ORIGINAL ARTICLE

Geochemical characterization of heavy metal contaminated area using multivariate factorial kriging

Joaquim C. B. Queiroz · José R. Sturaro · Augusto C. F. Saraiva · Paulo M. Barbosa Landim

Abstract This paper describes a geostatistical method, known as factorial kriging analysis, which is well suited for analyzing multivariate spatial information. The method involves multivariate variogram modeling, principal component analysis, and cokriging. It uses several separate correlation structures, each corresponding to a specific spatial scale, and yields a set of regionalized factors summarizing the main features of the data for each spatial scale. This method is applied to an area of high manganese-ore mining activity in Amapá State, North Brazil. Two scales of spatial variation (0.33 and 2.0 km) are identified and interpreted. The results indicate that, for the short-range structure, manganese, arsenic, iron, and cadmium are associated with human activities due to the mining work, while for the long-range structure, the high aluminum, selenium, copper, and lead concentrations, seem to be related to the natural environment. At each scale, the correlation structure is analyzed, and regionalized factors are estimated by cokriging and then mapped.

Keywords Heavy metal pollution · Amapá State, Brazil · Factorial kriging · Multivariate geostatistics

J. C. B. Queiroz
Department of Statistics, Federal University of Para,
Belém, Pará, Brazil

J. R. Sturaro · P. M. Barbosa Landim (✉)
Department of Applied Geology,
São Paulo State University-Rio Claro,
São Paulo, Brazil
e-mail: plandim@rc.unesp.br

A. C. F. Saraiva
Central Laboratory Eletronorte, Belém, Pará, Brazil

Introduction

Contamination of surface and groundwaters causes shortages and even restricts their use. Generally groundwaters can suffer direct undiluted contamination when pollution reaches the groundwaters directly through abandoned, deficiently constructed wells, and indirectly through diluted contamination. In the indirect case the pollution reaches the groundwaters after passing through areas with solid residue deposition (urban and industrial), petroleum leaks, or mining wastes. Fully understanding, estimating, and mapping spatial variations of heavy metal water pollution using efficient techniques enables accurate monitoring and remedial action.

Univariate geostatistical techniques have been widely used in Earth Science to analyze spatial patterns and variations in pollutant concentrations, but multivariate geostatistical studies are less common than univariate analysis. Two frequently observed features are correlations between variables and scale-dependent spatial variability. The combination of these two features means that regionalized variables can have different degrees of correlation at different spatial scales. To analyze multivariate spatial data sets, Matheron (1982) proposed a geostatistical method named factorial kriging analysis that involves semivariographic multivariate modeling, principal component analysis, and cokriging. This method allows correlation structures found at different scales to be distinguished and produces a group of regionalized factors summarizing the main data characteristics at each spatial scale. Factorial kriging analysis has been applied to geophysics (Chiles and Guilen 1984; Galli et al. 1984), geochemistry (Sandjiv 1984; Wakernagel and Butenuth 1989), soil science (Goulard 1989; Goovaerts 1991; Castrignanò et al. 2000; Lin et al. 2002), and hydrogeology (Rouhani and Wackernagel 1990).

This study applies factorial kriging to assess the potential contamination of heavy metals in the soil, surface waters and groundwaters caused by decades of intense mining activity in the district of Santana, State of Amapá, in northern Brazil. Analyses of the spatial scales of variation in contamination were applied in this study as a way to separate local and regional phenomena characteristics, and also to verify relationships seen between different metal concentrations. In this way, it was possible to ascertain whether the origin of contamination from certain metals is due to human activities or the natural environment.

Study area

The study area is located in the district of Santana, Amapá State, in the extreme north of Brazil, approximately between 50° and 55° W and 0° to 5°N (Fig. 1a). Santana City has a population of 80,439 inhabitants, the second largest in the state (IBGE 2000). In 1953, following the discovery of high quality manganese in Serra do Navio, about 200 km from Macapá, the state's capital (Fig. 1b, c), the Ore Trade Industry Inc (ICOMI) was established to exploit and commercialize the ore. In order to carry out the mining, ICOMI constructed a residential colony near the manganese mines in Serra do Navio and Santana with a complete infrastructure, including sanitation, recreation facility, schools, supermarket, hospital, and housing for the company's employees and their families, in addition to the industrial installation. The industrial Santana area with approximately 129 hectares, was defined to be strictly for industrial purposes. The area was basically used to stock manganese and iron ores, products (pellets/sinter and alloys), and raw materials (fuel, coke, etc.) that arrived and departed through the ICOMI port and rail terminal (PCA/Environmental Control Plan 2001). The manganese and chromite ore were transported by railway from the Serra do Navio mines to the ICOMI industrial area in the Port of Santana, a distance of approximately 200 km.

Geologically, the studied area, which extrapolates the perimeter of ICOMI, is over sediments of the Barreiras Formation consisting of silty organic clays, clay silts, and hard clay with occasional intercalations of fine and coarse sand. The hydrograph is an important element in the physical landscape and the area's economy. Santana Port fronts the Northern Channel of the Amazon River, where almost all its waters are discharged into the Atlantic Ocean. The study area is drained by several tributaries, which are generally narrow, sinuous water courses, just navigable by small boats. It is drained by the Elesbão 1 and Elesbão 2 tributaries that cross the ICOMI industrial area. It is totally covered by Quaternary sediments, silt rich in organic matter, and covered with a 0.30 to 0.50 m thick humus layer (PCA/Environmental Control Plan 2001).

Until its closure in 1997, the Serra do Navio deposit was one of the most important sources of high-grade manganese ore in the world. Tropical lateritic weathering processes from metasedimentary manganese protoliths in the Serra do Navio Formation (Paleoproterozoic) derived the high-grade manganese oxide ores. The ores were composed of manganese oxyhydroxides derived from metasedimentary protore lithologies rich in rhodochrosite, spessartine, rhodonite and tephroite. High-grade manganese oxide ore is composed chiefly of cryptomelane, pyrolusite and manganite. Through processes as dissolution, migration, minerals deposition and replacement formed also other type of manganese oxides, with high contents of iron, silica and aluminum.

Sampling and analysis

Due to lack of financial resources, logistical support, and local political questions, sampling did not totally adhere to the initially proposed plan for around 80 water samples. During the field work, it was only possible to acquire 42 water samples, with 26 from shallow wells inside the residential area, 10 from control drilling wells inside ICOMI's industrial area and 6 from superficial waters and creeks, locally so-called "igarapés". Moreover the results from 7 control drilling well water samples provided by ICOMI were used. Figure 2 exhibits the sampling area with the

Fig. 1 a The location of the state of Amapá; (**b**) and (**c**) the state of Amapá, showing the municipalities of Serra do Navio, where the mining occurred, and Santana, where the manganese ore (pellets/sinter) was shipped
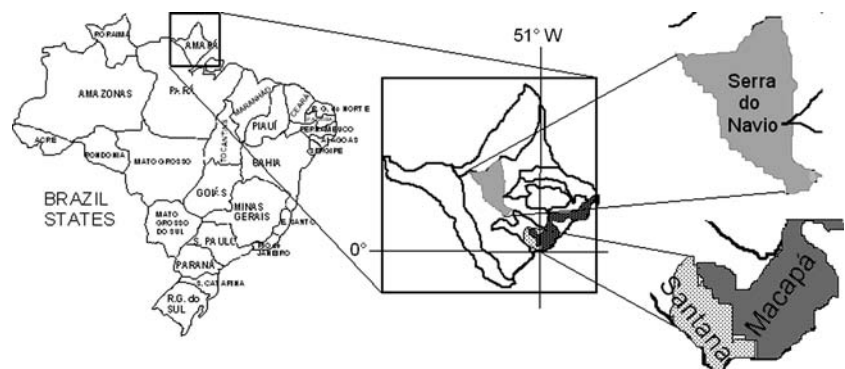
points of collection. For sample gathering a spherical sampling device and polyethylene bottle of 100 ml were used. For the sampling location points a global positioning system, Magellan type, was used. All water samples were analyzed by atomic spectrometry absorption in order to measure the concentrations, in mg/l, of 16 elements, shown in Table 1 (Queiroz 2003).

Geostatistical methods

Geostatistical methods are based on regionalized variable theory that states that variables in an area exhibit both random and spatially structured properties. The general assumption in spatial geostatistics is that the regionalized variable is second-order stationary.

Let $z_i(\mathbf{u})$, $i = 1,\dots,p$, where $\mathbf{u}$ is the vector of location coordinates for the regionalized variables p (i.e., variables related to some location space $\mathbf{u}$-measurement at all locations). In this case, $Z_i(\mathbf{u})$; $i = 1,\dots,p$, is a random function defined in the random variables set in a study area. A spatial increment $(z_i(\mathbf{u}) - z_i(\mathbf{u} + \mathbf{h}))$ is defined as the difference between z_i-values at sites $\mathbf{u}$ and $\mathbf{u} + \mathbf{h}$ for an interval lag distance and direction class $\mathbf{h}$-vector. Assuming that the variable is second-order stationary, the two moments: mean and variance, are independent of location $\mathbf{u}$ and depend only on the $\mathbf{h}$-vector.

The fundamental relations are

Mean-value vector:

$$m = E\{\mathbf{Z}(\mathbf{u})\}$$

Co-variance matrix:

$$\mathbf{C}(\mathbf{h}) = E[\{\mathbf{Z}(\mathbf{u}) - m\}^T \{\mathbf{Z}(\mathbf{u} + \mathbf{h}) - m\}]$$

Semi-variogram matrix:

$$\Gamma(\mathbf{h}) = 1/2 E[\{Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})\}^T \{Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})\}]$$

where transposed matrices are indicated by superscript T. For $\mathbf{h} = 0$ (lag zero), the co-variance $\mathbf{C}(\mathbf{h})$-matrix is equal to the classic variance–covariance $\mathbf{V}$-matrix:

$$\mathbf{C}(0) = E[\{\mathbf{Z}(\mathbf{u}) - m\}^T \{\mathbf{Z}(\mathbf{u}) - m\}] = \mathbf{V} \qquad (1)$$

Also, $\mathbf{C}(\mathbf{h})$ and $\Gamma(\mathbf{h})$ are related by the expression

$$\Gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - 1/2(\mathbf{C}(\mathbf{h}) + \mathbf{C}(-\mathbf{h})) \qquad (2)$$

The experimental semivariogram $\hat{\Gamma}(h)$ -matrix is a $(p \times p)$-matrix in which the diagonal and off-diagonal are, respectively, the direct values and cross semivariogram values for a lag $\mathbf{h}$:

$$\hat{\Gamma}(\mathbf{h}) = \begin{bmatrix} \hat{\gamma}_{11}(\mathbf{h}) & \cdots & \hat{\gamma}_{1p}(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ \hat{\gamma}_{p1}(\mathbf{h}) & \cdots & \hat{\gamma}_{pp}(\mathbf{h}) \end{bmatrix} \qquad (3)$$

An experimental semivariogram $\hat{\gamma}_{ij}(\mathbf{h})$ is computed as

$$\hat{\gamma}_{ij}(\mathbf{h}) = \frac{1}{2N_h} \sum_{\alpha=1}^{N_h} \{z_i(\mathbf{u}_\alpha) - z_i(\mathbf{u}_\alpha + \mathbf{h})\}\{z_j(\mathbf{u}_\alpha) - z_j(\mathbf{u}_\alpha + \mathbf{h})\} \qquad (4)$$

where N_h represents the number of pairs for a lag distance and direction class, $\mathbf{h}$.

In most cases, the experimental semivariogram is fitted by combining theoretical models. This can occur when data are available for a large extension and the experimental semivariogram along that extension can reveal several scales of spatial variability. Each variability scale can be represented by a semivariogram model so that the space variability is modeled by the sum of the semivariograms (nested). Multivariate fatorial kriging allows us to analyze
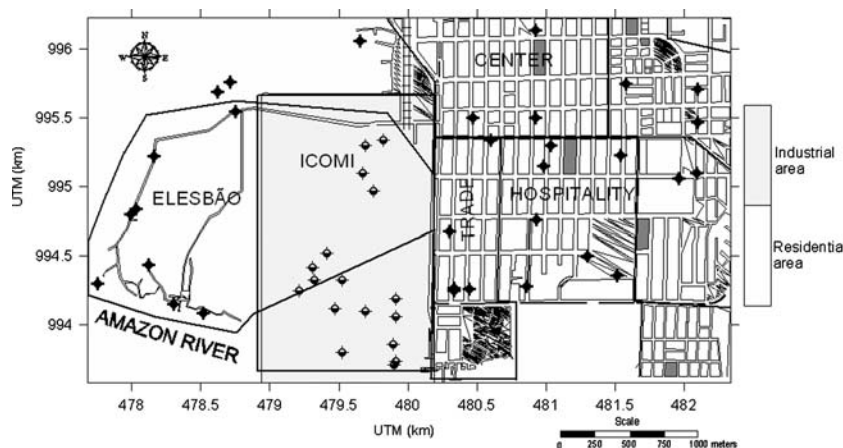
Fig. 2 Samples location map

Table 1 Statistics values for each element and maximum values allowed (in mg/l) for drinking water

Element	CONAMA (mg/l)	Measured values						
		Mean	Median	Standard deviation	Skewness	Min	Max	Pollution (%)
Fe	0.300	6.22	0.12	17.81	3.67	0.000	**85.41**	36.6
Mn	0.100	2.32	0.03	8.59	5.05	0.002	**51.44**	27.0
As	0.050	1.28	0.00	4.75	3.85	0.000	**22.92**	9.80
Al	0.100	1.45	0.21	4.19	3.27	0.008	**21.78**	75.5
Se[b]	0.010	1.60	0.00	1.75	0.18	0.000	**0.041**	46.3
Pb[b]	0.030	1.83	1.27	0.91	0.38	0.008	**0.04**	4.90
Cu[b]	0.020	0.36	0.00	0.68	3.35	0.000	**0.03**	4.90
Cd[a]	0.001	0.32	0.00	0.66	2.90	0.004	**0.003**	7.30
Ba	1.000	0.11	0.04	0.22	3.38	0.000	0.989	0.00
Sr		0.10	0.03	0.19	3.26	0.000	0.94	0.00
Mo		0.04	0.00	0.16	4.36	0.000	0.88	0.00
Zn	0.180	0.02	0.00	0.02	0.64	0.000	0.07	0.00
Co[b]	0.200	0.29	0.17	0.46	4.11	0.000	0.03	0.00
Ni[b]	0.025	0.23	0.12	0.41	2.99	0.018	0.02	0.00
Cr	0.050	0.14	0.00	0.31	4.40	0.000	0.002	0.00
Ag	0.010	0.20	0.00	0.35	2.27	0.000	0.001	0.00

Elements that show concentration levels above those established by the Brazilian environmental legislation are shown in bold

[a] (x 0.001 mg/l)

[b] (x 0.01 mg/l)

the relationships between $Z_i(\mathbf{u})$ in the spatial scales detected by the experimental semivariograms (nested).

One of the major practical difficulties found in multivariate factorial kriging is modeling the $p(p + 1)/2$ experimental semivariograms. To ensure that the variances of all finite linear combinations of the random $Z_i(\mathbf{u})$ functions are positive, the semivariogram functions-matrix must be conditionally negative semi-definite (Journel and Huibregts 1978). The formulation that has received the most attention is the linear model of coregionalization (LMC) which assumes that the original set of random correlated functions $\{Z_i(\mathbf{u}), i = 1,\ldots, p\}$ can be decomposed into an uncorrelated random functions $\{Y_k^u(\mathbf{u}), k = 1,\ldots, p; s = 1,\ldots,S$ (number of different space scales)$\}$ with transformation coefficients a_{is}^u:

$$Z_i(\mathbf{u}) = \sum_{s=0}^{S}\sum_{k=1}^{p} a_{ik}^u Y_k^u(\mathbf{u}) \qquad (5)$$

where $Y_k^u(\mathbf{u})$ are called regionalized factors with k denoting the different factors at a given spatial scale s. For a fixed scale s, the p regionalized factors have the same semivariogram function $\gamma_s(h)$. Also, $Y_k^u(\mathbf{u})$ are mutually orthogonal by construction:

$$1/2E[\{Y_k^u(\mathbf{u}) - Y_k^u(\mathbf{u}+\mathbf{h})Y_{k'}^{u'}(\mathbf{u}) - Y_{k'}^{u'}(\mathbf{u}+\mathbf{h})\}] = \gamma_s(\mathbf{h}),$$
$$\text{if } s = s' \text{ and } k = k'$$

and $= 0$ otherwise.

The direct and cross semivariograms $\gamma_{ij}(\mathbf{h})$ can be expressed as a linear combination of the basic semivariogram $\gamma_s(\mathbf{h})$ with coefficients b_{ij}^u:

$$\gamma_{ij}(\mathbf{h}) = 1/2E[\{Z_i(\mathbf{u}) - Z_i(\mathbf{u}+\mathbf{h})\}\{Z_j(\mathbf{u}+\mathbf{h}) - Z_j(\mathbf{u}+\mathbf{h})\}]$$
$$= \sum_{u=1}^{S}\sum_{u'=1}^{S}\sum_{k=1}^{p}\sum_{k'=1}^{p} a_{ik}^s a_{ik'}^{s'} 1/2E[\{Y_k^s(\mathbf{u})$$
$$- Y_k^s(\mathbf{u}+\mathbf{h})\}\{Y_{k'}^{s'}(\mathbf{u}) - Y_{k'}^{s'}(\mathbf{u}+\mathbf{h})\}]$$
$$= \sum_{s=0}^{S}\sum_{k=1}^{p} a_{ik}^s a_{jk}^s \gamma_s(\mathbf{h}) = \sum_{s=0}^{S} b_{ij}^s \gamma_s(\mathbf{h})$$
$$(6)$$

By grouping the regionalized factors $Y_k^s(\mathbf{u})$ with their respective semivariogram function $\gamma_s(\mathbf{h})$, each random function $Z_i(\mathbf{u})$ can be expressed as the sum of random orthogonal functions $Z_i^s(\mathbf{u})$ called spatial components:

$$Z_i(\mathbf{u}) = \sum_{s=0}^{S} Z_i^s(\mathbf{u}) \quad \text{with} \quad Z_i^s(\mathbf{u}) = \sum_{k=1}^{p} a_{ik}^s Y_k^s(\mathbf{u}) \qquad (7)$$

where $Z_i^s(\mathbf{u})$ represents the behavior of the random function $Z_i(\mathbf{u})$ at a given spatial scale s. Eq. 6 then becomes:

$$\gamma_{ij}(\mathbf{h}) = \sum_{s=0}^{S} \gamma_{ij}^s(\mathbf{h}) \quad \text{with} \quad \gamma_{ij}^s(\mathbf{h}) = b_{ij}^s \gamma_s(\mathbf{h}) \qquad (8)$$

where $\gamma_{ij}^s(\mathbf{h})$ are the cross semivariograms between spatial components $Z_i^s(\mathbf{u})$ and $Z_j^s(\mathbf{u})$.

The linear model of coregionalization, Eqs. 6 and 7 can be rewritten using matrix notation:

$$Z(\mathbf{u}) = \sum_{s=0}^{S} Y^s(\mathbf{u})(A^s)^T. \tag{9}$$

$$\mathbf{B}^s = \mathbf{A}^s(\mathbf{A}^s)^T; \tag{10}$$

with $Z(\mathbf{u}) = [Z_1(\mathbf{u}), \ldots, Z_p(\mathbf{u})]$, $Y^s(\mathbf{u}) = [Y_1^s(x), \ldots, Y_p^s(x)]$ and $\mathbf{A}^s(i,j) = a_{ij}^s$, and $\mathbf{B}^s$ is a positive semi-definite matrix of coefficients, b_{ij}^s, called the coregionalization matrix.

The linear correlation between any two spatial components $Z_i^s(\mathbf{u})$ and $Z_j^s(\mathbf{u})$ is then measured by the structural correlation coefficient:

$$\rho_{ij}^s = \text{Corr}\{Z_i^s(\mathbf{u}), Z_j^s(\mathbf{u})\} = \frac{b_{ij}^s}{\sqrt{b_{ii}^s \cdot b_{jj}^s}}. \tag{11}$$

The structural correlation coefficients matrix is defined as $\mathbf{R}_s = [\rho_{ij}^s]$.

Multivariate factorial kriging accounts for the regionalized nature of variables by analyzing the coregionalization $\mathbf{B}^s$-matrix or structural correlation coefficients $\mathbf{R}_s$-matrix separately. By so doing, each correlation structure is distinguished by filtering the structures belonging to the other spatial variation scales.

The practical implementation of multivariate factorial kriging involves the following procedures:

1. A coregionalization must be modeled that, in practice, involves the following steps:
   (i) Compute the $p(p+1)/2$ direct and cross experimental semivariograms;
   (ii) Chose the number (S) and characteristics (type and range) of the basic semivariogram functions $\gamma_s(\mathbf{h})$. Webster et al. (1994) recommend that before fitting a coregionalization model, Principal Components Analysis be performed on the variance–covariance $\mathbf{V}$-matrix (or correlation $\mathbf{R}$-matrix) of the original variables, and the spatial correlation structure of the principal components should be evaluated by computing their semivariograms.
   (iii) Fit the selected model to the experimental values, i.e., estimate the b_{ij}^s coefficients under the constraint of positive semi-definiteness of the coregionalization matrices-$\mathbf{B}^s$ (Pardo-Igúzquiza and Dowd 2002).

2. Multivariate methods, such as Principal Components Analysis, must then be applied to each matrix $\mathbf{B}^s$. Elements of each coregionalization matrix $\mathbf{B}^s$ thus represent the relative contributions of the basic model, $\gamma_s(\mathbf{h})$, to direct and cross semivariograms. A Principal Components Analysis in the coregionalization matrix $\mathbf{B}^s$ would lead to the following spectral decomposition:

$$\mathbf{B}^s = \mathbf{Q}^s \Lambda^s (Q^s)^T = \mathbf{A}^s(\mathbf{A}^s)^T \quad \text{with} \quad \mathbf{A}^s = \mathbf{Q}^s(\mathbf{\Lambda}^s)^{1/2} \tag{12}$$

where $\mathbf{Q}^s = [q_{ik}^s]$ is the orthogonal matrix of eingenvectors and $\Lambda^s = \lambda_k^s$ is the diagonal matrix of eigenvalues. The correlation coefficient between spatial component $Z_i^s(\mathbf{u})$ and t regionalized factor $Y_k^s(\mathbf{u})$ is

$$\rho_{ik}^s = \frac{q_{ki}\sqrt{\lambda_k^s}}{\sqrt{b_{ii}^s}}. \tag{13}$$

where b_{ii}^s is the sill of the basic model $\gamma_l(\mathbf{h})$ of the direct semivariogram. The correlation representation graph indicates that interrelations between metals change as a function of the spatial scale.

3. The regionalized factors $Y_k^s(\mathbf{u})$ are estimated by ordinary cokriging as (Pardo-Igúzquiza and Dowd 2002):

$$\hat{Y}_k^s(\mathbf{u}) = \sum_{i=1}^{p} \sum_{\alpha=1}^{n} \lambda_{i\alpha}(\mathbf{u}) Z_i(\mathbf{u}_\alpha) \tag{14}$$

where $\hat{Y}_k^s(\mathbf{u})$ is the ordinary cokriging estimator of the k-th regionalized factor at the s-th spatial scale, p is the number of variables, n is the number of data points surrounding $\mathbf{u}_\alpha$ used in the estimate, and $\lambda_{i\alpha}(\mathbf{u})$ is the weight assigned to the datum $Z_i(\mathbf{u}_\alpha)$. Minimizing the estimation subject to the unbiased conditions yields the following system of $p(n+1)$ linear equations, whose solution supplies the weights $\lambda_{i\alpha}(\mathbf{u})$:

$$\sum_{i=1}^{p}\sum_{\beta=1}^{n} \lambda_{\beta i}\gamma_{ij}(\mathbf{u}_\alpha - \mathbf{u}_\beta) - \mu_i = a_{ik}^s \gamma_s(\mathbf{u}_\alpha - \mathbf{u}_0)$$
$$\sum_{\beta=1}^{n} \lambda_{\beta i} = 0 \quad \text{for } i = 1, \ldots, p \text{ and } \alpha = 1, \ldots, n. \tag{15}$$

The estimator is unbiased as a consequence of the conditions imposed on the weights. The condition that the sum of the weights be zero means that the local mean of each regionalized factor is a set of zeros.

4. In the same way, an unbiased estimate of a spatial component may be obtained by cokriging (Pardo-Igúzquiza and Dowd 2002):

$$\hat{Z}_i^s(\mathbf{u}_0) = \sum_{i=1}^{p} \sum_{\alpha=1}^{n} \lambda_{i\alpha}(\mathbf{u}) Z_i(\mathbf{u}_\alpha). \tag{16}$$

The system to be solved is

$$\sum_{i=1}^{p}\sum_{\beta=1}^{n}\lambda_{\beta i}\gamma_{ij}(\mathbf{u}_\alpha - \mathbf{u}_\beta) - \mu_i = b_{ik}^s\gamma_s(\mathbf{u}_\alpha - \mathbf{u}_0).$$

$$\sum_{\beta=1}^{n}\lambda_{\beta i} = 0 \quad \text{to} \quad i = 1,\ldots,p \text{ and } \alpha = 1,\ldots,n$$

(17)

where $\gamma_s(\mathbf{u}_\alpha - \mathbf{u}_o)$ is the value of the s-th basic semivariogram between the α-th sampling point and $\mathbf{u}_o$, the point at which the estimate is made.

The elements of $b_{ik}^s\gamma_s(\mathbf{u}_\alpha - \mathbf{u}_o) = \gamma_{ik}(\mathbf{u}_\alpha - \mathbf{u}_o)$ are the cross semivariances between $Z_i(\mathbf{u}_\alpha)$ and $Z_i^s(\mathbf{u}_0)$. The condition where the sum of the weights is zero means that the local mean of each spatial component is a set of zeros.

Results and discussion

Table 1 presents the 16 variables and the statistics for each element. Iron shows a larger average value followed by manganese, aluminum and arsenic. The median of each variable is also presented and it is generally a unlike value from the corresponding mean indicating no variables with normal distributions. That was confirmed by observing that most elements exhibit accentuated positive asymmetry. Some elements presented values inside the limits established by the Brazilian environmental legislation (Table 1). The first column in Table 1 shows the maximum limits allowed by the environmental legislation (CONAMA 1986). Thus this study considered only 8 elements that show concentration levels above those established by the Brazilian environmental legislation. Aluminum presented the largest pollution percentage (75.5%), followed by selenium (46.3%), iron (36.6%), manganese (27%), arsenic (9.8%), cadmium (7.3%) lead (4.9%) and copper (4.9%). These elements were submitted to multivariate factorial kriging analysis. Metal concentrations are to some extent linearly correlated. The correlation matrix, **R**, listing the ordinary product-moment correlation coefficients is shown in Table 2. It shows strong correlations between lead and selenium ($r = 0.965$) and iron and cadmium ($r = 0.934$).

All selected variables are standardized to zero mean and unit variance. Relationships between the selected variables were first studied by the classic method of principal components analysis (PCA) applied to the correlation matrix. Table 3 summarizes the eigenvalues, the factor loadings and the explained variance for the first axis of PCA. The first three principal components account for 74.1% of the total variance. There were strong contributions from lead, selenium, cadmium and iron in first component, evidenced by the high values of the factorial loads for these elements highlighted in boldface in the Table 3. Arsenic and manganese contributed most to the third component, according to the biggest values of the factorial loads highlighted in Table 3, and the most important contributions in the second component were from aluminum, iron and cadmium in opposition to the copper variable (Table 3).

A linear coregionalization model was used to fit the semivariograms of the principal components. All the semivariograms have a nugget effect and show two scales of spatial variation. The solid lines in Fig. 3 show these fitted models, where C_0 is the nugget variance, C_1 is the sill for the first structure (short-range variance), C_2 is the sill for the second structure (large-range variance) and a_1 and a_2 are the amplitudes for the first and second structures, respectively.

Omnidirectional semivariograms were modeled using a nugget effect and two spherical schemes with ranges of 0.33 and 2.0 km, respectively. Both short and large-range structures are apparent in the first component whereas the short-range structure (0.33 km) plays an increasing role for the second and third principal components. All 36 direct and cross experimental semivariograms were computed and modeled as the sum of a nugget effect and two spherical schemes with a range of 0.33 and 2.0 km, respectively. Queiroz (2003) presents the 36 direct and cross experimental semivariograms fitted by linear coregionalization model.

In the selection of contrasting semivariograms for Fe, Cd, Pb, and Se (Fig. 4) the large-range structure (2 km) dominates the Pb and Se semivariograms and the short-range structure (0.33 km) dominates the Fe and Cd

Table 2 Variables correlation matrix

	As	Mn	Al	Fe	Pb	Se	Cd	Cu
As	1.000							
Mn	0.024	1.000						
Al	–0.094	–0.090	1.000					
Fe	0.070	0.166	**0.369**	1.000				
Pb	0.172	**0.336**	0.105	**0.415**	1.000			
Se	0.093	**0.344**	0.096	**0.405**	**0.965**	1.000		
Cd	0.085	0.196	**0.349**	**0.934**	**0.475**	**0.428**	1.000	
Cu	–0.046	–0.006	–0.026	0.050	**0.525**	**0.559**	0.047	1.000

Values in bold typeface: less than 5% significance

Table 3 Loading factors of principal components (PC)

components	As	Mn	Al	Fe	Pb	Se	Cd	Cu	λ	Explained variance (%)	Cumulated explained variance (%)
PC 1	0.142	0.387	0.303	**0.753**	**0.888**	**0.876**	**0.779**	0.460	3.202	40.0	40.0
PC 2	0.073	0.177	**−0.584**	**−0.560**	0.356	0.389	**−0.529**	**0.597**	1.607	20.1	60.1
PC 3	**−0.667**	**−0.508**	0.425	−0.059	0.005	0.066	−0.091	0.467	1.118	14.0	74.1

The high values of the factorial loads for the elements are highlighted in boldface

λ: eigenvalues

Fig. 3 Experimental omnidirectional semivariograms of the three principal components (lines linking points) and the linear coregionalization model fitted semivariograms (*solid lines*)
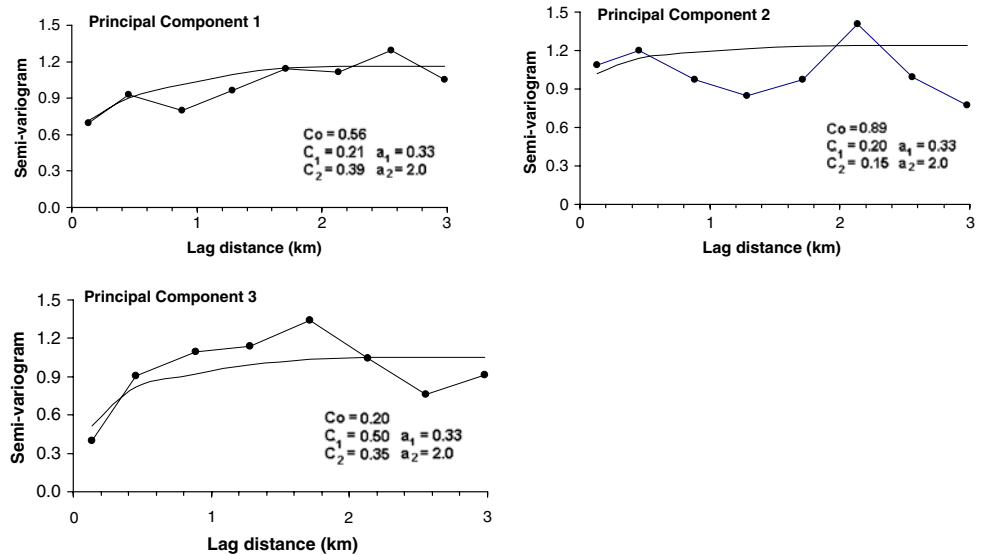
semivariograms. The relative contributions of both structures to the direct semivariograms reflects the role played by human and geological factors in the spatial pattern displayed by each element.

Sill values for the two models, i.e., the b_{ij}^s coefficients, were assembled into two coregionalization matrices $\mathbf{B}^1$ and $\mathbf{B}^2$. These describe the correlation structures of the variable to small (0.33 km) and large (2 km) spatial scales, respectively. The PCA results of these two coregionalization matrices provide the regionalized factors which are the principal components of each coregionalization matrix, in their respective "s" scale and spatial components, $Z_i^s(\mathbf{u})$.

Table 4 presents the results of the Principal Component Analysis of matrices $\mathbf{B}^1$ and $\mathbf{B}^2$. A strong correlation can be observed in iron and cadmium for the $\mathbf{B}^1$-matrix (small spatial scale) with the first regionalized factor. Manganese correlated strongly with the second regionalized factor and arsenic gave the strongest contribution to the third regionalized factor. This suggests a possible relationship between the contamination from these elements and contamination linked to human activities. This source of contamination generally occurs on a small spatial scale, the case described by the $\mathbf{B}^1$-matrix. For the $\mathbf{B}^2$-matrix (large

spatial scale), the first regionalized factor accounts for most of the variability (84.1%), and all variables correlate from 0.415 to 0.735, except copper with 0.097. This is interpreted as a dispersion effect, although lead, selenium and arsenic contribute the most. Copper is more strongly correlated with the second factor, and no single variable significantly contributes to the third factor which explains only 3.2% of the total variability. Aluminum has a stronger correlation with the first regionalized factor on the large scale than to the small scale. This indicates that arsenic, manganese, iron, and cadmium have stronger correlations to the small variability scale, the one linked to human activities. Meanwhile the high concentrations observed for aluminum, lead, selenium, arsenic and copper are due to the natural environment itself, deduced from their stronger correlations with the large scale.

It is clear in Table 4 that the correlation structure changes from one spatial scale to the other. Analyzing the coregionalization matrices brings to light the scale-dependent relationships that could not simply be detected by analyzing the conventional correlation $\mathbf{R}$-matrix (or variance-covariance $\mathbf{V}$-matrix). For example, the correlation between iron and cadmium is stronger in the small scale

Fig. 4 Experimental omnidirectional semivariogram of iron, cadmium, lead and selenium (lines linking points) and the linear model coregionalization fitted semivariograms (*solid lines*)
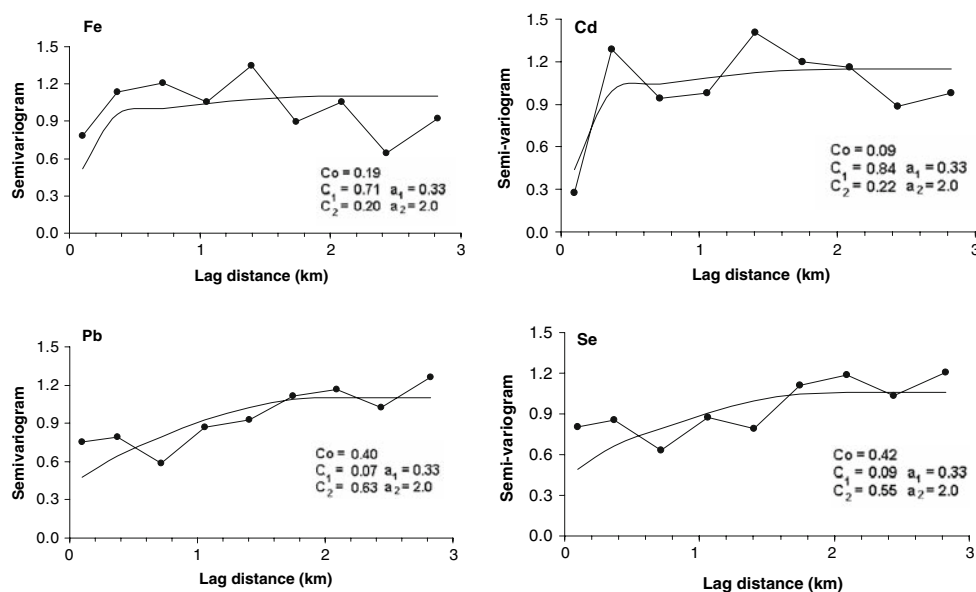

Table 4 Regionalized factors on the small (B^1) and large (B^2) spatial scale

Variables	Matrix B1 regionalized factor loads			Matrix B2 regionalized factor loads		
	Y_1^1	Y_2^1	Y_3^1	Y_1^2	Y_2^2	Y_3^2
As	0.000	–0.003	**0.517**	**0.663**	0.098	–0.116
Mn	0.001	**0.817**	0.006	0.447	–0.044	–0.144
Al	0.316	–0.007	–0.011	0.438	0.097	0.025
Fe	**0.803**	–0.002	0.002	0.417	0.077	0.038
Pb	0.223	0.005	–0.003	**0.735**	–0.147	0.092
Se	0.219	0.010	–0.009	**0.685**	–0.147	0.140
Cd	**0.836**	–0.001	0.005	0.415	0.078	–0.116
Cu	–0.006	–0.104	0.033	0.097	**0.520**	0.092
Eigenvalues	1.732	0.734	0.368	2.439	0.354	0.094
% Variance	52.7	22.3	11.2	84.1	12.2	3.2
Accumulated	52.7	75.0	86.2	84.1	96.3	99.5

Values in bold indicate higher factor loads

while selenium and lead have stronger correlation in the large scale. The load values of the first regionalized factors of the $\mathbf{B}^1$ and $\mathbf{B}^2$ matrix show more probability of some correlation to occur between arsenic and manganese on the large, rather than the small spatial scale. For the first regionalized factor in the large scale, it can be observed that the contribution, or load, of the arsenic variable is similar to that of the selenium variable. This relation between arsenic and selenium does not appear on the correlation matrix $\mathbf{R}$.

The regionalized factors Y_k^s (**u**) and spatial components Z_i^s (**u**), $s = 1,2$ and $i,k = 1, ..., 8$ were mapped through cokriging. Figures 5 and 6 show maps of the first regionalized factor at the small and large spatial scales, respectively. The value of any regionalized factor equals zero at a distance greater than 0.33 km on the small scale, and at a distance greater than 2.0 km on the large scale.

Regionalized Factor 1 on the small spatial scale, Y_1^1, which mainly denotes iron and cadmium, takes high values in an area belonging to ICOMI indicated by a rectangle (Fig. 5). Iron and cadmium association can be related with the chemical aggregation capacity with manganese present in the mineral deposit. The drainage is plotted on the maps in white colors (Figs. 6 and 8). The spatial pattern related to area drainage agrees well with low values of the regionalized factor 1 at the large spatial scale showing the effect of metal dispersion. This regionalized factor accounts for 84.1% of the variability with reasonable contribution from almost all element variables, except copper.

Arsenic components on the small and large spatial scales are shown in Figs. 7 and 8, respectively. On the small spatial scale, the contribution from arsenic is more relevant in only the third principal component (or third regionalized

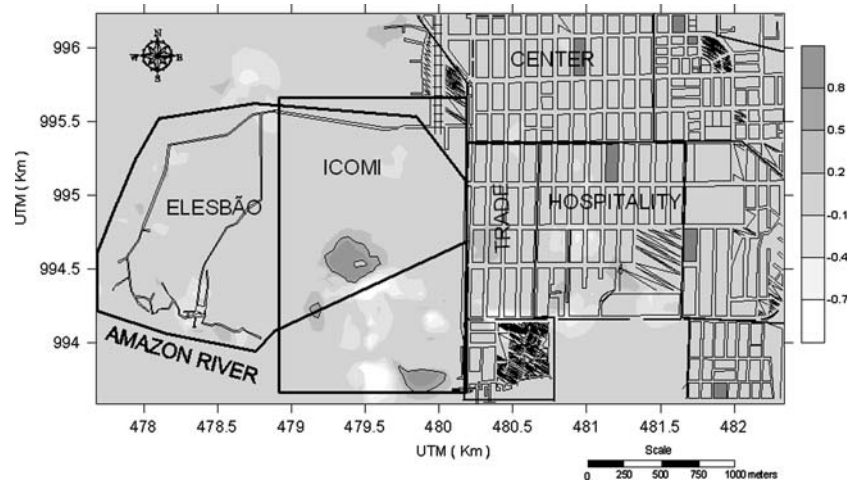**Fig. 5** Plot of first regionalized factor associated with the small spatial-scale spherical model

Fig. 6 Plot of first regionalized factor associated with the large spatial scale spherical model (*white lines* refer to drainage)
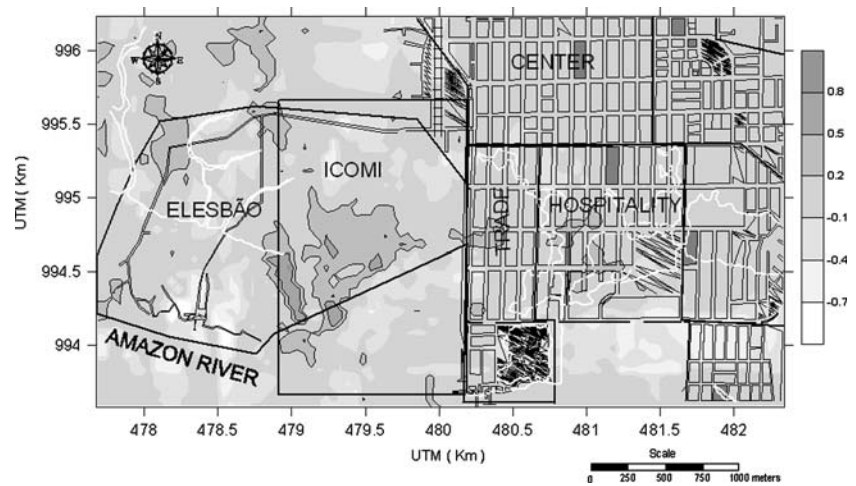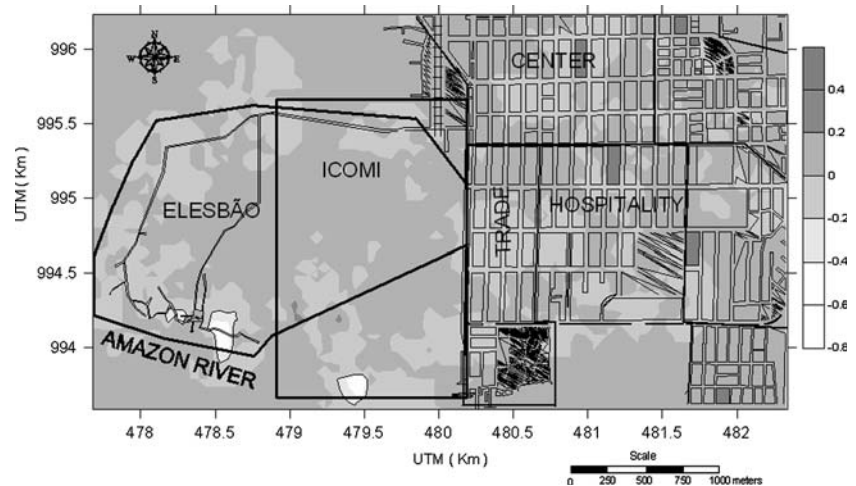

Fig. 7 Plot of the arsenic spatial component associated with the small spatial scale spherical model

factor) which accounts for only 11.2% of the total variance (Table 4). Two small negative anomalies can be seen on the small scale map, one within the ICOMI industrial area and the other in the Elesbão neighborhood by the Amazon River (Fig. 7). Webster et al. (1994) and Goovaerts et al.

(1993) show that low and high values, meaning positive and negative anomalies, observed in the regionalized factors maps are, in general, associated with the geology of the area due to interactions among metals and soil types. Due to the lack of data such as a reliable geology map of the

Fig. 8 Plot of the arsenic spatial component associated with the large spatial scale spherical model (*white lines* refer to drainage)
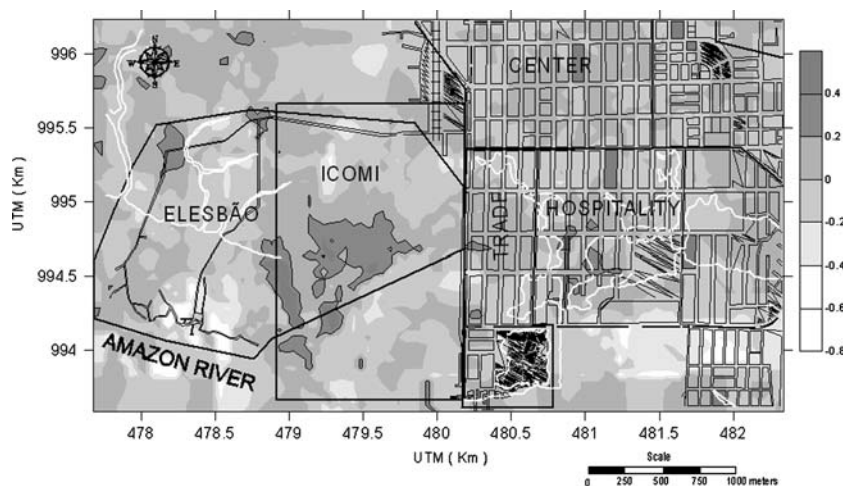
area in study, it was not possible to assess associations between type and soil use and the locations of negative and positive anomalies. On the large spatial scale, the contribution from arsenic is more relevant in the first principal component (or third regionalized factor) which accounts for only 84.1% of the total variance (Table 4). Figure 8, the large spatial scale map shows substantial areas with high positive values in the ICOMI area, the Elesbão neighborhood—to the northwest, and the residential area—in the Hospitality estate, which seem to indicate that arsenic is disseminated across the study area. The large spatial scale map also shows a certain similarity between spatial pattern and study area drainage.

Conclusions

Environmental characterization is fundamental in evaluating risk analysis. Multivariate factorial kriging yields a regionalized factor set that summarizes the main features of data at each spatial scale, thus allowing identification of pollutant sources and patterns and supplying quantitative measurements of the complex interactions between analyzed variables. Modeling semivariograms allows distinction between small and large range variations. The small-scale structure is likely to be due to local sources of contamination linked to human activities. The large-scale structure probably reflects regional changes in the natural environment. Results in this study could probably be improved by information on area geology and analysis of pollutant concentrations at several different soil levels. Even so, the manganese exploration and commercialization activities in Santana have probably contributed to the high concentrations of manganese, arsenic, iron, and cadmium, in the study area. The high concentrations of aluminum, selenium, lead, and copper seem to be more closely related to the natural environment.

References

Castrignanò A, Giugliarini L, Risaliti R, Martinelli N (2000) Study of spatial relationships among some soil physico-chemical properties of a field in central Italy using multivariate geostatistics. Geoderma 97:39–60

Chiles JP, Guilen A (1984) Variogrammes et krigeages pour la gravimétrie et reads magnétisme. La Sci Terre 20:455–468

CONAMA/Brazilian National Council of Environment (1986) Evaluation of environmental impact: social agents, procedures and tools. Brasilia/DF

Galli A, Gerdil-Neuillet F, Dadou C(1984) Factorial kriging analysis a substitute to spectral analysis of magnetic dates. Geostat Nat Resour Charact NATO-ASI 122(1):543–558

Goovaerts P (1991) Étude des relations entre propriétés physicochimiques du sol par la géostatistique multivariable. Compte Rendu Journées Géostat 1:247–261

Goovaerts P, Sonnet P, Navarre A (1993) Factorial kriging analysis of springwater contents in the Dyle River Basin, Belgium. Water Resour Res 29(7):2115–2125

Goulard M (1989) Inference in a coregionalization model. Geostatistic 1:397–408.

IBGE/Brazilian Institute of Geography and Statistics (2000) Geografia do Brasil-Região Norte, vol 1. Rio de Janeiro

Journel AG, Huijbregts C (1978) Mining geostatistics. Academic, London

Lin YP, Chang TK, Shih CW, Tseng CH (2002) Factorial and indicator kriging methods using a geographic information system to delineate spatial variation and pollution sources of soil heavy metals: Environ Geol 42:900–909

Matheron G (1982) Pour une analyse krigeante des données régionalisées. Rep. N-732, Centre de Géostatistique, Fontanebleaus, France

Pardo-Igúzquiza and Dowd PA (2002) FACTOR2D: a computer program for factorial kriging. Comput Geosci 28:857–875

PCA/Plane for Environmental Control (2001) Report elaborated by AMPLA ENGENHARIA/ICOMI

Queiroz JCB (2003) Utilização da geoestatística na quantificação do risco de contaminação por metais pesados, na área portuária de Santana–Amapá–Brazil [in Portuguese]. PhD thesis, São Paulo State University, Rio Claro, São Paulo

Rouhani S, Wackernagel H (1990) Multivariate Geostatistical approach to space–time data analysis. Water Resour 26(4):585–591

Sandjiv L (1984) The factorial kriging analysis of regionalized data. Its application to geochemical prospecting. Geostat Nat Resour Charact NATO-ASI 122(1):559–572

Wackernagel H, Butenuth C (1989) Caractérisation d'anmalies géochimiques par la géostatistique multivariable. J Geochem Explor 32:437–444

Webster R, Atteia O, Dubois J-P (1994) Coregionalization of trace metals in the soil of the Swiss Jura. Eur J Soil Sci 45:205–218