Application and improvement of soil spatial distribution mapping using advanced modeling techniques

Jasminka Alijagić* and Robert Šajn

Geological Survey of Slovenia, Dimičeva ulica 14, 1000 Ljubljana, Slovenia; (Jasminka.Alijagic@geo-zs.si; Robert.Sajn@geo-zs.si)
*corresponding author

Crossref
Similarity Check
Powered by iThenticate

Abstract

The main purpose of this contribution is to develop realistic prediction digital soil maps in order to increase their visuality, and to evaluate and compare the performance of different modeling techniques: a) Kriging, b) Artificial Neural Network – Multilayer Perceptron (ANN-MLP) and c) Multiple Polynomial Regressions (MPR). The following criteria were used to determine selection of the testing site for the modeling: (1) intensive metal ore mining and metallurgical processing; (2) geomorphological natural features; (3) regular geological setting, and (4) the remaining minefields. The success of Digital Soil Mapping and the plausibility of prediction maps increases with the availability of spatial data, the availability of computing power for processing data, the development of data-mining tools, geographical information systems (GIS) and numerous applications beyond geostatistics. Advanced prediction modeling techniques, ANN-MLP and MPR include geospatial parameters sourced from Digital Elevation Models (DEM), land use and remote sensing, applied in combination with costly and time-consuming soil measurements, developed and finally incorporated into the models of spatial distribution in the form of 2D or 3D maps. Innovative approaches to modeling assist us in the reconstruction of different processes that impact the entire study area, simultaneously. This holistic approach represents a novelty in contamination mapping and develops prediction models to help in the reconstruction of main distribution pathways, to assess the real size of the affected area as well as improving the data interpretation.

1. INTRODUCTION

The soil can be characterised as a natural purifying agent and a place where pollutants sink and/or accumulate. Metals occur in rocks and in soils whether in the soil solution, organically bound, bound to hydrous oxides of Fe, Mn and Al, exchangeable, acid-soluble (bound to carbonates, phosphates, etc.) or structurally bound in the residual fraction (KABATA-PENDIAS & PENDIAS, 2001). When the metal retention capacity of the soils is exceeded (ROMIĆ, 2012) and when soil conditions are changed (e.g. reaction kinetics, pH and Eh values, organic matter content, microorganisms, etc.), simple ions can be released into solution, become mobile and thus accessible to plants (KABATA-PENDIAS & PENDIAS, 2001). Sources of trace elements in groundwater can be natural, (originating from earth materials), or the elements can be introduced by industrial activities, landfill, mining and transportation (BRADL, 2005). Important sources of trace metals input into the soils are from the metals industry. Starting from mining and milling operations, concentrating, sintering and metal ore transportation to the disposal of tailings. Smelting processes can include atmospheric discharges, blown dust and cause acidification of water bodies (DUDKA & ADRIANO, 1997). It is known that the majority of ore minerals accommodate various other trace metals as minor inclusions, causing contamination with major and minor ore metals in areas surrounding mines and smelters (SHALLARI et al., 1998; CHOPIN & ALLOWAY, 2007; AL-KHASHMAN & SHAWABKEH, 2009).

The metal ore mining and processing are some of the largest releases of trace elements causing a considerable impact on the surrounding environment. Obsolete metal production technologies used for mass production generate an enormous quantity of mining waste which is considered hazardous due to the presence of toxic metals (SINGH et al., 2010; NAVARRO et al., 2008). Due to past and present mining activities, many worldwide rivers are significantly affected by multiple metal contamination (FERREIRA et al., 2016). Metal containing particles are quite often transported over large distances by water. This can affect the alluvial sediments several hundred kilometres away from the main source of contamination (ŠAJN et al., 2011; YI et al., 2012; ZHAO et al., 2013; PAVLOWSKY et al., 2017).

The soil maps can be generally understood as a graphic representation for transmitting information about the spatial distribution of soil attributes (YAALON, 1989). According to MCBRATNEY et al. (2003), traditional soil mapping techniques mostly depend on ground-based surveys but they rarely inform us about the spatial distribution of soil properties at the desired resolution over the landscape.

Complex interactions between the soil and geospatial parameters of site factors sourced from Digital elevation models (DEM) is crucial for understanding the main spatial distribution pathways. Their mutual coherence can be integrated by modeling techniques with an advanced prediction (ALIJAGIĆ, 2013). Environmental surveys are usually based on sampling on the relatively small and quite well-delimited system, but in general, the measurements represent a continuum in space from which the sample has been drawn (OLIVER & WEBSTER, 2014). A great potential for improving predictive soil mapping (PSM) techniques and understanding the relationships between soils and the environment, has been created after the increase in computer efficiency and capacity, geo-information technology, data availabi-

lity and demand for accurate and reliable maps (MCBRATNEY et al., 2003; SCULL et al., 2003; BAGHERI BODAGHABADI et al., 2015; ZHANG et al., 2017). It requires the creation of spatial soil information based on numerical models, which account for spatial and temporal variations in soil properties on the basis of soil information and related environmental variables (LAGACHERIE & MCBRATNEY, 2007).

The main purpose of this study was to develop a visualisation model of contamination using linear and nonlinear mathematical methods that combine a sparse chemical analysis and various geospatial parameters in Stavnja Valley (Bosnia and Herzegovina) but at the same time to evaluate and compare the performance of different modeling techniques: Kriging, Artificial Neural Network – Multilayer Perceptron (ANN-MLP) and Mul-

tiple Polynomial Regressions (MPR). For this reason, five specific purposes of the present study were (1) to include various geospatial parameters in linear and nonlinear modeling; (2) to identify the main distribution pathways and change in concentrations according to the distance from the source of contamination; (3) model verification; (4) to assess the real size of the affected area, and (5) to provide an optimal methodology for geochemical researchers in areas with restricted sampling conditions.

2. MATERIALS AND METHODS

2.1. Description of the study area

The river Stavnja has a length of about 35 km, located in Central Bosnia and Herzegovina, north of the capital Sarajevo (Fig. 1). Approximately 30,000 people inhabit this Valley, mostly settled

Figure 1. Location of the study area.

into the small cities of Vareš and Breza. The Stavnja River catchment has been known for intensive mining and metallurgic activities for more than 100 years. Besides these activities, the valley was selected for study due to several other reasons: its unique geological setting – especially the metallogenic belt, the geomorphological characteristics that affect the contamination halo and remaining minefields from the last war, 1992-1995.

The anthropogenic impact in the municipality of Vareš can be reflected through the three abandoned iron ore deposits (Smreka, Brezik and Droškovec), an abandoned Pb-Zn-Ba deposit of Veovača, the abandoned ironwork at Vareš, and brown coal mining in the southern part of the study area. High concentrations of particular elements are released into the environment not only by anthropogenic activities but also by natural erosion and weathering reactions of the parental rocks that add to the complexity of the environmental assessment (ALIJAGIĆ, 2013).

Problems of contamination along the river can be demonstrated as the anisotropic appearance between the layered (isotropic) lithological units, which can be solved using advanced linear and nonlinear modeling techniques instead of a denser sampling grid which is restricted by the remaining minefields at 5-6 % of the study area. The physiography of the region displays great variability. Contrasting landscapes are observed from the mountain environment in the northern part of the study with an altitude of more than 1000 m, to the southern part at approximately 400 m (Fig. 2).

2.2. Geological description of the study area

The outcropping stratigraphic sequence exposes rock formations spanning the Triassic to the Cretaceous, and more recent times. Ten major lithological units (Fig. 3) had been isolated, ranging from the oldest to the youngest in a N-S direction. The most important geotectonic unit of the Vareš metallogenic district belongs to the Central ophiolite zone. Vareš, siderite–haematite sedimentary exhalative (SEDEX) deposits at Smreka, Droškovac and Brezik are the *locus typicus* of mineralisation of the Mid-Triassic. The deposits contain hydrothermal, stratiform siderite–haematite–chert beds. The mineralisation forms part of the Anisian and Ladinian sequences and displays distinct vertical zoning, reflecting a gradual change of redox conditions in the depositional environment. The Veovača Pb, Zn, Ba deposit contains ore breccia or ore conglomerates with dm to m sized clasts cemented by barite and Pb–Zn sulphides. The barite from Veovača is typical for Triassic SEDEX deposits elsewhere in the Dinarides (PALINKAŠ et al., 2008).

2.3. Sampling, sample preparation and chemical analyses

Regarding the primary purpose of research, previous experience (ŠAJN, 2005, 2006; ALIJAGIĆ, 2008; ALIJAGIĆ & ŠAJN, 2006, 2010; STAFILOV et al., 2008a, 2008b), and some sampling difficulties (narrow gorge, inaccessible terrain and remained minefields), resulted in preparation of a sampling grid so that the soil profiles were raised transverse to the river flow (Fig. 4). Respect-
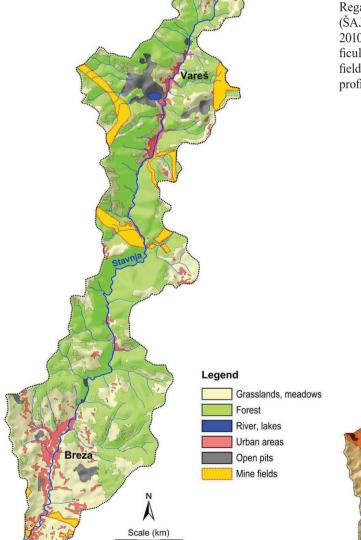

Legend

- Grasslands, meadows
- Forest
- River, lakes
- Urban areas
- Open pits
- Mine fields

Figure 2. Land use map.

Legend

- Quaternary alluvium
- Quaternary river terraces
- Miocene clastic series
- Miocene carbonate series
- Oligocene clastite complex
- Cretaceous flysch
- Jurassic and Cretaceous clastic carbonate series
- Jurassic and Cretaceous breccias and sandstones
- Triassic limestone
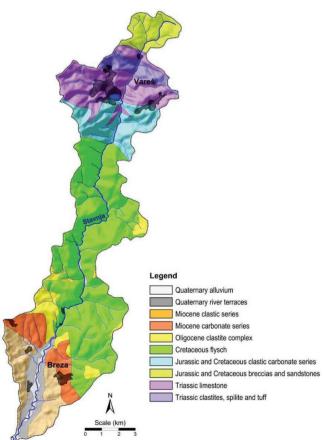- Triassic clastites, spilite and tuff

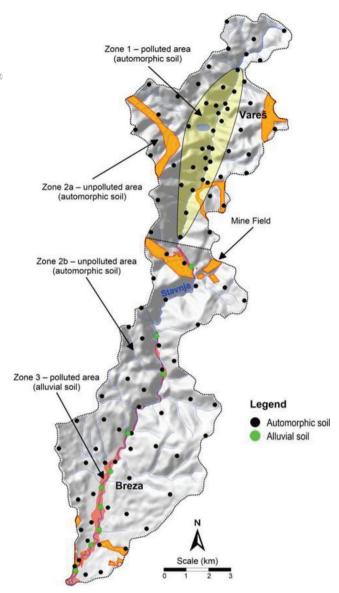**Figure 3.** Isolated lithological units.

Figure 4. Sampling sites with determined zones.

ing the changes of environmental chemism according to the distance from the source of contamination, a change in altitude, and a transport mode, but at the same time understanding better the geogenic and anthropogenic influence on the level of trace element changes, the entire survey area was covered with 12 profile lines set to cross the river flow. Some additional samples had been collected at the edges of the study area and some lithological units that were not covered by the profile lines. In the urban and industrial zones, the sampling grid is slightly more dense (Fig. 4). The total number of soil sampling sites is 111 but at each site, samples were collected from two depths: in the topsoil (0 - 5 cm) and subsoil (20 - 30 cm) respectively.

Initially, the samples were air-dried and subsequently dried in a fan drier at 40 °C. All samples were analysed in the ACME, Ltd. laboratory in Vancouver, Canada (ACME Labs, 2010, 2011). Determination of 36 chemical elements (Ag, Al, As, Au, B, Ba, Bi, Ca, Cd, Co, Cr, Cu, Fe, Ga, Hg, K, La, Mg, Mn, Mo, Na, Ni, P, Pb, S, Sb, Sc, Se, Sr, Th, Ti, Tl, U, V, W and Zn) was performed by inductively coupled plasma mass spectrometry (ICP-MS) after aqua regia digestion, during which we assisted with the results for precision and trueness based on ALIJAGIĆ (2013).

3. DATA ACQUISITION AND DATA PROCESSING

All data processing and calculations, geostatistical data interpretation and visualization (mapping) have been performed using the following software: Statistica (Stat Soft Inc., 2012), Autodesk MAP 3D (Autodesk Inc., 2012), ArcINFO (ESRI Inc., 2004) and Surfer (Golden Software Inc., 2012).

In recent years, geographical information systems (GIS) have been used for spatial data management and manipulation. For this purpose various spatial data were acquired by digitalization of existing topographic maps: (1) Vareš 4 (475-4) in scale 1:50,000; Vareš 4-1 (475-4-1); (2) Vareš 4-2 (475-4-2), Vareš 4-3 (475-4-3), Vareš 4-4 (475-4-4), Sarajevo 2-2 (525-2-2), Sarajevo 2-1 (525-2-1) in scale 1:25,000 (provided by the Geodetic survey of Bosnia and Herzegovina); (3) Google Earth maps (Google Inc., 2010); (4) maps Breza-Vareš, scale 1:50,000; (5) Breza-Podlugovi, Brgule, Karaule, in scale 1:10,000 (provided by BH Mine Action Centre, in Sarajevo); (6) geological maps Vareš (L 34-133) and Sarajevo (K 34-1) in scale 1:100,000 (provided by Geological Survey of Slovenia and Geological Survey Federation of Bosnia and Herzegovina); (7) 80 m SRTM Digital Elevation Model (CGIAR Consortium for Spatial Information, 2011); (8) 30 m ASTER Digital Elevation Model (U.S. Geological Survey, 2011a) and (9) Landsat multispectral satellite images – 7 bands (U.S. Geological Survey 2011b).

All the aforementioned maps were used to obtain as much geospatial data as possible to be included in the database. The database includes the following information for 222 samples: Identification number, Sample label, Sampling material, Land use units, Lithological units, Defined zones, Latitude, Longitude, Absolute distance from the ironworks chimneys (Iron open pits, Smreka and Brezik; Ironworks Vareš, Breza coal mine) (Fig. 5A), Elliptical distance from the ironworks chimneys (Fig. 5B), Distance from the river Stavnja (Fig. 5C), Altitude (Fig. 6A), Altitude above the Stavnja valley bottom (Fig. 6B), Terrain Slope (Fig. 6C), Aspect, Plan Curvature (Fig. 7A), Profile Curvature (Fig. 7B), Tangent Curvature (Fig. 7C), and Landsat spectral bands 1 - 7 (Table 1, Fig. 8). Beside the aforementioned geospatial data and multispectral image bands, this database includes the analytical data (26 selected elements: Ag, Al, As, Ba, Bi, Cd, Co, Cr, Cu, Fe, Ga, Hg, La, Mg, Mn, Mo, Ni, Pb, Sb, Sc, Th, Ti, Tl, V, W and Zn). These elements represent mostly athropogenically entered into environment (Pb-Zn-Hg-Cd-Cu-Bi-Ag-Sb-Mo-W-Mn-Ba-Fe-Ti) or they are mainly influenced by heterogeneous lithology (Ni-Cr-Co-Mg-Th-La-As-Sc-Al-Ti-V-Ga) which content is determined by the parent material and soil properties (ALIJAGIĆ et al., 2015).

For modeling of the spatial distribution of a group of elements with Artificial Neural Network and Multiple Polynomial Regression, a recall grid was used. This means that the whole study area was divided into 50 x 50 m grid cells. The total number of recall points is 41,471. This database includes only geospatial data and the Landsat multispectral image bands mentioned before.

3.1. Topographic and geological data

The maps are constructed with the limitation of representing 3D real-world objects into a 2D representation, which involves some distortion of the shape, area, distance and direction of the spatial objects. Since the topographic maps (scale 1:25,000 and 1:50,000) were ambulated in the late `70s, to get more realistic spatial data, the maps from Google Earth and multispectral satellite images were combined with them. Particular land-use units are digi-
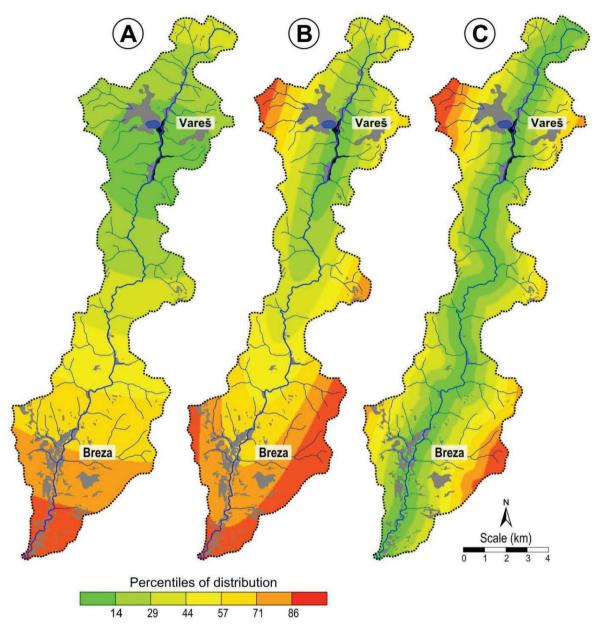
Figure 5. A – Absolute distance from the ironworks chimneys; B –Elliptical distance from the ironworks chimneys (ratio 1/5); C – Distance from the river Stavnja.

talized and isolated directly from Google Earth in the form of KML files. Such maps have been georeferenced, which means that we defined their existence in physical space, establishing locations in terms of map projections or a coordinate system. According to HILL (2006), the term is used both when establishing the relationship between raster or vector images and coordinates, and when determining the spatial location of other geographical features. This procedure is thus imperative to data modeling in the field of GIS and other cartographic methods. When data from different sources needs to be combined and then used in a GIS application, it becomes essential to have a common referencing system. Different maps may use different projection systems. Georeferencing tools contain methods to combine and overlay these maps with minimum distortion.

Beside the topographic maps, the basic geological map (scale 1:100,000) has also been digitalized and vectored. Isolation of the major lithological units is also an important step for the determination of the natural and anthropogenic background. The maps (1:25,000 and 1:10,000) of the remaining minefields and possible

minefields were obtained from the BH Mine Action Centre, in Sarajevo. Data about the minefields were isolated from them and incorporated into the previously digitalized topographic maps.

The given data were presented into three different projection systems: topographic, geological and minefield maps in the Gauss Kruger projection (zone 6) with a Hermanns-Kogel datum; the Google Earth and ASTAR in Unprojected Lat/Long projection with a World Geodetic System 1984 (WGS84) datum; and satellite images in the Universal Transverse Mercator (UTM 34N) projection with a WGS84 datum. All the aforementioned data have been converted into one projection system i.e. Gauss Krüger.

3.2. DEM and Terrain modeling (geomorphometry)

A Digital Elevation Model (DEM) is a quantitative representation of the Earth's surface providing basic information about the terrain relief (GUTH, 2006). A DEM is an important tool for geomorphometry, as its derived attributes (such as slope, aspect, drainage area and network, curvature, topographic index, etc.) are important parameters for information extraction or assess-
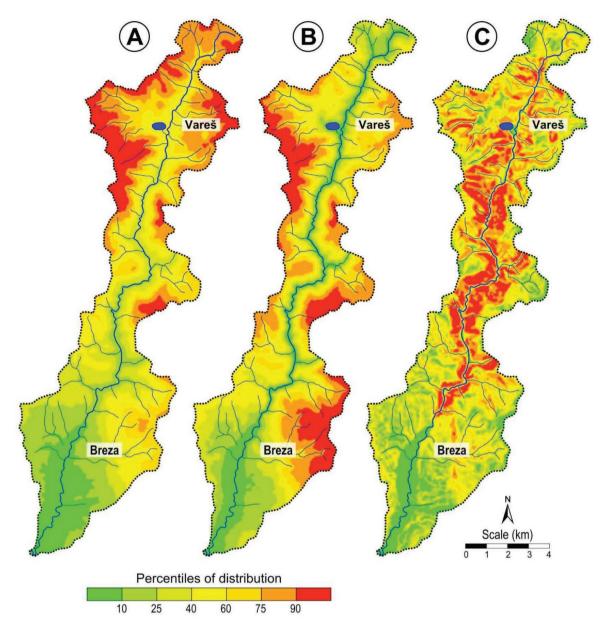
Figure 6. A – Altitude above sea level (absolute); B – Altitude above the bottom of the Stavnja river valley (relative); C – Terrain Slope.

ment of any process using terrain analysis (WOLOCK & PRICE, 1994). Providing all these geomorphometrical data, a DEM is prerequisite in different applications such as water flow modeling (JAIN & SINGH, 2003), volcanic hazards (VASSILOPOULOUA et al., 2002), terrain visualization and mapping (SPARK & WILLIAMS, 1996), flood simulation and management (RAMLAL & BABAN, 2008, HONGHAI & ALTINAKAR, 2011), climate and meteorological studies (THORNTON et al., 1997), etc. The outcomes of the models depend on the accuracy of the DEM (ZHANG & MONTGOMERY, 1994; MUKHERJEE et al., 2013). The accuracy of a DEM is especially important when used for the spatial prediction of soil attributes, as differences in topography play a crucial role in the distribution of pedological relationships (THOMPSON et al., 2001). Like all models, DEMs contain a certain degree of error. Errors in DEM datasets are usually related to (1) data errors, (2) measurement errors and (3) processing errors (BURROUGH, 1986; WISE, 1998; WECHSLER, 2003). They can also be characterised as gross errors, systematic errors and random errors. The main question to be answered here is to

which level of significance do certain errors extend (FISHER & TATE, 2006)?

Figure 5 provides the following geospatial information: (A) a cyclic distance from the Vareš ironwork (the main source of contamination); (B) an elliptical distance from the Vareš ironworks; and (C) the distance from Stavnja. These three distance shapes are prepared at a scale of 1:5. The cyclic and elliptical distances have been calculated from the ironworks chimney. For easier understanding, each distance is represented by a different colour in seven equal percentile classes. Several pieces of morphological spatial information: (A) an altitude; (B) an altitude above the bottom of the Stavnja valley, and (C) a slope, are provided in Fig. 6.

Terrain Aspect calculates the downhill direction of the steepest slope (i.e. the dip direction) at each grid node, with values eported as azimuths. The dip direction is perpendicular to the contour lines on the surface and is exactly opposite the uphill gradient direction. Terrain Slope is reported in degrees from zero (horizontal) to 90 (vertical). It is based on the direction of the
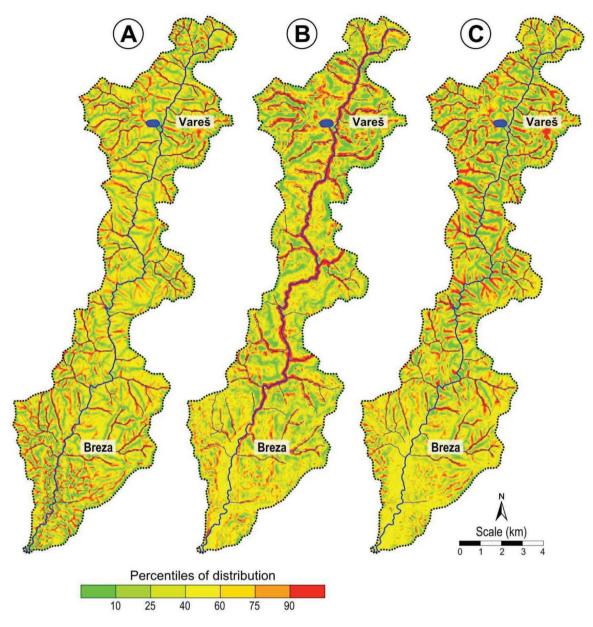
Figure 7. A – Plan terrain curvature, B – Profile terrain curvature, C – Tangent terrain curvature

steepest descent or ascent at that point (Terrain Aspect). This means that for a particular point on the surface the gradient direction can change.

Plan Curvature calculates the curvature of the surface in the horizontal plane or the curvature of the contour. Negative values indicate divergent water flow over the surface, and positive values indicate convergent flow. Profile Curvature measures the curvature of the surface in the direction of the gradient. It determines the downhill or uphill rate of change in the slope in the gradient direction (opposite of slope aspect direction) at each grid node. Grid files of Profile Curvature produce contour maps that show isolines of a constant rate of change of the steepest slope across the surface. Negative values are convex upward and indicate the accelerated flow of water over the surface. Positive values are concave upward and indicate slowed flow over the surface. Tangential Curvature measures curvature in relation to a vertical plane perpendicular to the gradient direction, or tangential to the contour. The negative and positive areas are the same as for Plan Curvature, but the curvature values are different. Tangential Cur-

vature measures the curvature of the surface in the vertical plane perpendicular to the gradient direction.

The mathematical definitions, the general review of the methods, and applications of topographic analysis (terrain slope, terrain aspect, plan curvature, profile curvature and tangential curvature) were taken from MITASOVA & HOFIERKA (1993) and MOOR et al. (1993).

3.3. Satellite images

Since 1972, Landsat satellites have continuously acquired space-based images of the Earth's land surface, coastal shallows, and coral reefs and have since provided worldwide science and resource management communities with an archive of space-based land remotely sensed data – a valuable resource for people who need this data for their work (U.S. GEOLOGICAL SURVEY, 2011c).

Landsats 5 and 7 each complete approximately 14 full orbits of the Earth each day. While each satellite has a 16-day full Earth coverage cycle, their orbits are offset to allow 8-day repeat coverage
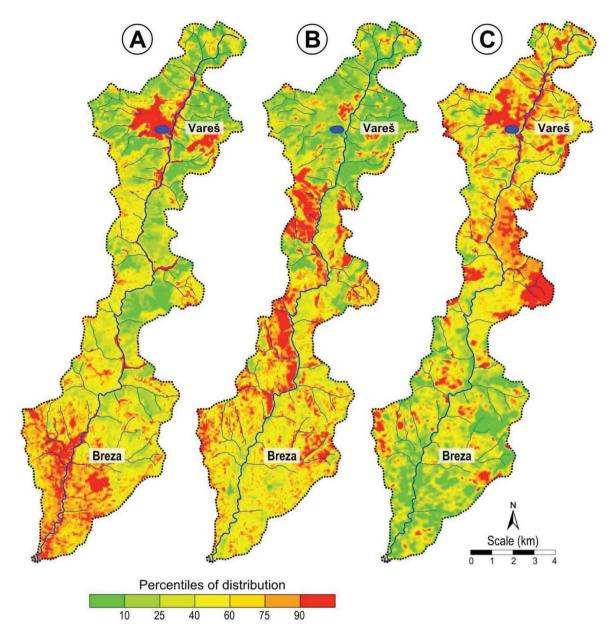
Figure 8. Relative intensity of radiation Landsat spectral bands. A – Visible spectrum, 0.45 – 0.69 μm (Bands 1 - 3); B – Infrared spectrum, 0.76 – 0.90 *μm (Band 4); C* – Thermal radiation, 10.4 - 12.5 μm.

Table 1. Landsat spectral bands.

Spectral bands	Wavelength (μm)	Resolution (ms)	Use
Band 1 (blue)	0.45–0.52	30	Bathymetric mapping; distinguishes soil from vegetation; deciduous from coniferous vegetation.
Band 2 (green)	0.52–0.61	30	Emphasizes peak vegetation, which is useful for assessing plant vigour.
Band 3 (red)	0.63–0.69	30	Emphasizes vegetation slopes.
Band 4 (IR)	0.76–0.90	30	Emphasizes biomass content and shorelines.
Band 5 (IR)	1.55–1.75	30	Discriminates moisture content of soil and vegetation; penetrates thin clouds.
Band 6 (thermal)	10.4–12.5	120/60	Useful for thermal mapping and estimated soil moisture.
Band 7 (IR)	2.08–2.35	30	Useful for mapping hydrothermally altered rocks associated with mineral deposits.

of any Landsat scene area on the globe. Landsat 7 carries the Enhanced Thematic Mapper Plus (ETM+), with 30-metre visible, near-IR, and shortwave infrared bands, a 60-metre spatial resolution thermal band, and a 15-metre panchromatic band (Table 1).

Moreover, since 2008 the USGS has freely provided all archived Landsat images, along with newly acquired Landsat 7 (launched in 1999) ETM+ SLC-off and Landsat 5 (launched in 1984). TM images with less than 40 % cloud cover, thereby enabling free access to multiple images of the same sectors. It depends what you need as there are many Landsat data formats: GeoTIFF, MTL, GCP, WO) (U.S. GEOLOGICAL SURVEY, 2011c).

In order to evaluate the capability of mapping contaminated areas from both Landsat TM and ETM data, we processed and analysed two available images for the study area (187 paths and

29 rows). Extracting information from satellite imagery often involves image interpretation techniques as well as GIS integration of other spatial data. Multispectral satellite imagery offers several advantages, such as a large number of data records, the availability of repeated images of a single place at different times, and the fact that virtually the entire planet is covered. We used images from 1990, in a period of intensive mining and smelting, and from 2005 after the civil war, a period when production was stopped.

3.4. Data transformation

Transforming data means performing the same mathematical operation on each piece of original data the main purpose of which is to reduce the difference between extreme values (GRINGARTEN & DEUTSCH, 2001). It is often observed that environmental variables are Log-normal (KRIGE, 1951, 1960; ROSE at al., 1979) or positively skewed (ZHANG et al., 1995; ZHANG & SELINUS, 1998), and data transformation is necessary to normalise such data sets. The Box-Cox transformation represents one of the most frequently used transformations in environmental sciences and geosciences where power transformation is needed (BOX & COX, 1962; JOBSON, 1991; ZHANG et al., 1998; MCGRATH, et al., 2004).

The Box-Cox transformation is given by:

$$y = \frac{X^{\lambda} - 1}{\lambda}; \; \lambda \neq 0$$

where y is the transformed value, and x is the value to be transformed.

3.5. Modeling techniques

3.5.1. Kriging

Kriging is one of the most popular geostatistical prediction methods successfully applied widely in the estimation and mapping of soil attributes in unsampled areas giving the best linear unbiased prediction of the concentration with minimal variance (LI et al., 2016). Kriging has become a quite effective tool for the study of hazard risk assessments and spatial uncertainty in general (GOOVAERTS, 1999, 2001), but at the same time it enables quantification of the main spatial characteristic of soil attributes and spatial interpolation methods considering only the neighbouring points of estimation data (CARLON et al., 2001; THEODOSSIOU & LATINOPOULOS, 2006). The most commonly used geostatistical prediction method is ordinary kriging that uses a semivariance function based on the assumption that nearby objects tend to be more similar than those that are farther apart.

An unbiased estimator of the semivariance function is half of the average squared difference between paired data values (OLIVER & WEBSTER, 2014):

$$y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z(x_1) - Z(x_1 + h) \right]^2$$

where γ(h) is the semivariance at a given distance h; and $N(h)$ is the number of sample pairs in h, and $z(x_i + h)$ and $z(x_i)$ are sample values at two points separated by h.

The variogram model mathematically specifies the spatial variability of the data set and the resulting grid file. The interpolation weights, which are applied to data points during the grid node calculations, are direct functions of the variogram model (WEBSTER & OLIVER, 2001). The graphical display is the final goal of many environmental scientists in the chain of steps that begins with sampling the prime objective of which is to utilise the power of the visual appearance (OLIVER & WEBSTER, 2014).

3.5.2. Artificial neural network – Multilayer Perceptron

An Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by biological nervous systems, which is both relatively easy to implement and very reliable (AITKENHEAD et al., 2012). It is composed of a large number of highly interconnected processing elements (neurons or nodes) working in unison to solve specific problems. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons (ŽIBRET et al., 2012).

The development of layered feed-forward ANNs began in the late 1950s by Rosenblatt's perceptron as the first model for learning with a teacher (i.e. supervised learning), which is often referred to as a single-layer perceptron. The multilayer perceptron (MLP) is one of the most popular feedforward ANNs that performs tasks such as function fitting and pattern recognition problems (ASCE, 2000). MLP can be used for the classification of linearly inseparable patterns and can also work as universal approximators. MLPs are feedforward neural networks (FNNs) with one or more layers of units between the input and output layers (DU & SWAMY, 2006; AGIRRE-BASURKO et al., 2006; DANANDEH & NOURANI, 2017).

The output of a neuron is given by:

$$\xi = \sum_{i=1}^{n} w_i x_i - b = w^T x - b$$

$$y = \sigma(\xi)$$

whereis the ith input, w_i is the link weight from the ith input, w = $(w_i ... w_n)^T$, x = $(x_1 ... x_n)^T$, b is a threshold or bias, and is the number of inputs. The activation function is usually some continuous or discontinuous function mapping the real numbers into the interval (-1,1) or (0,1) (DU & SWAMY, 2006). A different function can be used as an activation function but the most used is the sigmoidal activation function. A standard sigmoidal activation function has the following form:

$$\sigma(\xi) = \frac{1}{1 + e^{\Box \xi}}$$

The success of the method can be laid down for the following reasons: (1) They can model extremely complex systems and due to their nature can be used to model nonlinear natural systems (linearity in the sense of mathematical properties of additivity and homogeneity); when using linear algebra (such as most of the multivariate statistics) to describe nonlinear systems we always have to make approximations; (2) There is no limitation with the dimensionality of the problem; it can be arbitrary, depending on the CPU speed and memory, and (3) simple and successful operation due to well-developed learning algorithms (ŽIBRET et al., 2012).

3.5.3. Multiple polynomial regressions

Multiple regression, the term first used by Pearson (PEARSON, 1908 in KELECHI, 2012). refers to a regression application the main purpose of which is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable. Polynomial regression is a special case of multiple regression in which the relationship between the in-

dependent variable and the dependent variable is modelled as an mth order polynomial (OSTERTAGOVÁ, 2012). The general polynomial regression model is given by:

$$Y_i = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_m x^m + \varepsilon_i$$

where, for a set of I observations, Y_i is the predictand variable, β_0 is a regression constant, $\beta_1, \beta_2, \ldots, \beta m$ are the coefficients of the x, $x^2, \ldots, -x^m$ independent variables (predictors), m is the order of the polynomial, and ε_I is the model error (the difference between observations and predicted values).

Although polynomial regression fits a nonlinear model to the data, the estimation problem is linear because the regression coefficients are linear in the parameters (KERROU et al., 2017). The goal of the regression analysis is to determine the values of the parameters of the regression equation and then to quantify the goodness of the fit in respect of the dependent variable (AGIRRE-BASURKO et al., 2006). Two following values: R^2 and adjusted R^2 describe how well the model fits the data. All regression models provide information regarding the influence of the combined interactions of the estimator variables on the response. However, the major conceptual limitation of the regression techniques is that one can only ascertain relationships, but never be sure about the underlying causal mechanism.

4. RESULTS AND DISCUSSION

4.1. Linear mathematical methods vs. artificial neural networks (ANN-MP)

The collection of geospatial data has already begun with data acquisition. From the land use map and Digital Elevation Model (DEM), many geospatial data had been sourced and later used in modeling. The main purpose of using such parameters (data) is the construction of a spatial distribution map of particular elements as a final product. Geochemical maps, as final products are necessary for understanding both natural and anthropogenic distributions.

Due to the high cost and time-consuming nature of soil sampling, research in developing methods for the creation of soil maps from sparse soil data is becoming increasingly important. In recent years, the development of prediction methods (linear and nonlinear) that use secondary attributes sourced from the DEM, land use, and remote sensing in combination with sparse and expensive soil measurements has been sharpening the focus of this research. Consequently, the potential for applying such information to soil mapping is greater than ever before. Applying various modeling techniques, different prediction methods for soil types were compared, but we also choose the best combination of prediction method and secondary information. Various modeling techniques help us reconstruct different processes that influenced the entire study area simultaneously. Their main purpose is not only the isolation of hotspots with the highest concentrations but also a spatial distribution pattern of particular trace elements. Simultaneously they distinguish natural and anthropogenic influences as well as the transportation pattern, such as atmospheric or water transport of trace elements. Studying this pattern will help in better interpretation and understanding of the processes that occurred in whatever particular period of time, they are related to.

All the aforementioned data were used for preparing two spatial distribution models using two prediction methods, modeling by Multiple Polynomial Regression (MPR) and Artificial Neural Network - Multilayer Perceptron (ANN-MP), respec-

tively. For both modeling methods, a recall grid has been used. The whole study area was divided into 50 x 50 m grids. The total number of recall points is 41,471. Besides the standard position parameters, each particular recall point is also described by some new geospatial parameters. Both methods were treated using the same conditions and same software packages, Statistica 11 (Stat Soft Inc., 2012) and Surfer 11 (Golden Software Inc., 2012). The Kriging is a linear method and only considers the sparse measured data.

Modeling by ANN-MP was undertaken using a large amount of input data; 240 of the hidden units and 25 train networks. Aiming for the best model, we tried changing some input data; the number of neurons, as well as the number of training networks. Our experience showed that more neurons and more neural architectures produced better results. Models were constructed first for each particular element, then for groups of elements extracted by factor analysis. Each particular model is trained to 25 networks but only 5 logical networks were retained, and finally, an average model of 5 retained networks was calculated. Each training model contains a summary table with the following parameters: Training perfection, Test perfection, Validation perfection, All perfection, Training error, Test error, Validation error, Training algorithm, Hidden activation, and Output activation. Many data conversions, corrections and transformations had to be done prior to obtaining useful models. The main reason that we used so many different input data in modeling by MPR, lies in the fact that this method is very demanding compared to the ANN-MP.

The developed methods and procedures, especially ANN-MP, MPR as a control method and Kriging as a modified, worldwide successfully established method, have been used to predict the concentration of four leading elements Pb, Ni, Ti, and As from their previously identified geochemical association. Lead is a typical anthropogenic element, introduced solely by mining and smelting activities. Nickel and titanium are typical natural elements, whose enrichment depends mainly on the parental material degradation, and the vague distributiom of arsenic can be probably explained with its enrichment in independent outcropping rocks (AIJAGIĆ et al., 2015). Further processing of the factor score values is avoided because they represent the new synthetic variables which are calculated and presented as the standard values. This means that those methods can only determine relative enrichment relationships. In any case, with any certainty, we cannot judge the prediction of the absolute values.

The essence of the whole approach is that under the same conditions all three tested methods should give some stable distribution of results of the areas with anthropogenically enriched elements as well as examining natural enrichment. The expected results must be stable in all cases. Procedures, approaches, preparing data and calculations were being developed over time until they become stable. The stable procedures are procedures under the same conditions that correctly predict the distribution of elements that represent the geochemical characteristics of the landscape. The estimation of reliability of the predictions and model applicability lies in the fact that all the given models are repeatable. So the developed procedures are not random or specific to a particular item or an isolated area, but such models can be applied anywhere, even in the distribution of some other observations. We critically evaluated the methods according to the significance of the applied transformations, the similarity between models, as well as the stability of predictions. Basically, this represents the biggest contribution of this paper.
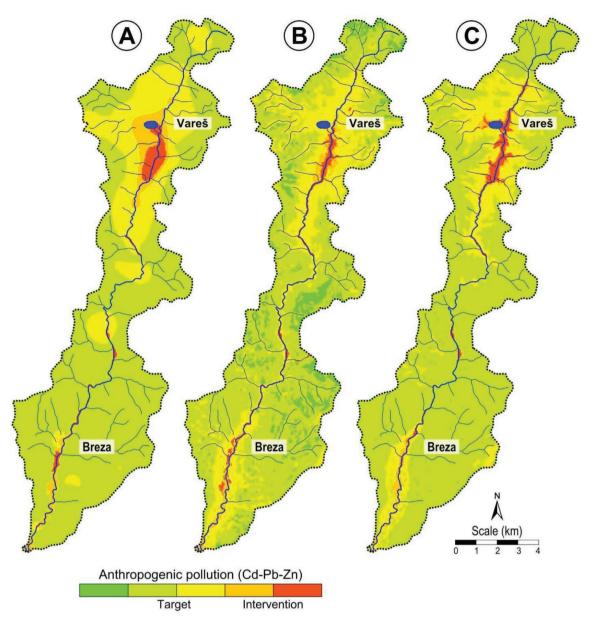
Figure 9. Spatial distribution of Cd–Pb–Zn pollution according to the NewDutch list (ESDAT, 2013) recommendation: A – Kriging; B – Multiple polynomial regression (MPR); C – Artificial neural network – Multilayer perceptron (ANN-MP).

Since computer equipment allows us to prepare a very large number of solutions in real-time, we used the principle which was used for knotting the Enigma coding system (the German coding system in the Second World War). The principle is very simple. Discard illogical solutions and focus on those that are logical. The basic principle (*principle of Ockham cat*) was applied: everything that happens around us must be logical and indeed simple.

4.2. Contamination and natural enrichment of chemical elements (application of models)

4.2.1. Anthropogenic impact

The spatial distributions of three common anthropogenic chemical elements Cd, Pb, and Zn according to the NewDutch list (ESDAT, 2013) are presented in Fig. 9. For a graphical display of spatial distribution, the maps with target and intervention values have been used, where different colours represent different concentration arrangements. In the scale used for its distribution we used five ranges: two green colours represent the ranges under the target values; the light green is one half of the target values;

the yellow range represents the target values, the orange is one half of the sum of both, target and intervention, and the red range represents the intervention values.

According to the NewDutch list (ESDAT, 2013), the target value for Cd is 0.8 mg/kg, for Pb 85 mg/kg, and for Zn 140 mg/kg. The intervention value for Cd is 12 mg/kg, for Pb 530 mg/kg, and for Zn 720 mg/kg. High concentrations of these three chemical elements exceed their intervention values only in Zone 1 and in alluvial soils. Several samples with their high concentrations are collected from these two units with maximum concentrations in the range of 4.0 – 7.2 mg/kg Cd, 880 – 1,700 mg/kg Pb and 1,500 – 3,100 mg/kg Zn.

The maximum concentration of Cd is 7.2 mg/kg, Pb 1,700 mg/kg, and Zn 3,100 mg/kg. If we compare these values to their target values, they exceed 9 times the maximum for Cd, 20 times for Pb, and 22 times for Zn, respectively. Even the intervention values are exceeded by more than 3 times for Pb, and 4 times for Zn. Almost the entire area is not contaminated with the aforementioned
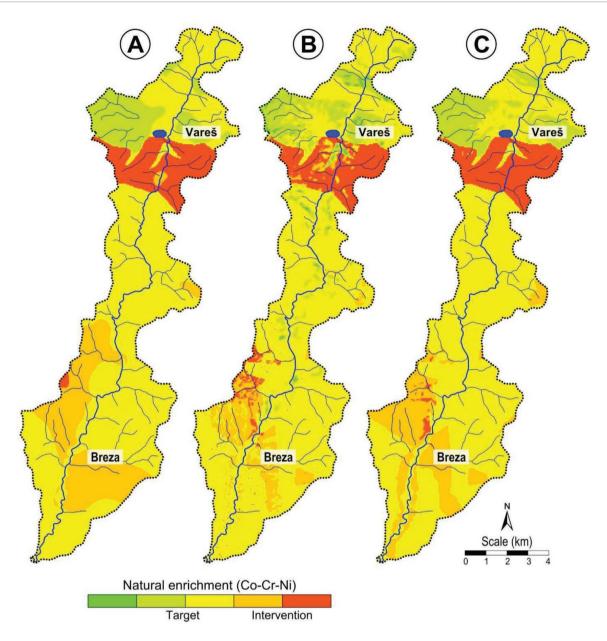
Figure 10. The spatial distribution of natural enrichment Co–Cr–Ni according to the NewDutch list (ESDAT, 2013) recommendation: A – Kriging; B – Multiple polynomial regression (MPR); C – Artificial neural network – Multilayer perceptron (ANN-MP).

group, except for the two isolated units, alluvial soils and the industrial zone 1?

Fine-grained clay-sized particles are transported far from the source of contamination and have been detected by the three applied methods. The concentrations under the target values are detected at 73 km² or 70 % of the entire study area with Kriging, 70 km² or 68 % with MPR and 77 km² or 74.5 % with ANN-MP. The first model extracts about 28 km² or 27 % between the target values and intervention values (yellow and orange colour), the second model about 31km² or 30 %, and the third model about 24 km² or 23 %. The range that represents concentrations higher than the intervention values are isolated on about 2.5 km² or 2.5 % of the study area with Kriging and ANN-MP and 2 km² or 2 % with MPR (Table 2).

4.2.2. Natural enrichment

The spatial distribution of natural enrichment is presented in the summarized map of Co, Ni, and Cr (Fig. 10). These prediction models show an arrangement in natural enrichment across the study area. For a graphical display of the spatial distribution, the maps with target and intervention values are presented, where different colours represent changes within them. The used colour scale is the same as for the anthropogenic impact (Fig. 9).

According to the NewDutch list (ESDAT, 2013), the target value for Ni is 35 mg/kg, Co 20 mg/kg, and Cr 100 mg/kg, but the intervention value for Ni is 210 mg/kg, Co 240 mg/kg, and Cr 380 mg/kg. The maximum concentration of Ni is 500 mg/kg, Co 64 mg/kg, and Cr 460 mg/kg. Comparing these values to the target values of the NewDutch list (ESDAT, 2013), natural enrichment exceeded 14 times for Ni, 3 times for Co, and 46 times for Cr. Compared to the intervention values, they exceed the values by 2.5 times for Ni and for Cr are more than the same again? It seems that the major enrichment is from Ni and Cr.

All three models identify one major hotspot with its intervention concentration, the Jurassic and Cretaceous flysch, but also some outcropping rocks on the Oligocene clastite complex. The first model isolates about 9 km² or 9 % of the total study area
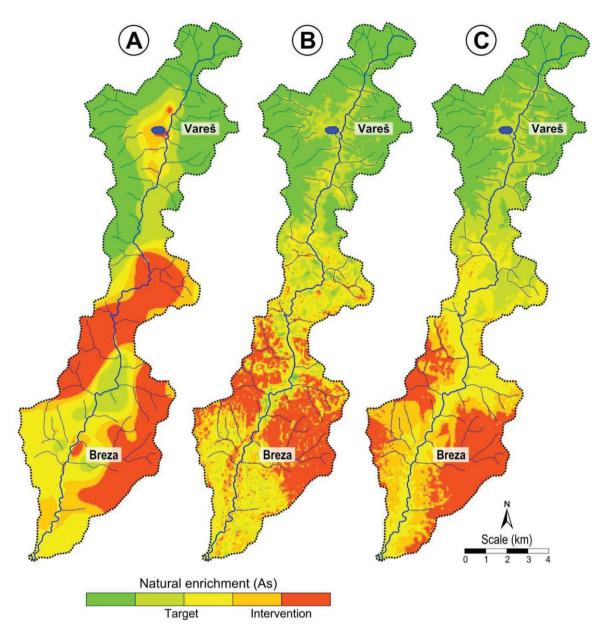
Alijagić and Šajn: Application and improvement of soil spatial distribution mapping using advanced modeling techniques 81

Geologia Croatica

Figure 11. The spatial distribution of Arsenic according to the NewDutch list (ESDAT, 2013) recommendation: **A** – Kriging; **B** – Multiple polynomial regression (MPR); **C** – Artificial neural network – Multilayer perceptron (ANN-MP).

($103.7 \ km^2$), under the range of target values (two green colours), the second model identifies about $11.2 \ km^2$ or 11 %, and the last model about $8.6 \ km^2$ or 8 %. Almost all the study area is enriched with these elements (yellow and orange colours), where all three applied models isolated the majority of the territory: Kriging and ANN-MP about $85 \ km^2$ or 82 % and MPR $82 \ km^2$ or 79 % (Table 2). Intervention values are observed at $10 \ km^2$ or 10 % of the entire study area with all applied models.

The spatial distribution of arsenic according to the NewDutch list (ESDAT, 2013) within all prediction methods is provided in Fig. 11. The target value for As is 29 mg/kg and the intervention value is 55 mg/kg. Comparing the values for As with the NewDutch list (ESDAT, 2013), its concentrations exceed the target values at $35 \ km^2$ with Kriging, $38 \ km^2$ with MPR and $41 \ km^2$ with ANN-MP (Table 2). For area between 18-$26 \ km^2$ its concentrations exceed the intervention value for As. Comparing the surfaces of all the selected elements (Pb, Zn, Cr, Ni) that includes major anthropogenic impact and natural enrichment, the natural

enrichment with arsenic is the greatest. This enrichment has been absolutely unexpected and needs special attention in future investigations.

Total contamination (As-Cd-Co-Cr-Ni-Pb-Zn) that includes anthropogenic impact and geogenic enrichment, respectively exceeds their target values over two-thirds of the entire study area, and one-third of their values exceed the intervention values. The total contamination is not a simple summary of these two types of contamination but it is necessary to mention that some spatial distributions are overlapping, and those final values seem lower than the summary of their particular distributions.

5. CONCLUSION

Along the Stavnja river valley, intensive mining and smelting activities have been occurring for more than 100 years. Diverse mineral occurrences, especially iron deposits and lead-zinc deposits make this study area interesting for geochemical research. Also, this study area was a place of intensive military operation

Geologia Croatica

Table 2. Areas of natural enrichment and pollution according to the NewDutch list (ESDAT, 2013) recommendation.

Anthropogenic origin (Cd-Pb-Zn)	Target level (km^2)	Intervention level (km^2)
Kriging	28	2.5
MPR	31	2.2
ANN-MP	24	2.6
Natural origin (Co-Cr-Ni)		
Kriging	85	9.7
MPR	82	10
ANN-MP	85	10
Natural origin (As)		
Kriging	35	26
MPR	38	20
ANN-MP	41	18
Total (As-Cd-Co-Cr-Ni-Pb-Zn)		
Kriging	64	37
MPR	67	30
ANN-MP	68	30

during the last war, 1992–1995, which resulted in numerous remaining minefields, thus making this research more complex.

The spatial distributions of particular trace elements are aiding the reconstruction of their main pathway in the study area and simultaneously isolating their hotspots. The problem of contamination along the river represents an anisotropic appearance between the isotropic lithological units and cannot be solved completely by standard interpolation kriging methods, based only on sparse soil measurements. Two new powerful linear and nonlinear modeling techniques are applied for solving this problem. Their arrangement in concentration across the Stavnja river valley is more realistic and picturesque because they include more geospatial and geomorphological data such as geological background, land use, aspect, slope, altitude, etc. Including all the aforementioned facts, results in better interpretation and understanding of the processes that occurred during the particular period of time that they are related to. Various modeling techniques help to reconstruct different processes that affected the entire area. Their main purpose is not only the isolation of hotspots with the highest concentrations. They simultaneously distinguish natural from anthropogenic influences as well as the transportation pattern (such as atmospheric or water transport).

Summarised maps that comprise solely three anthropogenic trace elements (Pb-Zn-Cd) and three natural trace elements (Ni-Co-Cr) according to the NewDutch list (ESDAT, 2013) provide information about the overall anthropogenic impact and natural enrichment of study area. The results show:

1. All provided models indicate that natural enrichment is much higher than the anthropogenic impact. About 71 % or 73 km^2 is under target values, in the range between target and intervention values about 26.5 % or 26 km^2, and over the intervention values by only about 2-2.5 % or 2-2.5 km^2 of entire study area for Pb-Zn-Cd.

2. Natural enrichment (Ni-Co-Cr) shows the opposite situation, whereby only about 9 % or 10 km^2 is under target values, about 84 km^2 or 80 % between intervention and target values, and 10 % or 10 km^2 over intervention values of the study area.

3. The three applied models are repeatable, which means that very similar spatial distribution can be obtained under the same

conditions, with an unrestricted number of applications. The very important characteristic of each particular model is that one model isolates all the hotspots simultaneously. On the other hand, their similarities are confirmation and validation of all the different models at the same time. An important success in such modeling is that all models are repeatable.

4. Natural enrichment with arsenic is the highest result. This enrichment has been absolutely unexpected and needs special attention in future investigations.

REFERENCES

ASCE (2000): Committee on the application of ANNs in hydrology artifical neural networks in hydrology I: preliminary concepts.– Journal of Hydrological Engineering, 5/2, 115–123.

AGIRRE-BASURKO, E., IBARRA-BERASTEGI, G. & MADARIAGA, I. (2006): Regression and multilayer perceptron-based models to forecast hourly O_3 and NO_2 levels in the Bilbao area.– Environmental Modeling & Software, 21, 430–446. doi: 10.1016/j.envsoft.2004.07.008

AITKENHEAD, M.J., COULL, M.C., TOWERS, W., HUDSON, G. & BLACK, H.I.J. (2012): Predicting soil chemical composition and other soil parameters from field observations using a neural network.– Computers and Electronics in Agriculture, 82, 108–116. doi:10.1016/j.compag.2011.12.013

ALIJAGIĆ, J. & ŠAJN, R. (2010): Distribution of chemical elements in an old metallurgical area, Zenica (Bosnia and Herzegovina).– Geoderma, 162/1–2, 71–85. doi: 10.1016/j.geoderma.2011.01.007

ALIJAGIĆ, J. & ŠAJN, R. (2006): Influence of ironworks on distribution of chemical elements in Bosnia and Herzegovina and Slovenia.– Geologija, 49/1, 123–132. doi:10.5474/geologija.2006.010

ALIJAGIĆ, J., ŠAJN, R. & ROKAVEC, D. (2015): Impact assessment of mining and metallurgical activities on the distribution of trace elements in the Stavnja valley, Bosnia and Herzegovina.– In: DANIELS, J.A. (ed.): Advances in environmental research – Nova Science Publisher, 42, 139–200.

ALIJAGIĆ, J. (2013): Application of multivariate statistical methods and artificial neural network for separation natural background and influence of mining and metallurgy activities on distribution of chemical elements in the Stavnja Valley (Bosnia and Herzegovina).– PhD Dissertation, University of Nova Gorica, 110 p.

ALIJAGIĆ, J. (2008): Distribution of chemical elements in an old Metallurgic area, Zenica (Central Bosnia).– Master dissertation, Masaryk University of Brno, 100 p.

AL-KHASHMAN, O.A. & SHAWABKEH, R.A. (2009): Metal distribution in urban soil around steel industry beside Queen Alia Airport, Jordan.– Environmental Geochemistry and Health, 31/6, 717–726. doi:10.1007/s10653-009-9250-9

AUTODESK, Inc. (2012): Autodesk MAP 3D, Version 2012 – Software. Autodesk, Inc. (http://www. autodesk.com)

BAGHERI BODAGHABADI, M., MARTÍNEZ-CASASNOVAS, J., SALEHI, M.H., MOHAMMADI, J., ESFANDIARPOOR BORUJENI, I., TOOMANIAN, N. & GANDOMKAR, A. (2015): Digital Soil Mapping Using Artificial Neural Networks and Terrain-Related Attributes.– Pedosphere, 25/4, 580–591. doi: 10.1016/S1002-0160(15)30038-2

BOX, G.E.P. & COX, D.R. (1962) An analysis of transformations.– Journal of the Royal Statistical Society, Series B, 26/2, 211–252. doi: 10.1111/j.2517-6161.1964.tb00553.x

BRADL, H.B. (2005): Sources and Origins of Heavy Metals.– Heavy Metals in the Environment, 1–27. doi: 10.1016/s1573-4285(05)80020-1

BURROUGH, P.A. (1986): Principles of Geographic Information Systems for Land Resources Assessment.– Oxford: Oxford University Press, 193 p.

CARLON, C., CRITTO, A., MARCOMINI, A. & NATHANAIL, P. (2001): Risk based characterisation of contaminated industrial site using multivariate and geostatistical tools.– Environmental Pollution, 111/3, 417–427. doi: 10.1016/S0269-7491(00)00089-0.

CHOPIN, E.I.B. & ALLOWAY, B.J. (2007): Distribution and mobility of trace elements in soils and vegetation around the mining and smelting areas of Tharsis, Ríotinto and Huelva, Iberian Pyrite Belt, SW Spain. Water, Air, and Soil Pollution, 182/1–4, 245–261. doi: 10.1016/j.scitotenv.2006.11.037

DANANDEH, M.A. & NOURANI, V. (2017): A Pareto-optimal moving average-multigene genetic programming model for rainfall-runoff modeling.– Environmental Modeling & Software, 92, 239–251. doi: 10.1016/j.envsoft.2017.03.004

DU, K.L. & SWAMY, M.N.S. (2006): Neural Networks in an Softcomputing Framework. Springer – Verlag, London, 566 p.

DUDKA, S. & ADRIANO, D.C. (1997): Environmental Impacts of Metal Ore Mining and Processing: A Review.– Journal of Environmental Quality, 26, 590–602. doi: 10.2134/jeq1997.00472425002600030003x

ESDAT (2013): Dutch target and intervention values, 2000 (the NewDutch list). Ministry of Infrastructure and the Environment. URL: http://www.esdat.com.au/.

ESRI, Inc. (2004): ArcINFO ver 9, Software.– Environmental research institute (http://www.esri.com/)

Alijagić and Šajn: Application and improvement of soil spatial distribution mapping using advanced modeling techniques   83

Geologia Croatica

FERREIRA, V., KORICHEVA, J., DUARTE, S., NIYOGI, D.K. & GUÉROLD, F. (2016): Effects of anthropogenic heavy metal contamination on litter decomposition in streams – A meta-analysis.– Environmental Pollution, 210, 261–270. doi: 10.1016/j.envpol.2015.12.060

FISHER, P.F. & TATE, N.J. (2006): Causes and consequences of error in digital elevation models.– Progress in Physical Geography, 30/4, 467–489. doi:10.1191/0309133306pp492ra

GOLDEN Software, Inc. (2012): Surfer (surface mapping system), ver 11 – Software.– Golden Software, Inc. (http://www.goldensoftware.com/)

GOOGLE Inc. (2010). Google Earth. Area Breza - Vareš (http://earth.google.com)

GOOVAERTS, P. (2001): Geostatistical modeling of uncertainty in soil science.– Geoderma, 103/1–2, 3–26. doi:10.1016/s0016-7061(01)00067-2

GOOVAERTS, P. (1999): Geostatistics in soil science: state of the art and perspectives.– Geoderma, 89/1, 1–45. doi:10.1016/s0016-7061(98)00078-0

GRINGARTEN, E. & DEUTSCH, C.V. (2001): Teacher's aide: variogram interpretation and modeling.– Mathematical Geology, 33/4, 507–534. doi:10.1023/a:1011093014141

GUTH, P.L. (2006): Geomorphometry from SRTM – comparison to NED. Photogrammetric Engineering and Remote Sensing, 72/3, 269–277. doi:10.14358/pers.72.3.269

HILL, L.L. (2006): Georeferencing: The Geographic Associations of Information, 2. Cambridge, MA: MIT Press, 280 p. doi: 10.7551/mitpress/3260.001.0001

HONGHAI, Q. & ALTINAKAR, M.S. (2011): A GIS-based decision support system for integrated flood management under uncertainty with two dimensional numerical simulations.– Environmental Modeling and Software, 26/6, 817–821. doi: 10.1016/j.envsoft.2010.11.006

JAIN, S.K. & SINGH, V.P. (2003): Water Resources Systems Planning and Management, 51, 858 p.

JOBSON, J.D. (1991): Applied Multivariate Data Analysis.– Regression and Experimental Design, Springer-Verlag, New York, 1, 662 p. doi: 10.1007/978-1-4612-0921-8

KABATA-PENDIAS, A. & PENDIAS, H. (2001): Trace elements in soils and plants. Third edition.– CRC Press LLC, Boca Raton, Florica, 403 p. doi: 10.1201/b10158

KELECHI, A.C. (2012): Regression and Principal Component Analyses: a Comparison Using Few Regressors.– American Journal of Mathematics and Statistics, 2/1, 1–5. doi: 10.5923/j.ajms. 20120201.01

KERROU, J., DEMAN, G., TACHER, L., BENABDERRAHMANE, H. & PERROCHET, P. (2017): Numerical and polynomial modeling to assess environmental and hydraulic impacts of the future geological radwaste repository in Meuse site (France).–Environmental Modeling & Software, 97, 157–170. doi: 10.1016/j.envsoft.2017.07.018

KRIGE, D.G. (1960): On the departure of ore value distributions from lognormal models in South African gold mines.– Journal of the South African Institute of Mining and Metallurgy, 61, 231–244.

KRIGE, D.G. (1951): A statistical approach to some basic mine valuation problems on the Witwatersrand.– Journal of the Chemical, Metallurgical and Mining Society of South Africa, 52/6, 119–139.

LAGACHERIE, P. & MCBRATNEY, A.B. (2007): Chapter 1. Spatial soil information systems and spatial soil inference systems: perspectives for Digital Soil Mapping.– In: LAGACHERIE, P., MCBRATNEY, A.B. & VOLTZ, M. (eds.): Digital Soil Mapping, an introductory perspective. Developments in soil science, Amsterdam, 31, 3–24. doi: 10.1016/S0166-2481(06)31001-X

LI, Y., ZHU, A.-X., SHI, Z., LIU, J. & DU, F. (2016): Supplemental sampling for digital soil mapping based on prediction uncertainty from both the feature domain and the spatial domain.– Geoderma, 284, 73–84. doi: 10.1016/j.geoderma.2016.08.013

MCBRATNEY, A.B., MENDONÇA SANTOS, M. & MINASNY, B. (2003): On digital soil mapping.– Geoderma, 117, 3–52. doi: 10.1016/S0016-7061(03)00223-4.

MCGRATH, D., ZHANG, C. & CARTON, O.T. (2004): Geostatistical analyses and hazard assessment on soil lead in Silvermines area, Ireland.– Environmental Pollution, 127/2, 239–248. doi:10.1016/j.envpol.2003.07.002

MITASOVA, H. & HOFIERKA, J. (1993): Interpolation by Regularized Spline with Tension: II. Application to Terrain Modeling and Surface Geometry Analysis.– Mathematical Geology, 25/6, 657–669. doi:10.1007/bf00893172

MOOR, I.D., LEWIS, A. & GALLANT, J.C. (1993): Terrain properties: Estimation Methods and Scale Effects.– In: JAKEMAN, A.J. (ed.): Modeling Change in Environmental Systems, John Wiley and Sons, New York, 189–214.

MUKHERJEE, S., JOSHI, P.K., MUKHERJEE, S., GHOSH, A., GARG, R.D. & MUKHOPADHYAY, A. (2013): Evaluation of vertical accuracy of open source Digital Elevation Model (DEM).– International Journal of Applied Earth Observation and Geoinformation, 21, 205–217. doi: 10.1016/j.jag.2012.09.004

NAVARRO, M.C., PÉREZ-SIRVENT, C., MARTÍNEZ-SÁNCHEZ, M.J., VIDAL, J., TOVAR, P.J. & BECH, J. (2008): Abandoned mine sites as a source of contamination by heavy metals: a case study in a semi-arid zone.– Journal of Geochemical Exploration, 96/4, 183–193. doi: 10.1016/j.gexplo.2007.04.011

OLIVER, M.A. & WEBSTER, R. (2014): A tutorial guide to geostatistics: Computing and modeling variograms and kriging.– Catena, 113, 56–69. doi:10.1016/j.catena.2013.09.006

OSTERTAGOVÁ, E. (2012): Modeling using polynomial regression.– Procedia Engineering, 48, 500–506. doi: 10.1016/j.proeng.2012.09.545

PALINKAŠ, L.A., ŠOŠTARIĆ, S.B. & PALINKAŠ, S.S. (2008): Metallogeny of the Northwestern and Central Dinarides and Southern Tisia.– Ore Geology Reviews, 34, 501–520. doi:10.1016/j.oregeorev.2008.05.006

PAVLOWSKY, R.T., LECCE, S.A., OWEN, M.R. & MARTIN, D.J. (2017): Legacy sediment, lead, and zinc storage in channel and floodplain deposits of the Big River, Old Lead Belt Mining District, Missouri, USA.– Geomorphology, 299, 54–75. doi:10.1016/j.geomorph.2017.08.042

PEARSON, H.W. (1908): The place of the great raised beaches in geology.– Science, new ser., 27, 189–190.

RAMLAL, B. & BABAN, S. (2008): Developing a GIS based integrated approach to flood management in Trinidad, West Indies.– Journal of Environmental Management, 88/4, 1131–1140. doi:10.1016/j.jenvman.2007.06.010

ROMIĆ, M. (2012): Bioavailability of trace metals in terrestrial environment: Methodological issues.– European Chemical Bulletin, 1/11, 489–493. doi: 10.17628/ecb.2012.1.489-493

ROSE, A.W., HAWKES, H.E. & WEBB, J.S. (1979): Geochemistry in mineral exploration, 2nd edition.– Academic Press, London, 657 p.

SCULL, P., FRANKLIN, J., CHADWICK, O.A. & MCARTHUR, D. (2003): Predictive soil mapping: A review.– Progress in Physical Geography, 27, 171–197. doi: 10.1191/0309133303pp366ra

SHALLARI, S., SCHWARTZ, C., HASKO, A. & MOREL, J.L. (1998): Heavy metals in soils and plants of serpentine and industrial sites of Albania.– Science of the Total Environment, 209, 133–142. doi: 10.1016/S0048-9697(98)80104-6

SINGH, K.P., BASANT, N., MALIK, A. & JAIN, G. (2010): Modeling the performance of "up-flow anaerobic sludge blanket" reactor based wastewater treatment plant using linear and nonlinear approaches – A case study.– Analytica Chimica Acta, 658/1, 1–11. doi: 10.1016/j.aca.2009.11.001

SPARK, R.N. & WILLIAMS, P.F. (1996): Digital Terrain Models and the visualization of structural geology.– Computer Methods in the Geosciences, 15, 421–446. doi: 10.1016/S1874-561X(96)80032-0

STAFILOV, T., ŠAJN, R., BOEV, B., CVETKOVIĆ, J. & MUKAETOV, D. (2008a): Geochemical atlas of Kavadarci and the environment – Sts. Cyril and Methodius University, Faculty of Natural Science, Skopje, 109 p.

STAFILOV, T., ŠAJN, R., PANČEVSKI, Z., BOEV, B., FRONTASYEVA, M.V. & STRELKOVA, L. (2008b): Geochemical atlas of Veles and the environs.– Sts. Cyril and Methodius University, Faculty of Natural Sciences and Matematics, Skopje, 123 p.

STATISTICA Stat Soft, Inc. (2012) (data analysis software system): version 11 – Software. Stat Soft, Inc. (www.statsoft.com)

ŠAJN R., HALAMIĆ, J., PEH, Z., GALOVIĆ, L. & ALIJAGIĆ, J. (2011): Assessment of the natural and anthropogenic sources of chemical elements in alluvial soils from the Drava River using multivariate statistical methods.– Journal of geochemical exploration, 110/ 3, 278–289. doi:10.1016/j.gexplo.2011.06.009

ŠAJN, R. (2006): Factor analysis of soil and attic-dust to separate mining and metallurgy influence, Meža Valley, Slovenia.– Mathematical Geology, 38/6, 735–747. doi: 10.1007/s11004-006-9039-7

ŠAJN, R. (2005): Using attic dust and soil for the separation of anthropogenic and geogenic elemental distributions in an old metallurgic area (Celje, Slovenia).– Geochemistry, Exploration Environment Analysis, 5/1, 59–67. doi: 10.1144/1467-7873/03-050

THE CGIAR Consortium for Spatial Information (2011): SRTM 90m Digital Elevation Data. Database: SRTM_40_04 (http://srtm.csi.cgiar.org)

THEODOSSIOU, N. & LATINOPOULOS, P. (2006): Evaluation and optimisation of groundwater observation networks using the Kriging methodology.– Environmental Modeling & Software, 21/7, 991–1000. doi: 10.1016/j.envsoft.2005.05.001

THOMPSON, J.A., BELL, J.C. & BUTLER, C.A. (2001): Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling.– Geoderma, 100, 67–89. doi:10.1016/s0016-7061(00)00081-1

THORNTON, P.E., RUNNING, S.W. & WHITE, M.A. (1997): Generating surface of daily meteorological variable over large regions of complex terrain.– Journal of Hydrology, 190/3, 214–250. doi:10.1016/s0022-1694(96)03128-9

U.S. GEOLOGICAL SURVEY (2011a): ASTER Global Digital Elevation Data (30 m). Database: 20130113055119_468788446 (http://gdex.cr.usgs.gov/gdex/)

U.S. GEOLOGICAL SURVEY (2011b): Global Land Survey multispectral landsat satelite images P187-R29. Databases: P188R029_5x19920914 and LE71880292007259ASN00 (http://earthexplorer. usgs.gov/)

U.S. GEOLOGICAL SURVEY (2011c): Landsat project description and Land Remote Sensing (http://landsat.usgs.gov/)

VASSILOPOULOUA, S., HURNI, L., DIETRICH, V., BALTSAVIAS, E., PATERAKI, M., LAGIOS, E. & PARCHARIDIS, I. (2002): Orthophoto generation using IKONOS imagery and high-resolution DEM: a case study on volcanic hazard monitoring of Nisyros Island (Greece).– ISPRS Journal of Photogrammetry and Remote Sensing, 57, 24–38. doi:10.1016/s0924-2716(02)00126-0

WEBSTER, R. & OLIVER, M.A. (2001): Geostatistics for environmental scientists.– John Wiley and Sons Ltd, Chichester, 271 p. doi: 10.1002/9780470517277

WECHSLER, S.P. (2003): Perceptions of Digital Elevation Model Uncertainty by DEM Users.– URISA Journal, 15/2, 57–64.

WISE, S.M. (1998): The effect of GIS interpolation errors on the use of digital elevation models in geomorphology.– In: LANE, S.N., RICHARDS, K.S. & CHANDLER, J.H. (eds.): Landform modeling, monitoring and analysis. John Wiley and Sons Ltd, Chichester, 466 p.

WOLOCK, D.M. & PRICE, C.V. (1994): Effects of digital elevation model map scale and data resolution on a topography-based watershed model.– Water Resources Research, 30/11, 3041–3052. doi:10.1029/94wr01971

YAALON, D.H. (1989): The Earliest Soil Maps and Their Logic. Bulletin of the International Society of Soil Science.– International Society of Soil Science, Wageningen p 24.

YI, Q., DOU, X.D., HUANG, Q.R. & ZHAO, X.Q. (2012): Pollution Characteristics of Pb, Zn, As, Cd in the Bijiang River.– Procedia Environmental Sciences, 13, 43–52. doi: 10.1016/j.proenv.2012.01.004

ZHANG, C.S. & SELINUS, O. (1998): Statistics and GIS in environmental geochemistry- some problems and solutions.– Journal of Geochemical Exploration, 64/1–3, 339–354. doi: 10.1016/S0375-6742(98)00048-X

ZHANG, C.S., SELINUS, O. & SCHEDIN, J. (1998): Statistical analyses on heavy metal contents in till and root samples in an area of southeastern Sweden.– The Science of the Total Environment, 212/2–3, 217–232. doi: 10.1016/S0048-9697(97)00341-0

ZHANG, C.S., ZHANG, S., ZHANG, L.C. & WANG, L.J. (1995): Background contents of heavy metals in sediments of the Changjiang River system and their calculation methods.– Journal of Environmental Sciences, 7/4, 422–429.

ZHANG, W. & MONTGOMERY, D.R. (1994): Digital elevation model grid size, landscape representation, and hydrologic simulations.– Water resources research, 30/4, 1019–1028. doi: 10.1029/93WR03553

ZHANG, L., LYMAN, S., MAO, H., LIN, C.J., GAY, D.A., WANG, S. & WANIA, F. (2017): A synthesis of research needs for improving the understanding of atmospheric mercury cycling – Atmospheric Chemistry and Physics Discussions, 1–29. doi: 10.5194/acp-2017-375

ZHAO, X., DAI, J. & WANG, J. (2013): GIS-based evaluation and spatial distribution characteristics of land degradation in Bijiang watershed.– SpringerPlus, 2, S8. doi: 10.1186/2193-1801-2-S1-S8

ŽIBRET, G., ŠAJN, R., ALIJAGIĆ, J. & STAFILOV, T. (2012): Use of neural networks in the geochemical data interpretation.– Zeitschrift für Geologische Wissenschaften, 40/4–5, 253–266.