



Comprehensive source identification and apportionment analysis of five heavy metals in soils in Wenzhou City, China

Shuai Shao · Bifeng Hu · Yunhan Tao · Qihao You · Mingxiang Huang · Lianqing Zhou · Qiuxiao Chen · Zhou Shi

Received: 17 March 2020 / Accepted: 8 March 2021 / Published online: 2 April 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract The source identification and apportionment of heavy metals (HMs) is a vital issue for restoring contaminated soil. In this study, qualitative approaches [a finite mixture distribution model (FMDM) and raster-based principal components analysis (RB-PCA)] and a quantitative approach [positive matrix factorization (PMF)] were composed to identify and apportion the sources of five HMs (Cd, Hg, As, Pb, Cr) in Wenzhou City, China, using several crucial auxiliary variables. An initial ecological risk assessment suggested that the ecological risk level in the study area was generally considered low, with the greatest contamination contributions coming from Cd and Hg. The result of the FMDM showed that Cd and Pb fit a single log-normal distribution, Hg fit a double

log-normal mixed distribution, and As and Cr presented a triple log-normal distribution. Each element was identified and separated from its natural or anthropogenic sources. A map of RB-PCA combined with an analysis of corresponding auxiliary variables suggested that the three main contribution sources in the entire study area were parental materials, industrial and agricultural mixed pollution, and mining exploration activities. Each element was discussed, using the PMF model, with regard to its quantitative contributions. Parental materials contributed to all elements (Cd, Hg, As, Pb, Cr) at 89.22%, 7.31%, 35.84%, 84.81% and 27.42%, respectively. Industrial emissions and agricultural inputs mixed pollution contributed 2.94%, 80.77%, 15.93%, 4.79%, and

S. Shao · L. Zhou · Z. Shi
Key Laboratory of Environment Remediation and Ecological Health, Ministry of Education, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China

B. Hu
Department of Land Resource Management, School of Tourism and Urban Management, Jiangxi University of Finance and Economics, 330013 Nanchang, China

B. Hu
URSOLS, INRAE, 45075 Orleans, France

Y. Tao
School of Earth Sciences, Lanzhou University, Lanzhou 730030, China

Q. You
Eco-Environmental Science & Research Institute of Zhejiang Province, Hangzhou 310012, China

M. Huang
Information Center of Ministry of Ecology and Environment, Beijing 100035, China

Q. Chen (✉)
Department of Regional and Urban Planning, College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China
e-mail: chen_qiuxiao@zju.edu.cn

25.63%, respectively. Mining activities contributed 7.84%, 11.92%, 48.23%, 10.40% and 46.95%, respectively, to the five HMs. Such result could be used efficiently to generate scientific decisions and strategies in terms of decision-making on regulating HM pollution in soils.

Keywords Source identification · Source apportionment · Heavy metals · RB-PCA · FMDM · PMF

Introduction

Heavy metals (HMs) are of great significance in soil environments, owing to their high ecotoxicity and degradation resistance once becoming excessively persistent in soils (Hu et al., 2020; Marchant et al., 2017; Saby et al., 2006; Xia et al., 2019; Xie et al., 2018). Originally, HMs come from natural backgrounds, i.e. so-called endogenous sources, which refers to, e.g. a complex parental rock weathering process (Hu et al., 2017; Ogunlaja et al., 2019; Schneider et al., 2016). However, as industrialization and urbanization has progressed, anthropogenic activities, known for ‘exogenous sources’ such as various industrial enterprises emitting solid wastes and sewage, mines exploitation, fertiliser inputs, and traffic emissions, have become major contamination sources (Wei & Yang, 2010). The combined endogenous and exogenous contaminations have resulted in complex spatial variability, making soil remediation difficult to carry out (Uría et al., 2009; Marchant et al., 2011). Without having a knowledge of the contamination sources, removing pollutants in the soil environment can only be temporary and consumes significant time and resources (Proshad et al., 2019; Song et al., 2018; Sungur et al., 2019). Thus, identifying and apportioning the contamination sources in a soil environment can be necessary and highly valuable (Chen et al., 2018; Dong et al., 2019).

Source identification and apportionment generally refers to utilizing multivariate models to identify and apportion the contamination sources of HMs in soils. It can not only efficiently contribute to remediation of soil contamination, but also assists in relevant management for decision-makers (Luo et al., 2015; Manoli et al., 2002). The approaches can be divided into

qualitative and quantitative approaches, according to recent research (Zhi et al., 2016). The qualitative approaches include multivariate statistical analysis techniques consisting of methods like principal component analysis (PCA) (Dai et al., 2018; Gülten 2019), cluster analysis (CA) (Facchinelli, et al., 2001), factor analysis (FA) (Dong et al., 2019), a support vector machine (SVM) (Chen et al., 2013), and tracing techniques consisting of stable isotope ratio technique (Glaser et al., 2005). Based on the various algorithms for data analysis, multivariate statistical methods are widely used, owing to their advantages in low costs and easy employment (Davies, 1997; Nikos et al., 2012). However, these models can also suffer flaws, such as the negative effects of abnormal values, and mismatches between the apportioned source and the real emitting source (Huang et al., 2015). As for the stable isotope ratio techniques, isotope tracing techniques can accomplish a contamination circle for determining integrated routes, but their application can be limited by e.g. expensive costs, low concentrations of HMs, or minor differences between ratio of isotopes (Zhu et al., 2017). As a result, more comprehensive profile analysis methods are required.

A finite mixture distribution model (FMDM) is a stochastic approach which separates an entire integrated data distribution into several subgroups of-distributions, while also estimating the proportions of natural and anthropogenic sources of HMs in soils according to comparisons between the relevant parameters of the subgroups of-distributions and soil background values (Lin et al., 2009; Shao et al., 2018). Raster-based principal components analysis (RB-PCA) is an improved version of the PCA, and combines geostatistical interpolation with a statistical approach, instead of interpolating each component through scores calculated by a dimensionality reduction (Wang et al., 2018). The RB-PCA replaces the original points with each HM’s interpolation raster map as a data source, and calculates the score of each raster with a higher precision map, helping to identify contribution sources more efficiently.

As a main type of quantitative apportion, receptor models—including positive matrix factorization (PMF) (Mehr et al., 2017), absolute principal component score/multiple linear regression (Huang et al., 2018), and ‘UNMIX’ (Liao et al., 2020)—do not need to acquire prior knowledge regarding source profiles, but quantitatively apportion contamination sources by

outputting valid contribution values of every source regarding every HM, which exactly compensates for the drawbacks of the qualitative methods. A PMF model considers the bias and uncertainties during the entire calculation process (Liang et al., 2017). The factor contributions' and factor profiles' non-negative limitations also make it more plausible as compared to other multivariate statistical models. Therefore, it can be the prior choice for apportioning potential contamination sources (Hu et al., 2019; Lv et al., 2018).

Overall, this study suggested an integrated proposal combining qualitative and quantitative means to identify and apportion potential contamination sources of HMs (Cd, Hg, As, Pb, Cr) in soils in the study area. The main objectives of this study were to: (1) generally assess the ecological risk level of the study area through a potential ecological risk index (RI); (2) qualitatively identify the natural and anthropogenic sources, and interpolate/map the spatial location of each source using the FMDM and RB-PCA, as assisted with natural and social auxiliary datasets; and (3) quantitatively apportion each source with factor contribution loadings of every HM, using PMF.

Materials and methods

Sampling and chemical analysis

Located in the southeast of the Zhejiang province (27°03'–28°36' N, 119°37'–121°18' E), Wenzhou City covers a total land area of 11,784 square kilometres (Fig. 1). With a topography that decreases from west to east, Wenzhou City consists of low hills and plains near the sea. Wenzhou City is a vital port, not only in Zhejiang province but also across all over China, which has a long history of industrial development and business trade.

In this study, a total of 1474 topsoil (0–20 cm) samples were collected in the study area. The collecting process can be simply described as: (1) dividing the study area into different strata according to land use types; and (2) applying grid to set the sample sites in each of strata. For land use types except residential land, one sample point is collected within a square of 1 km × 1 km, but for residential land areas where enterprises were clustering, the number of samples within a square are augmented to ensure that the

sample statistics are efficient. As for the sample details, they obeyed a basic principle of setting a soil sample at an intersection point, and combining it with five subgroups of samples collected from five locations within 5 m. The specific coordinates of the sampling locations were recorded with a GPS.

Auxiliary variables

Several types of auxiliary datasets were provided and categorised, respectively, as natural environmental, and economical social datasets. This study obtained information regarding the soil parental material in the study area from a 1:20,000 soil map of Zhejiang Province published by the National Soil Survey of China. For the anthropogenic factors, industrial enterprises, ore mines generally have most serious impacts, as they cause multi-environmental pollution through contaminated airflows or sewage irrigations during processing (Men et al., 2018; Tian et al., 2018; Zhang et al., 2018). In this study, 1846 enterprises within the study area were collected from the General Survey of Key Pollution Sources in Zhejiang Province in 2016, and were categorised into four main categories according to National Industry Classification standard documents: GB/4754–2011. In addition, 285 mines were surveyed and divided into seven different categories according to their quantity and impacts on soil HMs.

Fertiliser inputs and traffic emissions also have vital impacts on the HMs in soils. Different types of fertilisers contain certain contents of HMs and can influence the aggregation and transfer process, as well as the forms in which the HMs exist in soil (Yan, 2008). In detail, nitrogen fertilisers such as urea contain Cd, Pb, Cu and Zn; phosphate fertilisers such as superphosphate involve Cd, Hg, As, Pb, and Cr, while potash fertilizers such as potassium chloride include Cd, Pb, Cu and Zn; Moreover, organic fertilisers can also have impacts on HMs for it interference the viability of HMs in the soil therefore result in contamination. In this study, we collected the fertiliser data from the Survey of Heavy metals in Agricultural Products Areas in Zhejiang Province along with sample points collection process and calculated the contents of four fertiliser inputs (nitrogen fertiliser, phosphate fertiliser, potash fertiliser, and organic fertiliser) on the farmland of each town or village in Wenzhou City. Moreover, particulate matter

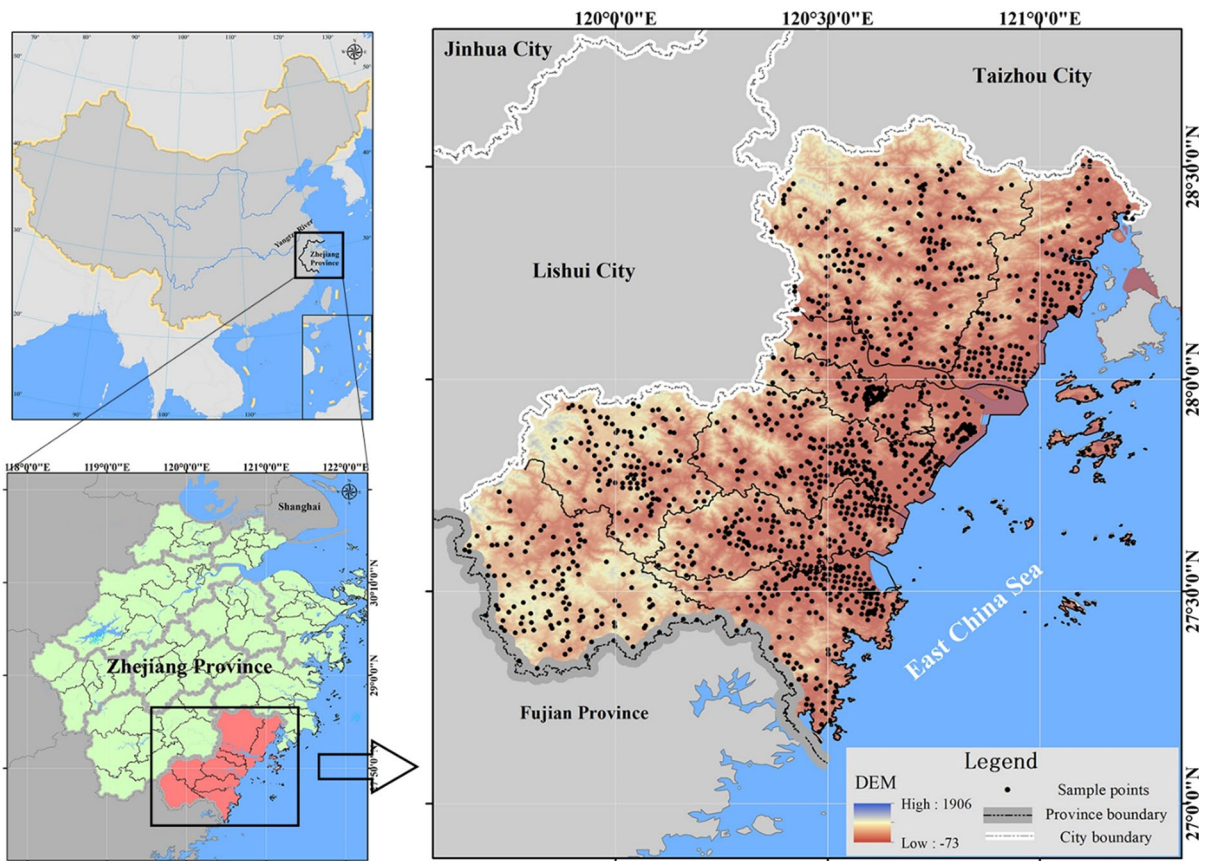


Fig. 1 Maps of study area and sampling sites

such as dusts containing HMs generated by vehicle fuels and brake pad wear can also enter the soil on both sides of the road during vehicle driving (Aminiyan et al., 2018; Li et al., 2017; Nabulo et al., 2006; Zhu et al., 2015). In this study, traffic system vector datasets including different levels of roads were obtained from the Institute of Geography and Natural Resources Research of the Chinese Academy of Sciences for further analysis.

Methodology

Buffer areas and fishnets construction and analysis

As a prerequisite for the quantitative analysis of the anthropogenic auxiliary datasets, fishnets and buffer areas were constructed to help analyse the impacts of potential contamination sources. For industrial

enterprises, firstly we created fishnets of 1.5 km covering the study area and counted the samples falling into the units and then implemented the HMs contents of sample points spatially joined with each unit; therefore, each unit can have the attributes of every point falling within, and then the mean values of HMs of each unit were calculated. Similar to ore mines, we constructed buffer areas with a radius of 2 km (Kosharna & Korzhnev, 2018) and made the HMs of the sample points spatial join with the buffer areas of different kinds of ore mines as well, and then the mean values of HMs were calculated in buffer areas of different kinds of mines. As for traffic emissions, we initially set different threshold distances to different levels of road as the radius of the line buffer polygons (Pirsahab et al., 2016), then set the HMs of the sample points spatial join with the buffer polygons of different levels of road, and finally

calculated the mean values of HMs in different levels of roads.

Potential ecological risk index

The potential ecological RI was deemed as an accurate approach for assessing the ecological risk caused by HMs (Kashif et al., 2020). It considers the concentrations, categories, levels of toxicity, and sensitivities of HMs together, and classifies them into different potential ecological levels using a quantitative methodology. Therefore, it has been used in most ecological studies. The calculation process can be shown as in the equation below (Hakason, 1980):

$$RI = \sum_{i=1}^m E_r^i = \sum_{i=1}^m T_r^i \times \frac{C^i}{C_n^i} \tag{1}$$

E_r^i is the potential ecological RI of the heavy HM i , and T_r^i is the toxicity index of each HM. For Cd, Hg, As, Pb, and Cr, their toxicity indexes are 30, 40, 10, 5, and 2, respectively (Hakason, 1980). C^i denotes the concentration of HM i , C_n^i denotes the background value of HM i , and m is the genre of HM. The classification of the potential ecological risk (E_r^i) and hazard quotients of the HMs can be seen in Table 1.

Finite mixture distribution model (FMDM)

Given that for a random variable x , a mixture distribution consists of m components, and the distribution of the i th individual component is determined by a specific probability density function (i.e. pdf) $f_i(x)$, we can express the general pdf $f(x)$ for the mixture distribution as (Lin et al., 2009):

$$f(x) = \sum_{i=1}^m \pi_i f_i(x) = \pi_1 f_1(x) + \dots + \pi_m f_m(x) \tag{2}$$

$$\sum_{i=1}^m \pi_i = 1 (0 \leq \pi_i \leq 1) \tag{3}$$

Here, π_i presents the mixed weights of every subgroups of-distribution.

It is known that most natural processes, and especially HMs in soils, usually express as a normal distribution or a log-normal distribution (Zhi et al., 2016). Therefore, we used the log-normal distribution to describe the contents of HMs in soils coming from different sources, which can be illustrated as follows:

$$f_i(x|\mu_m, \sigma_m) = \frac{1}{\sqrt{2\pi}\sigma_m x} e^{-\frac{(\ln x - \mu_m)^2}{2\sigma_m^2}}, x > 0 \tag{4}$$

In the above, μ_m and σ_m denote the mean and standard deviation of every subgroups of-distribution, respectively. Such parameters were calculated using the expectation maximization algorithm. As a result, we applied a Chi-square goodness-of-fit test to test the null hypothesis H_0 , to ensure that the assumed model could be fit with the observed distribution. In particular, the process of calculating the cut-off value between the i th and $(i + 1)$ components can be presented as follows:

$$\pi_i \int_{a_0}^{+\infty} f_i(x) dx = \pi_{i+1} \int_{-\infty}^{a_0} f_{i+1}(x) dx \tag{5}$$

Raster-based principal components analysis (RB-PCA)

Traditional PCA obeys a principle of using an orthogonal transformation to convert a set of observations of variables that might be correlated with a set of values of linearly uncorrelated variables (Shao et al. 2018). Here, we improved the accuracy of cluster manipulation by replacing the original point variables (which exactly referred to the content of each HM)

Table 1 Classification of potential ecological risk (E_r^i) and hazard quotient (risk index, RI)

E_r^i	Classification	RI	Classification
< 40	Low	< 150	Low
≥ 40 and < 80	Moderate	≥ 150 and < 300	Moderate
≥ 80 and < 160	High	≥ 300 and < 600	High
≥ 160 and < 320	Serious	≥ 600	Serious
≥ 320	Severe	–	–

with inverse distance weighted (IDW) prediction maps (z_1, z_2, \dots, z_k). As the results of IDW prediction maps tend to be the most authentic values within interpolation methods, replacing individual points scores with successive raster scores can prove to have high validation accuracy.

$$f = a_1z_1 + a_2z_2 + \dots + a_kz_k \quad (6)$$

Here, a_1, a_2, \dots, a_k denotes the respective coefficient of every cluster score.

Positive matrix factorization (PMF)

As a multivariate FA tool which is originally intended to address chemical mass balance problems, the PMF model can be summarised as (Pekey & Dogan, 2013):

$$X = GF + E \text{ or } x_{ij} = \sum_{k=1}^p g_{ik}f_{kj} + e_{ij} \quad (7)$$

In the above, $G(m \times p)$ denotes the factor contribution, and x_{ij} is one of the factors; $G(m \times p)$ represents the matrix of the factor profile, whereas g_{ik} represents the contribution from factor k to receptor i , f_{kj} represents the concentration of group j in factor k ; $E(m \times n)$ is the residual error matrix, and e_{ij} represents every residual error item.

$$Q = \sum_{i=1}^n \sum_{j=1}^m \left[\frac{x_{ij} - \sum_{k=1}^p g_{ik}f_{kj}}{u_{ij}} \right]^2 \quad (8)$$

PMF model uses parameter Q to minimize outputs derived from factor contributions and profiles. There can be two versions of Q , categorised as: (1) $Q(\text{true})$ is the goodness-of-fit parameter calculated by including all points; and (2) $Q(\text{robust})$ is the goodness-of-fit parameter calculated excluding points not fit by the model, but defined as samples in which the uncertainty-scaled residual is above four.

Under the constraint of a non-negative condition, the objective function Q could be minimised to produce more precise factor contributions and profiles, as enhanced by the 'Multilinear engine-2'. In the calculation process, the uncertainty could thus be obtained based on an element-specific method detection limit, and the error percentage could be measured using standard reference material (Norris and Duvall 2014).

Combined mechanism of three models

The three models interpret source identifications and apportionments of the HMs in soil with different mechanics; they are also complementary to each other. The FMDM analysis process starts from the statistical distribution analysis of each element (Hu & Cheng, 2013). It sufficiently identifies contamination sources for each HM, especially for dissociating between natural and anthropogenic sources, whereas RB-PCA conducts a qualitative approach to identify potential contamination sources within the study area. By combining the high scores in the spatial distribution of each principal component and conducting auxiliary multiple variables analysis, the sources could be determined (Dong et al., 2018; Qu et al., 2013). In this study, a combination of FMDM and RB-PCA approaches will not only separate natural or anthropogenic sources statistically within the study area, but also can provide further qualitative insights into the spatial patterns of different contamination sources. Apart from first two approaches, PMF adds quantitative results to the apportionment scheme, by outputting the defined sources' contributions to every element (Liang et al., 2017). As a result, an integrated contamination source identification and apportionment is achieved (Fig. 2).

Data analysis

In this study, all of the statistical analysis and correlation tests were conducted using the Statistical Package for Social Sciences (SPSS) Statistics 25 (IBM Corp.; 2016; IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp). The construction of fishnets and buffer areas, expression of spatial distribution, spatial join, and interpolation of auxiliary variables were all accomplished by ArcGIS 10.3 (ESRI, ArcGIS 10.3, Redlands, CA, USA). In the RB-PCA, the basic IDW interpolation of every element was fulfilled using a gstat package, whereas the PCA was conducted using a psych package in the software R Studio. The FMDM was completed in R Studio with the Mclust package, and the PMF model was utilised based on the US-EPA PMF 5.0 program (Norris and Duvall 2014).

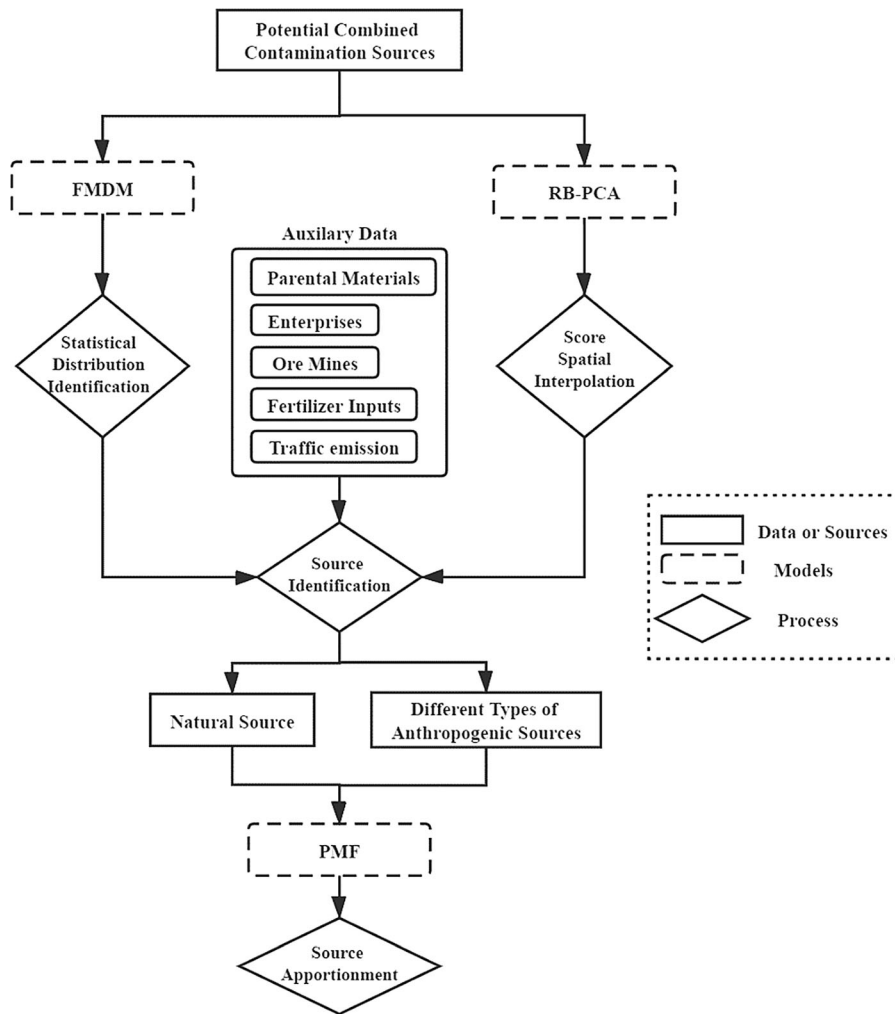


Fig. 2 The flowchart of combination of models in the study

Results and discussion

Summary statistical analysis of heavy metals (HMs) in soils

The basic summary of the statistical analyses of Cd, Hg, As, Pb, and Cr from the 1474 sample points is given in Table 2. The average concentrations of the five HMs were 0.28, 0.16, 8.73, 37.13, and 59.09 mg/kg, respectively. Except for Pb, the means of the other HMs were all high above the background values (Wang et al., 2007), indicating that apart from natural factors such as parental materials, other anthropogenic activities should have influences on the spatial heterogeneity of HMs in soils. (Dong et al., 2018).

The coefficient of variation (CV) revealed the degree of the dispersion of the statistics (Brock, 1953). The CVs of Cd, Hg, and As all exceeded a value of 100%, which reflected that there would be extreme high values in these HMs.

Potential ecological risk assessment of HMs in soils

The potential ecological coefficient E_r^i and hazard quotient RI were calculated and are summarised in Table 3. According to the mean value of E_r^i , the potential ecological risk assessment can be ranked as Hg > Cd > As > Pb > Cr. In particular, the mean values of E_r^i of Hg and Cd were 57.44 and 50.33,

Table 2 Summary statistics for HMs in topsoil (mg/kg)

Element	Cd	Hg	As	Pb	Cr
Valid number	1474	1474	1474	1474	1474
Mean	0.28	0.16	8.73	37.13	59.09
Median	0.20	0.10	6.80	33.50	57.90
STD ^a	0.62	0.17	11.71	23.65	29.71
CV ^b	219.59%	105.60%	134.08%	63.69%	50.27%
Range	0–17.00	0–2.80	0.80–216.00	11.90–518.00	2.80–416.00
Background value	0.178	0.127	5.86	41.6	56.9
Background interval	0.095–0.260	0.029–0.225	3.67–8.04	31.0–52.3	3.6–110.2

^aStandard deviation,^b Coefficient of variation**Table 3** Potential ecological coefficient (E_r^i) and hazard quotient (RI) of HMs

	E_r^i of each HM					RI
	Cd	Hg	As	Pb	Cr	
Mean	50.33	57.44	15.31	4.57	2.13	129.78
Min	16.85	31.5	1.36	1.72	6.28	57.71
Max	2865.17	881.89	368.6	62.25	14.62	4192.53

respectively, accounting for 44.26% and 38.78% of the RI, respectively. Thus, Hg and Cd can be recognised as greater ecological contamination contributors, with a moderate ecological risk in soils. Apart from Hg and Cd, the other HMs were below the threshold (40) and can therefore be treated as being of low ecological risk in soils. The value of RI ranged from 57.71 to 4192.53, with a mean value of 129.78. According to the rule of hazard quotient classification, the mean value was far below the first threshold value of 150, indicating that the general ecological risk of the entire study area was considered to be safe. Nevertheless, special attention should be paid to the maximum value of the RI, which reached a high value of 4192.53, far above the highest threshold of 600. In addition, the main contributors Cd and Hg had respective contributions of 2865.17 and 881.89 to this value, suggesting there should be abnormally extreme high values of these two HMs. In that regard, anthropogenic activities such as industrial emissions, mine exploitation, urban life wastes, and agricultural inputs could all be potential sources (Han et al. 2016) (Table 4).

Auxiliary variables analysis

Natural variables analysis

Land use cover and population density The map used for the land use cover and population density distribution is shown in Fig. 3. As can be seen, woodland and arable land predominantly covered the entire study area, taken up the total areas of Wenzhou City up to 62.15% and 23.33%, respectively. Grassland scattered over the northwest areas, commonly associated with woodland areas and taken up 2.38% areas. The residential areas were located mostly by the coast and near rivers, taken up 8.60% areas. The P values of most HMs of different land use cover showed significant differences. The quantitative analysis showed the Cd in residential land were on average higher than in other regions, which reflected that anthropogenic influences, particularly from human activities and life wastes, had certain impacts on these values (Jiang et al., 2017).

Soil parental materials The parental materials' and subgroups of parental materials' spatial distributions are shown in Fig. 4. Slope wash was dominant within the entire study area, whereas other parental materials were mainly distributed in the coastal areas. According to Table 5, the P values of Cd and Pb of different parental materials showed significant differences which indicates that the content of both HMs varies evidently from different types of parental materials. Particularly, Cd has a relatively high content in fluvial faces deposit, lacustrine faces, and slope washes, whereas Pb originates from mud faces and channel faces with a higher content. Other parental materials also had different contributions to the various HMs. Slope wash, especially for sand

Table 4 Mean values for HMs' contents in diverse land use cover (mg/kg)

Land use cover	Area cover ratio	Number	Cd	Hg	As	Pb	Cr
Arable land	23.33%	714	0.283	0.165	8.730	37.132	59.135
Wood land	62.15%	197	0.288	0.166	8.799	37.150	59.090
Grass land	2.38%	20	0.286	0.161	8.814	37.358	58.233
Water areas	3.48%	0	–	–	–	–	–
Residential land	8.60%	226	0.319	0.165	8.784	37.070	58.489
Unused land	0.04%	0	–	–	–	–	–
Ocean areas	0.02%	0	–	–	–	–	–
P value ^c	–	–	0.006	0.002	0.04	0.15	0.08

^c Kruskal–Wallis Test—P value

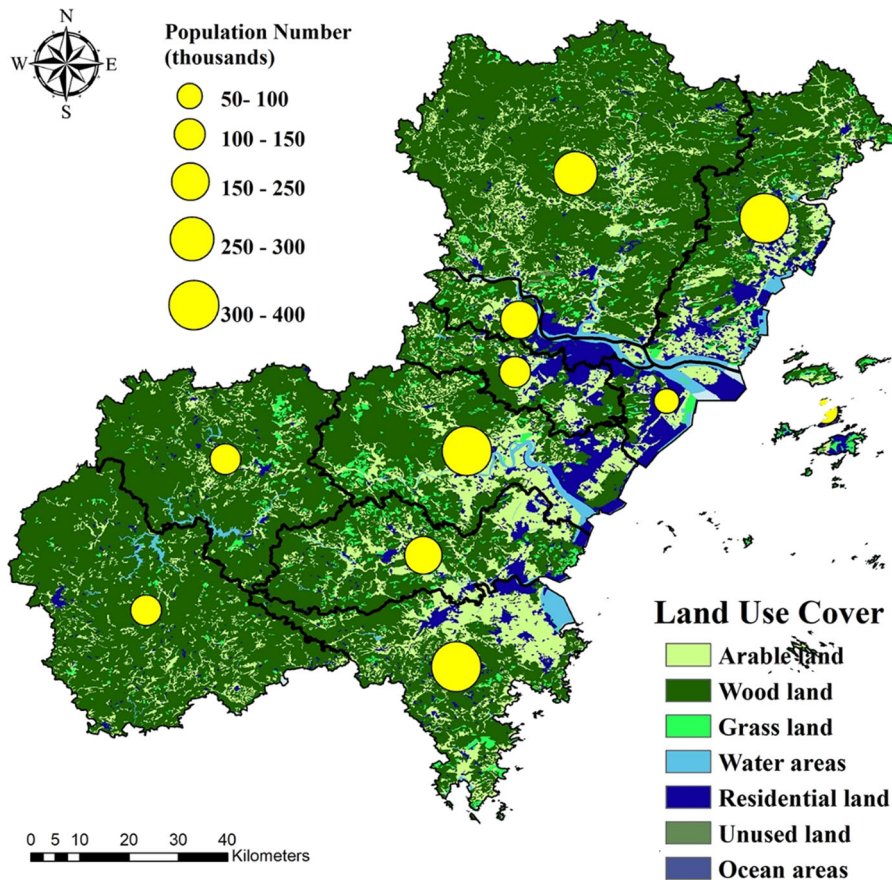


Fig. 3 Map of land use cover and population density in the study area

shale, had the most multiple contributions for Hg, As, and Pb.

Anthropogenic variables analysis

Industrial enterprises and mines The map of enterprises spatial distribution illustrated in Fig. 5 showed that most enterprises were clustered in the

coastal and main city areas. The map of kernel density clearly showed the degree of aggregation status; the highest density areas were also located along the riverside and urban areas. Fishnet maps of the four main categories, as shown as Fig. 6, delivered accurate quantity aggregation information for the different types of enterprises (Li et al., 2019). Figure 6a shows that the chemical manufacturing

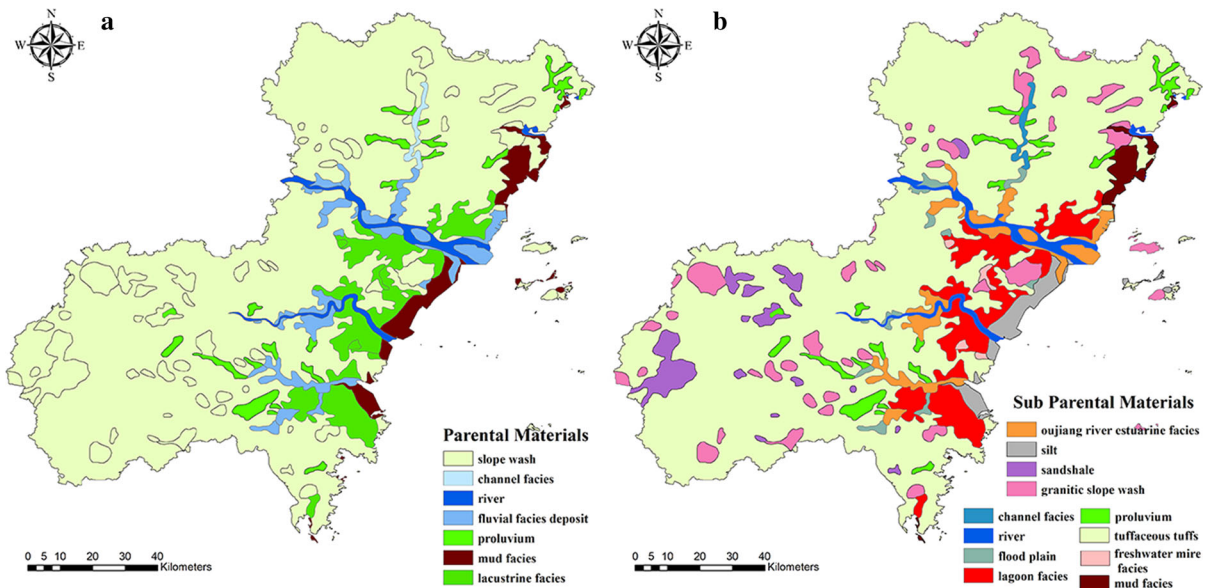


Fig. 4 Maps of soil parental materials and subgroups of parental materials

Table 5 Mean values for HMs' contents in different parental materials and their subgroups (mg/kg)

Parental material	Number	Subgroups of parental material	Number	Cd	Hg	As	Pb	Cr
Fluvial facies deposit	98	Flood plain	28	0.30	0.13	8.05	35.32	54.98
		Oujiang River Estuarine facies	70	0.24	0.16	9.13	34.93	59.24
Lacustrine facies	345	Freshwater mire facies	11	0.16	0.19	8.45	34.55	66.35
		Lagoon facies	334	0.30	0.16	9.80	36.10	55.41
Mud facies	102	silt	59	0.22	0.15	7.00	42.83	66.55
		Mud facies	43	0.17	0.17	6.84	33.41	61.90
Slope wash	849	Granitic slope wash	87	0.24	0.15	7.77	34.68	66.01
		Tuffaceous tuffs	723	0.31	0.17	8.70	37.63	58.84
		Sandshale	39	0.20	0.19	9.95	46.28	63.54
Channel facies	5	Channel facies	5	0.16	0.14	7.54	27.08	62.16
Proluvium	36	Proluvium	36	0.26	0.14	7.92	36.98	56.98
River ^d	39	River	39	-	-	-	-	-
P Value ^e	-	-	-	0.012	0.327	0.189	0.039	0.256

^d 39 sample points collected in 2016 were in the position where the soil map (1990) showed in the river

^eKruskal–Wallis Test—P value

enterprises were mainly distributed along the river and the south coast areas, and between one and five enterprises were mainly contained in each grid. Figure 6b illustrates that metal products showed a similar spatial distribution trend as in Fig. 6a, but with additional enterprises in each grid. As for Fig. 6c, life wastes enterprises were scattered in the entire study

area, whereas Fig. 6d shows that very few other enterprises were clustered in the Middle East coastal areas.

A statistical analysis of the HMs average contents in different enterprise categories is given in Table 6. The P values of most HMs within buffer areas of different enterprise categories also showed great

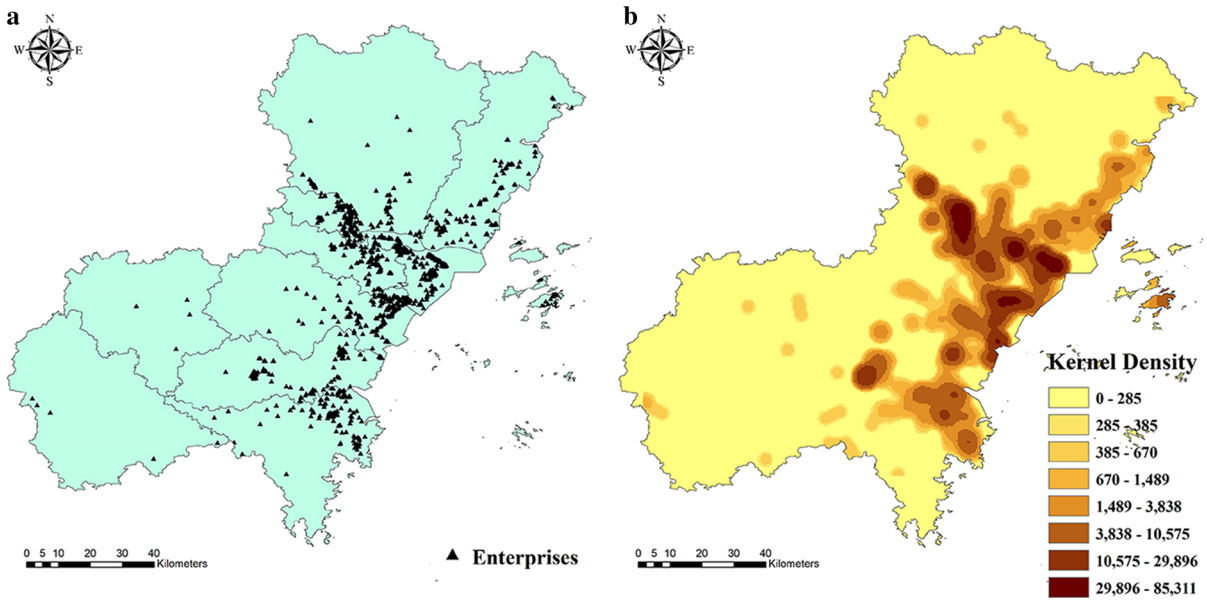


Fig. 5 Maps of enterprises' spatial distribution and kernel density

differences. As compared to 942 samples which had no industrial influences, the average contents of HMs in the chemical manufacturing, metal products, and life wastes enterprises were all above the original ones, confirming that enterprises with environmental wastes had impacts on the HMs in soils. Besides, it was also clear that with an increasing quantity of enterprises in a cluster, the HMs aggregated more in the soil. In particular, for metal products, the increasing trend between enterprise quantity and HM content was evident for Cd, Hg, and As, which were also the main elements in the industrial emissions.

The map of mine distribution (Fig. 7) showed the spatial location of every mine. As opposite from the cluster conditions for enterprises, the mines in Wenzhou City were generally scattered in the west side of the entire study area. Table 7 also shows that basically each type of mines all had different levels of influences on the HMs in soils, as the mean contents of HMs near mines were generally high above those far from mines. In particular, according to the largest value gaps between the average contents near mines and those far from mines, it was reasonable to believe mine exploitation had the most impact on Cd, As, and Cr (within the five elements).

Fertiliser inputs and traffic emissions The map of different fertilisers inputs in all towns and villages

(Fig. 8) revealed that high inputs of four fertilisers were basically centred in the north and middle counties of the study area, except for the southwest parts of the study area, with few little potash fertiliser inputs. Table 8 validates the influences that the fertiliser inputs have on the HMs in soils, since the P values of HMs within regions of different kinds of fertilisers inputs showed significant differences as well. With increases in the value intervals of different fertilisers, the contents of the HMs also increased. For the factor of traffic emissions, as the sample numbers falling into the line buffer areas were so limited, the impacts that traffic emissions (Table 9) have on the HMs in soils cannot be summarised clearly in our study. Accordingly, we ignored this factor in the discussion.

Source identification and apportionment

FMDM

The FMDM fitting results showed that Cd and Pb conformed to a log-normal distribution, whereas Hg, As, and Cr conformed to a log-normal mixture distribution. All of the five HMs passed the significance level test ($P > 0.05$). Cd and Pb fit the single log-normal distribution model shown in Fig. 9a, d. They both presented as only one single group.

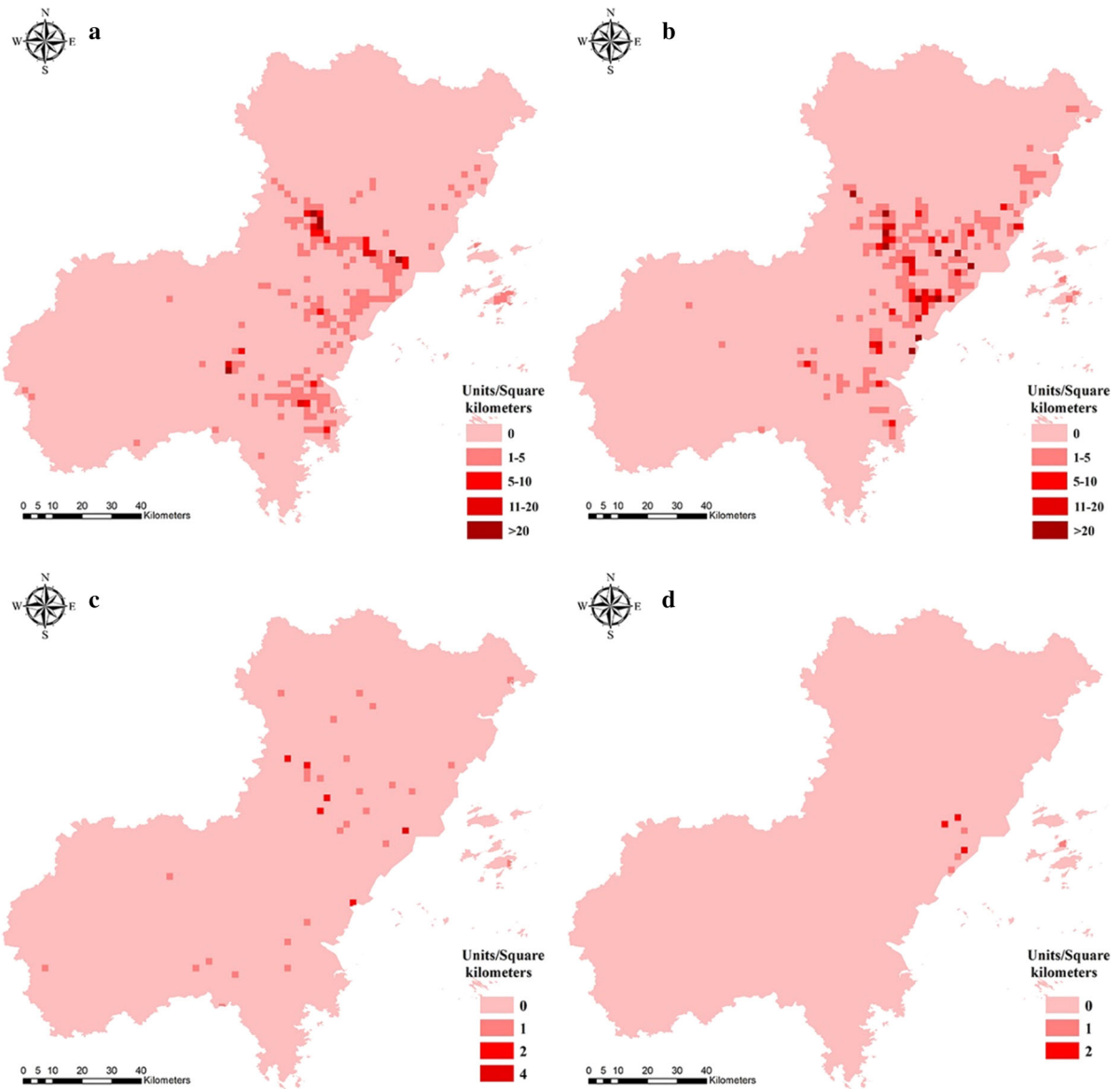


Fig. 6 Fishnet maps of enterprises divided as four main categories: **a** chemical manufacturing, **b** metal products, **c** life wastes, **d** others

According to Table 10, for Cd, its mean value was slightly higher than the background value but still within the background interval, revealing that it was basically coming from natural parental materials, but might have been slightly influenced by other anthropogenic activities (Hu et al., 2018). As for Pb, in cases where its mean value was lower than the background, it was believable that most Pb in such study areas could be generally attributed to natural resources.

In contrast, Hg exactly fit the double log-normal mixed distribution model shown in Fig. 9b. The FMDM model identified Hg as two main groups, where one was accounted for 81%, and the other accounted for 19%. As the result of the first group's mean value was 0.12, i.e. below the background value as 0.13, this source can be identified as a natural resource as well. For the other group, its mean value was 0.36, which was not only higher than the background value, but was also out of the range of

Table 6 Mean values for HMs' contents in buffer areas of different enterprise categories (mg/kg)

Categories	Enterprises quantity	Sample quantity	Cd	Hg	As	Pb	Cr
Chemical manufacturing ^f	1–10	309	0.33	0.17	9.26	38.07	56.15
	11–20	10	0.20	0.13	7.25	34.49	63.48
	> 20	9	0.33	0.20	9.69	37.79	60.34
	total	329	0.33	0.18	9.21	37.95	56.49
Metal products ^g	1–10	354	*0.28	*0.18	*8.73	37.15	59.06
	11–20	7	*0.50	*0.20	*12.68	35.63	57.67
	> 20	25	*0.59	*0.24	*12.71	37.17	60.07
	total	387	0.31	0.18	9.04	37.03	57.24
Life wastes ^h	total (0–4)	40	0.29	0.18	8.73	37.28	58.72
Others ⁱ	total (0–4)	31	0.23	0.15	9.87	40.16	63.04
All categories	0	942	0.26	0.17	8.57	36.04	58.97
	1–10	456	*0.32	*0.16	*8.10	*36.90	*58.16
	11–20	41	*0.40	*0.17	*8.22	*37.52	*59.61
	> 20	35	*0.46	*0.27	*8.75	*42.31	*68.84
P Value ^j	–	–	0.034	0.159	0.023	0.049	0.077

^fSpecified as chemical fibre manufacturing, chemical raw materials and chemical products manufacturing, 0Leather and footwear manufacturing, petroleum and gas engineering, pharmaceutical manufacturing

^gSpecified as metal products industry, coloured metal smelting and rolling processing industry

^hSpecified as public facilities management industry, environmental protection and renovation industry

ⁱOnly refers to warehousing industry

^jKruskal–Wallis Test—P value

*Indicates that with the increase in the quantity of enterprises, the contents of HMs also increase

background interval, showing that it was seriously influenced by one possible anthropogenic source.

As and Cr were identified as triple-normal mixed distribution models, as shown in Fig. 9c, e. There were three groups both identified for these two elements. For As, the account rate was 41%, 42%, and 17% for each group. The same as in the previous analysis, three potential sources could be identified (one natural source and two anthropogenic sources). In addition, as the second group's mean value was still within the background interval range, but the third one was out of such range, the second group for As can be identified as moderately polluted, whereas the third group can be recognised as severely polluted. The same analysis can also apply for Cr, i.e. it can be recognised as one natural source accounting for 48%, and two moderately polluting sources, each accounting for 26%.

RB-PCA

Obtained from HMs of 1474 sample points, the IDW prediction map of five HMs provided reliable basis for

RB-PCA, as both low values of ME and RMSE of five HMs revealed fine prediction accuracy (Huang, 2020). The sum of the cumulative variance showed that the five HMs could be divided into two main components (> 85%) (Wang et al., 2018). However, such results also showed that the contributions of Hg to both two components were so minor that could be ignored. In this case, here we justified two components to three, and the rotation matrix is shown in Table 11. Three components accounted for 35.46%, 55.38%, and 9.16% of the total, respectively.

The contribution of the first principle component (PC1) had factor loadings of 0.79 for both Cd and Pb, indicating that these two elements came from the same pollution source. The OK spatial interpolation map (Fig. 10a) showed that high scores of PC1 were generally scattered all over the study area. The second principal component (PC2) takes up 55.38%, with high factor loadings of 0.83 and 0.63 for As and Cr, respectively. Figure 10b shows that the high-scoring clusters were mainly distributed in the west part of the study area. With factor loadings of 0.90 and 0.48, the

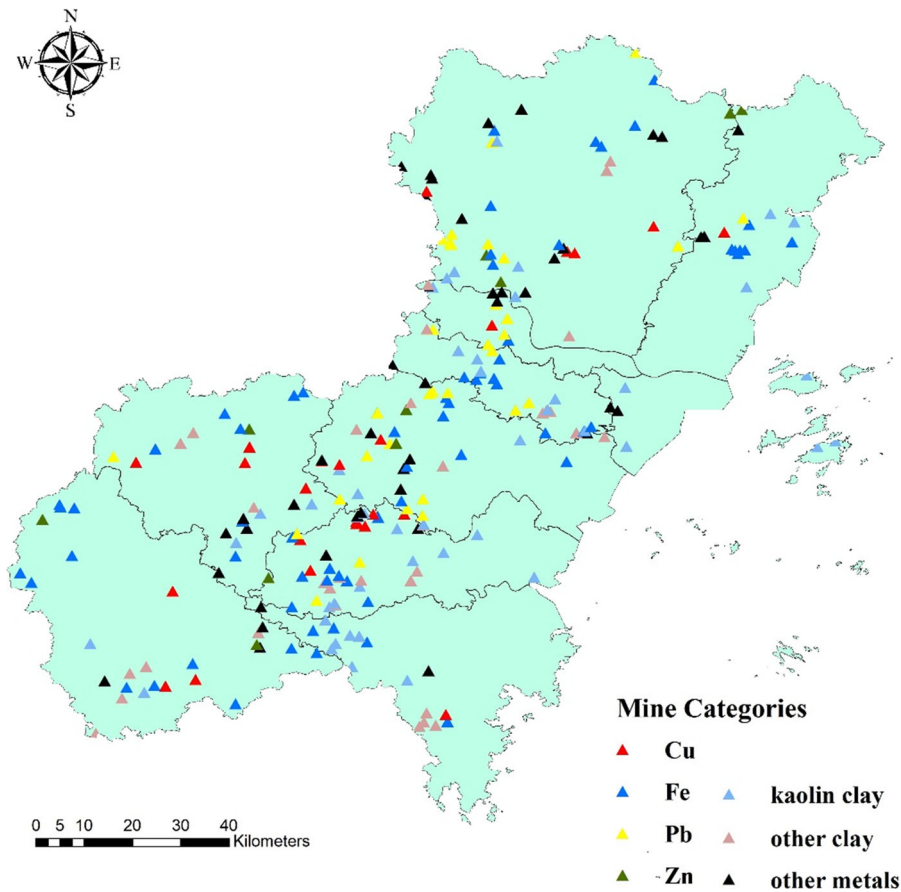


Fig. 7 The map of mine distribution of different categories

Table 7 Mean values for HMs' contents in buffer areas of different mine categories (mg/kg)

Categories	Mine quantity	Sample number	Cd	Hg	As	Pb	Cr
Copper mine	30	18	0.41	0.16	10.18	38.56	59.41
Iron mine	67	76	0.39	0.16	9.84	37.61	60.37
Lead mine	39	32	0.40	0.16	10.05	38.07	59.52
Zinc mine	22	10	0.40	0.15	11.21	35.27	60.69
Kaolin clay mine	46	47	0.40	0.16	10.02	38.16	59.48
Other metal mines ^k	47	21	0.39	0.16	9.96	37.89	60.14
Other clay mines	34	30	0.39	0.16	9.92	37.93	59.77
All	285	234	0.38	0.16	9.79	37.68	60.86
Far from mines (2 km)	0	1240	0.28	0.16	8.73	37.13	59.09
P Value ^l	–	–	0.002	0.016	0.033	0.178	0.086

^k Include Ti, Mn, Au, Ag, Nb, Y, Mo; Other metal mines include feldspar, mica, little clay, pyrophyllite, alunite, fluorite

^lKruskal–Wallis Test—P value

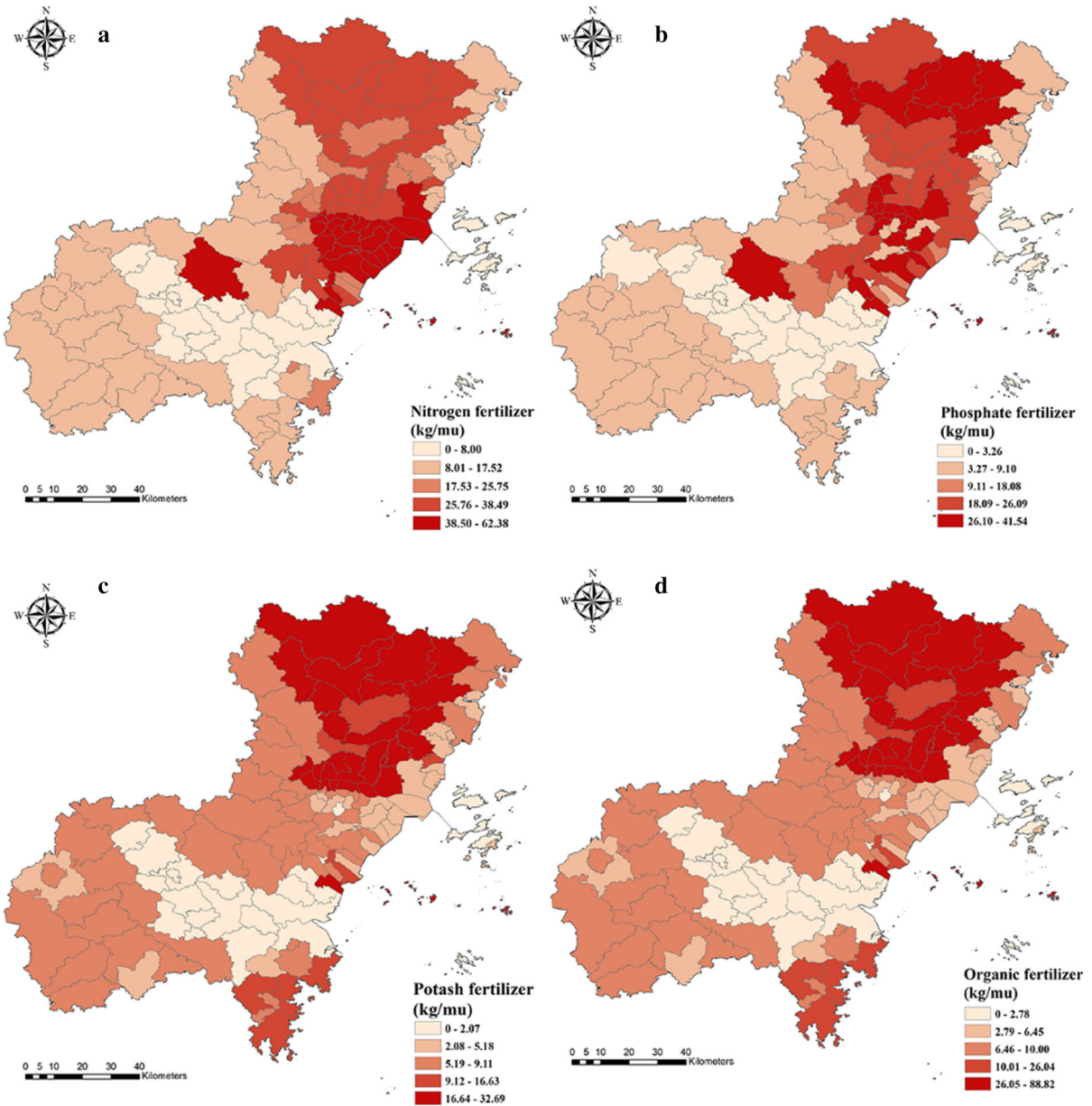


Fig. 8 The map of contents of different fertilisers in towns or villages

third component (PC3) contains Hg and Cr. Figure 10c shows high scores generally centred in the north and middle parts of the study area, where the residents and industrial enterprises were located.

PMF

Considering a few experiments results of setting the factor number of PMF model from 1–5, and pairing

with the previous two model analysis, we set the factor number to three in which the accuracy test shows the highest, and then validated the model fitting status and uncertainty. The basic model fitting parameters of the PMF are shown in Table 12. Both the general linear fitting status and the uncertainty diagnostics of the five HMs were quite satisfactory, with the highest R^2 (of Hg) reaching 0.99, and the lowest (Pb) still above 0.73. The KS test P values were all above 0.05, indicating an

Table 8 Mean values for HMs' contents in different fertilisers value interval (mg/kg)

Fertilizer categories	Inputs interval	Sample number	Cd	Hg	As	Pb	Cr
Nitrogen	0–8.00	336	0.28	0.17	8.57	37.02	58.56
	8.01–17.52	549	0.28	0.16	8.73	37.13	59.09
	17.53–25.75	109	0.28	0.16	8.73	37.15	58.99
	25.76–38.49	234	0.28	0.16	8.73	37.13	59.09
	38.50–62.38	246	0.30	0.16	8.91	36.36	58.64
Phosphate	0–3.26	347	0.28	0.17	*8.57	37.02	*58.56
	3.27–9.10	448	0.28	0.16	*8.63	37.13	*59.09
	9.11–18.08	557	0.28	0.17	*8.73	37.07	*59.09
	18.09–26.09	268	0.28	0.16	*8.73	37.13	*59.77
	26.10–41.54	206	0.28	0.16	*8.84	37.13	*60.03
Potash	0–2.07	328	0.28	0.16	8.57	*36.02	58.56
	2.08–5.18	232	0.28	0.16	8.67	*36.77	58.57
	5.19–9.11	614	0.28	0.16	8.73	*38.13	59.09
	9.12–16.63	103	0.28	0.17	8.73	*38.37	59.09
	16.64–32.69	192	0.28	0.19	8.73	*38.39	59.09
Organic	0–2.78	923	0.28	0.16	*8.43	*37.02	*58.09
	2.79–6.45	110	0.28	0.17	*8.57	*37.02	*58.56
	6.46–10.00	165	0.28	0.17	*8.71	*37.85	*58.79
	10.01–26.04	74	0.28	0.16	*8.73	*38.13	*59.79
	26.05–88.82	202	0.28	0.16	*8.73	*38.13	*60.07
P Value ^m	–	–	0.147	0.167	0.055	0.047	0.003

^mKruskal–Wallis Test—P value;

* Indicates that with the increase in the value interval of different fertilizers, the contents of HMs also increase

Table 9 Mean values for HMs contents in different road thresholds (mg/kg)

Road type	Threshold distance	Sample number	Cd	Hg	As	Pb	Cr
Country road	200	27	0.32	0.21	8.52	37.48	68.05
Highway	150	9	0.27	0.29	7.67	37.44	69.39
Province road	100	29	0.23	0.20	7.78	37.53	62.48
Railroad	100	2	0.30	0.05	5.25	30.15	50.00
County road	50	62	0.28	0.16	9.44	36.24	58.40
None	/	1345	0.28	0.16	8.74	37.17	58.82
ⁿ Kruskal–Wallis Test—P value	P Value ⁿ	–	0.332	0.432	0.161	0.250	0.552

accurate fit of the PMF modelling (Lv et al. 2018). Within 50 Bootstrap runs, the DISP LD in Q and %dQ remained pretty low level, revealing a stable model output (Norris and Duvall 2014). In addition, to make the contribution gap clearer, we rotated the positive matrix, by setting the F peak = −0.5 (Han et al., 2016).

According to the results from Table 13 and Fig. 11, Hg was exclusively monopolised by factor 1 with the highest contribution of 80.77%, and comparably low contributions of other factors. Factor 2 highly contributed 89.22% and 84.81% for both Cd and Pb, respectively, and contributed 7.31%, 35.84%, and 27.42% for Hg, As, and Cr, respectively. Factor 3 had less than great contributions for each HM, but was

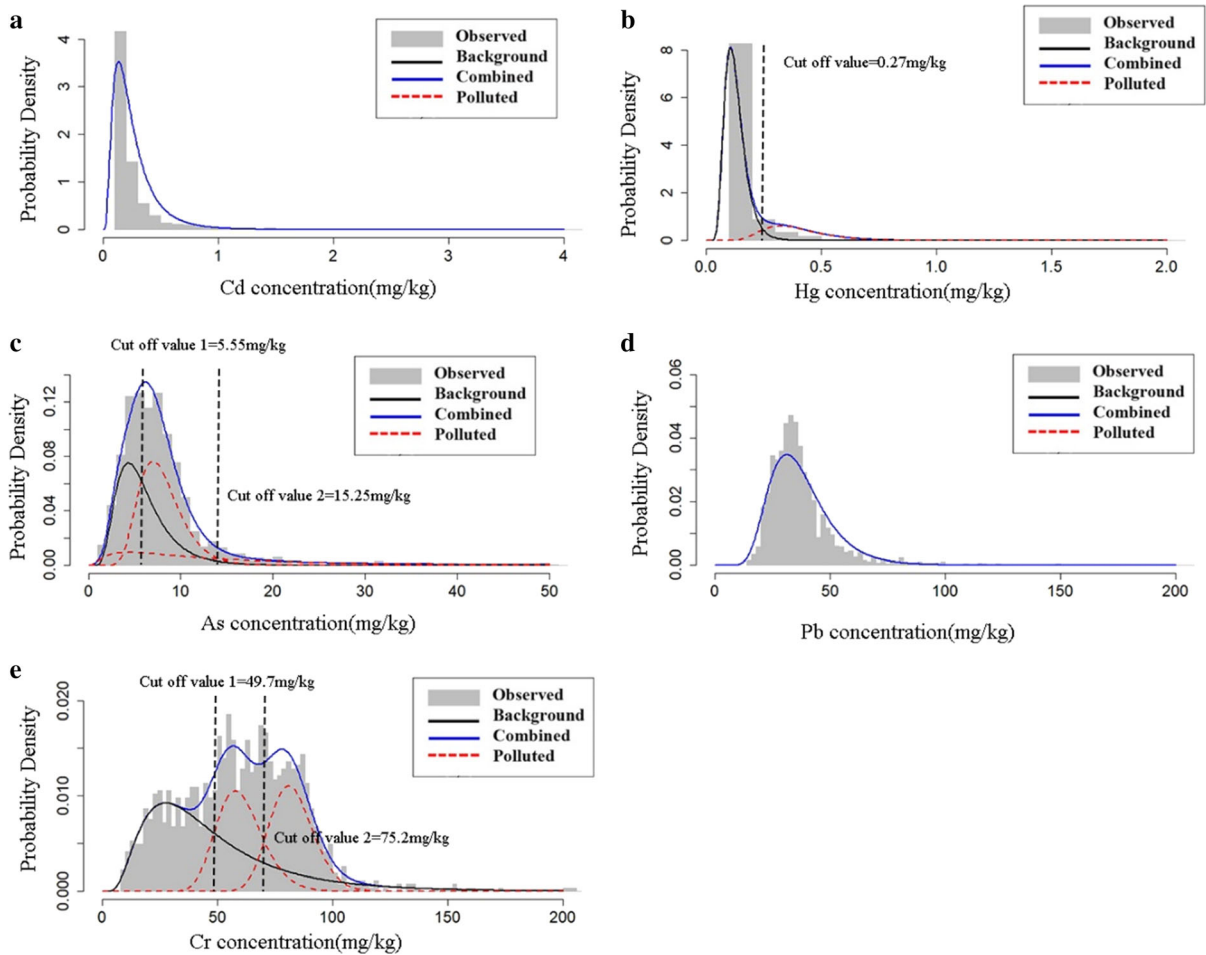


Fig. 9 Finite mixture distribution model (FMDM) fitting of Cd **a**, Hg **b**, As **c**, Pb **d**, Cr **e** in soils

Table 10 The parameters, goodness of fit, and cut-off values for finite mixture distribution model (FMDM)

Element	Group	MP ^o	Mean	Std ^p	DF ^q	χ^2	P Value	Cut-off	BV ^r	BI ^s
Cd	1	100%	0.21	0.446	4	53.64	0.16	/	0.18	0.09–0.26
Hg	2	81%	0.12	0.124	4	27.98	0.19	0.27	0.13	0.03–0.23
		19%	0.36	0.159						
As	3	41%	5.33	0.211	8	51.78	0.36	5.55	5.86	3.67–8.04
		42%	7.65	0.091						
		17%	11.34	0.975						
Pb	1	100%	35.20	0.119	5	46.32	0.14	/	41.6	31.0–52.3
Cr	3	48%	40.46	0.382	8	42.35	0.17	49.7	56.9	3.6–110.2
		26%	59.50	0.028						
		26%	82.04	0.013						

^oMmeans mixed proportions;

^pMeans standard of deviation;

^qMeans degree of freedom;

^rMeans background value,

^sMeans background interval

especially high for As and Cr, with contributions of 48.23% and 46.95%, respectively. In that regard, the high contributions of each factor to the applied HMs

exactly matched the results of the previous two models.

Table 11 Raster-based principal component analysis (RB-PCA) rotation matrix parameters

Element	IDW-raster		Component matrix			Rotated component matrix		
	ME ^t	RMSE ^u	PC1	PC2	PC3	PC1 (35.46%)	PC2 (55.38%)	PC3 (9.16%)
Cd	5.14E-03	0.67	0.63	-0.41	0.34	0.79	0.22	-0.11
Hg	9.87E-04	0.18	0.53	0.26	-0.69	0.12	-0.01	0.90
As	3.77	6.71	0.43	0.42	0.60	0.15	0.83	-0.10
Pb	2.01	1.28	0.64	-0.53	-0.16	0.79	-0.16	0.24
Cr	0.28	3.21	0.39	0.71	-0.03	-0.17	0.63	0.48

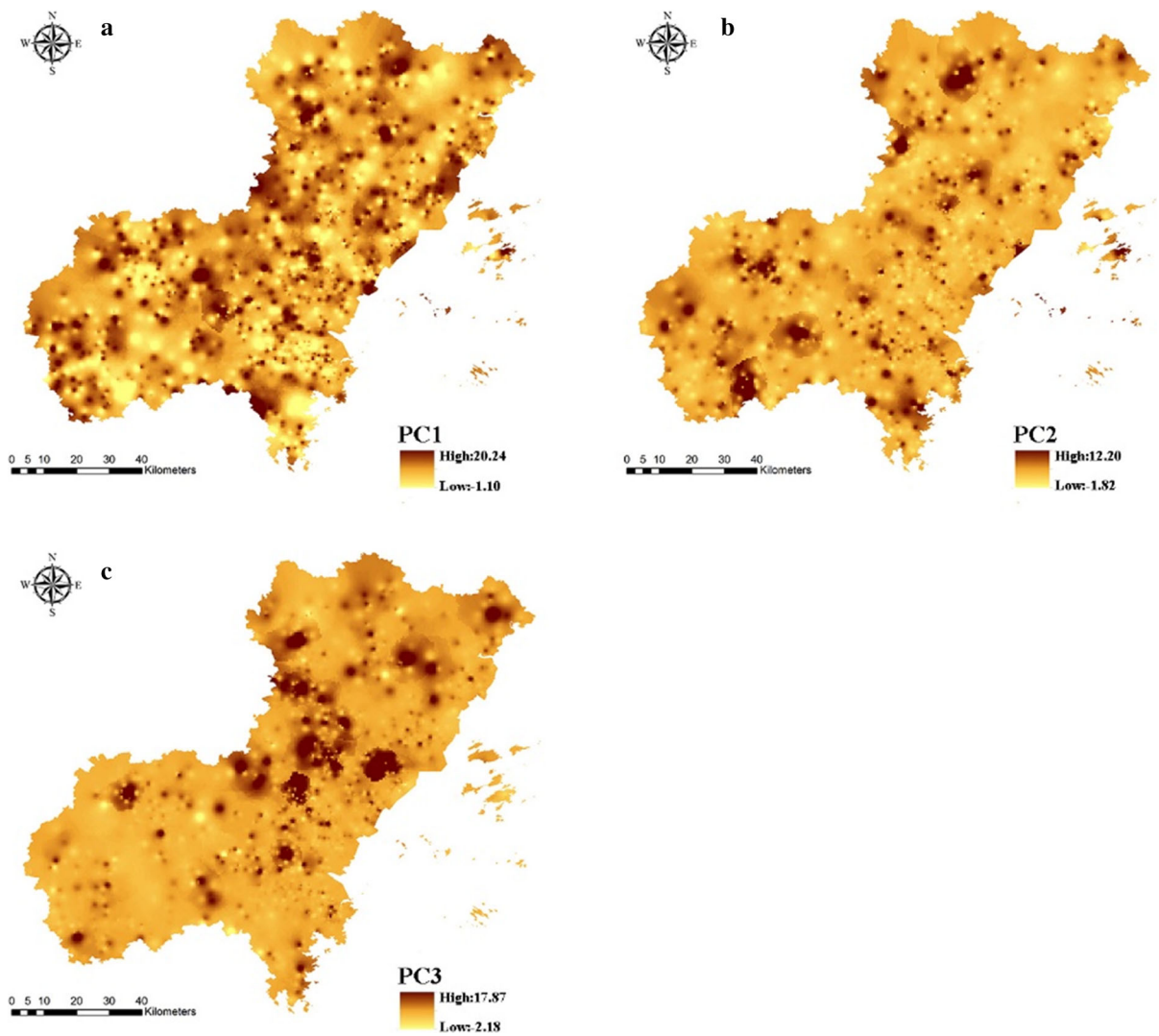
^tMeans Mean Error;^uMeans Root Mean Square Error**Fig. 10** Raster interpolation maps of three principal components

Table 12 The fitting parameters and uncertainty of the positive matrix factorization (PMF) model

Elements	Fitting parameters				Uncertainty diagnostics		
	Intercept	slope	P ^v	R ²	Bootstrap run	DISP LD in Q ^w	DISP %dQ ^x
dCd	0.15	0.85	0.05	0.81	50	−3.93	−2.91
Hg	0.00	0.98	0.28	0.99			
As	5.27	0.74	0.35	0.76			
Pb	13.44	0.63	0.16	0.73			
Cr	11.42	0.80	0.27	0.87			

^vMeans Kolmogorov–Smirnov Test—P value;

^w Means DISP (Base Model Displacement) largest decrease in Q;

^xMeans DISP (Base Model Displacement) [(Q_{robust} − Q_{true})/ Q_{True}]

Table 13 Source profile and contribution of the positive matrix factorization (PMF) model

Elements	Source Profile (mg/kg)			Source Contribution (%)		
	Factor1	Factor2	Factor3	Factor1	Factor2	Factor3
Cd	0.01	0.08	0.08	2.94	89.22	7.84
Hg	0.11	0.01	0.06	80.77	7.31	11.92
As	0.34	0.90	4.45	15.93	35.84	48.23
Pb	1.78	20.37	15.02	4.79	84.81	10.40
Cr	3.35	4.41	51.70	25.63	27.42	46.95

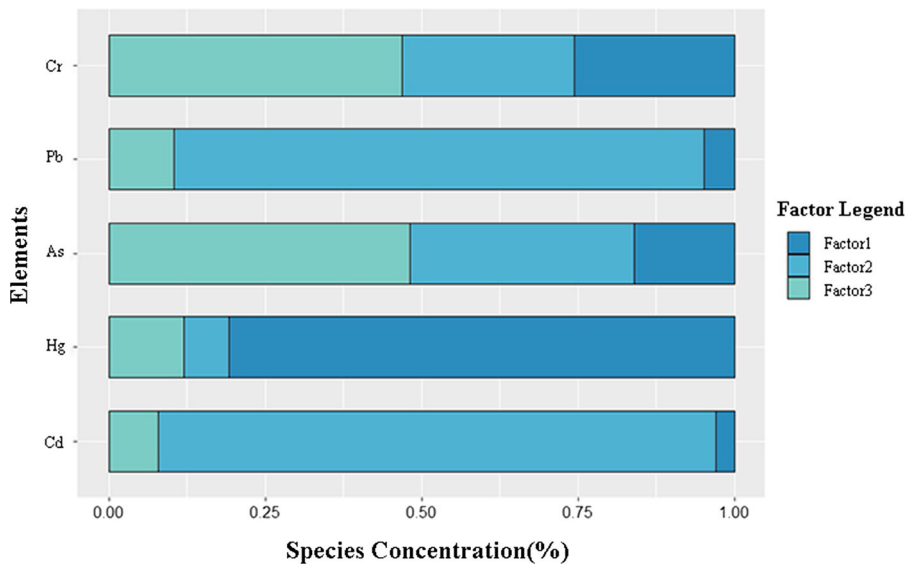


Fig. 11 The contribution percentage stacked bar chart

Comprehensive interpretation of contamination sources

By combining the results of three models, we found that there are mainly three contamination sources. The first contamination source was identified as natural lithogenic source with especially high contributions to Cd and Pb. The qualitative results of the PCA illustrated that the spatial distribution (Fig. 10a) of PC1 was scattered all over the study area, meaning it had no cluster trend, which was deemed as the symbol of anthropogenic activities. Moreover, the quantitative analysis (Table 5) also showed slope wash (containing tuffaceous tuffs and sand shale) taking up the largest parental materials portion of the study area (Fig. 4) with high values of Cd and Pb, matching the spatial analysis result. The results of the FMDM also showed the single log-normal statistical status of these two elements. As their mean values were close to the background values, it was certified that the first contamination source should be a natural lithogenic source. Moreover, according to the quantitative results of the PMF, the natural lithogenic source contributed especially high for Cd and Pb as 89.22% and 84.81%, respectively, and for Hg, As, and Cr as 7.31%, 35.84%, and 27.42%, respectively.

The second contamination source was attributed to industrial and agricultural mixed pollution, especially Hg-related pollution. As the results of the PCA suggested, the OK score interpolation map of PC3 (Fig. 10c) clustered on the urban areas, which is similar to the areas where most of the metal products and chemical manufacturing enterprises were located at (Fig. 6a and Fig. 6b); moreover, the distribution of fertilisers inputs (Fig. 8) also presents similar distribution as Fig. 10c. Besides, the corresponding quantitative statistical analysis of industrial enterprises (Table 6) revealed that metal products enterprises had severe impacts on Hg. The FMDM suggested that Hg had two main sources: one was a natural source, as discussed previously, and the other one was evidently from industrial Hg contamination which is mostly from enterprises of metal products, chemical manufacturing etc. In cases where the Hg was also influenced by potash and organic fertiliser inputs (Fig. 8c, d) and their high inputs areas also matched the urban areas, it was probable that the fertiliser inputs may also have had impacts associated with the enterprise emissions effect. According to the

quantitative results of the PMF, industrial emissions and fertiliser inputs mixed pollution contributed particularly high for Hg at 80.77%, and comparably low for Cd, As, Pb, and Cr at 2.94%, 15.93%, 4.79%, and 25.63%, respectively.

As for the third contamination source, it was dominated by mine exploitation activities. The PCA results suggested that according to Fig. 10b, the high scores were mainly clustered in the west part of the study area, generally matching the trend of where the mines were located at in the study area, as shown in Fig. 7. Table 5 also shows that mine exploitation had significant impacts on As and Cr; therefore, PC2 can be deemed as a main contribution from mine exploitation. The FMDM results also illustrated that As and Cr had a triple log-normal mixed distribution, which would have been influenced by mine exploitation activities, in addition to the natural source and industrial emissions sources previously discussed. The PMF results showed that this factor has a high contribution for As and Cr at 48.23% and 46.95%, respectively, and of 7.84%, 11.92%, and 10.40% for Cd, Hg, and Pb, respectively.

Conclusion

This study employed qualitative and quantitative approaches (including FMDM, RB-PCA, and PMF) and applied multiple auxiliary variables to identify and apportion potential contamination sources of HMs (Cd, Hg, As, Pb, and Cr) in soils in Wenzhou City, China. The potential ecological risk assessment showed that the general ecological risk level of the entire study area was low, but Hg and Cd could be threatening, as they were considered as the main contributors and had extremely high values at the maximum level. The results indicated that three main sources had been identified and apportioned. Parental materials contributed to all elements, and especially for Cd and Pb, with high contributions up to 89.22% and 84.81%, respectively, and contributions for Hg, As, and Cr of 7.31%, 35.84%, and 27.42% respectively. Industrial emissions mixed with fertiliser inputs were another vital contamination factor in the study area, owing to large proportions of metal products enterprises and related fertiliser inputs. These contributed to extremely high levels for Hg, with contributions of 80.77%, 2.94%, 15.93%, 4.79%, and

25.63% for Cd, As, Pb, and Cr, respectively. Mine exploitation was the third important contamination source; it had high contributions for As and Cr at 48.23% and 46.95%, respectively, and at 7.84%, 11.92%, and 10.40% for Cd, Hg, and Pb, respectively. Such results can help decision-makers in making more efficient and scientific decisions in regulating HM pollution.

Acknowledgments This work was supported by National Key Research and Development Program (2018YFC1800201). We also acknowledge the support received by Bifeng Hu from the China Scholarship Council (under grant agreement No. 201706320317) for 3 years' Ph.D. study in INRAE and Orléans University in Orléans, France.

Author contributions SS was involved in conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing—original draft, and visualization. BH was involved in conceptualization, validation, formal analysis, writing—review & editing, and visualization. YT was involved in software, formal analysis, formal analysis, and visualization. QY was involved in software, investigation, and validation. MH was involved in data curation and funding acquisition. LZ was involved in investigation and project administration. QCh was involved in conceptualization, methodology, formal analysis, writing—review & editing, and supervision. ZS was involved in conceptualization, methodology, validation, writing—review & editing, and funding acquisition.

Funding This work was supported by National Key Research and Development Program (2018YFC1800201).

Declarations

Conflicts of interests The authors have no financial or proprietary interests in any material discussed in this article.

Informed consent Informed consent was obtained from all individual participants included in the study.

Consent to publication The participant has consented to the submission of the case report to the journal.

References

Abanuz, G. Y. (2019). Application of multivariate statistics in the source identification of heavy-metal pollution in roadside soils of Bursa Turkey. *Arabian Journal of Geosciences*, *12*, 382. <https://doi.org/10.1007/s12517-019-4545-3>

Aminiyan, M. M., Baalousha, M., Mousavi, R., Aminiyan, F. M., Hosseini, H., & Heydariyan, A. (2018). The ecological risk, source identification, and pollution assessment of heavy metals in road dust: a case study in Rafsanjan, SE

Iran. *Environmental Science and Pollution Research*, *25*, 13382–13395. <https://doi.org/10.1007/s11356-017-8539-y>

Brock, J. E. (1953). Variation of coefficients of simultaneous linear equation. *Quarterly of Applied Mathematics*, *11*, 234–240.

Chen, H. Y., Teng, Y. G., Wang, J. S., Song, L. T., & Zuo, R. (2013). Source apportionment of trace element pollution in surface sediments using positive matrix factorization combined support vector machines: Application to Jinjiang River China. *Biology Trace Element Research*, *151*, 462–470. <https://doi.org/10.1007/s12011-012-9576-5>

Chen, X. D., & Lu, X. W. (2018). Contamination characteristics and source apportionment of heavy metals in topsoil from an area in Xi'an city, China. *Ecotoxicology and Environmental Safety*, *151*, 153–160. <https://doi.org/10.1016/j.ecoenv.2018.01.010>

Dai, L. Y., Wang, L. Q., Li, L. F., Liang, T., Zhang, Y. Y., Ma, C. X., et al. (2018). Multivariate geostatistical analysis and source identification of heavy metals in the sediment of Poyang Lake in China. *Science of the Total Environment*, *621*, 1433–1444. <https://doi.org/10.1016/j.scitotenv.2017.10.085>

Davies, B. E. (1997). Heavy metal contaminated soils in an old industrial area of wales, Great Britain: Source identification through statistical data interpretation. *Water, Air & Soil Pollution*, *94*, 85–87. <https://doi.org/10.1023/A:1026478427782>

Dong, B., Zhang, R. Z., Gan, Y. D., Cai, L. Q., Freidenreich, A., Wang, K. P., et al. (2019). Multiple methods for the identification of heavy metal sources in cropland soils from a resource-based region. *Science of the Total Environment*, *651*, 3127–3138. <https://doi.org/10.1016/j.scitotenv.2018.10.130>

Dong, R. Z., Jia, Z. M., & Li, S. Y. (2018). Risk assessment and sources identification of soil heavy metals in a typical county of Chongqing municipality, Southwest China. *Process Safety and Environmental Protection*, *113*, 275–281. <https://doi.org/10.1016/j.psep.2017.10.021>

Facchinelli, A., Sacchi, E., & Mallen, L. (2001). Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. *Environment Pollution*, *114*, 313–324. [https://doi.org/10.1016/S0269-7491\(00\)00243-8](https://doi.org/10.1016/S0269-7491(00)00243-8)

Glaser, B., Dreyer, A., Bock, M., Fiedler, S., Mehring, M., & Heitmann, T. (2005). Source apportionment of organic pollutants of a highway-traffic-influenced urban area in Bayreuth (Germany) using biomarker and stable carbon isotope signatures. *Environmental Science of Technology*, *39*, 3911–3917. <https://doi.org/10.1021/es050002p>

Hakason, L. (1980). An ecological risk index for aquatic pollution control: A sediment logical approach. *Water Research*, *14*, 975–1001. [https://doi.org/10.1016/0043-1354\(80\)90143-8](https://doi.org/10.1016/0043-1354(80)90143-8)

Han, D. M., Cheng, J. P., Hu, X. F., Jiang, Z. Y., Mo, L., Xu, H., Ma, Y., Chen, X., & Wang, H. (2016). Spatial distribution, risk assessment and source identification of heavy metals in sediments of the Yangtze River Estuary China. *Marine Pollution Bulletin*, *115*, 141–148. <https://doi.org/10.1016/j.marpolbul.2016.11.062>

Hu, B.F., Jia, X.L., Hu, J., Xu, D.Y., Xia, F., & Li, Y. (2017). Assessment of heavy metal pollution and health risks in the soil-plant-human system in the Yangtze River Delta,

- China. *International Journal of Environmental Research and Public Health*, 14(9): 1042. <https://doi.org/10.3390/ijerph14091042>
- Hu, B. F., Shao, S., Fu, Z. Y., Li, Y., Ni, H., Chen, S. C., et al. (2019). Identifying heavy metal pollution hot spots in soil-rice systems: A case study in South of Yangtze River Delta, China. *Science of the Total Environment*, 658, 614–625. <https://doi.org/10.1016/j.scitotenv.2018.12.150>
- Hu, B. F., Shao, S., Ni, H., Fu, Z. Y., Hu, L. S., Zhou, Y., et al. (2020). Current status, spatial features, health risks, and potential driving factors of soil heavy metal pollution in China at province level. *Environmental Pollution*. <https://doi.org/10.1016/j.envpol.2020.114961>
- Hu, Y. N., Wang, H. F., Dong, L. R., Huang, B. A., Borggaard, O. K., Hansen, H. C. B., et al. (2018). Source identification of heavy metals in peri-urban agricultural soils of southeast China: An integrated approach. *Environment Pollution*, 237, 650–661. <https://doi.org/10.1016/j.envpol.2018.02.070>
- Hu, Y. N., & Cheng, H. F. (2013). Application of stochastic models in identification and apportionment of heavy metals pollution sources in the surface soils of a large-scale region. *Environmental Science and Technology*, 47, 3752–3760. <https://doi.org/10.1021/es304310k>
- Huang, H. (2020). The comparison of interpolation methods and pollution assessments of heavy metals in mini scale block. *Environmental Ecology*, 2(3), 34–40. (in Chinese).
- Huang, Y., Deng, M. H., Wu, S. F., Jan, J., Li, T. Q., Yang, X. E., et al. (2018). A modified receptor model for source apportionment of heavy metal pollution in soil. *Journal of Hazard. Materials*, 354, 161–169. <https://doi.org/10.1016/j.jhazmat.2018.05.006>
- Huang, Y., Li, T. Q., Wu, C. X., He, Z. L., Japenga, J., Deng, M., et al. (2015). An Integrated approach to assess heavy metal source apportionment in pen-urban agricultural soils. *Journal of Hazard Materials*, 299, 540–549. <https://doi.org/10.1016/j.jhazmat.2015.07.041>
- Jiang, Y. X., Chao, S. H., Liu, J. W., Yang, Y., Chen, Y. J., Zhang, A. C., et al. (2017). Source apportionment and health risk assessment of heavy metals in soil for a township in Jiangsu Province, China. *Chemosphere*, 168, 1658–1668. <https://doi.org/10.1016/j.chemosphere.2016.11.088>
- Kashif, A., Said, M., Wajid, A., Ishtiaq, J., & Atta, R. (2020). Occurrence, source identification and potential risk evaluation of heavy metals in sediments of the Hunza River and its tributaries Gilgit-Baltistan. *Environmental Technology & Innovation*, 18, 100700. <https://doi.org/10.1016/j.eti.2020.100700>
- Kosharna, S., & Korzhnev, M. (2018). Heavy metals on technogenic objects of mining and processing enterprises of the Kryvyi Rig Basin. *Visnyk of Taras Shevchenko National University of Kyiv-Geology*, 2, 92–97.
- Li, F., Liu, S. Y., Li, Y., & Shi, Z. (2019). Spatiotemporal analysis and source apportionment of heavy metals in soils in a developed industrial city. *Environment Science*, 40, 0250–3301. (in Chinese).
- Li, Y. X., Song, N. N., Yu, Y., Yang, Z. F., & Shen, Z. Y. (2017). Characteristics of PAHs in street dust of Beijing and the annual wash-off load using an improved load calculation method. *Science of the Total Environment*, 581–582, 328–336. <https://doi.org/10.1016/j.scitotenv.2016.12.133>
- Liang, J., Feng, C. T., Zeng, G. M., Gao, X., Zhong, M. Z., Li, X. D., et al. (2017). Spatial distribution and source identification of heavy metals in surface soils in a typical coal mine city, Lianyuan, China. *Environment Pollution*, 225, 681–690. <https://doi.org/10.1016/j.envpol.2017.03.057>
- Liao, S. Y., Jin, G. Q., Muhammad, A. K., Zhu, Y. W., Duan, L., Luo, W. X., Jia, J., et al. (2020). The quantitative source apportionment of heavy metals in peri-urban agricultural soils with UNMIX and input fluxes analysis. *Environmental Technology & Innovation*. <https://doi.org/10.1016/j.eti.2020.101232>
- Lin, Y. P., & Cheng ShyuChang, B. Y. G. S. T. K. (2009). Combining a finite mixture distribution model with indicator kriging to delineate and map the spatial patterns of soil heavy metal pollution in Chunghua County, Central Taiwan. *Environment Pollution*, 158, 235–244. <https://doi.org/10.1016/j.envpol.2009.07.015>
- Luo, X. S., Xue, Y., Wang, Y. L., Cang, L., Xu, B., & Ding, J. (2015). Source identification and apportionment of heavy metals in urban soil profiles. *Chemosphere*, 127, 152–157. <https://doi.org/10.1016/j.chemosphere.2015.01.048>
- Lv, J. S., & Liu, Y. (2018). An integrated approach to identify quantitative sources and hazardous areas of heavy metals in soils. *Science of the Total Environment*, 646, 19–28. <https://doi.org/10.1016/j.scitotenv.2018.07.257>
- Manoli, E., Voutsas, D., & Samara, C. (2002). Chemical characterization and source identification/apportionment of fine and coarse air particles in Thessaloniki, Greece. *Atmospheric Environment*, 36, 949–961. [https://doi.org/10.1016/S1352-2310\(01\)00486-1](https://doi.org/10.1016/S1352-2310(01)00486-1)
- Marchant, B. P., Saby, N. P. A., & Arrouays, D. (2017). A survey of topsoil arsenic and mercury concentrations across France. *Chemosphere*, 181, 635–644. <https://doi.org/10.1016/j.chemosphere.2017.04.106>
- Marchant, B. P., Saby, N. P. A., Jolivet, C. C., Arrouays, D., & Lark, R. M. (2011). Spatial prediction of soil properties with copulas. *Geoderma*, 162, 327–334. <https://doi.org/10.1016/j.geoderma.2011.03.005>
- Mehr, M. R., Keshavarzi, B., Moore, F., Reza, S., Lahijanzadeh, A., & Kermani, M. (2017). Distribution, source identification and health risk assessment of soil heavy metals in urban areas of Isfahan province Iran. *Journal of African Earth Science*, 132, 16–26. <https://doi.org/10.1016/j.jafrearsci.2017.04.026>
- Men, C., Liu, R. M., Wang, Q. R., Guo, L. J., & Shen, Z. Y. (2018). The impact of seasonal varied human activity on characteristics sources of heavy metals in metropolitan road dusts. *Science of the Total Environment*, 637–638, 844–854. <https://doi.org/10.1016/j.scitotenv.2018.05.059>
- Nabulo, G., Origina, H., & Diamond, M. (2006). Assessment of Lead, Cadmium, and Zinc contamination of road side soils surface films, and vegetables in Kampala City Uganda.

- Environmental Research*, 101(1), 42–52. <https://doi.org/10.1016/j.envres.2005.12.016>
- Nanos, N., & Martín, J. A. R. (2012). Multiscale analysis of heavy metal contents in soils: Spatial variability in the Duero river basin (Spain). *Geoderma*, 189–190, 554–562. <https://doi.org/10.1016/j.geoderma.2012.06.006>
- Norris, G., & Duvall, R. (2006). EPA Positive Matrix Factorization(PMF) 5.0 fundamentals and user guide. Resource document. U.S. Environmental Protection Agency. <https://www.epa.gov/air-research/epa-positive-matrix-factorization-50-fundamentals-and-user-guide>. Accessed in Apr 2014.
- Ogunlaja, A., Ogunlaja, O. O., Okewole, D. M., & Morenkeji, O. A. (2019). Risk assessment and source identification of heavy metal contamination by multivariate and hazard index analyses of a pipeline vandalised area in Lagos State Nigeria. *Science of The Total Environment*, 651(2), 2943–2952. <https://doi.org/10.1016/j.scitotenv.2018.09.386>
- Pekey, H., & Dogan, G. (2013). Application of positive matrix factorization for the source apportionment of heavy metals in sediments: A comparison with a previous factor analysis study. *Microchemical Journal*, 106, 233–237. <https://doi.org/10.1016/j.microc.2012.07.007>
- Pirsaheb, M., Almasi, A., Sharafi, K., Jabari, Y., & Haghghi, S. (2016). A comparative study of heavy metals concentration of surface soils at metropolis squares with high traffic: A case study: Kermanshah Iran. *Acta Medica Mediterranea*, 32, 891–897.
- Proshad, R., Kormoker, T., & Islam, S. (2019). *Distribution, source identification, ecological and health risks of heavy metals in surface sediments of the Rupsa River*. Toxin Reviews. <https://doi.org/10.1080/15569543.2018.1564143>
- Qu, M. K., Li, W. D., Zhang, C. R., Wang, S. Q., Yang, Y., & He, L. Y. (2013). Source apportionment of heavy metals in soils using multivariate statistics and geostatistics. *Pedosphere*, 23, 437–444. [https://doi.org/10.1016/S1002-0160\(13\)60036-3](https://doi.org/10.1016/S1002-0160(13)60036-3)
- Saby, N. P. A., Arrouays, D., Boulonne, L., Jolivet, C., & Pochot, A. (2006). Geostatistical assessment of Pb in soil around Paris France. *Science of the Total Environment*, 367(1), 212–221. <https://doi.org/10.1016/j.scitotenv.2005.11.028>
- Schneider, A. R., Morvan, X., Saby, N. P. A., Cancès, B., Ponthieu, M., Gommeaux, M., et al. (2016). Multivariate spatial analyses of the distribution and origin of trace and major elements in soils surrounding a secondary lead smelter. *Environmental Science and Pollution Research*, 23(15), 15164–15174. <https://doi.org/10.1007/s11356-016-6624-2>
- Shao, S., Hu, B. F., Fu, Z. Y., Wang, J. Y., Lou, G., Zhou, Y., et al. (2018). Source Identification and Apportionment of Trace Elements in Soils in the Yangtze River Delta, China. *International Journal of Environment Research and Public Health*, 15, 1240. <https://doi.org/10.3390/ijerph15061240>
- Song, H. Y., Hu, K. L., An, Y., Chen, C., & Li, G. D. (2018). Spatial distribution and source apportionment of the heavy metals in the agricultural soil in a regional scale. *Journal of Soils and Sediments*, 18, 852–862. <https://doi.org/10.1007/s11368-017-1795-0>
- Sungur, A., Soylak, M., & Ozkan, H. (2019). Fractionation, Source Identification and Risk Assessments for Heavy Metals in Soils near a Small-Scale Industrial Area (Çanakkale-Turkey). *Soil and Sediment Contamination: An International Journal*, 28(2), 213–227. <https://doi.org/10.1080/15320383.2018.1564735>
- Tian, S. H., Liang, T., Li, K. X., & Wang, L. Q. (2018). Source and path identification of metals pollution in a mining area by PMF and rare earth element patterns in road dust. *Science of the Total Environment*, 633, 958–966. <https://doi.org/10.1016/j.scitotenv.2018.03.227>
- Uría, A. F., Mateo, C. L., Roca, E., & Marcos, M. L. F. (2009). Source identification of heavy metals in pastureland by multivariate analysis in NW Spain. *Journal of Hazardous Materials*, 165(1–3), 1008–1015. <https://doi.org/10.1016/j.jhazmat.2008.10.118>
- Wang, B., Xia, D. S., Yu, Y., Chen, H., & Jia, J. (2018). Source apportionment of soil-contamination in Baotou City (North China) based on a combined magnetic and geochemical approach. *Science of the Total Environment*, 642, 95–104. <https://doi.org/10.1016/j.scitotenv.2018.06.050>
- Wang, S. F., Li, R. A., & Ye, Z. J. (2007). *Zhejiang soil, second edition*. Zhejiang Science and Technology Press.
- Wei, B. G., & Yang, L. S. (2010). A Review of heavy metal contaminations in urban soils, urban road dusts and agricultural soils from China. *Microchemical Journal*, 94(2), 99–107. <https://doi.org/10.1016/j.microc.2009.09.014>
- Xia, F., Hu, B. F., Shao, S., Xu, D. Y., Zhou, Y., Zhou, Y., et al. (2019). Improvement of Spatial Modeling of Cr, Pb, Cd, As and Ni in Soil Based on Portable X-ray Fluorescence (PXRF) and Geostatistics: A Case Study in East China. *International Journal of Environmental Research and Public Health*, 16, 2694. <https://doi.org/10.3390/ijerph16152694>
- Xie, W. S., Peng, C., Wang, H. T., & Chen, W. P. (2018). Bioaccessibility and source identification of heavy metals in agricultural soils contaminated by mining activities. *Environment Earth Sciences*, 77, 606. <https://doi.org/10.1007/s12665-018-7783-x>
- Yan, S. D. (2008). Source apportionment and spatial heterogeneity of agricultural non-point source pollution in Huangshi, Hubei Province. *Transaction of the Chinese Society of Agricultural Engineering*, 24, 225–228. (in Chinese).
- Zhang, X. W., Wei, S., Sun, Q. Q., Syed, A. W., & Guo, B. L. (2018). Source identification and spatial distribution of arsenic and heavy metals in agricultural soil around Hunan industrial estate by positive matrix factorization model, principle components analysis and geo statistical analysis. *Ecotoxicology and Environmental Safety*, 159, 354–362. <https://doi.org/10.1016/j.ecoenv.2018.04.072>
- Zhi, Y. Y., Li, P., Shi, J. C., Zeng, L. Z., & Wu, L. S. (2016). Source identification and apportionment of soil cadmium in cropland of Eastern China: A combined approach of models and geographic information system. *Journal of*

- Soils and Sediments*, 16, 467–475. <https://doi.org/10.1007/s11368-015-1263-7>
- Zhu, G. X., Guo, Q. J., Xiao, H. Y., Chen, T. B., & Yang, J. (2017). Multivariate statistical and lead isotopic analyses approach to identify heavy metal sources in topsoil from the industrial zone of Beijing Capital Iron and Steel Factory. *Environmental Science and Pollution Research*, 24, 14877–14888. <https://doi.org/10.1007/s11356-017-9055-9>
- Zhu, M. J., Tang, L., & Liu, D. Q. (2015). Summary of Monitoring and Evaluation of Soil Heavy Metals along the Traffic Road. *China Environment Monitor*, 31(03), 84–91. (in Chinese).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.