Journal of Hazardous Materials 180 (2010) 542–551

Contents lists available at ScienceDirect

Journal of Hazardous Materials

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

Defining and modeling the soil geochemical background of heavy metals from the Hengshi River watershed (southern China): Integrating EDA, stochastic simulation and magnetic parameters

Xu Zhou, Beicheng Xia*

School of Environmental Science and Engineering, Sun Yat-Sen University, Guangzhou 510275, PR China

ARTICLE INFO

Article history:
Received 10 January 2010
Received in revised form 15 April 2010
Accepted 16 April 2010
Available online 22 April 2010

Keywords:
Soil geochemical background
Potentially toxic heavy metals
Exploratory data analysis
Sequential indicator co-simulation
Rock magnetic parameter

ABSTRACT

It is crucial to separate the soil geochemical background concentrations from anthropogenic anomalies and to provide a realistic environmental geochemical map honoring the fluctuations in original data. This study was carried out in the Hengshi River watershed, north of Guangdong, China and the method proposed combined exploratory data analysis (EDA), sequential indicator co-simulation (SIcS) and the ratio of isothermal remnant magnetization ($S_{100} = -\mathrm{IRM}_{-100\,\mathrm{mT}}/\mathrm{SIRM}$). The results showed that this is robust procedure for defining and mapping soil geochemical background concentrations in mineralized regions. The rock magnetic parameter helps to improve the mapping process by distinguishing anthropogenic influences.

In this study, the geochemical backgrounds for four potentially toxic heavy metals (copper 200 mg/kg; zinc 230 mg/kg; lead 190 mg/kg and cadmium 1.85 mg/kg) Cu, Zn and Cd exceeded the soil Grade II limits (for pH < 6.5) from the Chinese Environmental Quality Standard for Soils (GB 15618-1995) (EQSS) which are 100, 200, 250 and 0.3 mg/kg for Cu, Zn, Pb and Cd, respectively. In particular, the geochemical background level for Cd exceeds standard six times. Results suggest that local public health is at high-risk along the riparian region of the Hengshi River, although the watershed ecosystem has not been severely disturbed.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The effects of anthropogenic activities on ecosystem functions and risks to human health have highlighted the need for measuring the geochemical background [1,2]. Therefore, as the main task of project 360 within the International Geosciences Program (IGCP), the geochemical background has become a priority in environmental changes issues [3,4]. The term "geochemical background" has yet to be clearly defined. However, it has been agreed that the geochemical background provides fundamental guidance for monitoring environmental changes caused by human activities on the surface environment [5–7].

A geochemical background characterized by spatio-temporal variability can be affected by the basic geology and soil type in the secondary environment [8]. Background data, particularly in mineralized regions, are represented by individual probabilistic patterns due to either natural enrichment or human activities. Thus, it is a complex task to correctly define the geochemical background and to more realistically delineate the fluctuations in the original data. Conventional statistical methods including mean calculations and probability distributions are no longer useful for dealing with spatial variables in a complex environment [9]. Both exploratory data analysis (EDA) [10–12] and spatial analysis have been recommended as an effective tool to map the geochemical background and allow further analysis of the processes in complex environments [13,14].

EDA techniques combining graphs, such as histograms, boxplots and normal quantile–quantile (Q–Q) plots, are helpful to allow rapid insight into the distributions and features of the data. It is appropriate for identifying anthropogenic influences because it eliminates some of the geogenic anomalies. The EDA technique was tested by Bounessah and Atkin [15], Reimann et al. [10], Panno et al. [11] and others and proved to be effective and robust for studying the geochemical characteristics of sediment, soil and ground water.

Spatial analysis techniques, like geostatistical methods, have been effectively used to investigate the spatial variability of environmental variables and incorporate this information into maps [16]. The procedure integrating exploratory data analysis (EDA) and the ordinary kriging (OK) has been applied to the regional back-

* Corresponding author at: College of Environmental Science and Engineering, Sun Yat-sen University, 135 Xin Gang Xi Road, Guangzhou 510275, Guangdong, PR China. Tel.: +86 20 8411 4591; fax: +86 20 8411 0267.
E-mail addresses: xuzhou228@yahoo.cn (X. Zhou), xiabch@mail.sysu.edu.cn (B. Xia).

0304-3894/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2010.04.068

ground levels of Ag, As, Cd, Cu, Hg, Ni, Pb, TI and Zn in soils from Catorce-Matehuala (Mexico) [17]. However, the kriging estimate is rarely met for environmental attributes, which typically display highly skewed histograms, because it relies on assumptions of normality in the distribution of prediction errors and homoscedasticity [18].

At present, sequential simulation algorithms of the stochastic simulation technique, such as the sequential Gaussian simulation (SGS), sequential indicator simulation (SIS) and direct sequential simulation (DSS) are used widely to delineate the spatial distribution and uncertainty of pollutants in a heterogeneous medium [19–22]. Generally, the distribution of heavy metal concentrations has the right skewed. DSS and SIS are more suited to application, as SGS requires strong assumptions of multi-Gaussian distribution which cannot be easily checked [20]. SIS does not require any assumptions of the shape of the conditional distribution and its mapping can be achieved with much flexibility in categorical data [23,24].

This paper will collocated co-simulations for soil geochemical background distributions of potentially toxic heavy metals (PTHMs) using the sequential indicator co-simulation (SIcS). In the upstream of the study area there lays Dabaoshan mine with a large rich polymetallic ore. This makes it difficult to separate the soil geochemical background from human activities in mineralized regions. In this paper, the ratio of isothermal remnant magnetization ($S_{100} = -\mathrm{IRM}_{-100\,\mathrm{mT}}/\mathrm{SIRM}$) is introduced to co-simulate the geochemical background of PTHMs. The environmental implication of this ratio is that the ratio increases as anthropogenic influence is enhanced. The range of S_{100} is 0–1. Robinson [25] indicated that the dominance of ferrimagnetic materials caused by anthropogenic influence when S_{100} is more than 0.7. And, Lecoanet et al. [26], Lu and Bai [27] and others had the agreement with the tested parameter in different regions.

Dabaoshan mine is a large sulfide deposit containing rich iron, copper, lead and zinc in China. Previous studies indicated that Cu, Zn, Pb and Cd content exceeded the soil Grade II limits (for pH < 6.5) from the Chinese Environmental Quality Standard for Soils (GB 15618-1995) (EQSS) after detecting Cr, Cd, Co, Ni, Cu, Zn, As, Sb, Hg and Pb content in the topsoil and water about the Dabaoshan mine [28,29]. Accordingly, the aims of this study are: (1) to define and visualize the regional soil geochemical backgrounds of copper, zinc, lead and cadmium based on EDA and SIcS; (2) to try and use the rock magnetic parameter to improve the mapping process for soil geochemical backgrounds in mineralized regions; and (3) to detect potential hazardous areas to give regional early warning information on human health and eco-environmental risks.

2. Materials and methods

2.1. Site description

The study area (E113°37′33″–113°55′58″, N24°15′26″–24°39′37″), with a total area of 1017.9 km^2, is located on the boundary of Qujiang County and Wengyuan County in Guangdong Province, China (Fig. 1). It includes the Hengshi River watershed and three additional watersheds—the Chuandu River watershed, Tielong watershed and Xitou watershed. These are considered a total geographical system due to the continuous geologic units and land use. The site of interest is characterized by typical humid subtropical conditions. The average annual temperature is 20.3 °C and precipitation is approximately 1762 mm. The weathering of surface rock is severe and the soil pH is lower than 6.5.

In the study area, irrigated agriculture and ore deposit mining are the dominant land use patterns. The minerals mainly consist of pyrite, pyrrhotite, chalco pyrite, and minor amounts of limonite, chalcocite, galena, sphalerite, calaverite and native bismuth. Since the 1970s, mining on Dabao Mountain has been the most important pollution source in the whole watershed. Acid mine drainage (AMD) has run into the Hengshi River, which is the primary water supply for agriculture and human requirements in the alluvial plain. Several high-risk sites influenced by heavy metals have been detected in the study area [30] and have posed adverse effects to local public health, such as cancer, neuropathies and growth problems [31,32].

2.2. Sampling and analysis

2.2.1. Sampling

The main soil types in the study area are red earth and mountain yellow earth. Sampling sites were selected according to soil type and land use, and come from eight groups: farmland, forest, grassland, garden plots, industrial mine land, roads, habitable land, land immediately adjacent to water, and all other types of land. The uncertainty of sampling is caused by gaps between the sampled sites. It is necessary that the interval of measurement be <5 km to ensure the enough high accuracy of spatial analysis [33,34]. And, the maximum sampling grid with 16 km^2 is required by the investigation of regional soil geochemical background of Multi-purpose Investigation and Regulation of Regional Geochemistry (Technical Standard of Geological Investigation in China Geological Survey (DD2005-01)). Accordingly, the average sampling grid was designed a square mesh with sampling points at approximately 2.5 km × 2.5 km with the actual sampling sites randomly selected near the target sampling sites in each grid cell. The samples were located using hand-held GPSs, ASTER Orthograph satellite images (data: 2006-10; row/path: 122/42, 43; pixel: 15 m), and topographic maps and geological profile maps at scales of 1:10,000 and 1:200,000, respectively. Sixty-seven samples were collected from the A horizon (top 20 cm) in the Hengshi River watershed. The sampling of sites could not be of uniform design because some are inaccessible.

Each sample was collected with a stainless steel sampler and a wooden shovel, and stored in a plastic bag. The soil samples were air-dried at room temperature, grinded, and then sieved to retain the <1.7 mm fraction using a nylon sieve. Each sample (typically 3–9 g, average weight 4.65 g) was packed into an 8 cm^3 nonmagnetic plastic box for a rock magnetic test. A sub-sample (100 g) of each soil sample was ground to <0.15 mm with a nylon sieve for heavy metal analysis.

2.2.2. Testing of total heavy metal concentrations

Total heavy metal concentrations were measured according to the standard method [35]. The soil samples for heavy metal analysis were digested by an aqua regia (HNO_3–HCl) acid digestion in a microwave oven. Total concentrations of Cu, Pb and Zn were measured by Inductively Coupled Plasma Optical Emission Spectrometer (ICP-OES) and total Cd concentration was measured by Graphite Furnace Atomic Absorption Spectrometry (GFAAS).

Reagent blanks were used to reduce the systematic error. A standard sample was used for every ten tested samples. The accuracy was evaluated by a standard reference material of a subtropical red soil from the Hengshi River watershed. The results of recovery were 85.83% (Cu), 124.24% (Pb), 88.89% (Zn) and 118.08% (Cd), respectively. The errors of the estimate for the measured metals were determined by the relative standard deviation (S.D. < 5%) based on three replicates of one sample randomly chosen.

2.2.3. Rock magnetic measurements

The magnetism of the samples was acquired in magnetic fields of 1 Tesla (T), followed by −100 milliTesla (mT) (reverse) exerted with a speed as constant as possible at room temperature using an ASC

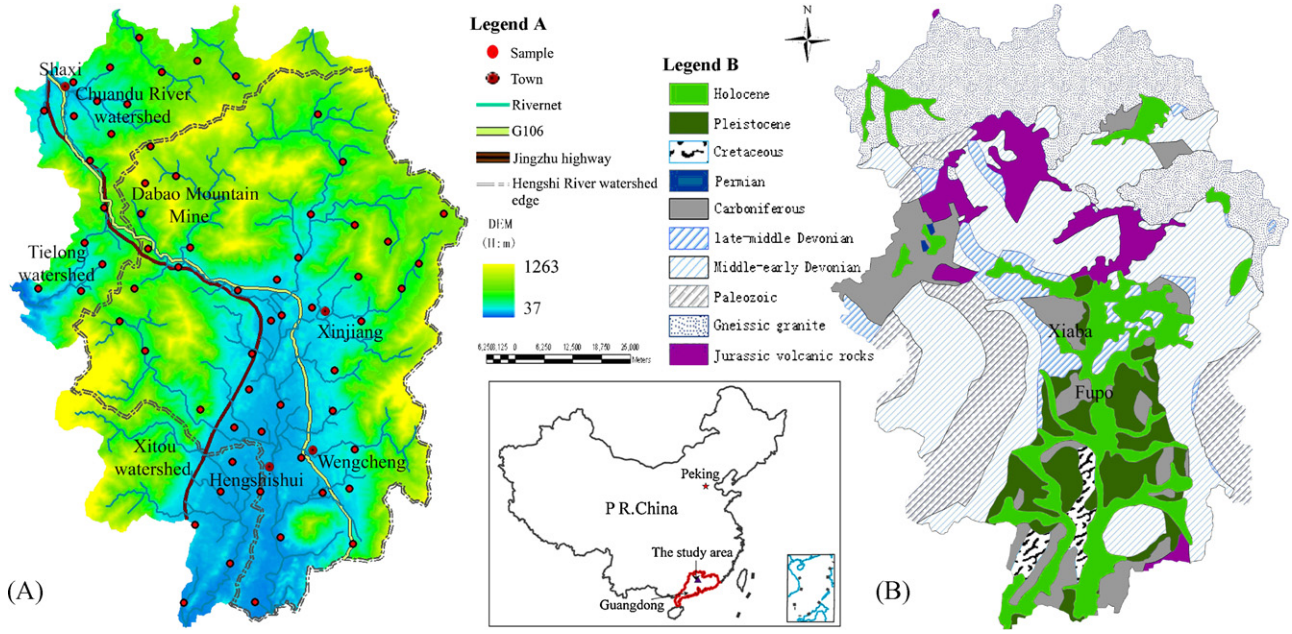

Fig. 1. Digital terrain elevation map of sampling sites (A) and geological profile map (B) of the study area.

IM10-30 pulse magnetizer. The values of isothermal remnant magnetization (IRM) were obtained along the same axial direction using a 2G Enterprises DC-SQUID cryogenic magnetometer housed in a magnetically shielded room (<500γ) at the Paleomagnetic Lab of the Key Laboratory of Marginal Sea Geology, South China Sea Institute of Oceanology, Chinese Academy of Science, Guangzhou, China. In this case, the IRM acquired at 1 T was regarded as the saturation isothermal remnant magnetization (SIRM) because most magnetic minerals are saturated in this field. S_{100}($S_{100} = -\mathrm{IRM}_{-100\,\mathrm{mT}}/\mathrm{IRM}_{1\,\mathrm{T}}$) was calculated [36].

2.3. EDA method

EDA was applied to Descriptive Statistics models using the SPSS software (version 17.0) for Windows. Graphs of experimental data contain histograms, density markers, distribution curves and skewness, as well as descriptions of summary statistics. Outliers were inspected through boxplots, which include a non-outlier range, lower quartile (25th percentile), median, upper quartile (75th percentile), outliers, and extreme values. Outliers exist outside the whiskers. Normal Q–Q plot was applied to look for the concentration intervals which imply the likely presence of different processes or multiple populations.

2.4. SIcS method

SIcS makes no prior assumptions about the shape or type of the conditional distribution. The results of the indicator simulation can be improved by integrating additional information. The soft data has to be coded into prior local probability values, and indicator kriging (IK) can be used to integrate that information into a posterior probability value [37]. In this study the first variable simulated was the PTHMs as hard data. S_{100} was the next variable to be co-simulated as soft data.

2.4.1. Indicator coding

The environmental information of continuous soil property z at n locations u_α, $z(u_\alpha)$ ($\alpha = 1, \ldots, n$) includes:

(1) The local hard indicator data $i(u_\alpha; z_k)$ originated from the local hard data (PTHMs) $z(u_\alpha)$:

$$i(u_\alpha; z_k) = \{1 \text{ if } z(u_\alpha) \le z_k;\ 0 \text{ otherwise}\} \quad (1)$$

where z_k is a discretization of K threshold values in the range of values of $z(u_\alpha)$.

(2) The local soft indicator data $y(u_\alpha; z_k)$ originated from ancillary information (S_{100}), providing prior probabilities about the value $z(u_\alpha)$:

$$y(u_\alpha; z_k) = \mathrm{Prob}\{Z(u_\alpha) \le z_k | \text{local ancillary information}\} \quad (2)$$

(3) Global prior information common to all locations u within the area A:

$$F(z_k) = \mathrm{Prob}\{Z(u) \le z_k\}, \quad \forall u \in A \quad (3)$$

2.4.2. Updating

To provide a model of uncertainty at location u, the local prior cumulative distribution function (cdf) must be updated into a posterior cdf using information supplied by neighboring local prior cdfs [37]:

$$[\mathrm{Prob}\{Z(u) \le z_k \,|\, n + n'\}]^*_{IK} = \lambda_0(u)F(z_k) + \sum_{\alpha=1}^{n} \lambda_\alpha(u; z_k) i(u_\alpha; z_k) + \sum_{\alpha'=1}^{n'} \nu_{\alpha'}(u; z_k) y(u'_\alpha; z_k) \quad (4)$$

where $\lambda_\alpha(u; z_k)$ is the weight attached to the n neighboring hard indicator data from Eq. (1), $\nu_{\alpha'}(u; z_k)$ are the weights attached to the n' neighboring soft indicator data from Eq. (2) and λ_0 is the weight attributed to the global prior cdf.

2.4.3. Markov-type approximation

In order to simplify the variogram inference of two auto covariance and one cross-covariance at each cutoff value z_k, the Markov–Bayes model is used.

$$C_{IY}(h, z_k) = B(z_k)C_I(h, z_k) \quad \forall h$$

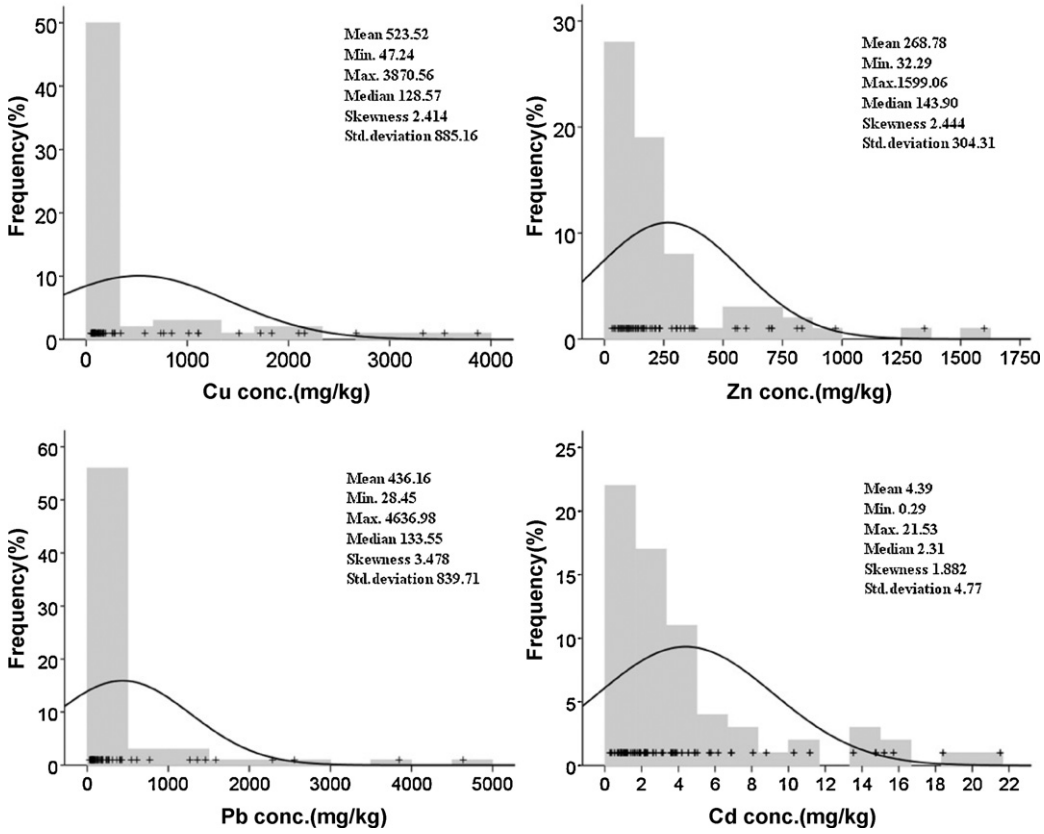

Fig. 2. Experimental data patterns of histograms, density markers, distribution curves and summary statistics of the concentrations of copper, zinc, lead and cadmium in soils.

$$C_Y(h, z_k) = \begin{cases} |B(z_k)|C_I(h, z_k), & \forall h = 0 \\ B(z_k)^2 C_I(h, z_k), & \forall h > 0 \end{cases} \quad (5)$$

The coefficients $B(z_k)$ are obtained from calibration of the soft y data to the hard z data.

2.4.4. SIS

SIS includes all available data within a given neighborhood of the location u, including the original data and all previously simulated values. In this study, the cutoff value z_k was set to be the geochemical backgrounds of the four heavy metals and their Grade II limits (pH < 6.5) from the Chinese EQSS (GB 15618-1995). For each heavy metal concentration cutoff value, z_k the experimental semi-variogram of the indicator code was computed and this was modeled using a linear combination of permissible semi-variogram models. A random path visiting each node of the regularly spaced grid covering the study area was established. At each un-sampled location, the procedure used was as follows:

(1) For a location u in the random path, each heavy metal ccdf value was estimated as a linear combination of neighboring indicator data by indicator kriging. ccdf is the probability that the heavy metal concentration was greater than the soil geochemical background of each heavy metal and their soil Grade II limits (pH < 6.5) from EQSS.
(2) The order relation deviations of the obtained probabilities were corrected and a continuous model for the prior ccdf for each

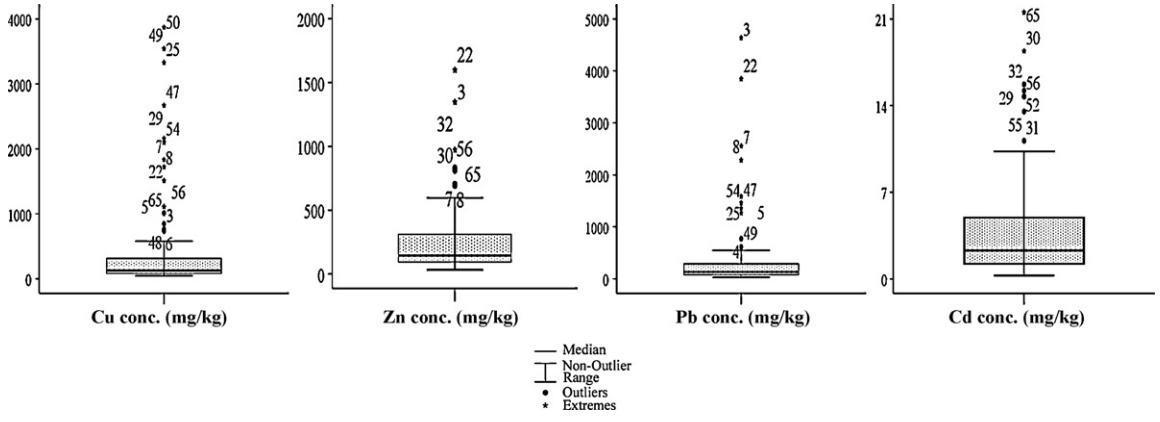

Fig. 3. Boxplots for the concentration of copper, zinc, lead and cadmium in soils.

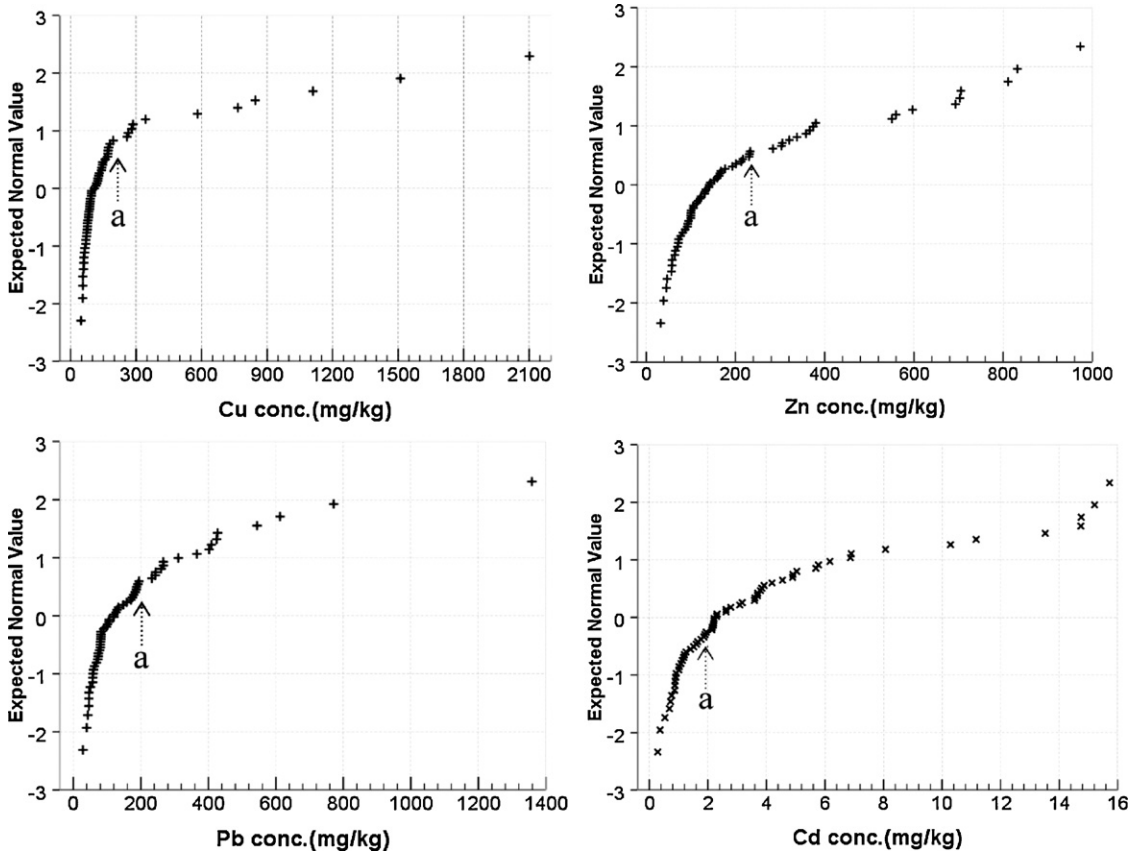

Fig. 4. Normal Q–Q plots of the eliminated outliers for copper, zinc, lead and cadmium (arrow a: the first bend of the slope).

heavy metal concentration at location u was constructed by interpolating or extrapolating ccdf values.

(3) A simulated value for each heavy metal concentration was randomly drawn from the prior ccdf for each heavy metals concentration at location u.

(4) The indicator code of the simulated values at location u was added into the prior ccdf modeling for the next location $u+1$.

Along the random path, repeating steps (1)–(4) until all nodes are visited and each node has received a simulated value will obtain a realization of the SIS.

Using the Markov–Bayes algorithm amounts to an indicator co-kriging, where S_{100} data cross-covariance is calibrated from the hard indicator covariance models. In all other aspects the Markov–Bayes algorithm is similar to SIS.

Repeating the SIcS L times, each time using a different path to visit all nodes of the grid defined over the study area, will result in L equiprobable realizations. Six hundred repeats of the simulation (the size of the grid cell is 15 m × 15 m) were carried out in this study and the search radii were set as ranges of the semi-variogram models. In this study, the experimental variogram and covariance models were obtained by using the software package GS+ (Gamma Design). The "Cosisim" program in the SGeMS (Stanford Geostatistical Modeling Software) software was used to perform the SIcS.

2.5. Uncertainty analysis of simulated results

The main objectives of this study are to define soil geochemical backgrounds and to detect areas that are potential hazardous to human health and the eco-environment. Consequently, the lowest threshold was chosen from soil geochemical background for each heavy metal, with its soil Grade II (pH < 6.5) limit from the Chinese EQSS used as the critical limit. The probability that the unknown heavy metal concentration $z(i)$ at i is greater than its critical limit (z_{cl}), denoted by Prob[$z(i) > z_{cl}$], can be calculated from the following equation:

$$\text{Prob}[z(i) > z_{cl}] = \frac{n(i)}{600} \quad (6)$$

The SIS was carried out 600 times and $n(i)$ is the number of realizations that show the heavy metal concentration as higher than the critical limit at node i. Eq. (6) can be used to assess the reliability of the geochemical maps for the four heavy metals at a given critical probability.

3. Results and discussion

3.1. Defining the soil geochemical background of PTHMs

3.1.1. Experimental data patterns

Environmental geochemical probability patterns are close to lognormal (positive skewness) for outliers from either natural enrichment or human activities. These outliers usually exist in mineralized regions or are concentrated in contaminated sites. Histograms provide a graphical data summary depending on the

Table 1
Comparison of concentrations of the four heavy metals (mg/kg) in soils among natural backgrounds (the Chinese EQSS), Grade II limits (the Chinese EQSS) and geochemical background of the watershed.

Heavy metals	Cu	Zn	Pb	Cd
Soil natural background	35	100	35	0.20
Grade II (pH < 6.5 for soils)	100	200	250	0.30
Soil geochemical background	200	230	190	1.85

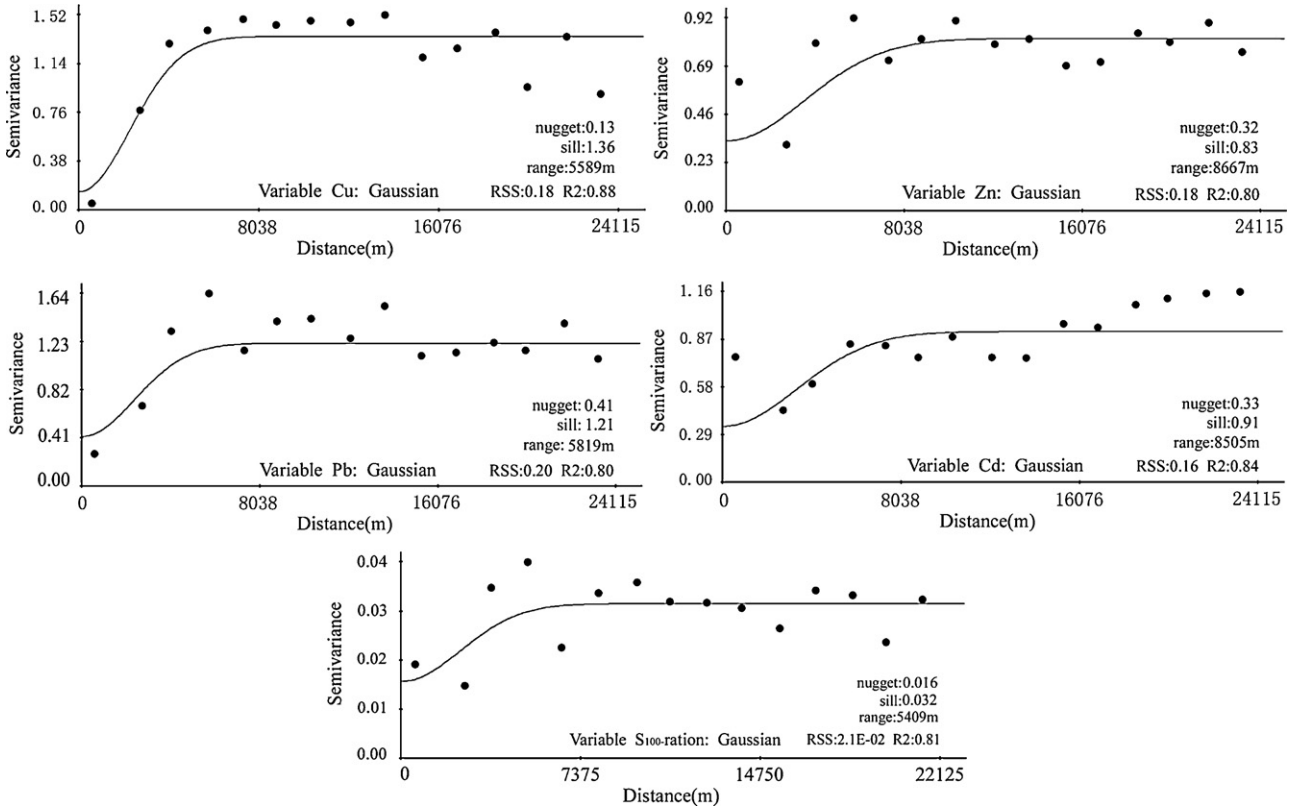

Fig. 5. Variograms of Log-transformed soil Cu data, Zn data, Pb data, Cd data and S_{100} raw data.

data structure without any assumptions [38]. Before any data analysis is applied, an insight to the inherent data patterns of the experimental heavy metal concentrations is displayed in Fig. 2. Each graph shows a histogram, density marker, distribution curve and summary statistics of the heavy metal concentrations.

Results show that the outliers as indicators of unusual processes more easily recognized in the density marker. One of the important features of the histograms is that the four heavy metals are right-skewed. Skewness characterizes the degree of asymmetry of a distribution around its mean. Generally, high skewness value indicates that the majority of its values are concentrated on the low value end representing the relatively uncomplicated environment in soils [39]. The skewness feature of Pb concentration suggests that the majority of the samples at lower values and that Zn and Cu concentration have the similar distributions. The pattern of Cd concentration show that different processes, such as lithological processes or contamination, lead to elevated metal concentration. The distribution reveals that anthropogenic pollution of Cd is more difficult to be separated from natural process than those of Pb, Zn and Cu due to the coexistence of different processes.

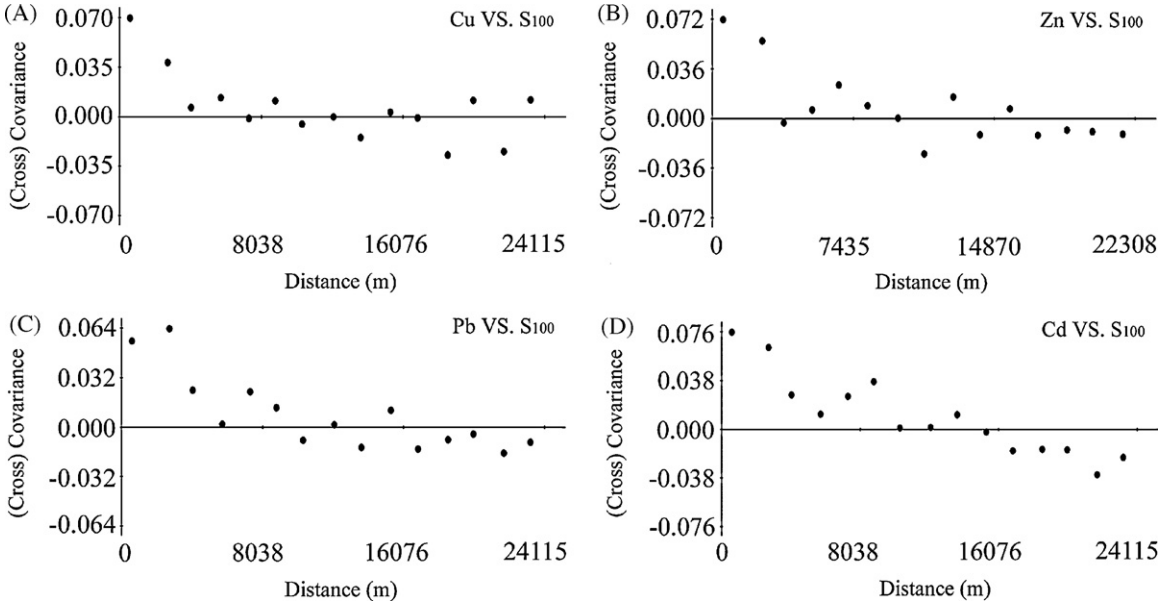

Fig. 6. Experimental indicator (cross) covariance for the critical limit of each heavy metal in soils.

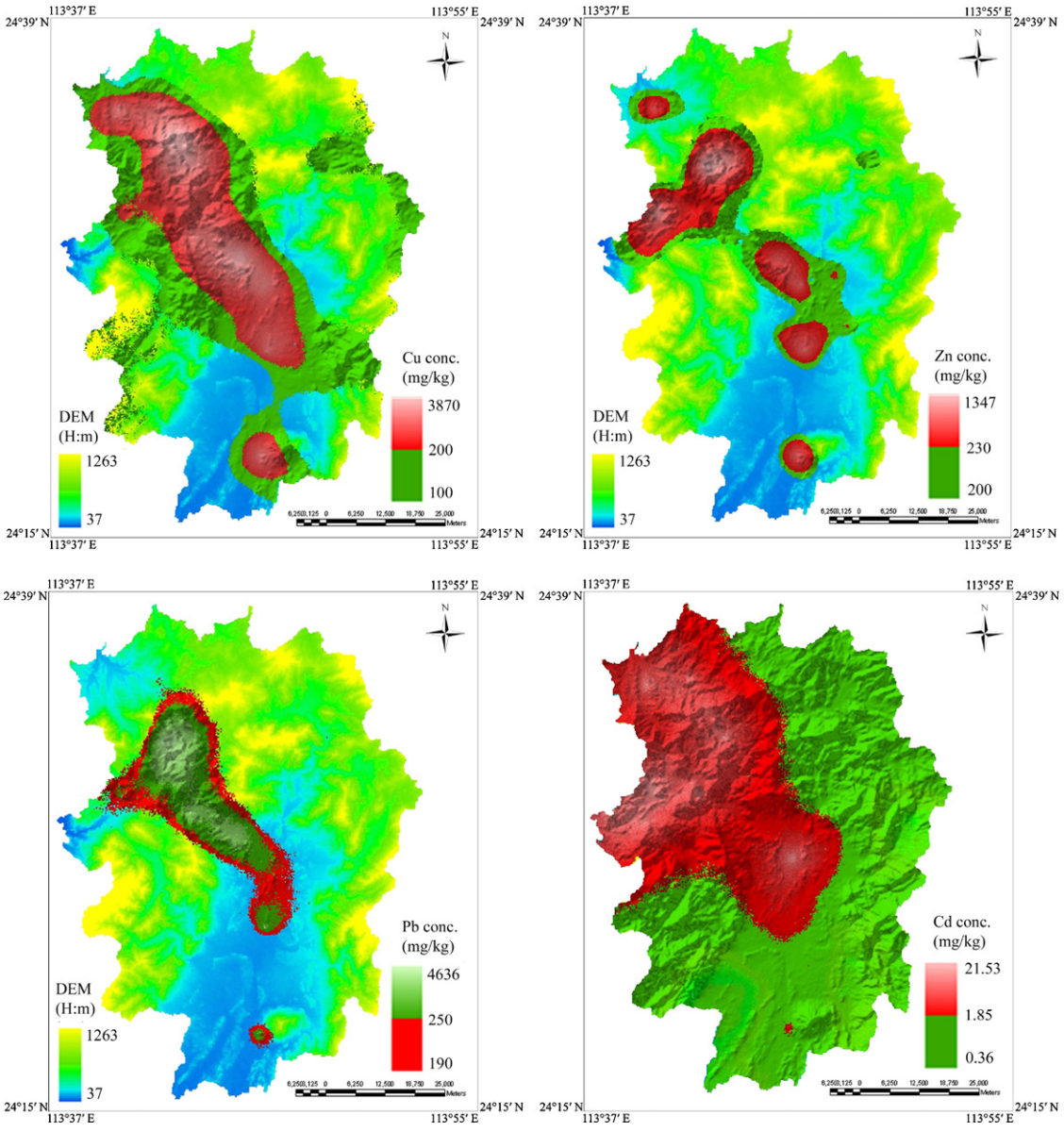

Fig. 7. Estimates of the critical limits of 95% confidence interval for soil geochemical background and Grade II limit of each heavy metal.

3.1.2. Detection of anomalous values

A graphical inspection of geochemical data is necessary before the initial data analysis because environmental geochemical data do not usually exhibit Gaussian pattern. A simulation of boxplots provides the explanation for observed behavior and is more appropriate for identifying anthropogenic influences because it eliminates some of the geogenic anomalies [15]. Outliers of the four heavy metal concentrations are identified outside the whiskers (the whiskers represent the 10th and 90th percentiles) of the boxplots. In Fig. 3, all the anomalies (outliers) are clearly observed outside the whiskers. The outliers are situated in the mining areas on mountain and in polluted sites on the alluvial plain. The 'hot spots' are located in the Dabaoshan mine (samples 3, 22 and 30), the Tielong watershed (samples 26 and 54), the riparian zone of the Hengshi River (samples 5–8) and in the Xiaba-Fupo zone (samples 47 and 65). The boxplot detected distinctly the peculiarities of the anomalous values from the raw data and provide more details of data structure than those of the histograms. It should be mentioned that the outliers should be removed from the datasets in the following statistical analysis. The anomalies may cause biased results and make statistical results unreliable, as they have a strong effect on the overall feature of the datasets. As demonstrated by Reimann et al. [10], the outlier limits of the boxplot are not necessarily symmetric around the median. It is more realistic for the right-skewed geochemical background than the assumption of symmetry.

3.1.3. Soil geochemical backgrounds for Cu, Zn, Pb and Cd

The geochemical background threshold can be defined by using a Q–Q plot prior to inspecting a clear and detailed visualization of the experimental data distribution, which also provides understanding of the probability features [10]. The measured values are plotted on the *X* axis and the expected values under a normal distribution are plotted on the *Y* axis. This procedure provides a powerful visualization of the data which shows breaks and inflection points that imply the likely presence of different processes [39]. Additionaly, studies indicated that the technique of Q–Q plot could avoid the eliminated impact of anthropogenic activities on the heavy metal concentrations in the soil [40].

Bauer and Bor [41] defined the first bend of the slope on the plot curve as the upper limit of any collective background data. A

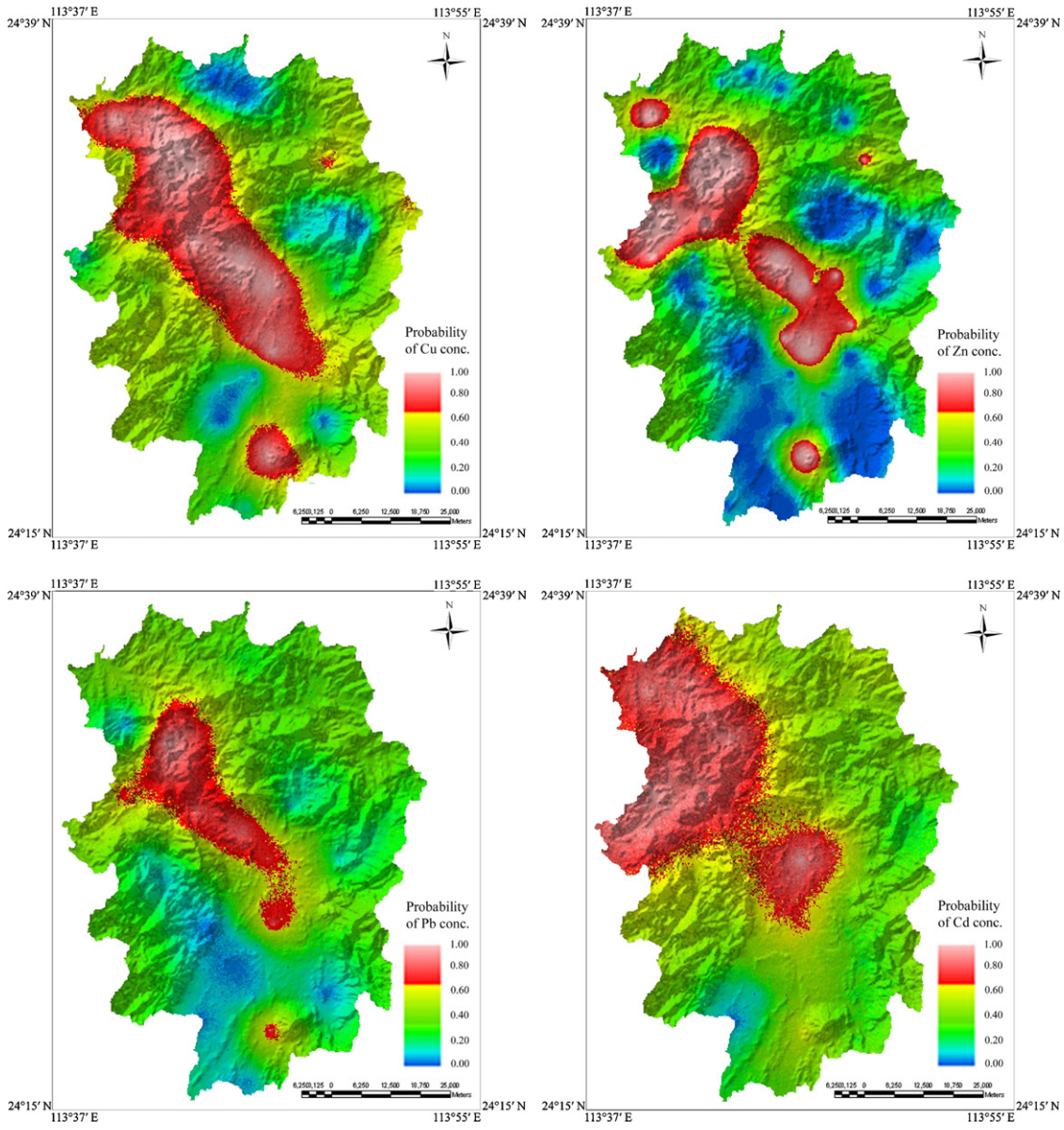

Fig. 8. Probabilities of each heavy metal in soil exceeding its own critical limit.

bend in the upper part of the graph (high values) can be used to distinguish between natural samples with comparatively low concentrations and anthropogenic samples with high concentrations [8]. The 'breaks' on the copper, zinc and lead plots obviously exist in normal Q–Q plots (Fig. 4) and their geochemical background can be distinctly detected. However, the 'inflection points' on the cadmium slope likely reflect the presence of different processes such as lithification, soil organic matter or anthropogenic influences. Searching for the inflection points needs to be aided by the environmental and geology investigations for determining cautiously the geochemical background of Cd contents.

The soil geochemical backgrounds of the four metals were determined from the first bend in the slope and are shown in Table 1. Results indicate that the geochemical backgrounds of the four heavy metals except Pb are above the Grade II limits at pH < 6.5 (EQSS) which are established to protect agriculture and human health. The upper background levels for Cd are six times greater than the Grade II limits.

Geochemical association of heavy metal in soils is mainly determined by the regional geological features, pedogenesis and the characteristics of heavy metal. It would influence regional geochemical background. The soil geochemical background of the four metals may vary greatly between the national scale and the watershed scale, as the land uses and parent rock of the watershed are much more complex than those of the whole territory of China. Therefore, previous pollutant results of the four metals from anthropogenic activities in soils could be over-estimated based on the national reference background values. The regional environmental studies have needed to be guided by the regional geochemical background to provide more realistic and scientific evaluations.

3.2. SIcS of soil geochemical backgrounds

3.2.1. Co-simulation of PTHMs with S_{100}

In the process of mapping the geochemical background, it is a complex task to correctly separate 'hotspots' derived from human activities from natural enrichment in the mineralized regions. For this reason it is necessary to collocated co-simulated process using S_{100} as the auxiliary information. The heavy metal concentrations and S_{100} were locally estimated by ordinary kriging, which gives a first approximation of their distributions. The kriging was per-

formed on a regular grid of 15 m × 15 m points, with 2393 columns and 3201 lines. In this study, a logarithmic transformation was applied to the measured metal concentrations to eliminate the positive skewness of their distribution. The variograms, used to model the dispersion, show an average behavior revealed by a spatial continuity with a range of 5409–8667 m. These models were fitted with Gaussian models (Fig. 5).

To check the Markov approximation, the hard indicator data were modeled against estimates of collocated S_{100} according to the cross-covariance model from Eq. (5) (Fig. 6A–D). The soft data for the minimal cutoff values for z_{Cu}, z_{Zn}, z_{Pb} and z_{Cd} provided information on whether the actual heavy metal concentrations originated from anthropogenic influences or not. The cross-covariance for each of the heavy metals is shown in Fig. 6. The soft indicator data were applied to perform a sequential indicator co-simulation for mapping the soil geochemical background of the four heavy metals because S_{100} has a similar spatial distribution and significant correlation with the metal concentrations.

From the parameters obtained above, the 'cosisim' program in the SGeMS software was used to perform the SIcS. Each realization is equiprobable and a set of realizations reveals the possible spatial distribution for the geochemical background of the metals in soils of the study area. Fig. 7 shows the estimated environmental geochemical background maps for the four heavy metals in mineralization-influenced soils in the study area. Features of the soil geochemical backgrounds for the four metals are presented in the similar spatial patterns. The areas of soils with elevated concentrations due to human activities extended from the Dabao Mountain mine to the riparian region on the middle-reaches of Hengshi River. The areas of elevated Cu and Cd concentrations are relatively wide along the riparian region of the Hengshi River. The spatial distributions agree with the graphic results based on the EDA analysis. Additionally, spatially discontinuous contaminated sites were detected in the riparian area downstream of the Hengshi River. It suggests that there are "safe" concentrations of Cu, Zn and Pb in soils downstream of the watershed, whereas a potential hazardous risk of the four heavy metals is possible pooled in the river sediments. The riparian area of the river needs more attention for restoration so that it can be the buffer area to contain the diffuse pollution from the contaminated sites.

3.2.2. Assessing the reliability of simulated results with uncertainty analysis

The joint probability calculated from the 600 realizations can be used to assess the spatial uncertainty of the soil geochemical background maps. The corresponding spatial uncertainty maps presented in Fig. 8 show that regions with a higher probability of elevated heavy metal concentrations are mainly located around the Dabaoshan mine and the middle-reaches of the Hengshi River, with conditional probability values over 0.80 in those regions. Downstream of the Hengshi River and the Xitou watershed, the conditional probability values are below 0.20. These mean that the simulated results of the four metals have the higher certainties in those polluted area and those unpolluted area, respectively. In contrast, the simulated results of the heavy metals have the lower reliabilities in the regions with conditional probability values approximately 0.50. In this study, the joint probability of the areas influenced by mining is over 0.70 and the spatial uncertainty of soil geochemical backgrounds are much smaller.

4. Conclusions

It is important for decreasing uncertainty of biogeochemical questions and models that a geochemical background should be quantified correctly. In this paper, the method integrated by EDA and SIcS was proposed for defining and modeling the geochemical background of Cu, Zn, Pb and Cd contents in mineralization-influenced soil from the Hengshi River watershed, Guangdong Province, China.

The combination of histogram, density marker, distribution curve and skewness, boxplot and Q–Q plot presented a much improved and quick insight into the distribution of empirical raw data and its characteristic. By applying the EDA technique, anthropogenic anomalies can be obviously detected through eliminating some of the geogenic influences. No statistical model is assumed in the procedure avoiding the need for transforming data, which is more reliable for the right-skewed geochemical background than the assumption of symmetry. However, defining the geochemical background in the Q–Q plot is still not a trivial task when the different processes coexist. In this case, the inflection point of geochemical background on the cadmium plot was determined with the aid of the contour interval of the SIcS. It provided more spatial details of environment and geology for identifying realistically the geochemical background.

The SIcS method based on non-parametric geostatistics does not require any assumptions of the shape of the conditional distribution and honors the fluctuations in original data. In particular, S_{100} as spatially correlated secondary information was introduced to separate the soil geochemical background concentrations from anthropogenic anomalies. This parameter is appropriate to solve the problem in mineralized soil as it displays a good spatial cross-covariance with the four metals. Also, the magnetic method is suitable to the environmental geochemistry on a large scale, for it is comparatively rapid and cost-effective, in dealing with much larger sample populations than other methods.

The study demonstrated that the two complementary techniques (EDA–SIcS) could give greater confidence for defining and modeling the soil geochemical background of heavy metal and provide a more reliable, realistic and cost-effective way for regional environment decision-making and safety management in the complex environment. The integrating method is suitable to researching for not only regional soil geochemical background, but also biogeochemical patterns and processes.

Acknowledgments

This study was financed by Guangdong Provincial Natural Science Foundation (No. 06202438). Our thanks are due to all people who provided help: Dr. Cuiping Zhang, Dr. Yulin Han, Meng Hu, Dr. Yanwu Zhou, Peng Zhao, Si Yang and Defu Tang from School of Environmental Science and Engineering, Sun Yat-Sen University. Also, we greatly appreciate the comments of the three reviewers and the helps of the editors.

References

[1] T.W. Parr, A.R.J. Sier, R.W. Battarbee, A. Mackay, J. Burgess, Detecting environment change: science and society-perspectives on long-term research and monitoring in the 21st century, Sci. Total Environ. 31 (2003) 1–8.
[2] A. Fairbrother, R. Wenstel, K. Sappington, W. Wood, Framework for metals risk assessment, Ecotoxicol. Environ. Safe 68 (2007) 145–227.
[3] N.S. Urquhart, S.G. Paulsen, D.P. Larsen, Monitoring for policy-relevant regional trends over time, Ecol. Appl. 8 (1998) 246–257.
[4] S.M. Rodrigues, M.E. Pereira, E. Ferreira da Silva, A.S. Hursthouse, A.C. Duarte, A review of regulatory decisions for environmental protection. Part II. The case-study of contaminated land management in Portugal, Environ. Int. 35 (2009) 214–225.
[5] R. Samlminen, T. Tarvainen, The problem of defining geochemical baseline. A case study of selected elements and geological materials in Finland, J. Geochem. Explor. 60 (1997) 91–98.
[6] E. Galán, J.C. Fernández-Caliani, I. González, P. Aparicio, A. Romero, Influence of geological setting on geochemical baselines of trace elements in soils, application to soils of South-West Spain, J. Geochem. Explor. 98 (2008) 89–106.
[7] D.J. Brus, F.P.J. Lamé, R.H. Nieuwenhuis, National baseline survey of soil quality in the Netherlands, Environ. Pollut. 157 (2009) 2043–2052.

[8] J. Matschullat, R. Ottenstein, C. Reimann, Geochemical background—can we calculate it? Environ. Geol. 39 (9) (2000) 990–1000.
[9] A.J. Sinclair, A fundamental approach to threshold estimation in exploration geochemistry: probability plots revisited, J. Geochem. Explor. 41 (1991) 1–22.
[10] C. Reimann, P. Filzmoser, R.G. Garrett, Background and threshold: critical comparison of methods of determination, Sci. Total Environ. 346 (2005 A) 1–16.
[11] S.V. Panno, W.R. Kelly, A.T. Martinsek, K.C. Hackley, Estimating background and threshold nitrate concentrations using probability graphs, Ground Water 44 (2006) 697–709.
[12] M. Hanesch, G. Rantitsch, S. Hemetsberger, R. Scholger, Lithological and pedological influences on the magnetic susceptibility of soil: their consideration in magnetic pollution mapping, Sci. Total Environ. 382 (2007) 351–363.
[13] C.S. Zhang, O. Selinus, Statistics and GIS in environmental geochemistry-some problems and solutions, J. Geochem. Explor. 64 (1998) 339–354.
[14] M.D. Cocker, Geochemical mapping in Georgia, USA: a tool for environmental studies, geologic mapping and mineral exploration, J. Geochem. Explor. 67 (1999) 345–360.
[15] M. Bounessah, B.P. Atkin, An application of exploratory data analysis (EDA) as a robust non-parametric technique for geochemical mapping in a semi-arid climate, Appl. Geochem. 18 (2003) 1185–1195.
[16] P.A. Burrough, GIS and geostatistics: essential partners for spatial analysis, Environ. Ecol. Stat. 8 (2001) 361–377.
[17] J.A. Chiprés, J. Castro-Larragoitia, M.G. Monroy, Exploratory and spatial data analysis (EDA-SDA) for determining regional background levels and anomalies of potentially toxic elements in the soils from Catorce-Matehuala, Mexico, Appl. Geochem. 24 (2009) 1579–1589.
[18] M.V. Meirvenne, P. Goovaerts, Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold, Geoderma 102 (2001) 75–100.
[19] A.S. Almeida, A.G. Journel, Joint Simulation of multiple variables with a Markov-type coregionalization model, Math. Geol. 26 (5) (1994) 565–588.
[20] K.W. Juang, Y.S. Chen, D.Y. Lee, Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils, Environ. Pollut. 127 (2004) 229–238.
[21] C. Franco, A. Soares, J. Delgado, Geostatistical modeling of heavy metal contamination in the topsoil of Guadiamar river margins (S Spain) using a stochastic simulation technique, Geoderma 136 (2006) 852–864.
[22] Y.C. Zhao, X.H. Xu, W.X. Sun, B. Huang, J.L. Darilek, X.Z. Shi, Uncertainty assessment of mapping mercury contaminated soils of a rapidly industrializing city in the Yangtze River Delta of China using sequential indicator co-simulation, Environ. Monit. Assess. 138 (2008) 343–355.
[23] P.A. Finke, D. Wladis, J. Kros, E.J. Pebesma, G.J. Reinds, Quantification and simulation of errors in categorical data for uncertainty analysis of soil acidification modeling, Geoderma 93 (1999) 177–194.
[24] D. D'or, H. Demougeot-Renard, M. Garcia, An integrated geostatistical approach for contaminated site and soil characterization, Math. Geosci. 41 (2009) 307–322.
[25] S.G. Robinson, The late pleistocene palaeoclimatic record of north Atlantic deep-sea sediments revealed by mineral–magnetic measurements, Phys. Earth Planet Inter. 42 (1986) 22–47.
[26] H.F.B. Lecoanet, F. Lévèque, Y. Kok, J.P. Ambrosi, Magnetic properties of salt-marsh soils contaminated by iron industry emission (South-East, France), J. Appl. Geophys. 48 (2001) 67–81.
[27] S.G. Lu, S.Q. Bai, Study on the correlation of magnetic properties and heavy metal content in urban soils of Hangzhou City, China, J. Appl. Geophys. 60 (2006) 1–12.
[28] Q.M. Chen, X.J. Zhang, M.A. Hu, Assessment of aquatic pollution of heavy metals in a copper–iron area, Environ. Sci. Tech. 29 (2006) 64–71 (In Chinese).
[29] S.M. Fu, Study on Geochemical Environmental Effects of Metallic Sulfide Deposit and its Mining in Dabaoshan, Sun Yat-Sen University, Northern Guangdong (China), 2007, pp 106.
[30] X. Zhou, B.C. Xia, Y.L. Han, Z.L. Zhang, X. Wei, ARM susceptibility mapping as a proxy for heavy metals loads of soils monitoring in the Dabaoshan Mine, South China, EPPH2009 special track within ICBBE, 14 June–16 June, Beijing. Available from: http://ieeexplore.ieee.org/stamp/stamp.jsp? arnumber = 05162406.
[31] P. Zhuang, M.B. McBride, H.P. Xia, N.Y. Li, Z.A. Li, Health risk from heavy metals via consumption of food crops in the vicinity of Dabaoshan mine, South China, Sci. Total Environ. 407 (2008) 1551–1561.
[32] X.J. Zou, R.L. Qiu, X.Y. Zhou, W.H. Zheng, Heavy metal contamination and health risk assessment in Dabao Mountain, China, Acta Sci. Circumst. 28 (7) (2008) 1406–1412 (In Chinese).
[33] J.A. Gili, J. Corominas, J. Rius, Using global positioning system techniques in landslide monitoring, Eng. Geol. 55 (2000) 167–192.
[34] J.F. Wang, L.F. Li, G. Christakos, Sampling and kriging spatial means: efficiency and conditions, Sensors 9 (2009) 5224–5240, doi:10.3390/s90705224.
[35] R.K. Lu, Analyse Methods of Soil and Agrochemistry, Soil Science Society of China, Chinese Agricultural Science and Technology Press, Beijing, 1999, pp. 474–489.
[36] R. Thompson, F. Oldfield, Environmental Magnetism, Allen&Unwin, Landon, 1986, pp. 25.
[37] C.V. Deutsch, A.G. Journel, Geostatistical Software Library and User's Guide, Oxford University Press, New York, 1998.
[38] C. Reimann, R.G. Garrett, Geochemical background-concept and reality, Sci. Total Environ. 350 (2005 B) 12–27.
[39] C.S. Zhang, D. Fay, D.M. Grath, E. Grennan, O.T. Carton, Statistical analyses of geochemical variables in soils of Ireland, Geoderma 146 (2008) 378–390.
[40] H.B. Zhang, Y.M. Luo, M.H. Wong, Q.G. Zhao, G.L. Zhang, Defining the geochemical baseline: a case of Hong Kong soils, Environ. Geol. 52 (2007) 843–851.
[41] I. Bauer, J. Bor, Lithogene, geogene und anthropogene Schwermetallgehalte von Lößböden an den Beispielen von Cu, Zn, Ni, Pb, Hg und Cd, Mainzer Geowiss Mitt. 21 (1995) 47–70.