Modelling Heavy Metal Background Concentrations In Bulgarian Monitoring Soil Quality Network

Yotova G.[1], Padareva M.[1], Hristova M.[2], Astel A.[3], Dınev N.[2] And Tsakovskı S.[1,*]

[1] Faculty of Chemistry and Pharmacy, University of Sofia "St. Kliment Ohridski", J. Bourchier Blvd. 1, 1164 Sofia, Bulgaria

[2] N. Poushkarov Institute of Soil Science and Agroecology, Sofia, Bulgaria

[3] Environmental Chemistry Research Unit, Institute of Biology and Environmental Protection, Pomeranian University, 22b Arciszewskiego Str., Słupsk, Poland

*corresponding author:

e-mail: tsakovski@gmail.com

Abstract

Defining baseline concentrations for toxic elements in soils is essential for recognizing and managing soil pollution. Estimation of heavy metal baseline concentrations in soils, especially on national scale, has been fraught with many difficulties to cover different soil types, natural and anthropogenic impacts on regional level. The present study deals with univariate and multivariate statistical treatment of 358 samples from the Bulgarian monitoring network for soil quality assessment. The surface soil samples were analyzed for 8 critical heavy metals: Cu, Zn, Cd, Pb, Ni, Cr, As, Hg. Two decision tools (box-plot method and Principal component analysis) were applied for identification of contaminated sites. Comparison between both methods was performed using GIS based maps.

Keywords: heavy metals, soil pollution, baseline values, statistics

1. Introduction

Soil is a vital non-renewable resource which can act as sink and reservoir for many environmental pollutants. Surface soil pollution by heavy metals is a significant environmental problem due to intense anthropogenic activities like industrialization, urbanization and agriculture (Hu *et al*., 2013; Micó *et al*., 2007; Saa *et al*., 2011). Heavy metal concentrations in soils are subject of high concern because of their accumulative and non-decay properties which could lead to toxic levels for the ecosystem (Facchinelli *et al*., 2001). Moreover the fate and behavior of heavy metals depends on soil type, land use etc. Methods for identification of heavy metal contamination in networks without geochemical information concerning major element content in soil minerals are often based on statistical approaches. They are divided in univariate (Reimann *et al*., 2005) and multivariate (Redon *et al*., 2013; Lado *et al*., 2008). The univariate methods define baseline values of heavy metals using descriptive statistics of each heavy metal, whereas multivariate approaches are oriented to revealing of natural and anthropogenic sources which contribute to heavy metal concentrations in soils. In many studies spatial distribution of heavy metals or contribution sources are presented by GIS based maps using geostatistical approaches (Facchinelli *et al*., 2001; Lado *et al*., 2008; Zheng *et al*., 2008) which is a valuable tool for identification of contaminated regions. The data set produced by the FOREGS Geochemical Baseline Mapping programme is the biggest European geochemical atlas covered 26 European countries (Salminen *et al*., 2006). The Bulgarian monitoring soil quality network is not included in FOREGS and the lack of national baseline environmental geochemical data is a serious obstacle for policy makers and soil management activities. The present study focused on identification of heavy metal contaminated sites in Bulgarian Monitoring Soil Quality Network by the use of univariate (box-plot method) and multivariate (Principal component analysis) approaches. Comparison between both methods and estimated baseline values of FOREGS data set was performed.

2. Materials and methods

2.1. Sampling and chemical analysis

The 348 topsoil samples of Bulgarian soil quality monitoring network were collected at a depth of 0-20cm and in the intersections of an orthogonal 16-km grid over the territory of whole country. The collected samples were analyzed after aqua regia digestion (AR) for 8 toxic and heavy metals: Cu, Zn, Cd, Pb, Ni, Cr, As and Hg. The Cu, Zn, Cd, Pb, Ni, Cr were analyzed according to ISO 11466 (AAS) and As, Hg according to CEN/TS 16171 (ICP-MS).

2.2. Statistical methods

Box plot method: As a first method for identification of extreme values the box plot of each variable was used. The box plot consists of four equal parts (quartiles). Approximately 50% of the results lie between the upper and lower quartiles, also called upper and lower hinges, and form the central box. The inner fence is defined as: Inner fence = (upper hinge + 1.5*HW) – (lower hinge – 1.5*HW), where HW is the hinge width (HW = Q3 – Q1). Any value beyond the inner fence is defined as an outlier (Reimann *et al*., 2005). After removing the outliers for each of the variables separately, the mean concentrations and the standard deviation were calculated to establish the baseline values as mean + 2*stdev (Redon *et al*., 2013). A site with higher concentration than the determined baseline value could be defined as contaminated. *Principal component analysis*: Principal component analysis (PCA) is a data reduction method with a main goal to estimate the internal relations in the data set. The input variables (heavy metal concentrations) are converted in new ones which are better descriptors of the data structure. The new variables, called principal components (PC) or latent factors are linear combinations of the original variables. The main task of PCA is to resolve the data matrix (m×n) into factor loadings matrix (f×n), factor score matrix (m×f) and matrix of residuals (m×n). The participation of each of the original variables to the new formed principal components is reflected in the factor loadings matrix, since the new coordinates of each object in the new space of variables are presented in the factor score matrix. The PCA was performed in Varimax rotation mode which maximises the sum of the variances of the squared loadings.

3. Results and discussion

In the Table 1 basic statistic parameters of input data set (Bulgarian Monitoring Soil Quality Network) and FOREGS data set are presented. The input data median values of all critical heavy metals are higher than respected ones in FOREGS data set. Comparison based on 90th percentile levels, which is also used as a baseline threshold value, shows that contents of Cr, Cu and Hg in Bulgarian network are almost two times higher than in the European ones. It could be concluded that background heavy metal concentrations for Bulgarian soil network differ significantly from European scale values from the FOREGS data set. Baseline values for heavy metals in Bulgarian Monitoring Soil Quality Network are presented in the Table 2. The box plot derived baseline values are in good agreement with 90th percentile values of the Bulgarian network. Difference, higher than 10%, is observed only for Cu. It is an indication for peculiar Cu distribution in input data set due to local pollution or geochemical anomalies. The interpolated spatial distribution of critical heavy metal concentrations together with contaminated sites identified by box-plot method are shown on Figure 1. In general identified contaminated sites for most heavy metals are located in Northwest and Southeast regions of Bulgaria. These regions are affected by mine activity but presence of contaminated sites by Ni and Cr is an evidence for their different geochemical background. For precise identification of heavy metal pollution in above mentioned regions more knowledge concerning geogenic origin is needed. Outside these regions the spatial distribution of heavy metals reveals some “contaminated” places which could be a result of local pollution or anomalies. On presented maps some specific heavy metal distribution patterns could be observed. Similar spatial distributions are observed within couples Ni-Cr, Cd-Pb and in less extent within Cu-Zn. Such similarity is an indication for common origin of respected couples (Fig. 1). PCA was performed on Bulgarian soil quality monitoring data set. The first five principal components explain more than 80% of input data variance (Table 3). The second latent factor (16.84%) could be attributed to mixed “Cu-Zn pollution” because of its high positive correlation with Cu and Zn

Table 1. Basic statistics of toxic and heavy metals top soil contents after aqua regia digestion (mg kg^{-1})

	Network	Count	Minimum	Median	Mean	Standard deviation	Percentile 90	Maximum
As	this study	348	0.04	6.70	8.25	11.3	14.3	159
	FOREGS	837	<5.0	6.00	9.88	15.8	20.0	220
Cd*	this study	348	0.02	0.16	0.23	0.31	0.39	4.32
	FOREGS	840	<0.01	0.145	0.284	0.71	0.48	14.1
Cr	this study	348	2.30	45.5	53.1	34.7	91.4	213
	FOREGS	837	1.00	22.0	32.6	89.3	53.0	2340
Cu	this study	348	3.60	23.7	31.3	30.7	58.0	351
	FOREGS	837	1.00	12.0	16.4	18.0	33.0	239
Hg*	this study	348	0.01	0.12	0.15	0.13	0.28	0.97
	FOREGS	833	0.005	0.037	0.061	0.10	0.115	1.35
Ni	this study	348	1.20	35.5	35.9	20.1	55.6	208
	FOREGS	837	<0.2	14.0	30.7	124	41.0	2560
Pb	this study	348	3.07	16.8	20.6	17.8	32.9	200
	FOREGS	837	<3.0	15.0	23.9	50.2	38.2	886
Zn	this study	348	1.26	64.4	63.5	22.3	87.1	162
	FOREGS	837	4.00	48.0	60.9	115	96.0	2270

*The soil contents of Cd and Hg in FOREGS data set are determined after complete dissolution.

Table 2. Baseline values for critical heavy metals in Bulgarian Monitoring Soil Quality Network (mg kg^{-1})

	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
Baseline values	14.5	0.36	91.1	42.3	0.31	61.5	29.8	98.6

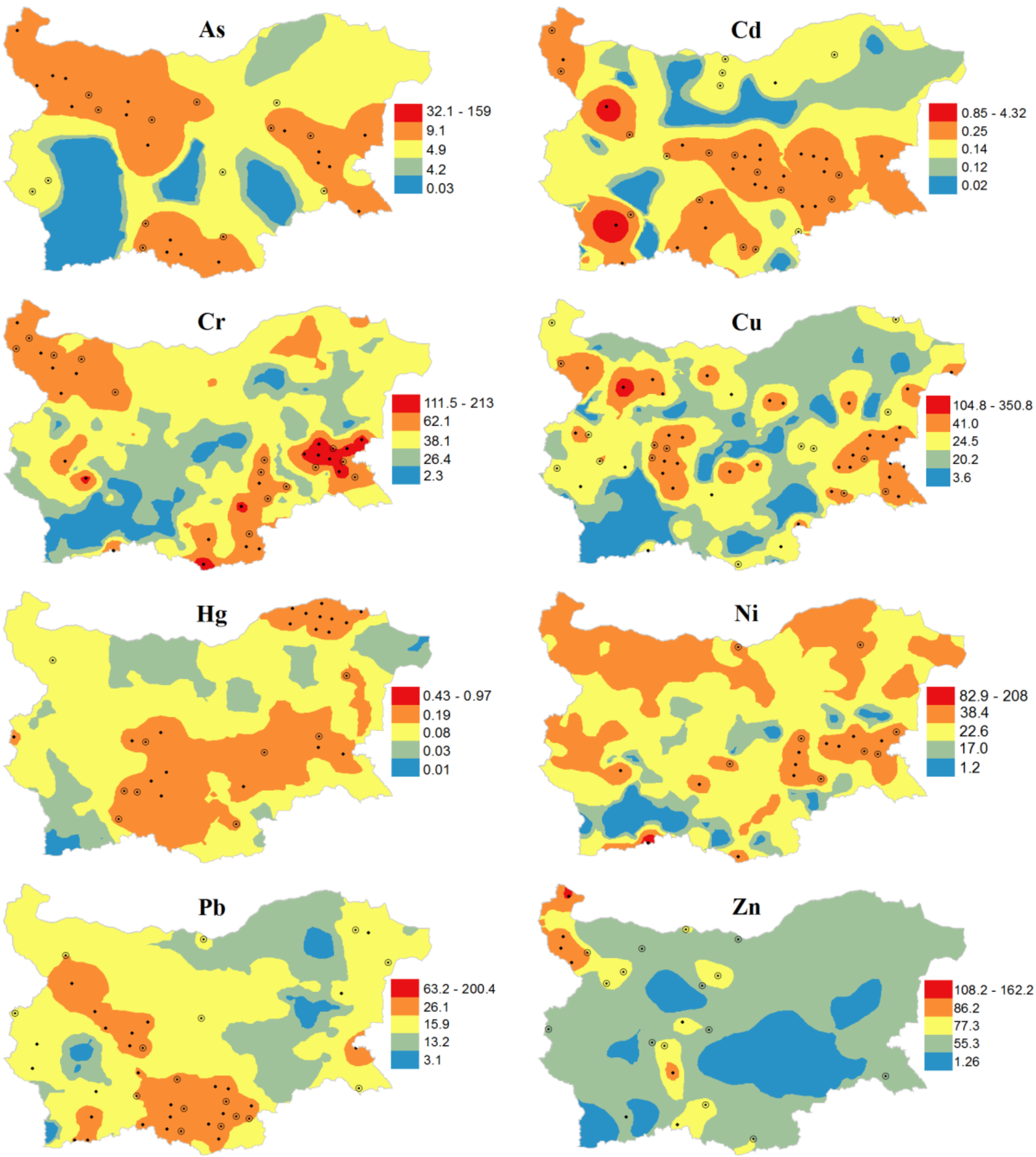

Figure 1. Spatial distribution of heavy metals and contaminated topsoil sites determined by the box-plot method (●- outliers, ⊙ - higher than baseline value).

Table 3. Factor loadings (loadings higher than 0.7 are marked)

	PC1	PC2	PC3	PC4	PC5
Cu	0.122	**0.878**	-0.075	0.144	0.022
Zn	0.106	**0.707**	0.288	-0.164	0.226
Cd	0.015	-0.046	**0.838**	0.050	-0.073
Pb	-0.080	0.169	**0.745**	0.010	0.159
Ni	**0.922**	0.110	-0.024	-0.055	-0.042
Cr	**0.920**	0.010	-0.037	0.089	0.050
As	0.003	0.149	0.058	0.009	**0.972**
Hg	0.028	0.021	0.057	**0.982**	0.007
% expl. var.	21.62	16.84	16.94	12.82	12.89

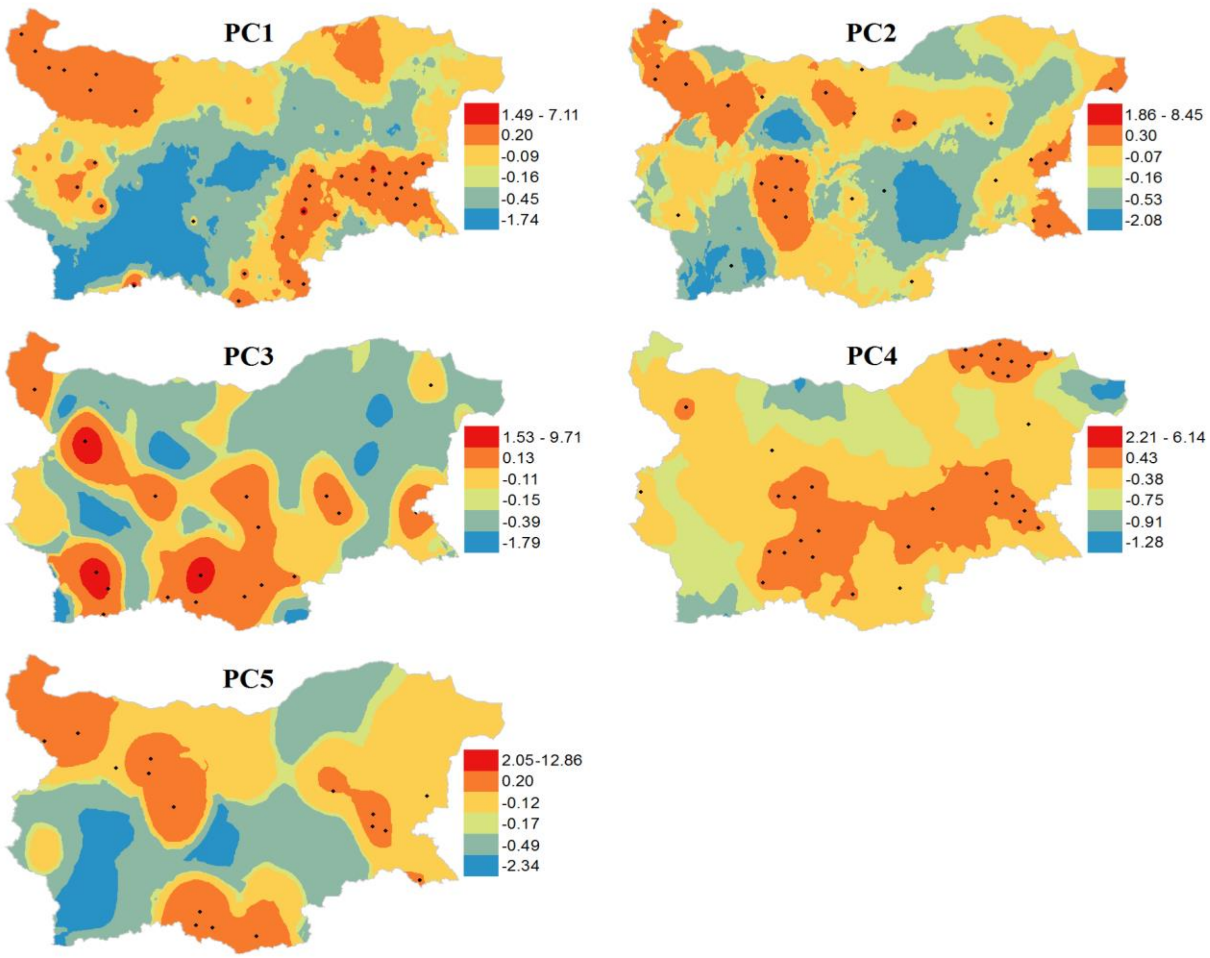

Figure 2. Spatial distribution of principal components and contaminated topsoil sites determined by the PCA

The third latent factor with 16.94% explained of the total variance has the similar origin and could be conditionally named “Cd-Pb pollution”. The last two latent factors reveal Hg (PC4 – 12.82%) and As (PC5 – 12.89%) pollution respectively. The extracted five latent factors represent natural (PC1) and anthoropogenic sources impact (PC-2-PC5) on heavy metal concentrations. A mapping of factor scores by ordinary kriging was performed to describe spatial characteristics of extracted five latent factors on national scale (Fig. 2). Sites with factor score values higher than 1 were arbitrary defined as contaminated ones. The comparison between proportions of contaminated sites defined by both statistical approaches clearly distinguishes heavy metals in three groups (Table 4). The first group of Cr, Hg and Zn has similar number of identified contaminated sites by both methods. For the second group

(As, Cd, Cu and Pb) univariate approach (box-plot method) defines almost twice more contaminated sites then PCA. The difference could be explained by the way the data set variation is presented. Multivariate approach (PCA) reveals natural and anthropogenic sources by complex latent factors. For instance, contribution to the Pb concentrations comes not only from "Cd-Pb pollution" (PC3) source but also from "Cu-Zn pollution" (PC2) and "As pollution" sources. The same holds threw for Cu concentrations which is frequently associated with geogenic origin of site (Lado *et al*., 2008). The ability to highlight the major sources influencing heavy metal concentrations makes PCA very good tool for identification of sites affected by the same contamination source. On the other hand, the box-plot method recognizes also sites which are subject of local pollution and/or local natural anomalies. The univariate approach is reliable tool for easily mobile elements like As and Cd (Redon *et al*., 2013; Lado *et al*., 2008).

Table 4. Proportion (%) of contaminated sites identified following the two decision tools used in this study

	As	**Cd**	**Cr**	**Cu**	**Hg**	**Ni**	**Pb**	**Zn**
Box plot method	9.2	12.1	10.1	15.8	8.0	5.5	12.1	6.6
PCA	4.6	5.2	9.8	8.9	9.8	9.8	5.2	8.9

4. Conclusions

Comparison between baseline values of Bulgarian monitoring network and FOREGS outlines the specificity of the national geogenic origin. The higher baseline values, especially for Cr, Cu and Hg, require detailed background information concerning identification of contaminated sites. There are five factors affecting concentrations of critical heavy metals: one with natural with geogenic origin (Ni, Cr) and four accounting for specific pollutions of (Cu, Zn), (Cd, Pb), As and Hg. The presented spatial distributions of heavy metal concentrations and latent factors along with identified contaminated sites by both decision tools provide complete picture of contaminated regions based on descriptive statistics and sources controlling heavy metal concentrations.

Acknowledgment

Support from the project No. DFNI E 02/7 12/12/2014 of the National Science Fund of Bulgaria is acknowledged.

References

Facchinelli A., Sacchi E. and Mallen L. (2001), Multivariate statistical and GIS-based approach to identify heavy metals sources in soils, *Environmental Pollution*, **114**, 313-324.

Hu Y., Liu X., Bai J., Shih K., Zeng E.Y. and Cheng H. (2013), Assessing heavy metal pollution in the surface soils of a region that had undergone three decades of intense industrialization and urbanization, *Environmental science and pollution research international*, **20**, 6150-6159.

Lado L.R., Hengl T. and Reuter H.I. (2008), Heavy metals in European soils: A geostatistical analysis of the FOREGS Geochemical database, *Geoderma*, **148**, 189-199.

Luo W., Wang T., Lu Y., Giesy J.P., Shi Y., Zheng Y., Xing Y. and Wu G. (2007), Landscape ecology of the Guanting reservoir, Beijing, China: multivariate and geostatistical analyses of metals in soils, *Environmental Pollution*, **146**, 567-576.

Micó C., Peris M., Recatalá L. and Sánchez J. (2007), Baseline values for heavy metals in agricultural soils in an European Mediterranean region, *Science of the Total Environment*, **378**, 13-17.

Micó C., Recatalá L., Peris M. and Sánchez J. (2006), Assessing heavy metal sources in agricultural soils of an European Mediterranean area by multivariate analysis, *Chemosphere*, **65**, 863-872.

Redon PO., Bur T., Guiresse M., Probst JL., Toiser A., Revel JC., Jolivet C. and Probst A. (2013), Modelling trace metal background to evaluate anthropogenic contamination in arable soils of south-western France, *Geoderma*, **206**, 112-122.

Reimann C., Filzmoser P. and Garrett R.G. (2005), Background and threshold: critical comparison of methods of determination, *Science of the Total Environment*, **346**, 1-16.

Saa G.R., Victoria J.A.R. and Molina R.M. (2011), Methods for establishing baseline values for heavy metals in agricultural soils: Prospects for Colombia, *Acta Agronómica*, **60**(3).

Salminen T. (2006), Geochemical atlas of Europe part 1: background information, methodology and maps. In: Vol. 1 of Geochemical Atlas of Europe, EuroGeosurveys & Foregs, Espoo, Finland.

Zhang C. (2006), Using multivariate analyses and GIS to identify pollutants and their spatial patterns in urban soils in Galway, Ireland, *Environmental Pollution*, **142**, 501-511.

Zheng YM., Chen TB. and He JZ. (2008), Multivariate geostatistical analysis of heavy metal in topsoils from Beijing, Chinia, *Journal of Soils and Sediments*, **8**, 51-58.