

A multivariate approach for anomaly separation of potentially toxic trace elements in urban and peri-urban soils: an application in a southern Italy area

Ilaria Guagliardi · Gabriele Buttafuoco · Domenico Cicchella · Rosanna De Rosa

Received: 11 March 2012 / Accepted: 3 August 2012 / Published online: 28 August 2012
© Springer-Verlag 2012

Abstract

Purpose Geogenic soil enrichment and anthropogenic pollution by potentially toxic trace elements (PTEs) are two processes acting together. Although it is often difficult, it is necessary to separate the two processes for risk assessment and understanding the environmental implications. The aim of this study was to analyse the soil concentrations of various PTEs in a southern Italy area in order to: (1) determine their different correlation structure to isolate sources of variation acting at different spatial scales and (2) to define potential anomalies based on the correlation structure.

Materials and methods In the urban and peri-urban area of Cosenza-Rende, 149 topsoil samples were collected (0.10 m) and analysed for different elements by X-ray fluorescence spectrometry. Principal component analysis and factorial kriging analysis were used to map the spatial distribution of PTEs in topsoil and to identify the main factors influencing their spatial variability.

Results and discussion Two groups of PTEs were identified: the first group included As, Pb and Zn; and the second one

Al, Co, Cr, Fe, La, Nb, Ni, Ti and V. The first group was related to anthropogenic causes, while the second one was more related to parent rock composition. The regionalized factors at different scales of variability allowed to aggregate and summarize the joint variability in the PTEs and consider the probable causes of soil pollution.

Conclusions The study allowed analysing and quantifying the sources (environmental or anthropogenic) of variation of PTEs acting at different spatial scale and defining the spatial anomalies based on the correlation structure associated at the different spatial scales.

Keywords Factorial kriging analysis · Pollution · Principal component analysis · Urban and peri-urban soil

1 Introduction

Natural concentrations of potentially toxic trace elements (PTEs) in soils mainly depend on the chemistry of the parent materials from which the soils are derived (Mirsal 2008; Hooda 2010; Wang et al. 2011). Anthropogenic activity may greatly increase the PTEs' concentrations in the environment, and the amount of anthropogenic pollutants released often exceeds contributions due to natural sources; as a matter of the fact, concentrations found in urban and peri-urban soil are, therefore, likely to be much higher than those found in soil from rural areas (Markus and McBratney 1996; Mirsal 2008). Sources of PTEs related to human activities in urban and peri-urban soils are traffic emissions, industrial discharges and urban development processes (Lado et al. 2008; Mahanta and Bhattacharyya 2011; Mirsal 2008; Norra et al. 2001; Odewande and Abimbola 2008). Among these sources, vehicular emissions are widely understood to be a significant and increasing source of soil and environmental pollution (Liu et al. 2009). Among pollutants, heavy metals have been the

Responsible editor: Willie Peijnenburg

I. Guagliardi · G. Buttafuoco (✉)
CNR—Institute for Agricultural and Forest Systems
in the Mediterranean (ISAFOM),
Via Cavour 4/6,
87036 Rende, CS, Italy
e-mail: gabriele.buttafuoco@cnr.it

D. Cicchella
Department of Geological and Environmental Studies,
University of Sannio,
Via dei Mulini, 59/A,
82100 Benevento, Italy

R. De Rosa
Department of Earth Sciences, University of Calabria,
Ponte Bucci,
87036 Rende, CS, Italy

subject of particular attention because of their long-standing toxicity, especially when exceeding specific thresholds (Coşkun et al. 2006).

Environmental monitoring of soil is of great immediate concern because of its effect on ecosystem (Schröder et al. 2004). It is mainly aimed to detect zones with both high concentrations of PTEs and relevant spatial extension in order to identify the geochemical anomalies of the territory (Wackernagel 2003) and the potential threat they represent to the ecosystem. Thus, it is necessary to distinguish if the high environmental concentrations of PTEs may be explained by the natural geochemical background levels or by the anthropogenic pollution (Cicchella et al. 2008).

Geostatistics (Matheron 1973) provides a valuable tool to study the spatial distribution of PTEs. It takes into account the spatial autocorrelation of data to create mathematical models to explain the spatial correlation of variables by variograms, which try to relate the data variance between two locations to their separation distance. The variogram is a model of spatial dependence and enables the estimation of the variable at unsampled locations by using an interpolation technique, known as kriging, which provides the ‘best’, unbiased, linear estimate of a regionalized variable in an unsampled location, where ‘best’ is defined in a least-squares sense (Chilès and Delfiner 1999). Geostatistical techniques are commonly used to generate soil maps and have been described in many texts (Chilès and Delfiner 1999; Goovaerts 1997; Webster and Oliver 2007; among many others).

Principal component analysis (PCA) is an appropriate method for both analysing large sets of correlated multivariate data and identifying the relation among the variates and then clustering (Wackernagel 2003; Webster 2001) when several attributes are determined on each soil sample. PCA allows to obtain several components, such as linear combinations of original PTEs. Some of these can then be used for studying a specific group of variables, providing information on the association of elements which is geochemically more significant than the study of individual variables (Jimenez-Espinosa et al. 1993). Many examples have already indicated the suitability of PCA applications in studies on soil science (Borůvka et al. 2005; Carroll and Oliver 2005; Sánchez-Marañón et al. 2011; Visconti et al. 2009; among many others). PCA does not take into account any spatial correlation that may exist between variables values, and to overcome these shortcomings, Matheron (1982) proposed a geostatistical method known as factorial kriging analysis (FKA), which allows to distinguish the correlation structure of the multivariate data at different spatial scale and yields a set of regionalized factors summarizing the main features of the data at each spatial scale (Goovaerts 1997). There are examples of geostatistical applications in studies on soil concentration of PTEs (Atteia et al. 1994;

Brus et al. 2002; Goovaerts and Webster 1994; Goovaerts et al. 1997; Juang et al. 2001; Lin et al. 2002; McGrath et al. 2004; Queiroz et al. 2008; Reis et al. 2007; Sollitto et al. 2010; Webster et al. 1994). To cite a few from the many in urban and peri-urban soil: Markus and McBratney (1996) investigated the occurrence and the spatial distribution of Pb, Zn, Cu and Cd in the topsoil of an inner-city part district area of Sydney (Australia); Cattle et al. (2002) evaluated different kriging methods for assessing the spatial distribution of lead contamination in urban soil and delineating between contaminated and uncontaminated soil; Saby et al. (2006) used a geostatistical approach for lead assessment in soil around Paris to distinguish Pb due to widespread pollution from geochemical background; Shi et al. (2008) identified the source of metals in urban soils and roadside in Shanghai through geostatistical and multivariate statistical analysis, Zawadzki and Fabijańczyk (2008) analysed the spatial distribution of lead content in urban soils using ordinary kriging and sequential Gaussian simulation; Zheng et al. (2008) identified the spatial variability and the main sources of heavy metals in Beijing soils by conducting multivariate statistical analyses including geostatistical analysis assisted with GIS tools; Alary and Demeougeot-Renard (2010) applied factorial kriging analysis to Cd and Zn concentrations in riverbed sediments to decompose the two variables into components of different spatial scales and map them separately; and Maas et al. (2010) used a multivariate geostatistical approach to identify the source of pollution in urban, suburban and agricultural soils in Annaba (Algeria).

The aim of this study was to analyse the soil concentrations of various PTEs in the urban and peri-urban area of Cosenza-Rende in order to: (1) determine their different correlation structure to isolate and display sources of variation acting at different spatial scales and to (2) define potential spatial anomalies based on the correlation structure associated at the different spatial scales.

2 Materials and methods

2.1 Case study

The study area, covering the Cosenza and Rende municipalities’ territory, is located in southern Italy (Calabria region; Fig. 1). Geologically (Fig. 2), the study area corresponds to a tectonic depression extending over 92 km² characterized by a tick succession of Pliocenic sediments made up of light brown and red sands and gravels, blue gray silty clays and silt interlayers; sediments overlap a Palaeozoic intrusive–metamorphic complex formed by paragneiss, biotite schists, gray phyllitic schists with quartz, chlorite and muscovite which, in some cases, are in the process of weathering (Le Pera et al. 2001).

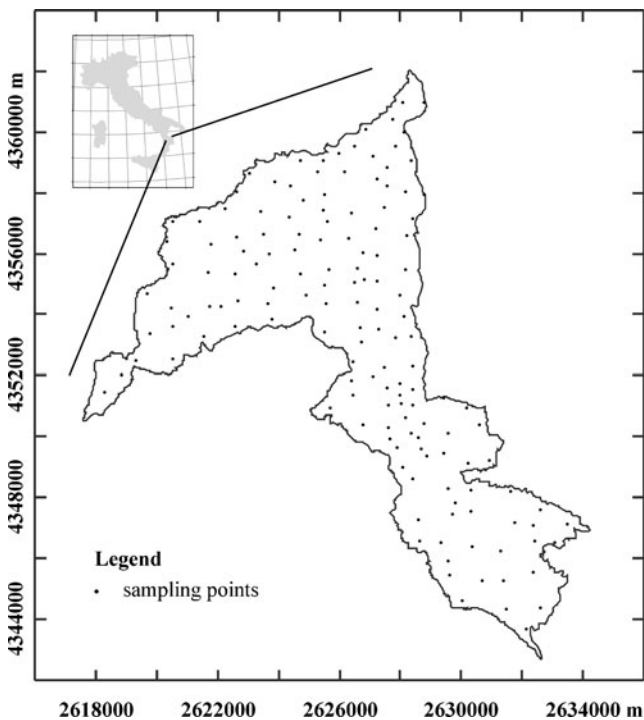
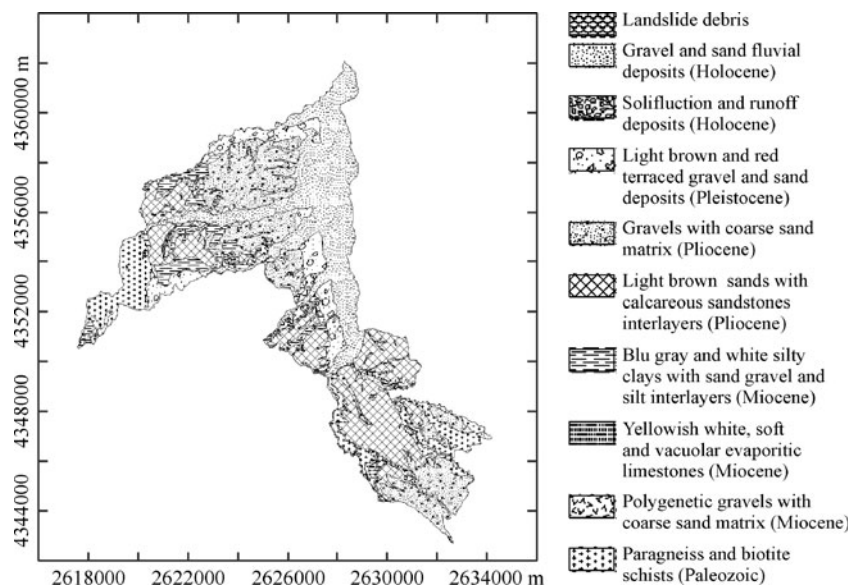


Fig. 1 Study area and sampling locations

Figure 3 shows the soil map (after ARSSA 2003) based on the World Reference Base for Soil Resources (FAO 1998). Soils ranged between fluvisols, luvisols, cambisols, vertisols, calcisols, arenosols, leptosols, umbrisols and phaeozems. Properties, dynamics and functions of the studied soils are highly variable. Their average values were: 17.59 % for clay content, 56.50 % for sand content, 6.84 for pH, 2.86 % for organic matter, 0.25 mS cm⁻¹ for electrical conductivity, 16.14 meq 100 g⁻¹ for CEC and 1.24 g cm⁻³ for bulk density.

Fig. 2 Geology of the study area



From a geomorphologic point of view, the study area is characterized by a flat part, including the urban area, surrounded by hills. The area is characterized by high population (more than 100,000 inhabitants with a density of about 1.900 inhabitants per km²) and heavy traffic, and there is a little industrial activity. Moreover, it includes parks, gardens and agricultural activity in the peri-urban areas.

2.2 Soil sampling and analytical methods

Samples were collected at 149 locations (see Fig. 1) chosen from topsoil in gardens, parks, flower beds and agricultural fields. At each location, topsoil samples (0–0.10 m depth from the surface) were taken from five locations at the corners and at the centre of a square with a side 20 m in length with a hand auger and combined to form a bulked sample. In addition, two duplicate pairs from every ten locations were collected and split in the laboratory to give replicates.

All samples of soil were dried, disaggregated, sieved through a 2-mm mesh sieve, and portions of these soil samples were ground in a mechanical agate grinder until fine particles were obtained. The total concentrations of PTEs were determined in each sample by X-ray fluorescence spectroscopy (XRF) by using a XRF spectrometer Philips PW 1480. The determinations included aluminium (Al), arsenic (As), barium (Ba), cobalt (Co), chromium (Cr), iron (Fe), lanthanum (La), niobium (Nb), nickel (Ni), lead (Pb), titanium (Ti), vanadium (V), yttrium (Y) and zinc (Zn).

The analytical quality control procedures undertaken to assess the precision and accuracy of the laboratory determinations were performed using certified international reference materials AGV-1, BCR-1, BR, DR-N, GA, GSP-1, NIM-G and analysis duplicates. The errors of the estimate

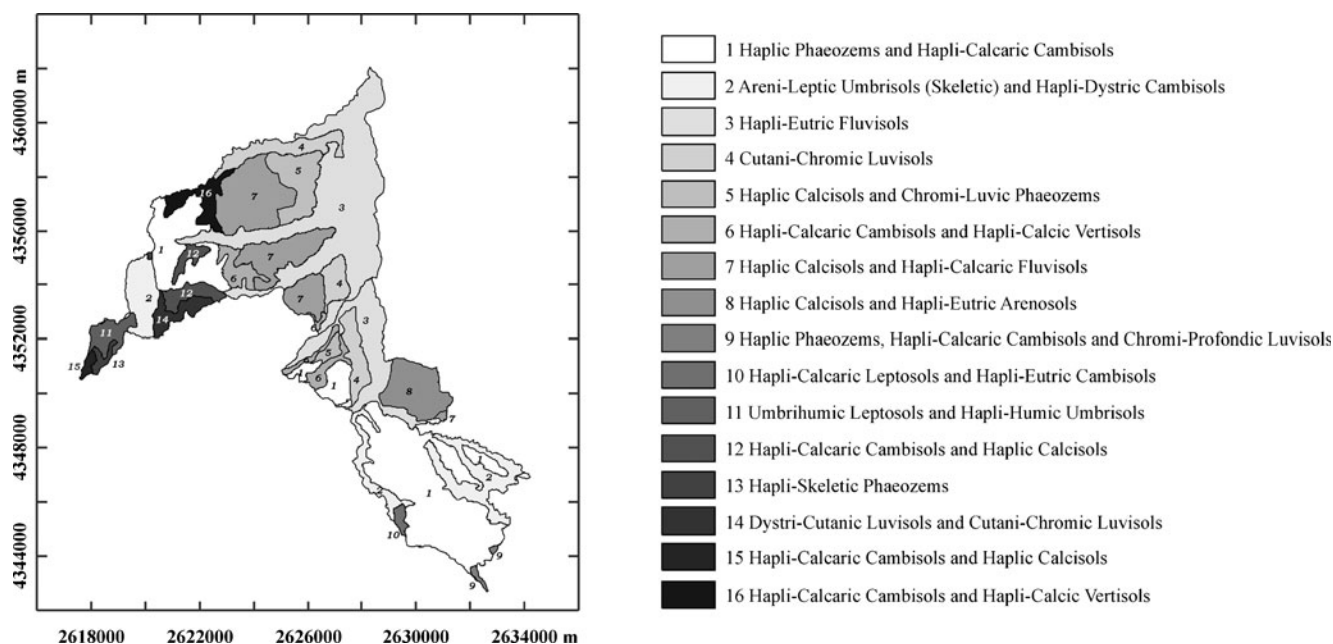


Fig. 3 Soil map of the study area

for the measured elements were determined by the relative standard deviation (<5 %) based on three replicates of one sample randomly chosen.

2.3 Principal component analysis

Principal component analysis (PCA) was introduced by Karl Pearson (1901), and it is the widely used method of multivariate data analysis owing to the simplicity of its algebra and to its straightforward interpretation (Wackernagel 2003). PCA consists in (1) transforming a set of variables Z_i ($i=1, \dots, N$) linearly correlated into uncorrelated principal components (PCs) Y_p ($p=1, \dots, N$) which partition optimally the total variance, and (2) ordering the PCs by decreasing explained variance. PCs are nothing more than the eigenvectors of a variance–covariance matrix or a correlation matrix (Davis 2002), and each PC extracts a maximal share of the total variance. The correlation c_{ij} between the original variables and a pair of principal components can be shown in a graphic display called circle of correlations. The correlation c_{ij} was computed using the following equation:

$$c_{ij} = a_{ij} \sqrt{(\lambda_j / \sigma_i^2)} \quad (1)$$

where a_{ij} is the i th element of the j th eigenvector, λ_j is the j th eigenvalue and σ_i^2 is the variance of the i th original variable. The coordinates of the variables are obtained using the values of correlation of the PTEs with the first (ordinate axis) and the second (coordinate axis) principal component. The intention underlying the use of PCA in geochemical exploration has generally been to separate

the associations inherent in the structure of the correlation matrix into a number of groups of elements that together account for the greater part in the original data (Jimenez-Espinosa et al. 1993).

2.4 Factorial kriging analysis

FKA is a geostatistical method developed by Matheron (1982). The theory underlying FKA (Goovaerts 1992; Goovaerts and Webster 1994; Wackernagel 2003) consists of decomposing the set of original second-order random stationary variables $\{Z_i(\mathbf{x}_\alpha), i=1, \dots, N; \alpha=1, \dots, n\}$ into a set of reciprocally orthogonal regionalized factors $\{Y_p^u(\mathbf{x}), p=1, \dots, N; u=1, \dots, S\}$ where S is the number of spatial scales, through transformation coefficients a_{ip}^u , combining the spatial with the multivariate decomposition:

$$Z_i(\mathbf{x}_\alpha) = \sum_{u=1}^S \sum_{p=1}^N a_{ip}^u Y_p^u(\mathbf{x}) \quad (2)$$

The basic steps of FKA include: (1) modelling the coregionalization of the set of variables using the linear model of coregionalization (LMC), (2) analysing the correlation structure between the variables, by applying PCA at each spatial scale, to obtain independent factors which synthesize the multivariate information and (3) estimating the values of these specific factors at each characteristic scale using cokriging and mapping them.

LMC was developed by Journel and Huijbregts (1978) and considers all the studied variables as the result of the

Table 1 Basic statistics of potential toxic trace elements in milligrams per kilogram of soil

	Min	Max	Mean	Median	Lower quartile	Upper quartile	St. Dev.	Skewness (-)	Kurtosis (-)	Exp. χ^2 (-)
Al	11.19	23.79	15.89	15.38	13.65	17.48	2.80	0.69	0.04	17.18
As	3.00	22.00	7.48	7.00	5.00	9.00	3.03	1.53	3.47	292.54
Ba	335.00	2,000.00	603.05	592.00	530.00	643.00	153.08	5.36	46.74	39.93
Co	6.00	40.00	17.05	16.00	13.00	20.00	5.91	1.21	2.02	80.48
Cr	46.00	309.00	90.54	86.00	73.00	103.00	31.90	2.90	15.44	38.11
Fe	3.11	10.58	5.47	5.13	4.42	6.16	1.47	1.03	1.01	19.68
La	13.00	80.00	37.71	37.00	31.00	42.00	11.24	0.92	1.91	32.88
Nb	6.00	35.00	14.49	14.00	11.00	15.00	5.13	1.67	3.87	166.20
Ni	18.00	82.00	34.67	33.00	28.00	40.00	9.93	1.16	2.95	31.40
Pb	8.00	708.00	63.67	31.00	20.00	69.00	85.00	3.89	22.56	328.73
Ti	0.45	1.18	0.73	0.71	0.61	0.83	0.15	0.47	0.20	26.77
V	54.00	239.00	107.36	102.00	87.00	123.00	30.89	1.28	2.46	25.77
Y	0.00	55.00	24.99	26.00	19.00	30.00	8.15	0.35	1.08	21.46
Zn	38.00	871.00	166.73	127.00	93.00	189.00	130.53	2.68	8.70	102.01

same independent physical processes, acting at different spatial scales u . The $N(N+1)/2$ simple and cross variograms of the N variables are modelled by a linear combination of S standardized variograms to unit sill $g^u(\mathbf{h})$. Using the matrix notation, LMC can be written as:

$$\Gamma(\mathbf{h}) = \sum_{u=1}^{N_s} \mathbf{B}^u g^u(\mathbf{h}) \tag{3}$$

Table 2 First five eigenvectors from the PCA of the standardized elements data

	PC1	PC2	PC3	PC4	PC5
Al	0.331	-0.014	0.129	0.160	0.296
As	-0.044	0.508	0.189	-0.431	0.063
Ba	0.125	0.369	-0.148	0.591	-0.240
Co	0.351	-0.025	-0.179	-0.079	0.063
Cr	0.299	0.052	-0.334	-0.314	-0.073
Fe	0.353	0.038	-0.148	0.102	0.196
La	0.253	0.083	0.541	0.046	-0.328
Nb	0.259	0.123	0.514	-0.171	-0.245
Ni	0.281	0.015	-0.362	-0.374	-0.384
Pb	-0.094	0.518	0.004	-0.196	0.522
Ti	0.326	0.087	0.064	0.256	0.332
V	0.359	-0.021	-0.091	-0.042	0.064
Y	0.285	-0.197	0.132	-0.016	0.107
Zn	-0.024	0.514	-0.208	0.217	-0.302
E.V.	7.219	2.477	1.228	0.958	0.559
%V.	51.60	17.70	8.80	6.80	4.00
Cumulative %	51.60	69.30	78.10	84.90	88.90

Eigenvalues (E.V.) and percentage of explained variance (%V.) are also reported

where $\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]$ is a symmetric matrix of order $N \times N$, whose diagonal and non-diagonal elements represent simple and cross variograms for lag \mathbf{h} ; $\mathbf{B}^u = [b_{ij}^u]$ is called coregionalization matrix, and it is a symmetrical semi-definite matrix of order $N \times N$ with real elements b_{ij}^u at a specific spatial scale u . The model is authorized if the functions $g^u(\mathbf{h})$ are authorized variogram models. In LMC, the spatial behaviour of the variable is assumed to result from superimposition of different independent processes working on different spatial scales. These processes may affect the behaviour of experimental semivariograms, which can be modelled by a set of functions $g^u(\mathbf{h})$. Fitting of LMC is performed by weighed least-squares approximation under the constraint of positive semi-definiteness of the \mathbf{B}^u , using an algorithm developed by Lajaunie and Béhaxëtéguy (1989). The best model was chosen, as suggested by Goulard and Voltz (1992), by comparing the goodness of fit for several combinations of functions of $g^u(\mathbf{h})$ with different ranges in terms of the weighted sum of squares.

Variogram modelling is sensitive to marked departures from normality because a few exceptionally large values may contribute to a high number of large squared differences (Webster and Oliver 2007). To produce the maps of PTEs, we used multi-Gaussian cokriging (Goovaerts 1997; Verly 1983; Wackernagel 2003). In the multi-Gaussian approach, regardless of the shape of the sample histogram, the data are transformed into a Gaussian-shaped variable (Y) with zero mean and unit variance by Gaussian anamorphosis using an expansion into Hermite polynomials $H_i(Y)$ (Chilès and Delfiner 1999). The transformed data are estimated at all unsampled locations using ordinary cokriging. Finally, the estimates are transformed back to the raw data through the mathematical model calculated in Gaussian anamorphosis.

Regionalized principal component analysis consists in decomposing each coregionalization matrix \mathbf{B}^u into eigenvalues and eigenvectors matrices (Wackernagel 2003):

$$\mathbf{B}^u = \mathbf{Q}^u \mathbf{\Lambda}^u \mathbf{Q}^{uT} = \left(\mathbf{Q}^u \sqrt{\mathbf{\Lambda}^u} \right) \left(\mathbf{Q}^u \sqrt{\mathbf{\Lambda}^u} \right)^T = \mathbf{A}^u \mathbf{A}^{uT} \quad (4)$$

where \mathbf{Q}^u is the matrix of eigenvectors and $\mathbf{\Lambda}^u$ is the diagonal matrix of eigenvalues for each spatial scale u ; $\mathbf{A}^u = \mathbf{Q}^u \sqrt{\mathbf{\Lambda}^u}$ is the matrix of order $N \times N$ of the transformation coefficients a_{ip}^u . The transformation coefficients a_{ip}^u in the matrix \mathbf{A}^u correspond to the covariances between the original variables $Z_i(\mathbf{x})$ and the regionalized factors $Y_p^u(\mathbf{x})$. The behaviour and relationships among variables at different spatial scales can be illustrated by mapping the regionalized factors $Y_p^u(\mathbf{x})$ estimated by cokriging (Wackernagel 2003).

All statistical and geostatistical analyses were performed by using the software package ISATIS[®], release 2011.

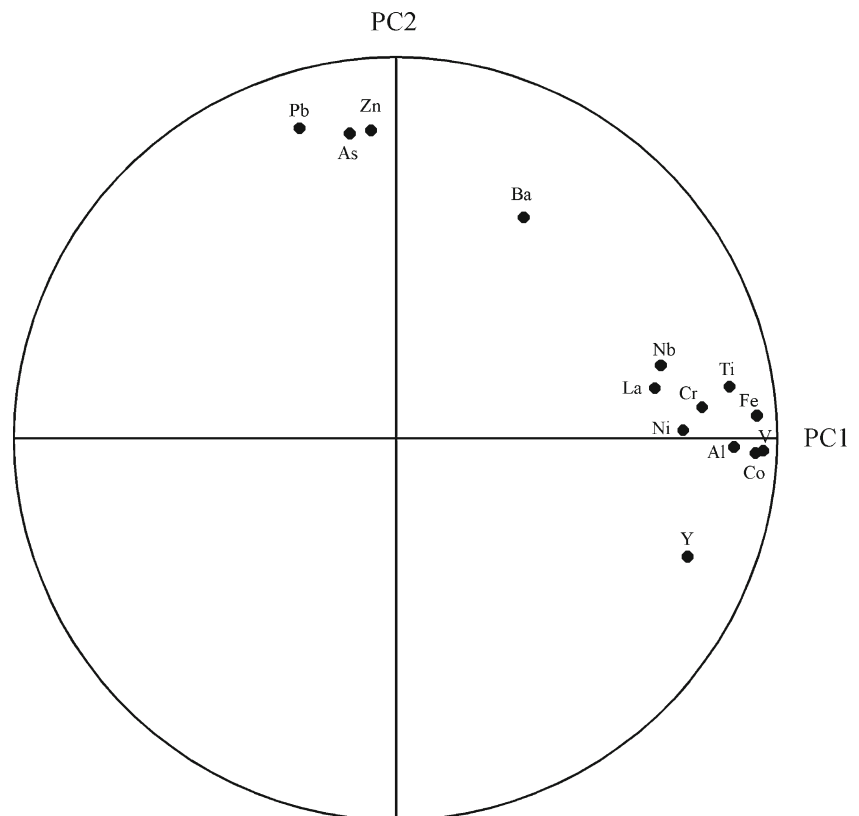
3 Results and discussion

The descriptive statistics for the 14 PTE data are presented in Table 1. Comparing values for mean, median and skewness, it can be seen that the distributions of variables vary from normal (e.g. Al and Fe) to very positively skewed (e.g. Zn, Pb and Ba; see Table 1). The distributions of PTEs departed

significantly from the Gaussian distribution as also shown by the χ^2 test for normality: only for Al and Fe the hypothesis that the data were normally distributed was accepted at the 0.10 and 0.05 probability levels because the experimental χ^2 values (see Table 1) were less than the theoretical χ^2 values of $\chi_{0.10}^2 = 21.06$ and $\chi_{0.05}^2 = 23.69$, respectively. For subsequent analyses, the PTE data were transformed to normality and standardized to zero mean and unit variance using a Gaussian anamorphosis by an expansion of Hermite polynomials restricted to the first 30 terms (Wackernagel 2003).

A summary of the results of PCA is presented in Table 2, where the elements of the PCs (loadings) for each variable in relation to the first five principal components are given as well as the eigenvalues of the PCs, and the percentages of explained variance for each of the components are also reported. The first two components account for almost 70 % of the total variance, and the third for an additional 9 %. The first four components are sufficient to explain the information contained in the correlation matrix because the fifth component has a variance percentage lower than what would be explained by each individual variable ($1/14 \approx 0.07$). The contribution of Al, Co, Cr, Fe, La, Nb, Ni, Ti, V and Y to PC1 ranges between 0.253 and 0.359 (see Table 2), which means that no variate makes a predominant contribution to that component. As, Pb and Zn make a similar contribution to PC2 with values ranging between 0.508 and 0.518 (see Table 2).

Fig. 4 Correlation circle of PTEs



The first two associate components were converted to the correlation between the original variables and the principal components by Eq. (1), then they were plotted in the unit circle in the plane of the first two principal components (Fig. 4). The circle of correlations showed the proximity of the PTEs inside the unit circle, and it was useful to evaluate their associations. The inspection of the plane of the first two principal components (see

Fig. 2) revealed two groups of PTEs. A group showed a positive correlation with the second principal component and included Pb, Zn and As, while a second group of nine PTEs (Al, Co, Cr, Fe, La, Nb, Ni, Ti and V) showed a positive correlation with the first component. Ba and Y were isolated, and they were discarded. Then, the two groups of PTEs were analysed separately by factorial kriging analysis.

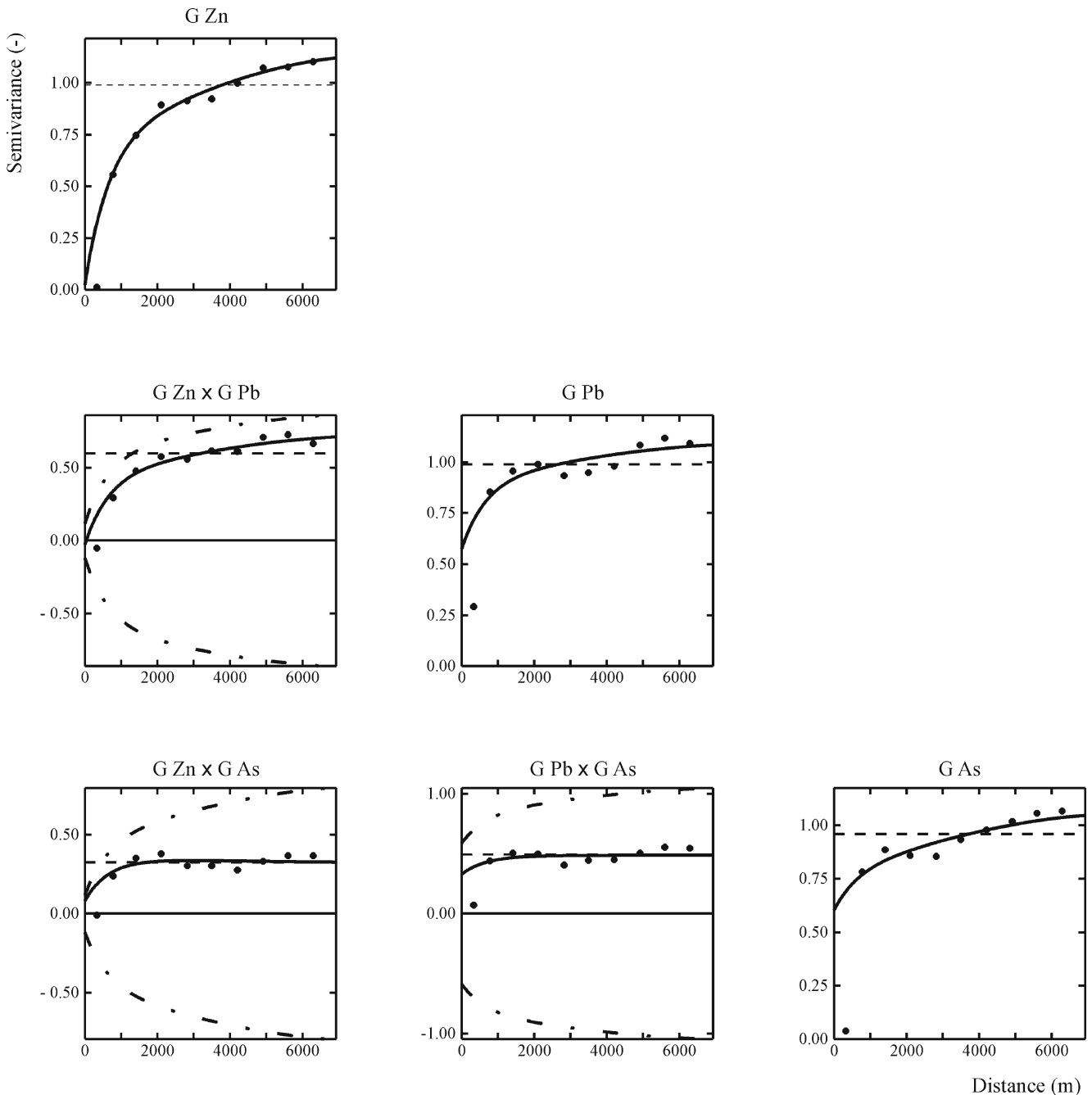


Fig. 5 LMC for the first group of PTEs (As, Pb and Zn). The experimental values are the *plotted points*, and the *solid lines* are of the model of coregionalization. The *dash-dotted lines* are the hull of perfect correlation, and the *dashed lines* are the experimental variances

3.1 Group I: As, Pb and Zn

The variographic analysis was carried out for the three elements (As, Pb and Zn) of the first group, and no anisotropy was evident in the maps of the two-dimensional variograms (not shown) to a maximum lag distance of 6,000 m. A nested isotropic LMC (Fig. 5) was fitted to the experimental variograms which included three basic structures: a nugget effect, an exponential model (Webster and Oliver 2007) with a practical range of 2,000 m and a spherical model (Webster and Oliver 2007) with a range of 8,000 m.

The appropriateness of the LMC used compared to alternative models was evaluated by cross-validation. The mean error and the variance of standardized errors for the selected model were close to 0 and 1: the mean error varied between 0.01 and 0.02, while the variance of standardized errors varied between 0.98 and 1.06.

The Gaussian values of the three PTEs belonging to ‘group I’ were interpolated by using multi-Gaussian cokriging at the nodes of a 25×25-m grid. Then, the estimates were back transformed to the row data through the mathematical model calculated in the Gaussian anamorphosis. Figure 6 shows only the maps of lead and zinc concentrations.

Table 3 reports the decomposition in the regionalized factors of variance–covariance matrices of LMC at each spatial scale. The loadings [coefficients of transformation of Eq. (2)] for each variable in relation to the regionalized factors are given as well as the eigenvalues and the percentage of explained variance for each of the regionalized factors. The sum of the eigenvalues at each spatial scale gives an estimate of the variance at that scale. The nugget was about 37 % of total variance (3.27), while the contribution of the shorter range component (2,000 m) of variation to the total variance was 37 %, and the contribution of the longer range component (8,000 m) was 26 %. Moreover,

Table 3 Decomposition of the coregionalization matrix in the regionalized factors (group I)

	G As	G Pb	G Zn	E.V.	%V.
Nugget effect					
Factor 1	0.7242	0.6881	0.0452	0.9201	76.64
Factor 2	0.6525	−0.7050	0.2779	0.2805	23.36
Exponential model (practical range=2,000 m)					
Factor 1	0.3162	0.5298	0.7870	1.1320	92.7
Factor 2	0.9053	−0.4164	−0.0834	0.0891	7.3
Spherical model (range=8,000 m)					
Factor 1	−0.0716	0.5500	0.8321	0.5863	68.89
Factor 2	0.9952	0.0951	0.0228	0.2647	31.11

The eigenvalues (E.V.) and the corresponding percentages of explained variance (%V.) for each spatial scale are also reported

the matrix of eigenvectors (regionalized factors) contains useful information only if some meaning can be attached to the principal components. The elements of an eigenvector represent the contribution of the original variate to that regionalized factor. An element of an eigenvector with a value near 1 means that the original variate makes a large contribution to that regionalized factor. Conversely, if an element of an eigenvector is near 0 the contribution to that regionalized factor is small. Therefore, by examining the eigenvectors it may be possible to give the regionalized factors a physical interpretation. However, as the regionalized factors are only mathematical constructs, with no direct physical meaning, interpretation is by no means assured. As a result, the regionalized factors producing eigenvalues greater than 1 were retained because when an eigenvalue is lower than 1, the associated factor has less explanatory value than any single PTE as its variance is inferior to the unit variance of each PTE. Only the first regionalized factor

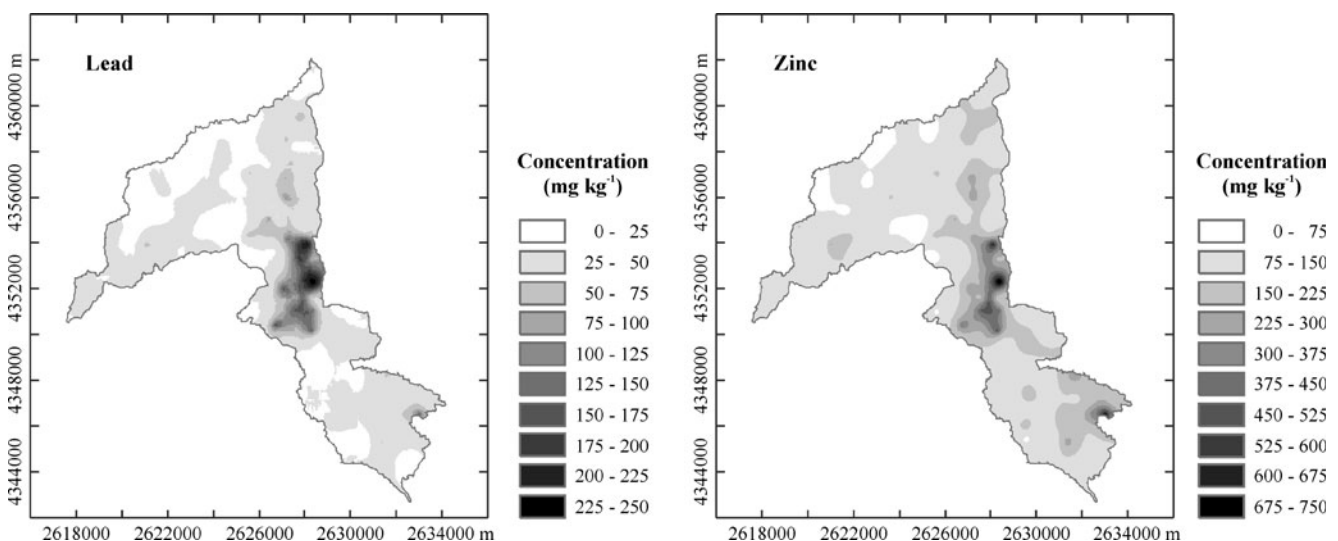


Fig. 6 Maps of lead and zinc concentrations

corresponding at short range (2,000 m) was considered for the final analysis, accounting for about 93 % of the variability at that spatial scale. The first regionalized factor at shorter range was interpolated by using ordinary cokriging and a 25×25-m grid (Fig. 7a). The first regionalized factor at short range, as a new variable, allows to aggregate and summarize the joint variability in the first group of PTEs (As, Pb and Zn) and to draw conclusions about the probable causes of soil pollution. The loading values for the first regionalized factor at short range (see Table 3) indicated the Zn as the variable most influencing the first regionalized factor at shorter range, followed by Pb and As. The contribution of the shorter range component (2,000 m) of variation (37 %) to the total variance was probably related to anthropogenic causes. The spatial distribution of the first regionalized factor at shorter range (see Fig. 7a) shows abrupt changes of values and a clear correspondence with the sources of pollution such as the main roads and urban areas (Fig. 8). Therefore, variation at shorter scale is related to the local traffic flow and high density housing. This result is confirmed by the maps of lead and zinc concentration (see Fig. 6) in which the higher values were located in an urbanized zone with a high density of road network (see Fig. 8). In this zone, the values of lead concentration were higher than those reported in literature ranging between 17 and 29 mg kg⁻¹ (Wedepohl 1995; Ure and Berrow 1982). The same considerations are applicable to As and Zn because their concentration values were higher than those reported in literature for the average upper continental crust values (Wedepohl 1995).

3.2 Group II: Al, Co, Cr, Fe, La, Nb, Ni, Ti and V

For the elements of the second group, the variographic analysis was carried out, and no anisotropy was evident in

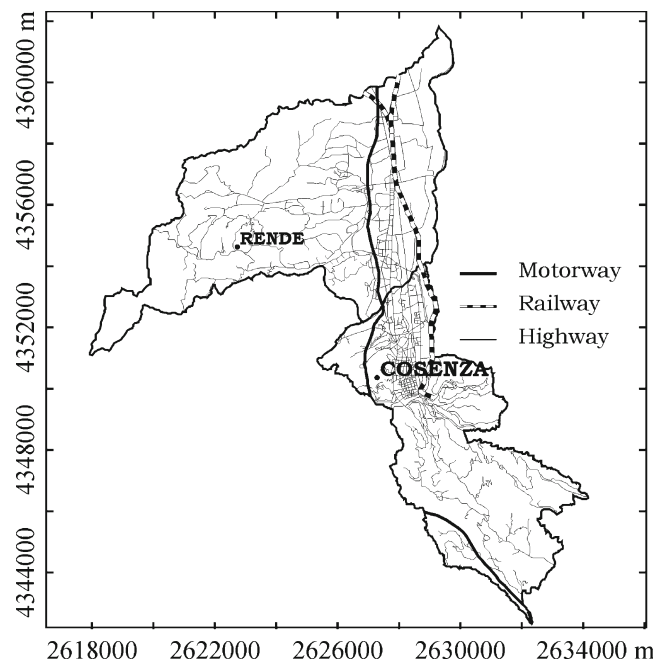


Fig. 8 Road network map of the study area

the maps of the two-dimensional variograms (not shown) to a maximum lag distance of 6,000 m. A nested isotropic LMC (not shown) was fitted to the experimental variograms which included three basic structures: a nugget effect, an exponential model (Webster and Oliver 2007) with a practical range of 2,000 m and a spherical model (Webster and Oliver 2007) with a range of 8,000 m.

The appropriateness of the LMC used compared to alternative models was evaluated by cross-validation. The mean error and the variance of standardized errors for the selected model were close to 0 and 1: the mean error varied between

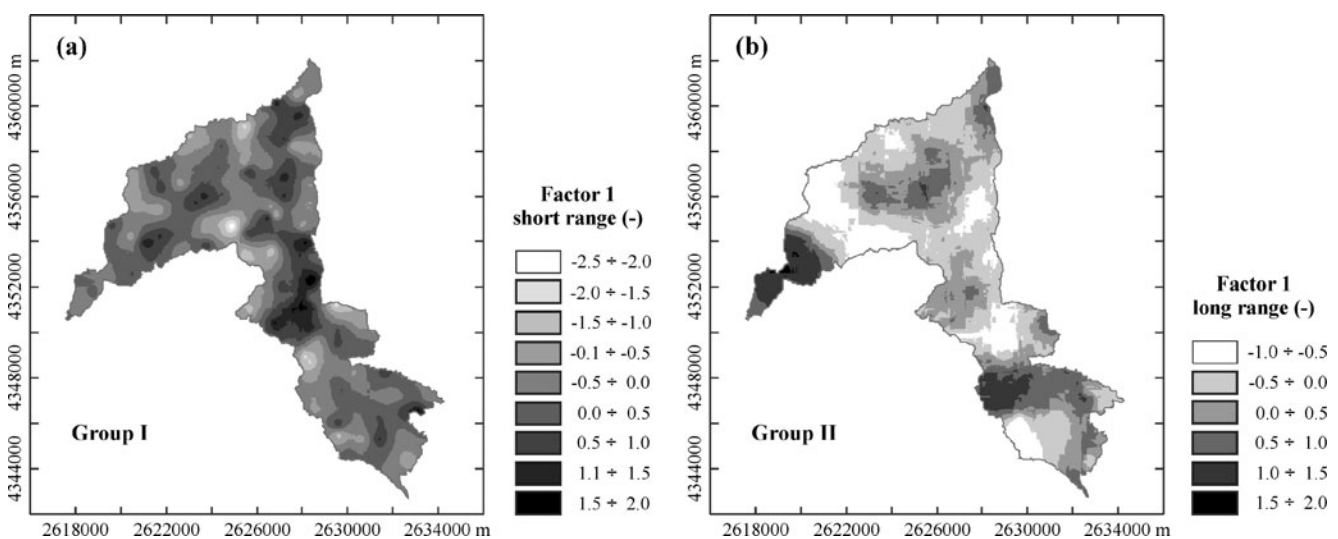


Fig. 7 Maps of the first regionalized factors at short range a and at long range b

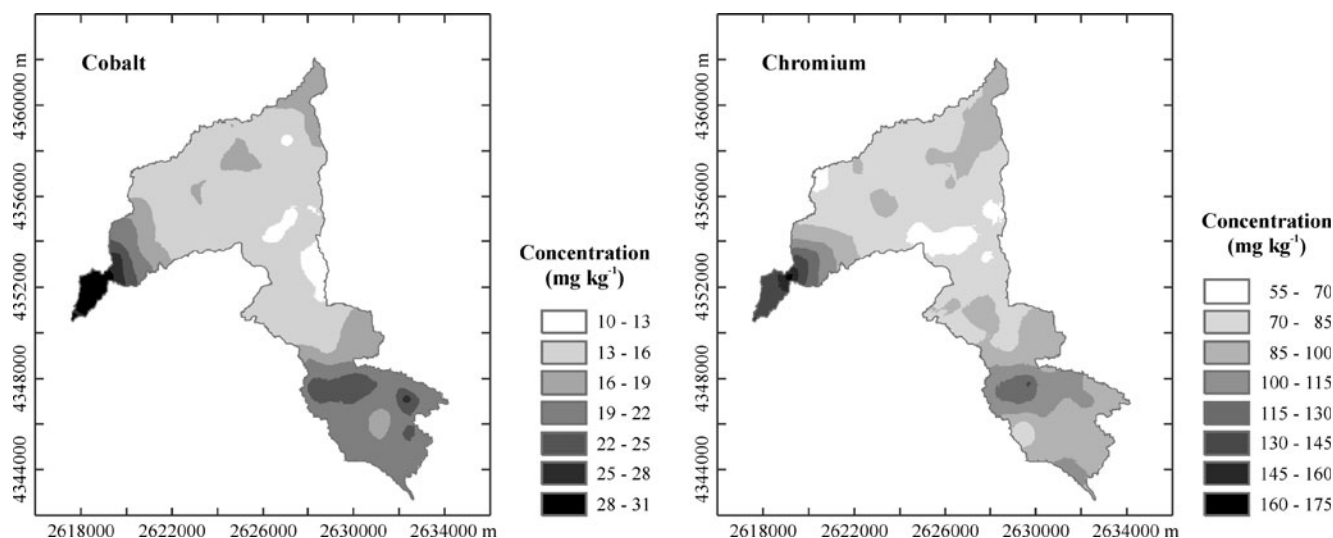


Fig. 9 Maps of cobalt and chromium concentrations

−0.01 and 0.01, while the variance of standardized errors varied between 0.96 and 1.10.

The Gaussian values of the PTEs belonging to ‘group II’ were interpolated by using multi-Gaussian cokriging at the nodes of a 25×25-m grid. Then, the estimates were back transformed to the row data through the mathematical model calculated in the Gaussian anamorphosis. Figure 9 shows only the maps of cobalt and chromium concentrations.

Table 4 reports the decomposition in the regionalized factors of variance–covariance matrices of LMC at each spatial scale. The loadings [coefficients of transformation of Eq. (2)] for each variable in relation to the regionalized factors are given as well as the eigenvalues and the percentage of explained variance for each factor. The nugget was roughly 53 % of total variance (8.86), while the contribution of the shorter range component (2,000 m) to the total variance was

Table 4 Decomposition of the coregionalization matrix in the regionalized factors (group II)

	G Al	G Co	G Cr	G Fe	G La	G Nb	G Ni	G Ti	G V	E. V.	%V.
Nugget effect											
Factor 1	0.285	0.331	0.307	0.311	0.393	0.408	0.309	0.270	0.359	3.510	74.26
Factor 2	0.040	−0.215	−0.454	−0.184	0.599	0.432	−0.382	0.043	−0.136	0.603	12.76
Factor 3	−0.379	−0.271	0.357	−0.304	0.493	−0.121	0.458	−0.304	−0.060	0.247	5.23
Factor 4	0.100	−0.043	0.096	0.021	0.391	−0.716	−0.294	0.448	0.163	0.142	3.00
Factor 5	−0.669	−0.182	0.335	0.257	−0.113	0.272	−0.293	0.408	−0.051	0.107	2.27
Factor 6	0.012	−0.668	−0.296	0.351	−0.085	−0.025	0.203	−0.049	0.539	0.055	1.15
Factor 7	0.329	−0.242	0.559	−0.198	−0.081	0.140	−0.506	−0.347	0.282	0.037	0.78
Factor 8	0.002	−0.063	−0.053	−0.736	−0.242	0.147	0.182	0.488	0.316	0.026	0.55
Exponential model (practical range=2,000 m)											
Factor 1	0.558	−0.145	−0.288	0.029	0.188	−0.164	−0.686	0.221	−0.053	0.344	67.65
Factor 2	−0.323	0.456	−0.454	−0.017	0.316	−0.440	−0.112	−0.420	−0.019	0.141	27.78
Factor 3	0.351	0.092	−0.434	0.204	−0.222	−0.290	0.580	0.285	−0.297	0.017	3.41
Factor 4	0.065	−0.371	−0.415	0.172	−0.183	0.050	0.109	−0.263	0.734	0.006	1.17
Spherical model (range=8,000 m)											
Factor 1	0.347	0.371	0.362	0.386	0.201	0.265	0.268	0.367	0.382	3.171	87.42
Factor 2	0.476	−0.160	−0.229	0.006	0.266	0.202	−0.635	0.351	−0.238	0.312	8.59
Factor 3	−0.462	−0.411	0.392	0.256	−0.146	−0.173	−0.177	0.562	−0.028	0.086	2.37
Factor 4	−0.344	0.131	0.043	−0.590	−0.075	0.628	−0.126	0.198	0.243	0.049	1.34
Factor 5	0.059	−0.157	0.515	0.154	−0.020	0.159	−0.486	−0.592	0.265	0.010	0.28

The eigenvalues (E.V.) and the corresponding percentages of explained variance (%V.) for each spatial scale are also reported

6 %, and the contribution of the longer range component (8,000 m) was 41 %. Only the regionalized factors producing eigenvalues greater than 1 were retained. The regionalized factor corresponding to the nugget effect was omitted because this is more affected by measurement error and variation at a scale smaller than 700 m (lag). Only the first regionalized factor corresponding at long range (8,000 m) was considered for the final analysis, accounting for about 87 % of the variability at that spatial scale (see Table 4). The first regionalized factor at longer range was interpolated by ordinary cokriging at the nodes of a 25 × 25-m grid (see Fig. 7b). The loading values for the first regionalized factor at long range (see Table 4) for the second group of PTEs indicated that Al, Co, Cr, Fe, Ti and V equally affected the first regionalized factor at long range with no element predominant. The same occurred for La, Nb and Ni. The contribution of the longer range component (8,000 m) of variation to the total variance was 41 % and only 6 % for the shorter range component (2,000 m) indicating that the variability of the second group of PTEs (Al, Co, Cr, Fe, La, Nb, Ni, Ti and V) was probably unrelated to anthropogenic causes, but rather to the predominant rock-forming elements constituting the soil parental materials. The map of the first regionalized factor at longer range (see Fig. 7b) shows a smooth change of values. In addition, high values were located in peri-urban areas (see Fig. 7b). These results were confirmed by the maps of cobalt and chromium (see Fig. 9) where, similarly to all elements of ‘group II’ (maps not shown), the higher values of elements’ concentration were located in peri-urban areas with a low road network density. The values of concentration of elements belonging to ‘group II’ were higher than those reported in literature (Wedepohl 1995), but similar to the ones in parent rocks.

4 Conclusions

This study allowed to analyse and quantify the relationships between potentially toxic trace elements and environmental and anthropogenic causes of soil pollution by using principal component analysis and factorial kriging analysis. The advantage of this approach is that it is flexible and can be used to make direct comparisons of soil pollution in different areas. The regionalized factor at different scale of variability allowed to aggregate and summarize the joint variability in the PTEs and to draw conclusions regarding the probable causes of soil pollution.

The study allowed to identify two groups of PTEs: the first group including As, Pb and Zn; and the second one Al, Co, Cr, Fe, La, Nb, Ni, Ti and V. The first group was related to anthropogenic causes, while the second was more related to rocks composition.

In the first group of PTEs, the results indicated Zn as the variable most influencing the first regionalized factor at

shorter range, followed by Pb and As. Moreover, the contribution of the shorter range component of variation to the total variance was probably related to anthropogenic causes because the spatial distribution of the first regionalized factor at shorter range showed a clear correspondence with the sources of pollution such as the main roads and urban areas.

In the second group of PTEs, the results indicated that Al, Co, Cr, Fe, Ti and V equally affected the first regionalized factor at long range with no element predominating. The contribution of the longer range component of variation to the total variance was predominant indicating that the variability of the second group of PTEs (Al, Co, Cr, Fe, La, Nb, Ni, Ti and V) was probably unrelated to anthropogenic causes, but rather to the predominant rock-forming elements that constituted the parental materials of the soil. Finally, the results of this study can be considered a useful contribution to identifying polluted areas and proposing remedial action aimed at reducing health risk above all to people.

Acknowledgments The authors thank the reviewers of this paper for providing constructive comments which have contributed to the improvement of the published version. The authors thank Emilio Catalano for his valuable technical support.

References

- Alary C, Demeougeot-Renard H (2010) Factorial kriging analysis as a tool for explaining the complex spatial distribution of metals in sediments. *Environ Sci Technol* 44:593–599
- ARSSA (2003) Carta dei suoli della Calabria. Rubettino Industrie Grafiche e Editoriali, Soveria Mannelli
- Atteia O, Dubois JP, Webster R (1994) Geostatistical analysis of soil contamination in the Swiss Jura. *Environ Pollut* 86:315–327
- Borůvka L, Vacek O, Jehlička I (2005) Principal component analysis as a tool to indicate the origin of potentially toxic elements in soils. *Geoderma* 128:289–300
- Brus DJ, de Gruijter JJ, Walvoort DJJ, de Vries F, Bronswijk JJB, Römkens PFAM, de Vries W (2002) Mapping the probability of exceeding critical thresholds for cadmium concentrations in soils in the Netherlands. *J Environ Qual* 31:1875–1884
- Carroll ZL, Oliver MA (2005) Exploring the spatial relations between soil physical properties and apparent electrical conductivity. *Geoderma* 128:354–374
- Cattle JA, McBratney AB, Minasny B (2002) Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. *J Environ Qual* 31:1576–1588
- Chilès JP, Delfiner P (1999) *Geostatistics: modeling spatial uncertainty*. Wiley, New York
- Cicchella D, De Vivo B, Lima A, Albanese S, Fedele L (2008) Urban geochemical mapping in the Campania region (Italy). *Geochem Explor Environ Anal* 8:19–29
- Coşkun M, Steinnes E, Frontasyeva MV, Sjobakk TE, Demkina S (2006) Heavy metal pollution of surface soil in the Thrace region, Turkey. *Environ Monit Assess* 119:545–556
- Davis JC (2002) *Statistics and data analysis in geology* (3rd edn). Wiley, New York
- FAO (1998) *World reference base for soil resources*. World Soil Resources Rep. 84. FAO, Rome

- Goovaerts P (1992) Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. *J Soil Sci* 43:597–619
- Goovaerts P (1997) *Geostatistics for natural resources evaluation*. Oxford University Press, New York
- Goovaerts P, Webster R (1994) Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. *Eur J Soil Sci* 45:79–95
- Goovaerts P, Webster R, Dubois J-P (1997) Assessing the risk of soil contamination in the Swiss Jura using indicator geostatistics. *Environ Ecol Stat* 4:31–48
- Goulard M, Voltz M (1992) Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Math Geol* 24:269–286
- Hooda PS (2010) Introduction. In: Hooda PS (ed) *Trace elements in soils*. Wiley, Chichester, pp 3–8
- Jimenez-Espinosa R, Sousa AJ, Chica-Olmo M (1993) Identification of geochemical anomalies using principal component analysis and factorial kriging analysis. *J Geochem Explor* 46:245–256
- Journel AG, Huijbregts CJ (1978) *Mining geostatistics*. Academic, San Diego
- Juang KW, Lee DY, Ellsworth TR (2001) Using rank-order geostatistics for spatial interpolation of highly skewed data in a heavy-metal contaminated site. *J Environ Qual* 30:894–903
- Lado LR, Hengl T, Reuter HI (2008) Heavy metals in European soils: a geostatistical analysis of the FOREGS Geochemical database. *Geoderma* 148:189–199
- Lajaunie C, Béhaxétéguy JP (1989) Elaboration d'un programme d'ajustement semi-automatique d'un modèle de corégionalisation—Théorie. Technical Report N21/89/G. ENSMP, Paris, p 6
- Le Pera E, Critelli S, Sorriso-Valvo M (2001) Weathering of gneiss in Calabria, southern Italy. *Catena* 42:1–15
- 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
- Liu H, Chen LP, Ai YW, Yang X, Yu YH, Zuo YB, Fu GY (2009) Heavy metal contamination in soil alongside mountain railway in Sichuan, China. *Environ Monit Assess* 152:25–33
- Maas S, Scheifler R, Benslama M, Crini N, Lucot E, Brahmia Z, Benyacoub S, Giraudoux P (2010) Spatial distribution of heavy metal concentrations in urban, suburban and agricultural soils in a Mediterranean city of Algeria. *Environ Pollut* 158:2294–2301
- Mahanta MJ, Bhattacharyya KG (2011) Total concentrations, fractionation and mobility of heavy metals in soils of urban area of Guwahati, India. *Environ Monit Assess* 173:221–240
- Markus JA, Mcbratney AB (1996) An urban soil study: heavy metals in Glebe, Australia. *Aust J Soil Res* 34:453–465
- Matheron G (1973) The intrinsic random functions and their applications. *Adv Appl Probab* 5:239–465
- Matheron G (1982) Pour une analyse krigeante des données régionalisées. Rapport N-732. Centre de Géostatistiques, École des Mines de Paris, Fontainebleau
- McGrath D, Zhang C, Carton OT (2004) Geostatistical analyses and hazard assessment on soil lead in Silvermines area, Ireland. *Environ Pollut* 127:239–248
- Mirsal IA (2008) *Soil pollution. Origin, monitoring & remediation*, 2nd edn. Springer, Berlin
- Norra S, Weber A, Kramar U, Stüben D (2001) Mapping of trace metals in urban soils. *J Soils Sediments* 1:77–97
- Odewande AA, Abimbola AF (2008) Contamination indices and heavy metal concentrations in urban soil of Ibadan metropolis, southwestern Nigeria. *Environ Geochem Health* 30:243–254
- Pearson K (1901) On lines and planes of closest fit to systems of points in space (PDF). *Philos Mag* 2:559–572
- Queiroz JCB, Sturaro JR, Saraiva ACF, Barbosa Landim PM (2008) Geochemical characterization of heavy metal contaminated area using multivariate factorial kriging. *Environ Geol* 55:95–105
- Reis AP, Menezes de Almeida L, Ferreira da Silva E, Sousa AJ, Patinha C, Fonseca EC (2007) Assessing the geochemical inherent quality of natural soils in the Douro river basin for grapevine cultivation using data analysis and geostatistics. *Geoderma* 141:370–383
- Saby N, Arrouays D, Bouillon L, Jolivet C, Pochot A (2006) Geostatistical assessment of Pb in soil around Paris, France. *Sci Total Environ* 367:212–221
- Sánchez-Marañón M, García PA, Huertas R, Hernández-Andrés J, Melgosa M (2011) Influence of natural daylight on soil color description: assessment using a color-appearance model. *Soil Sci Soc Am J* 75:984–993
- Schröder W, Pesch R, Schmidt G (2004) Soil monitoring in Germany. *J Soils Sediments* 4:49–58
- Shi G, Chen Z, Xu S, Zhang J, Wang L, Bi C, Teng J (2008) Potentially toxic metal contamination of urban soils and roadside dust in Shanghai, China. *Environ Pollut* 156:251–260
- Sollitto D, Romic M, Castrignanò A, Romic D, Bakic H (2010) Assessing heavy metal contamination in soils of the Zagreb region (Northwest Croatia) using multivariate geostatistics. *Catena* 80:182–194
- Ure AM, Berrow ML (1982) The elemental constituents of soils. In: Bowen HJM (ed) *Environmental chemistry*. Royal Society of Chemistry, London, pp 94–202
- Verly G (1983) The Multigaussian approach and its application to the estimation of local reserves. *Math Geol* 15:259–286
- Visconti F, de Paz JM, Rubio JL (2009) Principal component analysis of chemical properties of soil saturation extracts from an irrigated Mediterranean area: implications for calcite equilibrium in soil solutions. *Geoderma* 151:407–416
- Wackernagel H (2003) *Multivariate geostatistics: an introduction with applications*. Springer, Berlin
- Wang Z, Darilek JL, Zhao Y, Huang B, Sun W (2011) Defining soil geochemical baseline at small scales using geochemical common factors and soil organic matter as normalizers. *J Soils Sediments* 11:3–14
- Webster R (2001) Statistics to support soil research and their presentation. *Eur J Soil Sci* 52:331–340
- Webster R, Oliver MA (2007) *Geostatistics for environmental scientists*, 2nd edn. Wiley, Chichester
- Webster R, Atteia O, Dubois JP (1994) Coregionalization of trace metals in the soil in the Swiss Jura. *Eur J Soil Sci* 45:205–218
- Wedepohl KH (1995) The composition of the continental crust. *Geochim Cosmochim Acta* 59:1217–1232
- Zawadzki J, Fabijańczyk P (2008) The geostatistical reassessment of soil contamination with lead in metropolitan Warsaw and its vicinity. *Int J Environ Pollut* 35:1–12
- Zheng YM, Chen TB, He JZ (2008) Multivariate geostatistical analysis of heavy metals in topsoils from Beijing, China. *J Soils Sediments* 8:51–58