CRITICAL REVIEW OF THE ENVIRONMENTAL INVESTIGATION ON SOIL HEAVY METAL CONTAMINATION

ERSOY, A.

*Department of Mining Engineering, Adana Alparslan Turkes Science and Technology University, 01250 Sarıcam, Adana, Turkey
(e-mail: aersoy@atu.edu.tr; phone: +90-322-444-0188/2001)*

Abstract. Soil contamination by heavy metals has become a severe environmental issue in the world due to rapid development of urbanisation, industrial, mining, agricultural and natural processes, and chemical compounds. Reliable and quality results quantify the adverse effects of these factors. A precise and cost-efficient study depends on adequate background research, a well-planned sampling design and strategy, quality data, appropriate selection and implementation of analytical techniques and investigation. The investigation methods for heavy metal soil or land contaminations drive decision making and remediation which is very expensive. Therefore, this study offers comprehensive and comparative review on data organisation and treatment; guidelines, legislation of heavy metals; data analysis and investigation methods. The primary objectives of the review are to discuss the various stages involved in the investigation of heavy metals/land for site engineers and environmental scientists. Data analysis methods include exploring contamination indices, statistical and multivariate statistical analysis methods, interpolation techniques, geostatistical estimation, simulation, and combined methods. Strengths, weaknesses and the application scopes of these methods and the resulting models used are critical for success in environmental modelling.
Keywords: *data analysis methods, contamination indices, multivariate analysis, geostatistical simulation, and spatial interpolation methods*

Introduction

Heavy metal contamination in soil or land has become an increasingly common and serious problem and threat to every country of the world due to rapid development of technology, economy, public awareness, and society. In Europe, mineral oil and heavy metals are the main contaminants contributing by 50% to soil contamination. The management of contaminated sites is estimated to cost around six billion Euros annually (Panagos et al., 2013). In the past, soil contamination was not considered as important as air and water pollution, because it was often with wide range and was more difficult to be controlled and governed than air and water pollution. However, in recent years the soil contamination in developed countries has become an essential issue thus, more and more attention is paid to this issue which has become a significant topic of environmental protection worldwide (Su et al., 2014). Characteristics of heavy metal contamination of soils include wide distributions, strong latency, irreversibility, remediation hardness, high cost and complex heavy metal contamination. In the world's top ten environmental events, two events have been related to heavy metal contamination (Yang and Sun, 2009). These are soil, air, and water pollutions, caused by heavy metals, which are a serious threat to almost every country. Heavy metals have been effectively used by humans for thousands of years. Although, several adverse health effects of heavy metals have been known for a long time, exposure to heavy metals continues and is even increasing in some parts of the world, particularly in less developed countries (Jarup, 2013).

Heavy metals constitute an ill-defined group of inorganic chemical hazards, and at contaminated sites Pb, Cr, As, Zn, Cd, Cu, Hg and Ni are found most commonly. Source of heavy metal contamination in soil may be classified into two categories, natural and anthropogenic. The spatial distribution of naturally originating heavy metals is highly heterogeneous and different concentrations exist in different soils. Heavy metals have been used for thousands of years in a large variety of industrial products, which have been deposited for a long time as waste. The main anthropogenic sources of heavy metals are agricultural activities, metallurgical activities, mining operations, energy production, transportation, micro-electronic products, and waste disposal. They are found in different forms such as gaseous, particulate, aerosol/aqueous solid and emanate both diffuse and point sources. Literature research showed that exposure of human health from contaminated soil by heavy metals is directly implicated as ingestion, inhalation, skin contact and dermal absorption. Human health is also indirectly affected through contaminating the food, water, and atmosphere. Different contaminants originate different negative effects on human health and environment depending on their properties. These are dispersion solubility in water, bioavailability, carcinogenicity, bioaccumulation, and so on.

Sampling efficiency and representatives in soil contamination by heavy metals are affected by various factors which include sampling design and strategy, sampling location, depth, density, sampling stages and methods. These factors have been widely studied by many researchers (e.g. Coşkun et al., 2006; Davis et al., 2009; Maas et al., 2010; Sun et al., 2010; Wang and Lu, 2011; Lu et al., 2012; Shan et al., 2013; Kelepertzis, 2014; Haung et al., 2015; Mihailovic et al., 2015; Zhou et al., 2016; Moore et al., 2016). The factors optimisation and economic cost of soil sampling are typically analysed by geostatistical techniques and Geographic Information System (GIS) integrated multivariate statistical methods.

Toxic levels of heavy metals may be various in different countries because of different cultures and different protection methods of environment and commonly health. Thus, a large variation in environmental and human health regulations and their effects for heavy metal contaminants in soil may be observed trough the world. Regulations in the developed countries for soil contamination with heavy metals may be guidance and useful to investigate risk assessment and decision-making for the developing countries. Therefore, total concentration levels of heavy metal contaminants for soil quality guidelines and the protection of environmental health in the United Kingdom, the European Community, the Netherlands, Canada, and Australia are presented in *Tables 1-4,* respectively. There are also many environmental laws and regulations for metal contaminants in soil around the world including USA, Germany, Japan, China, Singapore, and Malaysia. For example, there are several federal and state sets of regulations and standards in the United States of America (USA). The most widely recognised methodology for risk assessment of an environmental contaminant developed by US Environment Protection Agency (USEPA, 2011). A wide variation in standards, regulations and their effects were observed throughout the world. In summary, most current legislations are still based on the total concentrations of contaminants in soil. Consequently, these regulations and limits may act as a guideline to purpose risk assessment methodologies, model tools, and exposure scenarios, especially for the developing or less developed countries.

Table 1. *Heavy metal guideline in soil contaminated land exposure assessment (CLEA, 2009)*

Heavy metals	Function of land use	CLEA soil guideline value (mg/kg)
Pb	Residential with home grown produce	200
	Residential without home grown produce	310
	Allotments	80
	Commercial	2300
Cr VI	Residential with home grown produce	21
	Residential without home grown produce	21
	Allotments	170
	Commercial	49
Cr	Residential with plant uptake	130
	Residential without plant uptake	200
	Commercial and industrial	5000
As	Residential with home grown produce	37
	Residential without home grown produce	40
	Allotment	49
	Commercial	640
Cd	Residential with home grown produce	22
	Residential without home grown produce	150
	Allotment	3.9
	Commercial	410
Hg	Residential	10
	Allotment	26
	Commercial	26
Ni	Residential	130
	Allotment	230
	Commercial	1800

Table 2. *Heavy metal soil and sediment guideline values in the Netherlands (The Ministry of Housing, 2011)*

Heavy metals	Target value (mg/kg)	Intervention value (mg/kg)
Pb	85	530
Cr	100	380
As	29	55
Zn	140	720
Cd	0.8	12
Cu	36	190
Hg	0.3	10
Ni	35	210

The key to effective quality assessment of soil contamination by heavy metals is in the use of investigation methods. There is currently a wide arrange of investigation methods used to evaluate soil contamination. A discussion of the advantages and limitations of different soil contamination assessment methods such as contamination indices, statistical analysis, spatial interpolation techniques, geostatistical methods and combined methods is presented. Contamination indices are the most widely used significant tools for the comprehensive evaluation and the grade of soil contamination. Many authors previously described several indices which are defined for evaluation of the degree of soil

contamination in recent publications (Wu et al., 2014; Kovalska et al., 2018). In this study, different aspects, and significant characteristics of the indices such as the similarities and differences, comparisons, advantages and disadvantages were briefly evaluated. This ensures the selection of appropriate indices in the environmental study of different soils.

Table 3. *Soil and quality guidelines for the protection of environmental health values in Canada for land use (Canadian Council of Ministers of the Environment, 2010)*

Heavy metals	Agricultural (mg/kg)	Residential/park land (mg/kg)	Commercial (mg/kg)	Industrial (mg/kg)
Pb	70	140	260	600
Cr	64	64	87	87
As	12	12	12	12
Zn	200	200	360	360
Cd	1.4	10	22	22
Cu	63	63	91	91
Hg	6.6	6.6	24	50
Ni	50	50	50	50

Table 4. *Heavy metal levels in soil in Australia (Department of Environment and Conservation, 2010)*

Heavy metals	Ecological level (mg/kg)	Residential/garden (mg/kg)	Residential/apartments/flats minimum soil access (mg/kg)	Parks/recreational/playing fields area (mg/kg)	Commercial/industrial (mg/kg)
Pb	600	300	1200	600	1500
Cr III	400	120000	48000	240000	60000
Cr VI	1	100	4000	200	500
As	20	100	400	200	500
Zn	200	7000	28000	4000	35000
Cd	3	20	80	40	100
Cu	100	1000	4000	2000	5000
Hg	1	15	60	30	75
Ni	60	600	2400	600	3000

Soil contamination prediction requires frequent use of statistics. Statistical analysis has been used for a long time to address soil contamination as univariate or classical statistics and multivariate statistical analysis. Univariate statistical tools present several facilities including improving understanding of data and soil contamination, providing data quality, organising, and grouping data information, and making inferences and estimations. Multivariate statistical analysis was not alone widely used in environmental studies. However, recently, the use of multivariate statistical analysis integrated with GIS have been successfully studied in the identification of metal sources, assessment of metal behaviour, soil quality, mapping of metal spatial distribution in regions (Saby et al., 2009; Lu et al., 2012; Shao et al., 2014; Haung et al., 2015; Zhou et al., 2016; Ali et al., 2016; Moore et al., 2016; Gabarron et al., 2017). GIS is a system designed to capture, store, manipulate, analyse, manage, and present all types of geographical data (ESRI, 1994). GIS is increasingly used as the most comprehensive tools for life and industry including mapping, environmental impact analyses, geological and mining studies, hydrology, archaeology, rural and urban planning, disaster management and mitigation, crime statistics, health and medical resource, management, transportation planning, agricultural

applications, climate and meteorology, telecom and network services, and many other areas. GIS provides spatial data entry, management, and retrieval, analysis, and visual functions.

Geostatistics contains different methods based on regionalised theory and stationary for the analysis, estimation and simulation of data correlated in space or time. Geostatistics was initially developed for mineral resource estimation and geological modelling (David, 1977; Isaaks and Srivastava, 1989; Goovaerts, 1997; Rossi and Deutsch, 2014), and later enhanced for spatial analysis of environmental issues (Burges and Webster, 1980; Goovaerts, 1999; Webster and Oliver, 2007; Oliver, 2010). A major aspect of geostatistical modelling is to quantitatively measure spatial variability by subsequent estimation and simulation.

Traditional interpolation or Inverse Distance Weighting (IDW) and geostatistical interpolation or Kriging methods have been increasingly used to estimate the spatial distribution of contaminants in soil for 1990's years (Zhang et al., 1995; Steiger et al., 1996; Journel, 1998; Meirvenne and Goovaerts, 2001). However, these methods have smoothing effects, results in less variance in the estimation than in the observed data. Recently, on the other hand, geostatistical simulations, the most commonly used Sequential Gaussian Simulation (SGS), have overcome the limitations intrinsic in conventional and kriging-based interpolation techniques. SGS reproduce original statistics, histograms and variograms of the spatial variability for the data without smoothing effects. SGS are the most frequently applied in mining industry and environmental studies (Goovaerts al., 1996; Soares, 2001; Meirvenne and Goovaerts, 2001; Pereira et al., 2001; Franco et al., 2006; Ersoy et al., 2008; De Almedia, 2010; Qu et al., 2013; Rossi and Deutsch, 2014; Garcia-Lorezo et al., 2014; Albuquerque et al., 2017; Zhang et al., 2017; Ersoy and Yünsel, 2018).

This review fills a knowledge gap in soil contamination by heavy metals. There are many review publications available that describe only single issue and provide few or no guidelines necessary focusing on practical applications of the environmental research. Descriptions, comparisons, advantages and disadvantages and integrations of the investigation methods or models for soil contamination by heavy metals to use in the spatial distribution, risk analysis and decision making are presented. The paper outlines soil guideline limits and characteristics of heavy metals, establishes investigation and application methods to soil contamination, explores data organisation and treatment and factor affecting the performance of the application methods, demonstrates validation test of estimation and simulation. The workflow approach of the review for contaminated site characterisation is presented in *Figure 1*. This study is presented to describe all important issues in an environmental study based on the workflow except for sampling issues and analytical techniques. These points help to evaluate results, risk assessment, and finally decision making and remediation for responsible authorities. The review may be used by a wide range practitioners, environmental scientists and engineers, and others involved in soil contamination by heavy metals in the world.

Investigation methods used in the examination of soil contamination by heavy metals

Contamination indices

Contamination indices are currently and widely used for the assessment of soil contamination. They also evaluate soil quality and the prediction of future ecosystem

sustainability especially for agricultural purposes. Moreover, the indices provide to determine the source of heavy metals, natural processes, or anthropogenic activities. The most widely applied indices are critically summarised from literature at the following. Uses, advantages, disadvantages and related references are given in *Table 5*. This tabulation is the key assessment of soil contamination by heavy metals. A recent study related to the description of a wide spectrum of contamination indices can be found in Kowalska et al. (2018).
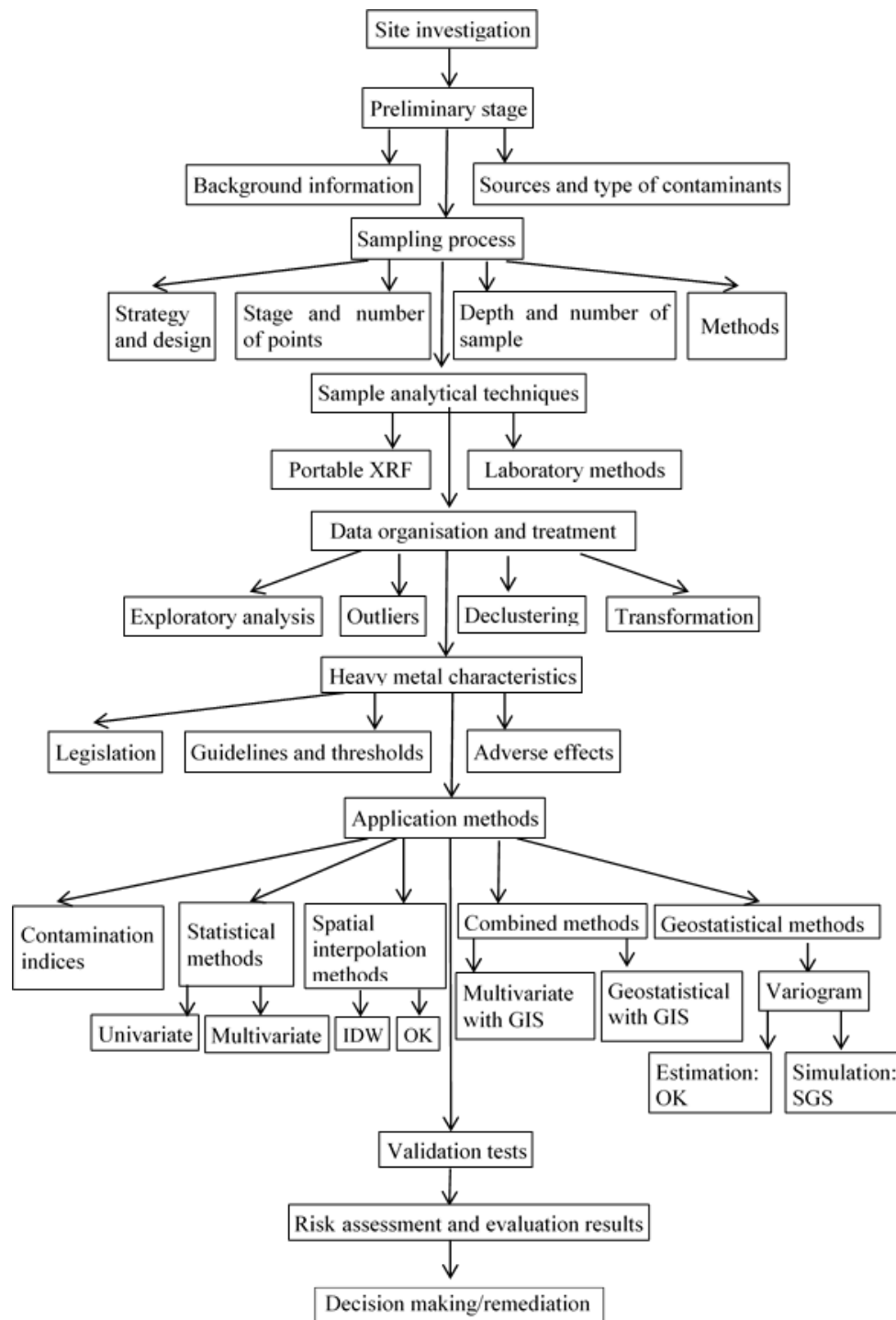
Figure 1. *Site investigation methods workflow in soil contamination by heavy metals*

Table 5. *The main characteristics of the most widely used contamination indices in soil contamination by heavy metals*

Index	Use extent	Advantages	Disadvantages	References
Igeo	• Simple, easy and most widely used • Contamination degree of single heavy metal • GB	• The comparison of the current and previous contamination • 1.5 multiple factors reduce lithogenic effect • Correct scale	• Bad GB selection, bad results • Ignores natural geochemical changes • Natural variability in GB	Chen et al., 2015 Karim et al., 2015 Sayadi et al., 2015 Wang et al., 2005 Su et al., 2014
PI	• Contamination degree of single heavy metal • Easy and widely use • GB	• Correct scale	• Ignores natural variability • Improper selection GB, wrong results	Begum et al., 2014 Chen et al., 2015 Karim et al., 2015 Sayadi et al., 2015
EF	• Identification of heavy metal origin • Comparison of heavy metal concentrations	• Predicts heavy metal origin and anthropogenic effect • Evaluation of the contamination by single heavy metal • Reduces heavy metal variability • Correct scale	• Results depend on GB selection • Assessment of uncontaminated contents	Inengite et al., 2015 Karim et al., 2015 Thabet et al., 2014 Sayadi et al., 2015 Varol., 2011 Wang et al., 2005
CF	• Soil quality • Toxic materials	• Single for each metal • Containing the difference between sample and reference values • Easy and direct application • Correct scale	• Ignores the variability of natural process and presence of heavy metals • No GB • Previous reference value necessary before contamination	Hakanson, 1980 Inengite et al., 2015
PIsum	• Evaluation of contaminants group for all contamination	• Integrates all heavy metals • Comparing for contamination in different soil • Depend on PI values	• Ignores the variation of natural process and the presence of heavy metals • Selection of GB is important • Missing of correct scale	Hakanson, 1980 Inengite et al., 2015
PLI	• Evaluation of degree of contamination • Easy and widely used	• Integrates a number of heavy metals • Comparing of contamination in different soil • Depends on PI values • GB use	• Related GB • Ignores natural process and presence of heavy metals	Karim et al., 2015 Thabet et al., 2014 Pejman et al., 2015
ExF	• Contaminated site point • All soil evaluation • Easy use	• Integrates all heavy metals	• Not widely used • Ignores natural process • No correct scale	Babelewska, 2010

GB: geochemical background value

Geoaccumulation Index (Igeo) was first introduced by Müller (1969). It provides the assessment of soil contamination by heavy metals depending on its content in a horizon by comparing differences between current and background concentrations. Igeo is defined by *Equation 1:*

$$Igeo = \frac{Log2\ Cn}{1.5\ GB} \qquad (Eq.1)$$

where Cn is the measured concentration of the heavy metal and GB is the geochemical background value of the heavy metal. 1.5 is a constant, providing for an analysis of the variability of the data due to natural processes. Igeo values has been classified into seven quality classes (Müller, 1969; Li et al., 2014).

Single Pollution Index (PI) is used to assess greatest heavy metal accumulation in soil. It is expressed as follows (*Eq. 2*):

$$PI = \frac{Cn}{GB} \qquad (Eq.2)$$

where Cn is the content of heavy metal and GB is the geochemical background value.

Enrichment Factor (EF) is a measure of the effects of heavy metal concentrations by anthropogenic activity in soil (*Eq. 3*). It is computed in the following formula (Sutherland, 2000).

$$EF = \frac{sample\ {}^{Cn}/_{Lv}}{background\ {}^{GB}/_{Lv}}$$ (Eq.3)

where Lv (sample) is the concentration of the reference element in the soil. GB is the value of geochemical background. Reference metals are Fe/Al/Ca/Ti/Sc/Mn. If the EF value is less than 1.5, there is no heavy metal contamination in soil occurred with natural processes. If EF is more than 1.5, the heavy metal contamination formed as a result of anthropogenic processes (Elias and Gbadegesin, 2011).

Contamination Factor (CF) measures the content of heavy metal from the surface of the soil and values of pre-industrial reference levels defined by Hakanson (1980). CF is calculated by *Equation 4:*

$$CF = \frac{Cm}{Cp-i}$$ (Eq.4)

where Cm is mean content of heavy metal of at least five samples and C_{p-i} is pre-industrial reference value.

Sum of Contamination (PI_{sum}) is defined as the sum of all determined concentrations of heavy metals, given by Gong et al. (2008). PI_{sum} is computed using *Equation 5:*

$$PI_{sum} = \sum_{i-1}^{n} PI$$ (Eq.5)

where PI is calculated values of single pollution index, n is the number of total heavy metals.

Pollution Load Index (PLI) is used for the sum of assessment of heavy metal contamination in soil (Varol, 2011). PLI is determined from *Equation 6:*

$$PLI = \sqrt[n]{PI_1 \times PI_2 \times PI_3 \times \ldots .. PI_n}$$ (Eq.6)

where n is the number of studied heavy metals and PI is calculated values of the single pollution index.

Exposure Factor (ExF) is used to measure the greatest heavy metal accumulation in the study area (*Eq. 7*). This is calculated from the following formula (Babelewska, 2010).

$$ExF = \sum \frac{C_n - C_{av}}{C_{av}}$$ (Eq.7)

where Cn is the content of heavy metal at the sampling point, C_{av} is average content of heavy metal in the soil profile.

Other indices of less use are Nemerow Pollution index, ($PI_{Nemerow}$: Gong et al., 2008); Average Single Pollution Index (PI_{avg}: Gong et al., 2008), Vector Modulus of pollution Index (PI_{vector}: Gong et al., 2008), Background Enrichment Factor (PIN: Caeiro et al., 2005), Multi Element Contamination (MEC: Adamu and Nganje, 2010), Contamination Security Index (CSI: Pejman et al., 2015), Probability of Toxity (MERMQ: Pejman et

al., 2015), Degree of Contamination (C_{deg}: Hakanson, 1980), Potential Ecological Risk (RI: Hakanson, 1980), Modified Degree of Contamination (mCd: Abrahim and Parker, 2008). These are well documented in the literature (e.g. Kowalska et al., 2018).

Evaluation of geochemical background and contamination indices

The selection and identification of proper reference values for uncontaminated soil is a key task which results in precisely assessing soil contamination by heavy metals, because overall quantitative assessment methods are dependent on reference values of background concentrations (Desaules, 2012). There are many background definitions and related terms in literature. These definitions and applications of background values in environmental geochemistry are discussed and well documented in the literature (e.g., Reimann et al., 2005; Wu et al., 2014). The following important points can be briefly given:

- No specific global and regional background levels of heavy metals can be described. Because natural and anthropogenic effects are different in different regions.
- The levels of background concentrations are based on the area and its scale.
- Background value is a range and not absolute value due to heterogeneity of the environment.
- Natural background may vary in earth crust due to human activities.

Selection of proper geochemical background (GB) is significant in the evaluation of heavy metal contamination (Varol, 2011). Application of various GB provides a more precise investigation of contamination index values. This may be based on the possibility of the contamination of individual sites (Karim et al., 2015).

Two types of GB were classified as reference and local (natural) (Kowalska et al., 2016). The average content of heavy metals can be changed due to local heterogeneity and soil type which may be described with the reference geochemical background (RGB). Local geochemical background (LGB) is the occurrence of natural process which is not affected by human activity (Reimann et al., 2005).

RGBs do not contain natural variability (Xu et al., 2015). Use of RGB is not always possible to recognise natural and anthropogenic effects (Kowalska et al., 2016). However, RGB provides global or regional models of heavy metal contamination (Karim et al., 2015). Calculation of contamination indices needs RGB for many purposes.

LGB contains heavy metal content in rocks and the average content of samples and considers a definite level of human activity (Karim et al., 2015). LGB application is recommended for individual sites under the effect of natural activities and anthropogenic impact (Kierczak et al., 2016). However, LGB may change significantly through lithogenic processes, and its level should be evaluated within geologically homogeneous area (Kowalska et al., 2018). Consequently, literature argued that RGB and LGB values can be used to have complete knowledge (Reimann and de Caritat, 2017; Kowalska et al., 2016).

Many literature research studies have demonstrated that selection of contamination indices are used for different purposes including contamination degree, heavy metal source, potential risk of heavy metal accumulation, ecological risk, the scale of total concentration (e.g. Dung et al., 2013; Guan et al., 2014; Baran et al., 2018). These

criteria are used in the calculation of contamination indices which are based on GB values (e.g. Igeo and EF), data (e.g. CF), and heavy metal content in the soil (e.g. ExF). Although there are clear similarities between contamination indices, they differ from each other due to the effects of several factors.

Igeo and PI are the most accurate and widely used to assess the level of contamination (Begum et al., 2014; Karim et al., 2015; Sayadi et al., 2015). The indices provide to compare previous and present contamination and to have correct scale.

EF makes difference between contamination sources of anthropogenic activities and natural processes (Kowalska et al., 2016). EF identifies low concentrations of heavy element variability (Karim et al., 2015). RGB values have frequently been used in the calculation of EF like Igeo and PI. Heavy metal concentration levels of the sample and reference values are mostly described by concentration variability.

The calculation of CF does not need GB (Li et al., 2016). However, CF distinguishes proportion difference between single heavy metal contamination and previous industrial reference values. CF ignores the variability of natural activities (Varol, 2011).

PIsum and PLI are used for overall soil contamination assessment. These indices are similar to PIavg, PIvector and PIN, which are applied to similar purposes. Their uses are easy and simple. They exhibit reasonable levels of heavy metal contamination (Inengite, 2015). The main weakness of them has individual scale.

Consequently, appropriate selection of contamination indices is based on the degree of contamination, purposes of use, soil type (e.g., farmland, forest, and urban site). Understanding knowledge of contamination index is a basic key task for environmental management, risk of environmental exposure, agricultural practises, ecosystem protection, identification of natural and anthropogenic sources.

Multivariate statistical analysis methods

Multivariate statistical analysis is often used for identifying sources of heavy metal contamination (Mostert et al., 2010). The Methods consist of principal component analysis (PCA), cluster analysis (CA), Pearson correlation analysis, factor analysis (FA), multiple linear regressions (MLR). PCA is the most frequently used multivariate statistical analysis method to reduce data dimension. This technique derives to determine the variance in the data with a small number of independent variables referred to principal components (Boruwka et al., 2005). The relationship between metal fractions and physical chemical properties is determined by PCA. Varimax rotation is applied to minimise the number of variables with a high loading on each component and operates the assessment of results (Mico et al., 2006). In another way, an orthogonal transformation technique is used to obtain the first principal component showing for the highest variance in the observed data. An eigenvalue decomposition in matrix is constructed with the highest eigenvalue which is the principal component of the data (Hou et al., 2017). Consequently, it is important to adequately treat and organise the data for multivariate statistical analysis. Appropriate transformation is necessary.

CA is the second most used multivariate statistical analysis method in the literature (Hou et al., 2017). Variables of the data set are divided into groups of similar features. CA algorithms minimise and maximise inter group variability. CA is used to confirm PCA results for soil contamination by heavy metals.

Other less commonly used methods are Pearson correlation, FA and MLR. Pearson Correlation analysis makes linear correlation between two variables. It corresponds a

correlation coefficient ranging from -1 to 1. -1 represents perfect negative linear correlation, 0 refers to no correlation and 1 indicates perfect positive correlation. Pearson correlation is useful for PCA and CA. Hou et al. (2017) pointed out that Pearson correlation is not a multivariate statistical analysis technique. Because it explains only single pairs of variables at a time. FA proposes to reduce data set dimension like PCA, but mathematical methods of them are different. FA uses a discrete model to provide n variables within latent variables (n > m) whereas PCA does not account the model (Jaliffe, 2002). FA is often used in human behaviour study related to the environment (Hou et al., 2017). However, FA has rarely been used for soil contamination of heavy metal (Romic and Romic, 2003). MLR has seldom been used for spatial distribution of heavy metals in soil. MLR combined PCA was used by Ali et al. (2016) for quantifying the origin of heavy metals.

In conclusion, combined geographic information system (GIS) and multivariate analysis have been recently used by increasing number of studies for the assessment of spatial distribution of heavy metals to quantify soil quality in regional scale (e.g., Huang et al., 2015; Lin et al., 2016; Moore et al., 2016; and Zhou et al., 2016). GIS is a compilation of computer hardware, software, spatial and non-spatial data, and users designed to efficiently capture, store update, manipulate, analyse, and display all forms of geographically referenced information. GIS software is interoperable, supporting many data formats used in the infrastructure life cycle. Its technology provides a central location to conduct spatial analysis, over by data, and integrate other applications or systems. The recent development of GIS is to capture digital data in the field and provide more efficient transfer from field to office. GIS technology is changing fast and moving from mainframe computer to workstation and to desktop-based PC systems. GIS is driven by jurisdictional, purpose or application requirements. Most phases of infrastructure life cycle are commonly affected and enhanced by the enrolments of GIS.

Spatial interpolation methods

Interpolation is the process of estimating the values of interest variables at unsampled areas. Spatial interpolation methods differ from classic modelling approaches since spatial methods provide knowledge about the geographic position of the sample point. In the spatial interpolation sampling points closer to each other, exhibit good correlations and more similarities than the points further away. In this study from literature review (e.g. Li and Heap, 2014; Xi et al., 2011; Hou et al., 2017) the most frequently used spatial interpolation methods, inverse distance weighting and ordinary kriging (OK) were reviewed.

Inverse distance weighting (IDW)

IDW is based on a linear combination of data set. The main advantages of IDW are fast and easy use, directly interpolation (*Table 6*). Thus, the method most widely used for environmental and mining studies. The important weakness of IDW is that it does not account a particular model of spatial correlation for the variables being interest. The interpolating equation is given as follows (*Eq. 8*):

$$Z_{xy} = \frac{\sum_{i=1}^{n} z_i d_{xyi}^{-\beta}}{\sum_{i=1}^{n} d_{xyi}^{-\beta}} \tag{Eq.8}$$

where Z_{xy} is the estimated value at an interpolated point, z_i is the control value for the i_{th} sample point, n is the total number of observed points used in interpolation, d_{xyi} is the distance between Z_{xy} and Z_i, and β is on exponent described by the user. As the distance increases the weight decreases and weighting power incorporates the weight decreases while the distance increases. The accuracy of IDW may be improved by selecting the optimal neighbouring points and exponent value to generate optimum arrangement between observed data and the prediction. Many soil quality survey studies revealed that integrated IDW with GIS and multivariate statistical analysis have been used in several regions quantifying soil contamination of heavy metals (e.g. Haung et al., 2015; Lee et al., 2006 and Zhang, 2006).

Non-geostatistical rarely used other spatial interpolation methods are nearest neighbours, triangular irregular network related interpolations, natural neighbours, regression models, trend surface analysis, thin plate splines, regression tree, local polynomial, and radial basis functions.

Kriging

Kriging is the geostatistical method that is the most widely used among spatial interpolation methods for spatial distribution in soil. Kriging is produced from regionalised variable theory and dependent on stochastic spatial variation model. Confidence intervals for the values of variables at unsampled locations are estimated by kriging. A linear combination of the observed values with weights gives the kriging predictor. There are many types of kriging that include simple kriging (SK), ordinary kriging (OK), factorial kriging, dual kriging, indicator kriging (IK), disjunctive kriging, model-based kriging. These refer to univariate kriging type, whereas universal kriging (UK), SK with varying local means, kriging with external drift, simple cokriging (SCK), OK, standardised ordinary cokriging, principal component kriging, collocated cokriging, kriging with strata, multivariate factorial kriging, IK with an external drift, indicator cokriging and probability kriging are classified as multivariate kriging types (Li and Heap, 2014).

Kriging equation is given as follows (*Eq. 9*):

$$Z(B) = \sum_{i=1}^{n} \lambda z(xi) \tag{Eq.9}$$

where Z(B) is the estimated area, λi is weight and Z(xi) is sample value. Ordinary kriging (OK) is the most frequently applied technique among the kriging types for environmental, geological and mining studies. Advantages and weakness features of OK is given in *Table 6*. The main characteristics of OK can be presented at the following points:

- OK is the best linear unbiased estimator.
- OK estimates unsampled locations in studied site
- OK measures estimation errors and uncertainty
- OK minimise the variance of the data. Its variance is based on data values. The error variance is poorly correlated with actual estimation error. Thus, kriging variance may not be used alone as a measure of local uncertainty.
- OK provides spatial structure
- OK requires variogram construction before operating spatial interpolation process. Thus, OK is significantly affected by variogram parameters such as nugget effect, sill, range, variogram model or shape, search radius, number of

neighbouring measurements. Sufficient data and appropriate distribution of data are necessary for variogram building.

- The biggest weakness of OK has smoothing effects. Interpolated surface is smooth which can cause low values to be overestimated and high values to be underestimated. This most probably resulted in the high contamination risk area. Underestimated and low risk are clean area overestimated.
- Error assessment of OK is based on variogram structure, distribution of data points and size of interpolated blocks.
- If data are sufficient and appropriate distribution to compute variogram, OK will provide a well interpolator for sparse data.
- OK does not require knowledge of the mean over the region of interest and operates under simple stationarity assumptions.
- OK is a robust estimator due to only requiring local stationarity.

Many case studies demonstrated that combined multivariate statistical analysis with kriging analysis can be a reliable and useful tool to determine spatial distribution and source of heavy metals, to quantify soil quality (Maas et al., 2010; Lu et al., 2012; Shao et al., 2014; Cai et al., 2015; Gabarron et al., 2017).

Geostatistical simulations

Simulation is defined as imitations of conditions. Simulation generates an equally probable realization representing spatial distribution of heavy metals and measuring of uncertainty of the area being studied. Exploratory statistics such as mean, median, variance, coefficient of variation, standard deviation, skewness, and kurtosis; histogram; the variogram (spatial dispersion variance) of the original data information are reproduced by simulations on real scale. Simulated realizations are constructed on a fine grid. Simulation characteristics of soil contamination play a key role in sampling strategy and designing, planning, decision making, implementation risk and scheduling in site assessments. Significant parameters can be derived from the distribution of local uncertainty such as exploratory statistics and probability of exceeding value or threshold limit. Thus, a simulation process is a significantly more completed model than the single estimated block or point model.

Simulations provide a variety of purposes including study of element concentration continuity, optimising sampling for advanced investigation, assessment of soil contamination estimation methods, site (environmental) planning, risk evaluation (e.g. financial) and any integration of the aims given here.

There are two types of geostatistical or stochastic simulations, unconditional and conditional. Unconditional simulation is simply an application of the general Monte Carlo Technique that simulate values and are generated with a particular covariance function and variogram. Several simulation techniques exist for practitioners. Four methods are common in use; they are sequential Gaussian simulation (SGS), simulated annealing, and simulation by turning bands and lower-upper decomposition. The first three methods can generally be conditional and lower-upper decomposition is often used for unconditional (Webster and Oliver, 2007).

SGS is the most widely used technique in environmental studies for site assessment particularly to quantify risk and quality of soil (Goovaerts, 2001; Qu et al., 2013; Zhang et al., 2017; Ersoy and Yünsel, 2018). Sequential indicator simulation, direct sequential

simulation, and sequential Gaussian cosimulation or joint simulation are an extension of SGS simulation models of several continuous variables and based on SGS algorithms. A growing number of many environmental researchers have also used these applications for soil contamination by heavy metals (Huang et al., 2015; Franco et al., 2006; Ersoy and Yünsel, 2019). Because applications of SGS are simple, flexible, and fast; thus, SGS is briefly reviewed here (*Table 6*). SGS algorithm can be found in the literature in details (Journel and Alabert, 1989; Deutsch and Journel, 1998).

Table 6. Main characteristics of IDW, OK and SGS

Method	Inverse distance weighted (IDW)	Ordinary kriging (OK)	Sequential Gaussian simulation (SGS)
Advantages	• Fast and easy use • Direct interpolation • Widely used	• Best linear unbiased estimator • Measures estimation errors/uncertainty • Estimates unsampled locations • Provide spatial structure • Robust estimator • Local stationary • Knowledge mean of the region studied • Most widely used • Good variogram, good estimation	• No smoothing effect • Quantify uncertainty • Probabilistic map present risk assessment • Reproduce statistics, variogram, histogram and contour plots • Maps show contaminated and uncontaminated areas • Assessment of spatial structure • Evaluation of sampling strategy and design • Most frequently used
Weakness	• Does not measure errors • Smoothing effect • Performance based on size of search area • Select of weighting parameters • Neighbouring points	• Smooth effects on results • Bad variogram, bad estimation • Minimise variance lower than data variance • Performance based on variogram • Quality of data • Size of interpolated blocks	• Great tutorial, expertness and experience necessary • Long-time computerising • More trial and error • Reproductions and number of realisations • Results depend on sampling process • Data organisation and treatment • Neighbouring and search parameters • Block characteristics • Variogram and its parameters

A schematic diagram (*Fig. 2*) exhibits the basic and summary steps involved in the process of SGS. The main advantages of SGS include:

- SGS does not have smoothing effects unlike traditional interpolation methods such as kriging, IDW. SGS ensures to evaluate exactly the high and small values in the data.
- SGS generates maps representing an equally probable spatial distribution and to quantify uncertainty of heavy metals for site exploration.
- SGS produces maps showing contaminated areas and uncontaminated areas across the site.
- SGS is a probabilistic approach that provides probabilistic maps; exhibits a different description of regions into safe and hazardous.
- SGS reproduce descriptive statistics, histogram, variogram and contour plots of spatial characteristics. These are correlated with the same spatial characteristics of original data. This refers to validation tests of SGS.

In conclusion, literature studies demonstrated that risk and quality assessment in decision making should not be based on only kriging estimates. SGS should be operated in uncertainly assessment especially soil contamination with heavy metals. Because SGS ensures local variations in values of a contaminant particularly including design and strategy, estimation procedures, site planning and any risk assessment (e.g. financial).
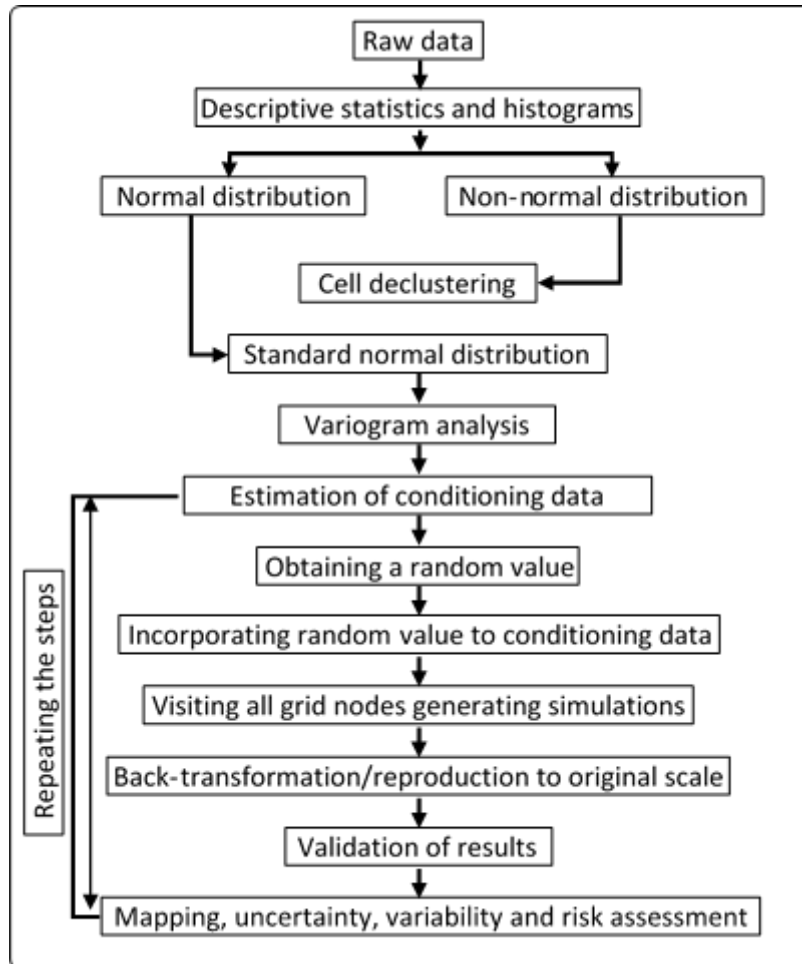
Figure 2. Schematic diagram showing the basic steps in the process of SGS

Factors affecting performance of investigation methods (estimation and simulation)

Various factors affect the performance of estimation and simulation methods (Isaaks and Srivastava, 1989; Burrough and Mcdonell, 1998; Zimmerman et al., 1999; Schloeder et al., 2001; Verly, 2005; Wang et al., 2005; Wu et al., 2006; Stahl et al., 2006; Hengl, 2007; Li and Heap, 2011; Xie et al., 2011; Rossi and Deutsch, 2014). These factors can be classified into four groups including sampling process, data organisation and treatment, variogram modelling and model variogram parameters, and cross validation. The factors were usually encountered in the literature.

Sampling process

The quality of soil contamination estimates is dependent on the available data based on the quality of sampling procedures. If the samples are not representative to form sample bias which will directly affect the final contamination estimate the result will not be reliable and accurate.

The estimate performance is generally measured using errors (Sinclair and Blackwell, 2002; Li and Heap, 2011, 2014; Rossi and Deutsch, 2014). There are no perfect measurements (Neufeld, 2005). The relatively large mass of a sample is reduced

to a small subsample from which a few grams are taken to make chemical analysis. There must be a difference between subsample content, the original sample and the analysed (assay) sample. The difference refers to sample error. Rossi and Deutsch (2014) presented that there are two forms of error. One is present due to the intrinsic properties and the material being sampled. The other comes from inappropriate sampling procedures and preparation. In the literature, there are several errors in measurement that are most commonly used including fundamental error, increment delimitation error, increment extraction error, mean or average variance, coefficient of variation, mean absolute error, mean squared error, root mean squared error, relative mean absolute error and relative root mean square error. Fundamental error results from constitution heterogeneity of the material being sampled. This error is random with a mean of zero. However, delamination and exaction errors are mean of non-zero, the errors resulted from improper sampling, and thus bias is related to the sampling procedure.

A variety of issues related to sampling process need to be considered sample collection, handling, preparation, and analysis. However, in this review sampling density or size and sampling design are focused factors especially in environmental studies because these factors were met in previous literature.

Literature studies have usually argued that when the sample density increased, the errors decreased. But as this reached a threshold number of sample density, collected addition of further samples does not improve the performance of estimation methods (Li and Heap, 2008, 2011). However, size of study site plays a significant role on sampling density. Different researchers studied different scales for small survey area or for larger survey area. When sample density is big enough, most estimation methods generate similar results (Burrough and Mcdonell, 1998). Estimation methods produce better results as the sample density increases (Isaak and Srivastava, 1989; Stahl et al., 2006). The effects of sample density on the errors are based on the type of estimation methods (Hengl, 2007; Li et al., 2014; Li and Heap, 2011, 2014). In practical applications, after a number of threshold samples further increase in sample size may not significantly contribute to the accuracy of the estimations and simulations. If sample size does not reach the threshold, it will still be a critical factor. Thus, the impacts of the sample density are controlled by data organisation and treatment as discussed below.

Data organization and treatment

Data is a key factor to influence on the performance of estimation and simulation methods. Major factors related to data quality are summarised here including exploratory data analysis, outliers and declustering.

Exploratory data analysis

The descriptive tabulated and graphical forms have been used to characterise data nature and quality since about 1940. The descriptive summary statistics include mod, median, mean, standard deviation, variance, and standard error of mean, coefficient of variation, skewness, and kurtosis. Histogram, probability, and quantile-quantile plots are graphical applications to assess data distribution. These statistical tools are very useful to construct data quality for several reasons, including understanding of the data and soil contamination, to summarise information, to provide inferences, estimations, and simulations.

Normal distribution of the input data can significantly affect estimation and simulation accuracy. If the data are not normally distributed, lognormal transformation is frequently used e.g., lognormal kriging. Other transformation methods may also be applied to obtain normal distribution, resulting in Gaussian kriging and multi-Gaussian kriging (Cressie, 1993). SGS requires normal sore transformation of the original data with zero mean and unit variance.

Histogram is the most basic statistical tool used in exploratory data analysis. Three factors should be considered: arithmetic or logarithmic scaling, arithmetic scaling is appropriate, whereas logarithmic scale clearly represents highly skewed data distribution; range of data; and number of boxes. The mean value is influenced by outliers; the median is affected by missing data. If coefficient of variation is higher than 2 the distribution should be combined with high and low values together.

All data values are given on the probability plot. Different statistical populations can be interpreted. Probability plot is also useful to determine data distribution, straight line on arithmetic scale presents a normal distribution; a straight line on logarithmic scale corresponds a lognormal distribution.

Outliers and declustering

Outliers are extreme values (a small number of very low or very high values) inconsistent with the majority of data values; outliers have significant influence on descriptive statistics and measures of spatial continuity. The extreme values should be removed from the data. There are different ways to identify the outliers: many geostatistical techniques require a transformation of the data that reduce the effect of outliers. Probability plots are useful to check extreme values. Another method is cutting (capping) for identification of outliers. Cutting values higher than cutting threshold (outlier threshold) can regulate to the outlier threshold itself. Outliers can come dramatically to increase the nugget variance of experimental variogram which would mislead us. If outliers are suspected, they should be removed by the identifying technique as explained above.

It is difficult in some situations that whether those values are outliers or not. In such cases, called outliers should be retained. Dowd (1984) and Genton (1998) recommended using the robust variogram estimators which reduce the effects of outliers.

Data are rarely collected randomly. Soil data commonly have relatively high concentration in source zone, thus histogram of raw concentrations is biased. The effect of clustering should be removed to obtain unbiased histogram and summary statistics. Declustering techniques are applied to provide the true form of the spatial distribution and descriptive statistics. Three declustering methods are most commonly used in literature: cell declustering (Journel, 1983), the nearest-neighbour declustering and polygonal declustering (Isaak and Srivastava, 1989). These techniques are well documented in the related references.

Variogram

Estimation (e.g. kriging) and simulation (e.g. SGS) are based on experimental variogram modelling. The variogram parameters consist of lag distance, nugget effect, variance, sill, range, and fitting model. Their definitions can be found in most geostatistical textbooks and software (e.g., Chiles and Delfiner, 2012; Geovariances, 2014; GSLIB: Deutch and Journel, 1998). Oliver and Webster (2014) showed that

reliability of the experimental variogram is affected by several factors including sample size, lag interval and bin width, marginal distribution of the data, anisotropy, and trend.

Accuracy of the experimental variogram is based on the size of the sample. Generally, more data represent more reliability. Webster and Oliver (2014) pointed out that variograms computed from less than 100 data are unreliable. When the sampling interval decreases, the size of the sample increases. Sampling interval provides the applicability of the experimental variogram. If the sampling interval is wider than the correlation range of the study, pure nugget effect will occur in the experimental variogram. This result in estimation is not reliable. Grid sampling is very appropriate in the field to produce well estimation maps and simulation realizations. However, if the grids are coarse, the process may lose the short-range variation significant for the variogram. The solution is that extra samples in the grid should be collected.

Choosing lag distance and bin width significantly affects judgement of the experimental variogram. If lag distance and bin width is short, there will be many variogram models which subject to wide error. In contrast, there will be few models with large lag distance and bin width. In practise, the experimental values should be selected plausibly.

Another important factor influencing variogram is data distribution. Variances incredibly increased as the data positively skewed. Histograms, box-plots and descriptive statistics should be computed to determine the data distribution. Many soil data are strongly and positively skewed in the literature, skewness value of data is greater than 1, and the data will be transformed to logarithms. If the skewness value is between 0.5 and 1, transformation to square roots will make normal distribution. Outliers cause seriously skewed distribution of the data as discussed earlier.

In many cases, data variation is not isotropic, displays anisotropy evidence. Practitioners may ignore or not to detect anisotropy make model improperly. Thus, biased estimations or simulations may emerge. Both directional and omnidirectional experimental variograms of variable/variables should be analysed separately. If the directional variogram reveals anisotropy, we should narrow direction angle to state its expression and to provide the direction of maximum continuity.

Trend can be defined as gradual variation in spatial data. In practise, it is difficult to decide on the trend. Detection of global trend is done by making map from the data with proper software in an easy manner. The map presents gradual continuous variation. Experimental variogram shows increasing changes more steeply when the lag distance increases. We can confirm the presence of trends mapping residuals. Thus, the main direction of the trend can be identified. The experimental variogram is constructed in the direction perpendicular to the principal direct of the trend.

In environmental studies, environmental authorities require predictions of contaminants in blocks that are appropriate size for remediation. Block size is significant for farmer applications of agricultural chemicals such as lime, fertilisers, and pesticides. Oliver and Webster (2014) pointed out that block size is typically 24 m in Europe. In practise, block kriging is more suitable than point estimates (punctual kriging) because block kriging variance is typically much less than punctual kriging variance.

Mapping software uses a moving window based on a chosen size of neighbourhood block in a study area. Although there are no rules to describe the estimation neighbourhood, the following guidance is given by Oliver and Webster (2014):

- If there is good (dense) data, the variogram will have a negligible nugget variance. The radius of the neighbourhood can be chosen close to the range.

- If a variogram has a large nugget variance, the radius of the neighbourhood can be chosen higher than the range.
- Minimum and maximum number of neighbourhood nearest data to the target is usually recommended 7 and 25, respectively.
- If the data are sparse and unevenly distributed, the neighbourhood is divided into octants. Each octant has at least two data points.
- If the data are irregularly scattered, the neighbourhood will be moved to predict a field of values for mapping.

Cross validation for kriging

Cross validation can be performed to cross-check the candidate models and data by statistical and graphical results. Mean error (ME), mean standardised squared error (MSSE) and mean squared deviation ratio (MSDR- the mean of squared errors (MSE) divided by the referring Kriging variances) is calculated. Perfectly the mean error should be zero, thus Kriging results are unbiased, even poorly produced model. The MSE should be at minimum level, and is composed of good statistics; however, the MSE will not determine a true model. The MSDR is the most testing criteria, it should be 1. If the MSDR is close to 1, this kriging model should be chosen.

Verification of graphical results are the base map, scatter diagram of observed data versus estimated value, histogram of the standardised estimation errors (SEE) and scatter diagram of standardised estimation error versus estimated value. Outliers are depicted on the base map. Scatter plots of observed data versus estimated value show conditional bias and variability. This indicates that true and estimated data should match exactly with each other. Nearly all frequencies of SEE should be equal zero or around zero. Scattered points must be distributed around zero within acceptable limits in graph of estimated value versus SEE. More information related to cross validation of kriging results may be obtained from literature (e.g. Ersoy et al., 2004; Webster and Oliver, 2007).

Cross validation for simulation

Verification of the simulation process can be carried out using a number of tests including summary statistics, histograms, variograms and contour maps. Summary statistics of the simulated data are compared to summary statistics of the raw data. The comparison should be reasonable in good agreement. The histograms of the observed data should be quite similar to the histograms of the simulated realizations. It should be noted that each histogram from produced realizations (e.g., hundred simulated realizations) must be randomly chosen. Simulated data or reproduction variograms are compared to the variograms of the observed data. The comparison should be a good reproduction of spatial variability. The contoured map of the simulated mean data and countered observed data should be presented. The simulation mean values should plausibly reproduce the intrinsic charter of the observed data. More details for validation of simulation results can be found in Ersoy and Yünsel (2019).

Conclusions

The main conclusions of the study are the following points:
- Adequate background research and information identifying sources and types of contaminants should be gathered in preliminary stage for site investigation.

- Adequate and good quality data are essential for the assessment of soil contamination by heavy metals.
- Good quality data are controlled by data organisation and treatment such as exploratory data, outlier analysis and declustering.
- There are wide ranges of variations in regulations and soil threshold limits (or heavy metal concentration guideline in soil) for different countries due to different politics, culture, and objectives. Briefly, the most regulations are still dependent on the total concentrations of heavy metals in soil.
- Literature studies demonstrated that there are some similarities and differences between contamination indices. Strengths and limitations of the indices have been compared. Igeo and EF are most commonly and universally applied for a range of contamination. The selection of appropriate index is a key task to understand the degree of contamination, the use of soil, and the aims of contamination indices characteristics. Geochemical background plays a significant role which is based on specific site and scale of contamination assessment. The contamination indices combined with multivariate statistical analysis provide discrimination between natural and anthropogenic heavy metals in soils.
- Multivariate statistical analysis is a better tool than classical univariate statistics identifying sources of heavy metal contamination. Literature research revealed that an integration of multivariate statistical analysis, GIS and geostatistical methods can be accurate and reliable to characterise spatial distribution of heavy metals and to determine their origin.
- Simulation and estimation have different purposes. Simulation provides local variations of a variable examined. Estimation has effect on environmental planning to identify contaminants on scale that physical dispersion can be achieved.
- In theory, simulation is superior to OK, because of its conditional expectation, whereas OK is conditionally biased. In general, performance of kriging techniques is better than interpolation methods (e.g., IDW) or geostatistical methods. However, conventional interpolation methods and kriging techniques have smoothing effects, whereby small values are typically overestimated, while high values are underestimated. Kriging estimates that local error variance is at a minimum level, less than original data variance, thus generally contains bias as a measure of reliability. This has a negative effect on soil contamination or environmental risk assessment. The solution is geostatistical simulation that overcomes kriging and other interpolation methods. SGS is most commonly used for spatial distribution, uncertainty, and risk assessment of heavy metals in soil contamination. The literature review revealed that risk assessment in decision making should not be dependent only kriging estimates. SGS should be carried out in uncertainty assessment, typically heavy elements contamination in soil. Because simulation process results in local variations in values of heavy elements particularly including sampling design and strategy, estimation procedures, site planning and risk assessment (especially financial).
- Performances of geostatistical estimation and simulation are affected by many factors including sampling process, data organisation and treatment, variogram and its parameters; block, search area and neighbourhood characteristics. These factors have been critically and comparatively reviewed.

- Multivariate simulation should be applied for soil contamination to create maps assessing uncertainty and representing an equally probable spatial distribution of the heavy metals in soil for future study. These maps will show contaminated and uncontaminated areas in the study site for future studies.

- It is also recommended for future research studies that sequential co-simulation integrated with local singularity analysis is effective, powerful, and useful tool to generate maps for risk analysis and to quantify uncertainty of soil contamination by heavy metals. Uncertainty quantification of soil contamination is a key issue for decision making in environmental analyses.

REFERENCES

[1]   Abrahim, G. M. S., Parker, R. J. (2008): Assessment of heavy metal enrichment factors and the degree of contamination in marine sediments from Tamaki Estuary, Auckland, New Zealand. – Environmental Monitoring and Assessment 136: 227-238.

[2]   Adamu, C. L., Nganje, T. N. (2010): Heavy metal contamination of source soil in relationship to land use patterns: a case study of Benue State, Nigeria. – Materials Sciences and Applications 1: 127-134.

[3]   Albuquerque, M. T. D., Gerassis, S., Sierra, C., Taboada, J., Martín, J. E., Antunes, I. M. H. R., Gallego, J. R. (2017): Developing a new Bayesian Risk Index for risk evaluation of soil contamination. – Science Total Environment 603-604: 167-177.

[4]   Ali, M. H., Mustafa, A. R. A., El-Sheikh, A. A. (2016): Geochemistry and spatial distribution of selected heavy metals in surface soil of Sohag, Egypt: a multivariate statistical and GIS approach. – Environmental Earth Science 75: 1-17.

[5]   Babelewska, A. (2010): The impact of industrial emissions on heavy metal and sulphur contamination level within the area of the projected Jurassic National Park. – Studies and Reports of the Prof Wladyslaw Szafer Museum 20: 135-145.

[6]   Baran, A., Wieczorek, J., Mazurek, R., Urbanski, K., Klimkowicz-Pawlas, A. (2018): Potential ecological risk assessment and predicting zinc accumulation in soils. – Environmental Geochemistry and Health 40: 435-450.

[7]   Begum, K., Mohiuddin, K. M., Zakir, H. M., Moshfigur Rahman, M., Nazmul Hasan, M. (2014): Heavy metal pollution and major nutrient elements assessment in the soil s of Bogra City in Bangladesh. – Canadian Chemical Transactions 3: 316-326.

[8]   Boruwka, L., Vacele, O., Jechlicka, J. (2005): Principal component analysis as a tool to indicate the origin of potentially toxic elements in soils. – Geoderma 128: 289-300.

[9]   Burges, T. M., Webster, R. (1980): Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual Kriging. – Journal of Soil Science 31: 315-331.

[10]   Burrough, P. A., McDonnell, R. A. (1998): Principles of Geographic Information Systems. – Oxford University Press, Oxford.

[11]   Caeiro, S., Costa, M. H., Ramos, T. B., Fernandes, F., Silveria, N., Coimbra, A. (2005): Assessing heavy metal contamination in Sado Estuary sediment: an index analysis approach. – Ecological Indicators 5: 151-169.

[12]   Cai, L., Xu, Z., Bao, P., He, M., Dou, L., Chen, L., Zhou, Y., Zhu, Y. G. (2015): Multivariate and geostatistical analysis of the spatial distribution and source of arsenic

and heavy metals in the agricultural soil in Shunde, Southeast China. – Journal of Geochemical Exploration148: 189-195.

[13] Canadian Council of Ministers of the Environment (2010): Canadian Environmental Quality Guidelines Summary Table. – Canadian Council of Ministers of the Environment, Winnipeg.

[14] Chen, H., Teng, Y., Lu, S., Wang, Y., Wang, J. (2015): Contamination features and health risk of soil heavy metals in China. – Science of Total Environment 512-513: 143-153.

[15] Chiles, J. P., Delfiner, P. (2012): Geostatistics: Modelling Spatial Uncertainty. – John Wiley & Sons, New York.

[16] Contaminated Land Exposure Assessment (CLEA) (2009): Assessment of Potentially Toxic Elements. – Heavy Metal Guidelines in Soil, Technical Datasheet.

[17] Coskun, M., Steinnes, E., Frontsyeva, M. V., Sjobakk, T. E., Demkina, S. (2006): Heavy metal pollution of surface soil in the Thrace region Turkey. – Environment Monitoring Assessment 119: 545-556.

[18] Cressie, N. A. (1991): Statistics for Spatial Data. – John Wiley & Sons, New York.

[19] David, M. (1977): Geostatistical Ore Reserve Estimation. – Elsevier Scientific Publishing Co, Amsterdam.

[20] Davis, H. T., Aeilon, C. M., McDermott, S., Lawson, A. B. (2009): Identifying natural and anthropogenic sources metals in urban and rural soils using GIS-based data, PCA, and spatial interpolation. – Environmental Pollution 157: 2378-2385.

[21] De Almeida, J. A. (2010): Stochastic simulation methods for characterization of lithoclasses in carbonate reservoirs. – Earth-Science Reviews 101: 250-70.

[22] Department of Environment and Conservation (DEC) (2010): Assessment Levels for Soil, Sediment and Water. – DED, Western Australia.

[23] Desaules, A. (2012): Critical evaluation of soil contamination assessment methods for trace metals. – Science Total Environment 426: 120-131.

[24] Deutsch, C. V., Journel, A. G. (1998): GSLIB: Geostatistical Software and User's Guide. 2nd Ed. – Oxford University Press, New York.

[25] Dowd, P. A. (1984): The Variogram and Kriging: Robust and Resistant Estimators. – In: Verly, G., David, M., Journel, A. G., Marechal, A. (eds.) Geostatistics Natural Resource Characteristics. Springer, Dordrecht.

[26] Dung, T. T. T., Cappuyns, V., Swennen, R., Phung, N. K. (2013): from geochemical background determination to pollution assessment of heavy metals in sediments and soils. – Reviews in Environmental Science & Biotechnology 12: 335-353.

[27] Elias, P., Gbadegesin, A. (2011): Spatial relationships of urban land use, soils and heavy metal concentrations in Lagos mainland area. – Journal of Applied Sciences and Environmental Management 15: 391-399.

[28] Ersoy, A., Yünsel, T. Y. (2018): The assessment of soil contamination by heavy metals using geostatistical sequential Gaussian method. – Human and Ecological Risk Assessment: An International Journal 24: 2142-2161.

[29] Ersoy, A., Yünsel, T. Y. (2019): Geochemical modelling and mapping of Cu and Fe anomalies in soil using sequential Gaussian co-simulation and local singularity analysis: a case study from Dedeyazı (Malatya) region SE Turkey. – Geochemistry: Exploration, Environment, Analysis 19: 331-342.

[30] Ersoy, A., Yunsel, T. Y., Cetin, M. (2004): Characterisation of land contaminated by past heavy metal mining using geostatistical methods. – Archives of Environmental Contamination and Toxicology 46: 162-175.

[31] Ersoy, A., Yunsel, T. Y., Atici, U. (2008): Geostatistical conditional simulation for the assessment of contaminated land by abandoned heavy metal mining. – Environmental Toxicology 23: 96-109.

[32] ESRI (1994): Understanding GIS the Arc/INFO Method. – Environmental Systems Research Institute, Redlands, CA.

[33] Fronco, C., Soares, A., Delgado, J. (2006): Geostatistical modelling of heavy metal contamination in the top soil of Guadiamar River margins (South Spain) using stochastic simulation technique. – Geoderma 136: 852-864.

[34] Gabarron, M., Faz, A., Martinez, S., Zornoza, R. A., Costa, J. A. (2017): Assessment of metal behaviour in industrial soil using sequential extraction, multivariate analysis and a geostatistical approach. – Journal of Geochemical Exploration 172: 174-183.

[35] García-Lorenzo, M. L., Pérez-Sirvent, C., Molina-Ruiz, J., Martínez-Sánchez, M. J. (2014): Mobility indices for the assessment of metal contamination in soils affected by old mining activities. – Journal Geochemical Exploration 147: 117-129.

[36] Genton, M. G. (1998): Highly robust variogram estimation. – Mathematical Geology 30: 213-221.

[37] Geovariances (2014): Isatis Software Manual. – Geovariances, Fontainebleau.

[38] Gong, Q., Deng, J., Xiang, Y., Wang, Q., Yong, L. (2008): Calculating pollution indices by heavy metals in ecological geochemistry assessment and a case study in parks of Beijing. – Journal of China University of Geosciences 19: 230-241.

[39] Goovaerts, P. (1997): Geostatistics for Natural Resources Evaluation. – Oxford University Press, New York.

[40] Goovaerts, P. (1999): Geostatistics in soil science: state-of-the-art and perspectives. – Geoderma 89: 1-45.

[41] Goovaerts, P. (2001): Geostatistical modelling of uncertainty in soil science. – Geoderma 103: 3-26.

[42] Goovaerts, P., Soares, A., Gomez Hernandez, J., Froidevaux, R. (1996): Kriging vs stochastic simulation for risk analysis in soil contamination. – Proceedings of GeoENV I Geostatistics Environment. Application, Lisbon.

[43] Guan, Y., Shao, C., Ju, M. (2014): Heavy metal contamination assessment and partition for industrial and mining gathering areas. – International Journal of Environmental Research and Public Health 11: 7286-7303.

[44] Hakanson, L. (1980): An ecological risk index for aquatic pollution control: a sedimentological approach. – Water Research 14: 975-1001.

[45] Hengl, T. (2007): A Practical Guide to Geostatistical Mapping of Environmental Variables. – TRC Scientific and Technical Reports for Official Publication of the European communities, Luxembourg.

[46] Hou, D., O'connar, D., Nathanail, P., Tian, L., Ma, Y. (2017): Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: a critical review. – Environmental Pollution 231: 1188-1200.

[47] Huang, Y., Li, T., Wu, C., He, Z., Jopenga, J., Deng, M., Yang, X. (2015): An integrated approach to assess heavy metal source apportionment in Peri-urban agricultural soils. – Journal Hazardous Materials 299: 540-549.

[48] Inengite, A. K., Abasi, C. Y., Walter, C. (2015): Application of pollution indices for the assessment of heavy metal pollution in flood impacted soil. – International of Pure and Applied Chemistry 8: 175-189.

[49] Isaak, E. H., Srivastava, R. M. (1989): Applied Geostatistics. – Oxford University Press, New York.

[50] Jaliffe, I. T. (2002): Principal component analysis and factor analysis. – Principal Component Analysis 150-166.

[51] Jarup, L. (2013): Hazards of heavy metal contamination. – British Medical Bulletin 68: 167-182.

[52] Journel, A. G. (1988): Nonparametric Geostatistics for Risk and Additional Sampling Assessment. – In: Keith, L. H. (ed.) Principles of Environmental Sampling. American Chemical Society, Washington, DC.

[53] Journel, A. G., Alabert, F. (1989): Non-Gaussian data expansion in the earth sciences. – Terra Nova 1: 123-134.

[54] Karim, Z., Qureshi, B. A., Mumtaz, M. (2015): Geochemical baseline determination and pollution assessment of heavy metals in urban soils of Karachi, Pakistan. – Ecological Indicators 48: 358-364.

[55] Kelepertis, E. (2014): Accumulation of heavy metals in agricultural soils of Mediterranean: insight from Argolida basin, Peloponnese, Greece. – Geoderma 221: 82-90.

[56] Kiereczak, J., Pedziwiatr, A., Waroszweski, J., Modelska, M. (2016): Mobility of Ni, Cr, and Co in serpentine soils derived on various ultrabasic bedrocks under temperature climate. – Geoderma 268: 78-91.

[57] Kowalska, J., Mazurek, R., Gasiorek, M., Setlak, M., Zaleski, T., Waroszweski, J. (2016): Soil pollution indices conditioned by medieval metallurgical activity: a case study from Krakow (Poland). – Environmental Pollution 218: 1023-1036.

[58] Kowalska, J. B., Mazurek, R., Gasiorek, M., Zaleski, T. (2018): Pollution indices as useful tools for the comprehensive evaluation the degree of soil contamination. A review. – Environment Geochemical Health 40: 2395-2420.

[59] Lee, C. S. I., Li, X., Shi, W., Cheung, S. C. N., Thornton, I. (2006): Metal contamination in urban, suburban, and country park soils of Hong Kong: a study based on GIS and multivariate statistics. – Science Total Environment 356: 45-61.

[60] Li, J., Heap, A. (2008): A review of spatial interpolation methods for environmental scientists. – Geoscience Australia 137: Record 2008/23.

[61] Li, J., Heap, A. (2011): A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. – Ecological Information 6: 228-241.

[62] Li, J., Heap, A. D. (2014): Spatial interpolation methods applied in the environmental sciences: a review. – Environmental Modelling & Software 53: 173-189.

[63] Li, Z., Ma, Z., Kuijp, T. J. V. D., Yuan, Z., Huang, L. (2014): A review of soil heavy metal pollution from mines in China: pollution and health risk assessment. – Science of Total Environment 468-469: 843-853.

[64] Li, M., Yang, W., Sun, T., Jin, Y. (2016): Potential ecological risk of heavy metal contamination in sediments and macrobenthos in coastal wetlands induced by freshwater releases: a case study in the Yellow River Delta, China. – Marine Pollution Bulletin 103: 227-239.

[65] Lin, W. C., Lin, Y. P., Wang, Y. C. (2016): A decision making approach for delineating sites, which are potentially contaminated by heavy metals via joint simulation. – Environment Pollution 2011: 98-110.

[66] Lu, A., Wang, J., Qin, X., Wang, K., Han, P., Zhang, S. (2012): Multivariate and geostatistical analyses of the spatial distribution and origin of heavy metals in the agricultural soils in Beijing, China. – Science Total Environment 425: 66-74.

[67] Maas, S., Scheifler, R., Benslama, M., Crini, N., Lucot, E., Brahmia, Z., Benayacoub, S., Giradoux, P. (2010): Spatial distribution of heavy metal concentrations in urban, suburb and agricultural soils in a Mediterranean city of Algeria. – Environmental Pollution 158: 2294-2301.

[68] Meirvenne, M. V., Goovaerts, P. (2001): Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. – Geoderma 102: 75-100.

[69] Mico, C., Recatala, L., Peris, M., sanchez, J. (2006): Assessing heavy metal sources in agricultural soils of an European Mediterranean area by multivariate analysis. – Chemosphere 65: 863-872.

[70] Mihailovic, A., Budinki-Petkovic, L., Popov, S., Ninkov, J., Vasin, J., Ralevic, N., Vasic, M. V. (2015): Spatial distribution of metals in urban soil of Novi Sad, Serbia: GIS based approach. – Journal of Geochemical Exploration 150: 104-114.

[71] Ministry of Housing (2011): Spatial Planning and Environment Protection, Directorate-General for Environmental (625). – Department of Soil Protection, New Dutch List.

[72] Moore, F., Sheykhi, V., Salori, M., Bagheri, A. (2016): Soil quality assessment using GIS-based approach. – Journal of Geochemical Exploration 150: 104-114.

[73] Mostert, M. M., Ayoko, G. A., Kokot, S. (2010): Application of chemometrics to analysis of soil pollutants. – Trends Analytical Chemistry 29: 430-445.

[74] Müller, G. (1969): Index of geoaccumalation in sediments of the Rhine River. – Geojournal 2: 108-118.

[75] Neufeld, C. T. (2005): Guide to Sampling. Guide Book Series No. 2. – Centre for Computational Geostatistics, University of Alberta, Edmonton.

[76] Oliver, M. A. (ed.) (2010): Geostatistical Applications in Precision Agriculture. – Springer, Dordrecht.

[77] Oliver, M. A., Webster, R. (2014): A tutorial guide to geostatistics: computing and modelling variogram and Kriging. – Catena 113: 56-69.

[78] Panagos, P., Liedekerke, M. V., Yigini, Y., Montanarella, L. (2013): Contaminated sites in Europe: review of the current situation based on data collected through a European network. – Journal of Environmental and Public Health 2013: 1-11.

[79] Pejman, A., Gholamrez, N. B., Saeedi, M., Baghvanda, A. (2015): A new index for assessing heavy metals contamination in sediments: a case study. – Ecological Indicators 58: 365-373.

[80] Qu, M., Li, W., Zhang, C. (2013): Assessing the risk costs in delineating soil nickel contamination using sequential Gaussian simulation and transfer functions. – Ecological Informatics 13: 99-105.

[81] Reimann, C., de Caritat, P. (2017): Establishing background variation and threshold values for 59 elements in Australian surface soil. – Science Total Environment 578: 633-648.

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

[83] Romic, M., Romic, D. (2003): Heavy metal distribution in agricultural top soil in urban area. – Environmental Geology 43: 795-805.

[84] Rossi, M. E., Deutsch, C. V. (2014): Mineral Resource Estimation. – Springer Science Business Media, Dordrecht.

[85] Saby, N., Arrouays, D., Boulonne, L., Jolivet, C., Pochot, A. (2006): Geostatistical assessment of Pb in soil around Paris, France. – Science Total Environment 367: 212-221.

[86] Sayadi, M. H., Shabani, M., Ahmadpour, N. (2015): Pollution index and ecological risk of heavy metals in the surface soils of Amir-Abad Area in Birjand City, Iran. – Health Scope 4: 121-137.

[87] Schloeder, C. A., Zimmerman, N. E., Jacobs, M. J. (2001): comparison of methods for interpolating soil properties using limited data. – Soil Society of American Journal 65: 470-479.

[88] Shan, Y., Tysklind, M., Hao, F., Ouyang, W., Chen, S., Lin, C. (2013): Identification of sources of heavy metals in agricultural soils using multivariate analysis and GIS. – Journal Soils Sediments 13: 720-729.

[89] Shao, X., Huong, B., Zhao, Y., Sun, W., Gu, Z., Qian, W. (2014): Impacts of human activities and sampling strategies on soil heavy metal distribution in a rapidly developing region of China. – Ecotoxicological Environment Safety 104: 1-8.

[90] Sinclair, A. J., Blackwell, G. H. (2002): Applied Mineral Inventory Estimation. – Cambridge University Press, Cambridge.

[91] Soares, A. (2001): Direct sequential simulation and cosimulation. – Mathematical Geology 33: 911-26.

[92] Stahl, K., Moore, R. D., Floyer, J. A., Asplin, M. G., McKendry, I. G. (2006): Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. – Agricultural and Forest Meteorology 139: 224-236.

[93] Steiger, V. B., Webster, R., Schulin, R., Lehmann, R. (1996): Mapping heavy metals in polluted soil by disjunctive Kriging. – Environmental Pollution 94: 205-215.

[94] Su, C., Jiang, L., Zhang, W. (2014): A review on heavy metal contamination in the soil worldwide: situation, impact and remediation techniques. – Environmental Sceptics and Critics 3: 24-38.

[95] Sun, C., Bi, C. Z., Chen, Z. I., Wang, D. Q., Zhang, C., Sun, Y. D., Yu, Z. J., Zhou, D. (2010): Assessment of environmental quality of heavy metals in agricultural soils of Chongming Island, Shanghai City. – Journal Geographic Science 20: 135-147.

[96] Sutherland, R. A. (2000): Bed sediment-associated trace metals in an urban stream, Oahu, Hawaii. – Environmental Geology 39: 611-627.

[97] Thabet, A. M., Mohamed, A. K. M., Ragab, R., Mahmoud, A. G. (2014): Application of pollution indices for evaluation of heavy metals in soil close to phosphate fertilizer plant, Assiut, Egypt. – Assiut University Bulletin Environment Research 17: 45-55.

[98] United States Environmental Protection Agency (USEPA) (2011): Clean-up Regulations and Standards. – USEPA, Washington, DC.

[99] Varol, M. (2011): Assessment of heavy metal contamination in sediments of the Tigris River (Turkey) using pollution indices and multivariate statistical techniques. – Journal of Hazardous Materials 195: 355-364.

[100] Verly, G. (2005): Grade control classification of ore and waste: a critical review of estimation and simulation based procedures. – Mathematical Geology 37: 451-475.

[101] Wang, H., Lu, S. (2011): Spatial distribution, source identification and affecting factors of heavy metals contamination in urban-suburban soils of Lishui city, China. – Environment Earth Science 64: 1921-1929.

[102] Wang, H., Liu, G., Gong, P. (2005): Use of cokriging to improve estimates of soil salt solute spatial distribution in the yellow river delta. – Acta Geographica Sinica 60: 511-518.

[103] Webster, R., Oliver, M. A. (2007): Geostatistics for Environmental Scientists. 2nd Ed. – John Wiley & Sons, Chichester.

[104] Wu, J., Norvell, W. A., Welch, R. M. (2006): Kriging only highly skewed data for DTPA-extractable soil Zn with auxiliary information for pH and organic carbon. – Geoderma 134: 187-199.

[105] Wu, J., Teng, Y., Lu, S., Wang, Y., Jiao, X. (2014): Evaluation of soil contamination indices in a mining area of Tiangxi, China. – PLoS ONE 9: 1-14.

[106] Xie, Y., Chen, T. B., Lei, M., Yang, J., Guo, Q. J, Song, B., Zhou, X. Y. (2011): Spatial distribution of heavy metal pollution estimated by different interpolation methods: accuracy and uncertainty analysis. – Chemosphere 82: 468-476.

[107] Xu, G., Liu, j., Pei, S., Hu, G., Kong, X. (2015): Geochemical background and ecological risk of heavy metals in surface sediments from the west Zhoushan Fishing Ground of East China Sea. – Environmental Science and Pollution Research 22: 20283-20294.

[108] Yang, Y. B., Sun, L. B. (2009): Status and control countermeasures of heavy metal pollution in urban soil. – Environmental Protection Science 35: 79-81.

[109] Zhang, C. (2006): Using multivariate analysis and GIS to identify pollutants and their spatial patterns in urban soils in Galway, Ireland. – Environmental Pollution 142: 501-511.

[110] Zhang, R., Rahman, S., Vance, G. F., Munn, L. C. (1995): Geostatistical analyses of trace elements in soils and plants. – Soil Science 159: 383-390.

[111] Zhang, S., Shao, M., Li, D. (2017): Prediction of soil moisture scarcity using sequential Gaussian simulation in an arid region of China. – Geoderma 295: 119-128.

[112] Zhou, J., Feng, K., Pei, Z., Meng, F., Sun, J. (2016): Multivariate analysis combined with GIS to source identification of heavy metals in soils around an abandoned industrial area. – Eastern China Ecotoxicology 25: 380-388.

[113] Zimmerman, D., Powlik, C., Ruggles, A., Armstrong, M. P. (1999): An experimental comparison ordinary and universal Kriging and inverse distance weighting. – Mathematical Geology 31: 375-390.