Geostatistical and multivariate statistical analysis of heavily and manifoldly contaminated soil samples

Kristin Schaefer · Jürgen W. Einax · Vasil Simeonov ·
Stefan Tsakovski

Abstract The surroundings of the former Kremikovtzi steel mill near Sofia (Bulgaria) are influenced by various emissions from the factory. In addition to steel and alloys, they produce different products based on inorganic compounds in different smelters. Soil in this region is multiply contaminated. We collected 65 soil samples and analyzed 15 elements by different methods of atomic spectroscopy for a survey of this field site. Here we present a novel hybrid approach for environmental risk assessment of polluted soil combining geostatistical methods and source apportionment modeling. We could distinguish areas with heavily and slightly polluted soils in the vicinity of the iron smelter by applying unsupervised pattern recognition methods. This result was supported by geostatistical methods such as semivariogram analysis and kriging. The modes of action of the metals examined differ significantly in such a way that iron and lead account for the main pollutants of the iron smelter, whereas, e.g., arsenic shows a haphazard distribution. The application of factor analysis and source-apportionment modeling on absolute principal component scores revealed novel information about the composition of the emissions from the different stacks. It is possible to estimate the impact of every element examined on the pollution due to their emission source. This investigation allows an objective assessment of the different spatial distributions of the elements examined in the soil of the Kremikovtzi region. The geostatistical analysis illustrates this distribution and is supported by multivariate statistical analysis revealing relations between the elements.

Introduction

Many case studies of diverse metal pollution due to atmospheric deposition have been reported [1–3]. Mostly, quantitative analyses of the soil properties were performed. Furthermore, studies focusing on the interactions between soil and the biosphere (e.g., fungi, mosses, plants) have been done [4, 5]. However, investigations on iron smelters are rare. Currently, the volatile organic compounds in the ambient air of steel mills [6] are being investigated. As in most soil pollution studies, we carried out a quantitative analysis of the soil properties. However, in the present study we also applied various statistical approaches to prove the significance of our observations.

In the last decades of the twentieth century the annual production of the Kremikovtzi steel mill in Bulgaria was about one million tons of steel. In addition, the steel mill produced alloys and inorganic compounds, such as lime, barite, and dolomite. The 2007 investigation by Schulin et al. [7] considered only the vertical distribution of 19 elements in two different soil types in the southern region of the metallurgical complex of Kremikovtzi. The authors suggested that the soil types investigated (chromic luvisols and alluvial fluvisols) contain different metal concentrations due to their parent materials rather than due to the

K. Schaefer (✉) · J. W. Einax
Department of Environmental Analysis, Institute of Inorganic
and Analytical Chemistry, Friedrich Schiller University of Jena,
Lessingstraße 8,
07743 Jena, Germany
e-mail: k.schaefer@uni-jena.de

V. Simeonov · S. Tsakovski
Faculty of Chemistry, University of Sofia,
J Bourchier Blvd 1,
1126 Sofia, Bulgaria

emission from the steel mill. Furthermore, they found no variation of the element concentrations dependent on the depth. It was concluded that the distributions of As, Cr, and Ni were caused by geogenic processes and the distributions of Pb, Zn, and Cu were caused by atmospheric deposition. In contrast, the present study also includes the analysis for other pollutants such as tracers for steel, iron alloys, and the main inorganic products. We collected 65 samples in an area of about 64 km^2 in the Kremikovtzi steel mill region. After determination of 15 elements, we used geostatistical methods, cluster analysis, factor analysis, and source apportionment modeling to analyze the complex data set.

Source apportionment modeling originates from investigations in air quality studies [8] and is not well known in soil science. Recent reports have pointed out that it is difficult to distinguish between contributions from different sources to the pedosphere [9]. Davis et al. [10] used the combination of principal component analysis and geostatistics to find different sources but did not calculate a receptor model. Furthermore, traditional monitoring data interpretation methods such as cluster analysis, principal component analysis, and source apportionment modeling were used to identify and apportion sources of soil pollution around an iron smelting facility but were not combined with geostatistics [11]. The present research closes this gap by showing the possibility of combining source apportionment modeling and geostatistics. The only previous study dealing with the combination of these two methods did not have a preliminary step able to select the appropriate pollution tracers [12]. In the present investigation, geostatistical analysis as a preliminary step for selection of tracers is introduced.

Material and methods

Study area, soil sampling, preparation, and analysis

The Kremikovtzi steel mill is located about 20 km northeast of the Bulgarian capital, Sofia. We collected 65 samples in an area of about 64 km^2, of which 16 sampling points were located within the boundaries of the steel mill (Fig. 1). Owing to the anthropogenic interventions in this region, we chose an irregular grid fitted to the local conditions as the sampling design. Taking the samples on a regular sampling grid was impossible since buildings, streets, mounds, plants, or cultivated fields shape this area. Furthermore, the complete coverage of the region investigated was requested. To ensure the comparability of the samples, we took the samples on grassland from soil not treated by fertilizers and with comparable orographic conditions.

We took component soil samples, consisting of five subsamples, from the upper 0–20-cm layer. The soil
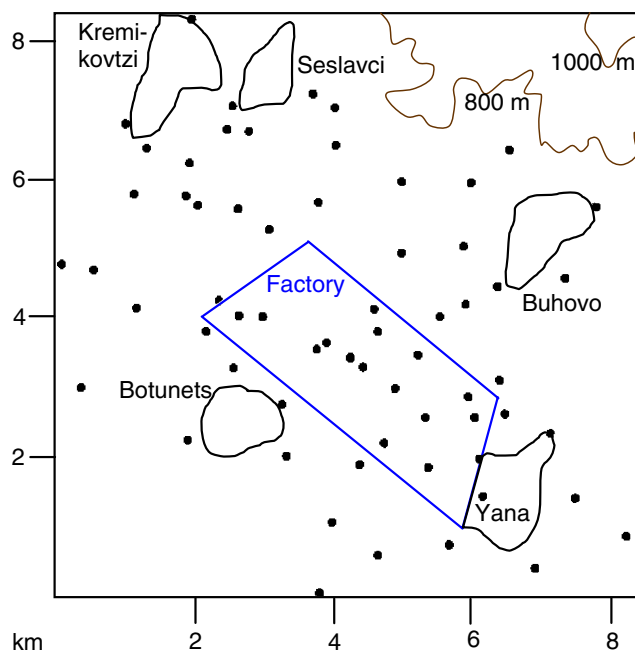

Fig. 1 Study area with the 65 sampling points in the Kremikovtzi steel mill region, Sofia, Bulgaria

samples were dried, homogenized, and passed through a 2-mm sieve. We performed a microwave (power 1,200 W) aqua regia digestion with a mixture of 21 mL HCl (12 M) and 7 mL HNO$_3$ (15.8 M) according to the German standard [13] with 0.5 g of the soil twice for each sampling site. After cooling, we made up the solutions to 100 mL with dilute HNO$_3$ (0.5 M). We determined the concentrations of 15 elements using a 5100 PC instrument from PerkinElmer for flow injection hydride generation atomic absorption spectroscopy, an AAS-3110 instrument from PerkinElmer for flame atomic absorption spectroscopy, and an Elan 6000 instrument from PerkinElmer for inductively coupled plasma mass spectrometry.

All metals examined could be detected in concentrations higher than the detection limits (Table 1) calculated according to the indirect method of the German standard [14]. Analyzing the certified reference material IAEA/Soil-7 (International Atomic Energy Agency, Vienna) verified the trueness (P=95%) of the analytical methods.

Overview of the theoretical fundamentals

Geostatistical methods

Geostatistical methods were used to analyze the spatial dependence of data and their correlation, and finally to estimate the concentration of the element of interest at an unsampled location. The statistical theory behind this analysis is the theory of regionalized variables, which means that the

Table 1 Overview of detection limits and standard deviations of the methods applied for the elements determined

Element	Method	Unit	Detection limit	Standard deviation of method
As	FI-Hydride AAS	µg/L	2.09	0.216
Ca	Flame AAS	mg/L	0.37	0.04
Cd	ICP-MS	µg/L	1.11	0.088
Co	ICP-MS	µg/L	5.50	0.777
Cr	ICP-MS	µg/L	1.90	0.359
Cu	ICP-MS	µg/L	5.09	0.902
Fe	Flame AAS	mg/L	3.61	0.243
K	Flame AAS	mg/L	0.45	0.042
Mg	Flame AAS	mg/L	0.15	0.015
Mn	Flame AAS	mg/L	0.98	0.108
Na	Flame AAS	mg/L	0.27	0.03
Ni	ICP-MS	µg/L	2.64	0.473
Pb	ICP-MS	µg/L	4.46	0.679
V	ICP-MS	µg/L	9.22	1.22
Zn	ICP-MS	µg/L	8.03	1.21

AAS atomic absorption spectroscopy, *ICP-MS* inductively coupled plasma mass spectrometry, *FI* flow injection

variance between two realizations of the random function $Z(X)$ depends only on the step width h between those locations and not on the absolute location in space [15]. Currently, geostatistical methods are used in different environmental analyses [16, 17], especially soil sciences [18, 19], or even human health studies [20]. There are two important parts of geostatistical analysis: first, the semivariogram analysis and, second, the estimation by means of kriging.

For semivariogram analysis the semivariance $\gamma(h)$ is calculated as

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(x_i) - z(x_{i+h})]^2, \tag{1}$$

where h is the step width, $\gamma(h)$ the semivariance, $n(h)$ the number of sample pairs at each step width h, $z(x_i)$ the realization of the random function at location i, and $z(x_{i+h})$ the realization of the random function at location $(i+h)$.

The semivariogram is the graphical representation of the semivariance $\gamma(h)$ as a function of the step width h. After the experimental semivariogram has been calculated, a theoretical semivariogram is modeled by fitting a function to the data points. We used exponential, Gaussian, linear, power, and spherical models in this investigation. The nugget effect, representing the microinhomogeneity of the data, the sill, which represents the variance of the data, and the range, up to which the data set is spatially correlated, are characterizing

parameters of those functions. Linear models are defined by a slope and may have a nugget effect. They are not characterized by a range or a sill, which means that they never reach a certain variance at a given distance. The parameters can be used to detect spatial structures within the data.

Kriging estimation can be used to describe the spatial distribution of the elements within a map. The results are usually represented as isoline plots. This technique uses a weighted moving average interpolation method, which means that closer points have more influence on the estimated value than more distant ones. The kriging estimation is calculated as follows:

$$z^*(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i), \tag{2}$$

where $z^*(x_0)$ is the estimate of the unknown value $z(x_0)$, λ_i are the weights of known neighboring points $z(x_i)$, and $z(x_i)$ are the known neighboring points.

The weights λ_i are calculated on the basis of the semivariogram function, so that the kriging variance is minimized:

$$\sigma_k^2 = \sum_{i=1}^{n} \lambda_i \gamma[z(x_i) - z(x_0)] + \mu, \tag{3}$$

where σ_k^2 is the kriging variance, $z(x_i) - z(x_0)$ is the distance between two known values $z(x)$ at sample location i and 0, μ is the Lagrange multiplier, and $\sum_{1=1}^{n} \lambda_i = 1$.

More detailed mathematical information can be found in [21, 22].

The results can be verified with the help of full cross-validation. In succession, all sampling points are eliminated and newly estimated according to Eqs. 2 and 3. To evaluate the cross-validation, two parameters were used. First, Pearson's correlation coefficient (r) calculated according to Eq. 4 should be 1 for a perfect estimation:

$$r = \frac{\sum_i^n [z(x_i) - z(\bar{x}_i)][z^*(x_i) - z^*(\bar{x}_i)]}{\sqrt{\sum_i^n [z(x_i) - z(\bar{x}_i)]^2 \sum_i^n [z^*(x_i) - z^*(\bar{x}_i)]^2}}, \tag{4}$$

where $z(x_i)$ is the observed known value, $z(\bar{x}_i)$ is the mean of the observed known values, $z^*(x_i)$ is the estimated known value, $z^*(\bar{x}_i)$ is the mean of the estimated known values, and n is the number of observations.

Furthermore, as a measure of uncertainty, the root mean square error of prediction (RMSEP) can be calculated for the cross-validation:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} [z(x_i) - z^*(x_i)]^2}{n}}. \tag{5}$$

For geostatistical analysis we used the software package R version 2.9.0 with the package gstat and Surfer 9.0.

Source apportionment modeling on absolute principal component scores

Source apportionment modeling can be used to quantify the contribution of different sources to the soil in the study area [8]. This technique is widely used for describing air pollution [23, 24] and aquatic systems [25, 26]. A multiple linear regression of the total concentrations on the absolute principal component scores (APCS) is done to use this method. First, a factor analysis is necessary to calculate the APCS. Factor analysis is used to reduce the dimensionality of a data set and to find hidden correlations or structures. Therefore, the matrix of all measured values $\mathbf{X}$ is divided into the product of a matrix of factor loadings $\mathbf{A}$ and a matrix of factor scores $\mathbf{F}$, and a matrix of residuals $\mathbf{Q}$:

$$\mathbf{X}_{mn} = \mathbf{A}_{mt}\mathbf{F}_{tn} + \mathbf{Q}_{mn}, \tag{6}$$

where m is the number of features, n is the number of objects, and t is the number of factors.

The covariance matrix of all features is computed, and then an eigenvalue analysis is performed to extract the factors. In this work we used varimax rotation to enable an easier interpretation of the factors [27].

The resulting factor scores are used to compute APCS for each sampling point. Z-scores have to be created, because the factor analysis is based on normalized data (Eq. 7). These Z-scores represent the absolute zero concentration at a so-called "zero day," on which none of the sources have yet influenced the samples.

$$(z_0)_i = \frac{0 - \overline{x}}{s}, \tag{7}$$

where s is the standard derivation, $\overline{x}$ is the mean, and $(z_0)_i$ is the concentration on the "zero day."

Absolute zero principal component scores and furthermore APCS are calculated with these resulting $(z_0)_i$ and the factor score coefficients. This step is followed by multiple linear regression of the total mass as a function of APCS as

$$M_k = \zeta_0 + \sum_{j=i}^{p} \zeta_j \mathrm{APCS}_{jk}, \tag{8}$$

where M_k is the particle mass record during observation k, ζ is the particle mass contribution, APCS_{jk} is the rotated APCS for component j at observation k, and $j...p$ is the number of pollution sources.

Furthermore, the element source profiles of the estimated source impact can be calculated. A detailed mathematical description can be found in [28]. We used the software package STATISTICA 6.1 for source apportionment modeling.

Results and discussion

Chemical analysis

The mean values and distribution measures of the elements examined are listed in Table 2. We found the range of the element concentration to be different. The values for Cd, Co, Cr, Cu, Fe, K, Na, Ni, and V were within 1 order of magnitude. The concentration of Ca ranged from 0.07 to 153 mg/g (4 orders of magnitudes) owing to the geogenic origin of this metal. In contrast, Cd, Mn, Pb, and Zn had 25–41 times as high concentrations in polluted samples as in unpolluted ones owing to their anthropogenic origin.

To evaluate these results considering the impact of the pollution, the optimum and action values of the new Dutchlist are presented for As, Cd, Co, Cr, Cu, Ni, Pb, and Zn in Table 3 [29]. The Dutchlist is seen as the unofficial guideline for risk assessment.

Over 90% of the sampling sites have higher contents of Cu and Pb than the optimum threshold (Table 3). Half of the sampling points exceed the optimum threshold for As, Cd, or Zn. For Co, Cr, and Ni the optimum threshold is exceeded for a couple of sampling sites. The situation for the action threshold is completely different. About 40% of the sampling points are highly contaminated with As and only 5–12% are contaminated with Cu, Pb, and Zn. The comparison of the results with another investigated iron-

Table 2 Mean values and distribution measures of the elements analyzed

Element	Unit	Mean value	Minimum value	Maximum value	Standard deviation
As	μg/g	56.0	6.99	154	34.3
Ca	mg/g	14.9	0.07	153	29.4
Cd	μg/g	1.17	0.20	6.07	1.04
Co	μg/g	16.4	9.39	28.5	4.28
Cr	μg/g	68.2	30.3	120	17.5
Cu	μg/g	85.7	34.9	292	56.8
Fe	mg/g	58.4	22.0	301	47.8
K	mg/g	7.4	1.71	11.8	2.31
Mg	mg/g	4.96	0.76	13.1	2.78
Mn	mg/g	4.30	0.58	15.4	3.95
Na	mg/g	0.313	0.108	1.42	0.216
Ni	μg/g	41.9	12.2	106	17.8
Pb	mg/g	0.316	0.039	1.62	0.349
V	μg/g	72.6	30.6	126	18.7
Zn	mg/g	0.236	0.055	1.67	0.256

Table 3 Number of sampling points exceeding the thresholds for As, Cd, Co, Cr, Cu, Ni, Pb, and Zn of the new Dutchlist (threshold in *parentheses*)

Element	No. of sampling sites exceeding the optimum value	No. of sampling sites exceeding the action value
As	46 (29 μg/g)	26 (55 μg/g)
Cd	30 (0.8 μg/g)	0 (12 μg/g)
Co	11 (20 μg/g)	0 (240 μg/g)
Cr	3 (100 μg/g)	0 (380 μg/g)
Cu	60 (36 μg/g	6 (190 (μg/g)
Ni	8 (35 μg/g)	0 (210 μg/g)
Pb	62 (0.085 mg/g)	8 (0.53 mg/g)
Zn	33 (0.14 mg/g)	3 (0.72 mg/g)

smelting facility in the UK shows a different kind of pollution from that found in the Kremikovtzi region. Whereas the area polluted with Cd, Cr, Cu, Ni, and Zn is smaller for the Kremikovtzi region, the area polluted with As is almost the same [30]. Furthermore, the maximum concentration of Pb is of about the same magnitude. The comparison with a German iron smelter shows comparable maximum concentrations for Pb and Zn, whereas the maximum concentrations for Cd, Cr, and Ni are lower for the Kremikovtzi region as well [31]. These comparisons suggest that Cr and Ni are not used for steel alloys in the Kremikovtzi region.

Geostatistical analysis

We used kriging estimation to interpolate the concentrations of the elements examined in the areas between sampling points. First, a semivariogram analysis for all elements was necessary. The elements were not transformed because the results should be compared with the source apportionment modeling. We used a visual fit of the possible theoretical semivariogram to the experimental semivariogram to select an adequate model. To confirm those results we performed a cross-validation for ordinary kriging. The correlation coefficients and the relative RMSEP values were estimated for each model chosen (see Table 4). The semivariograms showed a typical behavior for elements in heavily polluted areas [32, 33].

For most of the elements, a Gaussian model was found to be the most appropriate. In addition to the Gaussian model, we used three spherical models, three exponential models, one linear model, and one power model. The range for all elements except Cu and V was around 1 km or less. So, there was no high spatial correlation of the elements over the complete region.

To evaluate the results of the semivariogram analysis, we used the correlation coefficients and the RMSEP values. For a perfect estimation the correlation coefficient would be 1. For eight elements the correlation coefficients are above 0.6, which is acceptable in environmental analysis. For some elements the correlation coefficients are around 0.4 owing to the random, rather than normal distribution of these elements. The estimated RMSEP values are small for elements with a small concentration range and the RMSEP values were unsatisfactory for elements with a high range. For Ca, with a RMSEP of 179%, we found concentrations from 0.07 to 153 mg/g and for Co, with a RMSEP of 24%, we found concentrations from 9.39 to 28.5 μg/g (see Table 2). To avoid these problems, the calculations could be

Table 4 Results of semivariogram analysis for all elements including the results from cross-validation after kriging estimation for the given model

Element	Semivariogram				Cross-validation	
	Type	Nugget	Partial sill	Range (km)	r	Relative RMSEP (%)
As	Gaussian	95	740	0.74	0.62	47
Ca	Gaussian	100	700	0.6	0.41	176
Cd	Gaussian	0	1.2	0.6	0.84	49
Co	Spherical	0	18	0.9	0.38	24
Cr	Linear	225	25[a]		0.40	23
Cu	Spherical	0	4,500	2.8	0.85	34
Fe	Gaussian	880	2,200	1.2	0.84	44
K	Spherical	1.8	3.9	1.1	0.39	29
Mg	Gaussian	4.7	4.5	1.4	0.44	51
Mn	Gaussian	4.5	12	1.0	0.80	55
Na	Exponential	0	0.05	0.5	0.63	54
Ni	Exponential	100	220	0.74	0.05	43
Pb	Gaussian	0.025	0.13	1.0	0.88	51
V	Power	0	300	2.0	0.30	24
Zn	Exponential	0	0.09	0.72	0.64	82

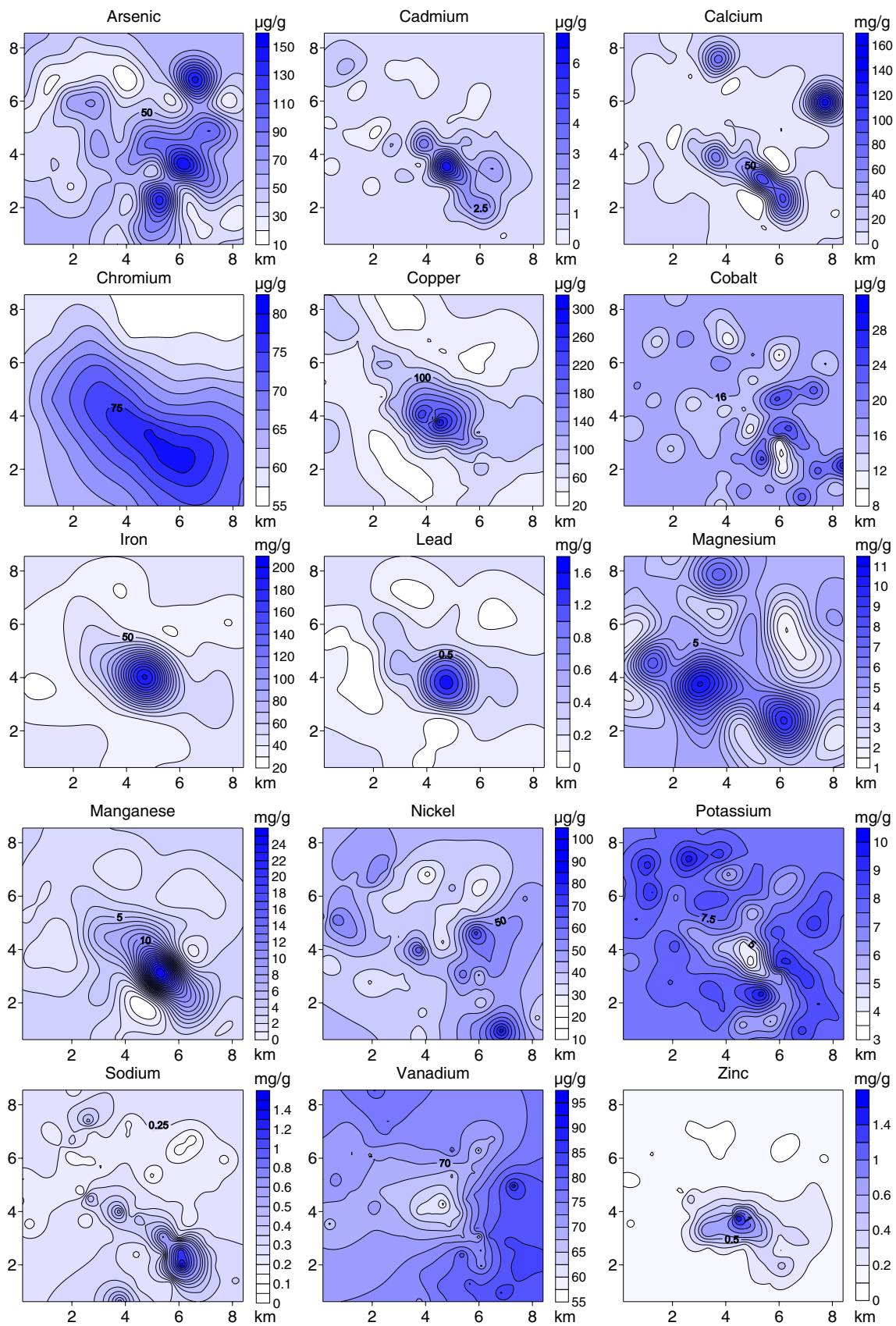*RMSEP* root mean square error of prediction

[a] Slope

Fig. 2 Isoline plots for all elements after kriging estimation
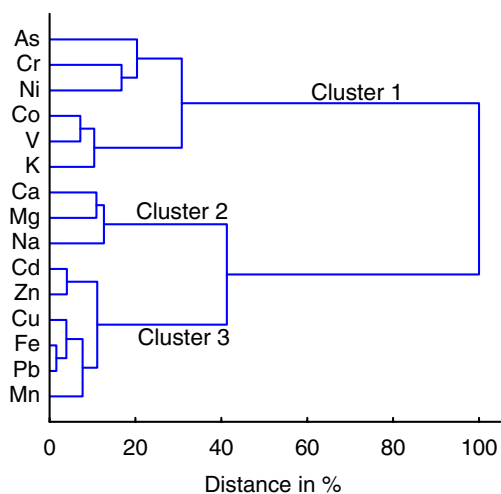
Fig. 3 Dendrogram of cluster analysis for all variables (Ward's method)

carried out with the lognormal transformed data. However, a calculation with the lognormal transformed data was not possible in this work because we wanted to apply the results of geostatistics to the source apportionment modeling. In summary, it can be said that the semivariogram analysis showed satisfactory results for the elements As, Cd, Cu, Fe, Mn, Na, Pb, and Zn.

In Fig. 2 the kriging estimation for all elements based on the semivariogram models listed in Table 4 are displayed as isoline plots to visualize the results.

The distributions of the elements differ a lot. Whereas Cd, Cu, Fe, Pb, and Zn have the highest concentration in the center of the factory, As, Co, K, Ni, and V show a rather haphazard distribution. K and V, for example, have their lowest concentration in the center of the steel mill. The region is not polluted by Co and Ni even though these elements are used for steel coating. High concentrations of Ca, Cr, Mg, and Na were detected in the southeastern part of the factory. The emission of the elements Ca, Mg, and Na is connected with lime production. Furthermore, the concentrations of Ca and Mg are high in the northern part

Table 5 Factor loading matrix after varimax rotation

Element	Factor 1	Factor 2	Communality
Cd	0.802	0.412	0.833
Cu	0.915	0.166	0.814
Fe	0.931	0.020	0.909
Mg	0.260	0.770	0.437
Mn	0.738	0.556	0.864
Na	0.021	0.913	0.643
Pb	0.978	0.083	0.952
Zn	0.797	0.315	0.769
Explained variance	65%	17%	

Table 6 Source apportionment for the different sources for each variable

Element	r	Relative RMSEP (%)
Cd	0.90	37
Cu	0.93	24
Fe	0.93	29
Mg	0.81	32
Mn	0.92	36
Na	0.91	28
Pb	0.98	21
Zn	0.85	53

of this area owing to their geogenic origin. For these two elements there are two different factors responsible for their distribution: the lime production and their geogenic origin. Elements usually connected with steel production—such as Cd, Cu, Fe, Pb, and Zn—have high concentrations in the area of the stacks of the iron smelter. The isoline plots of Fe and Pb almost look alike, which indicates that those elements are highly correlated. One of the main reasons is that the ore is rich in Pb. It is notable that Mn shows high concentrations in the center and as well in the southeastern part of the factory. Despite the lime production—indicated by Ca, Mg, and Na—the alloy production is located in this area. This explains the high concentrations of Cr and Mn in this southeastern part of the factory.

Cluster analysis

Cluster analysis is an unsupervised learning technique. In this study we used the agglomerative hierarchical cluster analysis according to Ward [34]. Applying cluster analysis for the elements examined shows a separation between elements emitted by the steel mill, which can also be distinguished into iron production (cluster 3), lime produc-
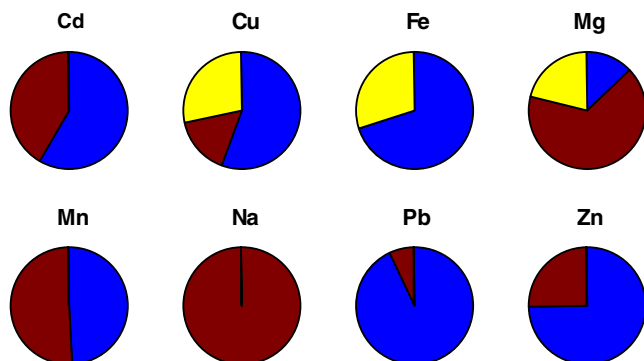

Fig. 4 Contribution of each source for the elements included in the source-apportionment modeling (*blue* source 1, *red* source 2, *yellow* unexplained)

tion (cluster 2), and elements originating from diffuse sources (cluster 1) (Fig. 3).

The elements in cluster 1 are those with a haphazard distribution according to the kriging estimation. The fact that As, Cr, and Ni originate from geogenic sources [7] was confirmed independently by cluster analysis. Usually Cr and Ni would be expected to be grouped with Fe because they are essential constituents in steel alloys. There are two specific production-related reasons as to why this is not the case in the Kremikovtzi region. Firstly, the Ni content is very low in the ore [35]. Secondly, the amount of low-alloy steel production is small and the electric arc furnaces are better equipped with dust-removing filters than the other production facilities. The elements in clusters 2 and 3 are the ones with high concentrations in the areas of the factory used for iron production (Cd, Cu, Fe, Mn, Pb, and Zn) and for lime production (Ca, Mg, and Na).

Source apportionment modeling

Factor analysis

We performed a factor analysis to conduct source apportionment modeling on principal components. The results of the geostatistical and cluster analysis showed that a reduction of variables for the source apportionment modeling was necessary. The factor analysis was performed with the elements Cd, Cu, Fe, Mg, Mn, Na, Pb, and Zn because those elements were not haphazardly distributed. Ca was removed because of its high RMSEP. The factor loadings after varimax standard rotation are listed in Table 5.

It is possible to distinguish between two factors explaining about 82% of the variance together. Factor 1 represents the elements that are emitted by the steel factory and factor 2 represents the elements emitted by the lime production. The communalities are high except for Mg.

Absolute principal component scores

We performed source apportionment modeling for the two main factors detected using factor analysis. The two factors represent the source iron production (factor 1) and lime–alloy production (factor 2). The resulting model has a correlation coefficient of 0.94 and an RMSEP of 26%. Source 1 explains 65% of the computed model, source 2 explains 12%, and about 23% cannot be connected to either source. After identifying the two sources, we considered the impact of them for each metal. Therefore, we calculated the contribution of each polluting source to the metal concentration. The correlation coefficients and the RMSEP values for each metal are presented in Table 6. The impact of each source on the metals examined is illustrated in Fig. 4.

Except for Zn, the RMSEP values are below 40%. The correlation coefficients are above 0.8 for all elements except Mn and Zn.

Only the concentrations of Cu, Fe, and Mg are influenced by sources outside the Kremikovtzi steel mill. Fe and Pb are mainly emitted by source 1 (more than 70%) and just slightly by source 2. Na and Mg are mainly emitted in the southeastern part (source 2) of the factory. The other elements—Cd, Cu, Mn, and Zn—are emitted by both sources but with different intensities. Whereas Cd and Mn are emitted by both sources almost equally, Cu and Zn are emitted threefold more by source 1 than by source 2. Those results prove as well that source 1 is the iron production and source 2 is a combination of lime and alloy production.

Conclusion

The combination of the geostatistical analysis and the source apportionment modeling is very beneficial. The kriging plots allow a graphical description of the area, which helps one interpret the emission sources. Furthermore an objective variable reduction can be accomplished. As source apportionment modeling with all elements was unsatisfactory, finding an objective tool which allows an element reduction was very important. We pointed out two main emission sources with the help of source apportionment modeling. One is the iron smelter and the other one comes from the southeastern part of the factory, which is a combination of lime and manganese–iron–alloy production. To separate those two in more detail, a study including more sampling points inside the territory of the steel mill should be done.

References

1. Koptsik S, Koptsik G, Livantsova S, Eruslankina L, Zhmelkova T, Vologdina Zh (2003) J Environ Monit 5:441–450
2. Yukselen MA (2002) Environ Geol 42:597–603
3. Vike E (2005) Water Air Soil Pollut 160:145–159
4. Cabala J, Krupa P, Misz-Kennan M (2009) Water Air Soil Pollut 199:139–149
5. Frontasyeva MV, Steinnes E (1995) Analyst 120:1437–1440
6. Ciaparra D, Aries E, Booth MJ, Anderson DR, Almeida SM, Harrad S (2009) Atmos Environ 43:2070–2079
7. Schulin R, Curchod F, Mondeshka M, Daskalova A, Keller A (2007) Geoderma 140:52–61

8. Hopke PK (1995) In: Einax J (ed) The handbook of environmental chemistry, vol 2, part H: chemometrics in environmental chemistry - applications. Berlin, Springer

9. Steinnes E, Frontasyeva MV, Gundorina SF, Pankratova YS (2008) Chem Anal (Warsaw) 53:877–886

10. Davis HT, Aelion CM, McDermott S, Lawson AB (2009) Environ Pollut 157:2378–2385

11. Simeonov V, Einax J, Tsakovski S, Kraft J (2004) Cent Eur J Chem 3:1–9

12. Wang K, Shen Y, Zhang S, Ye Y, Shen Q, Hu J, Wang X (2009) Environ Geol 56:1041–1050

13. DIN EN 13346:2000 (2001) Characterization of sludges – determination of trace elements and phosphorus – aqua regia extraction methods

14. DIN 32645 (2008) Chemical analysis – decision limit, detection limit and determination limit under repeatability conditions – terms, methods, evaluation

15. Matheron G (1965) Les variables régionalisées et leur estimation. Masson, Paris

16. Carlucci R, Lembo G, Maiorano P, Capezzuto F, Marano CA, Sion L, Spedicato MT, Ungaro N, Tursi A, D'Onghia G (2009) Estuar Coast Shelf Sci 83:529–538

17. Mohr K, Holy M, Pesch R, Schröder W (2009) Umweltwiss Schadst Forsch 21:454–469

18. Sigua GC, Hudnall WH (2008) J Soils Sediments 8:193–202

19. Zhao Y, Xu X, Darilek JA, Huang B, Sun W, Shi X (2009) Environ Geol 57:1089–1102

20. Goovaerts P (2008) Math Geosci 40:101–128

21. Journel AG, Huijbregts CH (1978) Mining geostatistics. Academic, New York

22. Cressie NCA (1991) Statistics for spatial data. Wiley, New York

23. Tauler R, Viana M, Querol X, Alastuey A, Flight RM, Wentzell PD, Hopke PK (2009) Atmos Environ 43:3989–3997

24. Olson DA, Hammond DM, Seila RL, Burke JM, Norris GA (2009) Atmos Environ 43:5647–5653

25. Simeonov V, Sarbu C, Massart DL, Tsakovski S (2001) Mikrochim Acta 137:243–248

26. Simeonov V, Einax JW, Stanimirova I, Kraft J (2002) Anal Bioanal Chem 374:898–905

27. Einax JW, Zwanziger HW, Geiss S (1997) Chemometrics in environmental analysis. VCH, Weinheim

28. Thursten GD, Spengler JD (1985) Atmos Environ 19:9–25

29. Layla Resources (2001) The new Dutchlist. http://www.contaminatedland.co.uk/std-guid/dutch-l.htm

30. Harber AJ, Forth RA (2001) Environ Geol 40:324–330

31. Einax JW, Soldt U (1999) Chemom Intell Lab Syst 46:79–91

32. Ersoy A, Yunsel TY, Cetin M (2004) Arch Environ Contam Toxicol 46:162–175

33. Myers JC (1997) Geostatistical error management: quantifying uncertainty for environmental sampling and mapping. Thomson, New York

34. Ward JH Jr (1963) J Am Stat Assoc 58:236–244

35. Vassileva M, Ruskov K (2006) Ann Univ Min Geol St Ivan Rilski 49:23–31