MULTIVARIATE ESTIMATION AND SIMULATION FOR ENVIRONMENTAL DATA MODELLING:

PROCESSING OF HEAVY METALS CONCENTRATION DATA IN SOIL

Barbara Namyslowska-Wilczynska [1]* ***and Artur Wilczynski*** [2]

*[*1] Institute of Geotechnics and Hydrotechnics, Wroclaw University of Technology, Poland*
e-mail: wilczynska@i10.igh.pwr.wroc.pl
[2] Institute of Electrical Power Engineering, Wroclaw University of Technology, Poland
e-mail: artwil@see.ie.pwr.wroc.pl

ABSTRACT

Results applying ordinary kriging and cokriging techniques as well as the turning bands simulation method to the survey of heavy metal pollution of the superficial layer (at the depth of 0-20 cm) of soil in a selected mining region of Upper Silesia (S Poland) are presented. The multivariate structural analysis, estimation and conditional simulation was performed on data coming from the regional monitoring of soils.
Based on estimated and the simulated cadmium and zinc heavy metal soil concentrations, the most polluted zones and places in the Dabrowa Górnicza region, where environmental monitoring should be instituted again, were determined.

Keywords: heavy metal, pollution, soil, estimation, kriging, cokriging, conditional simulation

1 INTRODUCTION

The main soil problems in EU countries are irreversible losses because of increasing soil sealing and soil erosion, and continued deterioration due to local contamination and diffuse contamination (acidification and heavy metals) (European Environment Agency, 1999). Local soil contamination by heavy metals, which are especially high in areas with heavy industries and military bases and in high density urban areas within the EU and the Accession countries have been observed. In many cases, the higher values occur because of natural geological or soil-forming conditions. Strategies for soil protection and systems for monitoring soil contamination are not adequately developed at the European or national level, compared with air and water (European Environment Agency, 1999).

The largest areas most affected by contamination are located in north-west Europe (Nord-Pas de Calais in France, Rhein-Ruhr region in Germany, Belgium, Netherlands). Other such regions include the Saar region in Germany, northern Italy and the corner of Poland, the Czech Republic and the Slovak Republic, with Cracow and Katowice at its centre. Information available for 20 European countries reveals that the estimated total of contaminated sites exceeds 1.5 million, most of which are located in the 13 EU Member States (European Environment Agency, 1999). Many Accession Countries have initiated national inventories of contaminated sites and air, water and soil pollution sources. They have also enacted legislation.

A subject of interest to the authors of this paper is the level of heavy metal pollution of the soil surface-layer in selected areas of the Katowice Province in Upper Silesia, which for years has been the most industrialized region in south-western Poland. Upper Silesia with its characteristic lead-zinc-cadmium complex stands out as an anomaly from the general geochemical picture of Poland (Lis & Pasieczna, 1995a; 1995b; 1999). The natural environment there is adversely affected by ore-bearing deposits occurring close to the surface (rock outcrops) in some areas, by the historical and current mining of the zinc-lead deposits (in the Silesian-Cracow region zinc and lead have been mined since the beginning of the 12th century), by the zinc-lead ore processing and metallurgical plants operating there and by heavy motor traffic in the region.

The soil environment in Poland has been monitored since the end of the 80s. Initially a sparse grid covering the whole country was used (Lis & Pasieczna, 1995a). Later a denser grid was used in the most heavily industrialized regions to make maps of areas of high soil contamination (Upper Silesia, Lower Silesia, the Cracow area, and the Walbrzych area). An awareness of the need to create environmental databases and to use the processed data for environmental quality assessments has been growing.

The processing, analysis and estimation of an increasing amount of environmental and environment protection data stored in continuously updated databases requires more and more effective estimation tools.

The authors have tested different spatial statistics methods representing linear geostatistics (variogram, kriging and ordinary cokriging) and nonlinear geostatistics (indicator variogram, indicator kriging and conditional *turning bands* simulation) (Armstrong & Dowd, 1994; Armstrong, 1998; Isaaks & Srivastava, 1989; ISATIS, 2001; Rivoirard, 1994; Wackernagel, 1995). Environmentalists (European Environment Agency, 1999), as well as the present authors, have most often studied the cadmium, lead and zinc concentrations, whose variability in the soil has been estimated for the Bytom, Będzin, Dabrowa Górnicza, Piekary Sl. and Tarnowskie Góry areas (Namyslowska-Wilczynska & Wilczynski, 1997; 1999a; 1999b; 1999c; 2000; 2001). In addition copper and lead concentrations have been determined for the Legnica-Glogów Copper Basin (Namyslowska-Wilczynska & Pyra, 2000). Also potential changes in the heavy metal pollution of soil over time have been investigated (Namyslowska-Wilczynska & Rusak, 1999) by taking advantage of available environmental databases storing data on soil sampling for 6 heavy metals: cadmium, zinc, lead, copper, nickel and chromium (Namyslowska-Wilczynska, 1999; Namyslowska-Wilczynska & Wilczynski, 2001). It has been determined that the samples have identical parameters which indicates isotopy (Wackernagel, 1995).

In unsampled areas the average estimated heavy metal concentration was used for one of the investigated elements, which means that a one-dimensional model was used. The question of whether the estimation's accuracy could be increased if a multivariate estimation model was built arose. In such cases when estimating the average concentration of one metal (the primary variable), the concentration of the second metal (the secondary variable) is taken into account if it is highly correlated with the first metal's concentration (Wackernagel, 1995).

With the aim of producing better pollution modelling results, a multidimensional model of cadmium and zinc soil concentration variability was built and the model parameter values were used in the estimation calculations (original data) and simulations (Gaussian data). The question is which of the obtained pollution pictures will represent the reality more truthfully?

2 CHARACTERIZATION OF STUDY AREA

An area in the neighbourhood of a steel mill (Figure 1) and other smaller industrial plants in Dabrowa Górnicza, where ore-bearing dolomite outcrops occur, was investigated. The Dabrowa Górnicza Municipality is situated within a large tectonic unit - the Silesian-Cracow monocline (Lis & Pasieczna, 1995b; 1999). The ore-bearing dolomites belong to middle Triassic formations which include zinc and lead ores. They are mostly finely-crystalline, porous, often cracked dolomites similar to breccia in make-up. Their pores and caverns are usually filled with zinc and lead sulphides, sphalerite, galena, pyrite, marcasite, calcite, limonite or barite. Geochemical surveys of the soils, including cultivated land, carried out in the Katowice Province showed the cadmium concentration ranging from 0.3 to 143 ppm and the zinc concentration nearly 3 times higher than in other regions of Poland. The zinc concentration varies from 300 to 3300 ppm.

The cadmium and zinc soil pollution in the agricultural area was analyzed. The deposits forming the surface soil layer were clayey (loamy sand and light, medium and heavy clays). The sample size for the 20 cm thick soil surface layer was 152. The samples were taken by the Institute of Environment Research and Control in Katowice. The spatial extent of heavy metal (lead, copper, nickel, chromium) soil contamination in the Dabrowa area was determined by applying the kriging estimation techniques given in (Namyslowska-Wilczynska, 1999; Namyslowska-Wilczynska & Wilczynski, 1999c; 2001).

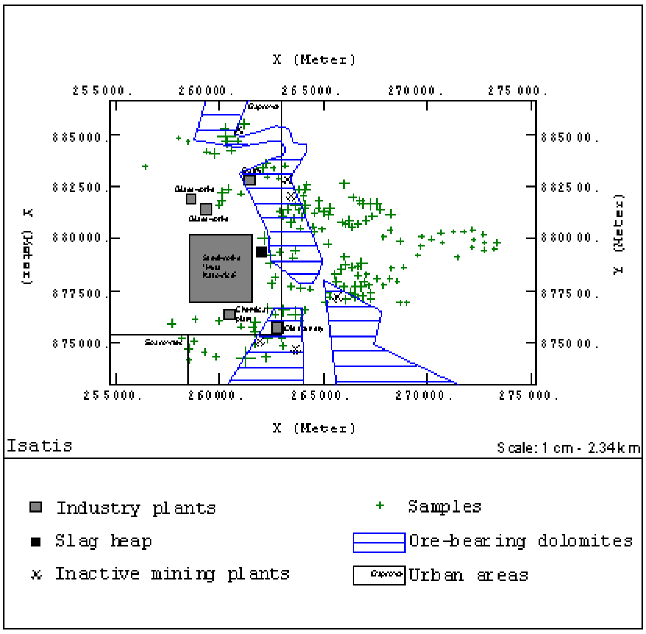

Figure 1. Analysed region of the Dabrowa Górnicza with the soils sampling points marked.

The histograms of the soils' heavy metal concentration distribution more or less exhibit pronounced positive skewness (Namyslowska-Wilczynska, 1999, Table 1). The distribution of the metals is truncated. The cadmium concentration histogram is bimodal while the zinc one is thinner. Both distributions are logarithmically-normal. The average cadmium concentration of 4.12 ppm is only slightly higher than the threshold limit value (TLV) of 3 ppm set by the Institute of Soil-Cultivation, Fertilizing and Pedology in Pulawy (Kabata-Pendias, Motowicka-Terelak, Piotrowska, Terelak & Witek, 1993) but the highest cadmium concentrations (over 17 ppm) exceed the TLV several times. This corresponds to third (average) degree and locally to fifth degree soil contamination. The average zinc concentration (420.58 ppm) exceeds the TLV for this metal in soil (200-300 ppm), which corresponds to second degree soil contamination. But in places the soil's zinc concentration exceeds 3800 ppm, which corresponds to a fifth degree contamination.

Table 1. Basic statistical parameters of the studied heavy metal soil concentrations.

Analyzed Parameter	Count [n]	Minimum [ppm]	Maximum [ppm]	Mean $\bar{x}$ [ppm]	Standard deviation s [ppm]	Variance s^2 $[ppm]^2$	Skewness	Kurtosis	Variation Coef. [%]
cadmium	152	1	17	4.12	2.25	5.05	2.05	11.51	54.63
zinc	152	37	3820	420.58	489.32	239437.43	4.31	25.84	116.34

The variability of zinc concentration was much larger (V_{Zn} ~ 116%) in comparison with cadmium concentration (V_{Cd} ~ 55%) (Table 1).

Table 2. Correlation matrix.

cadmium	1	0.66
zinc	0.66	1

The correlation coefficient *r* for the cadmium and zinc concentration assumes a statistically significant value of 0.66, which indicates a close correlation between the two elements (Table 2).

3 APPLIED SPATIAL STATISTICS METHODS

Geostatistical studies of the state of cadmium and zinc soil pollution were carried out by applying ordinary (point) kriging and cokriging estimation techniques and the conditional *turning bands* simulation method. The foundations of the theory relating to both geostatistical and (conditional and unconditional) simulation methods, including examples of their application, can be found in the textbooks (Armstrong & Dowd, 1994; Armstrong, 1998; Isaaks & Srivastava, 1989; Namyslowska-Wilczynska, 1993; Rivoirard, 1994; Wackernagel, 1995).

By applying kriging it is possible to calculate the "best" local estimates of average heavy metal concentration Z^* in unsampled areas. The qualification "best" means the minimized estimation variance σ_k^2 (the lowest value of estimations' standard deviation σ_k).

Cokriging is the traditional technique used for integrating several variables that are mutually correlated during the process of estimating the average values of the parameters considered (e.g. heavy metal concentration) (Wackernagel, 1995). The estimation of one variable at a given point of a considered area consists of the linear combination of all the variables whose values are available at the neighbouring points. Cokriging requires a continuous multidimensional variogram model. An ordinary cokriging estimator is a linear combination of weights w_α^i with data representing different variables at sampling points in the neighbourhood of point x_0. Each variable is determined on the basis of a sampling set, with a different possible sample size n_i and the estimator is defined as:

$$Z_{i_0}^*(x_o) = \sum_{i=1}^{N} \sum_{\alpha=1}^{n_i} w_\alpha^i Z_i(x_\alpha) \tag{1}$$

where: index i_0 applies to a given variable from set N of variables; sample size n_i depends on index i of the variables. For the main variable, the weights add up to 1, whereas for the auxiliary variables they amount to 0:

$$\sum_{\alpha=1}^{n_i} w_\alpha^i = \delta_{ii_0} = \begin{cases} 1, & if \quad i = i_0 \\ 0, & otherwise \end{cases} \tag{2}$$

The cokriging variance is as follows:

$$\delta_{ck}^2 = \sum_{i=1}^{N} \sum_{\alpha=1}^{n_i} w_\alpha^i \gamma_{ii_0}(x_\alpha - x_0) + \mu_{i_0} - \gamma_{i_0 i_0}(x_0 - x_0) \tag{3}$$

The studied area of Dabrowa Górnicza is characterized by isotopy, which means that the available measurements for different $Z_i(x)$ variables, i.e. cadmium and zinc concentrations, were taken at the same soil sampling points. By carrying out the cokriging procedure with the cross-variogram function, a numerical pollution model, consisting of two estimated cokriging surfaces (averages Z^*) of cadmium and zinc concentrations and two local estimation accuracy surfaces (estimations' standard deviation σ_k) for the two metals, is obtained. The calculations were based on original data derived from the chemical analysis of soil samples.

In the next step of the study, the conditional *turning bands* simulation method was applied to more accurately represent the spatial extent of soil pollution and the variation in the metal concentration (Armstrong & Dowd, 1994; Isaaks & Srivastava, 1989; Kabata-Pendias, Motowicka-Terelak, Piotrowska & Terelak, Witek, 1993). The multigaussian *turning bands* method is described in more detail in (Armstrong & Dowd, 1994; ISATIS, 2001; Namyslowska-Wilczynska, 1995).

The simulation methods produce or generate realizations of the random function characterized by a covariance C. Sometimes we also have a set of data points $Z(x_\alpha)$. A simulation is conditional if it coincides with a random function at these data points. A random function is said to be multigaussian if any linear combination of its variables follows a gaussian distribution. In the stationary case, the multigaussian random function has its spatial distribution totally characterized by its mean and its covariance. The *turning bands* method is used to simulate stationary random functions and enables the generation of stochastic images, i.e.

repeated realizations $z_s(x)$ of the multidimensional Gaussian function $Z_s(x)$. Each realization $z_s(x)$ is a sum of 100 independent one-dimensional realizations generated along a line of rotation.

The *turning bands* method is a device designed to reduce a multidimensional simulation to unidimensional one. It is sufficient to simulate a stationary unidimensional random function X with covariance:

$$C_1(h) = \frac{\delta}{\delta r}\left[rC_3(r)\right] \tag{4}$$

$$Y(x) = X(\langle \Theta, x \rangle) \tag{5}$$

The simulation of the covariance C_3 is obtained by summing the realizations of the random function (one-dimensional simulations) on a given number of lines of the covariance C_1 in the three-dimensional space. Each line is called a *turning band.*

The simulation calculations were based on cadmium and zinc concentration data converted into Gaussian distribution values by applying a Gaussian anamorphosis (transformation function) modelled by the Hermite Polynomials expansion. An experimental cross-semivariogram of the converted cadmium and zinc concentrations was calculated and approximated by the complex variogram function. The parameters of this model were taken into account in the conditional *turning bands* simulation of the pollution. The count of turning bands was assumed to be 100. The seeds for the random number generator amounted to 125793. Fifty conditional simulations were run. The simulation results were converted from the Gaussian scale to the original scale.

As a result, many alternative numerical models of contamination variation, i.e. 50 simulation surfaces for the particular metals, were obtained. In addition, an averaged (based on 50 conditional simulations) map, a map of the standard deviation of realizations, a map of the smallest realizations and a map of the biggest realizations for the particular grid nodes, as well as a map of the probability of exceeding the highest threshold limit values for heavy metal soil pollution (acc. to Institute of Soil-Cultivation, Fertilizing and Pedology in Pulawy) (Kabata-Pendias, Motowicka-Terelak, Piotrowska, Terelak & Witek, 1993) were calculated.

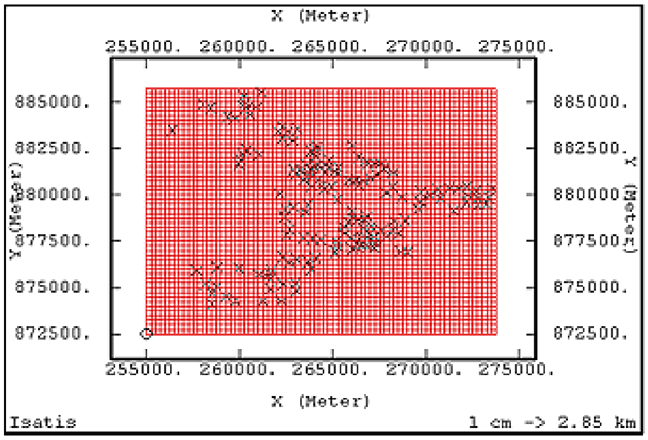

Figure 2. Grid of nodes covering the study area in the region of Dabrowa Górnicza (the dimensions of elementary grid - 250m x 250m).

A 250x250 m elementary grid of nodes was used to cover the investigated area of Dabrowa Górnicza for ordinary point kriging, cokriging and conditional *turning bands* simulation calculations (Figure 2). There were 4104 grid nodes (76 along axis x and 54 along axis y). The studied area is within the rectangular coordinates interval: X: 255000-275000 m and Y: 872500-885000 m. The geostatistical software in the ISATIS 3.4.3, 2001 package (by Geovariances-Avon Cedex and Ecole des Mines de Paris, France) was used for the calculations (ISATIS, 2001). During the performed kriging, cokriging and conditional simulation calculations the neighbourhood was assumed to be unique.

4 MULTIVARIATE ESTIMATION

4.1 Cross-variogram modelling

A multivariate estimation of average heavy metal concentration Z^* was done taking into account the cadmium and zinc concentration cross-semivariogram model parameters (Figure 3). Simple experimental variograms of the two metals concentration were also calculated. The assumed lag value amounted to 1000 m, but the number of lags was 8. We tested several geostatistical models. The model consisting of 3 basic structures was chosen to fit the experimental cadmium/zinc cross-variogram (Figure 3). The modelling of variograms was performed on 152 samples.

The first basic structure (linear) is used with a coefficient (slope) of:
- 0.91 in the cadmium variogram
- 72836.99 in the zinc variogram,
- 161.06 in the cadmium/zinc cross-variogram
and a scale factor of 2057.52m.

The second basic structure is used with a nugget effect value of:
- 1.74 in the cadmium variogram,
- 27810.99 in the zinc variogram,
- 219.74 in the cadmium/zinc cross-variogram;

The third basic structure (exponential) is used with a sill variance of:
- 0.51 in the cadmium variogram,
- 84226.25 in the zinc variogram,
- 83.17 in the cadmium/zinc cross-variogram
and a scale factor of 2057.52m.

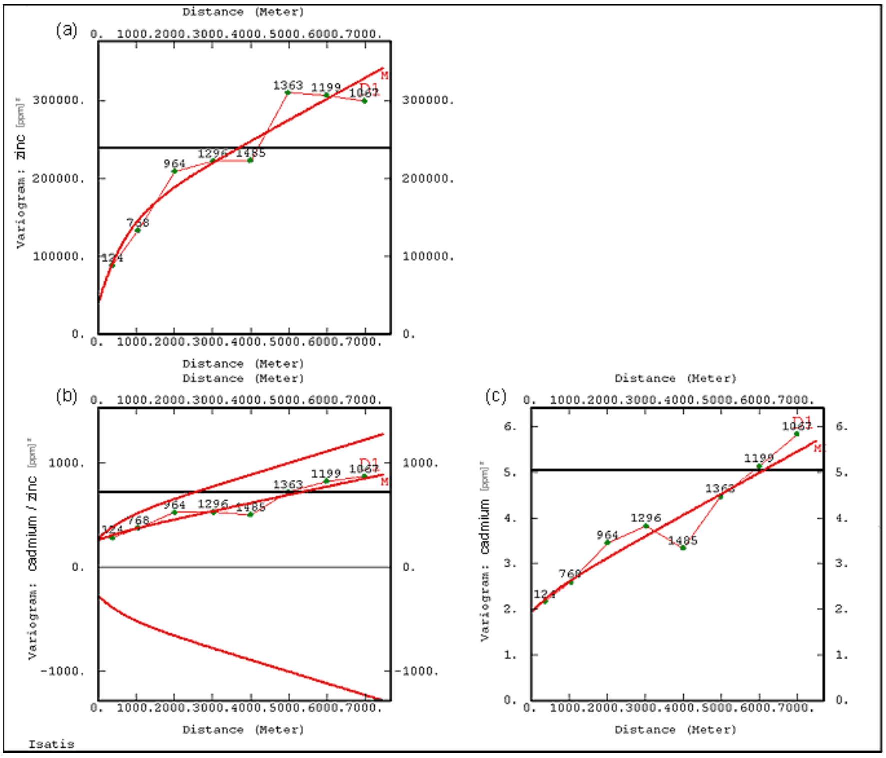

Figure 3. Simple experimental variograms of zinc (a) and cadmium (c) concentrations with the fitted theoretical models. Experimental cross-variogram cadmium/zinc (b) concentration with the approximated theoretical models (exponential, nugget effect, linear).

The results of the performed cross-validation are as follows:

- for cadmium/zinc the standardized error (mean) is 0.002 and the variance of the standardized error of estimation is 1.441 on 152 test data (0.029 and 0.788 on 148 robust data); The error (mean) is 0.005 and the variance of the error of estimation is 2.15 on 152 test data (0.048 and 1.135 on 148 robust data).
- for zinc/cadmium the standardized error (mean) is 0.004 and the variance of the standardized error of estimation is 1.160 on 152 test data (0.108 and 0.338 on 145 robust data); The error (mean) is 2.04 and the variance of the error of estimation is 67031.32 on 152 test data (21.72 and 23328.32 on 145 robust data).

4.2 Estimation results

The global results of calculations connected with the used estimation techniques are presented in Tables 3 (for cadmium concentration) and 4 (for zinc concentration).

The estimation pictures of soil pollution in the area of Dabrowa Górnicza obtained by kriging and cokriging differ for the two metals. The surface area of the cadmium contaminated soil is much more extensive (Figures 4, 6) than that for zinc (Figures 8, 10), lead (Namyslowska-Wilczynska & Wilczynski, 1999c; 2001), chromium or nickel (Namyslowska-Wilczynska, 1999; Namyslowska-Wilczynska & Wilczynski, 2001).

4.2.1 Cadmium grade estimation

In both the kriging (Figure 4) and cokriging maps (Figure 6) five subcentres of raised cadmium concentration can be distinguished. The boundaries between the particular centres are fluid, sometimes blurred. Zones with the largest cadmium contamination surface areas occur in the southern part of the studied area (Figures 4, 6). The highest estimated (by kriging and cokriging) cadmium averages range from 5.38 to >=9.17 ppm. The lowest values of the estimations' standard deviation σ_k, ranging from <1.48 to 1.78 ppm, are found within the boundaries of the subareas under environmental monitoring (Figures 5, 7).

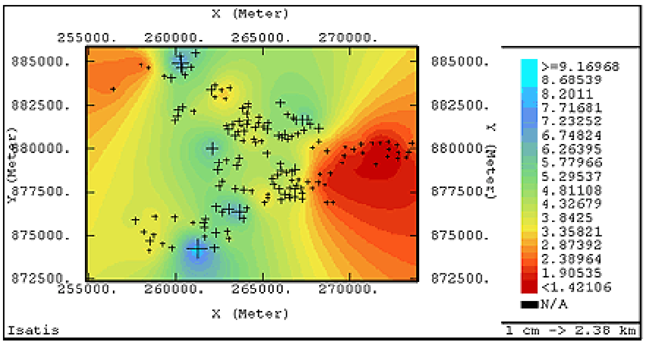

Figure 4. Raster map of estimated averages Z* of Cd concentration (ppm) with the marked original data (using Cd kriging).

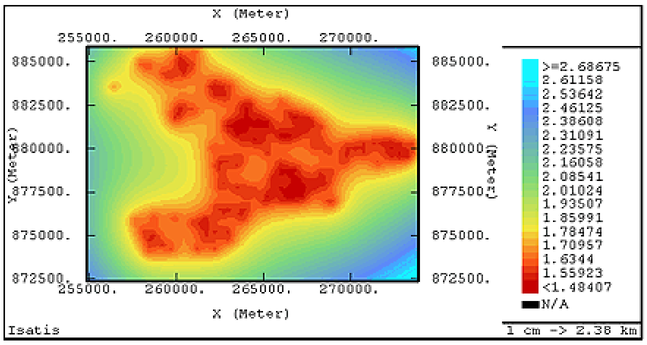

Figure 5. Raster map of standard deviation values σ_k of Cd concentration (ppm) estimates (using Cd kriging).

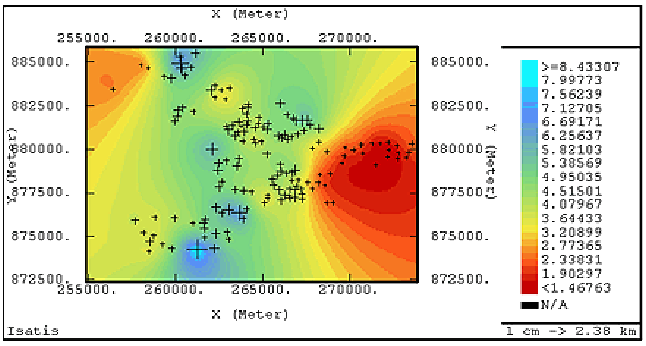

Figure 6. Raster map of estimated averages Z* of Cd concentration (ppm) with the marked original data (using Cd/Zn cokriging).

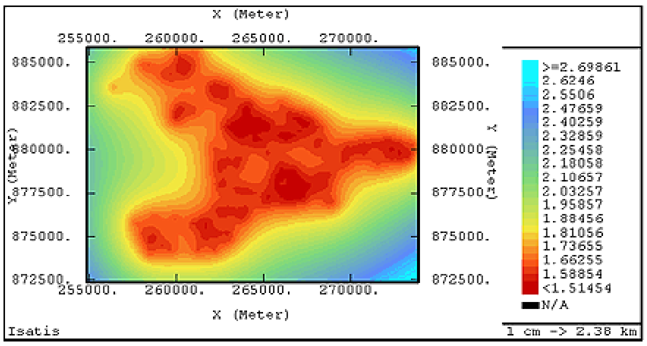

Figure 7. Raster map of standard deviation values σ_k of Cd concentration (ppm) estimates (using Cd/Zn cokriging).

The picture of soil contamination obtained by cokriging shows a spatial distribution of cadmium concentration similar, on the whole, to that obtained by kriging. Kriging yields higher estimates of Z^* averages (the range of highest Z^* values: 6.74->=9.17 ppm (Figure 4). Similar values of σ_k (range 1.48 >= 2.69 ppm) were obtained by kriging and cokriging (Figures 5, 7).

The use of cokriging gives slightly lower global values of estimated Z* averages, maximum and minimum values of Z* and the estimations' standard deviation σ_k for cadmium than the kriging calculations (Table 3). Only the minimum value of deviation σ_k is slightly higher in this case.

Table 3. Global values of cadmium concentration statistics calculated on the basis of the estimation techniques - the Dabrowa Górnicza region

Used techniques	Geostatistical parameter	Number of grid nodes n	Mean $\bar{x}$ [ppm]	Minimum x_{min} [ppm]	Maximum x_{max} [ppm]	Variance s^2 [ppm]^2	Standard deviation s [ppm]
Kriging Cd	estimated average value Z*	4104	3.82	1.42	9.17	1.30	1.14
Cokriging Cd/Zn	estimated average value Z*	4104	3.78	1.36	8.63	1.56	1.25
Kriging Cd	estimation's standard deviation σ_k	4104	1.87	1.48	2.69	0.06	0.25
Cokriging Cd/Zn	estimation's standard deviation σ_k	4104	1.83	1.51	2.46	0.04	0.21

4.2.2 Zinc grade estimation

In the case of zinc, cokriging also gives a picture of the spatial distribution of zinc concentration (Figure 10) very similar to that obtained by kriging (Figure 8). Two larger contamination centres and two, less distinct, smaller ones occur. In contrast to the cadmium estimation, the boundaries of the zones for zinc are sharper, with borders of lower value Z^* averages. A raster map of the Z^* averages (for zinc) obtained by cokriging shows high zinc pollution variability, with the zinc concentration ranging widely from 814 to 3124 ppm, the highest values of Z^* occurring in the interval: 1777->=3124 ppm (Figure 10). The picture yielded by kriging shows a slightly wider range of values of Z^* averages (847->=3216 ppm) (Figure 8). For cokriging, the range of lowest σ_k values obtained is at a slightly different level of σ_k <235-343 ppm (Figure 11) as it is in the case of kriging (<191-309 ppm) (Figure 9). However for cokriging, the lower limit of σ_k values begins at a higher level (Figure 11). The results of the standard deviation value (σ_k) calculations by kriging and cokriging can be explained by a very high variation coefficient V_{Zn}, which can be as high as ~116% as well as high coefficient V_{Cd} of approximately 55%. Similarly as in the case of the cadmium concentration, the lowest values of σ_k (<1,48-1,73 ppm) (Figures 5, 7) for the zinc concentration (<191-309 ppm and <235-343 ppm) were obtained in the areas of densest sampling (Figures 9, 11). Using the cokriging technique gives only slightly lower estimated Z* averages (Figure 10), compared to the kriging results (Figure 8). The sub-areas of the lowest zinc σ_k values constitute larger surfaces than when cokriging was used. The very large variability in zinc concentration is probably the reason, and taking into account this parameter's value (as a secondary variable) doesn't decrease σ_k values during the estimation of the Z* averages for cadmium.

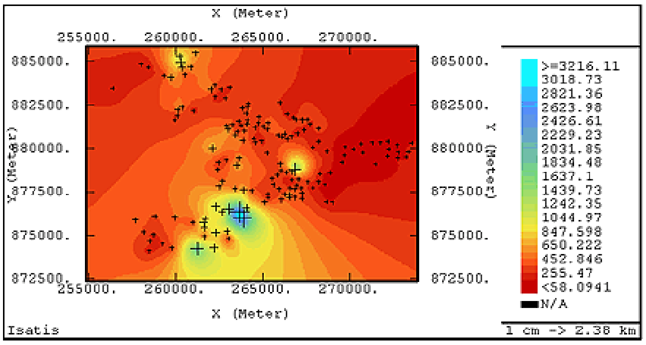

Figure 8. Raster map of estimated Z* averages of Zn concentration (ppm) with the original data marked (using Zn kriging).

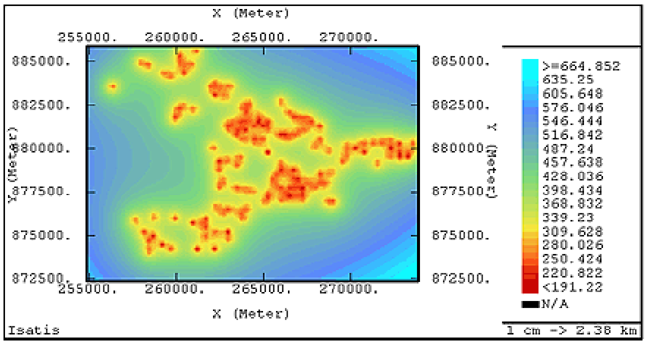

Figure 9. Raster map of standard deviation values σ_k of Zn concentration (ppm) estimates (using Zn kriging).

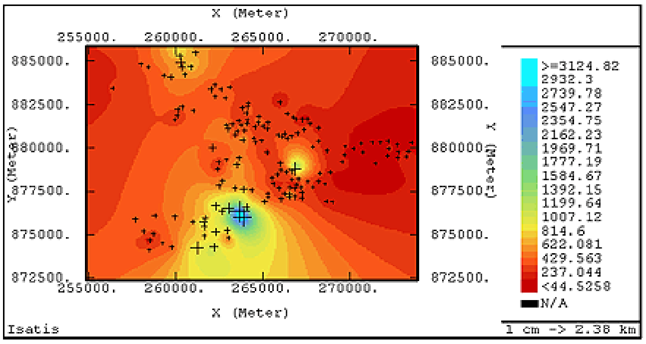

Figure 10. Raster map of estimated Z* averages of Zn concentration (ppm) with the original data marked (using Zn/Cd cokriging).

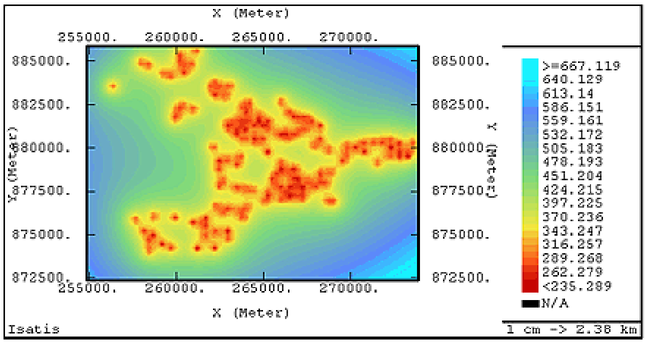

Figure 11. Raster map of standard deviation values σ_k of Zn concentration (ppm) estimates (using Zn/Cd cokriging).

A comparison of basic, global statistical parameter values for the applied estimation techniques gives results (Table 4) similar to that for cadmium concentration (Table 3). Using cokriging, lower values of estimated Z* averages, as well as maximum, minimum Z* values for zinc compared with kriging results, have been obtained too. Note the very high average of deviation σ_k for zinc (Table 4). The estimation's standard deviation values σ_k are also lower for cokriging (except for minimum value of σ_k) (Table 4).

Table 4. Global values of zinc concentration statistics calculated on the basis of the estimation techniques - the Dabrowa Górnicza region

Used techniques	**Geostatistical parameter**	**Number of grid nodes n**	**Mean $\bar{x}$ [ppm]**	**Minimum x_{min} [ppm]**	**Maximum x_{max} [ppm]**	**Variance s^2 [ppm]^2**	**Standard deviation s [ppm]**
Kriging Zn	estimated average value Z*	4104	399.14	58.09	3216.11	73703.46	271.48
Cokriging Zn/Cd	estimated average value Z*	4104	390.05	11.01	3161.97	89153.35	298.59
Kriging Zn	estimation's standard deviation σ_k	4104	437.31	191.22	664.85	9530.62	97.62
Cokriging Zn/Cd	estimation's standard deviation σ_k	4104	427.91	249.32	601.85	7912.03	88.95

5 MULTIVARIATE SIMULATION

Conditional *turning bands* simulation was applied in the next stage of investigating the heavy metal pollution of soil. For this purpose the original, determined heavy-metal data (from chemical analyses) were converted into Gaussian distribution values (Figures 12, 13) using Gaussian anamorphosis (Section 3). This transformation of the analysed metal concentrations gives a correspondence between the sorted data (vertical axis) and the corresponding frequency quantile in the gaussian scale (horizontal axis).

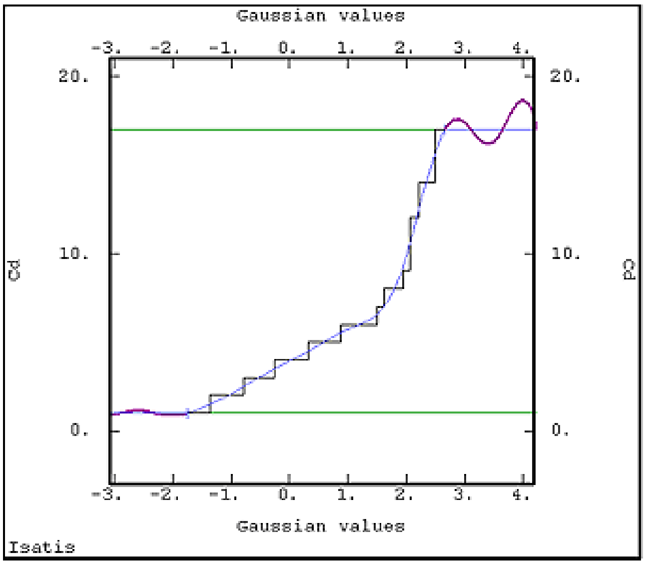

Figure 12. Experimental anamorphosis with the theoretical model (transformed data of cadmium concentration, ppm).

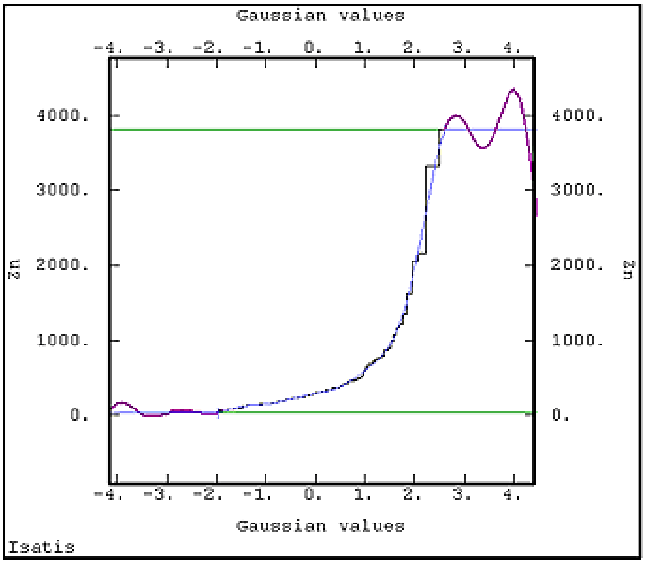

Figure 13. Experimental anamorphosis with the theoretical model (transformed data of zinc concentration, ppm).

5.1 Gaussian cross-variogram modelling

Simple experimental variograms and an experimental cross-variogram of the cadmium and zinc concentrations were converted into Gaussian values and these were approximated by a model consisted of three structures: Bessel J, nugget effect and spherical (Figure 14).

The first basic structure (Bessel J) is used with a sill variance of:
- 0.56 in the cadmium (Gaussian) variogram,
- 0.60 in the zinc (Gaussian) variogram,
- 0.83 in the cadmium/zinc (Gaussian) cross-variogram

and a scale factor of 2057.52m.

The second basic structure is used with a nugget effect value of:
- 0.34 in the cadmium (Gaussian) variogram,
- 0.32 in the zinc (Gaussian) variogram,
- 0.20 in the cadmium/zinc (Gaussian) cross-variogram.

The third basic structure (spherical) is used with a sill variance of:
- 0.27 in the cadmium (Gaussian),
- 0.15 in the zinc (Gaussian),
- 0.18 in the cadmium/zinc (Gaussian) cross-variogram

and a range of 2057.52m.

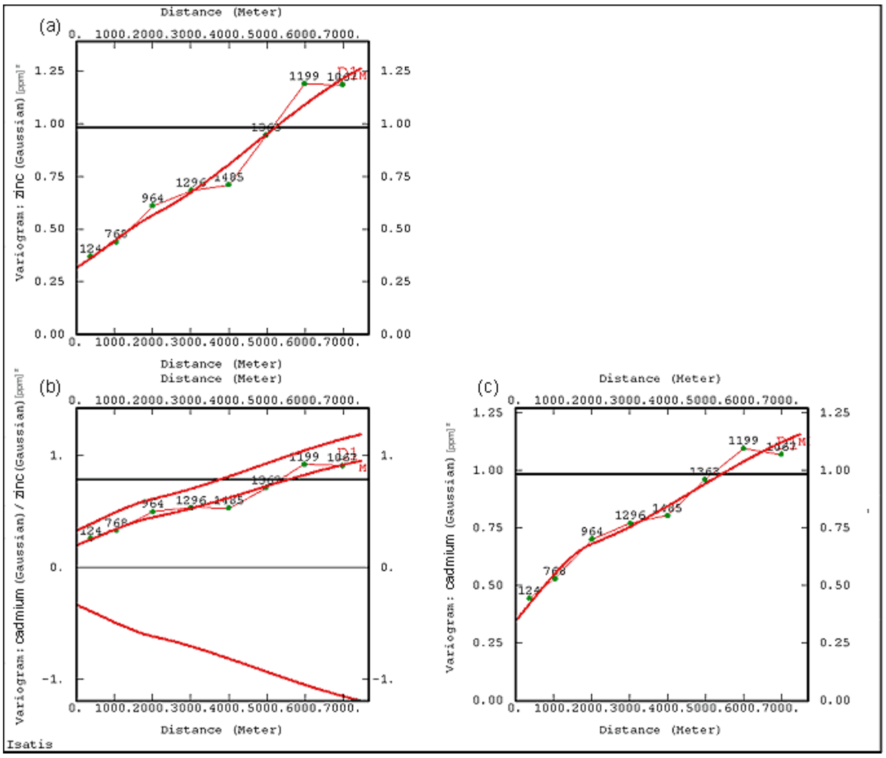

Figure 14. Simple experimental variograms (a, c) and cross-variogram (b) of transformed data of zinc and cadmium concentrations, approximated by the theoretical models (Bessel J, nugget effect, spherical).

For the fitted geostatistical model we obtained the following cross-validation results:

- for cadmium/zinc (gaussian variables) the standardized error (mean) is 0.000 and the variance of the standardized error of estimation is 1.058 on 152 test data (0.010 and 0.900 on 149 robust data); The error (mean) is 0.00003 and the variance of the error of estimation is 0.27 on 152 test data (0.00601 and 0.23 on 149 robust data).
- for zinc/cadmium (gaussian variables) the standardized error (mean) is 0.000 and the variance of the standardized error of estimation is 0.897 on 152 test data (0.024 and 0.726 on 149 robust data); The error (mean) is 0.00044 and the variance of the error of estimation is 0.21 on 152 test data (0.01154 and 0.17 on 149 robust data).

5.2 Results of the conditional *turning bands* simulation

The same elementary grid size used for kriging, cokriging (Section 3, Figure 2) and the parameters of Gaussian cross-variogram (Figure 14) was used for the conditional simulation calculations. The global results of the conditional simulation (turning bands) are given in Tables 5 (for the cadmium concentration) and 6 (for the zinc concentration).

5.2.1 Cadmium grade simulation

The multivariate simulation model with the cadmium concentration as the main variable yields a highly accurate representation of the cadmium concentration fluctuations (Figures 15-18). All the cadmium contamination subcentres obtained by cokriging, which previously appeared on the map of estimated *Z** averages for cadmium (Figure 6) become distinct. By averaging of 50 simulations for cadmium one can see on the map that the highest simulated values range from 5.29 to >=7.08 ppm (Figure 15). The boundaries between the particular subcentres are often blurred. These subcentres are much more distinct than in the kriging and cokriging contamination pictures (Figures 4, 6). The highest realization standard deviation values: from 2.40->=3.28 ppm occur in the north and south of the studied region (Figure 16). A map of the biggest cadmium concentration realizations indicates the possibility of more intensive contamination with a values range of 12.52->=17 ppm (Figure 17). Whereas the map of the smallest realizations shows a possible contamination of 3.28->=4.32 ppm (Figure 18).

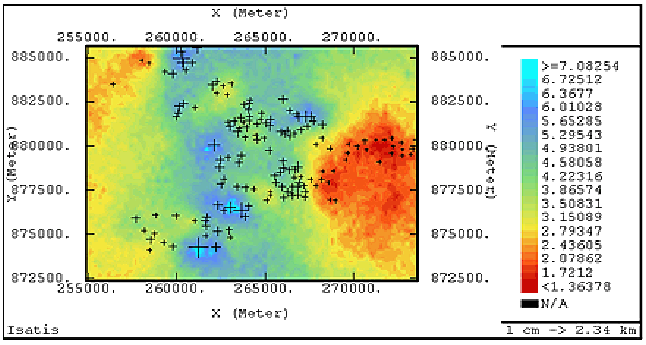

Figure 15. Raster map of multivariate conditional cadmium/zinc simulation of cadmium concentration (ppm) - the averaged picture from 50 simulations.

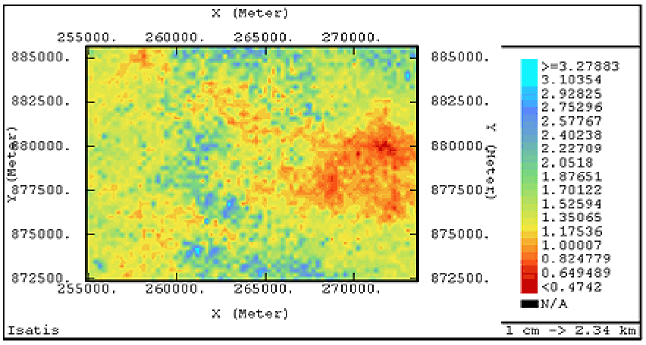

Figure 16. Raster map of standard deviation of realizations for cadmium concentration (ppm) - multivariate conditional cadmium/zinc simulation.

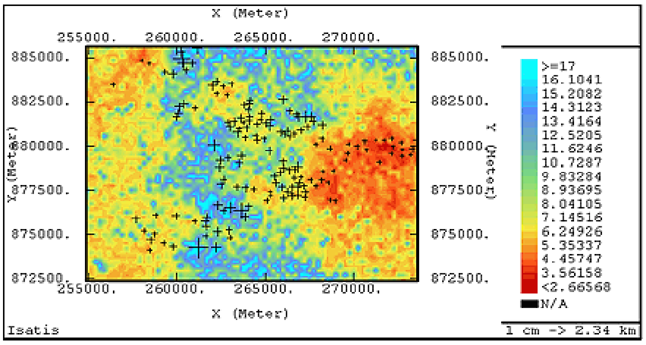

Figure 17. Raster map of the largest simulated values for cadmium concentration (ppm) in particular grid nodes (multivariate conditional cadmium/zinc simulation).

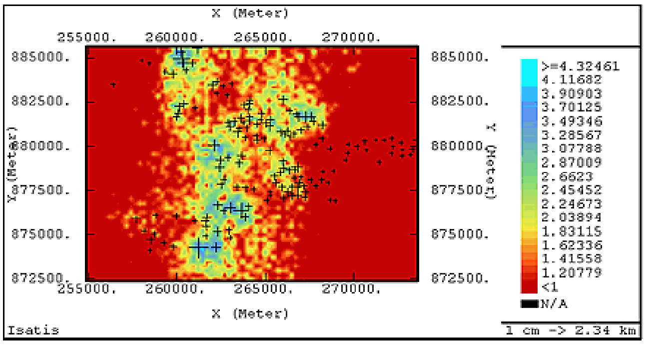

Figure 18. Raster map of the smallest simulated values for cadmium concentration (ppm) in particular grid nodes (multivariate conditional cadmium/zinc simulation).

The arrangement of zones of the smallest simulated Z_s values (Figure 18) is similar to the averaged map obtained on the basis of 50 realizations for the cadmium (Figure 15).

The global average of the cadmium concentration calculated on the basis of simulated Z* values using the turning bands method (Table 5) indicates an almost identical degree of soils pollution as that shown by the kriging and cokriging computations (Table 3). However the range of simulated Z_s values for the cadmium is narrower (Figure 15) than in case of the estimated Z* values (Figures 4 and 6).

Table 5. Global values of cadmium concentration. Statistics calculated on the basis of the conditional cadmium/zinc turning bands simulation (for 50 realizations) of the Dabrowa Górnicza region

Geostatistical parameter	Number of grid nodes n	Mean $\bar{x}$ [ppm]	Minimum x_{min} [ppm]	Maximum x_{max} [ppm]	Variance s^2 $[ppm]^2$	Standard deviation s [ppm]
Averaged simulated value $\bar{Z}_s$	4104	3.73	1.27	7.22	1.18	1.09
Realizations' standard deviation σ_s	4104	1.52	0.47	3.65	0.21	0.46
Maximal value of simulated values $Z_{s,max}$	4104	8.71	2.75	17.00	14.09	3.75
Minimal value of simulated values $Z_{s,min}$	4104	1.30	1.00	3.95	0.27	0.52

Simulation maps made for the cadmium pollution thresholds i.e. 5 ppm and 10 ppm, (Figures 19 and 20) show interesting results. The map of the probability of exceeding the 5 ppm threshold shows a similar level of cadmium soil pollution (Figure 19) as the map averaged from 50 simulations for this metal (Figure 15). The general picture of the cadmium pollution presented in this map (Figure 19) is similar to the image of the largest simulated Z_s values (Figure 17). The highest probability values for exceeding such a threshold are in an interval of 0.64->0.86. If a higher threshold limit value for cadmium is assumed e.g. 10 ppm, the highest probability values sharply decrease to 0.15->=0.18 (Figure 20). The probability map for this threshold shows five very small potential zones (occurring separately) where this contamination limit could be exceeded. Their surfaces were miniaturized and can indicate the cadmium pollution source. Zones of the increased cadmium concentration are more distinctly marked in the simulation images (Figures 15, 19) than in the maps of Z* averages, obtained by means of kriging or cokriging (Figures 4, 6).

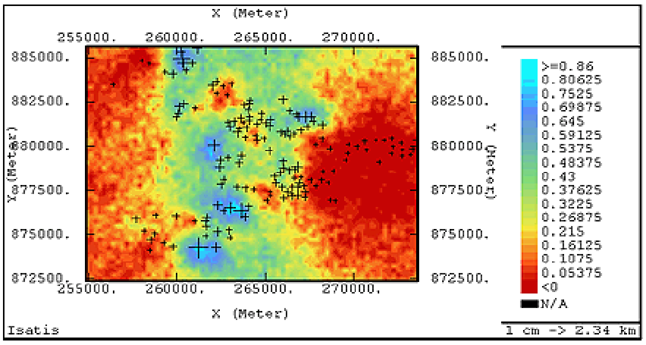

Figure 19. Probability raster map of cadmium concentration (the threshold - 5 ppm) using multivariate conditional cadmium/zinc simulation.

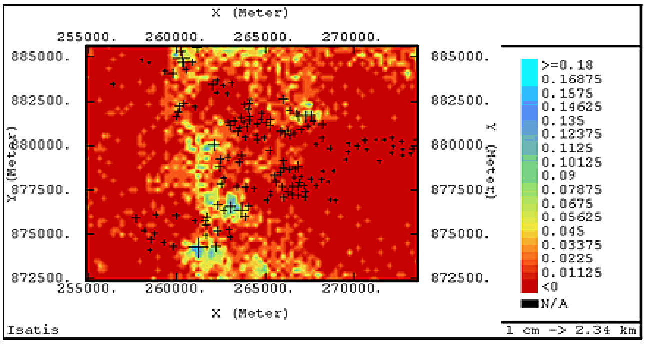

Figure 20. Probability raster map of cadmium concentration (the threshold - 10 ppm) using multivariate conditional cadmium/zinc simulation.

5.2.2 Zinc grade simulation

The multivariate simulation of soil pollution with the zinc concentration as the main variable (Figures 21-24) and the cadmium concentration as the auxiliary variable generally gives a similar zinc concentration picture to that obtained by kriging and cokriging (Figures 8, 10). The pollution areas calculated by means of the simulation method are more distinct (Figure 21). Two contamination centres are visible: one is more extensive and consists of two subcentres (two separate centres in the case of kriging and one centre for cokriging) and the others which are much less distinct. The highest simulation values are in an interval of 1080->1535 ppm (Figure 21). If one compares the averaged contamination simulation picture (Figure 21) with the one obtained by kriging and cokriging (Figures 8, 10), one will notice the absence of the smaller centre of raised zinc concentration that occurs in the middle part of the studied area in Figures 8 and 10. The simulation results prove the existence of much weaker zinc pollution (lower Z_s realization value of 444 ppm) in this place.

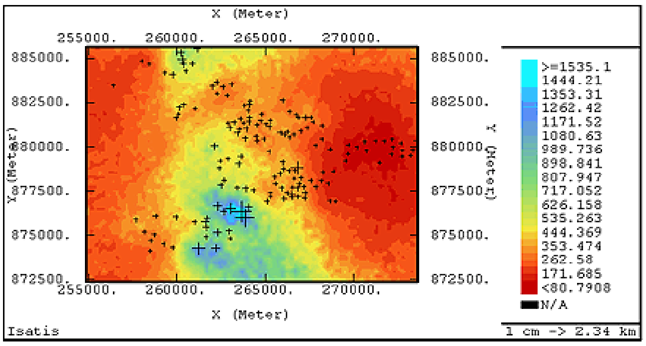

Figure 21. Raster map of multivariate conditional zinc/cadmium simulation of zinc concentration (ppm) - the averaged picture of 50 simulations.

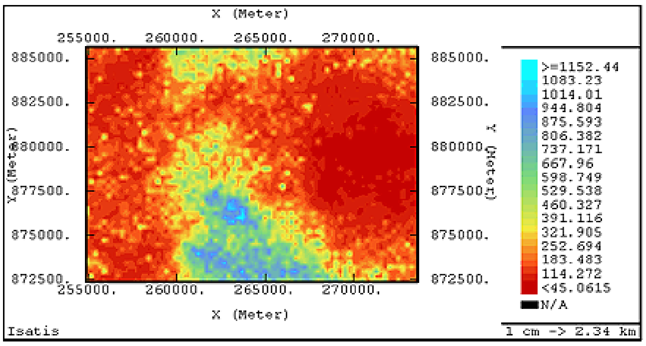

Figure 22. Raster map of standard deviation of realizations for zinc concentration (ppm) - multivariate conditional zinc/cadmium simulation.

The estimations' standard deviation values σ_k in kriging and cokriging conditions are in an interval of 191.22->=667.11 ppm (Figures 9, 11), whereas the realizations' standard deviation values range very widely: <45->=1152 ppm (Figure 22). The lowest σ_k values were obtained in the environmentally monitored areas; with intervals of <235.28-343.25 ppm for kriging and cokriging results (Figure 9, 11). The lowest σ_s values were <45->322 ppm for *turning bands* simulation results (Figure 22).

The highest realizations' standard deviation values of 806->1152 ppm occur in the zones with the highest zinc contamination (Figures 21, 22) in the southern part of the investigated area. Two simulation maps of the largest and smallest zinc concentration realizations for the particular grid nodes show distinct sub-areas of more intensive contamination. In the former case, the highest simulated zinc values range from 2678>=3820 ppm (Figure 23) and in the latter case: 341-479 ppm (Figure 24). As for cadmium, the first map presents a possible area of more intensive zinc pollution (Figure 23), while the second map (Figure 24) shows similar borders to its contaminated area on that in the averaged picture of 50 simulations for the zinc (Figure 21).

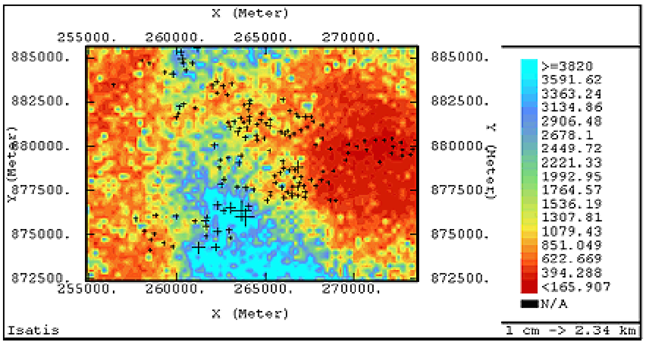

Figure 23. Raster map of the largest simulated values for zinc concentration (ppm) in particular grid nodes (multivariate conditional zinc/cadmium simulation).

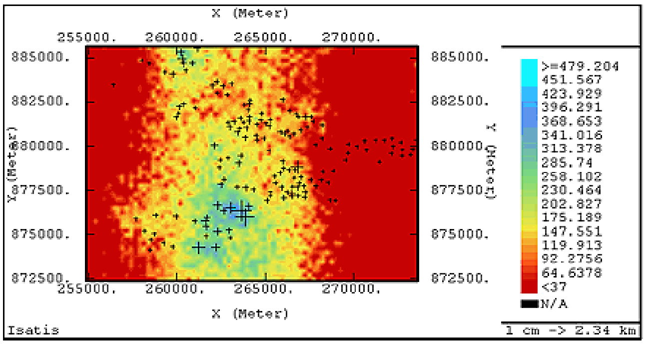

Figure 24. Raster map of the smallest simulated values for zinc concentration (ppm) in particular grid nodes (multivariate conditional zinc/cadmium simulation).

The global results of the simulation calculations performed for zinc concentration (Table 6) presents a picture of lower soil pollution by this metal. A markedly lower average value based on simulated Z_s values in particular grid nodes, compared to kriging or cokriging calculations, was obtained (Table 4).

Table 6. Global values of zinc concentration. Statistics calculated on the basis of the conditional zinc/cadmium turning bands simulation (for 50 realizations) - the Dabrowa Górnicza region

Geostatistical parameter	Number of grid nodes n	Mean $\bar{x}$ [ppm]	Minimum x_{min} [ppm]	Maximum x_{max} [ppm]	Variance s^2 $[ppm]^2$	Standard deviation s [ppm]
Averaged simulated value $\bar{Z}_s$	4104	367.09	79.82	1759.91	52793.00	229.77
Realizations' standard deviation σ_s	4104	252.92	47.22	1183.30	47063.54	216.94
Maximal value of simulated values $Z_{s,max}$	4104	1389.40	182.18	3820.00	1190952.37	1091.31
Minimal value of simulated values $Z_{s,min}$	4104	95.46	37.00	408.40	4788.78	69.20

Even for the assumed soil zinc pollution threshold of 1000 ppm, which is high compared with the legally allowable contamination standards of 200-300 ppm (Kabata-Pendias, Motowicka-Terelak, Piotrowska, Terelak & Witek, 1993), the probability of exceeding this threshold is rather high (interval 0.50-0.64) (Figure 25). This could occur in three places: in one case the pollution zone is quite large while in the other cases it is very small. There is a similarity to the averaged map obtained from 50 realizations, even though the surface zinc pollution was lower (Figure 21).

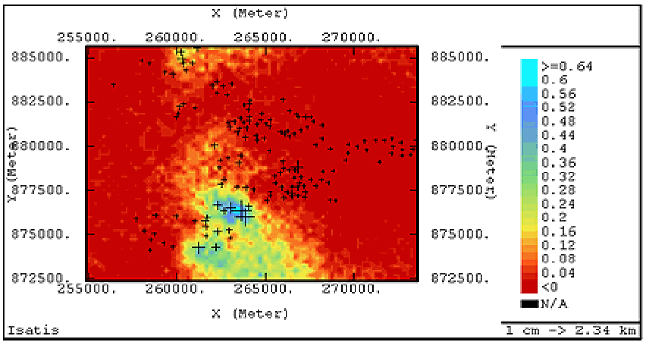

Figure 25. Probability raster map of zinc concentration (the threshold - 1000 ppm) using multivariate conditional zinc/cadmium simulation.

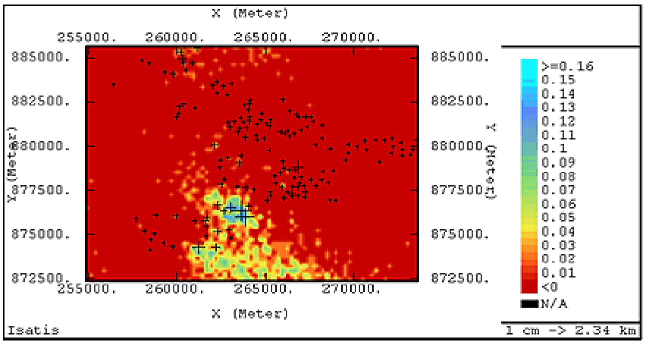

Figure 26. Probability raster map of zinc concentration (the threshold - 3000 ppm) using multivariate conditional zinc/cadmium simulation.

If the threshold is increased to 3000 ppm (Figure 26), the sub-areas of raised zinc concentration are still visible, but the contaminated surface area is much smaller, being reduced to points in the southern part of the studied region. Exceeding this high threshold is linked with a range of very low probability values (0.12->=0.16).

Most of the centres representing the highest cadmium, zinc and lead metal concentrations show a similar spatial distribution (Namyslowska-Wilczynska & Wilczynski, 1999c; 2001). The distribution of some of the anomalies in lead, cadmium and zinc concentration is partly correlated with the spatial characteristics obtained for the other metals studied (copper, nickel and chromium) (Namyslowska-Wilczynska, 1999).

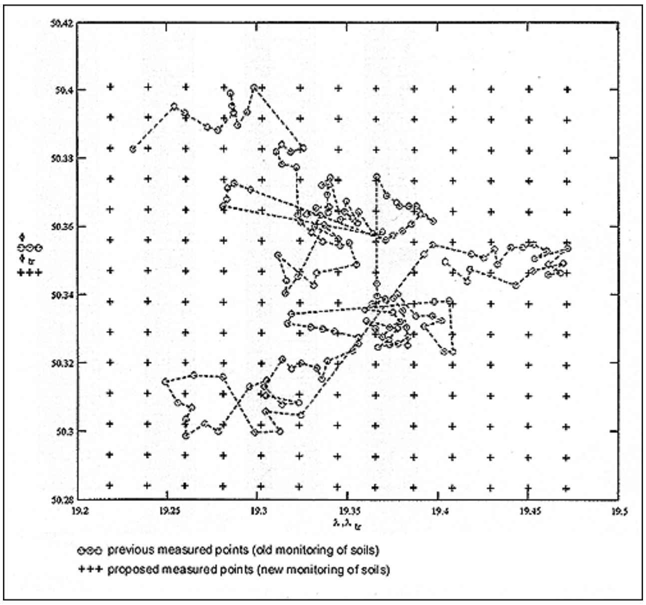

Figure 27. Scheme of the proposed sampling grid of soils in the region of Dabrowa Górnicza.

The results obtained from these geostatistical studies now form the basis for planning a sampling grid for a new environmental monitoring scheme in the region of Dabrowa Górnicza. A uniform sampling grid of soils with two variants of sample size (N = 182, which is cheaper and N = 266, which is more expensive) has been proposed (Namyslowska-Wilczynska, 1999). A possible sampling grid scheme (cheaper) with the forecasting points' coordinate values converted into the geographic coordinates is presented in Figure 27.

6 CONCLUSION

The methods of *kriging*, *cokriging* and *conditional turning bands simulation*, which were applied to study the levels of cadmium and zinc pollution of the surface soil layer constitute highly effective tools for processing and analysing data. With these methods of raised heavy metal concentration anomalies could be identified in the area of Dabrowa Górnicza.

Considering the fact that there is a relatively close correlation between cadmium and zinc, a model of multivariate estimation and conditional simulation were proposed for the estimation of the extent, degree and character of soil contamination. As a result, the accuracy of the calculated metal concentration averages and thus the reliability of the contamination picture obtained was increased.

If cokriging is used, slightly lower estimated Z^* averages and estimations' standard deviation σ_k for the both analyzed metals are obtained.

The conditional turning bands simulation results (averaged picture based on 50 realizations) indicate lower zinc pollution than that shown by the kriging and cokriging images for this metal. For cadmium, the second metal, the simulation pollution pictures obtained are similar to the kriging and cokriging images. But the simulated zinc concentration values are obtained at much higher realization standard deviation values compared to that of cadmium.

While kriging and cokriging techniques allow one to quickly determine the boundaries of sub-areas with higher heavy metal concentrations, conditional turning bands simulation reflects their fluctuation well. The latter method has the advantage of producing different kinds of simulation maps e.g. average pictures from simulated values, maps of maximum and minimum simulated values as well as maps of the probability of exceeding assumed pollution thresholds at the particular grid nodes.

The spatial extent of cadmium and zinc soil pollution in the area of Dabrowa Górnicza obtained by the methods used only partially coincide, mainly in the southern part of region and a little in the northern one.

With regard to cadmium, an extensive soil area with raised levels of this metal was identified. The maximum cadmium concentration values are linked to the close proximity of metallurgical plants (both inactive and active) and to the consequences of past zinc-lead mining operations. The distribution of the high cadmium concentration identified in the soils is also influenced by the deposition of ore-bearing dolomites close to the surface and by a waste dump situated next to the metallurgical plant.

The higher contamination (distinctly higher zinc centres) of the soils in the area of Dabrowa Górnicza by zinc is linked to the metallurgical plants located there and with the mining industry (historical mining centres). The deposition of dust emitted by the industrial plants and the spread of the dust from the settling ponds by wind contributes greatly to the pollution of the soil environment.

Tested geostatistical methods are an efficient tools for processing and analysing of any kind of data contained in thematic databases, including environmental databases.

7 REFERENCES

Armstrong M. & Dowd P.A. (Eds.) (1994) *Geostatistical Simulations*, Dordrecht: Kluwer Academic Publisher.

Armstrong M. (1998) *Basic Linear Geostatistics*, Berlin: Springer.

European Environment Agency (1999) *Environment in the European Union at the turn of the century*, Environmental assessment report No. 2, Copenhagen Denmark.

Isaaks E.H. & Srivastava R.M. (1989) *An Introduction to Applied Geostatistics*. Oxford, New York, USA: Oxford University Press.

ISATIS (2001) *Isatis Software Manual*, Avon Cedex, France: Geovariances and Ecole des Mines de Paris (Centre de Geostatistique).

Kabata-Pendias A., Motowicka-Terelak T., Piotrowska M., Terelak H. & Witek T. (1993) *Assessment of Soil and Plant Contamination with Heavy Metals and Sulphur*. General Guidelines for Agriculture (in Polish), Pulawy, Poland: Institute of Soil-Cultivation, Fertilizing and Pedology.

Lis J. & Pasieczna A. (1995a) *Geochemical Atlas of Poland, scale: 1 : 2 500 000*, Warsaw, Poland: National Geological Institute.

Lis J. & Pasieczna A. (1995b) *Geochemical Atlas of Upper Silesia, scale 1 : 200 000*, Warsaw, Poland: National Geological Institute.

Lis J. & Pasieczna A. (1999) *Detailed Geochemical Map of Upper Silesia, scale 1 : 25 000*, Slawków, Warsaw, Poland: National Geological Institute.

Namyslowska-Wilczynska B. (1993) *Variability of Copper Ore Deposits in the Foresudetic Monocline Against the Background of Geostatistical Methods*, Monography of Institute of Geotechnics and Hydrotechnics, Wroclaw University of Technology, Series 64/21, Wroclaw, Poland.

Namyslowska-Wilczynska B. (1995) Selected Methods of Geostatistical Simulation (in Polish), *2nd Conference on Computer Aiding of Scientific Research KOWBAN'95* (pp. 251-262), Wroclaw, Poland.

Namyslowska-Wilczynska B. & Wilczynski A. (1997) *Geostatistical Studies of Soil Heavy Metal Pollution in Selected Areas of Upper Silesia (in Polish). Environment Protection*, publication under scientific auspices of Civil Engineering Committee of Polish Academy of Sciences 2(65), pp. 9-18, Lower-Silesian Section of Polish Association of Sanitary Engineers and Technicians.

Namyslowska-Wilczynska B. & Wilczynski A. (1999a) An assessment of soil contamination with heavy metals using indicator geostatistics in Upper Silesia (S Poland). *Proceedings of The Third Annual Conference of the International Association for Mathematical Geology* (pp. 905-910), Barcelona, Spain.

Namyslowska-Wilczynska B. & Wilczynski A. (1999b) Application of Indicator Geostatistics to Modelling of Environment Heavy Metal Pollution (in Polish), 4th *All-Polish Scientific Conference General and Particular Environmental Engineering Problems* (pp. 461-474), Scientific Papers of Engineering Building and Environmental Engineering Faculty, No. 15, Series: Environmental Engineering, Koszalin Polytechnic, Ustronie Morskie, Poland.

Namyslowska-Wilczynska B. (1999) *Determination of Soil Sampling Points in Farmland in Dabrowa Górnicza Area by Means of Geostatistical Analysis (in Polish)*, Report of Institute of Geotechnics and Hydraulic Engineering, Wroclaw University of Technology, series: SPR, No. 6/99, Wroclaw, Poland.

Namyslowska-Wilczynska B. & Wilczynski A. (1999c) Geostatistical methods and turning bands simulation assessment of heavy metals pollution of soils in Dabrowa Górnicza area. *International Conference on "Spatial Information Management in The New Millenium"* (pp. 97-107), Cracow, Poland.

Namyslowska-Wilczynska B. & Rusak K. (1999) Time factor in the studies of soils pollution by heavy metals (in Polish). *6th Conference on Computer Aiding of Scientific Research* (pp. 215-227), Polanica Zdrój, Poland.

Namyslowska-Wilczynska B. & Wilczynski A. (2000) An application of the turning-bands-simulation method for the modeling of environmental data. *Book of Abstracts. CODATA 2000, 17th International CODATA Conference* (pp. 177-178), Baveno, Italy.

Namyslowska-Wilczynska B. & Pyra J. (2000) Use of kriging estimation methods for assessment of state of soil copper and lead pollution in Legnica-Glogów copper basin area (in Polish), *3rd Forum of Environmental Engineering on Mathematical Modelling Modelling in Strategy of Environment Management* (pp. 222-234), Nałęczów, Poland: Wydawnictwo Ekoinżynieria Lublin,.

Namyslowska-Wilczynska B. & Wilczynski A. (2001) Geostatistical Studies of Distribution of Heavy Metals Concentration in Soils. *Journal - Geoinformatica Polonica 2*, pp. 51-65, Cracow, Poland: Polish Academy of Arts and Science.

Rivoirard J. (1994) *Introduction to disjunctive kriging and non-linear geostatistics.* Oxford: Clarendon.

Wackernagel H. (1995) *Multivariate Geostatistics*, Berlin, Heidelberg, New York: Springer - Verlag.