Applied Geochemistry 48 (2014) 193–206

Contents lists available at ScienceDirect

Applied Geochemistry

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

Determination of regional soil geochemical baselines for trace metals with principal component regression: A case study in the Jianghan plain, China

Shuang Zhang [a,b], Dong Yang [c], Fanglin Li [b,*], Huayong Chen [a], Zhengyu Bao [b], Bin Huang [b], Dongfeng Zou [b], Jun Yang [d]

[a] *Key Laboratory of Mineralogy and Metallogeny, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China*
[b] *Faculty of Earth Sciences, China University of Geosciences, Wuhan 430074, China*
[c] *China Southwest Geotechnical Investigation and Surveying Institute, Chengdu 610081, China*
[d] *Hubei Institute of Geological Survey, Wuhan 430030, China*

ARTICLE INFO

Article history:
Available online 31 July 2014
Editorial handling by M. Kersten

ABSTRACT

The soil geochemical baseline is an important index in environmental assessment. Detailed baseline studies are necessary in large areas with complex geological settings, landforms and soil types. The Jianghan plain, a major industrial and agricultural region located in central China, has a soil geochemical baseline that has yet to be fully defined. The objective of this paper is to study the baseline of Cd, Pb and Zn in the topsoil of the Jianghan plain in a subarea using principal component regression (PCR). A total of 9030 samples were collected from the surface layer, and 2 soil profiles and 2 sedimentary columns were sampled near the Yangtze and Han rivers. Fifty-two elements and two parameters were analyzed. Data processing and the creation of spatial distribution maps of the elements were performed using MapGIS, R and SPSS software. The results show that the distributions of Cd, Pb and Zn are mainly controlled by parent material, drainage system and soil type. The study area is divided into 3 subareas, with factors reflecting the physico-chemical characteristics of the soil using factor analysis (FA). The geochemical baseline model is established in every subarea to predict the Cd, Pb and Zn values using principal component regression analysis (PCR); the exceptional values (as a result of anthropogenic input or mineralization) are distinguished by residuals (γ); and the natural background values and anthropogenic contributions are clearly distinguished. Therefore, the PCR method in these subareas is objective and reasonable, and the conclusion provides effective evidence of exceptional high values for further environmental assessment.

1. Introduction

The soil geochemical baseline is not only a reflection of soil properties, but is also an important standard for soil quality assessment and environmental protection and legislation (Cicchella et al., 2005; De Vivo et al., 2008; Gałuszka, 2007; Jarva et al., 2010; Xie and Ren, 1993). Baseline assessment can distinguish contaminated areas from uncontaminated ones and can be used to investigate anthropogenic influence on chemical changes in the environment (Baize and Sterckeman, 2001; López et al., 2008; Salminen et al., 2004; Tack et al., 2005; Tarvainen and Paukola, 1998; Wang et al., 2007b). In most countries, these assessments are usually conducted to determine the state of natural ecosystems prior to the implementation of a specific investment (Darnley, 1997; Salminen and Tarvainen, 1997; Wang et al., 2007a). In this paper, the soil geochemical baseline is employed with the understanding that "it represents a summation of natural and human influences, indicating the actual content of an element in the superficial environmental at a given point in time" (Salminen and Gregorauskiene, 2000; Tarvainen and Paukola, 1998; Teng et al., 2009). The data and the accompanying element distribution maps resulting from systematically documenting the concentration and spatial distribution of elements on the Earth's surface that can be a reference against which any future changes can be quantified (Wang, 2012).

Different predictive methods for defining soil geochemical baselines have been employed since the end of the previous century. These are: (a) Statistical methods. The iterative 2SD-technique is often used when elements are normally or lognormally distributed (Gałuszka, 2007). There is usually no single background, but rather

---

[*] Corresponding author. Tel.: +86 13545223600.
E-mail address: fanglinli@cug.edu.cn (F. Li).

http://dx.doi.org/10.1016/j.apgeochem.2014.07.019
0883-2927/

a series of multiple overlapping distributions in the geochemical data (Reimann and Garrett, 2005). Thus, fractal and relative cumulative frequency curve methods are chosen (Cheng et al., 1994). Cicchella et al. (2005) discussed baseline geochemical maps of elements that are harmful to human health in the metropolitan and provincial areas of Napole using the multi-fractal inverse distance weighted (IDW) interpolation method and spectral analysis (S-A) using the GeoDAS software; (b) Normalization procedures. Using a reference such as Al, Li and soil organic matter to normalize the geochemical data is also an available method at both regional and small scales (Covelli and Fontolan, 1997; Wang et al., 2011); (c) Substitution methods. Sometimes, baseline values obtained from samples in deep layers or pure background areas are considered as references for the surface layers or the whole study area; (d) Integration methods. Some methods that combine two or more methods have been proposed (Grünfeld, 2007; Teng et al., 2009). There are other publications discussing the establishment of the baseline and the separation of anthropogenic from natural origins (Chen et al., 1997; De Vivo et al., 2008; Guillén et al., 2011; Lattin et al., 2003; Reimann et al., 2008; Yang et al., 2009).

Usually there is no single concentration that can or should be considered representative of an element in soil, or other environmental media. There are many natural reasons why a given soil sample would contain an elevated concentration that is unexpectedly high and that does not conform to a man-made statistical model, and yet is not contaminated (polluted). Some geochemists have proposed that geochemical baseline concentrations depend not only on the dominant soil-forming factors (parent rock, climate, topography, biota and time) but also on sample material, grain size and the extraction method (Amorosi and Sammartino, 2007; Galán et al., 2008; Salminen and Gregorauskiene, 2000). It is important that regulators recognize that a geochemical baseline depends on location and scale (Reimann and Garrett, 2005). Because the parent material strongly influences the soil chemical properties, a geochemical baseline should be determined in separate geological regions (Darnley, 1997; Jarva et al., 2010). In their latest work, Smyth and Johnson (2011) studied the distribution and controlling factors of iodine in soils of Northern Ireland using a regional map. Cohen et al. (2012) studied the element spatial patterns in Cyprus across highly variable geology and soil types. These studies provide important references on baseline studies in large areas, with a great many samples from complex geological units, landforms and soil types. It is necessary to develop a useful method to meet the needs of accurate baseline computing for these studies.

We focus on a geochemical baseline establishment method for a large complex area using the Jianghan plain as an example. The plain stretches along with the Yangtze and Han rivers and is also one of the most important agricultural regions in China. However, no available soil reference values have been presented to establish a baseline for potentially toxic heavy metals. The Multipurpose Regional Geochemical Survey (1:250,000) (MRGS) is a development work of the International Geochemical Mapping (IGCP259) and Global Geochemical Baseline (IGCP360) projects in China. The MRGS is conducted by the China Geological Survey and aims to study the basal geology, resource potential, and ecological environment of the plain. This work provides a sufficient database for a detailed study on soil geochemical baselines. It has been found that an area of nearly one thousand square kilometers along the Yangtze and Han rivers is high in Cd (>0.3 ppm) (Yang et al., 2009). In this paper, heavy metals that are considered hazardous to human health (e.g. Cd, Pb and Zn) were determined in 9030 surface soil samples. The spatial distributions and material sources were systematically studied, as were the controlling factors of spatial distribution and variation. The sampled area was divided into 3 subareas using factor analysis due to the large survey region (approximately 36,160 km^2), with its complex geological setting and soil compositions. Principal component regression (PCR) analysis and Geographical Information System (GIS) spatial analysis techniques were used to establish a geochemical model of the surface soil geochemical baseline and to identify abnormal samples in the Jianghan plain.

2. Study area

The Jianghan plain, with an area of approximately 36,160 km^2, is located in Hubei province in central China (110°13′–114°15′E, 29°26′–31°20′N) (Fig. 1). The climate of this area is described as subtropical monsoon. The middle reaches of the Yangtze River and the lower reaches of Han River run through this alluvial plain, with includes many other rivers and lakes. The plain's climate and abundant water resources make it famous for its agriculture and aquatic products. The region exports grain, cotton, edible oil and fish. The landforms mainly consist of erosion aggradational, rolling and low aggradational plains (Fig. 2a). Approximately 60% of the area is paddy soil, and the rest is moist, brown–red[1] and yellow–brown soils (Fig. 2b). The region is also known for its mining industry. Most of the mines in the plain are medium-to large-size evaporite minerals salt deposits (halite and brine). Other mines contain clay and rare metal deposits. All the above that mentioned might be sources contributing Cd, Pb, and Zn to the surface soils.

All formations from the Mesoproterozoic to Quaternary are found in the Jianghan plain (Fig. 2c) (Li and Ma, 2008). Approximately 80% of the underlying strata of this area is Quaternary and 17% is Jurassic to Tertiary, leaving 3% as Mesoproterozoic to Lower Triassic strata. Sporadic Mesoproterozoic epimetamorphic rock series are distributed in the anticlinal core of the northeast fold belt and near the south boundary; a small amount of Cambrian–Triassic carbonates are distributed in the border fold belt; Jurassic–Cretaceous–Tertiary clasolites outcrop in the northern border; and a Quaternary loose river–lake deposit consisting of fine-grained clay and sandy soil covers most of the study area. The landforms and soil types of the Jianghan plain are controlled by its geological setting and drainage system, respectively (Figs. 1 and 2). The source areas of the Yangtze and Han rivers are the Yangtze block and the Qinling orogenic belt, respectively, which have different rock types, strata and mineral compositions.

3. Methods

3.1. Sampling

In accordance with the MRGS sampling procedures, 9030 topsoil samples (0–25 cm depth) and 2224 subsoil samples (150–170 cm depth) were collected in the Jianghan plain. The 0–25 cm depth samples provide information of regarding potential anthropogenic contamination of the top layer. The 150–175 cm depth samples are collected to indicate natural element concentrations because experiments have shown that there is very little anthropogenic effect on this depth (Zhao et al., 2008). The enrichment or depletion of elements between the two layers can also be observed (Cohen et al., 2012). The sampling grids of the surface and deep soil were 4 (2 × 2 km) and 16 (4 × 4 km) km^2, respectively. Each grid was then divided into four sub-cells (1 × 1 km and 2 × 2 km). The samples for analysis were composites of the four samples collected in each sub-cell center. To investigate the horizontal distribution of the elements in the soil across the two rivers, two northeast trending soil profiles were additionally sampled crossing the Yangtze River and Han River (profile I and profile II), respectively

---

[1] For interpretation of color in Fig. 2, the reader is referred to the web version of this article.

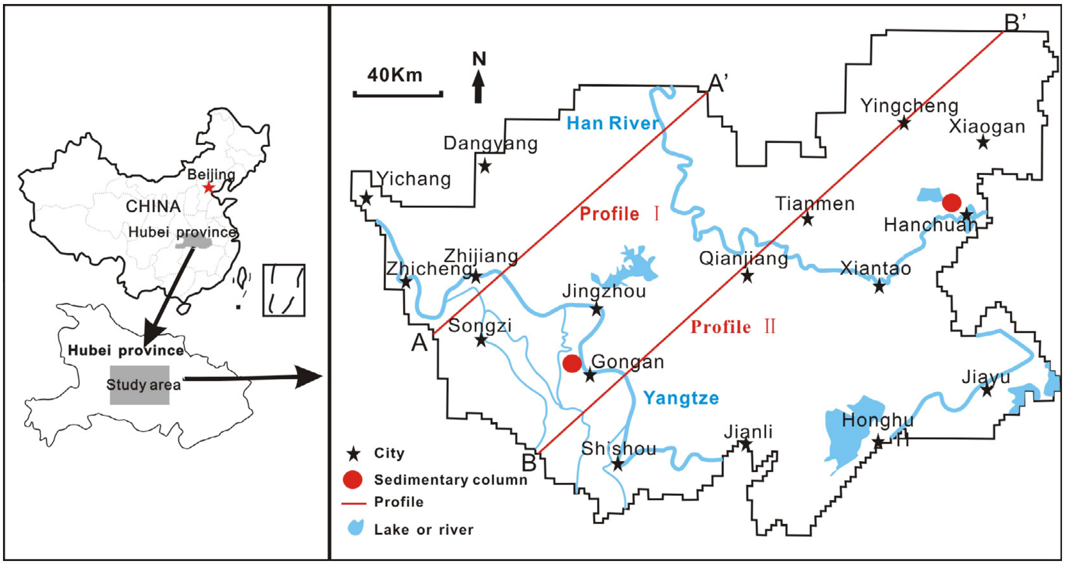

Fig. 1. Sketch map of the Jianghan plain and sample locations.

(Fig. 1). The profiles pass through areas of different types of strata, landforms and soils in the Jianghan plain. Two sedimentary columns (from the cities of Gongan and Hanchuan, Fig. 1) were sampled to study the vertical distribution of Cd, Pb and Zn. The depth of each sedimentary column was approximately 170 cm, with samples being collected. The samples were collected every 15 cm. Finally, ~1 kg of soil was collected and placed in clean polyethylene bags, and more than 0.5 kg of soil was stored in a bottle for transport to the laboratory.

3.2. Analytical techniques

Geochemical analyses were conducted at the Hubei Geological Research Laboratory. Fifty-two elements or oxides (i.e. Ag, As, Au, B, Ba, Be, Bi, Br, Tc, Cd, Ce, Cl, Co, Cr, Cu, F, Ga, Ge, Hg, I, La, Li, Mn, Mo, N, Nb, Ni, P, Pb, Rb, S, Sb, Sc, Se, Sn, Sr, Th, Ti, Tl, U, V, W, Y, Zn, Zr, SiO_2, Al_2O_3, Fe_2O_3, K_2O, Na_2O, CaO and MgO), organic carbon and pH were determined (Table 1). An approximately 40 g sample was used for the pH. The other samples were crushed and sieved to a <0.074 mm fraction. A 10–20 g sample was stored in a plastic bottle for the analysis of As, Sb, Bi, Hg and Se. An approximately 40 g sample was stored in a paper bag for the analysis of Au. Another 30–40 g sample for the remaining analyses was stored in a glass bottle and placed in the dryer for 2 h at 105 °C. Detailed quality control procedures were implemented based on Specifications of Analysis Technology on Eco-Geochemical Sample Evaluation (trial implementation) (DD2005-03). The analytical precision was controlled using 1 in-house replicate and 2 blind duplicates for every sub-batch of 50 unknown samples. The accuracy was determined using 4 in-house reference materials (GSS-1, 2, 3, 8) for every sub-batch of 50 unknown samples. The analytical quality met the procedures requirements with a precision and accuracy within ±10%.

4. Data processing

4.1. Statistics of element concentrations

A statistical evaluation of the soil data was performed, including the determination of relevant descriptive parameters of centering and dispersion, such as minimum, maximum, mean, first (Q_1), median (Q_2), third (Q_3), fourth (Q_4) quartiles, SD (standard deviation) and CV (coefficient of variation) (Table 2). The histograms of element values in topsoil and deep soil were generated in SPSS (Fig. 1S (Supplementary material)) (http://www.spss.com.cn/). Maps of elements in topsoil and deep soil were generated in MapGIS (Figs. 3 and 4, respectively) (statistics using Q_1, Q_2, Q_3 and 95% as interval boundaries) (http://english.mapgis.com.cn/en/) and the change of element concentration in the profiles (Figs. 5 and 6) and sediment columns are showed in the line chart (Fig. 7).

4.2. Calculation of soil geochemical baseline

Although progress has been made in previous estimations of soil geochemical baselines, the following questions regarding the calculated methods are still debated. (A) Sample representativeness: Most of the data for baseline research currently come from the MRGS databases in China. The elemental geochemistry patterns obviously differ in soils across wide survey areas (Cohen et al., 2012; Li and Ma, 2008; Smyth and Johnson, 2011). For this reason, samples in background areas and in the deep layer soil are most likely unsuitable for the surficial baseline determination of the whole survey area (Desaules, 2012; Tapia et al., 2012). (B) Subarea division: When the distribution of elements in soil is dominated by geologic units or other factors, especially in large-scale baseline studies, it is necessary to divide the large survey area into subareas depending on the controlling factors. (C) Data processing: In some studies the background is considered a baseline value, and the outliers are removed artificially, which may result in reference value errors (Meklit et al., 2009).

To solve these problems, the following three tasks were performed in the baseline study of the Jianghan plain:

(A) Samples of the topsoil were selected. Considering the geochemical patterns and changes in the layers, topsoil samples can provide more information about the surface environment.

(B) Subarea division. The study area was divided into subareas using factor analysis (FA) representing the controlling factors in the Spatial Analysis section of MapGIS (http://english.mapgis.com.cn/en/).

(C) The PCR method was chosen to calculate the baseline values of Cd, Pb and Zn. The baseline values of the three elements are determined by geochemical parameters using PCR in R

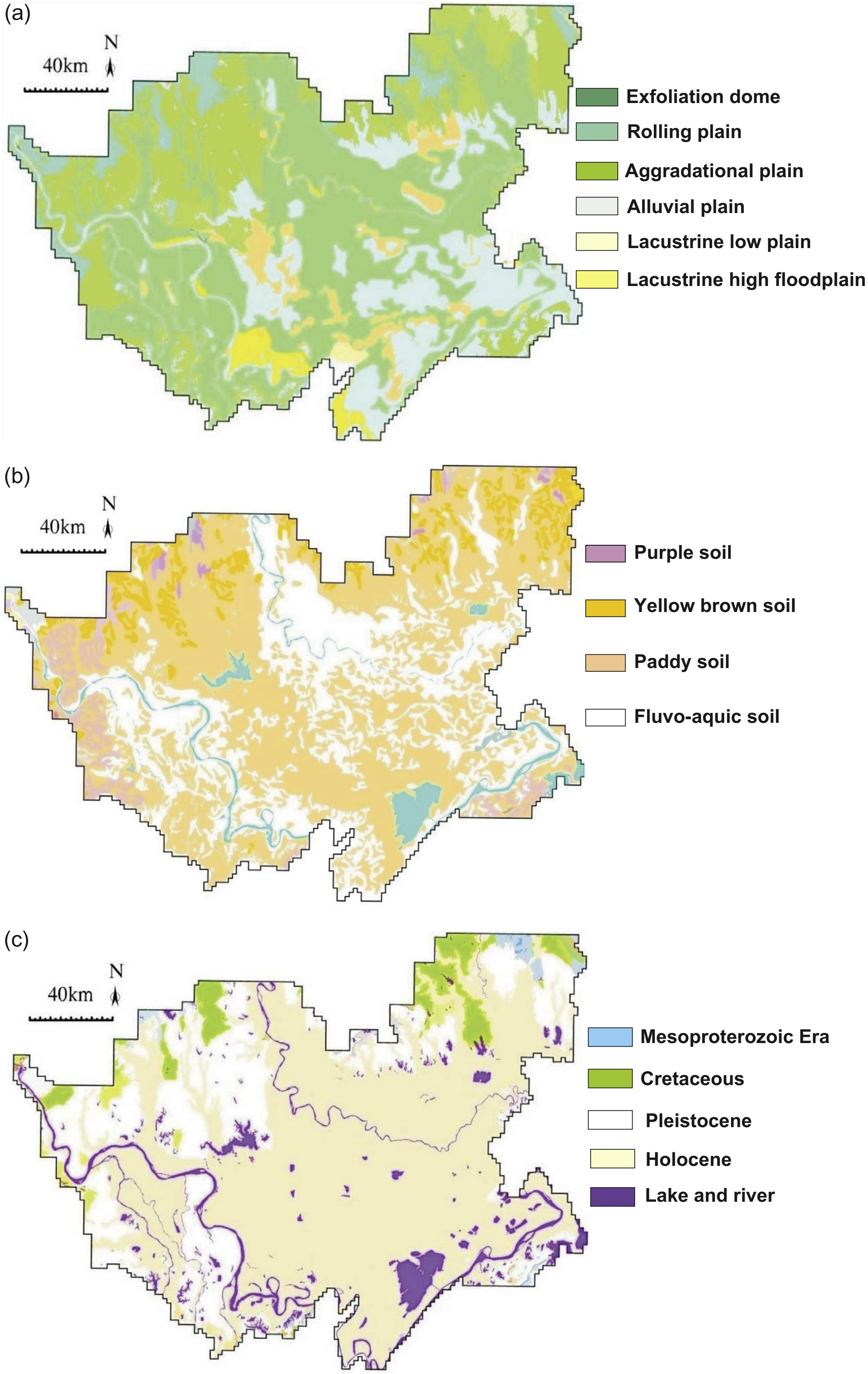

Fig. 2. (a) Simplified landform map of the Jianghan plain; (b) soil map of Jianghan plain; (c) simplified geological map of the Jianghan plain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Analysis methods for 52 elements and 2 parameters.

Number	Sample (g)	Method	Instrument	Element
1	0.1	Digested in $HNO_3 + HCl + HF + HClO_4$	ICP-MS	Be, Cd, Co, Cu, La, Li, Ni, U, Ge, Ce, Tl, Sc
2	4.0	Pressed powder	XRF	Ba, Cr, Mn, Nb, P, Pb, Rb, Sr, Th, Ti, V, Y, Zn, Zr, Ga, Cl, Br, SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO, Na_2O, K_2O
3	0.5	Digested in aqua regia, KBH_4 reduction, hydrogenation	AFS	As, Sb, Bi, Hg, Se
4	1.0	Alkali fusion	POL, ISE	Mo, W, F
5	0.1	Counter electrode spectrography	ES	Ag, B, Sn
6	10	Thiourea, activated carbon	ICP-MS	Au
7	0.5	Tube furnace	IA	S
8	0.2	Tube furnace	Non-aqueous titration	Total carbon
9	1.0	Digested in $K_2CrO_7 + H_2SO_4$	VOL	N
10	0.5	Semi-fusion in $Na_2CO_3 + ZnO$	COL	I
11	0.5	Digested in $K_2CrO_7 + H_2SO_4$	VOL	Organic carbon
12	10	Water immersion	pH meter	pH

ICP-MS, Inductively coupled plasma mass spectrometry; XRF, X-ray fluorescence; AFS, Atomic Fluorescence Spectroscopy; POL, polarography; ISE, ion selective electrode; ES, emission spectrometry; IA, iodimetric analysis; VOL, volumetry; COL, colourimetry.

Table 2
Relevant parameters of heavy metal elements in the soil of the Jianghan plain and a comparison with reference values (ppm, reference values are from Li and Ma, 2008).

Elements	Topsoil (0–25 cm depth)			Deep soil (150–170 cm depth)		
	Cd	Pb	Zn	Cd	Pb	Zn
Minimum	0.05	16.2	24	0.02	12.3	16.2
5th percentile	0.122	21.6	47.9	0.047	19	53.8
25th percentile	0.17	26.2	61.5	0.082	22.9	64
Median	0.29	28.9	86.5	0.185	25.4	77.4
75th percentile	0.36	31.5	105.2	0.24	27.9	94.5
95th percentile	0.454	36.9	124.1	0.294	32.3	112.1
Maximum	3	252.1	400	1.153	325.8	399.7
Arithmetic mean	0.28	29.16	85	0.171	25.66	79.84
Standard deviation	0.13	6.04	25.85	0.11	7.61	20.36
Coefficient of variation	0.46	0.21	0.30	0.64	0.30	0.26
Hubei province background	0.172	26.7	83.6	0.137	27.1	92.6
National background	0.097	26.0	74.2	0.084	24.7	71.1

software (http://www.r-project.org/), which includes two stages of principal component analysis (PCA) and regression analysis. The outliers can be removed by residual error (γ).

4.2.1. Division of subareas

As mentioned previously, the distribution of soil elements is controlled by parent material, landform, physico-chemical parameters and many other factors. The key reason for geochemical division is to more easily define an aggregative indicator reflecting all factors using FA (Albanese et al., 2010; Covelli and Fontolan, 1997; Grünfeld, 2007). Variables for the FA should meet the three conditions, be: (a) stable and hardly contaminated in the surface soil; (b) representative of the properties of the parent material; and (c) correlated to the spatial distribution of different types of soils and landforms.

The parent material of the soil is a principal control. Quaternary sediments, most of which are fluvial, underlay 80% of the study area. Studies have shown that the chemical composition of fluvial deposits is dominated by the supracrustal rocks in the catchments (Galán et al., 2008). The upstream region of the Yangtze River is underlain by the Yangtze block, while the Han River region is underlain by the Qinling orogenic belt. The supracrustal rocks of the two geologic units are different, so the sedimentary compositions of the regions along Yangtze River and Han River are different.

Landforms also exert influence on the distribution of heavy metals. The redistribution of material, energy and water on the earth's surface is dominated by the landform type during soil formation. The soils developed in the low plains are fine-grained and high in clay minerals because their parent materials were deposited in relatively low energy hydrological environment. The enrichment of metal elements can be attributed to the strong adsorption of the clay minerals.

Based on the above discussion, the distribution of major elements that dominate the properties of the rock and the behavior of the trace element match the distribution of the strata and the drainage system in the Jianghan plain. Oxides, which are the major components of soil, such as SiO_2, MgO, CaO, Na_2O, K_2O, Al_2O_3 and Fe_2O_3, can reflect most of the parent material's characteristics. For instance, CaO can indicate the different sedimentary sources of the Yangtze and Han rivers, and SiO_2/Al_2O_3 indicates the degree of weathering and the clay content (Li and Ma, 2008). Hence, the oxides can serve as a group of reference indicators. The pH value is another significant physico-chemical parameter. Soil pH is crucial for many chemical processes, such precipitation-dissolution and the sorption-resolving process. Thus, the pH value has an influence on the heavy metal speciation in soil. Some researchers have proposed that the content of water-soluble Cd has a negative correlation with pH values ranging from 4.5 to 7.2 (Xu et al., 2007). Organic matter, which can strongly adsorb heavy metal ions, is also considered an important indicator. In the Jianghan plain, Cd ($r = 0.629$, $p < 0.01$) and Zn ($r = 0.676$, $p < 0.01$) show a positive correlation with pH which means that soils with high natural total Cd and Zn concentration usually have high soil pH value. The positive correlation coefficients amongst the Cd, Zn and Al_2O_3/SiO_2, which reflect the particle size, reach 0.646 and 0.873 at the 0.01 level

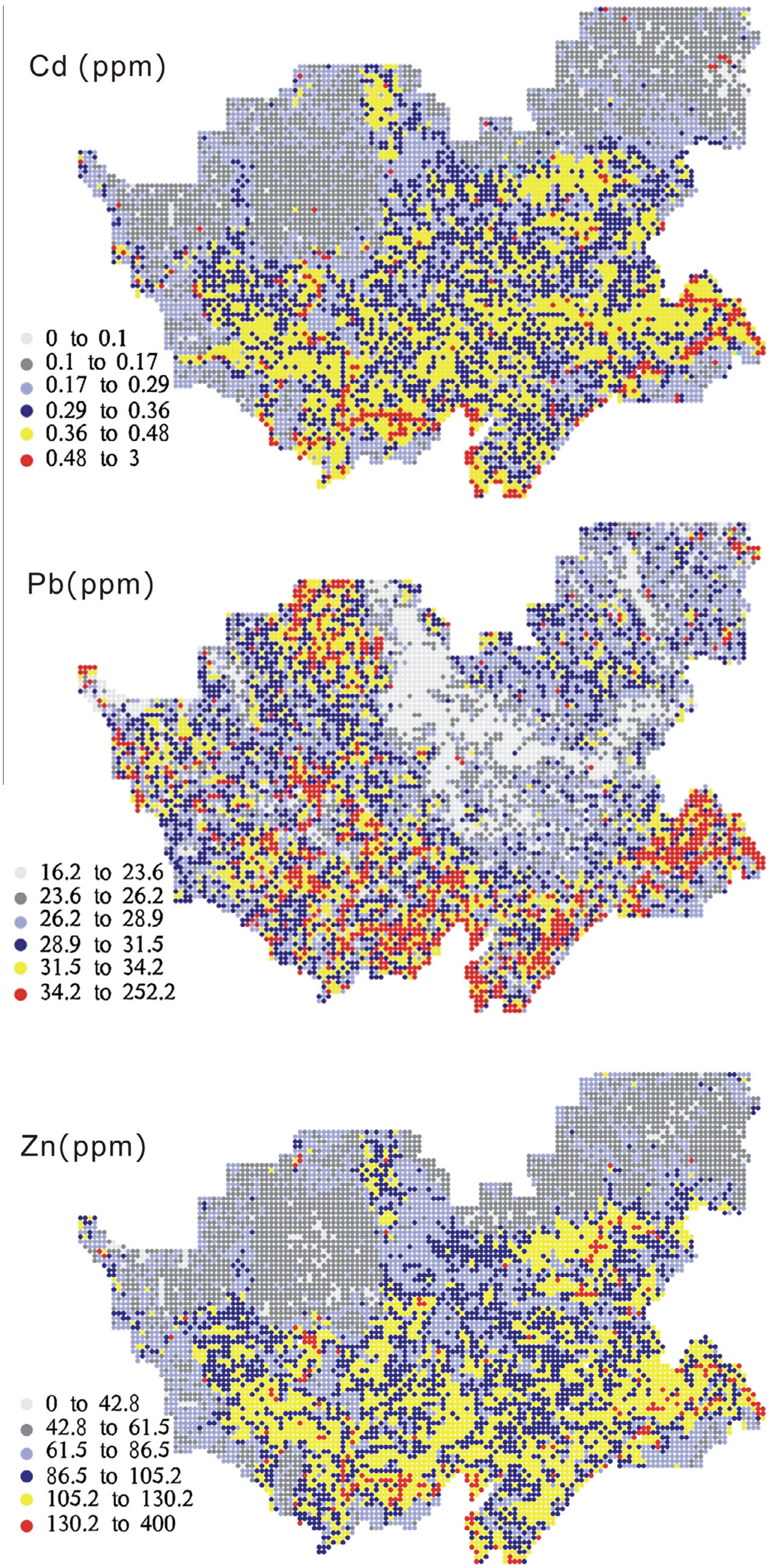

Fig. 3. Distribution of Cd, Pb and Zn in the topsoil of the Jianghan plain (using quartiles, the red spots are outliers). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

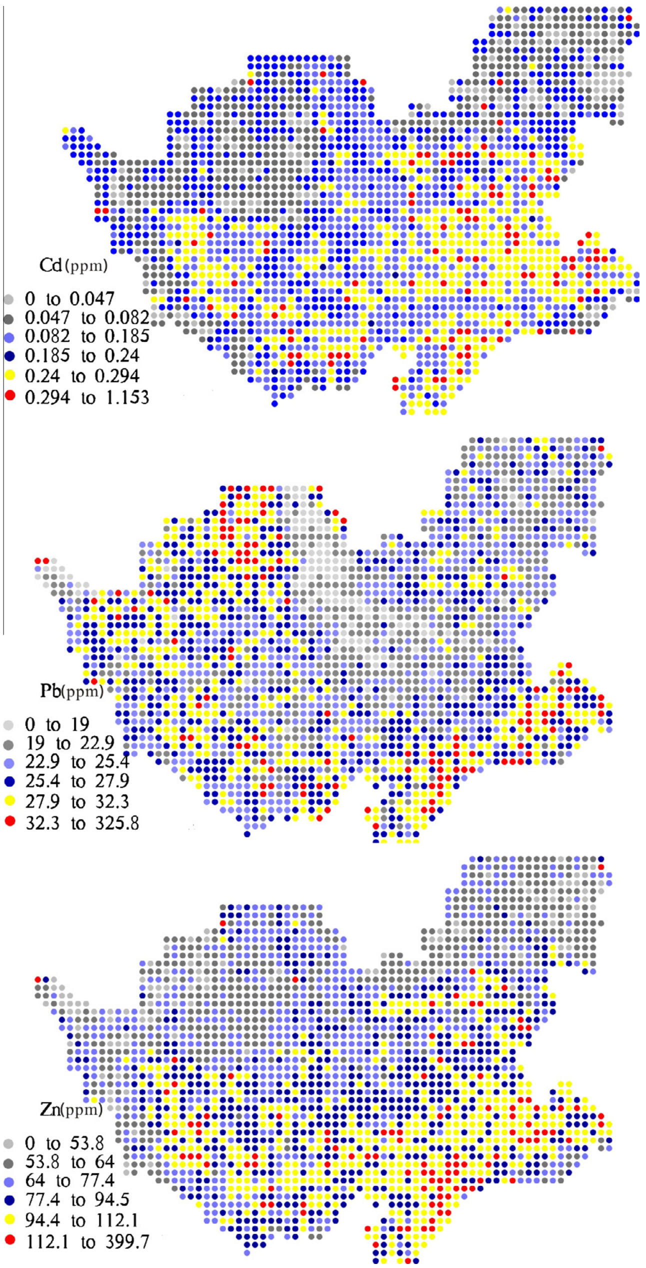

Fig. 4. Distribution of Cd, Pb and Zn in the deep soil of the Jianghan plain (using quartiles, the red spots are outliers). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Zhao et al., 2008). These indicate that the soil pH, organic carbon and Al_2O_3/SiO_2 values are variables that are correlated with the heavy metal content. With the contents of Na_2O, Ba and Mn being high in Han River region only, they can be indicators of Han River control. In this research, 7 major elements as well as Ba, Mn, organic carbon and pH are chosen as the variables for the factor analysis.

As the sample composition has to sum to a constant (i.e. 100% or 1 million ppm) in the closed data set, an increase in the amounts of one oxide or element must result in decrease in those of others. Thus it is necessary to open the data and destroy the effects of closure in the multivariate analyses. Three approaches are used to address the closure issue, including: (1) using the concentrations in ppm units for major elements and organic carbon instead of weight percent oxide; (2) undertaking a centred log-ratio (clr)

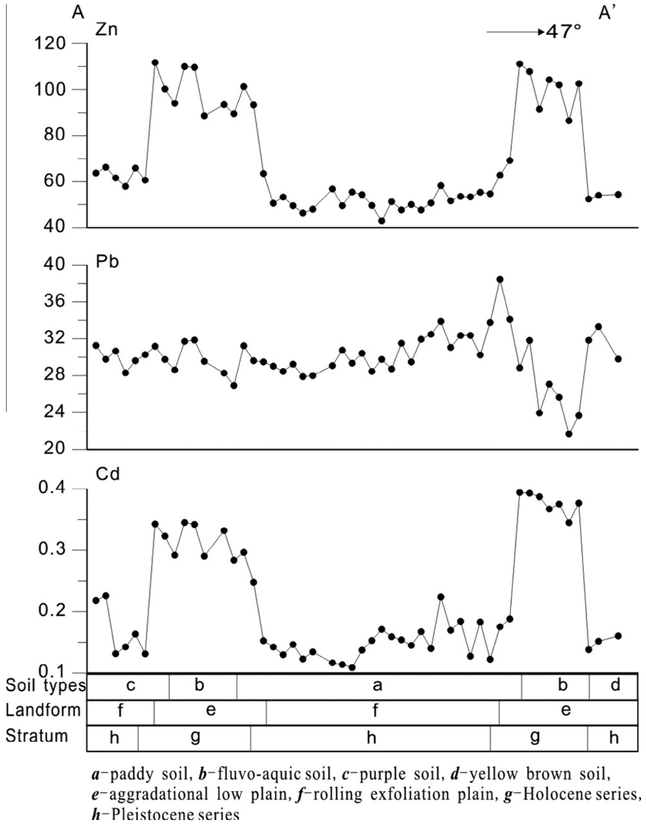

Fig. 5. Change of Cd, Pb and Zn in soil profile I of the Jianghan plain (*Y* axis: ppm).

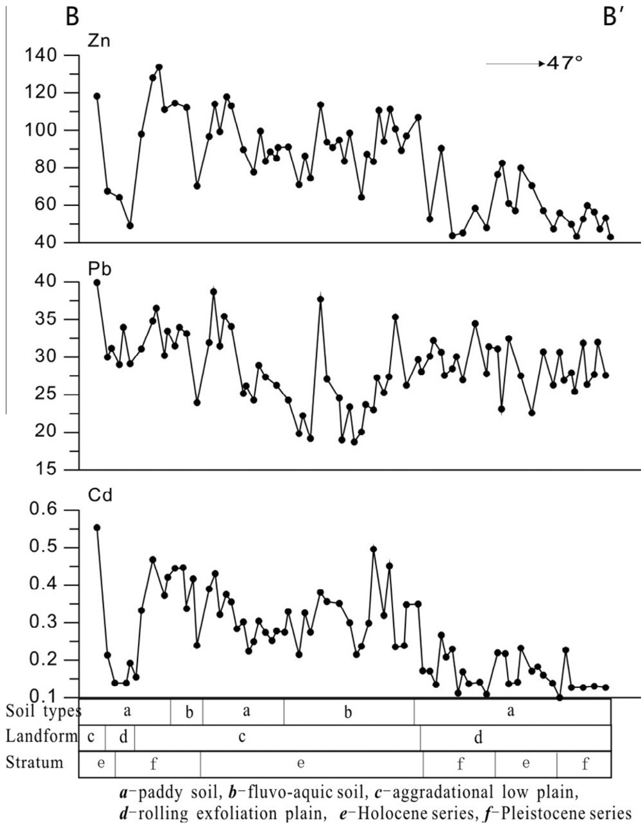

Fig. 6. Change of Cd, Pb and Zn in soil profile II of the Jianghan plain (*Y* axis: ppm).

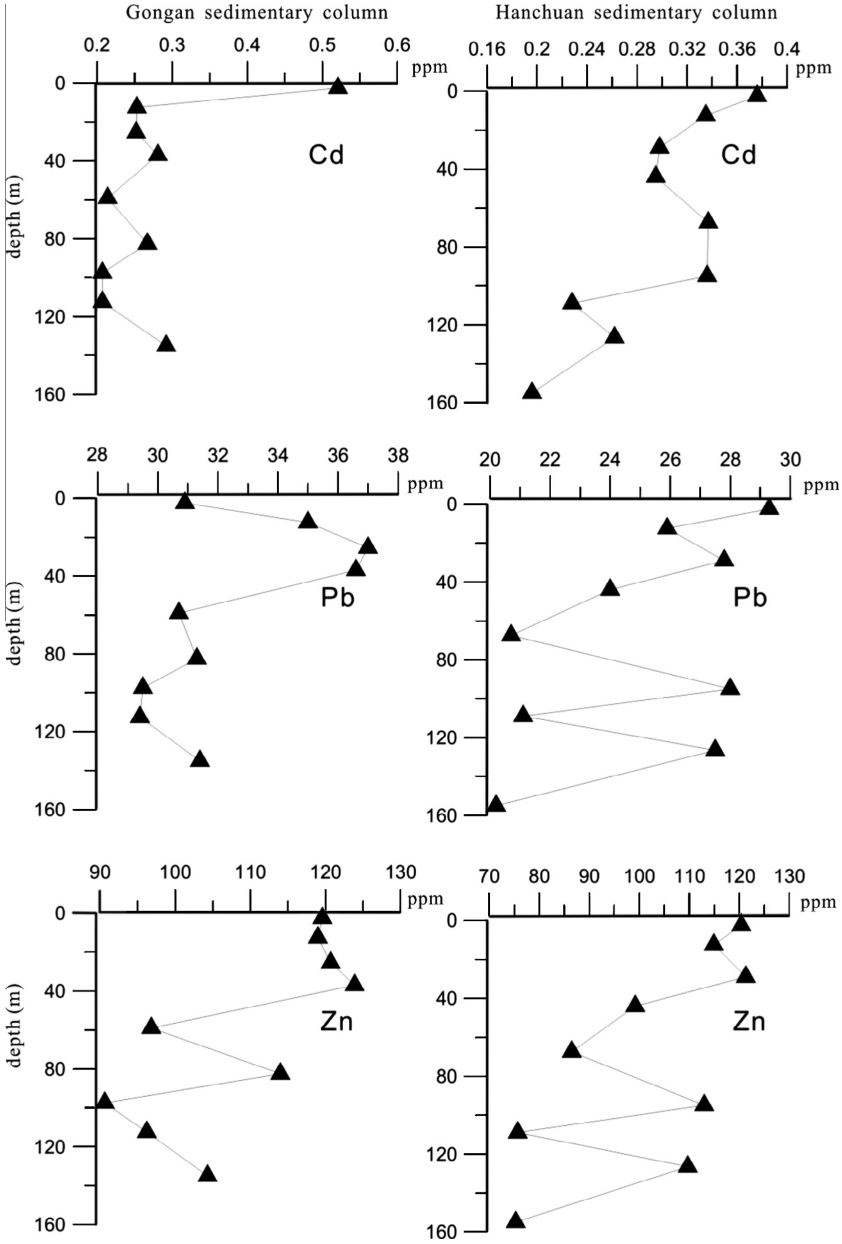

Fig. 7. Depth profiles of Cd, Pb and Zn soil concentrations.

transformation of selected variables except pH (Reimann et al., 2008); and, (3) adding the pH data into them. Then the processed data are taken as effective variables in the following factor analysis. The principal component method and Varimax rotation method are used in carrying out a factor analysis. Then the samples are classified into groups according to their factor-scores using MapGIS software (i.e. if a sample has its highest scores on factor 1, it is sorted into partition 1, discussed further below).

4.2.2. Determination of baselines

Major elements are the most important constituents of soil. Some researchers propose that the heavy metals sorption is clearly related to clays, metal oxides and hydroxides, metal carbonates and phosphates in soil (Bradl, 2004; Yang et al., 2009). For the major elements, the anthropogenic inputs are usually minute. Furthermore, to compensate for grain-size and mineralogy effects on trace element concentrations and to assess if anomalous metal contributions are present, a common approach is to normalize the geochemical data using Al as a grain-size proxy in baseline studies (Covelli and Fontolan, 1997). Fe, Li, organic carbon and Cs have been selected as normalizers in different conditions and the baseline values predicted by the normalizer using simple linear regression where there is a single predictor variable (Jiang et al., 2013). Based on the above discussion, in this study multiple geochemical parameters are selected from the MRGS database in whose context to study the Cd, Pb and Zn distributions. They have the characteristics that not significantly affected by anthropogenic processes employed in the study area, having stable chemistry properties and good correlation with geogenic Cd, Pb and Zn. The following parameters are selected: SiO_2, Al_2O_3, pH, organic carbon, Sc, Th, Zr, Nb, La, Ce, V, Ga, Be, Y, Ti, B, Co, Rb and Ba. The contents of SiO_2 and Al_2O_3 reflect the development level the of soil; the soil pH and organic carbon are variables reflecting the content of heavy metals in soil; Sc, Th, Zr, Nb, La, Ce, V, Ga, Be, Y and Ti bearing minerals, are insoluble and stable during weathering, erosion and transportation, are chemically inactive, indicating properties of soil origin; although B, Co, Rb and Ba are relatively active, they are easily adsorbed by clays, and their distributions reflect the properties of soil origin.

The transformations for the PCA and PCR are addressed. To open the closed number system, the oxide data are converted to elements. The units of concentrations of all the variables (in the form of elements) are unified in ppm and a clr transformation is undertaken before the PCA (Reimann et al., 2008). For the dependent variables (Cd, Pb and Zn), as trace and minor elements (<10%), a logarithmic transform will suffice to allow for closure (Filzmoser et al., 2009). For every subarea, PCR is then accomplished in R package 'robustbase'. First, PCA is used to reduce the dimension of the variables. Through PCA, the first several components are obtained according to the scree plot and the explained variability in every partition. After that the baseline values of Cd, Pb and Zn are predicted by regressing these elements against the principal components in every subarea. A regression analysis model is established as follows:

$$Y = \sum_{i=1}^{k} \beta_i X_i + \beta_0 + \gamma$$

where X is the matrix of selected predictor PC scores, Y is the vector of Cd, Pb or Zn data, k is the number of components selected, β_i are coefficients, β_0 is the constant, γ are the vector of residuals.

Considering that background is a range, the constant β_0 is equated to the geochemical baseline (average background). And the function of the PC's provides the background variation around the baseline due to the subarea specific variability influenced by the processes explained by the PCs.

To limit the effect of outliers a robust regression procedure, least trimmed squares (LTS), is applied to the data (Reimann et al., 2008). The principle components selected and the target elements (Cd, Pb and Zn) are considered independent variables (X) and induced variables (Y), respectively. Because the induced variables are logarithm values, the baseline values are back transformations of the constants. The residuals (γ), which should approach a normal distribution, are the differences between the actual values and the predicted values. Then the exceptional values are identified using a boxplot and their locations are identified in maps.

5. Results

5.1. Descriptive statistic

The Jianghan plain has higher average and median values of Cd, Pb and Zn than both the national background and the Hubei province background (Li and Ma, 2008) (Table 2). In addition, the element concentrations in the surface soil are higher than in the deep layers. The average value and median value of Cd in the surface soil are 2.5 and 1.6 times greater than those in the deeper layers, respectively, perhaps due to the possible contamination in the surface soil or the vertical fractionation of heavy metal caused by variations in organic carbon, physico-chemical or redox conditions. The range of heavy metal contents in the soil of the Jianghan plain is variable, as follows: Cd (0.05–3 ppm), Pb (16.2–252.1 ppm) and Zn (24–400 ppm) in the surface soil and Cd (0.02–1.15 ppm), Pb (12.3–325.8 ppm) and Zn (16.2–400 ppm) in the deep layer soil. All the maximum values are much higher than the 95th percentile values. These elements show different histograms shapes (Fig. 1S (Supplementary material)). Cd and Zn display a bimodal distribution and Pb displays a unimodal distribution both in the surface soil and deep soil (Fig. 1S (Supplementary material)). The bimodal distributions for Cd and Zn indicate the presence of multiple data populations. The studies on the origin of high Cd and Zn zones along the Yangtze River and Han River show that there is very little anthropogenic contamination of Cd, Pb and Zn in this area except some point-source pollution (Ma et al., 2005). The Cd, Pb and Zn enriched supracrustal rocks from the Yangtze block and Qinling orogenic belt is the main source of Cd, Pb and Zn in soil of Jianghan plain (Gao et al., 1998). So the two data populations are mainly related to the Yangtze and Han drainages.

5.2. Horizontal distribution and vertical distribution

The horizon distribution has similar shape in the topsoil and deep soil for every element. But the areas with high value are more extensive in the topsoil than in the deep soil (Figs. 3 and 4). We choose the topsoil map to describe (Fig. 3). The distribution characteristics of the Cd and Zn in the soil are similar. The element contents of the soil in the low aggradational plain are the highest. The soils sampled from the Yangtze River drainage system show higher element content than those from the other river systems. As mentioned in Section 5.1, the content of Cd and Zn are controlled by the sources of the Yangtze and Han drainages, which is accounting for the two data populations in the histogram plots. The distribution of Pb is related to the drainage systems. Soils with a high Pb content (more than 34.2 ppm) are distributed along the Yangtze River, while soils with a low content (less than 23.6 ppm) distributed along the Han River system and the drainage systems of the Dabie Mountain. The landform of the Jianghan plain is controlled by its geologic setting (Fig. 2a and c). Hillock plains and the low hilly areas are located on the Pleistocene series sediments and the low plain area is located on the Holocene series sediments.

In profile I, the samples from the low plain, especially along the Yangtze and Han rivers, are relatively high in Cd and Zn, but the samples from the rolling plain and hill areas are relatively low (Fig. 5). It is notable that soils developed on Holocene series exhibit higher Cd and Zn levels than those developed on Pleistocene series. However, the samples from the rolling low plain near the Han River have the highest Pb content, and near the Yangtze River the change of Pb content is small. In profile II (Fig. 6), the Cd and Zn contents decrease from southeast to northwest along the profile. The Holocene series samples in the low plain show the highest contents of Cd and Zn. Nevertheless, the Pb concentrations show no systematic pattern. As the survey results of the two sediment columns show (Table 3), the Cd content is the highest in surface soil (0.521 ppm) and decreases to a steady level from 5 cm to 150 cm in the Gongan sedimentary column. The Pb content increases to 35–37 ppm from 5 cm and 43 cm, and the Zn content is elevated in the top 43 cm and then falls to between 91 ppm and 114 ppm. Compared with the Gongan sedimentary column, the Cd, Pb and Zn contents in the Hanchuan column are unstable in every layer and generally decrease with increasing depth (Fig. 7).

5.3. Subarea division

In the factor analysis, the major element, Ba, Mn, organic carbon and pH data from the Jianghan plain research were subjected to Kaiser–Meyer–Olkin (KMO) and Bartlett's tests, which meet the conditions ($KMO = 0.82$, $P < 0.001$) of the Kaiser Standard. The factor-loading matrix with Varimax rotation is shown in Table 4. Three factors are extracted on the basis of the scree plot for a variance contribution reaching 85% (Fig. 2S (Supplementary material)). The factor-loading and factor score distribution are shown in Figs. 3S–5S (Supplementary material) and 8, respectively. Factor 1 is a soil parent material factor characterized by high Fe, Ca, Mg, K, pH, Si and Mn loadings (Fig. 3S (Supplementary material)). The samples with high scores in factor 1 are developed along the drainage systems reflecting their different origin from those of others (Fig. 8a). Factor 2 that dominated by Al and Ba is a Han River factor

Table 3
Cd, Pb and Zn contents in soil at different depths in the Jianghan plain (ppm).

Gongan sedimentary column				Hanchuan sedimentary column			
Depth (cm)	Cd	Pb	Zn	Depth (cm)	Cd	Pb	Zn
0–5	0.521	30.9	119.6	0–5	0.376	29.3	120.4
5–20	0.253	35	119	5–20	0.335	25.9	114.9
20–31	0.252	37	120.7	20–38	0.298	27.8	121.3
31–43	0.281	36.6	123.9	38–50	0.295	24	99.2
43–75	0.214	30.7	96.8	50–85	0.337	20.7	86.5
75–90	0.267	31.3	114	85–105	0.336	28	113
90–105	0.207	29.5	90.6	105–113	0.228	21.1	75.7
105–120	0.207	29.4	96.2	113–140	0.262	27.5	109.7
120–150	0.292	31.4	104.3	140–170	0.196	20.2	75.3

$W(b)/1 \times 10^{-6}$.

Table 4
Orthogonal rotated component of the factor analysis of soil in the Jianghan plain.

	F1	F2	F3
Fe	0.71	0.35	0.52
Mg	0.92	0.30	−0.10
Ca	0.91	−0.04	−0.32
Na	0.09	0.43	−0.79
K	0.77	0.55	0.09
pH	0.88	0.16	−0.26
Ba	0.27	0.91	−0.11
Mn	0.72	0.28	0.03
Al	0.54	0.63	0.47
Si	−0.90	−0.32	−0.19
Organic C	−0.12	0.16	0.74
λ	6.34	1.94	1.04
Contribution rate (%)	57.63	17.60	9.42
Cumulative (%)	57.63	75.23	84.65

(Fig. 3S (Supplementary material)). Thus the distribution of soil with high factor 2 scores is driven by Han River (Fig. 8b). Factor 3, a soil type factor (Fig. 8c), shows positive loading of organic carbon and negative loading of Na (Fig. 3S (Supplementary material)), which are controlled by the soil types. Then the samples are classified into three groups according to their maximum factor-scores. Partition 1 is in the Yangtze River flood plain and is dominated by high F1 scores, partition 2 is in the Han River flood plain, characterized by high F2 scores, and partition 3 is in the rest of the area, where high F3 scores dominate (Fig. 8d).

5.4. Baseline values

Through PCA four, five and six components were extracted in each subarea (Figs. 6S–8S (Supplementary material)). The scores

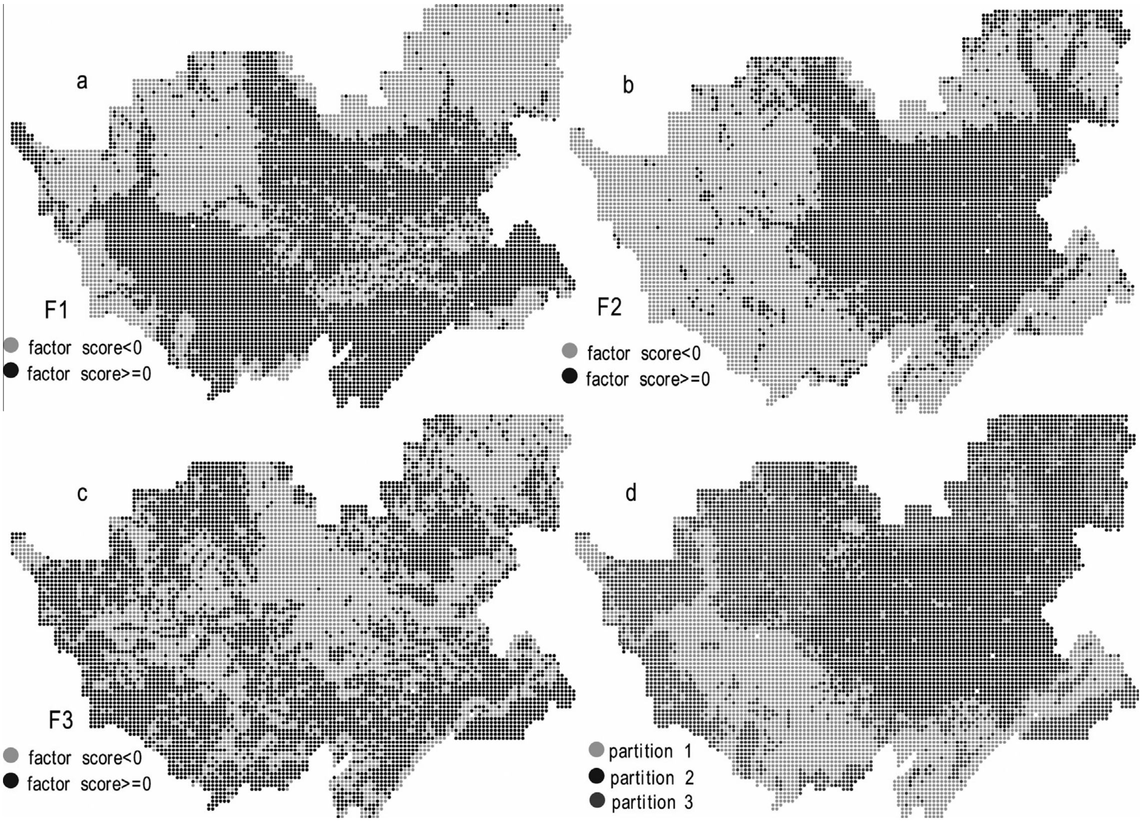

Fig. 8. The factor scores distribution and subarea division of Jianghan plain (a, b and c are distribution of the factor scores of F1, F2 and F3, respectively; d is the subarea division of Jianghan plain; detailed sample scores in the F1, F2 and F3 are shown in Fig. 5S).

and loadings of the principal components can provide information on the geological or geochemical processes. To address this issue, the results of PCA of the clr transformed data are shown in biplot. It is noted that the elements can be subdivided into three groups in subarea 1 (Fig. 9S (Supplementary material)). The first group, including zirconium and silicon which usually concentrate in fast-settling coarse and heavy minerals, has high scores in PC1. The second group characterized by high score in PC2 mainly consists of elements which have low solubilities in waters and preferential enriched in felsic igneous rocks, e.g. Ti, Nb, La, Y and Ce. Intermediate solubility elements (group 3; i.e. Ba, Rb, organic carbon, Al, Ga, Be, V, Sc, Co and B) exhibit the lowest scores in both PC1 and PC2. These elements are probably concentrate in fine-grained sedimentary rocks and have strong correlation with organic carbon that adsorbed easily by clay. PC1 is mainly influenced by group 1 (positive loadings) and group 3 elements (negative loadings) and to a lesser extent by group 2 elements (negative loadings) (Table 5). While PC2 is weakly dominated by the first two groups (positive loadings). In subarea 2, the grouping is almost same as that of subarea 1 excepting adding pH to group 1 and Th to group 3. All the groups exhibit low scores in both PC1 and PC2. PC1 is mainly influenced by group 1 (positive loadings) and group 3 elements (negative loadings). PC2 is dominated by the group 1 and group 2 elements (negative loadings) (Fig. 10S (Supplementary material)). In subarea 3, the PC1 and PC2 contain less information of organic carbon and Th but more information of pH than those of subarea 1. So the second group includes Ti, Nb, La, Y, Th and Ce and the third group includes Ba, pH, Al, Rb, Ga, Be, V, Sc, Co, and B. Both the high scores of PC1 and PC2 are driven by the third group. PC1 is influenced by group 3 elements (negative loadings) and PC2 is influenced by group 1 and group 2 elements (positive loadings) (Fig. 11S (Supplementary material)).

Table 5
Principal component of partition 1 of soil in the Jianghan plain (the other 2 partitions are in Appendix).

	PC1	PC2	PC3	PC4
B	−0.20	0.09	0.22	−0.34
Be	−0.28	−0.10	0.11	0.16
Ba	−0.19	−0.33	0.23	0.24
Co	−0.30	0.08	−0.13	−0.01
La	−0.18	0.36	−0.01	0.53
Nb	−0.24	0.32	0.02	−0.29
Rb	−0.29	−0.19	0.17	0.04
Ti	−0.24	0.27	−0.22	−0.32
V	−0.30	−0.05	−0.16	−0.13
Y	−0.15	0.31	0.34	−0.18
Zr	0.20	0.40	0.16	−0.11
Ce	−0.17	0.40	0.09	0.44
Ga	−0.30	−0.14	0.07	−0.02
Sc	−0.30	0.04	−0.08	0.12
Org. C	−0.15	−0.23	0.26	−0.20
Si	0.25	0.02	0.40	0.13
Al	−0.29	−0.18	0.11	0.04
pH	−0.06	−0.06	−0.61	0.06
λ	9.77	2.83	1.89	0.88
Contribution rate (%)	54.25	15.75	10.51	4.90
Cumulative (%)	54.25	70.00	80.51	85.41

Y: The Yttrium element.

The coefficients of the PCs and estimated baselines from the regression analyses for each subarea are shown in Table 6 and 3S. The baseline values are transformed from the constants. The baselines of Cd are 0.34 ppm, 0.29 ppm, and 0.17 ppm, respectively. The baselines of Pb are 30.2 ppm, 25.7 ppm, and 29.5 ppm, respectively. The Zn baselines are 97.7 ppm, 91.2 ppm and 58.9 ppm, respectively. All three elements have the highest baselines in subarea (partition) 1 and the lowest baselines for Cd and Zn in subarea (partition) 3, subarea 2 and subarea 2 has the lowest Pb baseline. The exceptional values, possibly anthropogenically influenced, are identified by the exceptional residuals (the error between the actual value and the predicted value) in the boxplots. Exceptional values of Cd, Pb and Zn identified by this procedure are shown on Fig. 9. The outliers are homogeneously distributed with the partition method compared with the statistics using quartile. No large area with high outliers of Cd, Pb and Zn emerge, indicating, probably, that there is not widespread regional anthropogenic contamination, but that there is some significant local point source contamination.

6. Discussion

6.1. Factors influencing the spatial distribution of trace elements

Compared to the National and Hubei province background, the higher element concentration of Jianghan plain are mainly due to the parent material (Qinling orogenic belt and Yangtze block) and soil properties (especially the fine-grained soil with a high clay mineral content). The sedimentation environment of the study area exhibits weak hydrodynamism, such as Quaternary alluvial, lacustrine and alluvial lacustrine environments, which formed the fine-grained soil with higher clay mineral content. In this research, clay minerals account for 60% of the soil mass in some areas (Li and Ma, 2008). Because of the fine grain size, large surface areas and strong adsorption of clay minerals, the heavy metal content is higher in the soils of plain areas than those of other landforms (Donoghue et al., 1998; Bradl, 2004; Rice, 1999; Salminen and Gregorauskiene, 2000). Though the ranges of Pb and Zn values are large their CVs are low. For Pb the data are unimodal and the frequency of the high values is low. The double-peak, bimodal, profiles of Cd and Zn indicate that at least two populations are present, which reflects the complexity of soils of the Jianghan plain. Thus it is not reasonable to determine the baseline values using traditional methods.

Based on the distribution maps, the spatial distributions of the Cd, Pb and Zn are influenced by the landform types and soils' parent material. The heavy metal concentrations of the surface soil in

Table 6
Data from the regression analysis in the Jianghan plain.

		Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Constant	Baseline (ppm)
Subarea 1	Cd	−0.018	0.0048	−0.0011	−0.01			−0.47	0.34
	Pb	−0.014	−0.0026	0.007	−0.0074			1.48	30.20
	Zn	−0.021	−0.0024	−0.0048	−0.0081			1.99	97.72
Subarea 2	Cd	−0.031	0.042	−0.021	0.011	0.032		−0.534	0.29
	Pb	−0.014	−0.0076	0.025	−0.0067	0.01		1.41	25.70
	Zn	−0.026	0.012	0.00068	0.0023	0.0042		1.96	91.20
Subarea 3	Cd	−0.019	−0.017	0.013	0.052	−0.076	0.013	−0.78	0.17
	Pb	−0.0092	−0.0012	0.0017	0.011	0.0025	−0.0081	1.47	29.51
	Zn	−0.026	−0.017	0.02	0.0056	−0.47	−0.0055	1.77	58.88

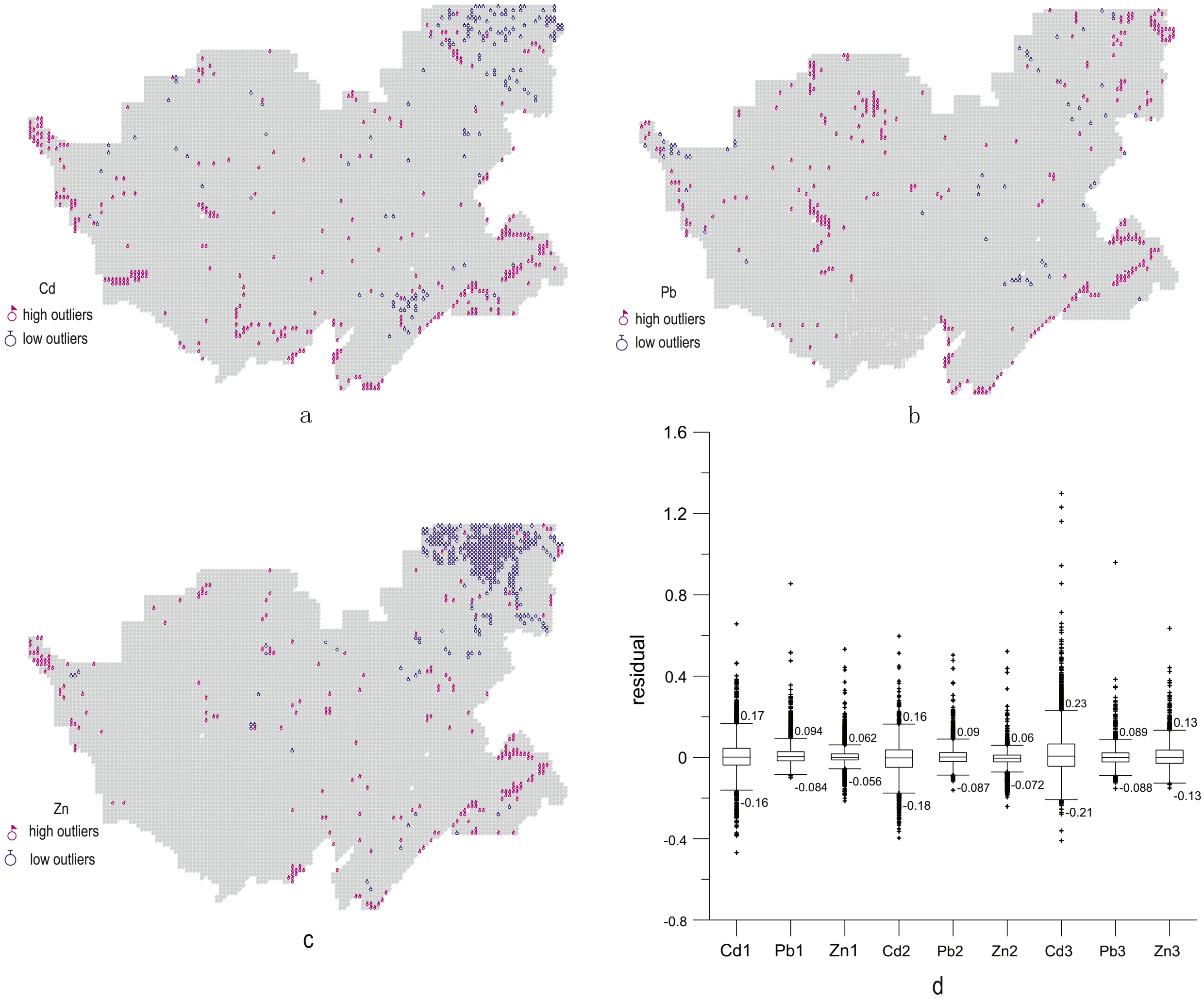

Fig. 9. Outliers of Cd, Pb and Zn data in the topsoil of the Jianghan plain (a, b, and c are the outliers of Cd, Pb and Zn data, respectively; d is the boxplot of residuals of Cd, Pb and Zn in three subareas).

the Jianghan plain have the following distributional features: (a) the soils in Yangtze drainage system have higher Cd, Pb and Zn contents than those in the other drainage systems. As mentioned in Section 5.1, both the supracrustal rocks of the two drainages are enriched in Cd, Pb and Zn. The high content of heavy metals in the soil is probably due to the high pH value resulting from high CaO and MgO concentrations of the crust of Yangtze block; (b) the soils in the low plain are higher in Cd, Pb and Zn contents than those in the other landform types; and (c) the soils in the Holocene series sediments are higher in Cd, Pb and Zn than those in the Pleistocene series sediments. This is probably driven by the clay in the Holocene sediments.

The reason for the variable vertical distribution in the Gongan and Hanchuan soil profiles may be that the columns are near rivers, where the sedimentary environment and source are unstable due to numerous floods. Previous studies have also suggested that Cd from anthropogenic sources exists mainly in the surface soil, and its concentration is very low in deep layers. For instance, approximately 80% of the Cd from anthropogenic has accumulated in the 0–30 cm layer in the irrigated soil of Shenyang, China (Xu et al., 2007). Thus the samples from deep soil could not provide the accurate information of surface environment. It is unreasonable to determine the geochemical baseline value for some metal elements such as Cd, Pb and Zn using deep layer samples instead of surface samples in the Jianghan plain.

As mentioned above, the Cd, Pb and Zn contents are controlled by both natural (mainly) and to a lesser extent anthropogenic processes. The natural processes, including the geological setting, landform, soil type, and the sources of sediment in the alluvial plains (influenced by Yangtze and Han rivers), are variable; the anthropogenic factors, which affect the distribution and variation of heavy metals in the surface soil, include industrial activity, mining, urban construction and agricultural activities. Industrial waste exported from mines located in the west of the Jianghan plain is transported into the rivers. Agricultural fertiliser can also change the content and valence of the elements in soil (Chen et al., 1997).

6.2. About the baseline calculation method

In the FA, the three factors indicated are related to: parent materials; drainage systems; and soil types, respectively. Therefore, it is reasonable to base a partition on the factors, which reflect most of the physico-chemical information and factors influencing the elements distributions. After the subarea division, the data of Cd, Pb and Zn are shown in histogram plots (Fig. 12S (Supplementary material)). In each subarea the data are normal distribution except the data of Cd in partition 2 which has a bimodal distribution. The subarea division potentially solves the problem of multiple data populations. Then prediction of the baselines of Cd, Pb and Zn by regression analyses is controlled by the selected elements.

Thus this baseline prediction method is more effective than the traditional ones, especially when the target element has a bimodal distribution. Additionally, the median values of Cd, Pb and Zn in each subarea are very different from those in the whole area (Fig. 13S (Supplementary material)). The medians of Cd and Zn data in subarea 1 and 2 (Cd, 0.35 ppm and 0.32 ppm; Zn, 102.8 ppm and 92.3 ppm) are higher than those in the whole subarea (Cd, 0.29 ppm; Zn, 86.5 ppm); Subarea 2 has notable lower Cd (0.16 ppm) and Zn (57.8 ppm) medians than the whole area; The differences among the 4 medians of Pb (whole area, 28.9 ppm; subarea 1, 31.1 ppm; subarea 2, 26.6 ppm; subarea 3, 29.7 ppm) are relatively smaller than those of Cd and Zn, and the subarea 2 has the lowest median value. Compared with the backgrounds obtained through the iterative method (0.26 ppm, 29.4 ppm and 80.4 ppm for Cd, Pb and Zn, respectively) (Li and Ma, 2008), the predicted baselines in this paper are preferred and more reasonable. The baseline values will be useful in land use planning, assessment of soil contamination and other works. For example, the baselines can be used to develop trigger values in assessing potential soil contamination and have been used to determine upper and lower guideline values in basic risk assessment (Jarva et al., 2010). The method presented here not only avoids anthropogenic factors in the calculation and gives an accurate estimation of the baseline value, but it also makes full use of the database. The proxies extracted by PCA provide much more geological information than a single normalizer.

7. Conclusion

Based on the relationships amongst Cd, Pb and Zn and the dominant factors of their behavior, the Jianghan plain is divided into 3 subareas by factor analysis, which reflect the parent materials, drainage systems and soil types controlling. The baselines in every subarea are predicted by PCR analysis. The principle components provide an abundance of synthesised information regarding the samples, making the baseline more quantitative and reasonable. This fitted geochemical model is an attempt to predict the soil geochemical baseline. The study shows that the distributions of Cd, Pb and Zn and the outliers are heterogeneous and differ in every subarea. Consequently, predicting the soil baseline on a subarea basis in the Jianghan plain is advisable, and it provides an important reference for environmental protection and legislation. This paper also proposes an approach for geochemistry baseline studies in large areas with a complex environment. This method should be developed, and a database of soil baselines in the Jianghan plain should be established.

Acknowledgements

This research is financially supported by the China Geological Survey (06010017). The authors thank the executive editor Michael Kersten, Robert G. Garrett and an anonymous reviewer for improving the research; we thank Bin Zhao and Mingzuan Xu for their help with the field work; and we thank Deyi Xu, Qiang Ma, Salah Fadlallah Awadelseid, Zhixuan Han, Jingjing Gong, Ganggang Yang, Zhide Ou and Yongdong Mao for improving the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.apgeochem.2014.07.019.

References

Albanese, S., De Vivo, B., Lima, A., Cicchella, D., Civitillo, D., Cosenza, A., 2010. Geochemical baselines and risk assessment of the Bagnoli brownfield site coastal sea sediments (Naples, Italy). J. Geochem. Explor. 105 (1–2), 19–33.

Amorosi, A., Sammartino, I., 2007. Influence of sediment provenance on background values of potentially toxic metals from near-surface sediments of Pocoastal plain (Italy). Int. J. Earth Sci. 96 (2), 389–396.

Baize, D., Sterckeman, T., 2001. Of the necessity of knowledge of the natural pedogeochemical background content in the evaluation of the contamination of soils by trace elements. Sci. Total Environ. 264 (1–2), 127–139.

Bradl, H.B., 2004. Adsorption of heavy metal ions on soils and soils constituents. J. Colloid Interface Sci. 277 (1), 1–18.

Chen, T.B., Wong, J., Zhou, H.Y., Wong, M.H., 1997. Assessment of trace metal distribution and contamination in surface soils of Hong Kong. Environ. Pollut. 96 (1), 61–68.

Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. J. Geochem. Explor. 51 (2), 109–130.

Cicchella, D., De Vivo, B., Lima, A., 2005. Background and baseline concentration values of elements harmful to human health in the volcanic soils of the metropolitan and provincial areas of Napoli (Italy). Geochem.: Explor. Environ., Anal. 5 (1), 29–40.

Cohen, D.R., Rutherford, N.F., Morisseau, E., Zissimos, A.M., 2012. Geochemical patterns in the soils of Cyprus. Sci. Total Environ. 420, 250–262.

Covelli, S., Fontolan, G., 1997. Application of a normalization procedure in determining regional geochemical baselines. Environ. Geol. 30 (1), 34–45.

Darnley, A.G., 1997. A global geochemical reference network: the foundation for geochemical baselines. J. Geochem. Explor. 60 (1), 1–5.

De Vivo, B., Lima, A., Bove, MA., Albanese, S., Cicchella, D., Sabatini, G., Di Lella, LA., Protano, G., Riccobono, F., Frizzo, P., Raccagni, L., 2008. Environmental geochemical maps of Italy from the FOREGS database. Geochem.: Explor. Environ., Anal. 8 (3-4), 267–277.

Desaules, A., 2012. Critical evaluation of soil contamination assessment methods for trace metals. Sci. Total Environ. 462, 120–131.

Donoghue, J.F., Ragland, P.C., Chen, Z.Q., Trimble, C.A., 1998. Standardization of metal concentrations in sediments using regression residuals: an example from a large lake in Florida, USA. Environ. Geol. 36 (1), 65–76.

Filzmoser, P., Hron, K., Reimmn, C., 2009. Univariate statistical analysis of environmental (compositional) data: problems and possibilities. Sci. Total Environ. 23 (407), 6100–6108.

Galán, E., Fernández-Caliani, J.C., González, I., Aparicio, P., Romero, A., 2008. Influence of geological setting on geochemical baselines of trace elements in soils. Application to soils of South-West Spain. J. Geochem. Explor. 98 (3), 89–106.

Gałuszka, A., 2007. A review of geochemical background concepts and an example using data from Poland. Environ. Geol. 52 (5), 861–870.

Gao, S., Luo, T., Zhang, B., Zhang, H., Han, Y., Zhao, Zi, Hu, Y., 1998. Chemical composition of the continental crust as revealed by studies in East China. Geochim. Cosmochim. Acta 62 (11), 1959–1975.

Grünfeld, K., 2007. The separation of multi-element spatial patterns in till geochemistry of southeastern Sweden combining GIS, principal component analysis and high-dimensional visualization. Geochem.: Explor. Environ., Anal. 7 (4), 303–318.

Guillén, M.T., Delgado, J., Albanese, S., Nieto, J.M., Lima, A., De Vivo, B., 2011. Environmental geochemical mapping of Huelva municipality soils (SW Spain) as a tool to determine background and baseline values. J. Geochem. Explor. 109 (1–3), 59–69.

Jarva, J., Tarvainen, T., Reinikainen, J., Eklund, M., 2010. TAPIR—Finnish national geochemical baseline database. Sci. Total Environ. 408 (20), 4385–4395.

Jiang, J.B., Wang, J., Liu, Sh.Q., Lin, Ch.Y., He, M.Ch., Liu, X.T., 2013. Background, baseline, normalization, and contamination of heavy metals in the Liao River Watershed sediments of China. J. Asian Earth Sci. 73, 87–94.

Lattin, J.M., Carroll, J.D., Green, P.E., 2003. Analyzing multivariate data. Brooks.

Li, F.L., Ma, Zh.D., 2008. Report about Cd and other elements of the alluvial economic region in Jianghan valley. (unpublished).

López, M., González, I., Romero, A., 2008. Trace elements contamination of agricultural soils affected by sulphide exploitation (Iberian Pyrite Belt, Sw Spain). Environ. Geol. 54 (4), 805–818.

Ma, Zh.D., Zhang, D.C., Bi, X.Y., Liu, G.Q., Ren, L.M., Quan, H.L., 2005. Origin of cadmium high-value zones along the Yangtze River and Hanjiang River in Wuhan, Hubei, China. Geol. Bull. China 24 (8), 740–743.

Meklit, T., Van Meirvenne, M., Verstraete, S., Bonroy, J., Tack, F., 2009. Combining marginal and spatial outliers identification to optimize the mapping of the regional geochemical baseline concentration of soil heavy metals. Geoderma 148 (3–4), 413–420.

Reimann, C., Garrett, R.G., 2005. Geochemical background—concept and reality. Sci. Total Environ. 350 (1), 12–27.

Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R., 2008. Statistical data analysis explained: applied environmental statistics with R.

Rice, K.C., 1999. Trace-element concentrations in streambed sediment across the conterminous United States. Environ. Sci. Technol. 33 (15), 2499–2504.

Salminen, R., Gregorauskiene, V., 2000. Considerations regarding the definition of a geochemical baseline of elements in the surficial materials in areas differing in basic geology. Appl. Geochem. 15 (5), 647–653.

Salminen, R., Tarvainen, T., 1997. The problem of defining geochemical baselines. A case study of selected elements and geological materials in Finland. J. Geochem. Explor. 60 (1), 91–98.

Salminen, R., Bogatyrev, I., Chekushin, V., Glavatskikh, S.P., Gregorauskiene, V., Niskavaara, H., Selenok, L., Tenhola, M., Tomilina, O., 2004. Geochemical baselines of nickel and chromium in various surficial materials in the Barents region, NW Russia and Finland. Geostand. Geoanal. Res. 28 (2), 333–341.

Smyth, D., Johnson, C.C., 2011. Distribution of iodine in soils of Northern Ireland. Geochem.: Explor. Environ., Anal. 11 (1), 25–39.

Tack, F.M.G., Vanhaesebroeck, T., Verloo, M.G., Van Rompaey, K., Van Ranst, E., 2005. Mercury baseline levels in Flemish soils (Belgium). Environ. Pollut. 134 (1), 173–179.

Tapia, J., Audry, S., Townley, B., Duprey, J.L., 2012. Geochemical background, baseline and origin of contaminants from sediments in the mining-impacted Altiplano and Eastern Cordillera of Oruro, Bolivia. Geochem.: Explor. Environ., Anal. 12 (1), 3–20.

Tarvainen, T., Paukola, T., 1998. Use of geochemical databases to delineate risk areas for contaminated groundwater. J. Geochem. Explor. 64 (1–3), 177–184.

Teng, Y., Ni, S., Wang, J., Niu, L., 2009. Geochemical baseline of trace elements in the sediment in Dexing area, South China. Environ. Geol. 57 (7), 1649–1660.

Wang, X.Q., 2012. Global Geochemical baselines: understanding the past and predicting the future. Earth Sci. Front. 19 (3), 007–018 (In Chinese with English Abstract).

Wang, Q., Dong, Y., Zhou, G., Zheng, W., 2007a. Soil geochemical and environmental background values of agricultural regions in Zhejiang province. J. Ecol. Rural Environ. 5 (26), 590–597 (In Chinese with English Abstract).

Wang, X., Zhang, Q., Zhou, G., 2007b. National-scale geochemical mapping projects in China. Geostand. Geoanal. Res. 31 (4), 311–320.

Wang, Z., LandonDarilek, J., Zhao, Y., Huang, B., Sun, W., 2011. Defining soil geochemical baselines at small scales using geochemical common factors and soil organic matter as normalizers. J. Soils Sediments 11 (1), 3–14.

Xie, X., Ren, T., 1993. National geochemical mapping and environmental geochemistry–progress in China. J. Geochem. Explor. 49 (1–2), 15–34.

Xu, S., Guo, S., Hu, X., 2007. Revaluation of soil heavy metals pollution in Zhangshi Irrigation Area of Shenyang and analysis of Cd forms in soil. Chin. J. Appl. Ecol. 18 (9), 2144–2148 (In Chinese with English Abstract).

Yang, D., Li, F., Xia, J., 2009. Sources analysis of Cd in Qingjiang stream sediment, Hubei province. Geol. Sci. Technol. Inform. 1 (28), 99–104 (In Chinese with English Abstract).

Zhao, B., Ouyang, J.P., Li, F.L., Xu, M.Z., Zhang, D.C., Yang, J., 2008. Distribution characters of element Cd in different physiognomy type alluvial soils of Yangtze River (Gong'an segment). Geol. Sci. Technol. Inform. 27 (5), 43–46 (In Chinese with English Abstract).