Journal of Geochemical Exploration 196 (2019) 1–7

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

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

Spatial distribution mapping of Hg contamination in subclass agricultural soils using GIS enhanced multiple linear regression

Xiaolu Jiang[a,b], Bin Zou[a,b,*], Huihui Feng[b], Jingwen Tang[c], Yulong Tu[b], Xiuge Zhao[d]

[a] Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, Ministry of Education, Central South University, Changsha, Hunan 410083, China
[b] School of Geosciences and Info-Physics, Central South University, Changsha, Hunan 410083, China
[c] Chinese National Engineering Research Center for Control and Treatment of Heavy Metal Pollution, Central South University, Changsha, Hunan 410083, China
[d] State Key Lab of Environmental Criteria and Risk Assessment, CRAES, Beijing 100012, China

ARTICLE INFO

Keywords:
Hg
Soil pollution
Mapping
Subclass agricultural lands
Multiple linear regression

ABSTRACT

In order to accurately determine the spatial distribution of soil Hg pollution in one agricultural land, 104 topsoil samples were studied by means of three subclass agricultural lands, including vegetable field, paddy field and orchard field in the southeast of China. The method of multiple linear regression modeling combined with the spatial analysis is adopted to recognize the driving factors of soil Hg contamination in each subclass agricultural land. Consequently, corresponding spatial distribution maps of estimated soil Hg concentrations in vegetable field and paddy field are plotted. Results demonstrate greater mean soil Hg concentrations in vegetable field and paddy field over orchard field, but both of them are lower than the maximum permissible concentration in the Chinese Environmental Quality Standard for agricultural soils. In vegetable field, soil Hg concentrations are closely interrelated with chimneys which have the least distances to their around soil samples. In paddy field, both the soil sampling slopes and distances from soil samples to the nearest roads are the dominant driving factors of soil Hg concentration variance. The spatial distribution of soil Hg concentrations reveals several localized hotspots in the north for the vegetable field, while in the paddy field, spreads evenly across the space. These results suggest that while the exhaust emission from industrial chimneys might drive the contamination difference of soil Hg in vegetable field in the study area, surface runoff and traffic contamination from the roads might be other factors causing those spatial variances of soil Hg contamination in paddy field. Quantification of soil Hg concentrations in agricultural lands at subclass level may provide a more accurate basis for taking appropriate prevention and controlling action to protect agricultural soil quality and human health.

1. Introduction

The toxic impact of Hg in agricultural soils has gradually become a world concern, because of its mobility, volatility, accumulation through food chain, and ultimate harm to human health (De Simone et al., 2015; Ottesen et al., 2013). Accurately identifying spatial distribution of Hg contamination become more important to improve the understanding of the soil environmental conditions and implement the control and prevention of soil Hg pollution in agricultural land (Micó et al., 2006). In soil Hg contamination investigation, traditional soil surface sampling is a common and reliable way to obtain Hg concentration data in agricultural soil (Li et al., 2008). However, this site based traditional sampling method is cost intensive and representativeness limited, which is insufficient for a wide range of continuous soil Hg contamination survey in a regional area.

Regarding this issue, researchers usually rely on traditional spatial interpolation methods to construct continuous surface distribution of soil Hg concentration in agricultural land based on a discrete set of sampling data (Lu et al., 2012; Xu et al., 2014; Zhou et al., 2015), such as Kriging, inverse distance interpolation, spline interpolation (Cheng et al., 2007; Xie et al., 2011; Yang and Wang, 2008). Theoretically, these methods could overcome the defects of sparse sampling somewhat, but their accuracies are still far from being satisfactory in practice. This is because not only the soil Hg concentrations at sampling locations, but numerous auxiliary geo-environment elements might also strongly affect the spatial distribution of soil Hg contamination (Huang and Jin, 2008; Wuana and Okieimen, 2011). And then, new machine learning methods have been gradually used in the Christoforidis estimation of soil Hg concentration, such as neural network interpolation models (*e.g.*, generalized regression neural network, radial basis

* Corresponding author at: School of Geosciences and Info-Physics, Central South University, Changsha, Hunan 410083, China.
E-mail address: 210010@csu.edu.cn (B. Zou).

https://doi.org/10.1016/j.gexplo.2018.10.002
Received 15 May 2017; Received in revised form 5 September 2018; Accepted 3 October 2018
Available online 04 October 2018
0375-6742/

function (RBF) neural network) (Hu et al., 2007; Lu et al., 2004; Zou et al., 2015) and fuzzy C-means clustering models (Amini et al., 2005). However, these methods benefit from abundant samples which are not suitable for soil Hg pollution investigation. Further, a series of statistical analysis approaches superior to the simple interpolation or new machine learning have been adopted for driving factors recognition, which typically comprise linear regression modeling (*e.g.*, geographically weighted regression) (Fang et al., 2016; Yang et al., 2015; Zhai et al., 2018; Zou et al., 2016), and non-linear modeling (*e.g.*, generalized additive model) (Zou et al., 2017a). Among them, the multiple linear regression (MLR) is recognized as a relatively simple and applicable one for soil Hg concentration estimation by fusing the auxiliary driving factors identified (Caeiro et al., 2005; D'Hose et al., 2014; Zeng et al., 2011).

In terms of main class agricultural land (including all the three types of agricultural lands, *i.e.*, vegetable field, paddy field and orchard field), current soil Hg concentration mapping technique based on driving factors identification has been able to cursorily estimate the spatial distribution of soil Hg concentration in a large-scale area (Yuan et al., 2014). However, due to the complexity of soil Hg contamination sources and transportation mechanism, the formation and dispersion processes of soil Hg contamination may be discrepant within the main class agricultural land over a large geographical area. These differences can be attributed to the terrain, crop types, irrigation types, and the distribution effects of industry enterprises and roads (Han et al., 2006; Micó et al., 2006). This inversely means the inconsistency of the driving factors of soil Hg contamination in the inner subclass agricultural lands (*i.e.*, vegetable field, paddy field and orchard field) (Gagiu et al., 2015). Therefore, to meet the need of fine agriculture governance, it is necessary to further improve the accuracy of soil Hg concentration estimation by more accurately identifying the different driving factors of soil Hg concentration variance in the various subclass agricultural lands.

Thus, this study was to test the hypothesis that whether factors driving soil Hg concentration variations at main class and sub-class level agricultural lands are different, and consequently depict their impacts on spatial distribution mapping of soil Hg contamination across the study area. Results of this study are expected to contribute or provide some insight to the soil conservation strategy during fine agriculture governance.

2. Materials and methods

2.1. Study area

The study area is located in the north of Pearl River Delta, the southeastern part of China (Fig. 1). As a traditionally agricultural-based economy city, the study area has large agricultural land which is mainly cultivated as vegetable field, paddy field, and orchard field. The rapid industrialization in this area has considerably increased chimneys' exhaust emissions, solid waste accumulation and sewage discharges of Hg contamination, which posed obvious effects on local agriculture (Qiu et al., 2009). The area is surrounded by mountains to the north and south, with a wide plain area between them. This area has a typical subtropical climate with an annual temperature average approximately 22 °C, the mean annual precipitation 1897 mm, and the prevailing wind from southeast to northwest. These geo-environment elements further cause soil Hg contamination diffusion through surface runoff and atmospheric deposition in different subclass agricultural lands (Cai et al., 2012).

2.2. Soil samples and chemical analysis

The sampling route was designed by means of a completely randomized layout, on the basis of the agricultural production, industrial distribution, waste discharging, roads and river networks, according to FOREGS Geochemical Mapping Field Manual (Salminen et al., 1998). The total 104 topsoil (0–20 cm) samples were collected from the agricultural land, including vegetable field (53), paddy field (40), and orchard field (11) (Fig. 1). Each main soil sample, consisted of nine subsamples, were randomly taken from the surroundings of each site, pooled and homogenized, then reduced to the weight of 300 g to form a representative sample (Micó et al., 2006). Each sample is at least 200 m apart from the next one. The longitudes and latitudes of sampled locations were recorded by a global positioning system (GPS) receiver.

Each sample was sealed and transported to the laboratory through a polyethylene bag, air dried for 24 h and redried at 45 °C in an oven until constant weight, and finally stored in plastic reagent bottles. All plastic instruments used for sample collection were cleaned sequentially with a phosphate-free detergent, rinsed with distilled water, then with 10% nitric acid, and finally with distilled water. The samples were then passed through a 100-mesh sieve to analyze Hg, and 2 mm sieve to analyze soil pH. For determining concentration of soil Hg, 0.5 g dried soil of each sample was treated with concentrated HNO_3 (10 ml) and H_2SO_4 (20 ml) for 1 h at 70 °C. When being cool, the volume of the digest was kept to 100 ml with deionized water. An aliquot of the solution was added with 1 ml $K_2Cr_2O_7$ solution and analyzed by cold vapor atomic fluorescence spectroscopy (AFS-9230) after reducing Hg with a stannous chloride solution. Quality assurance and quality control procedures were conducted by using the standard reference material (GBW07401-GBW07408) with each batch of samples (one blank and standard for each 10 samples). Soil pH was measured after shaking 10 g in a suspension of soil: water at a ratio of 1:5 for 30 min, using a glass electrode (ISO-10390 2005) (Briki et al., 2015). All of the experiments were conducted at the Soil and Fertilizer Institute (Hunan, China).

2.3. Spatial data processing and statistical analysis

Fossil fuel-fired power plants, mining plants, processing of no-ferrous metals plants, cement plants, municipal and medical waste incinerators are the general sources of Hg contamination (Frentiu et al., 2013). However, investigations during soil sampling visit in this study actually had preliminarily demonstrated that the soil Hg contamination in study area might be influenced by the intense traffic nearby, proximity to industries, irrigation types, and land morphology (Chen et al., 2005; Guney et al., 2010; Mostert et al., 2012). As a result, the distance from roads, rivers and chimneys, elevation, slope and land use were selected as predictors potentially indicating the variations of soil Hg concentration in this study. The positions of chimneys were extracted from the point of interest (POI) of industrial enterprises in China. The river network, road network and the land use raster data (30 m resolution) including construction land, agricultural land (vegetable, paddy and orchard field), forest land, lawn, water area and unexploited land, were provided by the national geographical condition cloud platform. The DEM data (30 m resolution) were downloaded from the United States Geological Survey. Also, all these data were collected in the same time year.

Geographic information system (GIS) technique, including proximity analysis, buffer analysis, and land morphology analysis, was adopted in spatial data processing to extract geographical driving factors of the soil Hg concentration variance in agricultural land. The proximity analysis was used to determine the distances between the sampling sites and nearest roads (*DRD*), nearest rivers (*DR*), and nearest chimneys (*DC*). The buffer analysis was applied to calculate the proportion of construction land in all types of land (*CP*) by creating buffer polygons around sampling sites to a specified distance, which ranges from 25 to 200 m with an interval of 25 m. And for example, the *CP* (*100*) expressed the construction land portion in a buffer of 100 m. The land morphology analysis was used to extract the elevation (*SE*) and slopes (*SS*) at each sampling site from the DEM data. All the spatial analyses were conducted using ArcGIS 10.1.

The standard descriptive statistical parameters (*e.g.*, mean,

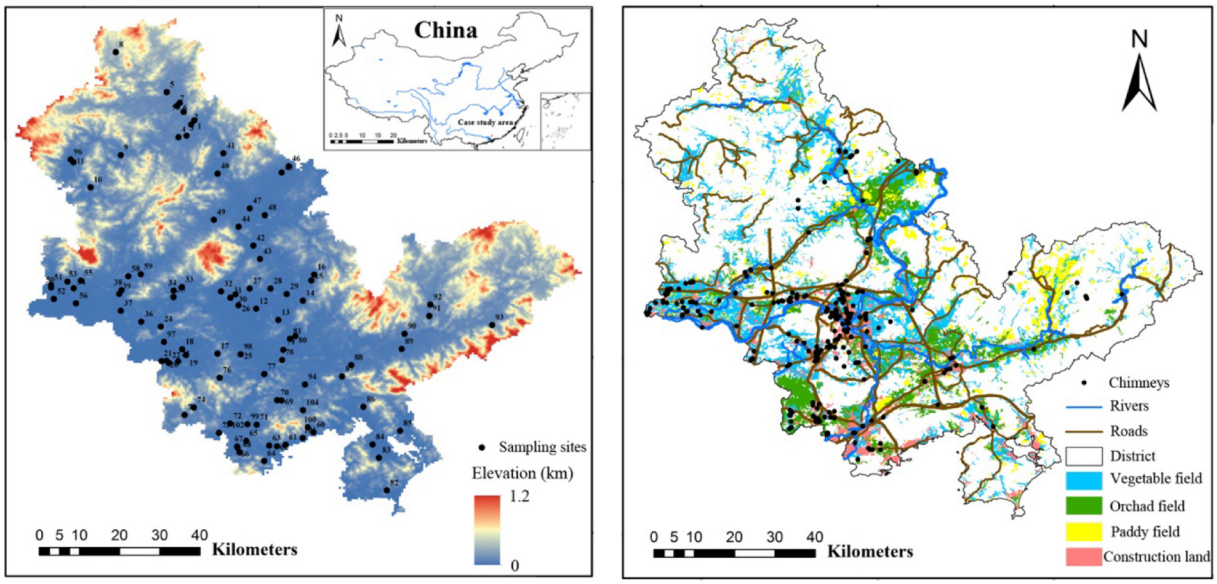

Fig. 1. Location of soil sampling sites and potential geo-environmental elements in the study area.

maximum, and standard deviation) of soil Hg concentration were calculated in order to evaluate the basic status of soil Hg contamination. Since the distribution of soil Hg concentration most often differs from normality, not all statistical methods are appropriate for analysis without previous transformations, so does the MLR analysis (Gagiu et al., 2015). Outliers removing, logarithm and square root transformation are three frequent data transformation methods (Schwertman and De Silva, 2007). Among them, square root transformation is found more effective and consequently utilized. In the process, Pearson's Chi-square test was employed for normality checking of the transformed soil Hg concentrations.

In addition, Pearson's correlation analysis, denoted by indices of correlation coefficient (r) and significance level (P), was adopted to initially identify the relationship among the soil Hg concentration and the potential driving factors (*i.e.*, pH, industrial sites, river network, road network, land use, DEM) (Francouría et al., 2009; Lu et al., 2010). This can contribute to determine accurate driving factors of soil Hg contamination in different subclass agricultural lands (*i.e.*, vegetable field, paddy field, and orchard field). Correlations between potential predictors were further investigated. Only the predictors without strong correlations can be input into the regression analysis at the same time to avoid multi-collinearity (Liu et al., 2011). The correlation analysis was conducted with IBM SPSS Statistics 19.0 for Windows.

2.4. Multiple linear regression modeling and soil Hg concentration mapping

Based on the modeling variables selection through the correlation analysis, MLR modeling was adopted to accurately identify the driving factors of soil Hg contamination in subclass agricultural lands (*i.e.*, vegetable field, paddy field, and orchard field), to compare with the results of driving factors identification in the main class agricultural land in order to verify the necessary and accuracy. The MLR model can be expressed as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon \quad (1)$$

where, y is the soil Hg concentration, x_0, x_1, … and x_p refer to the relevant driving factors, β_0 is the constant value, β_1, … β_p are regression coefficients, ε is the accidental error.

In the process of MLR analysis, the potential modeling variables selected in the correlation analysis were preferentially considered as the covariates to develop the MLR models through backward stepwise regression based on the ordinary least squares approach. All the variables were introduced to the model one by one until the model meets the requirement of significance ($P \leq 0.05$) and consequently the dominant driving factors were ascertained. As MLR modeling conditions, the Gauss-Markov assumption test was implemented to ensure the distribution similarity of soil Hg concentrations and values of related driving factors. The widely used '$n-1$ cross validation' was implemented to validate the reliability of MLR models. In this process, mean error (*ME*), mean absolute error (*MAE*), the mean relative error (*MRE*), and F test were employed (Zhai et al., 2017).

ArcGIS 10.1 was used to create fishnets (85,298 points) at a resolution of 1 km × 1 km in the agricultural land, including vegetable field (16,872 points), paddy field (40,296 points), and orchard field (28,130 points). The extracted model parameters at fishnet corners were then employed to estimate the soil Hg concentrations based on the MLR models in the main class and subclass agricultural lands, respectively. Then, all these estimated soil Hg concentrations at fishnet corners' locations were further interpolated as the continuous concentration surfaces. In addition, these spatial distribution maps of soil Hg concentration for main class and subclass agricultural lands were overlaid to demonstrate the differences caused by inner geographical driving factors.

3. Results

3.1. Descriptive statistics of Hg concentration in the soil samples

The descriptive statistics of soil Hg concentration are shown in Table 1. As displayed, Hg concentrations of 104 soil samples ranges from 0.003 to 0.773 mg·kg^{-1}. The values of soil pH vary from 4.0 to

Table 1
The descriptive statistics of soil Hg concentrations.

Soil uses	Mean (mg·kg^{-1})	Maximum (mg·kg^{-1})	Minimum (mg·kg^{-1})	Standard deviation	χ^2
Main class agricultural land (104)	0.181	0.773	0.003	1.355	2.260
Vegetable field (53)	0.153	0.394	0.003	0.170	2.071
Paddy field (40)	0.215	0.773	0.060	0.611	6.593
Orchard field (11)	0.101	0.325	0.009	0.138	1.665

χ^2 represents these test results in 95% confidence level according to percentile table of χ^2 distribution.

8.2. In general, for the main class agricultural land in this study area, the mean value of soil Hg concentration is 0.181 $mg{\cdot}kg^{-1}$, which is below the National Environment Quality Grade II Standard for soil, 0.30 $mg{\cdot}kg^{-1}$ (Zhang et al., 2014), even if the maximum value is up to 0.773 $mg{\cdot}kg^{-1}$ belonging to the serious pollution level. In fact, the distribution of soil Hg concentration significantly varies in three subclass agricultural lands according to their respective mean, maximum, and minimum. Among them, the average values of soil Hg concentration are 0.153 $mg{\cdot}kg^{-1}$, 0.215 $mg{\cdot}kg^{-1}$ and 0.101 $mg{\cdot}kg^{-1}$ in vegetable field, paddy field and orchard field, respectively. In vegetable field and orchard field, the minimum values of soil Hg concentration are both < 0.01 $mg{\cdot}kg^{-1}$, the maximum values are both < 0.40 $mg{\cdot}kg^{-1}$. These values are much lower than those in paddy field (0.060 $mg{\cdot}kg^{-1}$ and 0.773 $mg{\cdot}kg^{-1}$). It clearly indicates that the soil Hg concentration in paddy field is the highest, followed by the vegetable field, and then the orchard field. Moreover, the soil Hg concentration displays quite heterogeneous distributions across the paddy field and therefore higher standard deviation, contrasting with the homogeneous distributions of soil Hg concentration in the orchard field with lower standard deviation. Therefore, this study excluded four outliers from all the 104 soil samples, including one sample from vegetable field, two samples from paddy field, and one sample from orchard field. After outliers removing, square root transformation was done for the four sets of soil sampling data. Finally, the soil Hg concentration met normal distribution according to the results of Pearson's Chi-square test (χ^2) (Table 1).

3.2. Correlation based modeling variables selection

The Pearson's correlation analysis between predictors showed that the independent variables of the group including *DR* and *SE* are correlated with another group including *DRD* and *SS*. Besides, variables of group in various buffers are also strongly correlated. According to the Pearson's correlation coefficients, this study reserved *DRD*, *SS* and *CP* (*100*) as the potential driving factors with the exclusion of *DR*, *SE* and *CP* in other different buffer areas. Correlation analysis results between soil Hg concentrations and variable values denoting driving factors are partially shown in Table 2. As illustrated, driving factors vary at three subclass agricultural lands. In vegetable field, soil Hg concentration negatively correlates to *DC*, with coefficient -0.271 ($P \leq 0.05$). In paddy field, they are negatively correlated to *SS*, with coefficient -0.602 ($P \leq 0.05$). However, in orchard field, no obvious covariate is found. For the main class agricultural land, soil Hg concentration has significant positive (0.243 when $P \leq 0.05$) or negative (-0.209 when $P \leq 0.05$) relation to *DC* and *CP* (in a buffer of 100 m), respectively.

3.3. Multiple linear regression modeling and driving factors identification

Table 3 shows the coefficients and significance of soil Hg concentration estimation models through MLR modeling method in two sub-class and the main class agricultural lands (*i.e.*, vegetable field and paddy field). Clearly, the significant variables of models are not exactly the same in different subclass agricultural lands, which are *DC* in the vegetable field, while *SS* and *DRD* in the paddy field. Therefore, in the vegetable field, the exhaust emission from industrial chimneys is the main driving factors of soil Hg contamination, while in the paddy field, combined with the action of surface runoff, traffic exhaust and solid waste from the roads are the main pathways of the soil Hg contamination accumulation. However, for the orchard field, significant anthropogenic effects on soil Hg contamination were not found in the case of this study. By contrast, in main class agricultural land, the significant variables of soil Hg concentration estimation model include *DC* and *SS*. Therefore, the driving factors of soil Hg contamination identified on the basis of main class agricultural land are the exhaust emission from industrial chimneys and the contamination accumulation along with surface runoff.

Table 3
Various multiple linear regression (MLR) model parameters for soil Hg concentration estimation in agricultural land.

Soil uses	Parameter	Coefficients	Sig.
Vegetable field	Total model	–	0.021
	(Constant)	0.326	0.000
	DC	-7.276×10^{-6}	0.021
Paddy field	Total model	–	0.013
	(Constant)	0.452	0.000
	SS	−0.034	0.011
	DRD	3.874×10^{-5}	0.033
Main class agricultural land	Total model	–	0.044
	(Constant)	0.379	0.000
	DC	5.486×10^{-6}	0.006
	SS	−0.021	0.045

DRD: the distances between the sampling sites and nearest roads; *DC*: the distances between the sampling sites and nearest chimneys; *SS*: the slopes of sampling sites.

Table 4 lists the validation results of soil Hg concentration estimation models to characterize the overall quality of the results with respect to the three models. As shown in Table 4, all the three MLR models passed the *F* Test, and the *MRE* values for model deviations in subclass agricultural land, including vegetable field and paddy field, are 3.79% and 2.05%, respectively. These results indeed outperform the MLR model for the main class agricultural land with the *MRE* value at 9.82%. Additionally, in terms of the model results of two subclass agricultural lands, the accuracy of the model in the paddy field is much higher than in the vegetable field according to the *ME*, *MAE*, and *MRE* values for model deviations.

3.4. Spatial distribution mapping of soil Hg concentration

The spatial distributions of soil Hg concentration for the vegetable field and paddy field aided with the two subclass agricultural land models in this study area are shown in Fig. 2a–b. The ranges of the simulated soil Hg concentration are classified into five levels taking the statistics into consideration. The mean values of soil Hg concentration

Table 2
Pearson correlation coefficients between soil Hg concentrations and potential driving factor values.

Covariates	Vegetable field		Paddy field		Orchard field		Main class agricultural land	
	r	*P*	*r*	*P*	*r*	*P*	*r*	*P*
pH	−0.077	0.583	0.022	0.896	0.088	0.808	−0.035	0.866
DRD	0.269	0.069	−0.034	0.131	0.102	0.744	0.026	0.806
DC	−0.271	0.021*	−0.010	0.951	−0.075	0.379	0.243	0.014*
SS	−0.021	0.880	−0.602	0.032*	−0.209	0.447	−0.032	0.112
CP (*100*)	−0.126	0.244	−0.033	0.763	0.511	0.078	−0.209	0.035*

DRD: the distances between the sampling sites and nearest roads; *DC*: the distances between the sampling sites and nearest chimneys; *SS*: the slopes of sampling sites; *CP* (*100*): construction land portion in a buffer of 100 m.
* Represents significant level at 0.05.

Table 4
The validation of soil Hg concentration estimation models.

Soil uses	*ME* ($mg \cdot kg^{-1}$)	*MAE* ($mg \cdot kg^{-1}$)	*MRE* (%)	*F* test
Vegetable field	0.001	0.005	3.79	8.26
Paddy field	0.004	0.004	2.05	7.24
Main class agricultural land	0.013	0.025	9.82	6.25

ME: mean error; *MAE*: mean absolute error; *MRE*: the mean relative error.

($0.225\ mg \cdot kg^{-1}$ in paddy field, $0.135\ mg \cdot kg^{-1}$ in vegetable field) illustrates that the paddy field soil is suffering from severe Hg pollution than vegetable field in this study area. Fig. 2a suggests that spatial distribution of soil Hg concentration demonstrates several localized hotspots in the north of the vegetable field. As shown in Fig. 2b, in paddy field, it is difficult to find any local areas with continuous highest or lowest values, instead the soil Hg concentration spreads evenly across the space. However, the spatial distribution of soil Hg concentration is more heterogeneous with higher standard deviation (0.067) in paddy field than in the vegetable filed (0.029).

The spatial distribution of soil Hg concentration aided with the main class agricultural land MLR model is shown in Fig. 2c. Furthermore, the spatial distributions of differences for the soil Hg concentrations from main class agricultural land MLR model and subclass agricultural land MLR models are shown in Fig. 2d and e. In fact, both in vegetable field (Fig. 2d) and paddy field (Fig. 2e), the difference is greater in the areas with higher soil Hg concentrations.

4. Discussion

Benefiting from the MLR modeling method, there is indeed a powerful way to identify driving factors causing soil Hg contamination in agricultural land so far (Du et al., 2015; Podolsky et al., 2015). However, it has to be acknowledged that while previous MLR based studies mostly focused on the factors identification at main class agricultural land (Han et al., 2006), the process difference of soil Hg contamination, accumulation and diffusion in the different subclass agricultural lands was ignored.

According to results of descriptive statistical analysis of this study, several soil Hg concentrations in paddy field are slightly higher than the National Environment Quality Grade II Standard for soil, $0.30\ mg \cdot kg^{-1}$. But in general, Hg contamination is not significant in agricultural land of the study area, which will not harm to human health at present (Zou et al., 2017b). High concentration coupled with great standard deviation values suggests the complex anthropogenic sources for soil Hg contamination in paddy field. Relatively, in orchard field, the soil Hg concentration displays quite homogeneous distribution across the space with lower standard deviations, which probably suggests the unapparent anthropogenic pollution sources (Li et al., 2008). Moreover, the results of correlation analysis in Table 2 actually also confirmed that the selected predictors for denoting the soil Hg concentration variation in orchard field of the study area are ineffectual.

Proving the credibility of MLR models developed in this study is essential to accurately identify the real driving factors and consequent to reveal the accumulation and transport process of soil Hg contamination of the study area. As demonstrated in Table 3, the MLR model developed in vegetable field shows that the soil Hg concentration is only correlated with *DC*. This suggests that the farther away from the nearest industrial sites is where the soil Hg concentration is lower. And this result is completely consistent with the diffusion law of Hg exhaust from industrial chimneys in previous studies. For paddy field, the MLR model shows that soil Hg concentration is significantly correlated with *SS* and *DRD*. This correlation relationship between soil Hg concentrations and the slope of soil sampling sites might be contributed to large runoffs in steep areas, which is helpful for the Hg contamination migration in paddy field (Jiang et al., 2016). In this way, the ability of runoffs for Hg contaminants transportation decreases gradually with the digressive slope, which can lead to more accumulation of Hg (Christoforidis and Stamatis, 2009; Wei and Yang, 2010). Relatively, the correlation relationship between soil Hg concentrations and the distances to the nearest roads might arise from the well-developed road system. Generally, the on-road transportation of the solid waste is one of main Hg contamination sources (Podolsky et al., 2015). In this study, roads across around paddy fields, and this can lead to the Hg contaminants from multiple roads being transferred into the center of paddy fields. As a result, Hg in the soils of these areas can be accumulated faster and its concentration is higher than in the areas nearby single roads or with less intersected roads.

The spatial distribution mapping results (Fig. 2a–b) well shows the continuous variety characteristics of soil Hg concentration across the vegetable and paddy fields. In vegetable field, the localized hotspots of soil Hg concentration in the north (Fig. 2a) could be attributed to the distribution of chimneys in this case study area (Fig. 1). The exhaust diffusion of Hg contamination from industrial sites has already been identified as a key driving factor in the MLR modeling step. Along with the prevailing wind from southeast to northwest of the study area, the Hg contaminants coming from the industrial chimneys in the midland area can heavily accumulate in the north of the vegetable field. For paddy field, the spatial variances of soil Hg concentration might result from the combined actions of the soil sampling slope and the roads in this area. Consequently, the ragged terrain characteristic and the developed roads network (Fig. 1) result in the even spatial distribution of soil Hg concentration in paddy field (Fig. 2b). Moreover, comparisons of the spatial distributions of soil Hg concentrations based on the MLR models for main class and subclass agricultural lands clearly demonstrated the large differences among them. Spatial patterns of the biased soil Hg concentrations (Fig. 2d–e) are highly consistent with those from the subclass MLR models based spatial distributions (Fig. 2a–b). This results actually further indicate the defect of main class MLR model in revealing the high soil Hg concentration due to misunderstood driving factors. Additionally, although the '$n-1$ cross validation' was implemented to validate the performance of MLR models in this study, the reliability of the MLR model for orchard field still needs in-depth discussion due to the lack of enough samples. Therefore, we can only suggest that significant anthropogenic effects on soil Hg pollution were not found in the orchard field in the case of this study currently. Certainly, another probable explanation for this might be that, in terms of agricultural situation in China, orchard fields are usually planted on a large area and mostly intensively clustered in the high-altitude hills. Therefore, the pollution risk will be easily considered and avoided during the site selection. In contrast, the distribution of vegetable and paddy fields usually located around people's homestead, which easily suffer from various anthropogenic contamination sources.

5. Conclusion

Compared to the utilization of one unified MLR model based investigation of soil Hg contamination for main class agricultural land traditionally, this study innovatively developed the GIS enhanced MLR models at subclass level agricultural lands. In this process, it reveals that subclass level factors driving the soil Hg contamination variation in agricultural lands are obvious different to those from main class level. With the subclass level specific MLR models, the spatial distributions of the soil Hg concentrations in vegetable and paddy fields can be more accurately mapped. In consequence, the soil Hg concentrations of paddy field are the highest, which is mainly derived from surface runoff and traffic contamination from the roads. Secondly, higher soil Hg concentrations can be found in the vegetable field, which might be caused by the exhaust emissions from industrial chimneys. However, although this study lays a solid foundation for fine mapping of potentially hazardous elements contamination in soil and consequent precise

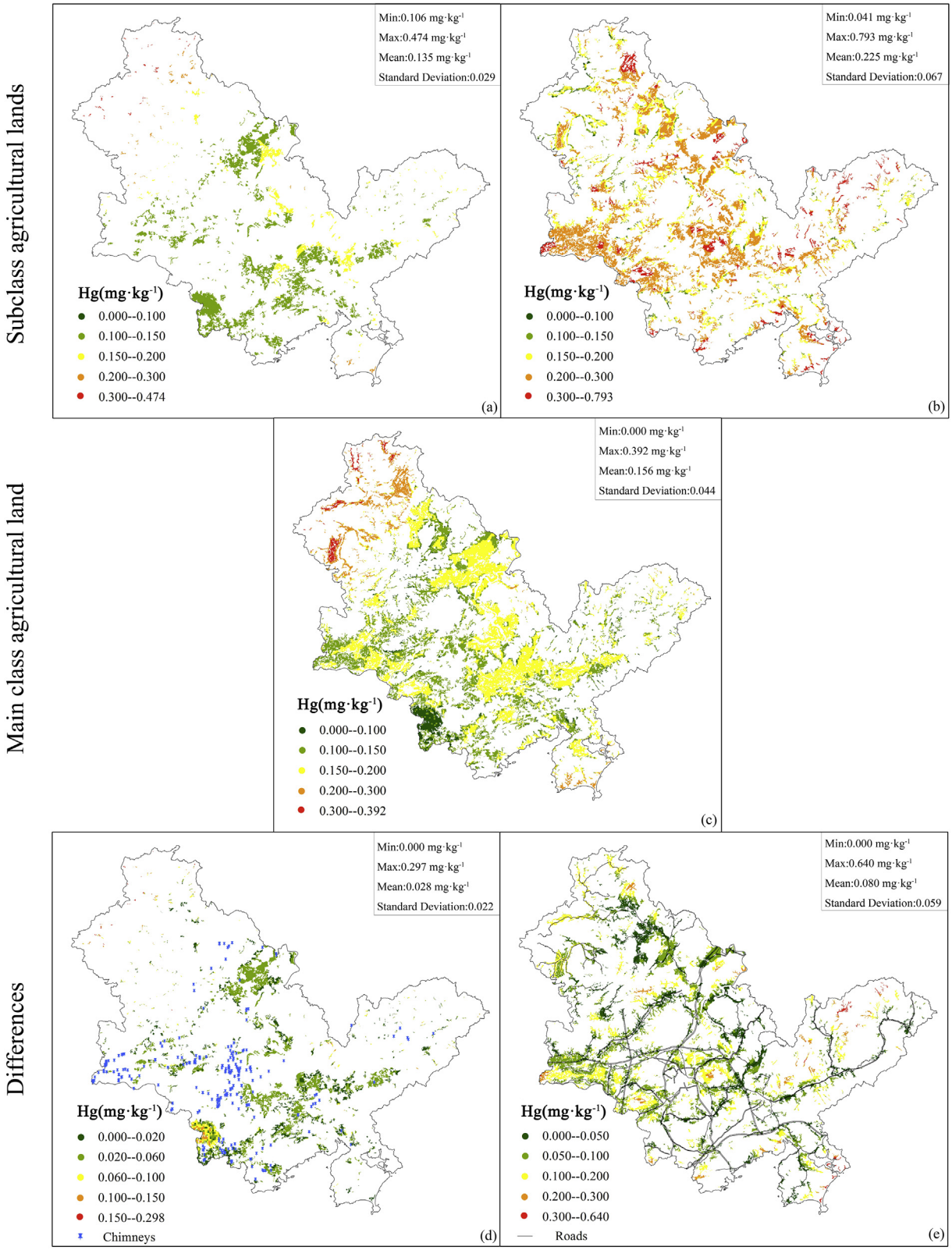

Fig. 2. Spatial distribution of soil Hg concentrations in vegetable field (a), paddy field (b) and main class agricultural land (c), and the differences in vegetable field (d) and paddy field (e).

prevention and controlling action, future work should be more concentrated on the inherent diffusion and transport mechanism enhanced semi-empirical modeling methods at higher spatial resolution.

Author contribution statement

B. Z. conceived the experiment(s), X. J. conducted the experiment (s), B. Z., X. J. and H. F. wrote the paper. B. Z., X. J., H. F., J. T., X. Z., analyzed the results, X. J. and Y. T. generated the maps, and all authors

reviewed the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgements

The author would like to thank the grants from the Postdoctoral Science Foundation of China (No. 2012M521558), the Special Program of the Postdoctoral Science Foundation of China (No. 2013T60780), and the Project of Innovation-driven Plan in Central South University (No. 2015CXS005).

References

Amini, M., Afyuni, M., Fathianpour, N., Khademi, H., Fluhler, H., 2005. Continuous soil pollution mapping using fuzzy logic and spatial interpolation. Geoderma 124 (3–4), 223–233.

Briki, M., Ji, H.B., Li, C., Ding, H.J., Gao, Y., 2015. Characterization, distribution, and risk assessment of heavy metals in agricultural soil and products around mining and smelting areas of Hezhang, China. Environ. Monit. Assess. 187 (12), 1–21.

Caeiro, S., Costa, M.H., Ramos, T.B., Fernandes, F., Silveira, N., Coimbra, A., Medeiros, G., Painho, M., 2005. Assessing heavy metal contamination in Sado Estuary sediment: an index analysis approach. Ecol. Indic. 5 (2), 151–169.

Cai, L., Xu, Z., Ren, M., Guo, Q., Hu, X., Hu, G., Wan, H., Peng, P., 2012. Source identification of eight hazardous heavy metals in agricultural soils of Huizhou, Guangdong Province, China. Ecotoxicol. Environ. Saf. 78 (2–8).

Chen, T.B., Zheng, Y.M., Lei, M., Huang, Z.C., Wu, H.T., Chen, H., Fan, K.K., Yu, Ke., Wu, X., Tian, Q.Z., 2005. Assessment of heavy metal pollution in surface soils of urban parks in Beijing, China. Chemosphere 60 (4), 542–551.

Cheng, J.L., Shi, Z., Zhu, Y.W., 2007. Assessment and mapping of environmental quality in agricultural soils of Zhejiang Province, China. J. Environ. Sci. 19 (1), 50–54.

Christoforidis, A., Stamatis, N., 2009. Heavy metal contamination in street dust and roadside soil along the major national road in Kavala's region, Greece. Geoderma 151 (3–4), 257–263.

De Simone, F., Cinnirella, S., Gencarelli, C.N., Yang, X., Hedgecock, I.M., Pirrone, N., 2015. Model study of global mercury deposition from biomass burning. Environ. Sci. Technol. 49 (11), 6712–6721.

D'Hose, T., Cougnon, M., De Vliegher, A., Vandecasteele, B., Viaene, N., Cornelis, W., Van Bockstaele, E., Reheul, D., 2014. The positive relationship between soil quality and crop production: a case study on the effect of farm compost application. Appl. Soil Ecol. 75, 189–198.

Du, P., Xie, Y., Wang, S., Zhao, H., Zhang, Z., Wu, B., Li, F., 2015. Potential sources of and ecological risks from heavy metals in agricultural soils, Daye City, China. Environ. Sci. Pollut. Res. 22 (5), 3498–3507.

Fang, X., Zou, B., Liu, X., Sternberg, T., Zhai, L., 2016. Satellite-based ground $PM_{2.5}$ estimation using timely structure adaptive modeling. Remote Sens. Environ. 186, 152–163.

Francouría, A., Lópezmateo, C., Roca, E., Fernándezmarcos, M.L., 2009. Source identification of heavy metals in pastureland by multivariate analysis in NW Spain. J. Hazard. Mater. 165 (1–3), 1008–1015.

Frentiu, T., Pintican, B.P., Butaciu, S., Mihaltan, A.I., Ponta, M., Frentiu, M., 2013. Determination, speciation and distribution of mercury in soil in the surroundings of a former chlor-alkali plant: assessment of sequential extraction procedure and analytical technique. Chem. Cent. J. 7 (1), 1–14.

Gagiu, C., Pica, E.M., Querol, X., Botezan, C.S., 2015. Analysis of predictors related to soil contamination in recreational areas of Romania. Environ. Sci. Pollut. Res. 22 (23), 18885–18893.

Guney, M., Zagury, G.J., Dogan, N., Onay, T.T., 2010. Exposure assessment and risk characterization from trace elements following soil ingestion by children exposed to playgrounds, parks and picnic areas. J. Hazard. Mater. 182 (1–3), 656–664.

Han, F.X., Su, Y., Monts, D.L., Waggoner, C.A., Plodinec, M.J., 2006. Binding, distribution, and plant uptake of mercury in a soil from Oak Ridge, Tennessee, USA. Sci. Total Environ. 368 (2–3), 753–768.

Hu, D.W., Bian, X.M., Li, S.M., Xu, Q., 2007. Spatial distribution of farmland heavy metals based on GRNN-ANN modeling. Chin. J. Soil Sci. 38 (2), 334–340.

Huang, S.W., Jin, J.Y., 2008. Status of heavy metals in agricultural soils as affected by different patterns of land use. Environ. Monit. Assess. 139 (1–3), 317–327.

Jiang, X.L., Zou, B., Tang, J.W., Tu, Y.L., 2016. Spatial distribution of As in vegetable field and paddy in southeast of Guangdong province. Trans. Chin. Soc. Agric. Eng. 32 (23), 263–268.

Li, Y., Gou, X., Wang, G., Zhang, Q., Su, Q., Xiao, G., 2008. Heavy metal contamination and source in and agricultural soil in central Gansu Province, China. J. Environ. Sci. 20 (5), 607–612.

Liu, H.B., Guo, P.T., Wu, W., Wang, Z.Y., 2011. Assessment of soil arsenic, chromium, mercury, and lead at an agricultural landscape scale. Soil Sediment Contam. Int. J. 20 (8), 995–1007.

Lu, W.Z., Wang, W.J., Wang, X.K., Yan, S.H., Lam, J.C., 2004. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. Environ. Res. 96 (1), 79–87.

Lu, X., Wang, L., Li, L.Y., Lei, K., Huang, L., Kang, D., 2010. Multivariate statistical analysis of heavy metals in street dust of Baoji, NW China. J. Hazard. Mater. 173 (1–3), 744–749.

Lu, A., Wang, J., Qin, X., Wang, K., Han, P., Zhang, S., 2012. Multivariate and geostatistical analyses of the spatial distribution and origin of heavy metals in the agricultural soils in Shunyi, Beijing, China. Sci. Total Environ. 425, 66–74.

Micó, C., Recatala, L., Peris, A., Sanchez, J., 2006. Assessing heavy metal sources in agricultural soils of a European Mediterranean area by multivariate analysis. Chemosphere 65 (5), 863–872.

Mostert, M.M., Ayoko, G.A., Kokot, S., 2012. Multi-criteria ranking and source identification of metals in public playgrounds in Queensland, Australia. Geoderma 173-174, 173–183.

Ottesen, R.T., Birke, M., Finne, T.E., Gosar, M., Locutura, J., Reimann, C., Tarvainen, T., 2013. Mercury in European agricultural and grazing land soils. Appl. Geochem. 33, 1–12.

Podolsky, F., Ettler, V., Sebek, O., Jezek, J., Mihaljevic, M., Kribek, B., Sracek, O., Vanek, A., Penizek, V., Majer, V., Mapani, B., Kamona, F., Nyambe, I., 2015. Mercury in soil profiles from metal mining and smelting areas in Namibia and Zambia: distribution and potential sources. J. Soils Sediments 15 (3), 648–658.

Qiu, Y., Guan, D., Song, W., Huang, K., 2009. Capture of heavy metals and sulfur by foliar dust in urban Huizhou, Guangdong province, China. Chemosphere 75 (4), 447–452.

Salminen, R., Tarvainen, T., Demetriades, A., Duris, M., Fordyce, F.M., Gregorauskiene, V., Kahelin, H., Kivisilla, J., Klaver, G., Klein, H., Larson, J., Lis, J., Locutura, J., Marsina, K., Mjartanova, H., Mouvet, C., O'Connor, P., Odor, L., Ottonello, G., Paukola, T., Plant, J.A., Reimann, C., Schermann, O., Siewers, U., Steenfelt, A., Van der Sluys, J., De Vivo, B., Williams, L., 1998. FOREGS Geochemical Mapping Field Manual. Geological Survey of Finland, Espoo, Finland.

Schwertman, N.C., De Silva, R., 2007. Identifying outliers with sequential fences. Comput. Stat. Data Anal. 51 (8), 3800–3810.

Wei, B., Yang, L., 2010. A review of heavy metal contaminations in urban soils, urban road dusts and agricultural soils from China. Microchem. J. 94 (2), 99–107.

Wuana, R.A., Okieimen, F.E., 2011. Heavy metals in contaminated soils: a review of sources, chemistry, risks and best available strategies for remediation. Int. Sch. Res. Netw. ISRN Ecol. 2011, 1–20.

Xie, Y.F., Chen, T.B., Lei, M., Yang, J., Guo, Q.J., Song, B., Zhou, X.Y., 2011. Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis. Chemosphere 82 (3), 468–476.

Xu, X., Zhao, Y., Zhao, X., Wang, Y., Deng, W., 2014. Sources of heavy metal pollution in agricultural soils of a rapidly industrializing area in the Yangtze Delta of China. Ecotoxicol. Environ. Saf. 108, 161–167.

Yang, X., Wang, L., 2008. Spatial analysis and hazard assessment of mercury in soil around the coal-fired power plant: a case study from the city of Baoji, China. Environ. Geol. 53 (7), 1381–1388.

Yang, Q., Chen, H., Li, B., 2015. Source identification and health risk assessment of metals in indoor dust in the vicinity of phosphorus mining, Guizhou province, China. Arch. Environ. Contam. Toxicol. 68 (1), 20–30.

Yuan, G.L., Sun, T.H., Han, P., Li, J., Lang, X.X., 2014. Source identification and ecological risk assessment of heavy metals in topsoil using environmental geochemical mapping: typical urban renewal area in Beijing, China. J. Geochem. Explor. 136, 40–47.

Zeng, F.R., Ali, S., Zhang, H.T., Ouyang, Y., Qiu, B.Y., Wu, F.B., Zhang, G.P., 2011. The influence of pH and organic matter content in paddy soil on heavy metal availability and their uptake by rice plants. Environ. Pollut. 159 (1), 84–91.

Zhai, L., Zou, B., Fang, X., Luo, Y., Wan, N., Li, S., 2017. Land use regression modeling of $PM_{2.5}$ concentrations at optimized spatial scales. Atmosphere 8 (1), 1–15.

Zhai, L., Li, S., Zou, B., Sang, H., Fang, X., Xu, S., 2018. An improved geographically weighted regression model for $PM_{2.5}$ concentration estimation in large areas. Atmos. Environ. 181, 145–154.

Zhang, H., Luo, Y., Li, Y., Zhou, Q., Liu, X., 2014. Screening of criteria for heavy metals for revision of the national standard for soil environmental quality of China. Acta Pedol. Sin. 51 (3), 429–438.

Zhou, P., Zhao, Y., Zhao, Z., Chai, T., 2015. Source mapping and determining of soil contamination by heavy metals using statistical analysis, artificial neural network, and adaptive genetic algorithm. J. Environ. Chem. Eng. 3 (4), 2569–2579.

Zou, B., Wang, M., Wan, N., Wilson, J.G., Fang, X., Tang, Y., 2015. Spatial modeling of $PM_{2.5}$ concentrations with a multifactoral radial basis function neural network. Environ. Sci. Pollut. Res. Int. 22 (14), 10395–10404.

Zou, B., Pu, Q., Bilal, M., Weng, Q., Zhai, L., Nichol, J.E., 2016. High-Resolution satellite mapping of fine particulates based on geographically weighted regression. IEEE Geosci. Remote Sens. Lett. 13 (4), 495–499.

Zou, B., Chen, J., Zhai, L., Fang, X., Zheng, Z., 2017a. Satellite based mapping of ground $PM_{2.5}$ concentration using generalized additive modeling. Remote Sens. 9 (1), 1–16.

Zou, B., Jiang, X.L., Duan, X.L., Zhao, X.G., Zhang, J., Tang, J.W., Sun, G.Q., 2017b. An integrated H-G scheme identifying areas for soil remediation and primary heavy metal contributors: a risk perspective. Sci. Rep. 7 (1), 1–11.