Journal of Geochemical Exploration 179 (2017) 25–35

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/gexplo

Layer-specific spatial prediction of As concentration in copper smelter vicinity considering the terrain exposure

Milutin Pejović[a,*], Branislav Bajat[a], Zagorka Gospavić[a], Elmira Saljnikov[b], Milan Kilibarda[a], Dragan Čakmak[c]

[a] Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, Belgrade, Serbia
[b] Institute of Soil Science, Teodora Drajzera 7, 11000 Belgrade, Serbia
[c] Department of Ecology, Institute for Biological Research "Siniša Stanković", University of Belgrade, 1060 Belgrade, Serbia

ARTICLE INFO

Keywords:
Pollution mapping
Arsenic
Terrain exposure
Exposure to wind
Spline-Than-Krige

ABSTRACT

Prevailing climatic conditions and local topography can be classified as the most influential environmental factors that affect the spatial dispersion of pollutants emanating from industrial sources. In this study, the combined effects of these factors were considered with respect to terrain exposure in order to explain the complex spatial trend of Arsenic (As) concentration that was atmospherically-deposited from one of the largest Copper Mining and Smelting Complexes in Europe, Bor in Serbia. Several exposure parameters were created and employed as spatial covariates within the so-called "Spline-Then-Krige" approach for producing maps of As concentration at three standard soil depth layers (0–5 cm, 5–15 cm and 15–30 cm). The exposure parameters were created to quantify two different aspects of terrain exposure: Geometrical (Proximity) and Topographical exposure. Regression analysis confirmed the presence of a significant statistical association between the As data and all exposure parameters. The trend model showed good overall accuracy explaining 52% of the variance in As data for the surface soil layer, 49% for the middle layer and 35% for the deepest layer. Relative importance analysis revealed the importance of considering a more general model that includes interactions between exposure parameters. The kriging interpolation improved, to some extent, the regression accuracy for all three layers with R^2 values ranging from 55% for the surface layer to the 36% for the deepest soil layer. The prediction maps show that As contamination levels are well above allowable Serbian agricultural concentration limits (As < 25 mg/kg) for approximately 78% of the mapping area, thereby indicating that long term smelting activity leaves significant consequences on soil even on deeper unexposed layers.

1. Introduction

Without a doubt, industrial mining has significant consequences on the environment and human health (Unit, 2013). Spatial extension and the magnitude of soil pollution in mining areas are conditioned by many environmental factors such as climatic conditions, relief, human or mining activity, the soil type, and land use. In geostatistics, environmental factors are approximated by spatial covariates. These are mainly maps in raster format, which could also be obtained as the outputs of some other environmental models. For example, Goovaerts et al. (2008) used an EPA Industrial Source Complex (ISC3) dispersion model (EPA, 1995) in combination with kriging and geostatistical simulation to delineate areas with high levels of dioxin TEQ_{DF} WHO_{98} in soil around an incinerator. Their dispersion model explained 47.3% of the variance found in the soil TEQ data, leaving the residuals suitable for geostatistical analysis. Dispersion models like ISC3 can take a wide range of parameters into account that pertain to meteorological conditions, the local topography, and the characteristics of the source (e.g., emission rate, stack height and diameter, particle diameter etc.) (De Visscher, 2013). These parameters are often inaccessible for long-term pollution processes; therefore soil scientists have to deal with only a few known parameters that are often related to relative distances from the source of pollution, terrain topography or common meteorological parameters including prevailing wind direction and wind speed. Žibret and Šajn (2008) presented successful implementation of the power function with negative exponent to model how the level of heavy metal concentrations in the air and soil decreases in relation to incremental increases of the distance from the source of pollution. (Saito and Goovaerts, 2001) incorporated the knowledge of the position of a pollution source and deviations from major wind direction into a kriging system to map the spread of pollutants from a known source.

In mountainous or hilly areas, the spatial variation of wind-

* Corresponding author at: Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia.
E-mail addresses: mpejovic@grf.bg.ac.rs (M. Pejović), bajat@grf.bg.ac.rs (B. Bajat), zaga@grf.bg.ac.rs (Z. Gospavić), kili@grf.bg.ac.rs (M. Kilibarda).

http://dx.doi.org/10.1016/j.gexplo.2017.05.004
Received 4 February 2016; Received in revised form 12 April 2017; Accepted 10 May 2017
Available online 19 May 2017
0375-6742/

deposited materials is highly affected by terrain topography. It is generally known that the amounts of wind-deposited materials tend to be greater on areas that are more directly exposed to wind flux. This fact has inspired researchers to develop many topographic indices with the aim to quantify topographic exposure to wind (Antonić and Legović, 1999; Lindsay and Rothwell, 2008; Winstral et al., 2002). Generally, all topographic exposure indices are based on Digital Elevation Model (DEM) analysis and tend to determine whether a particular area is sheltered by a distant topographic obstacle or not. There are several studies where topographic exposure indices were successfully used to model the spatial patterns of snow depths (Erickson et al., 2005; Plattner et al., 2004; Winstral et al., 2002).

Antonić and Legović (1999) introduced the aspect of topographic exposure to wind in their exploration of environmental pollution studies. They proposed the new comprehensive index, referred to as the Exposure toward the Wind Flux (EWF). EWF can be conceptualized as the angle between a plane orthogonal to the wind and a plane that represents the local topography at a grid cell. They utilized EWF to estimate the direction of an unknown air pollution source.

In this study, we considered different aspects of terrain exposure in order to explain the complex spatial trend of Arsenic (As) concentration that was atmospherically-deposited from one of the largest Copper Mining and Smelting Complexes in Europe, Bor in Serbia. Several exposure parameters were created and employed as covariates within the so-called “Spline-Then-Krige” (STK) approach (Malone et al., 2009; Orton et al., 2016) for producing maps of As concentration at three standard soil depth layers (0-5 cm, 5-15 cm and 15-30 cm). The created exposure parameters were grouped as follows: Geometrical (Proximity) exposure parameters and Topographical exposure parameters. The distances to the source of pollution and angular deviations from prevailing wind direction were utilized to create Geometrical (Proximity) exposure parameters. Furthermore, topographical exposure was quantified by using DEM and two DEM derivates: modified EWF index and the Morphometric Protection Index (MPI). A modification of EWF was performed to account for the location of the pollution source with the aim to emphasizing the effects of topographical exposure to the known source and not just limiting the index to wind direction. This study primarily aims to evaluate the effectiveness of using different exposure parameters for mapping atmospherically-deposited Arsenic on different soil depth layers. Relative importance analysis was performed to access the individual contribution of each exposure parameter in the trend model for each depth layer. By analyzing the role of exposure parameters in As variation on different soil depth layers, we also tried to realize the limit of significant influence of copper smelting in soil depth direction. To the best of our knowledge, this is the first study of its kind that evaluates the usage of different terrain exposure indices for mapping atmospherically-deposited pollutants from a known source.

2. Materials and methods

2.1. Case study area

The Copper Mining and Smelter Complex Bor (Lat = 44° 4′ N, Lon = 22° 6′ E) is the largest copper mine in Serbia and is also one of the largest in Europe. The entire complex is located adjacent to the town of Bor in the north-eastern section. The municipality of Bor is located in the central-eastern part of Serbia and covers an area of 856 km^2. The town contains a total of 35,000 inhabitants and an additional 20.000 people are settled in surrounding settlements belonging to the Municipality of Bor.

The climate of Bor is moderately continental. Air circulation is controlled by prevailing northwest and eastern winds. Winds from the northwest prevail during warmer months, whereas eastern and south-eastern winds prevail during colder periods of the year (Fig. 1a). Topographically, the area is mainly hilly and is covered with deciduous forests and agricultural lands.

The copper smelter, which is a part of the Mining-Metallurgical Complex Bor is recognized as the major pollution source in the region with emissions of over 200,000 t of SO_2 and 300 t of As per year (LEAP, 2003). Over the past 15 years, the annual SO_2 and As concentration detected in the air from urban and sub-urban areas surrounding the copper smelter in Bor greatly exceeded current threshold values thereby classifying the town Bor as one of the most polluted regions in Serbia and possibly Europe (Serbula et al., 2013). Meteorological activity in the area, such as winds, strongly influence the impact of the pollution produced by copper production in Bor. Correlation analysis between As concentration and basic meteorological parameters previously indicated that the most polluted areas in the town of Bor are the most influenced by east and northwest winds as well as changes in relative humidity and air temperature (Šerbula et al., 2010).

The case study area covers a region that is approximately 8 km West of Bor. The north–south transect of the survey area is about 20 km, while the east-west transect is about 10 km. In the period of prevailing eastern winds (colder parts of the year), the study area is located directly downwind from the copper smelter and should be highly susceptible to increased levels of pollution.

2.2. Materials

Data consists of 196 soil profiles that are randomly distributed over the entire study area (Fig. 2). Soil samples were collected based on generic horizons. Soil profiles were sampled from four soil layers including: O (organic soil horizon), A horizon, B (if it existed) and C. The thickness of the soil profiles varied considerably and consequently the number of samples per profile differed. The depth to the top of the C-horizon varied between 10 and 123 cm. The morphological characteristics of the explored area are very heterogeneous, which caused the formation of several soil types with different depths including: 1) Distryc Regosol (covering 11% of the study area) with a maximum depth of 20 cm that is typical for terrain with large slopes and higher altitudes; 2) Leptosol Eutric and Dystric soils (covering 47% of the study area) with a depth up to 40 cm that is primarily in the A horizon; and 3) Cambisol soils (covering 20% of the study area) that is formed from the Leptosol and is presented at the terrain with lower altitudes and valleys with the horizons: Amo - (B) v- (B) vC-C, or Amo - (B) v- (B) vR and deeper profiles depths.

The soil samples were digested with concentrated HNO_3, following the US EPA SW-846 Method 3050 (U.S. EPA, 1998) and then analyzed for As concentration using an iCAP 6300 ICP optical emission spectrometer. Geographic coordinates of profiles were determined using Global Navigation Satellite Systems (GNSS) within ± 3 m positional accuracy.

The box plots in Fig. 1b illustrate the vertical distribution of observed data classified according to standard soil depths intervals. Depth-wise data summary depicted on Fig. 1c shows similar, but more continuous form of vertical distribution of As data. It was produced by plotting functionality from “aqp” R package (Beaudette et al., 2013) exclusively created for soil profile data analysis. This functionality enables to specify the function as well as the depth interval along which the function will be calculated and plotted. In our case, this figure shows the change in median with depth bounded by 0.25th and 0.75th quartiles computed at 5 cm depth interval. As it is apparent from these figures, As data are characterized by pronounced decreasing trend in median with depth as well as with considerable higher variation in the upper soil layers. The abrupt change in the trend of median and interquartile range occurs at about the 30 cm depth. The numbers of profiles that contribute to the estimated median values are shown in percentages on the right vertical axes. It can be seen that < 50% of available data contribute to the estimate value for layers below 30 cm depths. Due to the fact that the number of observations sharply decreases below the depth of 30 cm and that the exposure parameters showed weak ability to explain the As variation at the deeper soil layers, we chose to

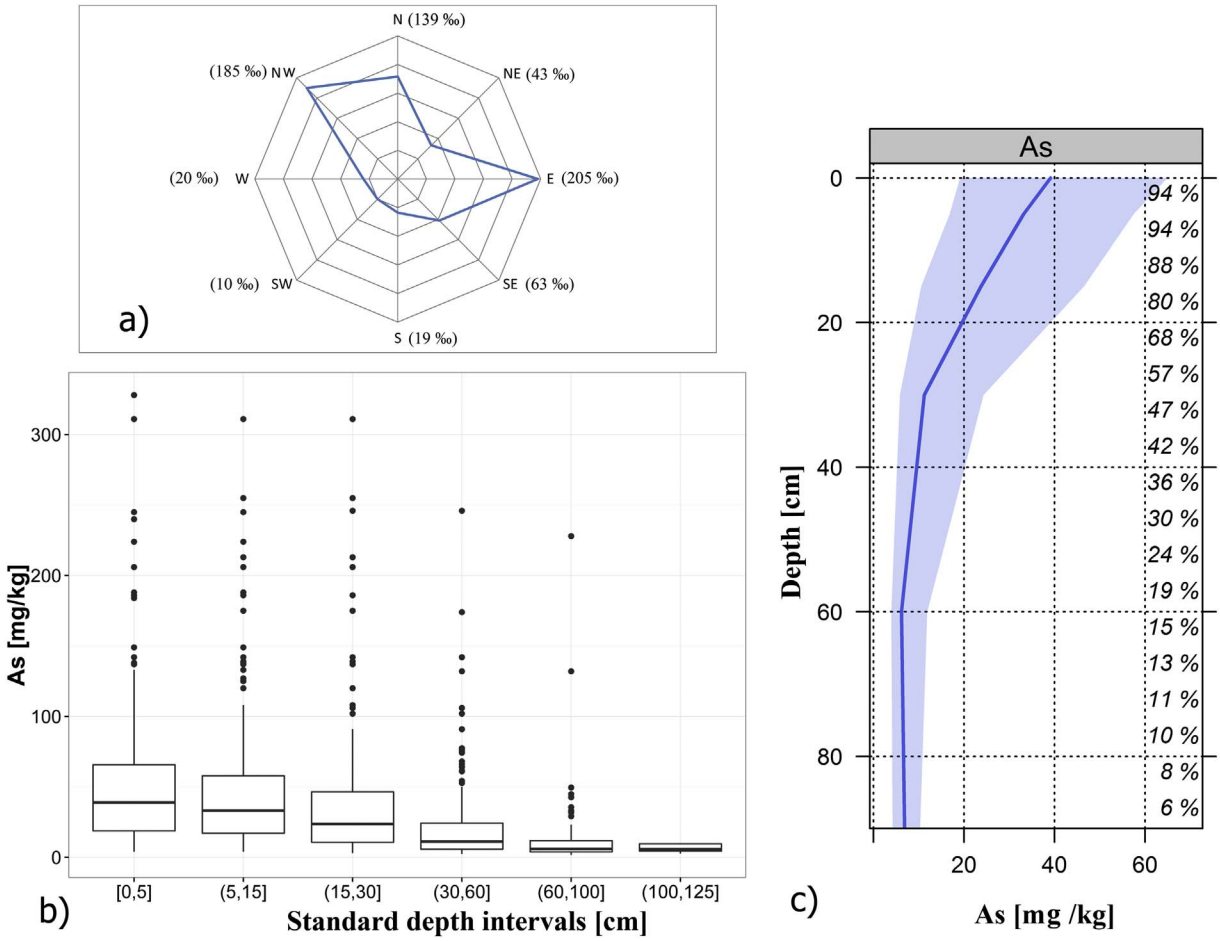

Fig. 1. a) Wind-rose diagram and wind frequency (‰) for the period 2003–2007; b) Box-plots of As data per standard soil depth intervals, c) Depth-wise aggregation of soil profiles data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

confine our analysis to the first three standard soil layers above this depth: 0–5 cm, 5–15 cm and 15–30 cm.

The presence of extreme observed values is an important characteristic of this data set. Fig. 2 depicts the spatial pattern of observations from the first soil layer, allocated within the 4th quartile (red circles: 80–280 mg/kg) with respect to the smelter location. The circle size depicted in the figure is proportional to the observed value. Terrain colors are utilized to represent the possible spatial coverage of plume dispersion.

High concentration and high variability in the As data at the upper soil layers combined with distinct differences between the upper and lower soil layers are generally considered to be indicators of external factors that have a pronounced influence on the soil. As a result, a hypothesis can be formed that the upper soil layers were indeed affected by long term pollution processes.

2.3. Terrain exposure

In this study, terrain exposure parameters aim to provide the numerical quantification of terrain exposure with regard to the location of the source of pollution, wind direction and topography. As mentioned above, the considered terrain exposure parameters have been divided into two groups: Topographical exposure and Geometrical (proximity) exposure. Table 1 summarizes the exposure parameters used in this study.

2.3.1. Topographical exposure

There are many existing indices that are suitable for explaining topographic exposures to wind. An exhaustive review of existing topographic wind related indices was outlined in studies reported by Lindsay and Rothwell (2008) and Winstral et al. (2002). In this study, we confined topographic exposure analysis to the following three parameters: Digital Elevation Model (DEM), Exposure toward the Source of pollution (ES) and Morphometric Protection Index (MPI).

2.3.1.1. DEM. Considering the assumptions that areas on higher altitudes are more exposed than lowlands areas, elevation was selected as the first topographic exposure parameter. A high resolution DEM with a grid size of 20 m was created by digitizing contours from 1:25.000 scale topographic map sheets. All other exposure parameters were computed in the same grid system.

2.3.1.2. Exposure toward the source of pollution (ES). The effects of topography along wind direction were considered through the modified EWF measure. By definition, the EWF index combines two simple exposure parameters to quantify topographic exposure to wind flux. These two parameters include relative terrain aspect and horizon angle:

$$EWF = \cos(\mu)\bullet sin(\beta) + sin(\mu)\bullet cos(\beta)\bullet cos(\delta - \gamma) \quad (1)$$

where μ represents the terrain slope, γ is the terrain aspect, δ is the azimuth of the dominant wind direction and β is the horizon angle in the wind direction (Fig. 3a).

The relative terrain aspect represents the orientation of the local terrain plane in relation to the selected wind direction. This is the angle between the land-surface aspect and the wind direction bounded between 0°, indicating an exposed location, and 180°, indicating sheltered location. The horizon angle quantifies the effects of upwind topography searching for the maximum elevation angle along the

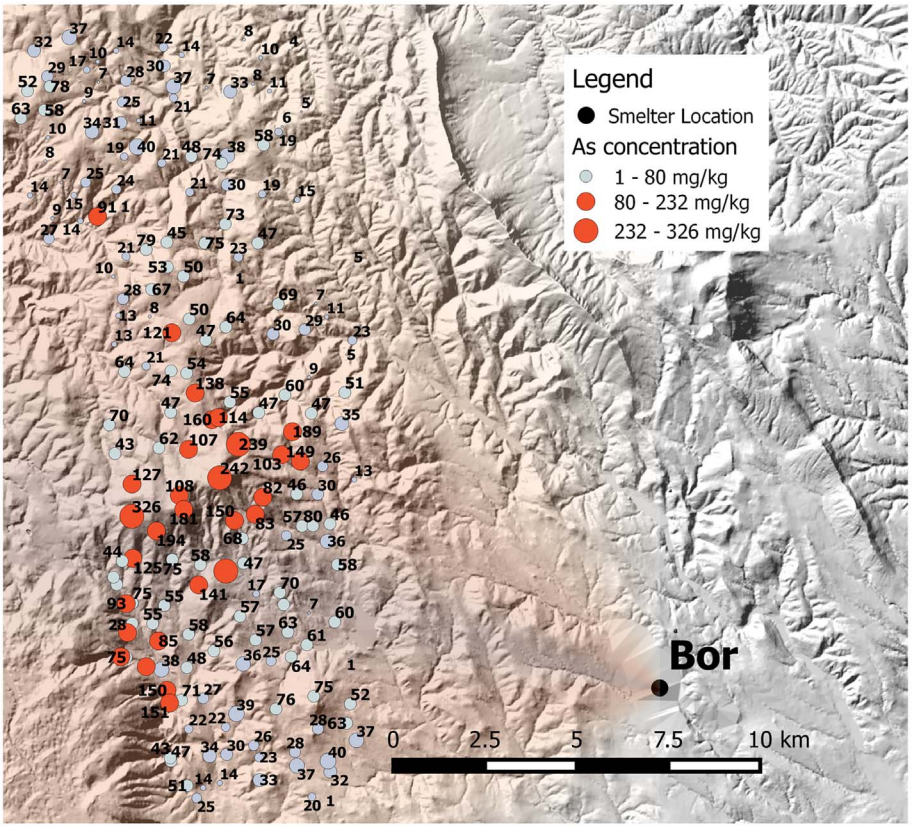

Fig. 2. Spatial disposition of extreme values observations relative to the location of smelter. Bor is in the lower-right corner; red circles represent the observations that belong to the fourth quartile. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Exposure parameters used in this study.

Name	Abbreviation	Group	Range
Digital Elevation Model	DEM	Topographical	300–1045
Exposure toward the Source	ES	Topographical	0.75–1.34
Morphometric Protection Index	MPI	Topographical	0–0.70
Down-wind Dilution	DD	Geometrical	0.20–0.65
Cross-wind Dilution	CD	Geometrical	0.39–1

direction of prevailing wind flux. The search distance has a crucial effect on the horizon angle estimation. According to the definition of horizon angle, the more exposed area is characterized by a negative horizon angle whereas a sheltered area is characterized with a positive horizon angle. Horizon angle has been used as the basis for many subsequently devised indices (Erickson et al., 2005; Winstral et al., 2002). Depending on the extent of the horizon angle search distance, EWF has been referred to the horizontal wind flux (zero search distance) or to the slope wind flux (search distance differs from zero).

The standard EWF index presumes a constant direction of wind flux, which participates in two terms of its formulation: namely the relative aspect and the horizon angle. Taking into account that the contaminated air flux starts from the one copper smelter stack and expands toward the explored region we assumed that: (1) the local terrain plane facing the source is more exposed to pollution than planes that are not; (2) the topographic obstacles founded within the direction of the source have a greater effect on redistribution of pollutants than the obstacles founded strictly in the upwind direction. Based on these assumptions, the EWF index was calculated for each grid cell with the adjustable wind direction. More specifically, the wind direction was defined as the

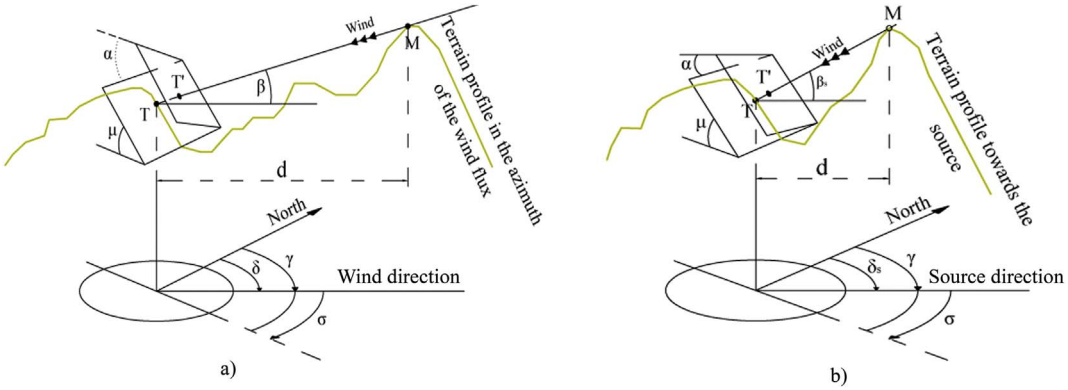

Fig. 3. a) Graphical representation of EWF components. Terrain at the point T has the maximum slope μ and terrain aspectγ. σ represents relative terrain aspect at point T for a given azimuth of the wind flux δ.β is the horizon angle of the point T for a search distance d. α is the angle between regression plane through the terrain point T and plane orthogonal to the wind. b) Equivalent graphical representation of ES components. Index "s" denotes the "source of pollution".

azimuth between each grid cell and the source of pollution. In this regard, the relative aspect becomes the angular distance between the land-surface aspect and direction to the source. At the same time, the horizon angle search path is also directed toward the source (Fig. 3b). This new parameter was denoted as Exposure toward the Source (ES).

2.3.1.3. Morphometric protection index (MPI). The influence of local (neighboring) topography was considered by the Morphometric Protection Index calculated for each grid cell. The calculation of MPI is equivalent to the *positive openness* described by (Yokoyama et al., 2002). It considers neighboring grid cells of DEM in eight directions (cardinal and diagonal) up to a given distance (we used radius of 200 m) while searching for the maximum horizon angle in each direction. The final MPI for one cell represents the averaged value of eight maximum horizon angles and quantifies how the neighboring relief protects that cell.

2.3.2. Geometric (proximity) exposure

The creation of Geometric (Proximity) exposure parameters was inspired by a dilution mechanism considered in the Gaussian dispersion model (Gaussian plume model). It assumes that dilution of plume emitted in the atmosphere could be considered in three directions (downwind, crosswind and vertical) (De Visscher, 2013). The downwind plume dilution is the result of mixing a plume with ambient air, while the dilution in the cross-wind direction is a result of a large number of negligible effects related to atmospheric motions. Taking into account all of the assumptions mentioned before, we presumed that areas are more geometrically (proximately) exposed if they are closer to the copper smelter and/or to prevailing wind direction. Therefore, the effects of dilution in downwind and crosswind directions in this study were approximated by Downwind Dilution (DD) and Crosswind Dilution (CD) parameters computed for each grid cell. These are modeled using a negative-exponential function, where the exponents are the distance to the smelter for DD and the directional departure from dominant wind direction for CD. Wind rose (Fig. 1a) shows that the prevailing winds blowing from east and northwest directions. However, in order to found the CD which is the most correlated with observed, wind direction was determined based on correlation analysis between first soil layer data and the Crosswind Dilution computed for several major wind directions in range 90° ± 30° along with increments of 5°. Finally, the wind direction of 105° was found to be most associated with data. Fig. 4 depicts the graphical representation of CD and DD parameters.

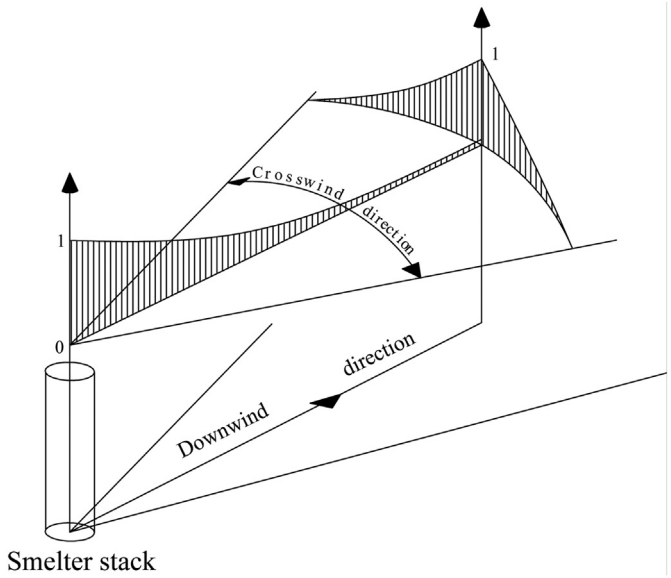

Fig. 4. Graphical representation of geometrical (proximity) measures.

2.4. Geostatistical mapping

Spline-Then-Krige (STK) refers to the 3D geostatiscal approach for producing a suite of digital maps for soil properties at different soil depths, which was first proposed by (Malone et al., 2009). In terms of methodology, it implies two separated modeling steps: 1) modeling the variation of soil properties with depth by fitting depth functions to the profile data; and 2) modeling of the spatial distribution of data predicted at a particular depth by a depth function. This approach was successfully used in many studies for the mapping of various soil properties (Adhikari et al., 2013; Lacoste et al., 2014; Mulder et al., 2016; Orton et al., 2016).

2.4.1. Vertical variation modeling

Arsenic from anthropogenic sources are more mobile than those from soil parent materials (Chlopecka et al., 1996). Depending on the soil conditions in the contaminated soil, As content decreased in the soil depths from 30 to 50 cm (Luo et al., 2008, Cappuyns et al., 2002) but to some extents it exhibits the same pattern noticed on the soil surface with an abrupt decrease when examining downwards in the profile. Variation in the soil profile was modeled by equal-area spline function proposed by Bishop et al. (1999). Equal-area spline function implies continuous vertical variation, which can be expected in our case, considering the fact that during the more than one hundred years of copper production in Bor, vertical leaching of deposited toxic materials in soil has certainly occurred. This function provides that, for each sampling layer (soil horizon), the average of the spline function equals the measured value for the horizon, i.e. the area above and below the fitted spline in any horizon are equal. (Bishop et al., 1999; Malone et al., 2009). Fig. 5 depicts an example of a fitted spline to the measured As data from profile No. 119. The colored horizontal bars represents the measured As concentration at different horizons (each bar corresponds to one horizon) while the vertical curve represents the equal-area spline depth function fitted to these data. In order to obtain the As concentration related to the selected fixed depth intervals (0–5 cm, 5–15 cm and 15–30 cm) spline function was averaged within these intervals. These intervals correspond to the standard soil depth intervals specified in *GlobalSoilMap* specifications (Arrouays et al., 2014).

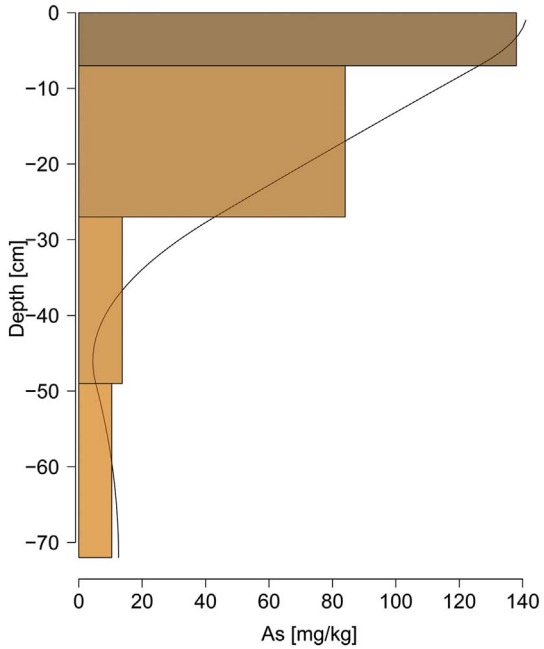

Fig. 5. Equal-Area spline depth function fitted to the data from profile No. 119.

Soil profiles containing only one sample layer were not modeled. Instead, they were considered as profiles with constant As concentration up to the depth of the sampling horizon. The equal-area spline function was fitted via the *mp.spline* function implemented in the GSIF R package (Hengl, 2015).

2.4.2. Trend modeling (analysis) and spatial prediction

The common model used in model based geostatistics describes the soil variation as sum of three major components: (1) the deterministic-trend component (2) a spatially correlated component (stochastic residuals) and (3) pure noise. The trend component represents the part of variation caused by influential environmental factors, typically expressed as linear combination of covariates known over the domain of interpolation. The stochastic component covers spatially correlated small-scale variations assumed as a realization of a spatially correlated random process (Webster and Oliver, 2007).

The first task in trend analysis was to identify the type of relationship between the As data and exposure parameters. It is convenient to represent the relationship between the target variable and the covariates using a linear model (Pebesma, 2006; Hengl et al., 2007a; Kilibarda et al., 2014). The adequacy of this specification was checked by examining the residual plots. Prior to model fitting, the exposure parameter values at the each profile location were extracted and joined to the spline-predicted As values for each depth increment. The interaction effects between each pair of exposure parameters were also considered to be included in the model. By doing this, it was enabled that the effects of one exposure parameter depends on the value of other exposure parameter.

Model selection was conducted by performing stepwise linear regression analysis using the *Akaike information criterion* (AIC) (Akaike, 1974) as a selection criterion. The AIC is one of the most common metric used for model selection. The AIC is defined as:

$$AIC = -2\log(\mathcal{L}(\theta \mid y) + 2K, \tag{2}$$

where $(\mathcal{L}(\theta|y)$ is the maximized log-*likelihood* and K is the number of parameters in the model. The first term of the Eq. (2) decreases as more parameters are added to the model, while the second term increases. In this way, AIC controls for overfitting by penalizing models that include too many predictors. The complete process of model selection was conducted on data from the first soil layer, considering the fact that the effects of atmospheric pollution are the most pronounced near the terrain surface.

As an integral part of trend analysis, we tried to realize the proportions to which geometric and topographic exposure parameters contributed to the prediction accuracy in each soil layer. This was assessed by computing measures for the relative importance of predictors. It is important to note that the term "predictor" is associated to the independent model variable, which could refer to the main effect or interaction effect as well. There are several measures for the relative importance of predictors in linear modeling theory that are all available in the "*relaimpo*" R package (Grömping et al., 2006). These measures provide information about the individual contribution of each predictor to the portion of explained variance (R^2). We used the most comprehensive and recommended measure called LMG, which was firstly proposed by (Lindeman et al., 1980).

2.4.3. Spatial prediction

Regression Kriging (RK) was adopted as a general statistical framework for spatial prediction. RK combines two conceptually different techniques, regression for trend estimation and simple kriging to interpolate stochastic residuals (Hengl et al., 2007b). If the measured values of a target variable are symbolized as$(s_i), i = 1 \ldots n$, where s_i represents spatial location and n is the number of realized measurements. The regression kriging prediction at unvisited location s_0, $\hat{z}(s_0)$ can be expressed as the sum of the fitted deterministic part (trend) $\hat{m}(s_0)$ and the interpolated residual $\hat{e}(s_0)$:

$$\hat{z}(s_0) = \hat{m}(s_0) + \hat{e}(s_0) \tag{3}$$

In practice, the trend coefficients are mainly obtained by ordinary least squares (OLS). However, this can cause bias in the estimates of the residual variogram (Cressie, 1993). One solution to reduce the bias, used in this study, is to estimate the residuals taking into account the spatial correlation between the observations (Hengl et al., 2004). For that reason, the usage of Generalized Least Squares (GLS) is recommended instead of the commonly used OLS. However, GLS implies an iterative procedure for the variogram estimation. In the first step, the trend model is estimated using OLS. The given OLS residuals are then used to construct the covariance function needed to obtain the GLS estimates. In the next step, the GLS residuals were used to update the covariance function in order to re-calculate the GLS residuals, from which an updated covariance function was computed. This procedure should be repeated until the trend coefficients no longer change. The final residual variogram was then estimated from the final GLS residuals and then modeled as a continuous function of lag distance.

For a given trend model and residual variogram, the prediction of a target variable at an un-sampled location s_0 is obtained by:

$$\hat{z}(s_0) = \sum_{k=0}^{p} \hat{\beta}_{GLS} \bullet x_k(s_0) + \sum_{i=1}^{n} \lambda_i \bullet \left[z(s_i) - \sum_{k=0}^{p} \hat{\beta}_{GLS} \bullet x_k(s_i) \right] \tag{4}$$

where $z(s_i)$ represents the observed values at the neighboring locations-s_i, $\hat{\beta}_{GLS}$ represents the estimated trend model coefficients, $x_k(s_0)$ are the known value of covariates at the predicted location, $x_k(s_i)$ are the known value of covariates at the location s_i and λ_i are the kriging weights.

2.4.4. Prediction accuracy assessment

Prediction accuracy was evaluated based on the leave-one-out cross-validation procedure. The following common statistical indices were calculated to evaluate the prediction accuracy:Mean Error (ME),

$$ME = \frac{1}{n} \sum_{i=1}^{n} [z(s_i) - \hat{z}(s_i)] \tag{5}$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [z(s_i) - \hat{z}(s_i)]^2} \tag{6}$$

and R squared (R^2):

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n} [z(s_i) - \hat{z}(s_i)]^2}{\frac{1}{n}\sum_{i=1}^{n} [z(s_i) - \overline{z}]^2} \tag{7}$$

3. Results and discussion

Table 2 shows the common descriptive statistics measures computed for aggregated profile data divided into a total of six standard depth increments. It is obvious that the measures of central tendency (mean, median) systematically decrease by depth. The mean values in the upper layers are almost double the mean value from layers below the 30 cm depth. This trend is even more pronounced when comparing median values. Decreases in mean (median) values are accompanied with decreases in variation (IQR and standard deviation), which results in small changes in coefficients of variation. The presence of extreme observed values, even in the deeper layers, is an important characteristic of this data set and produces considerable differences between calculated mean and median values.

The same statistical quantities computed on data predicted by the equal-area spline function and averaged over the same depth increments reveal that overall distribution remains almost unchanged after transformation to the continuous form (Table 3).

Table 2
Depth-wise summary of observations.

	min	1st. quartile	mean	median	3rd. quartile	max	IQR	sd	CV	obs.
0–5 cm	4.00	18.80	51.22	39.00	65.70	328.00	46.90	49.26	0.96	195
5–15 cm	4.00	17.20	45.21	33.20	57.90	311.00	40.70	45.04	1.00	195
15–30 cm	3.10	10.70	35.53	23.70	46.50	311.00	35.80	40.02	1.13	181
30–60 cm	2.40	5.80	20.90	11.20	24.30	246.00	18.50	26.85	1.28	135
60–100 cm	1.80	3.90	19.68	6.00	11.90	228.00	8.00	44.14	2.24	52
100–120 cm	2.70	4.50	6.99	5.80	9.60	10.10	5.10	2.84	0.41	6

3.1. Trend analysis and spatial prediction

In order to examine if the assumptions which justify the usage of linear regression are met, the residual plots were created (Fig. 6). For this purpose, only the model fitted to the first layer data (0–5 cm) was selected. First five graphs (except graph in lower-right corner) depict the relation between residuals and each exposure parameters separately, while the last graph shows the residuals against fitted values. The lack of systematic curvatures in first five graphs confirms the linearity of the relationship between As data and exposure parameters. The presence of increasing variation of residuals with the level of fitted values depicted in the last graph indicates the moderate violation of the assumption of constant error variance.

The final sub-group of predictors was selected by combining the backward and forward stepwise regression procedure. According to the stepwise regression analysis, all exposure parameters were included in the final model. In addition, the interaction effects between ES and DEM as well as between DEM and CD were also found to be useful predictors for As prediction. Considering this, the final model for each soil layer can be formulated as follows:

$$m_{As}(s_i) = \beta_1 \cdot DEM(s_i) + \beta_2 \cdot ES(s_i) + \beta_3 \cdot MPI(s_i) + \beta_4 \cdot DD(s_i) + \beta_5 \cdot CD(s_i) + \\ + \beta_6 \cdot ES(s_i) \cdot DEM(s_i) + \beta_7 \cdot DEM(s_i) \cdot CD(s_i) \quad (8)$$

The final model parameters for all three layers were obtained using the GLS method within the RK algorithm given above. The estimated model coefficients together with basic accuracy measures for each soil layer are given in Table 4. The values shown in brackets represent the corresponding OLS coefficient estimates. Considerable differences between OLS and GLS estimates indicate the existence of significant spatial clustering between the observations in each layer.

The asterisks following the estimated coefficients indicate the level of statistical significance according to the Wald test. Statistical significance for each predictor, except for DD, was confirmed for each soil layer. As expected, the variance in As data explained by trend models decrease with depth. R^2 values ranged from 0.52 for the first layer to the 0.49 and 0.35 for the second and third layer. This is not so evident in RMSE which takes the marginally smaller value for the deeper soil layers. However, it is not surprising, considering the fact that R^2 represents the relative measure whereas the RMSE value represents an absolute measure of fit. As the DEM participates in each interaction effect, trend model coefficients for ES and CD can vary according to the level of altitudes. To illustrate this effect, we reported the trend model coefficients from the first layer model estimated for four different levels of altitudes: 400, 600, 800 and 1000 m (Table 5).

Results for Relative Importance (RI) analysis are depicted in Fig. 7. There can be noted that CD appeared as the dominant predictor in each model. For the first two layers, it participates in R^2 with the portion > 40%. It has followed by DEM and two interaction terms, while ES, MPI and DD showed considerably poorer predictive contribution.

Once the trend model was defined for all soil layers, the obtained residuals were then analyzed for spatial dependence. The presence of spatial dependence in residuals justifies the usage of kriging to improve the accuracy of the prediction. The lag increment was set to the 550 m, which provided a sufficient number of point pairs for reliable variogram estimation. The effects of trend removal on the spatial dependence structure are shown in Fig. 8. Typically, the presence of a spatial trend in the observed data causes monotonically increasing differences in data as separation increase, which is reflected in the experimental variogram that never reaches the sill. On the other hand, the residual variograms more accurately reflect the spatially correlated random effects, reaching the sill at a particular distance. This is particularly expressed in the first soil layer, where the trend removal has the greatest effect (Fig. 8a). This result confirms the fact that the atmospherically deposited spatial trend is more expressed in the soil surface layer. Generally, the sill variance and the nugget variances are substantially reduced in each soil layer. The differences between the two variograms became smaller in the deeper soil layers. It is also important to note the considerable decrease of range parameter associated for the second two layers indicating abrupt changes in spatial correlation over the soil depth (Fig. 8b and c). Similarly, nugget variance for the first soil layer is also substantially higher than in the second two layers. This might be due to a higher variation in As data in surface soil.

The final prediction accuracy indices together with residual variogram parameters are reported in Table 6. Kriging interpolation slightly improved the trend model performance in each soil layer, while the accuracy between soil layers remained almost unchanged. The absolute prediction accuracy (RMSE) still remains almost equal for each soil layer. Decreasing in accuracy was replicated in R^2values ranging from 0.55 for the surface soil layer to 0.36 for the deepest layer. This is also supported with the coefficients of variation (CV) of predicted values, which take the values 0.6, 0.7 and 0.9 for first, second and third soil layers respectively. The mean errors indicate a negatively biased prediction for all layers. Considerable lower accuracy obtained for the third layer could also be noted. This might be due to the fact that a significant part of systematic variation still remains unexplained by the trend model. The obtained results are comparable with results outlined

Table 3
Depth-wise summary of spline-predicted data.

	min	1st. quartile	mean	median	3rd. quartile	max	IQR	sd	CV	obs.
0–5 cm	1.00	20.96	52.53	41.94	67.62	326.15	46.66	48.93	0.93	195
5–15 cm	2.19	18.40	46.81	35.36	60.40	313.87	42.00	45.23	0.97	195
15–30 cm	1.46	12.76	36.49	25.54	44.06	305.86	31.31	40.81	1.12	181
30–60 cm	1.63	5.74	21.13	10.88	23.78	243.10	18.04	30.42	1.44	135
60–100 cm	1.00	3.43	17.30	6.91	13.16	225.65	9.73	35.65	2.06	52
100–120 cm	3.89	4.46	7.20	7.10	9.68	11.01	5.22	3.08	0.43	6

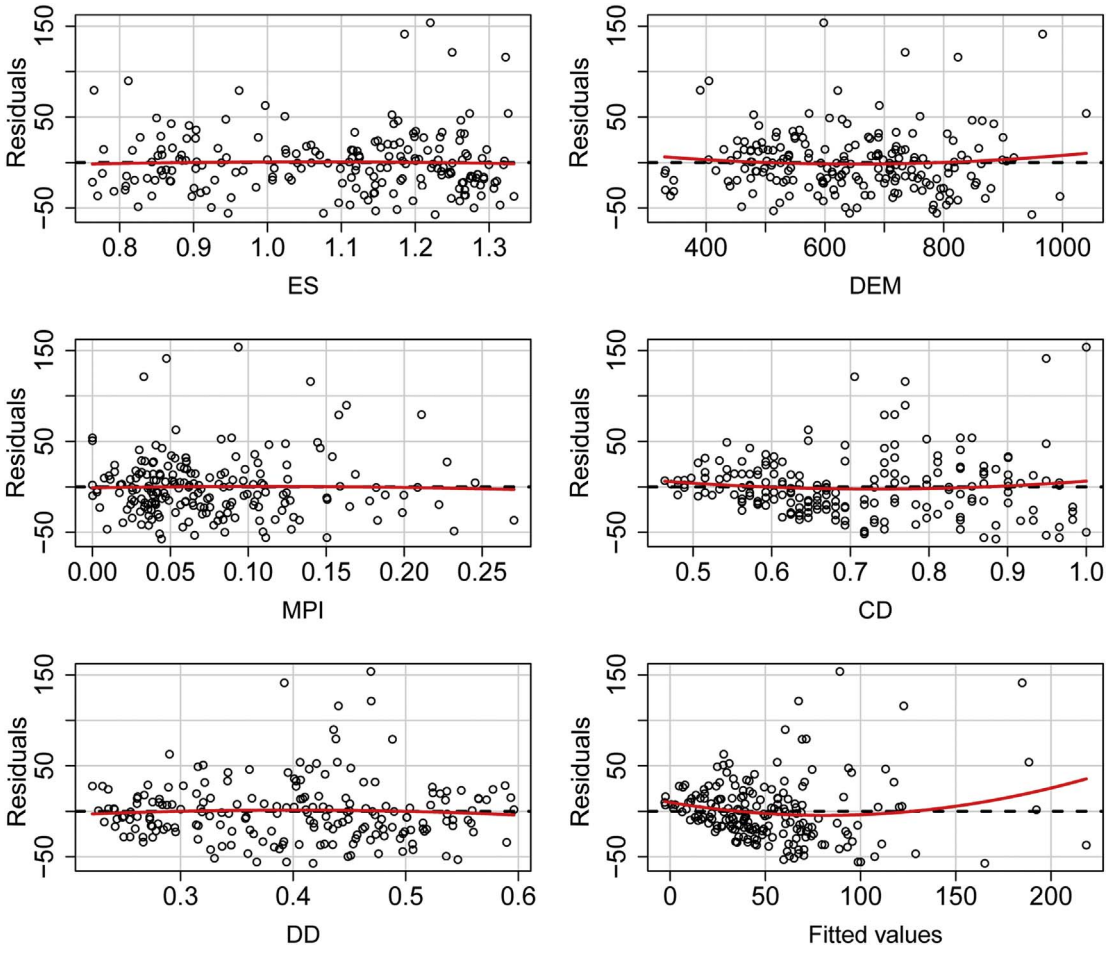

Fig. 6. Residual Plots for initial regression model fitted on data from the first soil layer (0–5 cm).

Table 4
The trend models coefficients and associated statistics.

Trend models:			
	0–5 cm	5–15 cm	15–30 cm
Constant	440.18**	394.96**	328.93**
	(369.88)	(385.96)	(327.38)
ES	− 285.37**	− 235.21**	− 192.61*
	(274.13)	(251.37)	(224.06)
DEM	− 0.85**	− 0.79**	− 0.64**
	(0.76)	(0.78)	(0.63)
MPI	115.63*	142.68**	144.30**
	(147.05)	(119.28)	(112.00)
CD	− 308.27*	− 269.65*	− 251.64*
	(209.49)	(220.80)	(204.37)
DD	80.41	51.06	74.03
	(76.53)	(47.46)	(79.58)
ES:DEM	0.44**	0.39**	0.32**
	(0.45)	(0.42)	(0.36)
DEM:CD	0.73**	0.65**	0.52**
	(0.58)	(0.58)	(0.45)
Observations	195	195	181
R^2	0.52	0.49	0.35
RMSE	33.7	32.2	32.7
F Statistic	30.01**	26.11**	13.77**

* $p < 0.05$.
** $p < 0.01$.

Table 5
Changing ES and CD coefficients according to the altitude level.

Elevation	ES	CD
400 m	− 109.37	− 16.27
600 m	− 21.37	129.73
800 m	66.63	275.73
1000 m	154.63	421.73

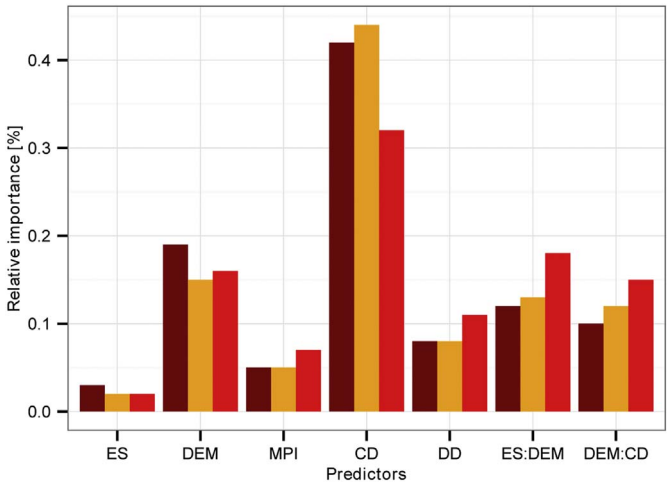

Fig. 7. Relative importance of predictors for each soil layer (0–5 cm – dark red, 5–15 cm – orange, 15–30 cm – red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in similar studies previously reported in literature by Adhikari et al., 2014, 2013; Goovaerts et al., 2008; Lacoste et al., 2014; Saito and Goovaerts, 2001. Moreover, the obtained results are in line with Beckett and Webster (1971), statement that values of R^2 higher than 0.7 are

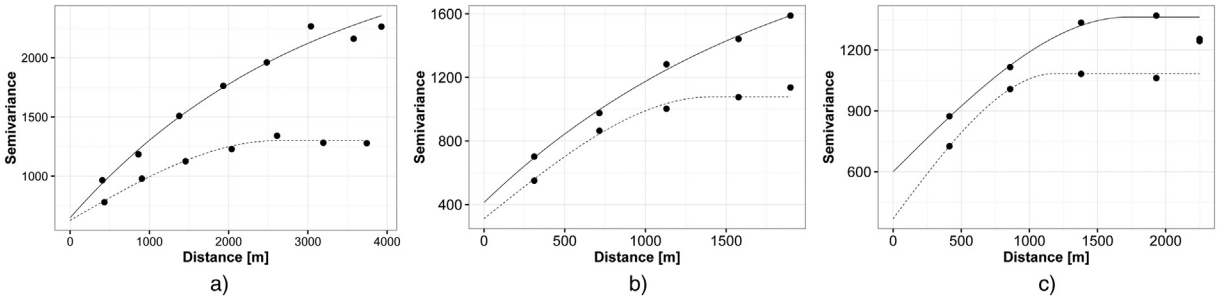

Fig. 8. Omnidirectional variogram models for observed data (black line) and residuals (gray line) for all soil layers: (a) 0–5 cm; b) 5–15 cm; c) 15–30 cm).

unusual and values of $R^2 < 0.5$ are quite common in soil attribute predictions.

The maps of final prediction for all layers are displayed in Fig. 9 (a-c). The exposed area with high As concentration in the central part dominate in all soil layers. The mean predicted value ranged from 58.1 mg/kg for the first soil layer to the 51.8 mg/kg and 41.6 mg/kg for the second and third layer respectively.

The prediction maps show that approximately 78% percent of the mapping area is above the allowable concentration limits for agricultural soils in Serbia (25 mg/kg) in accordance to regulations of the Republic of Serbia. This percentage, to some extent, decreases with depth (75% for 5–15 cm and 69% for 15–30 cm) suggesting that long term smelting activity leaves significant consequences on soil even on deeper unexposed layers. This also indicates that a certain amount of As can move downward in the profile through the process of water leaching (Adriano, 2001). It is important to note that the average distance between the explored area and the copper smelter is approximately 10 km, which also indicates a detrimentally large spatial consequence of smelting activity. On a larger scale, only the area that is to the west of the copper mine Bor is characterized to have elevated concentrations of As, which includes our case study area, (Fig. 10).

The far Southwestern part of the mapped area is also regarded as a highly contaminated area. Topographically, this area is characterized by a downhill front that is directly exposed to the smelter. However, the lack of observations in this area makes predictions unverifiable. Due to its location and topographic configuration, this area could be suitable for additional sampling and validation for this model.

Interactive Web-based maps were also created in order to obtain a better insight into the predicted spatial distribution of As concentration. The R package plotGoogleMaps (Kilibarda and Bajat, 2012) was used for creating these maps. These are available as interactive maps in HTML format at the web page URL:http://osgl.grf.bg.ac.rs/materials/Bor. In addition, the same maps in KML format, interactive point based maps as well as background data are also available at the same web page.

4. Conclusion

This paper reviews a method for geostatistical mapping of atmospherically-deposited pollutants from a known source by considering terrain exposure. The methodology applied is based on the so-called “Spline-Then-Krige” approach, which enables the production of a suite of maps for different soil depths. The exposure parameters were created to explain two different aspects of terrain exposure: Geometrical (Proximity) and Topographical exposure. Based on the obtained results, the main conclusion can be drawn as follows:

1. The equal-area spline depth function provides a reliable estimate of continuous vertical distribution of As data.
2. Stepwise regression analysis confirmed the utility of the predictive capability for all of the designed exposure parameters as well as for the two interactions: ES:DEM and DEM:CD. This confirms the hypothesis that there is an association between spatial spreading of Arsenic from the copper smelter in Bor and terrain exposure parameters.
3. The trend model showed good overall accuracy for all soil layers. The highest accuracy was obtained for the surface soil layer, where the model explained 52% of data variation. The trend model explained 49% of variations for the second layer, and for the third layer 35% of data variations.
4. The relative importance analysis showed that the trend models at each depth are highly controlled by the CD and DEM. Significant influences of interaction effects between ES and DEM as well as between DEM and CD at each depth indicate the importance of considering a more general model that includes interactions between exposure parameters.
5. The residual spatial dependence showed significant differences in structure between surface and other soil layers, indicating different effects of trend removal.
6. The kriging interpolation improved, to some extent, the regression accuracy for all three layers with R^2 ranging from 0.36 for the deepest layer to the 0.55 for the surface soil layer.
7. The relatively high RMSE values that follows the prediction on each soil layer indicates that a great portion of Arsenic data variation remains unexplained by the trend model, which implies that other variables in addition to the wind-driven process affects the Arsenic spatial distribution. However, in a situation when the wind indeed has an important role on spatial distribution of soil pollutants, integration of topographic exposure parameters could be useful for prediction, even on a deeper soil layers.
8. The obtained results were consistent with those reported by Goovaerts et al., 2008, suggesting that such an approach could be a promising alternative for complex air dispersion models. The direct comparison of these two approaches was not possible in this study due to missing data necessary for dispersion model characterization but it will certainly be the focus of a future study.

Table 6
Variogram parameters and final prediction accuracy indices.

	Nugget	Sill	Range	Nugget/Sill	R2	RMSE	ME	CV
0–5 cm	626.34	1299.88	2613.20	48.18	0.55	32.75	− 0.23	0.62
5–15 cm	311.68	1078.32	1410.88	28.90	0.51	31.70	− 0.26	0.67
15–30 cm	368.49	1084.21	1192.19	33.99	0.36	32.51	− 0.23	0.89

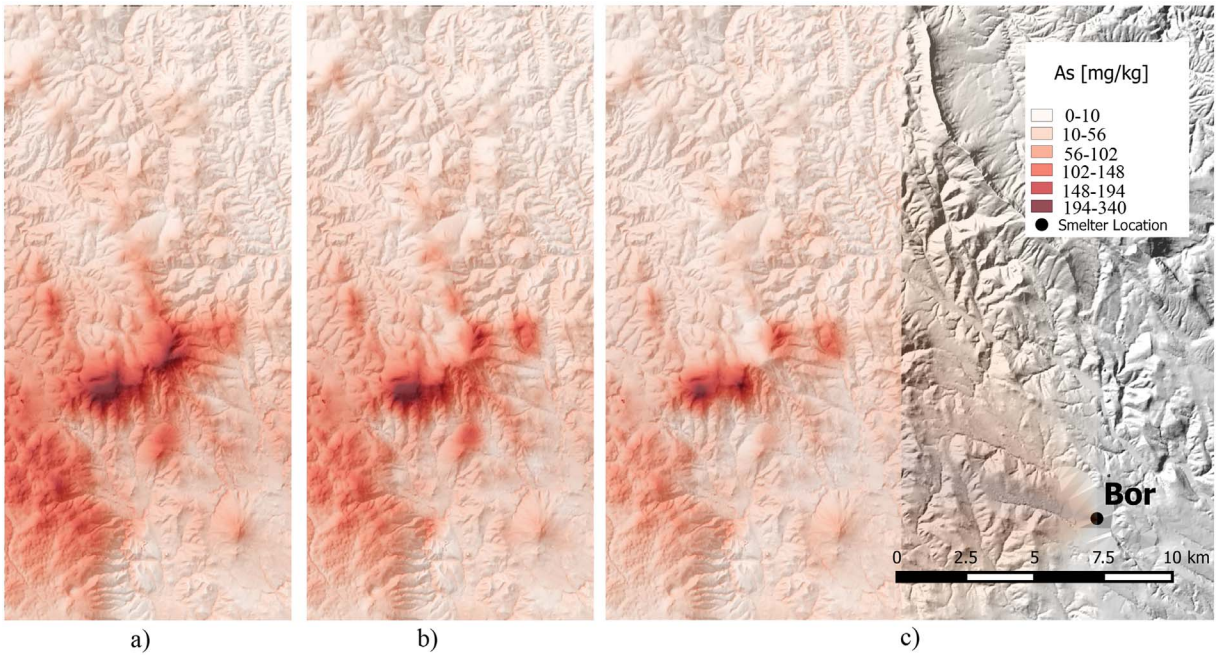

Fig. 9. Maps of predicted As concentration: a) 0–5 cm depth, a) 5–15 cm depth, a) 15–30 cm depth.

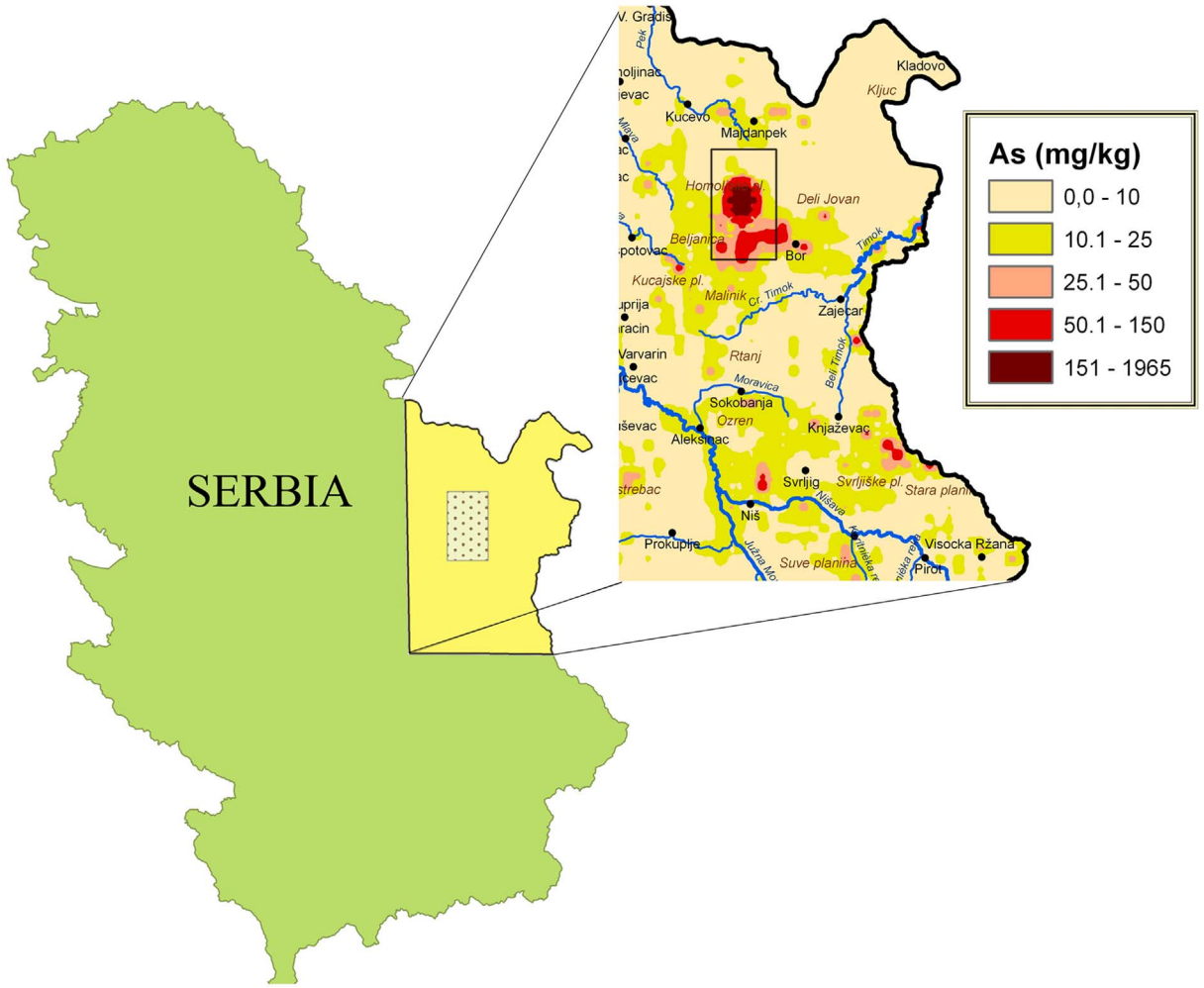

Fig. 10. Part of the broad map (Mrvić et al., 2009) of As concentrations for Serbia, the case study area is marked by rectangle.

Acknowledgments

This study was supported by the Serbian Ministry of Education and Science, under grants Nos. III 47014, TR 36035, TR 36009 and OI 173018.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.gexplo.2017.05.004. These data include the Background data set and KML file containing the Google map of As observations.

References

Adhikari, K., Hartemink, A.E., Minasny, B., Kheir, R.B., Greve, M.B., Greve, M.H., 2014. Digital Mapping of Soil Organic Carbon Contents and Stocks in Denmark.
Adhikari, K., Kheir, R.B., Greve, M.B., Bøcher, P.K., Malone, B.P., Minasny, B., McBratney, A.B., Greve, M.H., 2013. High-resolution 3-D mapping of soil texture in Denmark. Soil Sci. Soc. Am. J. 77, 860–876.
Adriano, D.C., 2001. Trace Elements in Terrestrial Environments: Biogeochemistry, Bioavailability, and Risks of Metals. Springer Science & Business Media.
Akaike, H., 1974. A new look at the statistical model identification. Autom. Control. IEEE Trans. 19, 716–723.
Antonić, O., Legović, T., 1999. Estimating the direction of an unknown air pollution source using a digital elevation model and a sample of deposition. Ecol. Model. 124, 85–95.
Arrouays, D., McBratney, A.B., Minasny, B., Hempel, J.W., Heuvelink, G.B.M., MacMillan, R.A., Hartemink, A.E., Lagacherie, P., McKenzie, N.J., 2014. The GlobalSoilMap project specifications. Glob. Basis Glob. Spat. Soil Inf. Syst. 9.
Beaudette, D.E., Roudier, P., O'Geen, A.T., 2013. Algorithms for quantitative pedology: a toolkit for soil scientists. Comput. Geosci. 52, 258–268.
Beckett, P.H.T., Webster, R., 1971. Soil variability: a review. Soils Fertil. 34, 1–15.
Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91, 27–45.
Cappuyns, V., Van Herreweghe, S., Swennen, R., Ottenburgs, R., Deckers, J., 2002. Arsenic pollution at the industrial site of Reppel-Bocholt (north Belgium). Sci. Total Environ. 295, 217–240.
Chlopecka, A., Bacon, J.R., Wilson, M.J., Kay, J., 1996. Forms of cadmium, lead andzinc in contaminated soils from southwest Poland. J. Environ. Qual. 25, 69–79.
Cressie, N., 1993. Statistics for Spatial Data: Wiley Series in Probability and Statistics.
De Visscher, A., 2013. Air Dispersion Modeling: Foundations and Applications. John Wiley & Sons.
EPA, S., 1995. User's Guide for the Industrial Source Complex (ISC3) Dispersion Models. (Volume I-User Instructions).
Erickson, T.A., Williams, M.W., Winstral, A., 2005. Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States. Water Resour. Res. 41.
Goovaerts, P., Trinh, H.T., Demond, A., Franzblau, A., Garabrant, D., Gillespie, B., Lepkowski, J., Adriaens, P., 2008. Geostatistical modeling of the spatial distribution of soil dioxins in the vicinity of an incinerator. 1. Theory and application to Midland, Michigan. Environ. Sci. Technol. 42, 3648–3654.
Grömping, U., et al., 2006. Relative importance for linear regression in R: the package relaimpo. J. Stat. Softw. 17, 1–27.
Hengl, T., 2015. GSIF: Global Soil Information Facilities.
Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007a. About regression-kriging: from equations to case studies. Comput. Geosci. 33, 1301–1315.
Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007b. About regression-kriging: from equations to case studies. Comput. Geosci. 33, 1301–1315. http://dx.doi.org/10.1016/j.cageo.2007.05.001.
Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75–93.
Kilibarda, M., Bajat, B., 2012. plotgooglemaps: the r-based web-mapping tool for thematic spatial data. Geomatica 66, 37–49.
Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E.J., Perčec Tadić, M., Bajat, B., 2014. J. Geophys. Res. Atmos. 119, 2294–2313. http://dx.doi.org/10.1002/2013JD020803.Received.
Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C., 2014. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213, 296–311.
LEAP, 2003. Local Environmental Action Plan.
Lindeman, R.H., Merenda, P.F., Gold, R.Z., 1980. Introduction to Bivariate and Multivariate Analysis. Scott, Foresman Glenview, IL.
Lindsay, J.B., Rothwell, J.J., 2008. Modelling channelling and deflection of wind. Adv. Digit. Terrain Anal. 383.
Luo, W., Lu, Y., Wang, G., Shi, Y., Wang, T., Giesy, J.P., 2008. Distribution and availability of arsenic in soils from the industrialized urban area of Beijing, China. Chemosphere 72 (5), 797–802.
Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152.
Mrvić, V., Zdravković, M., Sikirić, B., Čakmak, D., Kostić-Kravljanac, Lj, 2009. Content of hazardous and harmful elements. In: Mrvić, V., Antonović, G., Lj, Martinović (Eds.), Fertility Status and Content of Hazardous and Harmful Elements in the Soils of Central Serbia. Institute for Soil Science, Belgrade, pp. 75–145 (in Serbian).
Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D., 2016. National versus global modelling the 3D distribution of soil organic carbon in mainland France. Geoderma 263, 16–34.
Orton, T.G., Pringle, M.J., Bishop, T.F.A., 2016. A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. Geoderma 262, 174–186.
Pebesma, E.J., 2006. The role of external variables and GIS databases in geostatistical analysis. Trans. GIS 10, 615–632. http://dx.doi.org/10.1111/j.1467-9671.2006.01015.x.
Plattner, C., Braun, L.N., Brenning, A., 2004. Spatial variability of snow accumulation on Vernagtferner, Austrian Alps, in winter 2003/2004. Z. Gletscherk. Glazialgeol. 39, 43–57.
Saito, H., Goovaerts, P., 2001. Accounting for source location and transport direction into geostatistical prediction of contaminants. Environ. Sci. Technol. 35, 4823–4829.
Šerbula, S.M., Antonijević, M.M., Milošević, N.M., Milić, S.M., Ilić, A.A., 2010. Concentrations of particulate matter and arsenic in Bor (Serbia). J. Hazard. Mater. 181, 43–51.
Serbula, S.M., Kalinovic, T.S., Kalinovic, J.V., Ilic, A.A., 2013. Exceedance of air quality standards resulting from pyro-metallurgical production of copper: a case study, Bor (Eastern Serbia). Environ. Earth Sci. 68, 1989–1998.
Unit, S.C., 2013. Science for Environment Policy In-depth Report: Soil Contamination: Impacts on Human Health.
USEPA, 1998. SW-846: Test Methods for Evaluating Solid Waste, Physical and Chemical Methods. (Washington).
Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. John Wiley & Sons.
Winstral, A., Elder, K., Davis, R.E., 2002. Spatial snow modeling of wind-redistributed snow using terrain-based parameters. J. Hydrometeorol. 3, 524–538.
Yokoyama, R., Shirasawa, M., Pike, R.J., 2002. Visualizing topography by openness: a new application of image processing to digital elevation models. Photogramm. Eng. Remote. Sens. 68, 257–266.
Žibret, G., Šajn, R., 2008. Modelling of atmospheric dispersion of heavy metals in the Celje area, Slovenia. J. Geochem. Explor. 97, 29–41.