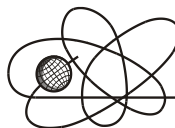




*Российская Академия Наук*

РОССИЙСКАЯ АКАДЕМИЯ НАУК

**ИНСТИТУТ ПРОБЛЕМ  
БЕЗОПАСНОГО РАЗВИТИЯ  
АТОМНОЙ ЭНЕРГЕТИКИ**



**ИБРАЭ**

RUSSIAN ACADEMY OF SCIENCES

**NUCLEAR SAFETY  
INSTITUTE**

Препринт ИБРАЭ № ИБРАЭ-2001-04

Preprint IBRAE-2001-04

**R. Parkin, M. Kanevski, M. Maignan, G. Raspa, E. Savelieva**

**MULTIVARIATE GEOSTATISTICAL MAPPING OF  
CONTAMINATION IN GENEVA LAKE SEDIMENTS.  
CASE STUDY WITH MULTIGEO**

Москва  
2001

Moscow  
2001

УДК 502.3

Паркин Р., Каневский М., Майгнан М., Распа Дж., Савельева Е.

МНОГОПЕРЕМЕННОЕ ГЕОСТАТИСТИЧЕСКОЕ КАРТИРОВАНИЕ ЗАГРЯЗНЕНИЙ В ДОННЫХ ОТЛОЖЕНИЯХ ЖЕНЕВСКОГО ОЗЕРА. ПРИМЕР ИССЛЕДОВАНИЯ С ПОМОЩЬЮ «MULTIGEO» (на англ.яз.). Препринт № IBRAE-2001-04. Москва: Институт проблем безопасного развития атомной энергетики РАН, 2001, 20 с. — Библиогр.: 5 назв.

#### Аннотация

В работе представлены геостатистические алгоритмы пространственного предсказания концентрации металла (Pb) с применением программного пакета “Multigeo”: обычный кригинг и обычный кокригинг. Иллюстрация методов проведена с использованием концентраций ряда металлов, измеренных в 200 точках Женевского озера, и концентрации свинца (Pb), измеренной в 175 точках Женевского озера. С помощью кросс-валидации было проведено сравнение предсказаний, выполненных как с помощью обычного кригинга, так и обычного кокригинга с различным числом дополнительных переменных (от кокригинга с единственной дополнительной переменной до кокригинга с восемью дополнительными переменными).

©ИБРАЭ РАН, 2000

Parkin R., Kanevski M., Maignan M., Raspa G., Savelieva E.

MULTIVARIATE GEOSTATISTICAL MAPPING OF CONTAMINATION IN GENEVA LAKE SEDIMENTS. CASE STUDY WITH MULTIGEO. Preprint IBRAE-2001-04. Moscow: Nuclear Safety Institute RAS, March 2001, 20 p. — Refs.: 5 items.

#### Abstract

This work presents geostatistical algorithms for the spatial prediction of concentration of metal (Pb) using software “Multigeo”: ordinary kriging and ordinary cokriging. The techniques are illustrated using concentrations of some metals measured at 200 points of Lemman Lake and concentration of metal Pb measured at 175 points of Lemman Lake. Cross-validation is used to compare the prediction performances of ordinary kriging and ordinary cokriging with the various numbers of secondary variables (from cokriging with the only secondary variable to cokriging with the eight secondary variables).

©Nuclear Safety Institute,

# Multivariate geostatistical mapping of contamination in Geneva lake sediments. Case study with Multigeo

*R. Parkin, M. Kanevski, M. Maignan (1), G. Raspa (2), E. Savelieva*

Institute of Nuclear Safety (IBRAE)  
B. Tulsкая 52, 113191 Moscow, Russia  
phone: (095) 955-22-31, fax: (095) 958-11-51, e-mail: [m\\_kanevski@ibrae.ac.ru](mailto:m_kanevski@ibrae.ac.ru)

(1) University of Lausanne, Switzerland

(2) University La Sapienza, Roma, Italy

## Contents

|   |    |
|---|----|
| Contents.....                               | 3  |
| 1 Introduction .....                        | 4  |
| 2 Multivariate Geostatistics .....          | 4  |
| 2.1 Direct and Cross Covariances .....      | 4  |
| 2.1.1 Cross covariance function .....       | 4  |
| 2.1.2 Delay effect .....                    | 5  |
| 2.1.3 Cross variogram .....                 | 5  |
| 2.2 Cokriging.....                          | 6  |
| 2.2.1 Isotopy and heterotopy .....          | 6  |
| 2.2.2 Ordinary cokriging .....              | 6  |
| 2.2.3 Simple cokriging .....                | 8  |
| 2.2.4 Collocated cokriging .....            | 8  |
| 3 The description of program Multigeo ..... | 9  |
| 3.1 Opportunities of program Multigeo.....  | 9  |
| 3.2 Structure .....                         | 9  |
| 3.3 Inputs and Output.....                  | 10 |
| 3.3.1 Files.....                            | 10 |
| 3.3.2 Folders.....                          | 10 |
| 3.3.3 Output.....                           | 10 |
| 3.4 Problems .....                          | 10 |
| 4 Case study.....                           | 11 |
| 4.1 Initial data.....                       | 11 |
| 4.2 Structural analysis.....                | 11 |
| 4.3 Geostatistical estimation.....          | 11 |
| 5 Conclusions .....                         | 12 |
| 6 Acknowledgements.....                     | 12 |
| 7 References .....                          | 13 |
| 8 Appendix .....                            | 14 |

# 1 Introduction

Our preprint concerns the contamination of Geneva Lake sediments. The aim of our work is to improve of quality of contamination prediction on the basis of additional information on variables. If there is excessive information on variables, the question arises: what information to use in variable estimation? The additional information can be useful (improves estimation, reduces uncertainty of estimation, etc...), but also it can be superfluous (does not give anything additional). The information on correlated variables should add something to estimation of a variable but complication of model can reduce this addition. Hence, it is necessary to find out: what quantity of useful additional data is necessary to use in variable estimation to improve it and (or) to reduce its uncertainty.

There are two methodological approaches to spatially distributed data treatment: deterministic (e.g., inverse distance squared, triangulations, splines, multiquadric equations) and statistical (e.g., geostatistics, stochastic simulations, fractal interpolations and modeling). As a matter of fact, some types of models explicitly take into account spatial continuity described by different methods – semivariograms, generalized covariance functions, etc. [1] and others do not take into account spatial continuity explicitly.

Geostatistics deals with the statistical treatment of data. It is assumed that measured data  $Z(\mathbf{u})$  are values of a random field  $Z(\mathbf{u})$ . It should be noted that random fields selecting while modeling of a regionalized variable is a matter of analytical convenience. This procedure does not imply that the phenomenon under study is really random. In order to use geostatistics we have to determine structure (covariance function) of the field by using available data.

The most popular geostatistical model is kriging (cokriging). Kriging (cokriging) belongs to the class of BLUE (best linear unbiased estimator) or BLUP (best linear unbiased predictor) estimators. It means that:

- 1) predicted mean must be equal to the real mean (if it is known the simple kriging (cokriging) is used, unknown - ordinary kriging (cokriging));
- 2) error of estimation is minimal. [2]

Multivariate geostatistics allows using the useful additional data on correlated variables for improving of prediction estimation of a variable and/or reduction of uncertainty of estimation.

This paper presents studying of influence of the number of additional variables and their mutual correlation with the main variable on result. Researches were carried out on the real data on Geneva Lake sediments contamination.

## 2 Multivariate Geostatistics

### 2.1 Direct and Cross Covariances

The cross covariance between two random functions can be computed not only at locations  $\mathbf{x}$  but also for pairs of locations separated by a vector  $\mathbf{h}$ . On the basis of an assumption of joint second-order stationarity a cross covariance function between two random functions is defined which only depends on the separation vector  $\mathbf{h}$ .

The interesting feature of the cross covariance function is that it is generally not an even function, i.e. its values for  $+\mathbf{h}$  and  $-\mathbf{h}$  may be different. This occurs in time series when the effect of one variable on another variable is delayed.

The cross variogram is an even function, defined in the framework of an intrinsic hypothesis. When delay-effects are an important aspect of the coregionalization, the cross variogram is not an appropriate tool to describe the data [4].

#### 2.1.1 Cross covariance function

The direct and cross covariance functions  $C_{ij}(\mathbf{h})$  of a set of  $N$  random functions  $Z_i(\mathbf{x})$  are defined in the framework of a joint order stationarity hypothesis

$$\begin{cases} \mathbb{E}[Z_i(\mathbf{x})] = m_i & \text{for all } \mathbf{x} \in D; i = 1, \dots, N \\ \mathbb{E}[(Z_i(\mathbf{x}) - m_i) \cdot (Z_j(\mathbf{x} + \mathbf{h}) - m_j)] = C_{ij}(\mathbf{h}) & \text{for all } \mathbf{x}, \mathbf{x} + \mathbf{h} \in D; i, j = 1, \dots, N \end{cases}$$

The mean of each variable  $Z_i(\mathbf{x})$  at any point of the domain is equal to a constant  $m_i$ . The covariance of a variable pair depends only on the vector  $\mathbf{h}$  linking a point pair and is invariant for any translation of the point pair in the domain.

A set of cross covariance functions is a positive definite function, i.e. the variance of any linear combination of  $N$  variables at  $n+1$  points with a set of weights  $\omega_\alpha^i$  needs to be positive. For any set of points  $\mathbf{x}_\alpha \in D$  and any set of weights  $\omega_\alpha^i \in \mathbb{R}$

$$\text{var}\left(\sum_{i=1}^N \sum_{\alpha=0}^n \omega_\alpha^i Z_i(\mathbf{x}_\alpha)\right) = \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=0}^n \sum_{\beta=0}^n \omega_\alpha^i \omega_\beta^j C_{ij}(\mathbf{x}_\alpha - \mathbf{x}_\beta) \geq 0$$

### 2.1.2 Delay effect

The cross covariance function is not a priori an even or an odd function. Generally for  $i \neq j$ , a change in the order of the variables or a change in the sign of the separation vector  $\mathbf{h}$  changes the value of the cross covariance function

$$C_{ij}(\mathbf{h}) \neq C_{ji}(\mathbf{h}) \quad \text{and} \quad C_{ij}(-\mathbf{h}) \neq C_{ij}(\mathbf{h})$$

If both the sequence and the sign are changed, we are back to the same value

$$C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h})$$

In particular, the maximum of the cross covariance function (assuming positive correlation between a given variable pair) may be shifted away from the origin of the abscissa in a certain direction by a vector  $\mathbf{r}$ . This shift of the maximal correlation is frequent with time series, where one variable can be an effect on another variable, which is not instantaneous. The time for the second variable to react to fluctuations of the first variable causes a *delay* in the correlation between the time series [4].

### 2.1.3 Cross variogram

The direct and cross variogram  $\gamma_{ij}(\mathbf{h})$  are defined in the context of a joint intrinsic hypothesis for  $N$  random functions, when for any  $\mathbf{x}, \mathbf{x} + \mathbf{h} \in D$  and all pairs  $i, j = 1, \dots, N$

$$\begin{cases} \mathbb{E}[Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})] = 0 \\ \text{cov}[(Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})), (Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x}))] = 2\gamma_{ij}(\mathbf{h}) \end{cases}$$

The *cross variogram* is thus defined as half the expectation of the product of the increments of two variables

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2} \mathbb{E}[(Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})) \cdot (Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x}))]$$

The cross variogram is obviously an even function and it satisfies the following inequality

$$\gamma_{ii}(\mathbf{h})\gamma_{jj}(\mathbf{h}) \geq |\gamma_{ij}(\mathbf{h})|^2$$

because the square of the covariance of increments from two variables is bounded by the product of the corresponding increment variances. Actually a matrix  $\mathbf{\Gamma}(\mathbf{h}_0)$  of direct and cross variogram values is a positive semi-definite matrix for any fixed  $\mathbf{h}_0$  because it is a variance-covariance matrix of increments.

It is appealing to investigate its relation to the cross covariance function in the framework of joint second-order stationarity. The following formula is easily obtained

$$\gamma_{ij}(\mathbf{h}) = C_{ij}(\mathbf{0}) - \frac{1}{2}(C_{ij}(-\mathbf{h}) + C_{ij}(\mathbf{h}))$$

which shows that the cross variogram takes the average of the values for  $-\mathbf{h}$  and for  $+\mathbf{h}$  of the corresponding cross covariance function. Decomposing the cross covariance function into an even and an odd function

$$C_{ij}(\mathbf{h}) = \underbrace{\frac{1}{2}(C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h}))}_{\text{even term}} + \underbrace{\frac{1}{2}(C_{ij}(\mathbf{h}) - C_{ij}(-\mathbf{h}))}_{\text{odd term}}$$

we see that the cross variogram only recovers the even term of the cross covariance function. It is not adequate for modeling data in which the odd term of the cross covariance function plays a significant role [4].

## 2.2 Cokriging

The cokriging procedure is a natural extension of kriging when a multivariate variogram or covariance model and multivariate data are available. A variable of interest is cokriged at a specific location from data about itself and about auxiliary variables in the neighborhood. The data set may not cover all variables at all sample locations. Ordinary cokriging requires at least one data value about the variable of interest, while simple cokriging, relying on its knowledge of the mean, can be performed with data solely about the auxiliary variables. At the other end, if all variables have been measured at all sample locations and if the variables are intrinsically correlated, then cokriging is equivalent to kriging.

### 2.2.1 Isotopy and heterotopy

The measurements available for different variables  $Z_i(\mathbf{x})$  in a given domain may be located either at the same sample points or at different points for each variable. The following situations can be distinguished

- *complete heterotopy*: the variables have been measured on different sets of sample points and have no sample locations in common;
- *partial heterotopy*: some variables share some sample locations;
- *isotopy*: data is available for each variable at all sampling points.

Complete heterotopy poses a problem for inferring the cross variogram or covariance model. Experimental cross variograms cannot be computed for completely heterotopic data. Experimental cross covariances, though they can be computed, are still problematic as they do not refer to the same points (and sometimes subregions) as the corresponding direct covariance values.

With partial heterotopy it is advisable, whenever possible, to infer the cross variogram or covariance function model on the basis of the isotopic subset of the data.

A particular case of partial heterotopy important for cokriging is when the set of sample points of the variable of interest is included in the sets of sample points of other variables, which serve as auxiliary variables in the estimation procedure. In this case, when the auxiliary variables are available at more points than the main variable, cokriging is generally of advantage [4].

### 2.2.2 Ordinary cokriging

The ordinary cokriging estimator is a linear combination of weights  $\omega_\alpha^i$  with data from different variables located at sample points in the neighborhood of a point  $\mathbf{x}_0$ . Each variable is defined on a set of samples of (possibly different) size  $n_i$  and the estimator is defined as

$$Z_{i_0}^*(\mathbf{x}_0) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \omega_\alpha^i Z_i(\mathbf{x}_\alpha)$$

where the index  $i_0$  refers to a particular variable of the set of  $N$  variables. The number of samples  $n_i$  depends upon the index  $i$  of the variables, so as to include into the notation the possibility of heterotopic data.

In the framework of a joint intrinsic hypothesis we wish to estimate a particular variable of a set of  $N$  variables on the basis of an estimation error, which should be nil on average. This condition is satisfied by choosing weights, which sum up to one for the variable of interest and which have a zero sum for the auxiliary variables

$$\sum_{\alpha=1}^{n_i} \omega_\alpha^i = \delta_{i_0} = \begin{cases} 1 & \text{if } i = i_0 \\ 0 & \text{otherwise} \end{cases}$$

Expanding the expression for the average estimation error, we get

$$\begin{aligned} \mathbb{E}[Z_{i_0}^*(\mathbf{x}_0) - Z_{i_0}(\mathbf{x}_0)] &= \mathbb{E}\left[ \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \omega_\alpha^i Z_i(\mathbf{x}_\alpha) - \underbrace{\sum_{\alpha=1}^{n_{i_0}} \omega_\alpha^{i_0} Z_{i_0}(\mathbf{x}_0)}_1 - \sum_{\substack{i=0 \\ i \neq i_0}}^N \underbrace{\sum_{\alpha=1}^{n_i} \omega_\alpha^i Z_i(\mathbf{x}_0)}_0 \right] = \\ &= \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \omega_\alpha^i \underbrace{\mathbb{E}[Z_i(\mathbf{x}_\alpha) - Z_i(\mathbf{x}_0)]}_0 = 0 \end{aligned}$$

For the variance of the estimation error we thus have

$$\sigma_E^2 = \mathbb{E}\left[ \left( \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \omega_\alpha^i Z_i(\mathbf{x}_\alpha) - Z_{i_0}(\mathbf{x}_0) \right)^2 \right]$$

Introducing weights  $\omega_0^i$  defined as

$$\omega_0^i = -\delta_{i i_0} = \begin{cases} -1 & \text{if } i = i_0, \\ 0 & \text{if } i \neq i_0 \end{cases}$$

which are included into the sums, we can shorten the expression of the estimation variance to

$$\sigma_E^2 = \mathbb{E}\left[ \left( \sum_{i=1}^N \sum_{\alpha=0}^{n_i} \omega_\alpha^i Z_i(\mathbf{x}_\alpha) \right)^2 \right]$$

Then, inserting fictitious random variables  $Z_i(\mathbf{0})$  positioned arbitrarily at the origin, increments can be formed

$$\sigma_E^2 = \mathbb{E}\left[ \left( \sum_{i=1}^N \left( \sum_{\alpha=0}^{n_i} \omega_\alpha^i Z_i(\mathbf{x}_\alpha) - Z_i(\mathbf{0}) \underbrace{\sum_{\alpha=0}^{n_i} \omega_\alpha^i}_0 \right) \right)^2 \right] = \mathbb{E}\left[ \left( \sum_{i=1}^N \sum_{\alpha=0}^{n_i} \omega_\alpha^i \underbrace{(Z_i(\mathbf{x}_\alpha) - Z_i(\mathbf{0}))}_{\text{increments}} \right)^2 \right]$$

Defining cross covariances of increments  $C_{ij}^l(\mathbf{x}_\alpha, \mathbf{x}_\beta)$  (which are not translation invariant) we have

$$\sigma_E^2 = \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=0}^{n_i} \sum_{\beta=0}^{n_j} \omega_\alpha^i \omega_\beta^j C_{ij}^l(\mathbf{x}_\alpha, \mathbf{x}_\beta)$$

In order to convert the increment covariances to variograms, the additional assumption has to be made that the cross covariances of increments are symmetric. With this hypothesis we obtain the translation invariant quantity

$$\sigma_E^2 = 2 \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \omega_\alpha^i \gamma_{i i_0}(\mathbf{x}_\alpha - \mathbf{x}_0) - \gamma_{i_0 i_0}(\mathbf{x}_0 - \mathbf{x}_0) - \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \omega_\alpha^i \omega_\beta^j \gamma_{ij}(\mathbf{x}_\alpha - \mathbf{x}_\beta)$$

After a minimization in which the constraints on the weights generate  $N$  parameters of Lagrange  $\mu_i$ , we have the ordinary cokriging system

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \omega_\beta^j \gamma_{ij}(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \mu_i = \gamma_{i i_0}(\mathbf{x}_\alpha - \mathbf{x}_0) & \text{for } i = 1, \dots, N; \alpha = 1, \dots, n_i \\ \sum_{\beta=1}^{n_i} \omega_\beta^i = \delta_{i i_0} & \text{for } i = 1, \dots, N \end{cases}$$

and the cokriging variance

$$\sigma_{CK}^2 = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \omega_\alpha^i \gamma_{i i_0}(\mathbf{x}_\alpha - \mathbf{x}_0) + \mu_{i_0} - \gamma_{i_0 i_0}(\mathbf{x}_0 - \mathbf{x}_0)$$

### 2.2.3 Simple cokriging

Ordinary cokriging has no meaning when no data is available for the variable of interest in a given neighborhood. On the other hand, simple kriging leans on the knowledge of the means of the variables, so that an estimation of a variable can be calibrated without having any data value for this variable in the cokriging neighborhood.

The simple cokriging estimator is made up of the mean of the variable of interest plus a linear combination of weights  $\omega_\alpha^i$  with the residuals of the variables

$$Z_{i_0}^*(\mathbf{x}_0) = m_{i_0} + \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \omega_\alpha^i (Z_i(\mathbf{x}_\alpha) - m_i)$$

To this estimator we can associate a simple cokriging system, written in matrix form

$$\begin{pmatrix} \mathbf{C}_{11} & \cdots & \mathbf{C}_{1j} & \cdots & \mathbf{C}_{1N} \\ \vdots & \ddots & & & \vdots \\ \mathbf{C}_{i1} & & \mathbf{C}_{ii} & & \mathbf{C}_{iN} \\ \vdots & & & \ddots & \vdots \\ \mathbf{C}_{N1} & \cdots & \mathbf{C}_{Nj} & \cdots & \mathbf{C}_{NN} \end{pmatrix} \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_i \\ \vdots \\ \mathbf{w}_N \end{pmatrix} = \begin{pmatrix} \mathbf{c}_{1i_0} \\ \vdots \\ \mathbf{c}_{ii_0} \\ \vdots \\ \mathbf{c}_{Ni_0} \end{pmatrix}$$

where the left hand side matrix is built up with square symmetric  $n_i \times n_i$  blocs  $\mathbf{C}_{ii}$  on the diagonal and with rectangular  $n_i \times n_j$  blocs  $\mathbf{C}_{ij}$  off the diagonal with

$$\mathbf{C}_{ij} = \mathbf{C}_{ij}^T$$

The blocks  $\mathbf{C}_{ij}$  contain the covariances between sample points for a fixed pair of variables. The vectors  $\mathbf{c}_{ii_0}$  list the covariances with the variable of interest, for a specific variable of the set, between sample points and the estimation location. The vectors  $\mathbf{w}_i$  represent the weights attached to the samples of a fixed variable [4].

### 2.2.4 Collocated cokriging

A particular heterotopic situation encountered in practice is when we have a variable of interest  $Z(\mathbf{x})$  known at a few points and an auxiliary variable  $S(\mathbf{x})$  known everywhere in the domain (or at least at all nodes of a given estimation grid).

XU ET AL. [3] called *collocated cokriging* a neighborhood definition strategy in which the neighborhood of the auxiliary variable is arbitrarily reduced to only one point: the estimation location  $\mathbf{x}_0$ . The simple cokriging system for such a neighborhood (using  $n$  data values for the primary variable) is written

$$\begin{pmatrix} \mathbf{C}_{ZZ} & \mathbf{c}_{ZS} \\ \mathbf{c}_{ZS}^T & \sigma_{SS} \end{pmatrix} \begin{pmatrix} \mathbf{w}_Z \\ \omega_S \end{pmatrix} = \begin{pmatrix} \mathbf{c}_{ZZ} \\ \sigma_{ZS} \end{pmatrix}$$

where  $\mathbf{C}_{ZZ}$  is the left hand matrix of the simple kriging system of  $Z(\mathbf{x})$  and  $\mathbf{c}_{ZZ}$  is the corresponding right hand side. The vector  $\mathbf{c}_{ZS}$  contains the cross covariances between the  $n$  data points on  $Z(\mathbf{x})$  and the point  $\mathbf{x}_0$  where the data value of  $S(\mathbf{x})$  is located that is included into the estimator.

In the case of intrinsic correlation the collocated cokriging can be expressed as

$$\begin{pmatrix} \sigma_{ZZ} \mathbf{R} & \sigma_{ZS} \mathbf{r}_0 \\ \sigma_{ZS} \mathbf{r}_0^T & \sigma_{SS} \end{pmatrix} \begin{pmatrix} \mathbf{w}_Z \\ \omega_S \end{pmatrix} = \begin{pmatrix} \sigma_{ZZ} \mathbf{r}_0 \\ \sigma_{ZS} \end{pmatrix}$$

To perform collocated cokriging with intrinsic correlation of the variable pair we only need to know the covariance function

$$C_{ZZ}(\mathbf{h}) = \sigma_{ZZ} \rho(\mathbf{h})$$

as well as the variance  $\sigma_{SS}$  of  $S(\mathbf{x})$  and the covariance  $\sigma_{ZS}$  between  $Z(\mathbf{x})$  and  $S(\mathbf{x})$ . [4]



### 3 The description of program Multigeo

MULTIGEO Software is a Geostatistical program that realizes the principal functions of Multivariate Linear Geostatistics [5].

#### 3.1 Opportunities of program Multigeo

The software acquires an sample/variable data table without limits to their number and permits the user to operate, on the subsets created by himself: variographic analysis, fitting a linear model of coregionalization, regionalized Component Principal Analysis, variable estimation, estimation and filtering of variable components, estimation and filtering of regionalized factors.

The implemented methodologies of the estimation are: *kriging*, *cokriging*, *collocated cokriging* and *factorial kriging*.

A cross-validation function, based on the principle “*leave one out*” permits the user to check the performance of a model using the statistical comparing of the estimation results and measured values.

The possibility to operate on the histograms, scatterplots and maps, dividing the data set in different subset, is convenient for the study of the phenomenon variability.

The software also consents the user to create new variables. There are three possibilities: numerical function of original variables, indicative (Boolean) variables and numerical attributes extracted from a georeferenced grid in correspondence of sample location.

The set of operations realized on a data subset constitutes a job *Job*. The intermediate and final results done in a *job* are memorized in a Data Base: the memorized objects are the data subset, coregionalization model and the estimations. The user can inspect, in any moment, the Data Base deleting some objects.

The initial data table during the job can be added by new variables (new columns in the data table): these are constituted, further than new variables, from the cross-validation and from the estimation or filtering of spatial components or factors, carried out on the sampled points [5].

#### 3.2 Structure

The software realizes its performances through eight functions that constitute the principal menu. In following is the list of these functions with a brief description.

##### FILE

Thorough this function is realized the communication between Multigeo and external part of the user: it opens a existing job, creates a new job, imports and exports a job, prints the output graphics, imports and exports a grid of values, exports subset data extracted from the current data table.

##### NEW VARIABLE

Thorough this function can be create three typology of new variables: indicative variables, numerical function variables, extracted variables from a regular grid of values.

##### DESCRIPTION & STATISTICS

It consents the user to editing the sample map and to carry out the histograms and scatterplots.

##### COREGIONALIZATION MODEL

It consents the user to compute experimental variograms and to fit automatically a linear model of coregionalization, after some interactive adjusting in order to explore the spatial structures.

##### REGIONALIZED PCA

It consents the user to carry out the regionalized Principal Component Analysis, i.e. the PCA on different spatial scale. It is realized, further than on the synthetic description of the links among the variables in study, on the realization of the factorial kriging.

##### ESTIMATE

Thorough this function the user operates the different estimation patterns: on grid nodes, on regular pixels, on random points, on sampled points.

##### CROSS-VALIDATION

This function carry out the estimation of variables in each sampled point without using the corresponding measure in the estimation [5].

### 3.3 Inputs and Outputs

#### 3.3.1 Files

Multigeo changes information with external part in files ASCII format. The changing information in input and in output is at most tables, grid of values and images. The first ones are files in GEOEAS format, the second ones are files SURFER format. The imagines, that normally content the elaborated graphical output, are exported as files in BMP format. One-third type of information, changed at internal of the ESTIMATE function, consists in a polygon series. The corresponding files are in BLN format, used by SURFER.

#### 3.3.2 Folders

The set of files that compose MULTIGEO software can be memorized in any folder. During the installation the software creates a folder JOBS and inside of it will create as many folders as are the jobs realized. The name of created folder corresponds with the job name used by user.

#### 3.3.3 Output

The graphical output realized during the execution of the functions can be printed or memorized as files in the GRAPH\_OUT folder created automatically by the software inside of each job-folder. [5]

### 3.4 Problems

Multigeo does not produce correlation coefficient variables and presents only graphic image of correlation. Program provides only the correlation matrix for various theoretical models of variables (nugget, spherical and i.e.).

The program does not allow considering monitoring networks.

In general Multigeo allow only carrying out the variables variogram analysis, the regionalized Principal Component Analysis (PCA) and the various estimation methodologies: *kriging*, *cokriging*, *collocated cokriging* and *factorial kriging*.

## 4 Case study

### 4.1 Initial data

The methods described above were used for the analysis of the spatial data on contamination by metals in Geneva Lake (Leman) sediments. From a set, which contained 32 names of metals, 9 (Pb, Cu, Cd, Zn, Cr, Ni, Be, B, Mn) metals were chosen.

Leman is situated in the southwest part of Switzerland. The amount of measurements of Pb was less than amount of measurements of other metals. Figure 1 shows the location of 175 samples of Pb concentration used in this study for ordinary kriging and cokriging. Figure 2 shows the location of 200 samples of concentration of other metals used in this study for cokriging. The technique of cokriging has allowed to use additional measurements of other metals for improvement of an estimation of Pb contamination and/or reduction of uncertainty of Pb contamination estimation.

The distribution of values of metals concentration is shown in figure 3. The values of Pb concentration range from 10 up to 100 and only two points lay in a range from 66 up to 100. The similar situation (presence of outliers) is as well observed in the distribution of values of Cd, Zn, Cu and, particularly, Mn.

The correlation of Pb with other metals is submitted in Figure 4 and Table 1. We see that Pb is linear correlated with Zn, Ni, Cd, Cr, Be, Cu and poorly correlated with B, Mn.

## 4.2 Structural analysis

To investigate spatial correlation between metals and their own spatial autocorrelations the structural analysis was carried out. Variogram and cross-variogram roses have the form of an ellipse and are extended along a direction of 22,5°. For all metals correlation coefficient increases from 15 km in the direction perpendicular to the direction of anisotropy (small axis of the ellipse) to 65 km along the direction of anisotropy (big axis of the ellipse). Both variograms and cross-variograms in all directions and in the direction of 22,5° were constructed.

With the help of Multigeo software [5] the models of variograms and cross-variograms were constructed. The theoretical positively determined functions are used while modeling variograms. One of such functions (model of spherical type) is used in the present work:

$$\gamma(h) = \begin{cases} c_0 + (c - c_0) \left( \frac{1.5h}{a} - \frac{0.5h^3}{a^3} \right) & h \leq a \\ c, & h > a \end{cases},$$

where  $c_0$  is nugget,  $c$  is sill,  $a$  is range.

The share of accident (noncorrelatedness) in spatial distribution of the data is determined by nugget.

As a result of further analysis we receive theoretical models of omnidirectional and in 22,5° direction variograms (Figures 5 and 6) and cross-variograms (Figures 7 and 8) as well, which values are submitted in Table 2. Selection parameters of models were carried out method with the help of least square.

## 4.3 Geostatistical estimation

The models described above (in the direction of 22,5°) were used in cokriging. Cokriging on a rectangular grid 40 on 40 cells with the size of the cell 1.475 on 0.825 km was carried out. Calculations were executed with the help of Multigeo software [5]. The obtained maps of estimations and the variances of Pb estimations are presented in figures 12, 13, 14 and 15.

For comparison of the estimations of ordinary kriging (OK) with ordinary cokriging (made for all variables) on the same grid using the same models of variograms were designed. Maps of estimation and variance of estimation of ordinary kriging (OK) are given in figure 12.

Comparing maps of estimations we see, that ordinary kriging (OK) has overestimated Pb values in areas where the measurements of Pb are not available. In these areas cokriging uses additional measurements of the other metals for estimation. It has allowed to improve the estimation of Pb by cokriging in areas where values of Pb are not present.

The estimation of ordinary kriging is more smeared. It means undervaluing of concentration in points with high pollution and overestimation of concentration in points with low pollution. On the contrary ordinary kriging (OK), cokriging has given considerably less smeared estimation. The supplement of additional variables for Pb estimation with cokriging improves the estimation.

As it was mentioned above, cokriging should reduce estimation variances in comparison with ordinary kriging. The maps of estimation variances given on figures 12 - 15 confirm this: in areas where there Pb measurements are absent, estimation variance of cokriging falls below 100. Which is not observed on the map of estimation variance of ordinary kriging. A nugget in zero of variograms and cross-variograms causes the high value of Pb variance assessment.

However, after the number of particular variables exceeds certain level Pb estimation cannot be improved, and even Pb values assessments are worsened. In the given work Pb estimation is worsened and is more smoothed over in case of adding to the three existing additional variables (Zn, Cu, B) the fourth one.

The analysis of the results of kriging and cokriging with various numbers of additional variables was conducted using the procedure of cross-validation. The results of cross-validation for various methods are presented in tables 3, 4 and 5, and on figures 9, 10 and 11. It is seen from them, that the best results are observed in the procedure of cokriging with the three particular additional variables (Zn, Cu, B).

## 5 Conclusions

In present work the approach to the mutual analysis of several spatially distributed variables is described and applied. The geostatistical model of cokriging is applied to the data on the contamination of Geneva Lake with metals. The values of contamination for different metals are correlated among themselves. To obtain the assessment of Pb contamination the measurements of eight metals (Zn, Cu, Mn, Be, B, Ni, Cr, Cd) were used. The results of estimation are compared with the results obtained by ordinary kriging using only one Pb variable and also by cokriging varying additional variables and their number. According to the results of comparison it is possible to draw the following conclusions:

1. The including of additional variables allows to correct the model of spatial correlation structure of the data and to obtain more the improved assessments in those areas where there are no measurements of the primary variable;
2. Combined consideration of two and more variables essentially improves final estimation of the basic variable in the field of high values and prevents their smearing;
3. Cokriging gives the smaller mistake of estimation and correspondingly smaller uncertainty of estimation in comparison with kriging;
4. Three different additional variables of the eight (Zn, Cu, B) are enough for cokriging to obtain better estimation and estimation variation.

To analyze and visualize of raw data and results Multigeo and GeoStat Office research programs were used.

## 6 Acknowledgements

The work was supported in part by INTAS grants 97-31726 and 99-00099.

## 7 References

1. Goovaerts, P. (1997) «Geostatistics for Natural Resources Evaluation», Oxford University Press.
2. M. Kanevski, R. Arutyunyan, L. Bolshov, V. Demyanov, S. Chernov, E.Savelieva, V. Timonin, M. Maignan, M.F. Maignan, Mapping of Radioactively Contaminated Territories with Geostatistics and Artificial Neural Networks. In Contaminated Forests I. Linkov and W.R. Schell (eds.), pp. 249-256, Kluwer Academic Publishers, 1999, printed in Netherlands.
3. Xu W, Tran TT, Srivastava RM & Journel AG (1992) *Integrating seismic data in reservoir modeling: the collocated cokriging alternative*. Proceedings of 67<sup>th</sup> Annual Technical Conference of the Society of Petroleum Engineers, paper SPE 24742, 833-842, Washington.
4. H. Wackernagel. Multivariate Geostatistics. Springer-Verlag, Berlin, 1995
5. Guiseppa Raspa, Massimiliano Ticci, Ismail Hoxha, Multigeo, User's Guide, Università La Sapienza Dipartimento ICMMPM Materie Prime, September 1999.

## 8 Appendix

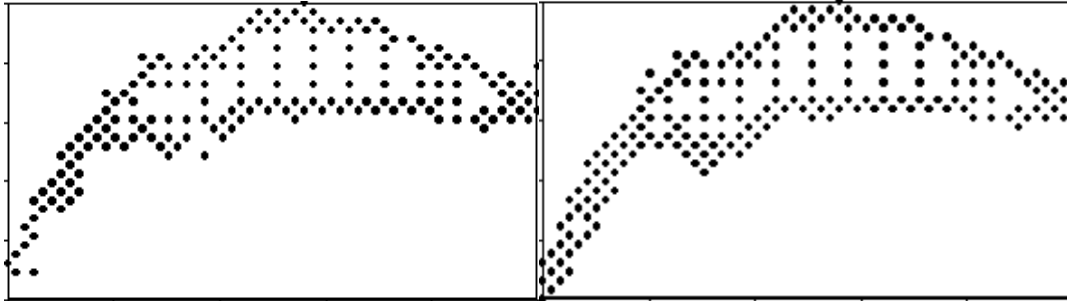


Figure 1. Points where Pb is known

Figure 2. Points where other metals are known

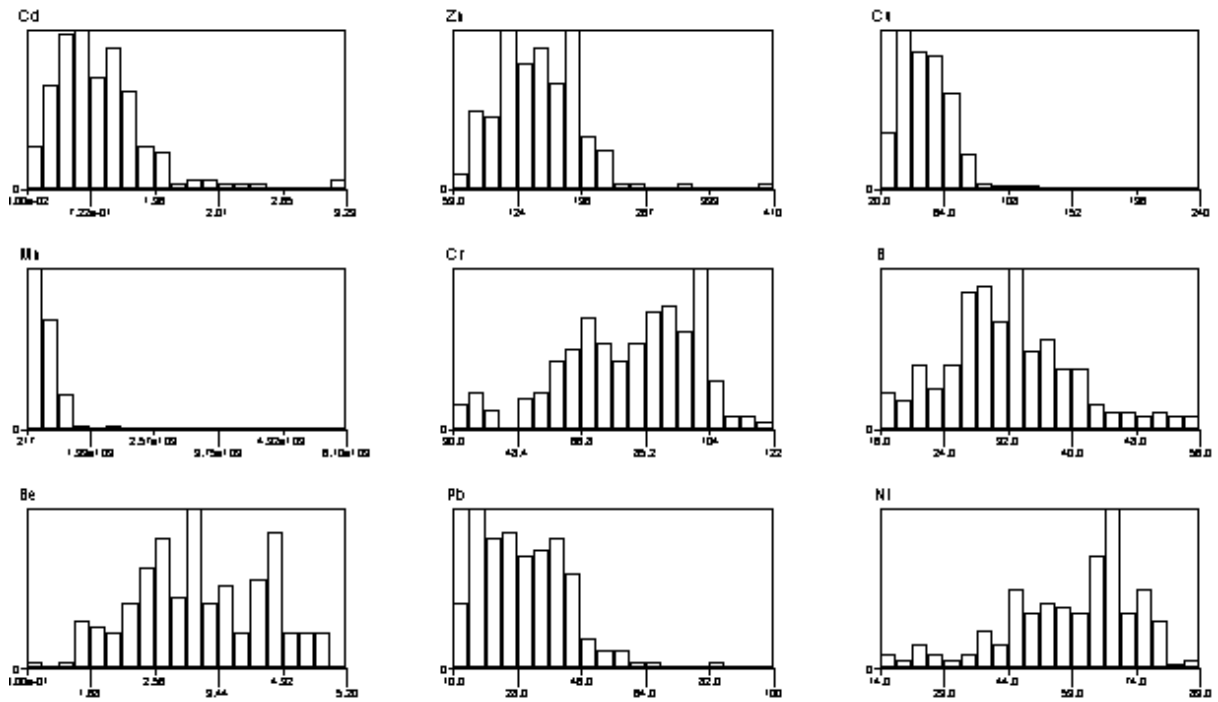


Figure 3. The frequency distribution of all concerned metals

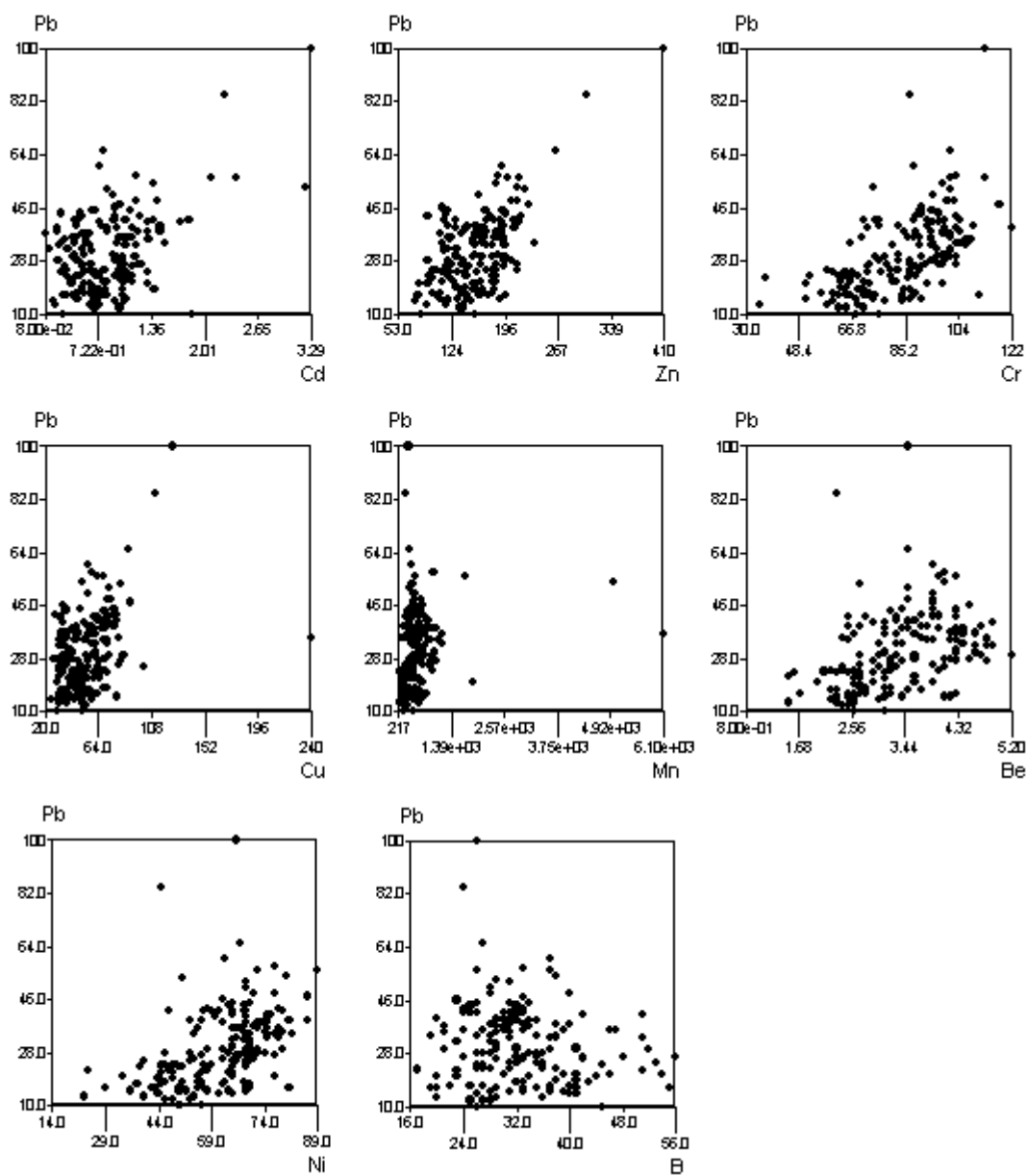


Figure 4. The correlation Pb with all concerned metals

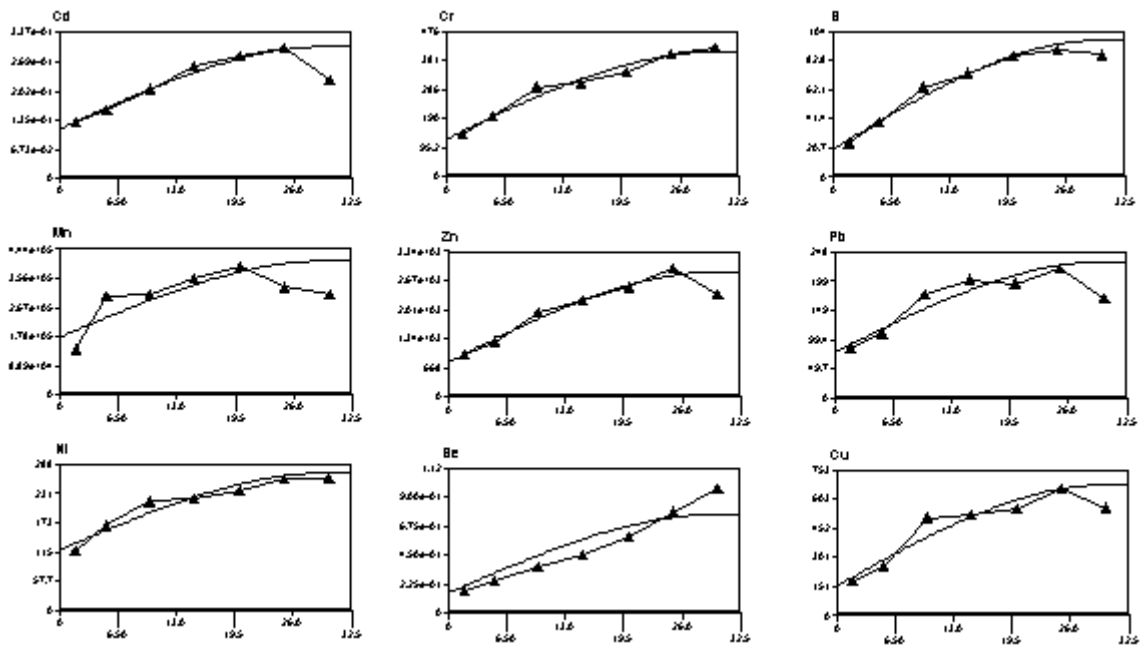


Figure 5. The omnidirectional variograms of all concerned metals

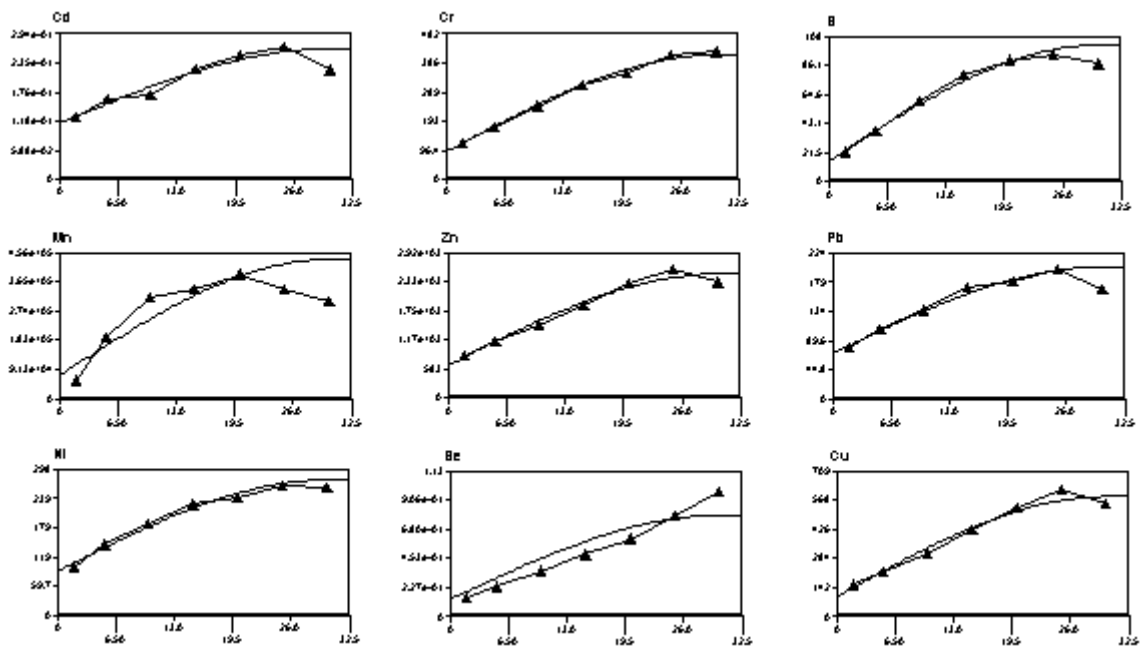


Figure 6. The variograms (in direction 22,5°) of all concerned metals

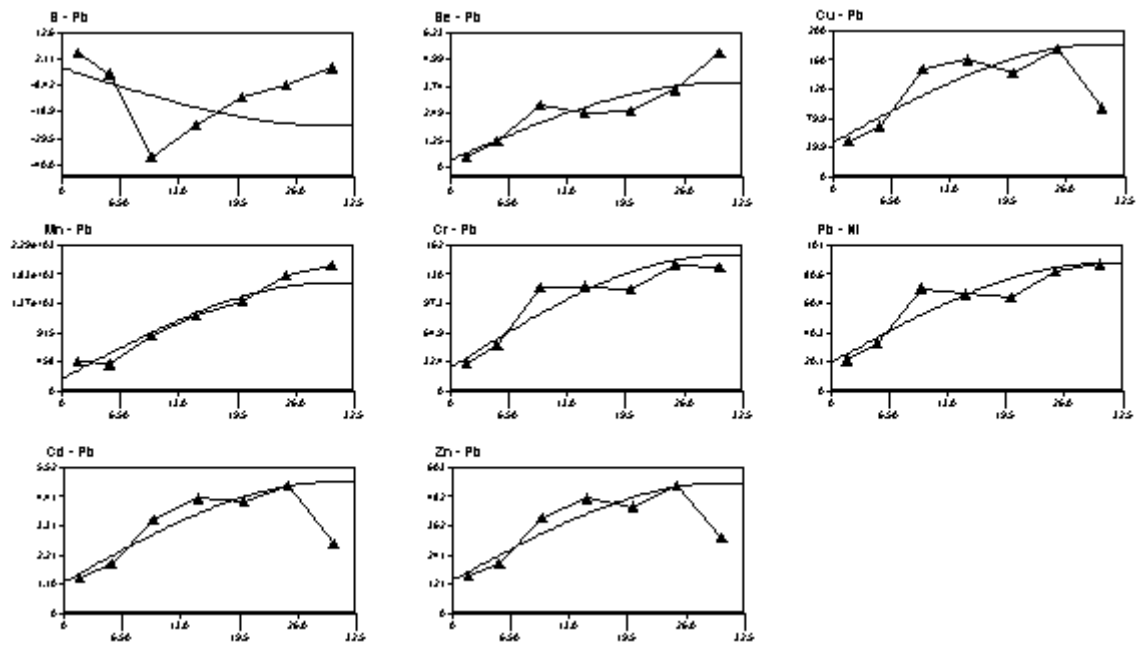


Figure 7. The omnidirectional cross-variograms of all concerned metals with Pb

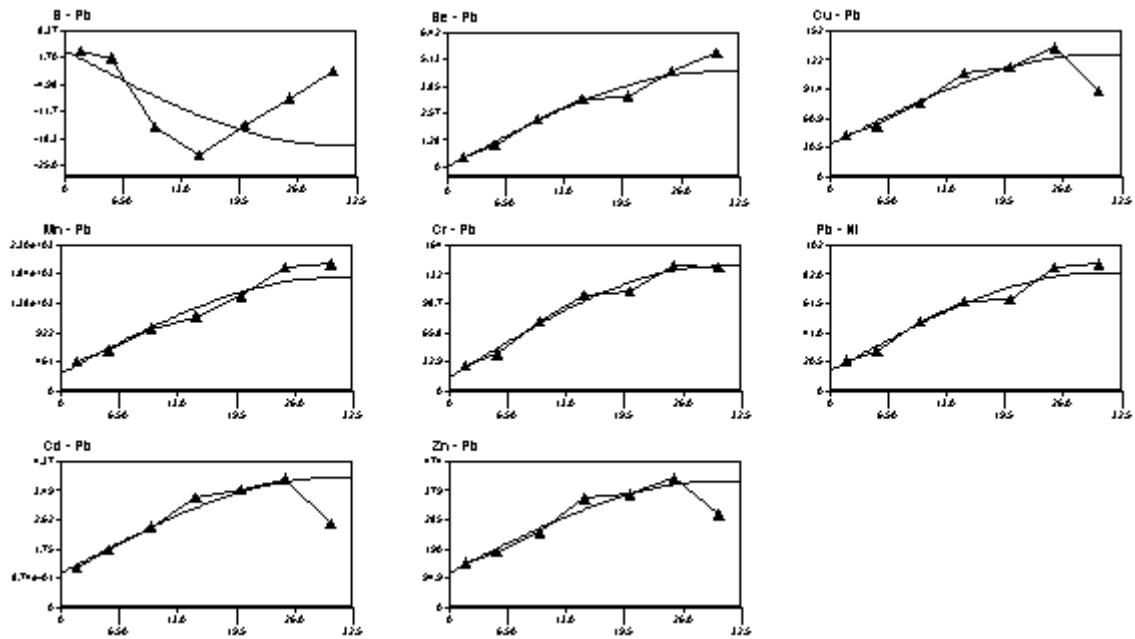


Figure 8. The cross-variograms (in direction 22,5°) of all concerned metals with Pb



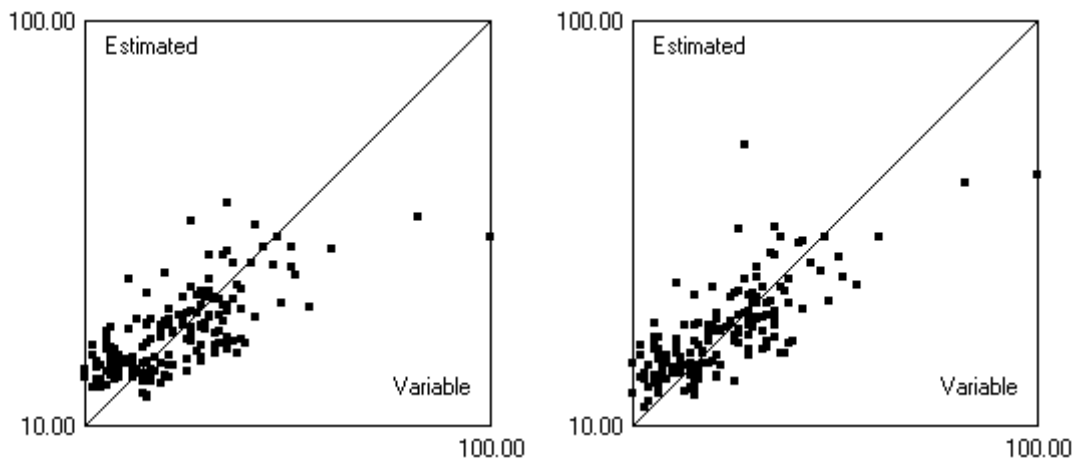


Figure 9. Scatter plot of cross-validation of kriging Pb (left) and cokriging Pb with Cu (right)

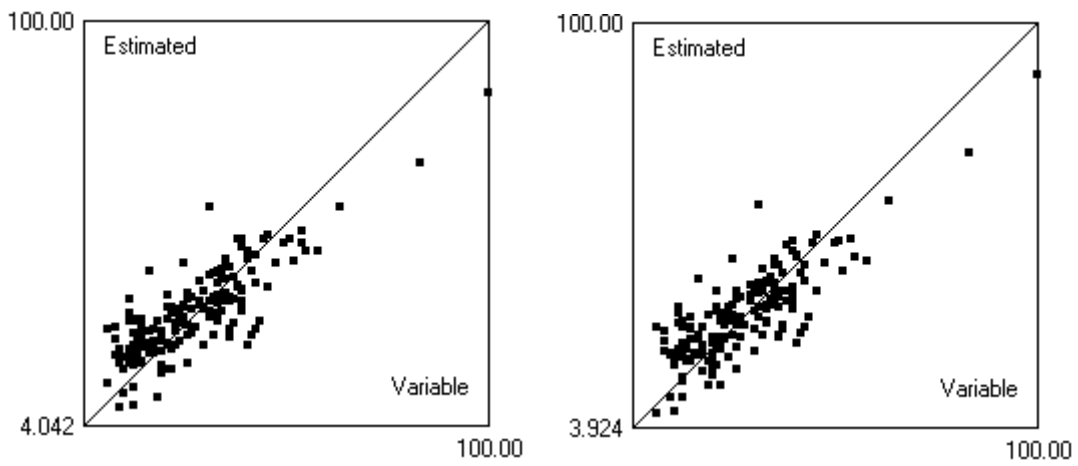


Figure 10. Scatter plot of cross-validation of cokriging Pb with Zn (left) and cokriging Pb with Zn, B (right)

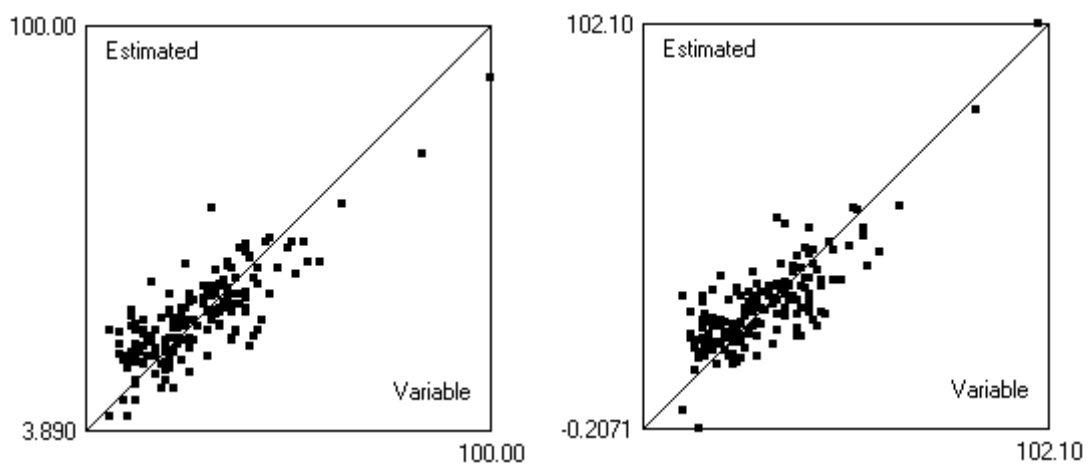


Figure 11. Scatter plot of cross-validation of cokriging Pb with Zn, Cu, B (left) and cokriging Pb with all (right)

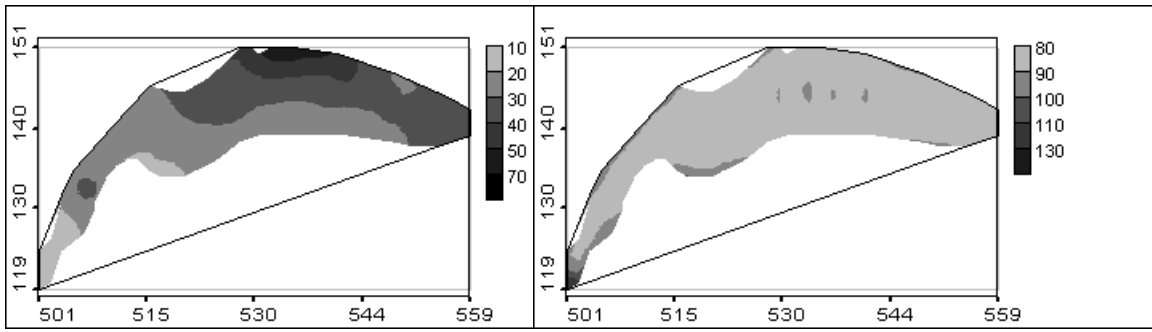


Figure 12. The kriging estimations of Pb (left) and the kriging variations of Pb (right)

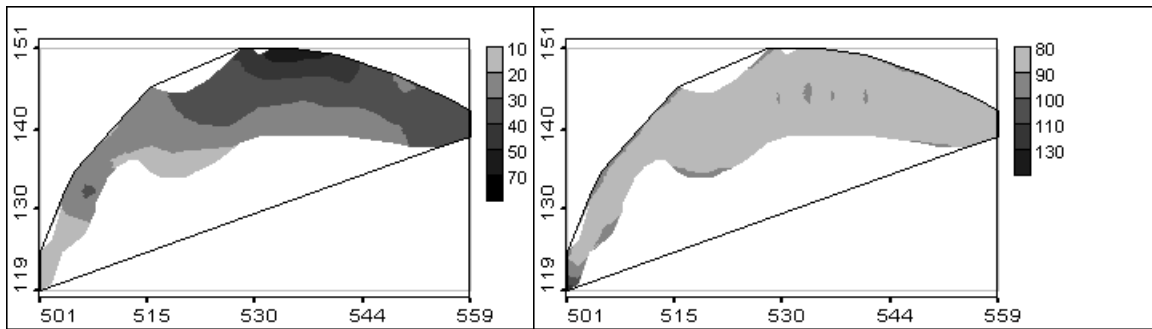


Figure 13. The cokriging estimations of Pb with Zn (left) and the cokriging variations of Pb with Zn (right)

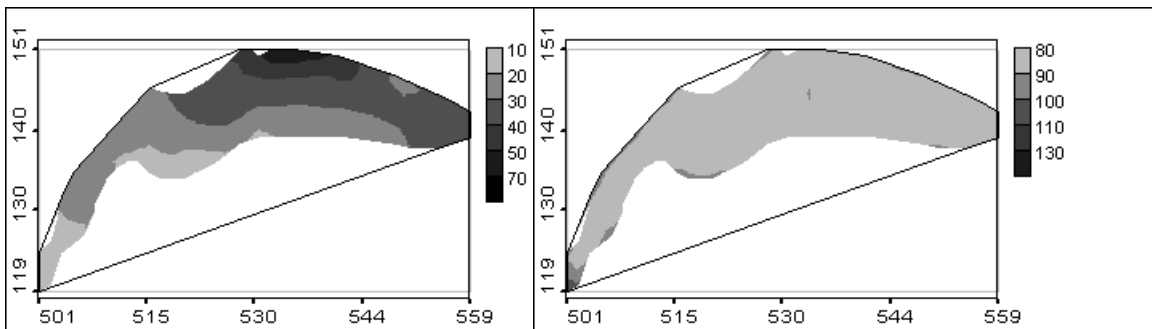


Figure 14. The cokriging estimations of Pb with Zn, Cu, B (left) and the cokriging variations of Pb with Zn, Cu, B (right)

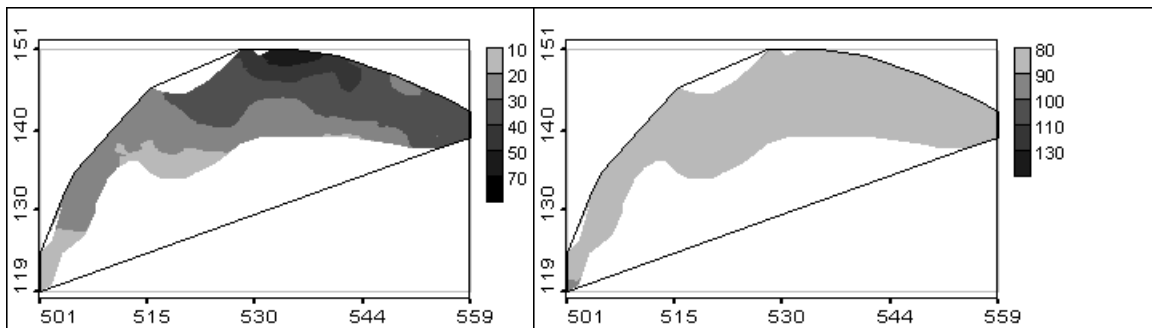


Figure 15. The cokriging estimations of Pb with all variables (left) and the cokriging variations of Pb with all variables (right)

**Table 1. The correlation coefficient of Pb and all concerned metals**

|           |           |           |           |           |           |           |          |           |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------|-----------|
|           | <i>Zn</i> | <i>Cu</i> | <i>Mn</i> | <i>Cd</i> | <i>Ni</i> | <i>Be</i> | <i>B</i> | <i>Cr</i> |
| <i>Pb</i> | 0.616     | 0.378     | 0.177     | 0.463     | 0.435     | 0.379     | -0.126   | 0.557     |

**Table 2. The values of model variograms and cross-variograms (in direction 22,5°)**

| <b>Variable</b> | <b>Nugget</b> | <b>Model</b> | <b>Sill</b>  | <b>Radius</b> |
|-----------------|---------------|--------------|--------------|---------------|
| Cd              | 0.1138        | Spherical    | 0.1498       | 30            |
| Zn              | 662.5690      | Spherical    | 1821.3650    | 30            |
| Cu              | 97.6125       | Spherical    | 487.2636     | 30            |
| Mn              | 73418.4844    | Spherical    | 363132.03.12 | 30            |
| Cr              | 91.2362       | Spherical    | 320.0846     | 30            |
| B               | 16.0238       | Spherical    | 85.0293      | 30            |
| Be              | 0.1349        | Spherical    | 0.6537       | 30            |
| Ni              | 89.6864       | Spherical    | 187.0863     | 30            |
| Pb              | 70.2201       | Spherical    | 132.2419     | 30            |
| Pb – Cd         | 1.0079        | Spherical    | 2.8556       | 30            |
| Pb – Zn         | 112.3534      | Spherical    | 298.0497     | 30            |
| Pb – Cu         | 32.9582       | Spherical    | 94.6861      | 30            |
| Pb – Mn         | 273.1706      | Spherical    | 1507.5059    | 30            |
| Pb – Cr         | 14.7250       | Spherical    | 125.8600     | 30            |
| Pb – B          | 3.4318        | Spherical    | -23.2637     | 30            |
| Pb – Be         | 0.0168        | Spherical    | 4.5325       | 30            |
| Pb – Ni         | 13.3944       | Spherical    | 69.4055      | 30            |

**Table 3. The results of cross-validation of Pb using kriging**

| Errors                       | Kriging      |
|------------------------------|--------------|
| Mean error:                  | 0.001515     |
| Mean square error:           | 86.37        |
| Sqrt (mean square error):    | 9.293        |
| Mean standard error:         | -0.000578329 |
| Mean square std. error:      | 0.9661       |
| Sqrt (mean sqr. std. error): | 0.9829       |
| Mean estimation variance:    | 89.37        |
| Sqrt (mean estim. var.):     | 9.454        |

**Table 4. The results of cross-validation of Pb using cokriging one, two and three variables**

| Errors                       | Cokriging<br>Zn | Cokriging<br>Zn Cu | Cokriging<br>Zn B | Cokriging<br>Zn Cu B |
|------------------------------|-----------------|--------------------|-------------------|----------------------|
| Mean error:                  | 0.002741        | -0.02151           | 0.002758          | -0.02968             |
| Mean square error:           | 59.10           | 58.72              | 58.69             | 58.25                |
| Sqrt (mean square error):    | 7.688           | 7.663              | 7.661             | 7.632                |
| Mean standard error:         | 0.02322         | 0.01927            | -0.007022         | -0.007816            |
| Mean square std. error:      | 1.736           | 2.001              | 0.9211            | 0.9403               |
| Sqrt (mean sqr. std. Error): | 1.318           | 1.415              | 0.9597            | 0.9697               |
| Mean estimation variance:    | 112.61          | 111.72             | 132.81            | 131.58               |
| Sqrt (mean estim. var.):     | 10.61           | 10.57              | 11.52             | 11.47                |

**Table 5. The results of cross-validation of Pb at addition of the fourth additional variable and all variables**

| Errors                       | Cokriging<br>Zn Cu B Cd | Cokriging<br>Zn Cu B Mn | Cokriging<br>Zn Cu B Cr |
|------------------------------|-------------------------|-------------------------|-------------------------|
| Mean error:                  | -0.03234                | -0.04132                | -0.09369                |
| Mean square error:           | 58.66                   | 58.74                   | 61.37                   |
| Sqrt (mean square error):    | 7.659                   | 7.664                   | 7.834                   |
| Mean standard error:         | -0.008283               | -0.002159               | -0.004565               |
| Mean square std. error:      | 0.7589                  | 0.7548                  | 0.6727                  |
| Sqrt (mean sqr. std. Error): | 0.8712                  | 0.8688                  | 0.8202                  |
| Mean estimation variance:    | 133.23                  | 136.25                  | 140.36                  |
| Sqrt (mean estim. var.):     | 11.54                   | 11.67                   | 11.85                   |

| Errors                       | Cokriging<br>Zn Cu B Be | Cokriging<br>Zn Cu B Ni | Cokriging<br>All variables |
|------------------------------|-------------------------|-------------------------|----------------------------|
| Mean error:                  | -0.06861                | -0.07455                | -0.01525                   |
| Mean square error:           | 60.63                   | 58.81                   | 60.35                      |
| Sqrt (mean square error):    | 7.786                   | 7.668                   | 7.769                      |
| Mean standard error:         | -0.008059               | -0.1227                 | 0.003977                   |
| Mean square std. error:      | 0.7253                  | 4.690                   | 0.3972                     |
| Sqrt (mean sqr. std. Error): | 0.8517                  | 2.166                   | 0.6303                     |
| Mean estimation variance:    | 129.85                  | 79.12                   | 168.18                     |
| Sqrt (mean estim. var.):     | 11.40                   | 8.895                   | 12.97                      |