Soil Contamination Interpretation by the Use of Monitoring Data Analysis

Aleksander Maria Astel · Lyubka Chepanova ·
Vasil Simeonov

Received: 23 February 2010 / Accepted: 29 June 2010 / Published online: 13 July 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract The presented study deals with the interpretation of soil quality monitoring data using hierarchical cluster analysis (HCA) and principal components analysis (PCA). Both statistical methods contributed to the correct data classification and projection of the surface (0–20 cm) and subsurface (20–40 cm) soil layers of 36 sampling sites in the region of Burgas, Bulgaria. Clustering of the variables led to formation of four significant clusters corresponding to possible sources defining the soil quality like agricultural activity, industrial impact, fertilizing, etc. Two major clusters were found to explain the sampling site locations according to soil composition—one cluster for coastal and mountain

sites and another—for typical rural and industrial sites. Analogous results were obtained by the use of PCA. The advantage of the latter was the opportunity to offer more quantitative interpretation of the role of identified soil quality sources by the level of explained total variance. The score plots and the dendrogram of the sampling sites indicated a relative spatial homogeneity according to geographical location and soil layer depth. The high-risk areas and pollution profiles were detected and visualized using surface maps based on Kriging algorithm.

Keywords Soil contamination · Cluster analysis · Principal components analysis · Environmetrics · Spatial variation

A. M. Astel (✉)
Biology and Environmental Protection Institute,
Environmental Chemistry Research Unit Pomeranian
Academy,
Arciszewskiego Str, 22a,
76200 Slupsk, Poland
e-mail: astel@apsl.edu.pl
e-mail: AliAst@poczta.fm

L. Chepanova
Regional Laboratory Burgas, Environmental Executive
Agency, Ministry of Environment and Waters,
Burgas, Bulgaria

V. Simeonov
Analytical Chemistry, Faculty of Chemistry,
University of Sofia “St. Kl. Okhridski”,
J. Bourchier Blvd. 1,
1164 Sofia, Bulgaria
e-mail: vsimeonov@chem.uni-sofia.bg

1 Introduction

The traditional assessment of soil contamination is based on the regular routine of comparison of allowable threshold values with the results of monitoring. This approach is even a required action in environmental agencies, agricultural administration, and managing organization. Very often, solving a particular problem concerning the soil contamination or respective decision making is based solely on single results and not on a more generalized model about the state of the soil contamination in a certain region. The application of multivariate statistical approaches to the problem allows a better classification,

modeling, and interpretation of the soil monitoring data. This environmetric strategy makes it possible to detect relationships between the chemical pollutants and specific soil parameters, between sampling sites and, therefore, to achieve a stratification of the pollution. Further, it becomes possible to identify possible pollution sources and to construct apportioning models allowing the determination of the contribution of each identified source to the formation of the total pollutant mass (Stanimirova et al. 2006, 2009; Einax and Soldt 1995; Singh et al. 2008; Andrade et al. 2007; Buszewski and Kowalkowski 2006; Kemper and Sommer 2002; Terrado et al. 2007; Perez Pavon et al. 2008).

The aim of the present study is to assess the soil quality in the region of Burgas, Bulgaria by the application of two already classical multivariate statistical methods (cluster analysis and principal components analysis) in order to get information about some spatial distribution of the soil pollutants in the region (by comparing the linkage between the different sampling sites) and to identify possible pollution sources (by determining an appropriate number of latent factors undergoing logic interpretation). The region of Burgas is located close to the Bulgarian Black Sea coastal line and is characterized by high industrial and agricultural activity.

2 Experimental

2.1 Sampling and Chemical Analysis

The sampling campaign comprised the period 2004–2006 with annual time interval and various repeatability (varying between 1 and 3) in particular locations. Altogether 36 locations were chosen (Jasna Poljana (JP), Drachevo (D), Svoboda (S), Sarnevo (SA), Sozopol (SO), Rudina (R), Rechica (RE), Bjala (B), Marinka (M), Muglen (MA), Karageorgievo (KA), Zvezdec (Z), Vizica (V), Kosti (KO), Malko Tynovo (MT), Vratice (VR), Krushevec (KR), Medovo (ME), Kozichino (KZ), Polski Izvor (PI), Smolnik (SM), Karnobat (KB), Podvis (P), Terzijsko (T), Luljakovo (L), Topuzevo (TO), Sedlarevo (SE), Kotel (KT), Sadovo (SD), Slivovo (SL), Zornica (ZR), Prohod (PH), Raklica (RK), Kipilovo (KI), Trakijci (TK), and Samotinovo (ST)) and the adequate sampling map is shown in Fig. 1. All of them belong to the Bulgarian monitoring network for soil quality

assessment. To improve manuscript clarity, all samples descriptions were abbreviated as follows: i.e., Luljakovo, sample 1, surface layer: L-1-S; Terzijsko, sample 2, subsurface layer: T-2-SS.

Soil sampling was carried out by an accredited laboratory in accordance with ISO 17025 (ISO 2005) standard and recommended procedures for sampling, described in ISO (ISO 2009). In every location, there were several independent soil samples taken from two different depths (the surface layer (0–20 cm) and subsurface layer (20–40 cm)), according to the sampled area as required by the ISO 10381 standard so there were enough data for statistical assessment of the analytical procedures applied.

Machining of the samples, drying, grinding and sieving was carried out according to ISO 11464 (ISO 1994). To determine the total content of metals (arsenic (As), cadmium (Cd), lead (Pb), nickel (Ni), chromium (Cr), copper (Cu), zinc (Zn)) and phosphorus a fraction with grain size less than 65 μm was used, since for pH, total nitrogen (N_{tot}), and total organic carbon (TOC) a fraction of less than 2 mm was used.

Mass contents of As, Cd, Pb, Ni, Cr, Cu, Zn, and P in the soil samples were determined by validated non-standardized methods developed by the Regional Laboratory of the Ministry of the Environment and Waters (Chepanova et al. 2007, 2008). The methods were verified through participation in interlaboratory comparative laboratory tests reporting compatible results.

The soil samples were mineralized with aqua regia in a microwave oven at 180°C for 15 min (Chepanova et al. 2008). Metals were determined after appropriate dilution by Agilent 7500 ICP-MS (Agilent Technologies, USA) in the standard mode of measurement. Total phosphorus content was determined by spectrometric vanadate-molibdate reagent using Agilent XY Diode Array UV-VIS Spectrometer (Agilent Technologies, USA; Chepanova et al. 2007).

Quality of results was controlled by analysis of Certified Reference Materials NIST 2709 (San Joaquin soil) and CRM 142 (light sandy soil), having a similar matrix (baseline trace element concentration) as the studied soils. Following recovery was obtained for the elements: As (101–110%), Cd (97–100%), Pb (90–101%), Ni (93–98%), Cr (78%), Cu (94–99%), Zn (95–101%), and P (90–110%).

Other parameters like pH, total N, and TOC were determined according to the recommended standardized methods (ISO 2005a, b). pH was determined in

Fig. 1 Map of the sampling site location with major industrial pollutants (1 oil terminal, 2 reload sea port, 3 Kronospan wood factory, 4 industrial complex near Drachevo, 4 Promet industrial complex, 6 Lukoil oil refinery)

aqueous suspension (1:2.5) with a Microprocessor pH Meter pH 3000 (WTW Weinheim, Germany) since pH (H_2O) measurement is obligatory in the monitoring procedure. TC and the TOC were measured by an instrumental method with Shimadzu TOCN-4110 on-line total carbon/total nitrogen analyzer (Shimadzu Corp., Japan; EN 2001). The content of total nitrogen was determined by modified Kjeldahl method according to ISO 11261 (ISO 1995). Quality control was tested by analysis of CRM NCS DC 85104 and the recovery obtained was between 105–108% for N_{tot} and 90–98% for TOC, recalculated as an organic substance.

2.2 Data Analysis Methods

Hierarchical cluster analysis (HCA) and principal components analysis (PCA) were used for multivariate statistical modeling of the input data (Massart and Kaufman 1983; Vandeginste et al. 1997). The main

goal of the hierarchical agglomerative cluster analysis is to spontaneously classify the data into groups of similarity (clusters) searching objects in the n -dimensional space located in closest neighborhood and to separate a stable cluster from other clusters. Usually, the sampling sites are considered as objects for classification, each one determined by a set of variables (chemical concentrations). It is also possible to search for links between the variables turned to objects of classification. In order to achieve this series of procedures is necessary:

- 1 Normalization of the raw input data to dimensionless units in order to avoid the influence of the different range of chemical dimensions (concentration);
- 2 Determination of the distance between the objects of classification by application of some similarity measure, e.g., Euclidean distance or correlation coefficient;

- 3 Performing appropriate linkage between the objects by some of the cluster algorithms like single, average or centroid linkage;
- 4 Plotting the results as dendrogram;
- 5 Determination of the clustering pattern;
- 6 Interpretation of the clusters both for objects and variables.

Using cluster analysis, one could display the object similarity in a reliable way to make the initial interpretation of the dataset structure. But a more reliable display method proves to be PCA. It enables the reduction of the dimensionality of the space of the variables in the direction of the highest variance of the system, new variables being linear combinations of the previous variables, replacing the old coordinates of the factor space. The new coordinates are called latent factors or principal components. The interpretation of the new factors is the main goal of the chemists since they deliver useful information about latent relationships within the data set. The results are indicated by two sets—factor scores giving the new coordinates of the factor space with the location of the objects and factor loadings informing on the relationship between the variables. Usually, only statistically significant loadings (>0.70) are important for the modeling procedure. In the PCA procedure, the squared factor loading, commonly interpreted in the term Pearson's r , is the percent of variance in that indicator variable explained by the factor. In the presented case study, factor loadings higher than 0.5 were taken into consideration due to the size of data set ($n=170$, $R_{\text{crit}} (p=0.05) < 0.2$).

The new principal components (latent factors) explain a substantial part of the total variance of the system for an adequate statistical model. Usually, the first principal component explains the maximal part of the system variation and each additional PC has a respective contribution to the variance explanation but with less significance comparing to the previous one.

A reliable model in environmental studies requires usually such a number of PCs, so that over 70–75% of the total variation can be explained. In case of presented modeling, several rotation algorithms (raw and normalized Varimax, raw and normalized Biquartimax, etc.) have been tried in order to minimize the influence of the type of weighting scheme applied on factorial solution interpretation. Repeatable rotated factorial solutions were discovered and the difference between significant

factor loadings was not higher than 0.02 in every case. Based on it, negligible risk of making wrong decisions if they rely on the type of applied weighting schemes was proven. Finally, the normalized Varimax (as indubitably the most popular rotation method by far (Hervé 2010)) rotated PCA solution was interpreted that allows a better explanation of the system in consideration since it strengthens the role of the latent factors with higher impact on the variation explanation and diminishes the role of PCs with lower impact. Moreover, the sequence of rotated factors might not be longer arranged in an order of decreasing percentage of variance explained, although the total variance explained is equal before and after rotation. As established by Thurston (1974) and Cattell (1978), rotation strategy simplifies the structure of factors and therefore makes its interpretation easier and more reliable. All calculations were performed by the use of the software package STATISTICA 7.0 (Statsoft Inc., USA).

2.3 Visualization

All maps presented in this study were generated using Surfer 8.04 software (Golden Software, USA) dedicated for this purposes. Data were interpolated with Kriging method (Oliver and Webster 1991), which is one of several available Surfer's gridding algorithms. It interpolates data onto a regular grid (between the sampling points) and then shading algorithm creates a color or gray-scale map. Kriging method uses variogram to express the spatial variation and minimizes the error of predicted values which are estimated by spatial distribution of the predicted values. It implies relative stationarity of the system as is in case of soil. Comparing to other commonly used algorithm, as inverse distance, the accuracy of Kriging-based predictions are generally unaffected by the coefficient of variation (Yasrebi et al. 2009). As stated by Webster and Oliver (2006) for interpolation purposes, typical in soil and environmental surveys variograms computed on fewer than 50 data are of little information value and at least 100 data are needed. Nevertheless, in case of our research surface interpolation based on geographical location of 34 sample points, measured parameters and factor scores values was not dedicated for soil properties prediction purposes or logic-based modeling and was supported by further environmetric expertise. The total number of data was lower than optimal suggested by Webster and Oliver

(2006); however, we decided to present isoline plots as Kriging-based interpolations because of their extraordinary accuracy. Moreover, our choice was supported by several successful estimations of the spatial distribution of contaminants in various land areas studied by others (Cattle et al. 2002; Chang 2000; Largueche 2006; Lin et al. 2002; Liu et al. 2006).

3 Results and Discussion

The monitoring data set was separated into two major parts—data for the surface soil layer (surface sampling 0–20 cm) and for the subsurface soil layer (depth of 20–40 cm). Altogether, 340 samples from 36 different sampling sites were included in the data analysis. The chemical variables involved with abbreviations used in the statistical treatment were pH, N_{tot} , total phosphorus (P_{tot}), TOC, Cu, Zn, Pb, Cr, Ni, Cd, and As. Finally, the input matrix contains 340×11 elements. In Table 1, the basic statistics for the input data for both layers are given along with comparative values from several, although very rare, Bulgarian soil investigations found in the literature (Angelov 2008; Stefanov et al. 1995; Dinev et al. 2008; Malinova 2002; Schulin et al. 2007) while accurate shaded maps presenting spatial distribution of measured parameters maps according to surface and subsurface soil layers are presented on Figs. 2 and 3, respectively.

The comprehensive comparison between results presented in this study and in other studies is problematic since they are focused mainly on heavy metal contamination along a soil transect in the vicinity of the iron smelter (Schulin et al. 2007), gold-extracting factory (Dinev et al. 2008), metalliferous region (Stefanov et al. 1995) or on the reserve area (Angelov 2008). However, a few general remarks could be formulated. The concentrations of Ni in the studied samples were higher than in soil samples collected from Boatin Reserve or Etropoleška Stara Mountain while lower than in soil collected in the highly polluted area of Sofiiska Planina. Determined concentration of Cr was also higher than in case soil samples collected in Etropoleška Stara Mountain. The concentration of Zn in the studied samples was generally much higher than in soil from Yundola, Vitina and even from Kuklen and Dolni Voden which is metalliferous region. Based on limited comparison possibilities, it should be emphasized that the present study in the long term may be valuable as

a data source in database containing information about parameters of soil quality in Bulgaria. High content of several heavy metals in the studied samples indicates that possible correlation with an anthropogenic impact could not be neglected (Figs. 2 and 3).

The first step in the data analysis was performing of HCA. As already mentioned the normalized data (by the use of z-transformation) were subject to Ward's method of linkage after using the squared Euclidean distance as similarity measure. The upper and the lower soil layers were separately treated and both objects (sampling sites and periods) and variables (chemical components) were classified.

In Fig. 4, the hierarchical dendrograms for classification of the chemical variables for the surface and for the subsurface soil layer are presented.

The cluster formed for both cases are very similar. Four major clusters are identified as follows:

Surface layer: (Pb, Zn, Cu) (Cd, Ni, Cr, As)
(TOC, N_{tot}) (P_{tot} , pH)
Subsurface layer: (Pb, Zn, Cu, As) (Cd, Ni, Cr)
(TOC, N_{tot}) (P_{tot} , pH)

The only obvious difference is the linkage of arsenic to the group of Cd, Ni, and Cr in the upper layer and to the group of Pb, Zn, and Cu in the lower layer. Since both clusters are realized due to the anthropogenic sources in the region, it does not seem very surprising the relationship of As either to the one or to the other source. In principle, the hierarchical clustering reveals only one pattern of soil constitution (independent on the depth of sampling), e.g., the sources of the chemical constituents are originating dominantly from *anthropogenic* sources characteristic for the region of monitoring (transportation and burning activities with tracers lead, zinc, copper) or oil refinery activities (with tracers like cadmium, nickel, chromium) or correspond to *natural* sources related to the soil acidity and the content of soil nutrients. The link between trace metals and overall sources proposed was established based on independent soil studies published by others (Sanghi and Sasi 2001; Mandal and Sengupta 2006; Xiaoping 2007; Davies 1997; Simeonov et al. 2005; Škrbić 2004).

Although heavy metals (Pb, Zn, Cu, Cr, Ni, and other trace metals) occur naturally in soils, the link between trace metals and overall sources proposed could not be omitted due to acceptable agreement between chemical interpretation and geographical loca-

Table 1 Basic statistics for the surface soil layer data (all concentrations in mg kg⁻¹)

Variable	N	Mean	Median	Min	Max	SD	SF	Reference values from Bulgarian soil investigations						
								Boatin Reserve [24]	Kuklen and Dolni Voden [25]	Etropolekska Stara Mountain [26]	Yundola [27]	Vitina [27]	Staro Oriahovo [27]	Sofijska Planina [28]
Surface soil layer (0–20 cm)														
pH(H ₂ O)	170	6.9	6.8	5.0	8.9	0.8	−0.1	–	–	–	4.93	4.51	5.55	–
N _{tot}	170	1,611.0	1,496.0	224.0	3,917.0	654.9	0.9	–	–	–	–	–	–	–
P _{tot}	170	795.6	757.5	199.0	2,263.0	387.1	1.2	–	–	–	–	–	–	–
TOC	170	2.1	2.1	0.8	4.8	0.8	1.0	–	–	–	–	–	–	–
As	170	7.1	6.2	1.2	25.4	4.3	2.2	–	–	–	–	–	–	–
Cu	170	46.6	42.8	4.0	112.0	23.0	0.5	49.4	–	118.5	5.0	35.0	24.0	40.7
Cr	170	85.3	60.5	10.6	259.0	61.8	0.9	–	–	35.3	–	–	–	–
Ni	170	41.4	37.5	3.2	98.2	21.5	0.4	9.64	–	26.0	–	–	–	108
Cd	170	0.4	0.5	0.03	0.9	0.2	−0.2	0.85	280	0.6	–	–	–	0.5
Zn	170	63.0	62.7	21.3	178.0	23.1	1.5	55.5	30	98.1	46.0	18.0	62.0	–
Pb	170	21.5	20.8	9.5	57.6	7.6	1.8	41.4	3,500	39.5	22.0	44.0	19.0	–
Subsurface soil layer (20–40 cm)														
pH(H ₂ O)	170	6.9	7.0	5.3	8.8	0.8	−0.1	–	–	–	–	–	–	–
N _{tot}	170	1,546.4	1,518.0	451.0	3,780.0	602.2	0.8	–	–	–	–	–	–	–
P _{tot}	170	797.8	748.5	178.0	2,519.0	392.2	1.2	–	–	–	–	–	–	–
TOC	170	2.0	2.0	0.6	4.3	0.7	0.5	–	–	–	–	–	–	–
As	170	7.2	6.5	1.1	26.8	4.5	2.3	–	–	–	–	–	–	–
Cu	170	46.6	43.5	3.7	112.0	22.4	0.5	–	–	–	–	–	–	–
Cr	170	85.7	63.4	10.4	266.0	62.8	0.9	–	–	–	–	–	–	–
Ni	170	41.6	38.5	2.6	94.0	21.3	0.3	–	–	–	–	–	–	–
Cd	170	0.4	0.5	0.03	1.0	0.2	−0.0	–	–	–	–	–	–	–
Zn	170	63.6	63.5	21.5	184.0	22.0	1.2	–	–	–	–	–	–	–
Pb	170	21.4	21.0	9.5	51.1	7.2	1.4	–	–	–	–	–	–	–
Chronic <i>Alluvial</i> <i>Luvisola</i> <i>Fluvisola</i>														

SD Standard deviation, SF skewness factor

Fig. 2 Isolines plot for pH and all chemical variables (mg kg^{-1}) measured in surface soil layer after Kriging interpolation

tion of sampling stations. Other important information from the hierarchical clustering of the soil monitoring data is the spatial distribution and classification of the sampling sites of the region. In Fig. 5, the hierarchical dendrograms of the linkage between the sampling sites is shown for the surface and the subsurface soil layer, respectively.

As in the previous situation with the clustering of the chemical variables, two significant clusters are formed (both for the surface (S) and the subsurface (SS) soil layer). They correspond to the geographical location of the sampling sites. Thus, the cluster on the left side of the dendrograms includes dominantly *rural* and *industrial* sites located close to Burgas and

Fig. 3 Isoline plot for pH and all chemical variables (mg kg^{-1}) measured in subsurface soil layer after Kriging interpolation

in the vicinity of A4 highway (Burgas-Sofia). To the group of *rural/industrial* among other sites Karnobat, Smolnik, Vratitsa, Muglen, Karageorgievo, Medovo were included. The other cluster involves sites closer to the *coast* like Sozopol and Jasna Poljana as well as those located in the area of the Balkan *mountain range*: Kipilovo, Topuzevo, Sadovo, Sedlarovo, Podvis,

Terzijsko, Luljakovo, Rechnica, and Rudina. These two major patterns can be discriminated by the heavy metals as well as by the nutrients content. The comparison between averages for the different chemical constituents has indicated that the level of pollution for the coastal and mountain sites is significantly lower (lower median values both for nutrients and heavy metals,

Fig. 4 Dendrogram for hierarchical clustering of variables (surface and subsurface soil layers)

Mann-Whitney U test, $p < 0.001$) as compared with the averages for the typically *industrial* sites. Keeping in mind the presence of real polluting emitters in the region of monitoring (Fig. 1), it can be explained that the industrial sites are subject of serious anthropogenic impact due to the location of the Lukoil oil refinery, Kronospan wood factory and oil terminal near to the Gulf of Burgas as well as A4 highway (Burgas-Sofia) and the industrial complex Promet near to Drachevo sampling site. Parallel, the level of nutrients for the *rural* sites is higher than that of the coastal sites since agriculture activity in the valley located in the south from Balkan Mountain is very intensive. In Fig. 6, the comparison between chemical constitutes levels according to sampling location categorization (*rural/industrial* and *coast/mountains*) as well as soil layer (*surface/subsurface*) is presented.

As mentioned before, although investigated trace metals occur naturally in soils, higher concentration of Cu, Cr, Ni, Cd, Zn, and Pb in soil samples collected near to major rural or industrial centers proves that promoted indicators really signifies a real correlation to the anthropogenic impact.

Since principal components analysis gives the opportunity for specific interpretation of the data set structure the next step in the data analysis was performing PCA (normalized input data). Again, both soil layers (surface and subsurface) were considered. Based on Kaiser's criteria (Kaiser 1960) four latent factors, responsible for explaining over 70% of the total variance of the system for both types of samples, were chosen for interpretation. In Table 2, the factor

loadings values are given and the statistically significant ones (higher than 0.5) are marked. To enable comparison of factorial solution with and without weighting schemes applied (no Varimax rotation and normalized Varimax rotation, respectively), results of both modes are depicted in Table 2.

In agreement with expectations, in case of unrotated PCA solution, the majority of variables were correlated with the most informative first factor making its direct interpretation problematic. Varimax rotation significantly simplified the structure of factors and therefore made their interpretation much more reliable. In general, the results obtained confirm the conclusions from the hierarchical cluster analysis. The four latent factors (principal components) for the surface soil layer could be interpreted as follows:

- The first latent factor explaining nearly 21% of the total variance indicates strong correlation between Cu, Zn, and Pb and could be conditionally named “vehicle and industrial burning impact”; it reflects the influence of the road traffic and some industrial activities in the region of interest. The first factor's interpretation was supported by results published before, besides others, by Yun et al. (2000), Martínez et al. (2008), Terzano et al. (2007), and Simeonov et al. (2005).
- The second latent factor explains a lesser part of the total variance (about 16%) and shows the relationship between TOC and total amount of nitrogen; it could be accepted as a conditional “nitrogen nutrition” factor and informs on the agricultural impact on the upper soil level; the

Fig. 5 Dendrogram for hierarchical clustering of sites (surface and subsurface soil layers; because of limited resolution of the dendrogram only 50% of totally 170 samples were labeled)

negative correlation of the nutrients with arsenic confirms to a certain way the specific behavior of arsenic in the dendrograms presented and is, probably, a measure for the specific mobility of the pollutant in the surface soil layer; it seems that the organic compounds in the upper layer bind strongly arsenic;

- The third latent factor indicates the strong correlation between Cr, Ni, and Cd already shown by cluster analysis; this factor explains above 21% of the total variance and could be conditionally named “industrial anthropogenic” factor; the relationship found is in accordance with the presence of large industrial facilities in the region

Fig. 6 Visual assessment of medians differences between pH and other chemical constituents (N_{tot} , P_{tot} , TOC, As, Cu, Cr, Ni, Cd, Zn, Pb) levels according to sampling location categorization

(coast/mountains (C/M) and rural/industrial (R/I) as well as soil layer (surface (S) and subsurface (SS))

of interest. The third factor's interpretation was supported by results published before by others (Simeonov et al. 2005; Yun et al. 2000; Martínez et al. 2008; Terzano et al. 2008).

- The last latent factor from the series explains nearly 14% of the total variance and could be conditionally named “acidity” factor due to the strong correlation between soil acidity and total phosphorus, e.g., parameters bound also to nutrition processes, soil fertility, etc.

For the subsurface soil layer, the number of sufficient latent factors is also four and each one of them has almost the same contribution to the formation of the percentage of explained total variance. The results obtained resemble those from cluster analysis. However, the relative importance of the separate latent factors to the explanation of the total variance is different as compared to the sequence for the upper soil layer. The first latent factor (with 17% explanation of the total variance) is the conditional “soil nutrition and acidity” factor with strong correlation between total nitrogen, total phosphorus and pH. Next one (explained variance of 19%) is the “vehicle and industrial burning impact”, the third (with 23% explained variance) correspond entirely to

the conditional “industrial anthropogenic” factor. The last one (explained variance of 12%) indicates the negative correlation between the total organic carbon content and arsenic.

It seems that the structure of the monitored data for the soil quality in the region of interest shows slight difference for the upper and lower soil layer. The principal components analysis gives better information about sources determining the soil quality than cluster analysis. In Figs. 7 and 8, the most informative factor scores plots for both soil layers are shown indicating the location of the *rural/industrial* and *coast/mountains* sampling sites in the plane of the PCA-derived coordinates.

Two major groups of sites are visible and the careful decoding of the position of each sampling site indicates the separation according a geographical parameter—the one of the group is dominated by coastal and mountain sites and the other—by rural and industrial sites. The results of PCA calculations according to surface and subsurface layers representing the main impacts of pollution sources present in investigated area of Bulgaria were used to create shaded maps (Figs. 9 and 10). Variation of gray tones in grid maps represents changes in factor scores based on measured parameters values for soil in examined

Table 2 Factor loadings (surface and subsurface soil layers)

Variable	Surface soil layer (0–20 cm)				Surface soil layer (0–20 cm)			
	without rotation				Varimax rotated			
	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
pH(H ₂ O)	−0.49	−0.10	0.16	−0.70	0.10	−0.02	0.15	0.86
N _{tot} (mg kg ^{−1})	−0.73	−0.32	−0.19	0.06	0.37	0.59	0.34	0.30
P _{tot} (mg kg ^{−1})	−0.53	−0.32	0.01	−0.55	0.09	0.27	0.12	0.77
TOC (mg kg ^{−1})	−0.70	−0.48	−0.16	0.22	0.25	0.74	0.39	0.18
As (mg kg ^{−1})	0.03	0.75	0.35	−0.14	0.17	−0.80	0.21	−0.01
Cu (mg kg ^{−1})	−0.67	0.34	−0.19	0.08	0.69	0.04	0.33	0.11
Cr (mg kg ^{−1})	−0.67	−0.18	0.51	0.38	−0.00	0.27	0.90	0.04
Ni (mg kg ^{−1})	−0.77	0.14	0.26	0.21	0.37	0.09	0.75	0.13
Cd (mg kg ^{−1})	−0.54	0.26	0.55	−0.07	0.12	−0.27	0.70	0.30
Zn (mg kg ^{−1})	−0.55	0.40	−0.58	−0.07	0.88	0.06	−0.05	0.13
Pb (mg kg ^{−1})	−0.57	0.43	−0.39	0.08	0.80	0.01	0.14	0.02
Eigenvalue	3.94	1.60	1.36	1.07	2.30	1.76	2.34	1.58
Variance explained (%)	36%	14%	12%	10%	21%	16%	21%	14%
Cumulative variance (%)	36%	50%	62%	72%	21%	37%	58%	72%
Variable	Subsurface soil layer (20–40 cm)				Subsurface soil layer (20–40 cm)			
	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
pH(H ₂ O)	−0.50	−0.17	0.32	−0.61	0.86	0.09	0.05	−0.10
N _{tot} (mg kg ^{−1})	−0.67	−0.10	0.11	−0.08	0.47	0.30	0.40	0.11
P _{tot} (mg kg ^{−1})	−0.53	−0.34	0.31	−0.38	0.76	0.06	0.15	0.16
TOC (mg kg ^{−1})	−0.66	−0.48	0.13	0.18	0.42	0.14	0.50	0.53
As (mg kg ^{−1})	0.01	0.66	−0.38	−0.50	−0.04	0.12	0.03	−0.90
Cu (mg kg ^{−1})	−0.66	0.35	0.15	0.22	0.13	0.72	0.31	0.03
Cr (mg kg ^{−1})	−0.68	−0.30	−0.55	0.20	0.06	−0.00	0.93	0.13
Ni (mg kg ^{−1})	−0.80	0.05	−0.25	0.20	0.14	0.41	0.75	0.05
Cd (mg kg ^{−1})	−0.60	0.11	−0.55	−0.27	0.22	0.07	0.72	−0.43
Zn (mg kg ^{−1})	−0.52	0.51	0.48	0.12	0.21	0.86	−0.05	0.00
Pb (mg kg ^{−1})	−0.44	0.58	0.16	0.20	−0.04	0.75	0.11	−0.14
Eigenvalue	3.79	1.65	1.32	1.06	1.85	2.12	2.48	1.37
Variance explained (%)	34%	15%	12%	10%	17%	19%	23%	12%
Cumulative variance (%)	34%	49%	61%	71%	17%	36%	59%	71%

sampling points. It has to be strongly stressed that all values for points between main sampling sites were not obtained by performing measurements, but interpolated using the gridding algorithm mentioned earlier. Each sampling point was labeled in accordance to the codes given in Experimental and located exactly as in the site map (Fig. 1).

For surface soil layer, the highest contamination caused by vehicle and industrial burning is observed

close to Kramorie while caused by industrial activity is observed along a soil transect in the vicinity of Karnobat-Sarnevo-Drachevo line what is in agreement with location of the major industrial centers (Fig. 1). The highest soil acidification in the surface layer is observed close to Karageorg location, while the most intensive agricultural impact on the upper soil level is observed in the vicinity of Drachevo. The highest impact of vehicle and industrial burning,

Fig. 7 Biplot of the most informative factor scores: **a** PC2 vs PC3, **b** PC3 vs. PC4 (surface soil layer)

industrial production, and agriculture on the subsurface soil layer is observed in identical locations as discussed above for surface soil layer.

4 Conclusion

The study carried out indicates that in the coastal region of City of Burgas, Bulgaria, two various kinds

of soil could be identified according to their location: *rural/industrial* and *coast/mountains*. All soil samples collected at *rural/industrial* area were statistically stronger impacted by anthropogenic influence being enriched both in heavy metals (oil refinery, fuel combustion, local traffic net) and nutrients (agriculture). Samples collected in the area of Balkan Mountain as well as those collected in the coast were relatively less contaminated.

Fig. 8 Biplot of the most informative factor scores: **a** PC1 vs PC2, **b** PC1 vs. PC2 (subsurface soil layer)

Fig. 9 Isoline plot for factor scores values for surface soil layer after Kriging interpolation

Fig. 10 Isoline plot for factor scores values for subsurface soil layer after Kriging interpolation

Acknowledgments The authors acknowledge and express their thanks and appreciation to the Pomeranian Academy for financial support of this study (grant BW/7/9009028/10).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Andrade, J. M., Kubista, M., Carlosena, A., & Prada, D. (2007). 3-Way characterization of soils by Procrustes rotation, matrix-augmented principal components analysis and parallel factor analysis. *Analytica Chimica Acta*, 603(1), 20–29.
- Angelov, G. B. (2008). Heavy metal pollution in the Boatin Reserve (Bulgaria). *Turkish Journal of Botany*, 32, 155–160.
- Buszewski, B., & Kowalkowski, T. (2006). A new model of heavy metal transport in the soil using nonlinear artificial neural networks. *Environmental Engineering Science*, 23, 589–595.
- Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum Press.
- Cattle, J. A., McBratney, A. B., & Minasny, B. (2002). Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. *Journal of Environmental Quality*, 31, 1576–1588.
- Chang, T. K. (2000). Simulated annealing and kriging method for identifying the spatial patterns and variability of soil heavy metals. *Journal of Environmental Science and Health Part A*, 35(7), 1089–1115.
- Chepanova, L., Madgarova, I., & Aneva, Z. (2007). *Validation of a spectrophotometric method for determination of total phosphorus in soils and sediments using a microwave-assisted digestion*. Burgas, Bulgaria: Annual Reports of Assen Zlatarov University, 7, 36–43 (in Bulgarian).
- Chepanova, L., Aneva, Z., & Simeonov, V. (2008). Evaluation of aqua regia microwave-assisted digestion procedures for metal elements determination in environmental samples by FAAS and ICP-MS. *Ecological Chemistry and Engineering*, 15, 505–5119.
- Davies, B. E. (1997). Heavy metal contaminated soils in an old industrial area of Wales, Great Britain: source identification through statistical data interpretation. *Water, Air, and Soil Pollution*, 94, 85–98.
- Dinev, N., Banov, M., & Nikova, I. (2008). Monitoring and risk assessment of contaminated soils. *General and Applied Plant Physiology*, 34(3–4), 389–396.
- Einax, J. W., & Soldt, U. (1995). Geostatistical investigations of polluted soils. *Fresenius' Journal of Analytical Chemistry*, 351, 48–53.
- EVS-EN 13137 (2001) “Characterization of waste determination of total organic carbon (TOC) in waste, sludges and sediments”.
- Hervé, A., Factor rotations in factor analyses. <http://www.utd.edu/~herve/Abdi-rotations-pretty.pdf>. Accessed 17 January 2010.
- ISO 10381-6 (2009) “Soil quality-sampling-part 6: Guidance on the collection, handling and storage of soil under aerobic conditions for the assessment of microbiological processes, biomass and diversity in the laboratory”
- ISO 10390 (2005a) “Soil quality—determination of pH”
- ISO 11261 (1995) “Soil quality—determination of total nitrogen-modified Kjeldahl method”
- ISO 11464 (1994) “Soil quality—pretreatment of samples for physicochemical analyses”
- ISO 17025 (2005b) “General requirements for the competence of calibration and testing laboratories”
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151.
- Kemper, T., & Sommer, S. (2002). Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environmental Science & Technology*, 36, 2742–2747.
- Largueche, F.-Z. B. (2006). Estimating soil contamination with Kriging interpolation method. *American Journal of Applied Sciences*, 3(6), 1894–1898.
- Lin, Y. P., Chang, T. K., Shih, Ch W, & Tseng, Ch H. (2002). *Factorial and indicator Kriging methods using a geographic information system to delineate spatial variation and pollution sources of soil heavy metals*, 42(8), 900–909.
- Liu, T. L., Juang, K. W., & Lee, D. Y. (2006). Interpolating soil properties using Kriging combined with categorical information of soil maps. *Soil Science Society of America Journal*, 70, 1200–1209.
- Malinova, L. (2002). Contents of total and mobile forms of heavy metals in soil from stationary sample plots at Yundola, Vitinia and Staro Oriahovo. *Journal of Environmental Protection and Ecology*, 3(4), 834–841.
- Mandal, A., & Sengupta, D. (2006). Measurement of soil contamination due to heavy metals around a coal-fired thermal power plant in India. *Environmental Geology*, 51, 409–420.
- Martínez, J., Llamas, J., Miguel, E., Rey, J., & Hidalgo, M. (2008). Soil contamination from urban and industrial activity: example of the mining district of Linares (southern Spain). *Environmental Geology*, 54(4), 669–667.
- Massart, D. L., & Kaufman, L. (1997). *The interpretation of analytical chemical data by the use of cluster analysis*. New York: Wiley Interscience.
- Oliver, M. A., & Webster, R. (1991). How geostatistics can help you? *Soil Use and Management*, 7, 206–217.
- Perez Pavon, J. L., Garcia Pinto, C., Guerrero Pena, A., & Moreno Cordero, B. (2008). Headspace mass spectrometry methodology: application to oil spill identification in soils. *Analytical and Bioanalytical Chemistry*, 391, 599–607.
- Sanghi, R., & Sasi, K. S. (2001). Pesticides and heavy metals in agricultural soil of Kanpur. *India Bulletin of Environmental Contamination and Toxicology*, 67, 446–454.
- Schulin, R., Curchod, F., & Mondeshka, M. (2007). Heavy metal contamination along a soil transect in the vicinity of the iron smelter of Kremikovtzi (Bulgaria). *Geoderma*, 140, 52–61.
- Simeonov, V., Einax, J., Tsakovski, S., & Kraft, J. (2005). Multivariate statistical assessment of polluted soils. *Central European Journal of Chemistry*, 3(1), 1–9.

- Singh, K. P., Malik, A., Basant, A., & Ojha, P. (2008). Vertical characterization of soil contamination using multi-way modeling—a case study. *Environmental Monitoring and Assessment*, 146, 19–32.
- Škrbić, B., Some toxic substances in surface soil in Novi Sad area, http://www.bodenkunde2.uni-freiburg.de/eurosoil/abstracts/id993_Skrbic_full.pdf. Accessed 23 Jan 2010.
- Stanimirova, I., Zehl, K., Massart, D. L., Vander Heyden, Y., & Einax, J. W. (2006). Chemometric analysis of soil pollution data using the Tucker N-way method. *Analytical and Bioanalytical Chemistry*, 385, 771–779.
- Stanimirova, I., Kita, A., Malkowski, E., John, E., & Walczak, B. (2009). N-way exploration of environmental data obtained from sequential extraction procedure. *Chemometrics and Intelligent Laboratory Systems*, 96(2), 203–209.
- Stefanov, K., Seizova, K., Yanishlieva, N., Marinova, E., & Popov, S. (1995). Accumulation of lead, zinc and cadmium in plant seeds growing in metalliferous habitats in Bulgaria. *Food Chemistry*, 54, 311–313.
- Terrado, M., Kuster, M., Raldua, D., Lopez De Alda, M., Barcelo, D., & Tauler, R. (2007). Use of chemometric and geostatistical methods to evaluate pesticide pollution in the irrigation and drainage channels of the Ebro river delta during the rice-growing season. *Analytical and Bioanalytical Chemistry*, 387, 1479–1488.
- Terzano, R., Spagnuolo, M., Vekemans, B., De Nolf, W., Janssens, K., Falkenberg, G., et al. (2007). Assessing the origin and fate of Cr, Ni, Cu, Zn, Pb, and V in industrial polluted soil by combined microspectroscopic techniques and bulk extraction methods. *Environmental Science & Technology*, 41(19), 6762–6769.
- Thurstone, L. L. (1974). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Vandeginste, B. G. M., Massart, D. L., Buydens, L. M. C., De Jong, S., Lewi, P. J., & Smeyers Verbeke, J. (1997). *Handbook of chemometrics and qualimetrics; data handling in science and technology, parts A and B*. Amsterdam: Elsevier.
- Webster, R., & Oliver, M. A. (2006). Sample adequate to estimate variograms of soil properties. *European Journal of Soil Science*, 43(1), 177–192.
- Xiaoping, L., & Chunghang, H. (2007). Environment impact of heavy metals on urban soil in the vicinity of industrial area of Baoji city, P.R. China. *Environmental Geology*, 52, 1631–1637.
- Yasrebi, J., Saffari, M., Fathi, H., Karimian, N., Moazallahi, M., & Gazni, R. (2009). Evaluation and comparison of ordinary Kriging and inverse distance weighting methods for prediction of spatial variability of some soil chemical parameters. *Research Journal of Biological Sciences*, 4(1), 93–102.
- Yun, S. T., Choi, B. I., & Lee, P. K. (2000). Distribution of heavy metals (Cr, Cu, Zn, Pb, Cd, As) in roadside sediments, Seoul Metropolitan City, Korea. *Environmental Technology*, 21(9), 989–1000.