \\ \text{Prob}\{\text{Pb} < 65 | \text{Zn} < 100\} \geqslant \text{Prob}\{\text{Pb} < 65\}. \end{cases} \quad (11)$$

The probability maps in Fig. 5 are useful for site management. For instance, the decision-maker can distinguish three areas (Demougeot-Renard and de Fouquet, 2004): a *remediation area* that contains the units with a hazard probability greater than a given threshold α and for which cleanup is needed; a *low-risk area* containing the units with a hazard probability less than a second threshold β (with $\beta < \alpha$), so that they can be left on site without further investigation; and an *uncertain area* (units with a probability between β and α), for which the risk of

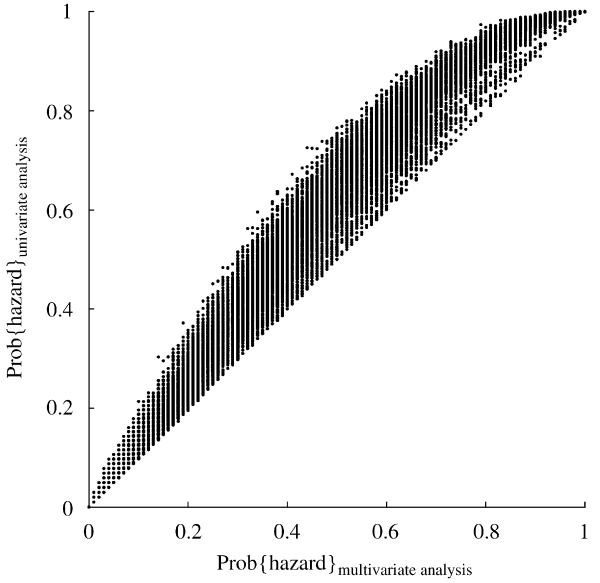

Fig. 6. Comparison of hazard probabilities obtained in Fig. 5.

Table 3
Number of remediation units in each area, calculated with probability thresholds $\alpha = 0.5$ and $\beta = 0.2$

	Remediation area	Uncertain area	Low-risk area
Multivariate analysis	8756	12,571	15,864
Univariate analysis	12,472	11,271	13,448

hazard is too low to be included in the remediation area, but too high to be left on site without further characterization (additional sampling). For given probability thresholds α and β, the extent of these three areas strongly depends on whether the cross-correlations between contaminant concentrations are accounted for or not (Table 3).

6. Conclusions

Many geostatistical simulation algorithms are still limited in multivariate situations, especially in what refers to the reproduction of cross-correlations between variables. In this paper, a Matlab co-simulation program (TBCOSIM) has been presented, which offers the following features:

- 3D simulation, either at gridded locations or scattered locations;
- no restriction on the number of these locations;
- no restriction on the number of nested structures;
- no restriction on the number of variables;
- handling of heterotopic datasets (variables known at different locations);
- use of stationary and intrinsic models;
- conditioning of the realizations to a set of data, using:
  - simple, ordinary or intrinsic co-kriging;
  - unique or moving neighborhood;
- availability of the most commonly used covariance models;
- back-transformation from normal values to original units;
- change of support (regularization) of the realizations.

The main advantages of TBCOSIM are the parallel procedure, in which all the variables are simulated simultaneously and a single co-kriging is used to condition all the realizations, and the avoidance of discretization for the simulation along the lines, which allows reproducing the simple and cross-covariances exactly and performing simulation at irregularly spaced locations in $\mathbb{R}^3$.

The drawbacks stem from the restrictions imposed by the choice of the random field model: multivariate normal distributions, simple and cross-covariances fitted by a linear model of coregionalization. The latter is widely used in spatial analysis, but may turn out to be insufficient for analyzing temporal or spatio–temporal data, as it excludes deferred correlations (delay effects). Program TBCOSIM can however be adapted to simulating more general models, e.g. Gaussian random fields with covariances fitted by a bilinear model of coregionalization that allows delay effects between the variables (Wackernagel, 2003, p. 207). It can also be extended to the simulation of plurigaussian and Gaussian-related random fields (Lantuéjoul, 2002; Emery, 2007a).

Acknowledgments

This research was funded by the National Fund for Science and Technology of Chile (FONDECYT) and is part of Project 1061103.

Appendix A. Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.cageo.2007.10.007.

References

Almeida, A.S., Journel, A.G., 1994. Joint simulation of multiple variables with a Markov-type coregionalization model. Mathematical Geology 26 (5), 565–588.

Atteia, O., Dubois, J.P., Webster, R., 1994. Geostatistical analysis of soil contamination in the Swiss Jura. Environmental Pollution 86 (3), 315–327.

Bogaert, P., 1999. On the optimal estimation of the cumulative distribution function in presence of spatial dependence. Mathematical Geology 31 (2), 213–239.

Bourgault, G., 1997. Spatial declustering weights. Mathematical Geology 29 (2), 277–290.

Carr, J.R., Myers, D.E., 1985. COSIM: a Fortran IV program for coconditional simulation. Computers & Geosciences 11 (6), 675–705.

Chilès, J.P., Delfiner, P., 1997. Discrete exact simulation by the Fourier method. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong' 96. Kluwer Academic, Dordrecht, pp. 258–269.

Chilès, J.P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York, 695pp.

Chilès, J.P., Lantuéjoul, C., 2005. Prediction by conditional simulation: models and algorithms. In: Bilodeau, M., Meyer, F., Schmitt, M. (Eds.), Space, Structure and Randomness. Springer, New York, pp. 39–68.

De Fouquet, C., 1994. Reminders on the conditioning kriging. In: Armstrong, M., Dowd, P.A. (Eds.), Geostatistical Simulation. Kluwer Academic, Dordrecht, pp. 131–145.

De Fouquet, C., 2001. Joint simulation of a random function and its derivatives. In: Kleingeld, W.J., Krige, D.G. (Eds.), Geostats 2000 Cape Town. Geostatistical Association of Southern Africa, Cape Town, pp. 84–93.

Demougeot-Renard, H., de Fouquet, C., 2004. Geostatistical approach for assessing soil volumes requiring remediation: validation using lead-polluted soils underlying a former smelting works. Environmental Science & Technology 38, 5120–5126.

Deutsch, C.V., Journel, A.G., 1992. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York, 340pp.

Dimitrakopoulos, R., 1990. Conditional simulation of intrinsic random functions of order k. Mathematical Geology 22 (3), 361–380.

Dowd, P.A., 1989. Generalised cross-covariances. In: Armstrong, M. (Ed.), Geostatistics. Kluwer Academic, Dordrecht, pp. 151–162.

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., 2006. A disjunctive kriging program for assessing point-support conditional distributions. Computers & Geosciences 32 (7), 965–983.

Emery, X., 2007a. Using the Gibbs sampler for conditional simulation of Gaussian-based random fields. Computers & Geosciences 33 (4), 522–537.

Emery, X., 2007b. Conditioning simulations of Gaussian random fields by ordinary kriging. Mathematical Geology 39 (6), 607–623.

Emery, X., Lantuéjoul, C., 2006. TBSIM: a computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method. Computers & Geosciences 32 (10), 1615–1628.

Freulon, X., de Fouquet, C., 1991. Remarques sur la pratique des bandes tournantes à trois dimensions (Remarks on the practice of turning bands in three dimensions). In: Cahiers de Géostatistique Fascicule 1, Centre de Géostatistique, Ecole des Mines de Paris. Fontainebleau, France, pp. 101–117.

Gneiting, T., 1999. The correlation bias for two-dimensional simulations by turning bands. Mathematical Geology 31 (2), 195–211.

Gómez-Hernández, J.J., Cassiraga, E.F., 1994. Theory and practice of sequential simulation. In: Armstrong, M., Dowd, P.A. (Eds.), Geostatistical Simulations. Kluwer Academic, Dordrecht, pp. 111–124.

Gómez-Hernández, J.J., Journel, A.G., 1993. Joint sequential simulation of multigaussian fields. In: Soares, A. (Ed.), Geostatistics Tróia' 92. Kluwer Academic, Dordrecht, pp. 85–94.

Goovaerts, P., 1993. Spatial orthogonality of the principal components computed from coregionalized variables. Mathematical Geology 25 (3), 281–302.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, 480pp.

Grzebyk, M., Wackernagel, H., 1994. Multivariate analysis and spatial/temporal scales: real and complex models. In: Proceedings of the XVIIth International Biometrics Conference, International Biometrics Society, Hamilton, Ontario, Canada, pp. 19–33.

Gutjahr, A., Bullard, B., Hatch, S., 1997. General joint conditional simulation using a fast Fourier transform method. Mathematical Geology 29 (3), 361–389.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London, 600pp.

Lantuéjoul, C., 1994. Non conditional simulation of stationary isotropic multigaussian random functions. In: Armstrong, M., Dowd, P.A. (Eds.), Geostatistical Simulations. Kluwer Academic, Dordrecht, pp. 147–177.

Lantuéjoul, C., 2002. Geostatistical Simulation, Models and Algorithms. Springer, Berlin, 256pp.

Mantoglou, A., 1987. Digital simulation of multivariate two- and three-dimensional stochastic processes with a spectral turning bands method. Mathematical Geology 19 (2), 129–149.

Matheron, G., 1973. The intrinsic random functions and their applications. Advances in Applied Probability 5, 439–468.

Myers, D.E., 1989. Vector conditional simulation. In: Armstrong, M. (Ed.), Geostatistics. Kluwer Academic, Dordrecht, pp. 283–293.

Oliver, D.S., 2003. Gaussian cosimulation: modelling of the cross covariance. Mathematical Geology 35 (6), 681–698.

Pardo-Igúzquiza, E., Chica-Olmo, M., 1993. The Fourier integral method—an efficient spectral method for simulation of random fields. Mathematical Geology 25 (2), 177–217.

Pardo-Igúzquiza, E., Chica-Olmo, M., 1994. Spectral simulation of multivariable stationary random functions using covariance Fourier transforms. Mathematical Geology 26 (3), 277–299.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the Gstat package. Computers & Geosciences 30 (7), 683–691.

Rivoirard, J., 1987. Two key parameters when choosing the kriging neighborhood. Mathematical Geology 19 (8), 851–856.

Rivoirard, J., 2004. On some simplifications of cokriging neighborhood. Mathematical Geology 36 (8), 899–915.

Robin, M.J.L., Gutjahr, A.L., Sudicky, E.A., Wilson, J.L., 1993. Cross-correlated random field generation with the direct Fourier transform method. Water Resources Research 29 (7), 2385–2397.

Shinozuka, M., Jan, C.M., 1972. Digital simulation of random processes and its applications. Journal of Sound and Vibrations 25 (1), 111–128.

Switzer, P., 1977. Estimation of spatial distributions from point sources with application to air pollution measurement. Bulletin of the International Statistical Institute 47 (2), 123–137.

Tompson, A.F.B., Ababou, R., Gelhar, L.W., 1989. Implementation of the three-dimensional turning bands random field generator. Water Resources Research 25 (8), 2227–2243.

Tran, T.T., 1994. Improving variogram reproduction on dense simulation grids. Computers & Geosciences 20 (7–8), 1161–1168.

Wackernagel, H., 2003. Multivariate Geostatistics—An Introduction with Applications. Springer, Berlin, 387pp.