Heliyon 9 (2023) e17246

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Source analysis and health risk assessment of heavy metals in agricultural land of multi-mineral mining and smelting area in the Karst region – a case study of Jichangpo Town, Southwest China

Zaiju Jiang [a], Shaozhang Yang [a,b,*], Sha Luo [a,b]

[a] Guizhou Coal Mine Geological Engineering Advisory and Geological Environment Monitoring Center, Guiyang, 550081, China
[b] Guizhou Rongyuan Environmental Protection Technology Co. LTD, Guiyang, 550081, China

ARTICLE INFO

Keywords:
Heavy metals in agricultural land
Self-organizing mapping (SOM)
Random forest (RF)
Priority-control pollutant source

ABSTRACT

In the Karst region of Southwest China, the content of soil heavy metals is generally high because of the geological background. Moreover, Southwest China is rich in mineral resources. A large number of mining and smelting activities discharge heavy metals into surrounding soil and cause superimposed pollution, which has drawn widespread concern. Due to the large variation coefficients of soil heavy metals in the Karst region, it is particularly essential to select appropriate analysis methods. In this paper, Jichangpo in Puding County, a Karst area with multi-mineral mining and smelting, is selected as the research object. A total of 368 pieces of agricultural topsoil in the study area are collected. The pollution level of heavy metals in agricultural soil is evaluated by the geological accumulation index (I_{geo}) and enrichment factor (EF). Absolute Factor Score/Multiple Linear Regression (APCS/MLR), geographic information system (GIS), self-organizing mapping (SOM), and random forest (RF) are used for the source allocation of soil heavy metals. Finally, the combination of APCS/MLR and health risk assessment model is adopted to evaluate the risks of heavy metal sources and determine the priority-control source. The results show that the average values of soil heavy metals in the study area (Cd, Hg, As, Pb, Cr, Cu, Zn, and Ni) exceed the background values of corresponding elements in Guizhou Province. Three sources of heavy metals are identified by combining APCS/MLR, GIS, SOM, and RF. Zn (63.47%), Pb (55.77%), Cd (58.98%), Hg (32.17%), Cu (14.41%), and As (5.99%) are related to lead-zinc mining and smelting; Cr (98.14%), Ni (90.64%), Cu (76.93%), Pb (43.02%), Zn (35.22%), Cd (28.97%), Hg (22.44%), and As (5.84%) are mixed sources (natural and agricultural sources); As (88.17%), Hg (45.39%), Cd (12.04%), Cu (8.66%), and Ni (6.72%) are related to the mining and smelting of coal and iron. The results of health risk assessment show that only As poses a non-carcinogenic risk to human health. 3.31% of the sampling points of As have non-carcinogenic risks to adults and 10.22% to children. In terms of carcinogenic risks, As, Pb, and Cr pose carcinogenic risks to adults and children. Combined with APCS/MLR and the health risk assessment model, the mining and smelting of coal and iron is the priority-control pollution source. This paper provides a comprehensive method for studying the distribution of heavy metal sources in areas with large variation coefficients of soil heavy metals in the Karst region. Furthermore, it offers a theoretical basis for the management and assessment of heavy metal pollution in agricultural land in the study area, which is helpful for researchers to make strategic decisions on food security when selecting agricultural land.

* Corresponding author. Guizhou Rongyuan Environmental Protection Technology Co. LTD, Guiyang, 550081, China.
E-mail addresses: jzj13885020165@163.com (Z. Jiang), ysz18785011721@163.com (S. Yang), ysz18785011721@163.com (S. Luo).

https://doi.org/10.1016/j.heliyon.2023.e17246
Received 21 February 2023; Received in revised form 10 June 2023; Accepted 12 June 2023
Available online 20 June 2023

1. Introduction

The soil pollution of arable land has become one of the most prominent environmental problems in China [1]. Heavy metals have polluted about one-fifth of China's arable land [2]. China bears 22% of the world's population upon 7% of the arable land. The arable land pollution in China threats the world's food security [3]. Heavy metals have been listed as the most severe pollutants on account of their high toxicity, bio-accumulation, and irreversible damage to humans [4]. Soil heavy metals enter the human body through the food chain and harm the human body. It arouses worries about the safety of agricultural products and widespread concerns about the heavy metals in cultivated soil [5–10]. Therefore, research on the accumulation degree, pollutant source, and health risks to humans of heavy metals in agricultural soil will make a significant contribution to food security.

Karst area accounts for 12% of the global land area. As one of the three largest Karst areas, Southwest China is an essential part of global Karst areas [10,11]. The area of carbonate rocks in Guizhou Province in Southwest China is about $1.1 \times 105\ km^2$, accounting for 73% of the province's total area [12]and ranking first in China. The soil in the Karst area is characterized by poor, discontinuous, and thin solum, leading to a shortage of cultivated land and an increment in its value [13]. The soil in the Karst area has the typical features of high geochemical background with the accumulation of heavy metals [6]. Meanwhile, the ecological environment in the Karst area is very fragile and easily affected by human activities. The bearing capacity of soil in the Karst area for heavy metals is limited. Southwest China is abundant in mineral resources. A large number of mining and smelting activities discharge heavy metals into the environment, which can cause superimposed pollution by migrating to the surrounding soil [9,14,15]. In Southwest China, the variation coefficients of soil heavy metals are generally large [6,9,11]. Previous studies have focused on the enrichment mechanism and distribution characteristics [16], health risk assessment [1,17], and source analysis [1,17,18] of soil heavy metals in karst areas. The source analysis methods of heavy metals mainly focus on geographic information system (GIS) mapping and receptor models (principal components analysis (PCA), positive matrix factorization (PMF)\UNMIX model) [1,9,17–19]. GIS and PCA can merely identify the potential source types of heavy metals rather than the source contributions [19]. PMF has the advantages of applying non-negative constraints and solving factor values while considering the uncertainty of heavy metals. However, PMF is sensitive to outliers, which may cause the deviation of results from actual values [10,20]. UNMIX can easily distinguish the source with the greatest contribution but may ignore the source with the lowest contribution [21]. APCS/MLR uses regression analysis to quantify the contribution of factors based on factor analysis, possessing the advantages of easy operation and fast calculation [21,22]. Previous studies show that in terms of predicting sources of soil heavy metals, APCS/MLR has a higher accuracy than PMF and UNMIX models and can effectively process strongly mutated data [20].

However, the receptor models do not consider the spatial change of data and require data to fit in with the algorithms. They cannot analyze variables such as soil type, land-use type, and soil parent material which are also significant factors affecting the accumulation of soil heavy metals [20]. Meanwhile, machine learning methods can reveal the complex relationship between soil heavy metals and environmental conditions and obtain more realistic results. Therefore, machine learning models are practical for identifying sources of soil heavy metals [23]. At present, there are few studies on machine learning methods for identifying sources of soil heavy metals [24]. SOM is an unsupervised artificial neural network, which relies on the competition among neurons to optimize the network step by step with the competitive learning strategy. It maintains the topological structure of input space by using the neighbor relationship function. SOM can project complex high-dimensional data sets into low-dimensional space, visualize them onto a two-dimensional surface, group the input objects with similar properties into the same category, and retain the initial topological structure. It applies to multi-dimensional data [25]. SOM can reveal the local relationship between variables and classify data with large variation coefficients. Besides, SOM supports the technology of giving information pictures of data by using reference vectors, which can clearly disclose the interdependence of variables [26]. RF is a decision tree model based on integrated learning, proposed by Leo Breiman in 2001 [27]. Using the bagging method to form integrated learners for training, RF can avoid data over-fitting. It has a specific resistance to noise, high prediction accuracy, and excellent performance in classification and regression [28].

We selected Jichangpo Town, Puding County, Guizhou Province, as the research object. Previous studies have shown that heavy metals such as Cd, Pb, Cu, and Cr in the soil of Puding County are enriched to varying degrees [29]. Jichangpo Town possesses various minerals, such as lead-zinc, coal, and iron. Multiple mines aggravate the pollution of soil heavy metals. However, the accumulation degree, source, and contribution rate of soil heavy metals in the study area are still unknown. Therefore, it is selected as the case of this study. Methods can be mutually verified and supplemented. This study combined traditional analysis methods (APCS/MLR and GIS) with machine learning methods (SOM and RF) to determine the contributions of heavy metal sources. The APCS/MLR and health risk assessment model are combined to assess the health risk of heavy metal sources. This study provides a comprehensive approach to the sources of heavy metals in karst areas. At the same time, our study is of great significance for the safe use of agricultural land and the protection of human health in karst areas with high heavy metal background values.

2. Materials and methods

2.1. Study area

Located in the central of Guizhou Province, Puding County is an agricultural county in the mountainous area with few farmlands per capita. A large amount of land cannot be cultivated. The distribution area of carbonate rock in the county accounts for 84% of the county's total area. Lead-zinc ores are abundant in Jichangpo Town and hosted in the Cambrian dolomite strata near the axis of the northeast Wuzhishan anticline in Jichangpo Town. Twelve ore spots have been found, and nine have been or are being mined with

mining rights. In addition, there are iron and coal mines in the study area. The distribution of mineral resources is shown in Fig. 1. The soil types in the study area mainly include yellow soil, yellow-brown soil, lime soil, paddy soil, and purple soil. The exposed strata are the Permian and Triassic strata (Fig. 1). The lithology is mostly limestone, dolomite, magmatite, and clastic rocks. The parent rock of soil formation is primarily carbonate rock.

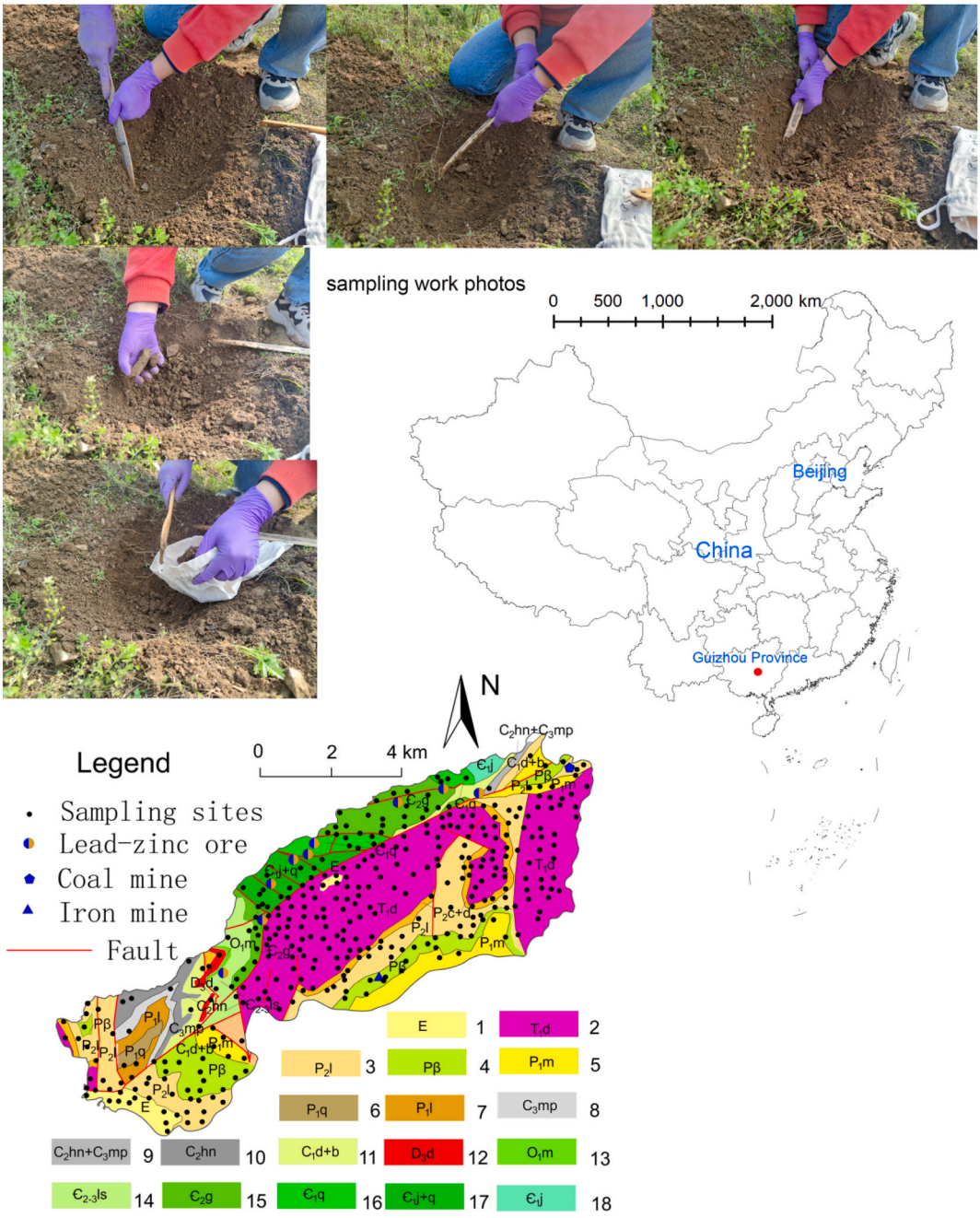


Fig. 1. Geological map and sampling map of the study area. 1. Paleogene, 2. Triassic Daye Formation limestone, 3. Permian Longtan Formation sandstone, 4. Permian Emeishan basalt formation, 5. Permian Maokou Formation, 6. Permian Qixia Formation, 7. Permian Liangshan Formation, 8. Carboniferous Mapping Group, 9. Carboniferous Mapping Group, Huanglong Group, 10. Carboniferous Huanglong Group, 11. Carboniferous Pazuo Group, Datang Group, 12. Devonian Daihua Group, 13. Ordovician Meitan Group, 14. Cambrian Loushanguan Group, 15. Cambrian Gaotai Formation, 16. Cambrian Qingxudong Formation, 17. Cambrian Qingxudong Formation, Jindingshan Formation, 18. Cambrian Jindingshan Formation.

2.2. Soil sample collection

The data in this paper was from the Geochemical Survey and Evaluation of Cultivated Land Quality in Guizhou Province. The sampling period was from January to February 2018. The total number of samples was 368.

The sampling used the land-use status map (1:50,000) with the second land survey data in Guizhou Province and the Ovital Map as working base maps, according to the "People's Republic of China Geology and Mineral Industry Standard Land Quality and Geochemical Evaluation Specification" (DZ/T 0295-2016) to design sampling points and determine the sampling depth. The density of sampling points ranged from 4 to 16 points/$km^{2.}$ Each sample was composed of a central sample and 4-6 sub-samples. Most sample points were arranged in an "X" shape, and a few ones were in the shape of an "S." Then the sampling points were input into the OvitalMap on a mobile phone. Sampling points were located according to the navigation service of the OvitalMap. The on-site sampling points could be adjusted within 100 m. Sampling points were set in the center of the cultivated land, the dominant land use type, the cultivated land for dominant crops, and areas with large soil thickness, with a sampling depth of 0–20 cm. In order to avoid the introduction of heavy metal elements before sampling, the soil was first dug with a shovel. The part of the soil that touched the shovel was removed with bamboo slices, and then samples were collected. During sampling, stones, insects, other debris and plant roots were removed. Live and dead roots were separated by visual inspection. Live roots were distinguished by their light color and high tensile strength. Roots were classified as dead when they became blackened and shriveled. Sub-samples and center samples were combined into one surface soil sample in equal proportion. The shovel and bamboo chips were cleaned up before collecting the next sample. All collected samples were packed in sealed kraft paper bags to avoid pollution and transported immediately to the laboratory. The distribution of sampling points is shown in Fig. 1.

Duplicate samples accounted for 2% of the total number of samples, which were collected at the same point by another sampling group after 3–5 days with the same method.

2.3. Sample test method

After being transported to the laboratory, The samples were then dried in a constant temperature drying oven below 60 °C. The samples were passed through a 20-mesh nylon sieve to remove impurities, such as rocks, plants, and animal debris. The dried samples were thoroughly mixed. 50 g of soil samples were taken out, ground by a ceramic mortar, and then were passed through a 20-mesh nylon sieve. The soil samples were divided into four parts by the quartile method. One part was for the pH test. The other parts were ground again and then were passed through a 60-mesh nylon sieve for the testing of soil organic matter and soil total nitrogen. 80 g of samples were taken out and ground by a pollution-free planetary ball mill (QXQM-4, Changsha Tianchuang Powder Technology Co. LTD, China), which inside the ball mill is agate ball, and then were passed through a 200 nylon mesh sieve. Then they were divided into five parts and packed for chemical analysis.

5.0 g of samples were taken out, uniformly put into a low-pressure polyethylene plastic ring, and placed on a press under the pressure of 30 MPa for 10s. An X-ray fluorescence spectrometer (ARL PERFORMX PerformX4200, Thermo Fisher Scientific, USA) was used to determine the contents of Cr, Cu, Ni, Pb, Zn, and TP (total phosphorus). 0.1000 ± 0.0002 g of soil samples were put into a 25 mL colorimetric tube. 10 mL 1 + 1 aqua rega containing tartric acid (50 g/L) was injected into the tube. Then the tube was put in a water bath and boiled for 1 h. After it was taken out and cooled down, it was diluted to 25 mL with 4.0 mol/L hydrochloric acid. After the solution was clarified, 5 mL of the supernatant was taken out and put into a 25 mL test tube. 5 mL of the mixed solution of 50 g/L thiourea and 50 g/L ascorbic acid was injected into the tube. The tube was shaken well, then stewing for 30min. As and Hg were determined by a two-channel atomic fluorescence spectrometer (BAF-2000, Beijing BaODE Instrument Co. LTD, China); 0.0500 ± 0.0002 g of samples were placed into a polytetrafluoroethylene crucible and added in 10 mL of mixed acid ($HNO_3 + HClO_4 + HF$). The crucible was covered and placed on a temperature-control electric hot plate overnight. The next day, the crucible was steamed until it was nearly dry. Then it was added HCl to dilute the solution to 25 mL. Cd was determined by an inductively coupled plasma mass spectrometry (ICP-MS, NexION 1000G, PE-Perkinelmer, USA). TN (total nitrogen) was determined by the kjeldahl method (auto-kjeldahls apparatus, K9860, Jinan Haineng Instrument Co. LTD, China). Soil pH was determined by a pH meter (PHSJ-4F, Shanghai Lei Chi Instrument Co. LTD, China) electrode method (ISE). Soil organic matter was determined by the high-temperature external thermal potassium dichromate oxidation - volumetric method (VOL) (burette,50 mL, Tianjin Tianbo Glass Co., LTD., China). The sample tests were completed at Sichuan Provincial Bureau of Geology and Mineral Exploration and Development, China. The quality of sample analysis was monitored with the first-grade standard materials of GBW07401-GBW07408 and GBW07425-GBW07428 (GSS1-8, GSS11-14). The recycling rate of each element was restrained in the range of 80%–120%.

2.4. Data collection

The elevation of sampling points was recorded during sampling. Information about the mineral species, longitude, latitude, and mining conditions of the study area was collected. The shortest distance between the sampling point and the mine was calculated using ArcGIS 10.8. The data on land-use types and soil types came from the Natural Resources Bureau of Puding County. The land-use types include paddy fields, dry land, tea gardens, and fruit gardens. According to the "Soil Classification System in China" (1992), the soil types contain yellow soil, yellow-brown soil, lime soil, paddy soil, and purple soil. Details are shown in Table S1. The data on pedogenic parent rock was from the 1:200,000 geological map of 91 bitmap assistants. The pedogenic parent rock includes dolomite, limestone, magmatite, and clastic rocks.

2.5. Assessment method of heavy metal pollution level

We used the geological accumulation index (I_{geo}) and enrichment factor method (EF) to evaluate soil heavy metals and reveal the accumulation and pollution degree of heavy metals in agricultural soil in the study area.

(1) Geo-accumulation index

I_{geo} takes into account not only the impact of geological background but also the impact of human activities on the background value of soil heavy metals. The calculation formula is shown in equation (1):

$$I_{geo} = \log_2\left[\frac{C_n}{1.5 \times B_n}\right] \tag{1}$$

I_{geo} is the geological accumulation index; C_n is the measured value of the element n in the sample; B_n is the geochemical background value of the element n; 1.5 is the correction coefficient. The judgment standard for the cumulative grade of soil heavy metals is referred to Ref. [30]. Details are shown in Table S2.

(2) Enrichment factor (EF)

EF can distinguish artificial and natural sources of soil heavy metals [20]. The calculation formula is shown in equation (2):

$$EF = \frac{C_i/B_i}{C_r/B_r} \tag{2}$$

C_i is the measured value of heavy metal elements in the soil, B_i is the background value of the corresponding element, C_r is the measured value of the reference element, and B_r is the background value of the reference element. Because Mn is stable and abundant in the crust, it is chosen as the reference element [6]. The judgment standard of the enrichment grade of soil heavy metals consults the literature [20]. Details are shown in Table S2.

2.6. Absolute factor score-multiple linear regression (APCS/MLR)

Principal component analysis (PCA) is a multivariate statistical method, which uses the idea of dimension reduction to separate multiple indicators into several components with little information loss. It can qualitatively identify pollution sources but cannot quantify source contributions [22]. APCS/MLR model is developed on the basis of PCA. First, the normalized factor fraction of heavy metal concentration APCS is obtained by PCA analysis, then APCS is converted into the concentration contribution of each pollution source to each sample [31,32]. The concrete steps are as following:

(1) Standardize all element contents, and get the normalized factor fraction by PCA analysis. The calculation formula is shown in equation (3):

$$Z_{ij} = \frac{C_{ij} - C_i}{\sigma_i} \tag{3}$$

where Z_{ij} denotes the standardized concentration, C_{ij} denotes the measured concentration of the heavy metal element, C_i denotes the average value of the element i, and σ_i denotes the standard deviation of the element i.

(2) For all elements, introduce an artificial sample with a concentration of 0. Calculate the factor fraction of the artificial sample. The calculation formula is shown in equation (4):

$$(Z_0)_i = \frac{0 - C_i}{\sigma_i} = -\frac{C_i}{\sigma_i} \tag{4}$$

(3) The APCS of each element is obtained by subtracting the factor fraction of the sample with 0 concentration from the factor fraction of each sample.

(4) Use the elemental concentration to perform multiple linear regression on APCS. The obtained regression coefficient converts APCS into the concentration contribution of each pollution source to each sample. The formula is shown in equation (5):

$$C_i = b_{0i} + \sum_{k=1}^{n} b_{ki} \times APCS_k \tag{5}$$

b_{0i} denotes the constant obtained from the multiple linear regression of the element i; b_{ki} denotes the regression coefficient of the source k to the element i; $APCS_k$ denotes the fraction of the adjusted factor k; $APCS_k \times b_{ki}$ represents the mass concentration contribution of the source k to C_i; The average value of $APCS_k \times b_{ki}$ of all samples represents the average absolute contribution of the source.

2.7. Self-organization mapping (SOM)

The SOM algorithm that comes with MATLAB 2018b was adopted for SOM analysis. The mechanism of SOM is similar to the human brain. It can self-organize neurons with the same weight vectors as input vectors and classify the spatial distribution of neurons with similar functions into clusters [33]. SOM is composed of input layers and output layers. By linking weight-connection, SOM forms a two-dimensional grid to cluster and visualize high-dimensional input data sets [34]. The data of soil heavy metals, soil pH, and soil organic matter is trained by the SOM algorithm after standardization (the mean value is 0, and the variance is 1). For the input data, the initialized network calculates the discriminant function value by comparing the distances between each neuron and the input. The specific neuron with the minimum discriminant function value is the winning neuron. After the winning neuron is activated, its weight is strengthened. The neurons in the topological neighborhood of the winning neuron will be activated to different degrees. The weight will be updated and shifted. This process keeps iterating until convergence. M represents the number of SOM neurons in the competitive layer, which is determined by formula (6) [35]:

$$M = 5\sqrt{N} \tag{6}$$

N denotes the total number of samples of the input layer.

We tested different map scales from 90 to 121 and compared QE and TE containing the output data of different neuron combinations. When the number of neurons in the SOM map was 110 (11 × 10), QE and TE were the smallest, which were 0.188 and 0.017, respectively. Therefore, 11 × 10 was selected as the number of neurons.

The results of SOM were further processed by K-means. It is an iterative clustering algorithm to divide data space into k clusters. Each sample was classified to the nearest group according to distance. The optional cluster was decided according to the minimum Davies-Bouldin index [26]. Finally, Two clusters were selected.

2.8. Random forest

Random forest is based on binary segmentation data to solve the problems of classification and regression. For classification, the Gini coefficient is used for data segmentation. In regard to regression, the weighted mean is adopted for training [36]. Training samples and characteristic variables are randomly selected from original samples with returns through the bootstrap resampling technology and generate some training data subsets with the same scales but containing different samples. Unselected data forms out-of-bag data sets for testing. As weak classifiers, the trained decision trees of the corresponding number constitute a random forest. The final result is determined by voting on the results of each decision tree in the random forest. The random forest model has few parameters to set and no requirement for data type and distribution. In the training process, it is easy to adjust and insensitive to outliers. It can not only prevent data from over-fitting but also model all covariant elements and determine the importance of each element. This study used % Inc MSE (mean square error increase) to measure variable importance. The greater the parameter, the more important the variable. The parameter may be negative, which indicates that the variable is not helpful for prediction [36,37].

We used the Random Forest toolkit in MATLAB 2018b to complete the training. Taking the contents of heavy metals as dependent variables and the natural and human factors as independent variables, 368 topsoil samples of cultivated land in the study area were divided into training sets and verification sets by a proportion of 7:3. After repeated tests, when the number of decision trees was ntree = 100, the number of prediction variables selected by each node was mtry = 3, and the minimum data value of each node was nodesize = 5, the fitting effect was the best. The overall fitting accuracy R^2 was Cd: 0.73, Hg: 0.79, As: 0.82, Pb: 0.84, Cr: 0.86, Cu: 0.93, Zn: 0.83, and Ni: 0.90, implying high prediction accuracy [38].

2.9. Human health risk assessment

Soil heavy metals enter the human body mainly through ways such as respiration, skin, and hand-to-mouth. Heavy metals will cause adverse effects on human health if they accumulate to a certain amount after long-term ingestion. The risks posed by heavy metals to human health include non-carcinogenic and carcinogenic risks. Hazard quotient (HQ) and cancer risk (CR) can quantitatively characterize non-carcinogenic and carcinogenic risks to humans that are exposed to heavy metals [39–41]. In this paper, the US EPA health risk assessment method is used to calculate the average daily exposure (ADD, $mg \cdot kg^{-1}\ d^{-1}$) under three exposure paths:

$$ADD_{ing} = \frac{c \times IngR \times EF \times ED \times CF}{BW \times AT} \tag{7}$$

$$ADD_{derm} = \frac{c \times SA \times AF \times ABS \times EF \times ED \times EF}{BW \times AT} \tag{8}$$

$$ADD_{inh} = \frac{c \times \mathrm{InhR} \times \mathrm{EF} \times \mathrm{ED}}{\mathrm{PEF} \times \mathrm{BW} \times \mathrm{AT}} \tag{9}$$

The meanings and values of above parameters are shown in Table S3.

According to the data of EPA Integrated Risk Information System (IRIS), Cd, As, Pb, Cr, and Ni pose both non-carcinogenic and carcinogenic risks, while Hg, Cu, and Zn pose only non-carcinogenic risks. Equations 10 and 11 are the calculation of single hazard quotient (HQ) and total hazard quotient (THQ) of heavy metals, and equations 12–14 are the calculation of single carcinogenic risk index (CR) and total carcinogenic risk index (TCR) of heavy metals.

$$\mathrm{HQ} = ADD/RfD \tag{10}$$

$$\mathrm{THQ} = \sum HQ_i \tag{11}$$

$$\mathrm{CR} = ADD \times SF \tag{12}$$

$$\mathrm{TCR} = \sum CR_i \tag{13}$$

The meanings and values of above parameters are shown in Table S4.

When THQ < 1, it indicates that the risk is small or negligible; when THQ > 1, it indicates a potential risk. The higher the HQ value, the higher the risk to the human body. When TCR $< 1 \times 10^{-6}$, it indicates that the risk to human health is slight; when TCR $1 \times 10^{-6} <$ TCR $< 1 \times 10^{-4}$, it indicates a risk; when TCR $> 1 \times 10^{-4}$, it indicates a significant risk [26].

2.10. Human health risk assessment based on source

By combining health risk assessment model with APCS/MLR, the carcinogenic and non-carcinogenic risks of heavy metal sources can be quantified. First, the quality contributions of heavy metal sources are calculated through the APCS/MLR model, then the health risks of different sources are calculated through the health risk assessment model [26]. The calculation formula for the acceptable daily intake (ADD) of the heavy metal source j is shown in equations 14–16. The calculation formula for non-carcinogenic risks of heavy metal sources is shown in equations 17 and 18, and the calculation formula for carcinogenic risks is shown in equations 19 and 20.

$$ADD^{j}_{ki,ing} = \frac{C^{j}_{ki} \times \mathrm{IngR} \times \mathrm{EF} \times \mathrm{ED} \times \mathrm{CF}}{\mathrm{BW} \times \mathrm{AT}} \tag{14}$$

$$ADD^{j}_{ki,derm} = \frac{C^{j}_{ki} \times \mathrm{SA} \times \mathrm{AF} \times \mathrm{ABS} \times \mathrm{EF} \times \mathrm{ED} \times \mathrm{EF}}{\mathrm{BW} \times \mathrm{AT}} \tag{15}$$

$$ADD^{j}_{kiinh} = \frac{C^{j}_{ki} \times \mathrm{InhR} \times \mathrm{EF} \times \mathrm{ED}}{\mathrm{PEF} \times \mathrm{BW} \times \mathrm{AT}} \tag{16}$$

$$HQ^{j}_{ki,t} = \frac{ADI^{j}_{ki,t}}{RfD_t} \tag{17}$$

$$\mathrm{THQ_j} = \sum HQ^{j}_{ki,t} \tag{18}$$

$$CR^{j}_{ki} = ADD^{j}_{ki,t} \times SF \tag{19}$$

$$TCR_j = \sum CR^{j}_{ki} \tag{20}$$

where c^{j}_{ki} denotes the mass contribution (mg/kg) of the heavy metal i from the source j in the sample k calculated by the APCS/MLR model; $\mathrm{THQ_j}$ denotes the non-carcinogenic risk of the heavy metal source j, and TCRj denotes the carcinogenic risk of the source j. Other parameters and values are the same as equations 7–13.

2.11. Data statistics and analysis

SPSS 22.0 (IBM Inc., USA) was used to complete the relevant calculations of APCS/MLR, and ArcGIS10.8 was used to draw the distribution map of sampling points and HMs contents. The SOM algorithm that comes with MATLAB 2018b was adopted for SOM analysis. The RF was trained in MATLAB2018b, and open-source code made available by Andrej Karpathy [42] of Stanford University was as a basis from which we created the full MATBAL implementation. other data was drawn by OriginPro (version 9.1, American Origin Laboratory Company).

3. Results and analysis

3.1. Analysis of descriptive statistical results of heavy metal contents in the soil

The test results of heavy metal concentration, soil pH, and soil organic matter in agricultural soil in the study area are shown in Table 1. The range of soil pH is 4.54–8.46, with an average value of 6.48. The soil is acidic. The concentration ranges of Cd, Hg, As, Pb, Cr, Cu, Zn, and Ni in soil are 0.11–6.25 mg/kg, 0.02–2.35 mg/kg, 2.23–1209.95 mg/kg, 5.58–843 mg/kg, 75.5–376 mg/kg, 12.1–355 mg/kg, 47.9–2455 mg/kg, and 10–161 mg/kg, respectively. According to the Screening Value of Soil Pollution Risk in Agricultural Land (GB15618-2018), except for Hg, Pb and Ni, the average values of other elements all exceed risk screening values. Referring to the background values of corresponding elements in Guizhou Province [43], the sampling points of Cd, Hg, As, Pb, Cr, Cu, Zn, and Ni in soil that exceed background values are 22.93%, 46.98%, 50.83%, 30.94%, 94.75%, 90.33%, 93.65%, 90.33%, and 75.14%, respectively. The average values of heavy metals all exceed the background values of corresponding elements, indicating that the soil heavy metals in the study area have accumulated to different levels. Variation coefficients reflect the variability and dispersion of soil elements. Elements with high variation coefficients (CV) may be affected by human activities [44]. The variation coefficients of Cd, Hg, As, Pb, and Zn are all greater than 1, indicating that these elements may be affected by human activities. While the variation coefficients of Cr and Ni are less than 1, which are 0.3 and 0.4, respectively, lower than other heavy metals in the study area. They are also strong variations [6] with high spatial heterogeneity. It may be due to the impacts of parent rocks and the process of soil formation [1,45].

3.2. Analysis of soil heavy metal accumulation

In order to clarify the enrichment degree of soil heavy metals in the study area and the impact of human activities on the enrichment, we calculated I_{geo} (Fig. 2(a)) and EF (Fig. 2(b)) of soil heavy metals in the study area. The results of I_{geo} show that for 98.89% of the sampling points of Cr and Ni, $I_{geo} \leq 1$, which means that Ni and Cr are mainly non-moderate pollution; for 72.65% of the sampling points of Cu, $I_{geo} \leq 1$, non-moderate pollution; for 22.38% of the sampling points of Cu, $1 < I_{geo} \leq 2$, moderate pollution, and for 4.97% of the sampling points of Cu, $2 < I_{geo} \leq 3$, moderate-severe pollution. For Cd, 4.4% of sampling points are moderate-severe pollution, and 95.6% are non-moderate pollution. For As, 19.89% of sampling points have $I_{geo} > 1$, meaning that they are at the level of moderate pollution or above; 0.55% of sampling points even have $4 < I_{geo} \leq 5$, meaning that they are at the level of severe or extremely severe pollution. For Hg, 17.40% of sampling points have $I_{geo} > 1$, meaning that they are moderately polluted or above; 1.38% have $I_{geo} \leq 4$, meaning that they are seriously polluted. For Pb and Zn, the proportion of sampling points having $I_{geo} > 1$ is 14.64%, meaning that they are at or above a moderate level of pollution; the proportion of samples with moderate-severe pollution is 10.77%; the proportion of sampling points with severe pollution is 3.59%; and the proportion of sampling points with severe and extremely severe pollution is 0.28%. The above analysis shows that point source pollution exists in As, Hg, Pb, Zn, Cd, and Cu. As in the soil of the study area is at the most severe pollution level, and samples are extremely polluted. It is followed by Pb and Zn, of which samples are heavily or extremely heavily polluted; followed by Hg and Cu, of which samples are heavily polluted; followed by Cd and Cu, of which samples are moderately or heavily polluted. 98.89% of the sampling points of Ni and Cr have $I_{geo} \leq 1$, meaning that these two elements mainly come from natural sources [46].

The calculation results of EF show that the proportions of elements at a moderate enrichment level and above are relatively small, which are As: 18.23%, Cd: 4.42%, Cr: 13.54%, Cu: 27.9%, Hg: 15.75%, Ni: 7.18%, Pb: 16.02%, and Zn: 19.33%. However, samples with extremely high concentrations of As and Hg in soil account for 0.55% and 0.28% of total samples, respectively; samples with strong enrichment levels of As, Hg, Pb, and Zn account for 1.66%, 0.83%, 0.28% and 0.28% of total samples, respectively; the proportions of elements at significant enrichment levels are: As: 4.97%, Cd: 0.28%, Cu: 6.91%, Hg: 4.7%, Pb: 5.52%, and Zn: 4.97%, indicating that point source pollution exists in As, Hg, Pb, Zn, and Cu.

From I_{geo} and EF, it can be concluded that human activities affect soil heavy metals in the study area to varying degrees. The effects on As, Hg, Pb, and Zn are the strongest, followed by Cd and Cu. Cr and Ni are hardly affected.

Table 1
Statistic characteristics of HM concentrations and pH in the soil.

	pH	Cd (gm/kg)	Hg (gm/kg)	As (gm/kg)	Pb (gm/kg)	Cr (gm/kg)	Cu (gm/kg)	Zn (gm/kg)	Ni (gm/kg)	Soil organic matter (g/kg)
Max	8.46	6.25	2.35	1209.95	843.00	376.00	355.00	2455.00	161.00	188.66
Mini	4.54	0.11	0.02	2.23	5.58	75.50	12.10	47.00	10.00	8.79
mean	6.48	0.64	0.24	53.40	73.01	159.40	99.51	267.14	59.78	45.63
CV	0.14	1.05	1.26	2.29	1.66	0.30	0.56	1.34	0.40	0.62
[a]background value	–	0.66	0.11	20.00	35.20	95.90	32.00	99.50	39.10	
[b]Risk screening value (GB 15618-2018)	≤5.50	0.30	1.30	40.00	70.00	150.00	50.00	200.00	60.00	
	5.5–6.50	0.30	1.80	40.00	90.00	150.00	50.00	200.00	70.00	
	6.5–7.50	0.30	2.40	30.00	120.00	150.00	100.00	250.00	100.00	
	>7.50	0.60	3.40	25.00	170.00	200.00	100.00	300.00	190.00	

[a] BV means background values of heavy metal (loid)s in the soil of Guizhou Province, which were obtained from CNEMC (1990) [45].
[b] The National Environmental Quality Standards for soil in China (GB15618-2018).

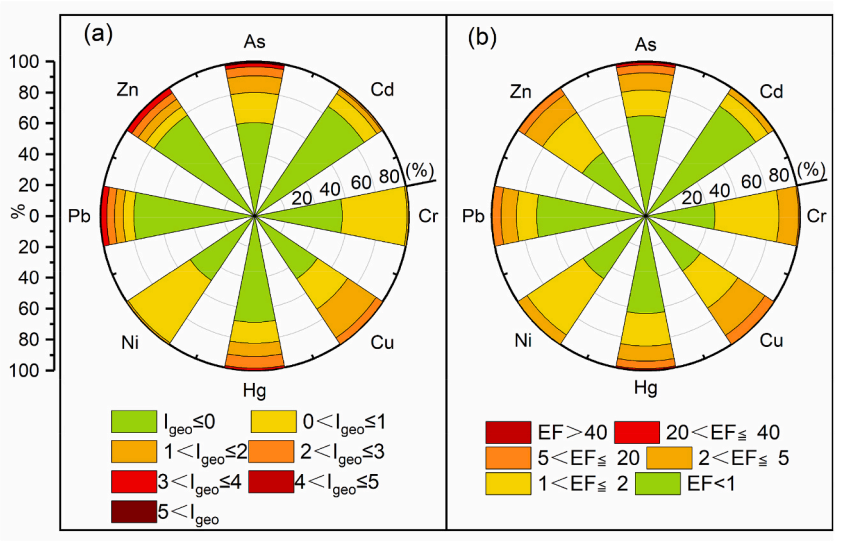


Fig. 2. Single Factor index (PI) and Potential Ecological Risk (Ei) grade statistical map of individual HMs in soil.

3.3. Source analysis of soil heavy metals

3.3.1. Correlation analysis

Heavy metal elements are positively correlated, indicating that they may have the same source [47]. The correlation analysis of soil heavy metals, pH, and soil organic matter is shown in Fig. 3. As and Hg are significantly and positively correlated ($P \leq 0.01$) with a correlation coefficient of 0.57. Hg and Cd are significantly and positively correlated ($P \leq 0.01$), with a correlation coefficient of 0.48. Zn is positively correlated with Cd and Pb ($P \leq 0.01$), and the correlation coefficients are 0.53 and 0.49, respectively. Cu is significantly and positively correlated with Ni and Cr ($P \leq 0.01$), and the correlation coefficients are 0.70 and 0.51, respectively; soil organic matter is significantly and positively correlated with Cr and Cu ($P \leq 0.01$), and the correlation coefficients are 0.35 and 0.3, respectively. Based on the above analysis, As, Hg, and Cd may come from the same source; Zn, Cd, and Pb may come from the same source; and Cu, Ni, and Cr may come from the same source. Soil organic matter represents soil fertility. The higher the content of organic matter is, the more fertilizer is applied. Elements with good correlation with the organic matter may be affected by agricultural activities [47]. Therefore, Cr and Cu may be affected by agricultural activities. The correlation between pH and heavy metals is weak, indicating that soil pH has little effect on the accumulation of soil heavy metals, which is consistent with the results of Qin et al. [1].

3.3.2. SOM analysis

The SOM diagram of soil heavy metals, soil organic matter, and pH is shown in Fig. 4. The color gradient variation of the SOM diagram represents the qualitative relationship between heavy metals. Yellow means high concentration. Black means low concentration. Each SOM matrix mapping represents a component value of 110 reference vectors. On the component plane, similar color

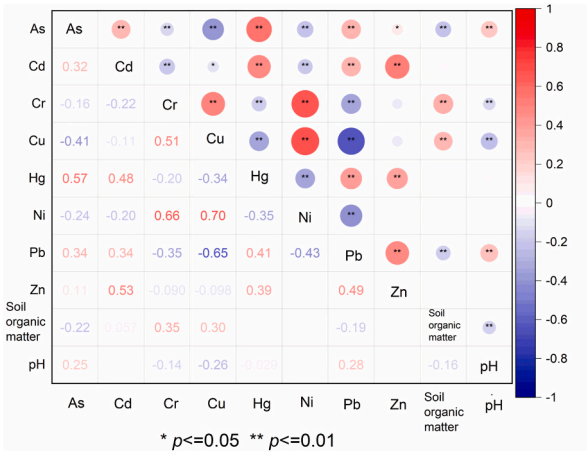


Fig. 3. Correlation analysis of soil heavy metals, soil organic matter, and soil pH.

gradients indicate a positive correlation between variables, and the sources of heavy metals may be the same [33]. It can be seen from Fig. 4 that the color gradients of Cu, Ni, and Cr are similar. The high concentration occupying neurons on the left indicates that these three elements have a strong positive correlation. The color change gradients of Pb, Zn, and Cd are similar. The high concentration appears in neurons on the upper right. The color change gradients of As and Hg are similar. The high concentration appears in the neurons in the upper figure.

The locations of soil samples were cluster analyzed by the K-mean algorithm and generated two different clusters (Fig. 5(a)). Cluster I includes 322 samples with the characteristic indicators of Cr, Cu, and Ni, mainly distributed in the northeast and southwest of the study area (Fig. 5(b)). The area has only one iron mine (Fig. 1), and the major human activity is agriculture. Cluster II is composed of 44 samples with the characteristic indicators of As, Cd, Hg, Pb, and Zn, distributed in the northeast of the study area in the shape of a belt (Fig. 5(b)). The area is a distribution area of the mining and smelting of lead-zinc and coal. These elements may be affected by mining activities.

3.3.3. APCS/MLR analysis

In order to further identify and quantify the sources of heavy metals in agricultural soil, PCA was used to determine the sources of heavy metals, and APCS/MLR was used to quantify the contributions of pollution sources. The principal component analysis of the KMO test was 0.63. For the Bartlett spherical test statistic, sig = 0 (0 < 0 01). Based on the principal component method, the correlation matrix was used for analysis. The extracted characteristic value was greater than 1. Three factors were extracted. The factor load matrix was orthogonally rotated by the maximum variance method, so that large factor loading was only concentrated on a few variables of each principal component, while the factor loading on remaining variables was 0 or close to 0. The total cumulative interpretation variance is 75.76% (Factor 1:40.09%, Factor 2:19.48%, Factor 3:16.20%) (Table 2), representing that these three principal components can interpret most of the information in heavy metal elements. The load values are classified into three types of "strong" load (>0.75), "moderate" load (0.75–0.50), and "weak" load (0.5–0.3) [48]. It can be seen from Table 3 that elements Zn, Pb, and Cd have strong loads in Factor 1, while Hg has a weak load. In Factor 2, Ni and Cr have strong loads, while Cu has a moderate load. As has a strong load in Factor 3, while Hg has a moderate load.

The calculation results of APCS/MLR model show that the correlation coefficients of the regression equation of As, Cd, Cr, Cu, Hg, Ni, Pb, and Zn (R^2) are 0.87, 0.65, 0.66, 0.56, 0.80, 0.80, 0.86, and 0.86, respectively, indicating that the APCS/MLR model has high accuracy [30]. According to the regression coefficients of multiple linear regression equation, the contributions of different sources of each heavy metal to the pollution level were calculated (Fig. 6).

From the average contribution rates of sources (Fig. 6(a)), the sources of Pb, Zn, and Cd are mainly from Factor 1, followed by Factor 2. The contribution rates of Factor 1 to Pb, Zn, and Cd are 55.77%, 63.47%, and 58.98%, respectively; the contribution rates of Factor 2 to Pb, Zn, and Cd are 43.02%, 35.22%, and 28.97%, respectively. Additionally, Factor 3 also contributes to Cd with a contribution rate of 12.04%. Factor 2 is the primary source of Cr, Ni, and Cu. The contribution rates of Factor 2 to Cr, Ni, and Cu are 98.14%, 90.64%, and 76.93%, respectively. Factor 1 also contributes to Cu with a contribution rate of 14.41%. The sources of As and Hg are mainly Factor 3. The contribution rates of Factor 3 to As and Hg are 88.17.9% and 45.39%. The contribution rates of Factor 1 and Factor 2 to Hg are 32.17% and 22.44%.

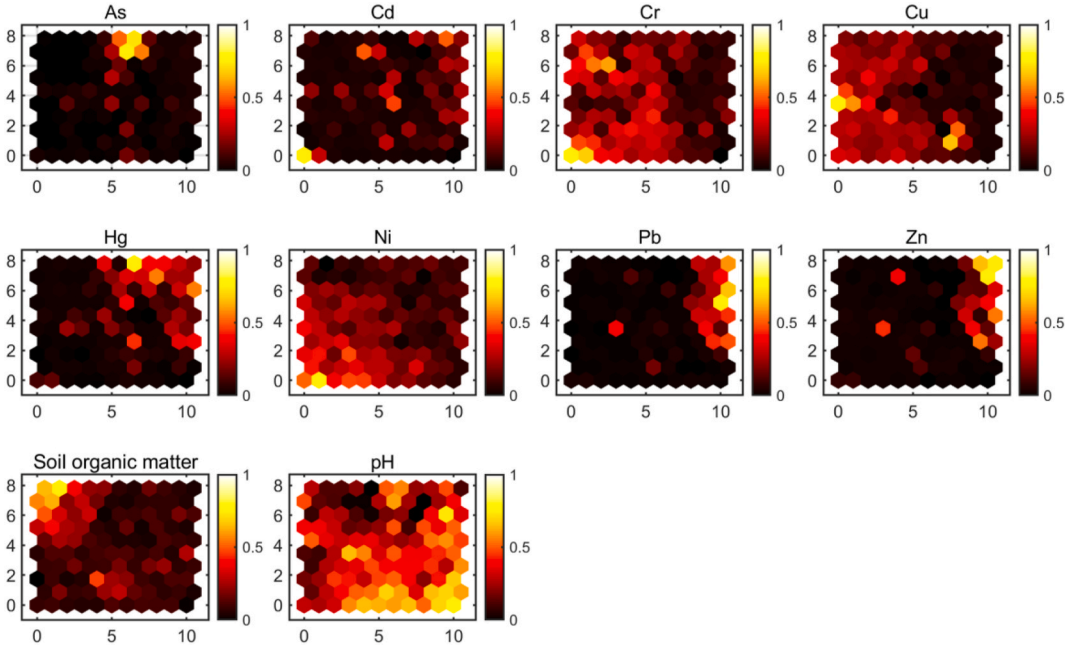
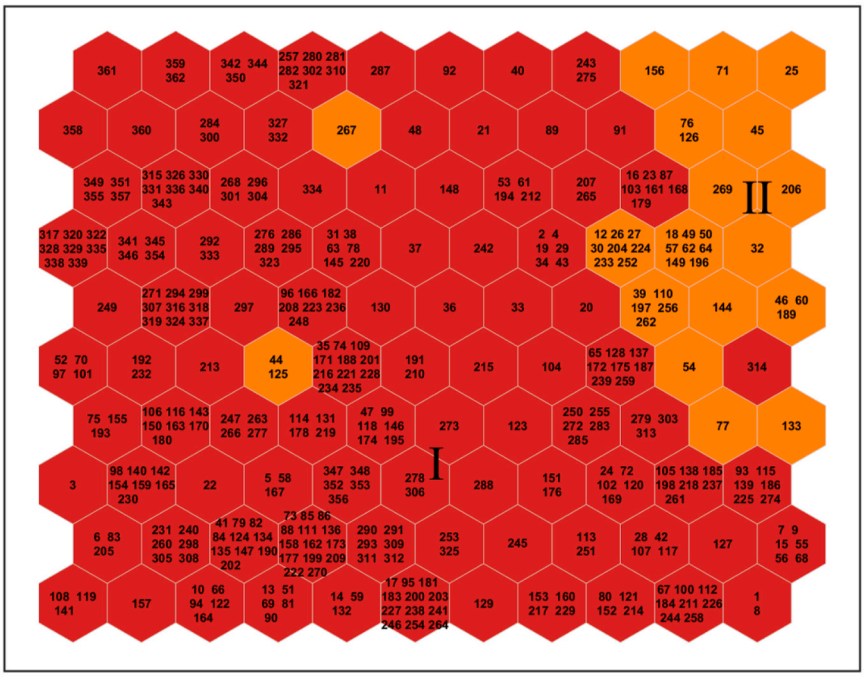


Fig. 4. SOM diagram of heavy metals concentration, Soil organic matter and pH.

(a)



(b)

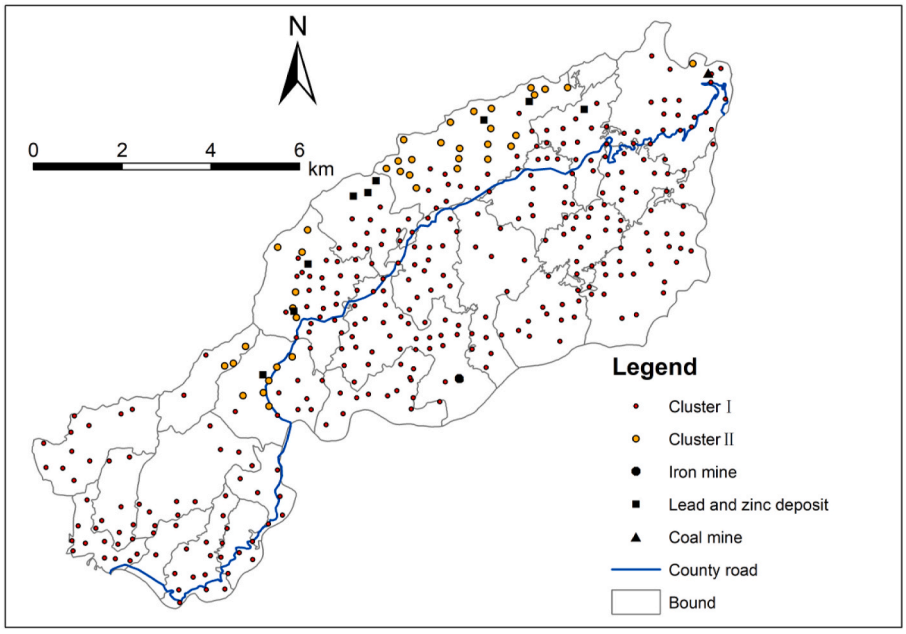


Fig. 5. Diagram of the pattern classification of two cluster groups based on SOM (a) and diagram of the distribution of the cluster pattern of sampling points in the study area (b).

Table 2
Cumulative contribution rates of variances and eigenvalues of the correlation matrix between soil elements.

Element	1	2	3	4	5	6	7	8
characteristic value	3.21	1.56	1.30	0.69	0.55	0.34	0.26	0.10
variance contribution rate %	40.09	19.489	16.20	8.59	6.91	4.30	3.20	1.24
cumulative contribution rate of variance %	40.09	59.57	75.76	84.35	91.26	95.56	98.76	100.00

Table 3
Factor matrix after maximum rotation of soil element variance.

Element	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
1	−0.10	0.76	−0.06	−0.22	0.46	−0.04	0.84	0.88
2	−0.06	0.22	0.81	0.70	−0.19	0.88	−0.39	−0.29
3	0.93	0.18	0.023	−0.15	0.74	−0.13	0.02	−0.02

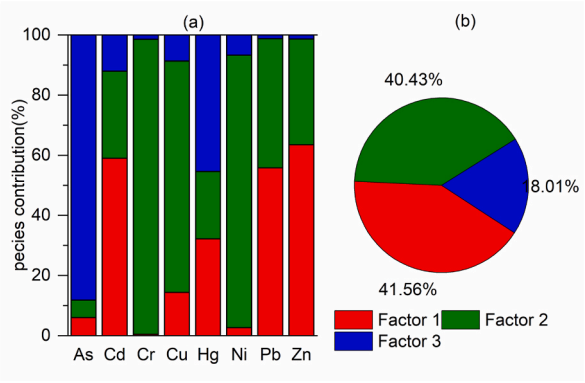


Fig. 6. Contribution rates of different sources to the contents of soil heavy metals.

3.3.4. RF analysis

The RF training results are shown in Fig. 7. %Inc MSE represents the weights of environmental influencing factors. The larger the weight value of the factor, the greater the contribution of this influencing factor to the accumulations of soil heavy metals in the study area. The distance to the coal mine (D2) is the most important factor interpreting As and Hg and the third most important factor interpreting Cd. The distance to the lead-zinc mine (D1) is an important factor interpreting Zn, the second most important factor interpreting Pb and Cd. The distance to the iron mine (D3) is an important factor interpreting Pb, and Zn. The soil-forming parent rock is the most important factor interpreting Cr and Zn, an important factor interpreting Cu, Hg, Ni, Pb, Zn and Cd. The elevation is the most important factor interpreting Cd, an important factor interpreting Pb and Zn. STP is the most important factor interpreting Cu and Ni, an important factor interpreting As and Cr. In short, the important factor interpreting As, Hg, Pb, and Zn is the distance to the mine; the important factors interpreting Cu, Ni, and Cr are the parent rock of soil and STP; the important factors interpreting Cd are the elevation, soil-forming parent rock, and the distance to the mine.

3.3.5. Source interpretation of multiple methods

According to the APCS/MLR model, the total contribution rate of Factor 1 is 41.56% (Fig. 6(b)). Factor 1 is mainly interpreted by Zn (63.47%), Pb (55.77%), and Cd (58.98%), followed by Hg (32.17%), Cu (14.41%), and As (5.99%) (Fig. 6(a)). The variation coefficients of Pb, Zn, and Cd are greater than 1. I_{geo} and EF manifest that serious point source pollution exists in these three elements, and they are affected by human activities. We used the ArcGIS ordinary Kriging interpolation method to figure out the spatial distribution characteristics of soil heavy metals. The spatial distribution graph of heavy metals shows that the high-value areas of Pb (Fig. S2 (d)) and Zn (Fig. S2 (c)) are distributed in belts in the northwest of the study area, which is also the sub-high-value distribution area of Cd with the distribution of nine lead-zinc mines. Correlation analysis and SOM algorithm imply that Pb, Zn, and Cd are homologous. After RF training, the distance to the lead-zinc mine, parent rock, and elevation are important factors interpreting Pb, Zn, and Cd. The sampling graph of the study area shows that the lead-zinc mine is distributed in the northwest of the study area, which has the highest elevation (Fig. S1). The stratum of this area is Cambrian, and the lithology is dolomite. The dolomite in the study area is also distributed. in this area. Therefore, it can be speculated that the effects of elevation and parent rock on the distributions of these three heavy metals are caused by lead-zinc mines. The above evidence testifies that Factor 1 mainly comes from lead-zinc smelting and mining. Moreover, previous research has confirmed that heavy metals, such as Pb, Cd, Cu, Zn, and As, are the major impurities in lead-zinc mines. During the smelting process of lead-zinc, these elements enter the surrounding soil with the dust discharged from lead-zinc

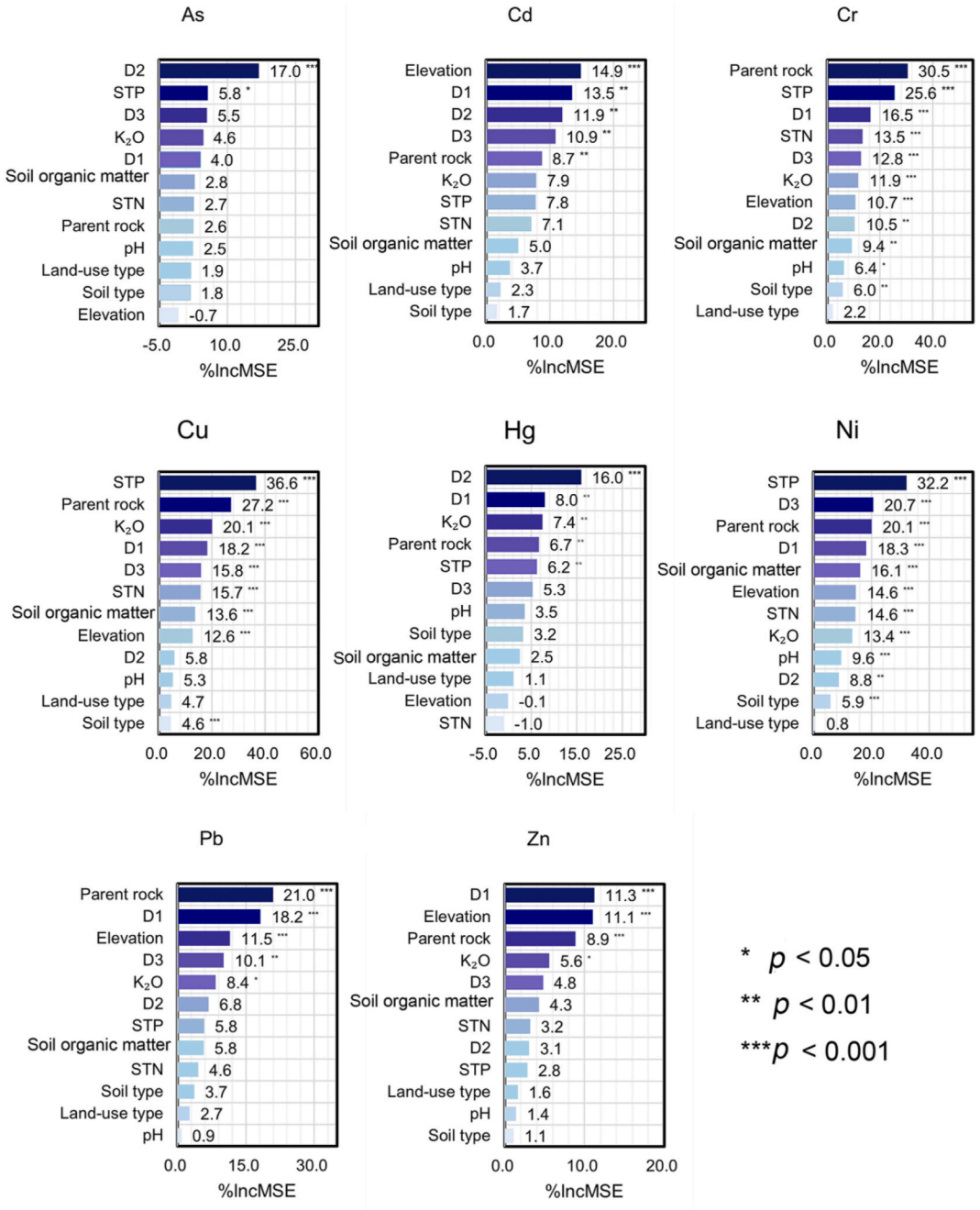


Fig. 7. Relative importance of variables obtained from the RF model (D1: The distance to the lead-zinc mine, D2:The distance to the coal mine, D3: The distance to the iron mine).

smelters, increasing the content of Pb, Cd, Cu, Zn, and As in the surrounding soil [44,49,50]. In addition, Pb is a symbol element of traffic sources. During driving, fuel combustion and the frictions of engines and tires will emit Pb, resulting in Pb accumulation in the surrounding soil [51,52]. However, some studies have shown that traffic contributes little to the Pb accumulation in the soil [19]. From the spatial distribution of heavy metal Pb (Fig. S2 (d)), no hot spot is found around the main road. Therefore, it is deduced that traffic contributes little to the Pb accumulation in the soil in the study area. So Factor 1 is defined as the source related to the mining and smelting of lead and zinc.

The total contribution rate of Factor 2 is 40.43% (Fig. 6(b)), which is mainly interpreted by Cr (98.14%), Ni (90.64%), and Cu (76.93%), followed by Pb (43.02%), Zn (35.22%), Cd (28.97%), Hg (22.44%), and As (5.84%) (Fig. 6(a)). Correlation analysis and SOM algorithm show that Cu, Ni, and Cr are homologous. Many studies have verified that Cu, Cr, and Ni in soil are natural sources

mainly from the parent rock of soil formation and are controlled by the soil weathering process [24,28]. From the spatial distribution map of heavy metals, it can be seen that the high concentration areas of Cr (Fig. S2 (f)), Ni (Fig. S2 (g)), and Cu (Fig. S2 (h)) are not located near mines. It indicates that these three elements are less affected by mining activities. The high-value areas of Cr and Ni are distributed in the Triassic System, Daye Formation, Permian System, Longtan Formation, Changxing Formation, etc. Parent rocks are limestone, sandstone, and shale. Previous research has shown that the heavy metals in sedimentary rocks are significantly affected by parent rocks and related to the compositions of parent rocks [53]. The high-value area of Cu is distributed in the Permian System Emeishan basalt formation. Studies show that the Emeishan basalt contains many copper minerals, leading to a high copper background value of the soil developed in the basalt [6,54]. According to the data on the network station of Puding County Natural Resources Bureau (http://www.aspd.gov.cn/zfbm/zrzyj/dwzz/index.html), copper mines have been detected in the Emeishan basalt in the study area. However, because of the intense variation in the grade and thickness, the copper mines currently have no mining value. The soil developed from copper-rich parent rocks is easy to enrich Cu [55]. Therefore, it can be concluded that Cu in the soil of the study area is affected by the copper mines in the Emeishan basalt. In addition, according to the investigation and interviews in the sampling process, the agricultural planting in the study area mainly applies compound fertilizer, nitrogen fertilizer, and farm manure. Cu, As, and Zn are added to animal feed, resulting in the existence of Cu, As, and Zn in animal excrement [28,56]. The contents of Cd and Pb in animal excrement are far higher than the standard values of organic fertilizer [6]. However, Cu and Zn are also contained in pesticides. The amount of Cu and Zn entering farmland with pesticides every year cannot be ignored. Furthermore, studies have shown that carbonate weathering is the main source of Cd in soil, and fertilizers and pesticides also contribute to the accumulation of Cd in agricultural soil [57,58]. Zhang et al. [59] confirmed that the main source of Cd in the farmlands of Puding County was agricultural activities, such as fertilization and pesticide spraying. Through the RF algorithm, the important factors interpreting Cu, Ni, and Cr are soil-forming parent rocks and STP, and K_2O is the third most important factor interpreting Cu. K_2O. STP in the soil is primarily affected by the properties of parent rocks and fertilization [60–62]. K_2O is an important product generated from the weathering of parent rocks during soil-forming [63]. Generally, the elements with good correlation with K_2O are natural sources. Based on the above analysis, Factor 2 is defined as the mixed source of natural and agricultural.

The total contribution rate of Factor 3 is 18.01% (Fig. 6(b)). As (88.17.9%) and Hg (45.39%) are the main pollution sources of this factor, followed by Cd (12.04%), Cu (8.66%), and Ni (6.72%) (Fig. 6(a)). Correlation analysis and SOM algorithm show that Hg and As are homologous, and the variation coefficients of Hg and As are greater than 1. I_{geo} and EF display that serious point source pollution exists in these two elements, indicating that they are affected by human activities. RF algorithm shows that the most important factor interpreting As and Hg is the distance to coal mines. From the spatial distribution map of As (Fig. S2 (b)), it can be seen that the high-value areas of As are distributed near coal and iron mines, while the high-value areas of Hg (Fig. S2 (a)) are distributed near the mines of lead and zinc in addition to coal and iron mines. Hg distributed near lead and zinc mines is interpreted in Factor 1. The content of As in coals in Guizhou is much higher than that in the United States and other countries [64]. Moreover, affected by the low-temperature hydrothermal solution, the content of Hg in coals in Guizhou is as high as 1.1 ppm, while the average content of Hg in coals in China is 0.19 ppm, and in the world it is only 0.1 ppm [65]. As and Hg in coals are released into the environment during the processes of mining, processing, and combustion [66,67]. Accidents of As poisoning caused by coal combustion occurred in some villages in Southwest Guizhou [68]. Previous studies have shown that iron mining leads to the accumulation of As, Hg, Cd, Cu, and other heavy metals in the surrounding soil to varying degrees [69,70]. Mining activities are an important contributor to the high enrichment of soil heavy metals in Karst areas of Southwest China [64]. Therefore, Factor 3 is defined as the source related to the mining and smelting of coal and iron.

3.4. Health risk assessment of soil heavy metals

This study adopted the health risk assessment model to evaluate the health risks of soil heavy metals in the study area. The results are shown in Tables S5 and S6. With regard to the non-carcinogenic risks to adults and children, the THQ of Cd, Hg, Pb, Cr, Cu, Zn, and Ni are all less than 1, representing that the non-carcinogenic risks to the human body can be ignored. The THQ of 3.31% of sampling points of As to adults is greater than 1, and the THQ of 10.22% of sampling points of As to children is greater than 1. Therefore, As is the dominant contributor to the THQ from soil heavy metals.

In terms of carcinogenic risks, the TCR of Cd and Ni for adults and children are all less than 10^{-6}, indicating that there are no carcinogenic risks. As, Pb, and Cr pose different degrees of carcinogenic risks to adults and children. Specifically, for adults, the TCR of 77.96% of As sampling points is in the range of 10^{-6}-10^{-4}. The TCR of 22.04% of sampling points of As is greater than 10^{-4}, meaning significant carcinogenic risks. The TCR of 15.98% of sampling points of Pb is in the range of 10^{-6}-10^{-4}, and the TCR of 0.55% of sampling points of Cr is in the range of 10^{-6}-10^{-4}. For children, the TCR of 53.44% of sampling points of As is in the range of 10^{-6}-10^{-4}; the TCR of 46.56% of sampling points of As is greater than 10^{-4}; the TCR of 24.79% of sampling points of Pb is in the range of 10^{-6}-10^{-4}, and the TCR of 51.79% of sampling points of Cr is in the range of 10^{-6}-10^{-4}. The non-carcinogenic and carcinogenic risks to children are all significantly higher than those to adults. These results are in accordance with previous studies [64], indicating that children are more vulnerable to the impact of soil heavy metals. It may be related to children's weak resistance and poor living habits, such as playing close to the ground, sucking fingers, etc. Therefore, children's hands and mouths should be kept clean, and their unhygienic habits, such as "eating hands," should be corrected [71].

3.5. Source-based health risk assessment

Health risk assessment is only conducive to understanding the level of health risks of heavy metals but cannot help people in effectively controlling health risk sources. Therefore, it is necessary to screen out the priority control factor in the pollution control of

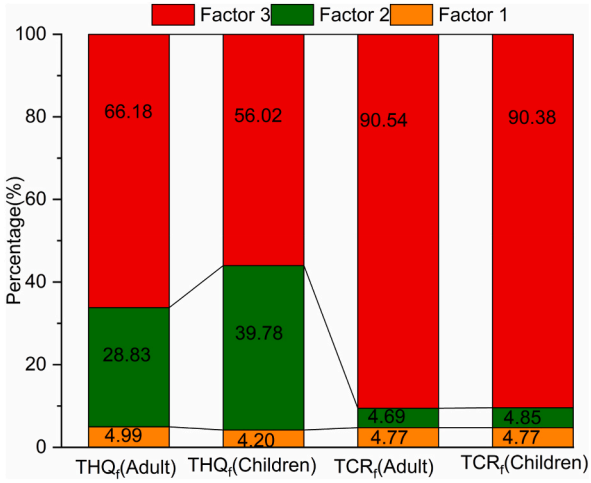


Fig. 8. Health risk assessment of soil heavy metal pollution sources.

soil heavy metals and identify and quantify the health risk of primary pollution sources [72]. In this paper, the contributions of different pollution sources to the content of soil heavy metals calculated by the APCS/MLR model are combined with the health risk assessment model to determine the non-carcinogenic and carcinogenic risks of heavy metal sources and their contribution rates (see Fig. 7). For adults and children, the THQj of three heavy metal sources is less than 1, indicating that non-carcinogenic risks can be ignored. For adults, Factor 3 poses the largest non-carcinogenic risks, contributing 66.18% to the total non-carcinogenic risks, followed by Factor 2, 28.83%, and Factor 1 poses the smallest contribution, 4.99% (Fig. 8). For children, the contribution rates of the three factors to the total non-carcinogenic risks are Factor 3 (56.02%) > Factor 2 (39.78%) > Factor 1 (4.20%). The range of TCRj of different heavy metal sources to adults is 5.03 E^{-06}-9-55 E^{-05}. The TCRj of the three factors is between 10^{-6} and 10^{-4}, indicating that the three heavy metal sources all pose carcinogenic risks. The range of TCRj of different heavy metal sources to children is 1.12 E^{-05}-2.13 E^{-04}. Factor 1 (1.12 E^{-05}) and Factor 2 (1.14 E^{-05}) pose carcinogenic risks. Factor 3 (2.13 E^{-04}) poses a significant carcinogenic risk. The three sources pose higher carcinogenic risks to children than adults. Therefore, special attention should be paid to preventing the carcinogenic risks of soil heavy metals to children. The carcinogenic and non-carcinogenic risks to adults and children from different sources of heavy metals have the same pattern, that is, Factor 3 > Factor 2 > Factor 1. For adults, Factor 3 : 90.54%, Factor 2: 4.69%, Factor 1: 4.77%; for children, Factor 3:90.38%, Factor 2:4.85%, Factor 1:4.77% (Fig. 8). Interestingly, the contribution rates of heavy metal sources are expressed as Factor 1 > Factor 2 > Factor 3, indicating that heavy metal sources with large contribution rates do not necessarily pose high health risks. It is due to the higher risks of As and Cr [73]. To sum up, Factor 3 (the source related to the mining and smelting of coal and iron) is the priority source of pollution control. To protect the environment and human health, the mining and smelting of coal and iron in the study area should be controlled, and normal mining industries must restrain heavy metal emissions.

4. Conclusion

In this study, except for Hg, Pb, and Ni, the average values of Cd, As, Cr, Cu, and Zn exceed the risk screening values in the "Risk Screening Values of Agricultural Land Pollution" (GB15618-2018). I_{geo} and EF show that point source pollution exists in As, Hg, Pb, Zn, and Cu. I_{geo} represents that point source pollution exists in Cd. Three sources of heavy metals are quantified in this study: Factor 1, the source of zinc mining and smelting, with a total contribution rate of 41.56% to the sources of heavy metals; Factor 2, the mixed source of natural and agricultural, with a total contribution rate of 40.43% to heavy metal sources; Factor 3, the source related to the mining and smelting of coal and iron, with a total contribution rate of 18.10% to heavy metal sources. The three sources of heavy metals are all of carcinogenic risks for adults and children. The source related to the mining and smelting of coal and iron poses an even more significant carcinogenic risk to children. Although Factor 3 has the lowest quality contribution rate, its health risk is the highest. Factor 3 (the mining and smelting of coal and iron) is the priority source of pollution control. This paper combines traditional source analysis methods (APCS/MLR and GIS) with machine learning methods (SOM and RF) to comprehensively consider the impacts of heavy metal concentrations, soil properties (pH, soil organic matter, K_2O, etc.), and environmental factors (elevation, distance to the mine, etc.) on the distribution and sources of heavy metals. However, the factors considered are still defective due to limited conditions. Future research should incorporate irrigation water, atmospheric sedimentation, population distribution, soil maturity, and other factors.

Author contribution statement

Shaozhang Yang: Conceived and designed the experiments.
Zaiju Jiang: Performed the experiments.
Sha Luo: Contributed reagents, materials, analysis tools or data.
Zaiju Jiang, Shaozhang Yang: Analyzed and interpreted the data; Wrote the paper.

Funding

This research was supported by Geochemical Survey and Evaluation of Cultivated Land Quality in Guizhou Province (Qian Geng Diao 2017-02).

Institutional review board statement

Not applicable.

Data availability statement

Not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgments

We would like to express appreciation to all of the participants from Division of Geochemical evaluation of cultivated land quality in guizhou province. At the some time, we would like to thank KetengEdit(www.ketengedit.com) for its linguistic assistance during the preparation of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e17246.

References

[1] Y. Qin, F. Zhang, S. Xue, T. Ma, L. Yu, Heavy metal pollution and source contributions in agricultural soils developed from karst landform in the southwestern region of China, Toxics 10 (10) (2022) 568.
[2] P. Liu, Z. Wu, X. Luo, M. Wen, L. Huang, B. Chen, et al., Pollution assessment and source analysis of heavy metals in acidic farmland of the karst region in southern China—a case study of Quanzhou County, Appl. Geochem. 123 (2020), 104764.
[3] G. Qin, Z. Niu, J. Yu, Z. Li, P. Xiang, Soil heavy metal pollution and food safety in China: effects, sources and removing technology, Chemosphere 267 (2021), 129205.
[4] N. Sarwar, M. Imran, M.R. Shaheen, W. Ishaque, M.A. Kamran, A. Matloob, et al., Phytoremediation strategies for soils contaminated with heavy metals: modifications and future perspectives, Chemosphere 171 (2017) 710–721.
[5] G. Luo, Z. Han, J. Xiong, Y. He, J. Liao, P. Wu, Heavy metal pollution and ecological risk assessment of tailings in the Qinglong Dachang antimony mine, China, Environ. Sci. Pollut. Control Ser. 28 (2021) 33491–33504.
[6] D.B. Senoro, C.E.F. Monjardin, E.G. Fetalvero, Z.E.C. Benjamin, A.F.B. Gorospe, K.L.M. de Jesus, et al., Quantitative assessment and spatial analysis of metals and metalloids in soil using the geo-accumulation index in the capital Town of Romblon province, Philippines, Toxics 10 (11) (2022) 633.
[7] X. Li, H. Liu, W. Meng, N. Liu, P. Wu, Accumulation and source apportionment of heavy metal (loid) s in agricultural soils based on GIS, SOM and PMF: a case study in superposition areas of geochemical anomalies and zinc smelting, Southwest China, Process Saf. Environ. Protect. 159 (2022) 964–977.
[8] E. Atikpo, E.S. Okonofua, N.O. Uwadia, A. Michael, Health risks connected with ingestion of vegetables harvested from heavy metals contaminated farms in Western Nigeria, Heliyon 7 (8) (2021), e07716.
[9] T.A. Laniyan, O.M. Morakinyo, Environmental sustainability and prevention of heavy metal pollution of some geo-materials within a city in southwestern Nigeria, Heliyon 7 (4) (2021), e06796.
[10] X. Li, T. Geng, W. Shen, J. Zhang, Y. Zhou, Quantifying the influencing factors and multi-factor interactions affecting cadmium accumulation in limestone-derived agricultural soil using random forest (RF) approach, Ecotoxicol. Environ. Saf. 209 (2021), 111773.
[11] Q. Yang, Z. Yang, Q. Zhang, X. Liu, X. Zhuo, T. Wu, et al., Ecological risk assessment of Cd and other heavy metals in soil-rice system in the karst areas with high geochemical background of Guangxi, China, Sci. China Earth Sci. 64 (7) (2021) 1126–1139.
[12] K. Luo, H. Liu, Q. Liu, Y. Tu, E. Yu, D. Xing, Cadmium accumulation and migration of 3 peppers varieties in yellow and limestone soils under geochemical anomaly, Environ. Technol. 43 (1) (2022) 10–20.
[13] B. Zhong, T. Liang, L. Wang, K. Li, Applications of stochastic models and geostatistical analyses to study sources and spatial patterns of soil heavy metals in a metalliferous industrial district of China, Sci. Total Environ. 490 (2014) 422–434.
[14] Q. Zhang, G. Han, M. Liu, T. Liang, Spatial distribution and controlling factors of heavy metals in soils from puding karst critical zone observatory, southwest China, Environ. Earth Sci. 78 (9) (2019).
[15] W. Pu, J. Sun, F. Zhang, X. Wen, W. Liu, C. Huang, Effects of copper mining on heavy metal contamination in a rice agrosystem in the Xiaojiang River Basin, southwest China, Acta Geochimica 38 (2019) 753–773.
[16] W. Ji, Y. Lu, M. Yang, J. Wang, X. Zhang, C. Zhao, et al., Geochemical characteristics of typical karst soil profiles in Anhui province, Southeastern China, Agronomy 13 (4) (2023) 1067.
[17] Z. Chen, J. Xu, R. Duan, S. Lu, Z. Hou, F. Yang, et al., Ecological health risk assessment and source identification of heavy metals in surface soil based on a high geochemical background: a Case Study in Southwest China, Toxics 10 (6) (2022) 282.
[18] E. Yu, H. Liu, F. Dinis, Q. Zhang, P. Jing, F. Liu, X. Ju, Contamination evaluation and source analysis of heavy metals in karst soil using UNMIX model and Pb-Cd isotopes, Int. J. Environ. Res. Publ. Health 19 (19) (2022), 12478.

[19] Y. Wang, G. Guo, D. Zhang, M. Lei, An integrated method for source apportionment of heavy metal (loid) s in agricultural soils and model uncertainty analysis, Environ. Pollut. 276 (2021), 116666.
[20] G. Guo, K. Li, D. Zhang, M. Lei, Quantitative source apportionment and associated driving factor identification for soil potential toxicity elements via combining receptor models, SOM, and geo-detector method, Sci. Total Environ. 830 (2022), 154721.
[21] S. Jain, S.K. Sharma, T.K. Mandal, M. Saxena, Source apportionment of PM10 in Delhi, India using PCA/APCS, UNMIX and PMF, Particuology 37 (2018) 107–118.
[22] Z. Xu, W. Mi, N. Mi, X. Fan, Y. Zhou, Y. Tian, Characteristics and sources of heavy metal pollution in desert steppe soil related to transportation and industrial activities, Environ. Sci. Pollut. Control Ser. 27 (2020) 38835–38848.
[23] Y. Hu, H. Cheng, Application of stochastic models in identification and apportionment of heavy metal pollution sources in the surface soils of a large-scale region, Environ. Sci. Technol. 47 (8) (2013) 3752–3760.
[24] T. Shi, J. Zhang, W. Shen, J. Wang, X. Li, Machine learning can identify the sources of heavy metals in agricultural soil: a case study in northern Guangdong Province, China, Ecotoxicol. Environ. Saf. 245 (2022), 114107.
[25] T. Kohonen, Self-organized formation of topologically correct feature maps, Biol. Cybern. 43 (1) (1982) 59–69.
[26] J. Li, G. Wang, F. Liu, L. Cui, Y. Jiao, Source apportionment and ecological-health risks assessment of heavy metals in topsoil near a factory, Central China, Exposure and Health 13 (1) (2021) 79–92.
[27] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[28] H. Zhang, S. Yin, Y. Chen, S. Shao, J. Wu, M. Fan, et al., Machine learning-based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China, J. Clean. Prod. 273 (2020), 122858.
[29] Q. Zhang, G. Han, M. Liu, T. Liang, Spatial distribution and controlling factors of heavy metals in soils from puding karst critical zone observatory, southwest China, Environ. Earth Sci. (9) (2019) 78.
[30] W.H. Zhang, Y. Yan, R.L. Yu, G.R. Hu, The sources-specific health risk assessment combined with APCS/MLR model for heavy metals in tea garden soils from south Fujian Province, China, Catena 203 (2021), 105306.
[31] J. Lv, Multivariate receptor models and robust geostatistics to estimate source apportionment of heavy metals in soils, Environ. Pollut. 244 (2019) 72–83.
[32] Z. Jin, J. Lv, Integrated receptor models and multivariate geostatistical simulation for source apportionment of potentially toxic elements in soils, Catena 194 (2020), 104638.
[33] Y. Xu, X. Wang, G. Cui, K. Li, Y. Liu, B. Li, Z. Yao, Source apportionment and ecological and health risk mapping of soil heavy metals based on PMF, SOM, and GIS methods in Hulan River Watershed, Northeastern China, Environ. Monit. Assess. 194 (3) (2022) 1–17.
[34] T. Li, X. Li, W. Luo, G. Cai, Combined classification and source apportionment analysis for trace elements in western Philippine Sea sediments, Sci. Total Environ. 675 (2019) 408–419.
[35] M.A.H. Bhuiyan, S.C. Karmaker, M. Bodrud-Doza, M.A. Rakib, B.B. Saha, Enrichment, sources and ecological risk mapping of heavy metals in agricultural soils of dhaka district employing SOM, PMF and GIS methods, Chemosphere 263 (2021), 128339.
[36] W. Zhou, H. Yang, L. Xie, H. Li, L. Huang, Y. Zhao, T. Yue, Hyperspectral inversion of soil heavy metals in Three-River Source Region based on random forest model, Catena 202 (2021), 105222.
[37] Y. Hu, H. Cheng, A method for apportionment of natural and anthropogenic contributions to heavy metal loadings in the surface soils across large-scale regions, Environ. Pollut. 214 (2016) 400–409.
[38] H. Zhang, A. Yin, X. Yang, M. Fan, S. Shao, J. Wu, et al., Use of machine-learning and receptor models for prediction and source apportionment of heavy metals in coastal reclaimed soils, Ecol. Indicat. 122 (2021), 107233.
[39] USEPA, Exposure Factors Handbook, Final, US Environmental Protection Agency, Washington, DC, 2011 [EPA/600/R-09/052F].
[40] USEPA, Integrated Risk Information System (IRIS). Visited: 2016-3-5, USEPA, 2016. Available at: https://www.cfpub.epa.gov/ncea/iris/search/index.cfm?keyword¼.
[41] B. Li, J. Deng, Z. Li, J. Chen, F. Zhan, Y. He, et al., Contamination and health risk assessment of heavy metals in soil and Ditch sediments in long-term mine Wastes area, Toxics 10 (10) (2022) 607.
[42] A. Karpathy, Random Forest for Matlab, 2012. GitHub.
[43] CNEMC(China National Environmental Monitoring Center), The Soil Background Value in China, China Environmental Science Press, Beijing in Chinese, 1990.
[44] S. Yang, Y. Qu, J. Ma, L. Liu, H. Wu, Q. Liu, et al., Comparison of the concentrations, sources, and distributions of heavy metal (loid) s in agricultural soils of two provinces in the Yangtze River Delta, China, Environ. Pollut. 264 (2020), 114688.
[45] F. Chen, Q. Wang, F. Meng, M. Chen, B. Wang, Effects of long-term zinc smelting activities on the distribution and health risk of heavy metals in agricultural soils of Guizhou province, China, Environ. Geochem. Health (2020) 1–16.
[46] M.J. Kang, Y.K. Kwon, S. Yu, P.K. Lee, H.S. Park, N. Song, Assessment of Zn pollution sources and apportionment in agricultural soils impacted by a Zn smelter in South Korea, J. Hazard Mater. 364 (2019) 475–487.
[47] H. Liu, S. Anwar, L. Fang, L. Chen, W. Xu, L. Xiao, et al., Source apportionment of agricultural soil heavy metals based on PMF model and multivariate statistical analysis, Environ. Forensics (2022) 1–9.
[48] J. Wang, G. Liu, H. Liu, P.K. Lam, Multivariate statistical evaluation of dissolved trace elements and a water quality assessment in the middle reaches of Huaihe River, Anhui, China, Sci. Total Environ. 583 (2017) 421–431.
[49] M.J. Kang, Y.K. Kwon, S. Yu, P.K. Lee, H.S. Park, N. Song, Assessment of Zn pollution sources and apportionment in agricultural soils impacted by a Zn smelter in South Korea, J. Hazard Mater. 364 (2019) 475–487.
[50] J. Li, G. Wang, F. Liu, L. Cui, Y. Jiao, Source apportionment and ecological-health risks assessment of heavy metals in topsoil near a factory, Central China, Water Quality, Exposure and Health (1) (2020) 13.
[51] Q. Guan, F. Wang, C. Xu, N. Pan, J. Lin, R. Zhao, et al., Source apportionment of heavy metals in agricultural soil based on PMF: a case study in Hexi Corridor, northwest China, Chemosphere 193 (2018) 189–197.
[52] L. Chai, Y. Wang, X. Wang, L. Ma, Z. Cheng, L. Su, Pollution characteristics, spatial distributions, and source apportionment of heavy metals in cultivated soil in Lanzhou, China, Ecol. Indicat. 125 (2021), 107507.
[53] X. Zhang, S. Wei, Q. Sun, S.A. Wadood, B. Guo, 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, Ecotoxicol. Environ. Saf. 159 (2018) 354–362.
[54] Y.A. Kettanah, Copper mineralization and alterations in Gercus basalt within the Gercus formation, northern Iraq, Ore Geol. Rev. 111 (2019), 102974.
[55] A.Y. Opekunov, M.G. Opekunova, S.Y. Janson, V.A. Bychinskii, V.V. Somov, S.Y. Kukushkin, E.E. Papyan, Mineral and geochemical characteristics of soils and bottom sediments in the area affected by mining dumps (a case study of the Sibay ore deposit), IOP Conf. Ser. Earth Environ. Sci. 817 (1) (2021), 012078 (IOP Publishing).
[56] R. Zhao, Q. Guan, H. Luo, J. Lin, L. Yang, F. Wang, et al., Fuzzy synthetic evaluation and health risk assessment quantification of heavy metals in Zhangye agricultural soil from the perspective of sources, Sci. Total Environ. 697 (2019), 134126.
[57] B. Pla, A. Zw, B. Xl, B. Mw, A. Lh, C. Bc, et al., Pollution assessment and source analysis of heavy metals in acidic farmland of the karst region in southern China—a case study of quanzhou county, Appl. Geochem. 123 (2020).
[58] H. Yuan an, K. He, Z. Sun, G. Chen, H. Cheng, Quantitative source apportionment of heavy metal (loid) s in the agricultural soils of an industrializing region and associated model uncertainty, J. Hazard Mater. 391 (2020), 122244.
[59] Q. Zhang, G.L. Hah, Speciation characteristics and risk assessment of soil heavy metals from puding karst critical zone, guizhou province, Huan Jing ke Xue= Huanjing Kexue 43 (6) (2022) 3269–3277.
[60] Shan Yang, Xunyang He, Yirong Su, Wei Zhang, Kelin Wang, Effects of lithology and land use patterns on karst soil fertility in Northwest Guangxi, Chin. J. Appl. Ecol. (6) (2010) 7.

[61] H. Zhang, L.N. Wu, K.C. Ouyang, Z.W. Feng, Responses of Riparian Soil Carbon and Phosphorus to Land Use Evolution in Karst Areas. Yangtze River, China (008), 053, 2022.
[62] J. Lv, Z. Zhang, S. Li, Y. Liu, Y. Sun, B. Dai, Assessing spatial distribution, sources, and potential ecological risk of heavy metals in surface sediments of the nansi lake, eastern China, J. Radioanal. Nucl. Chem. 299 (3) (2014) 1671–1681.
[63] Xiaomeng Cheng, Bin-Bin Sun, Chao Wu, Ling He, Dao-ming Zeng, Chen Zhao, Characteristics and health risk of heavy metals in farmland soil of typical pyrolysis mining area in central Zhejiang Province, Environ. Sci. 1 (2022) 43.
[64] J. Liu, B.S. Zheng, H.V. Aposhian, Y.S. Zhou, M.L. Chen, A.H. Zhang, M.P. Waalkes, Chronic arsenic poisoning from burning high-arsenic-containing coal in Guizhou, China, J. Peripher. Nerv. Syst. 7 (3) (2002) 208.
[65] L. Zheng, G. Liu, C.L. Chou, The distribution, occurrence and environmental effect of mercury in Chinese coals, Sci. Total Environ. 384 (1–3) (2007) 374–383.
[66] V.K. Mishra, A.R. Upadhyay, V. Pathak, B.D. Tripathi, Phytoremediation of mercury and arsenic from tropical opencast coalmine effluent through naturally occurring aquatic macrophytes, Water, Air, Soil Pollut. 192 (2008) 303–314.
[67] Y. Kang, G. Liu, C.L. Chou, M.H. Wong, L. Zheng, R. Ding, Arsenic in Chinese coals: distribution, modes of occurrence, and environmental effects, Sci. Total Environ. 412 (2011) 1–13.
[68] Y. Zhao, J. Zhang, W. Huang, Z. Wang, Y. Li, D. Song, et al., Arsenic emission during combustion of high arsenic coals from Southwestern Guizhou, China, Energy Convers. Manag. 49 (4) (2008) 615–624.
[69] S.M. Diami, F.M. Kusin, Z. Madzin, Potential ecological and human health risks of heavy metals in surface soils associated with iron ore mining in Pahang, Malaysia, Environ. Sci. Pollut. Res. 23 (2016) 21086–21097.
[70] D. Fazekašová, J. Fazekaš, Soil quality and heavy metal pollution assessment of iron ore mines in Nizna Slana (Slovakia), Sustainability 12 (6) (2020) 2549.
[71] J. Lv, Y. Wang, PMF receptor models and sequential Gaussian simulation to determine the quantitative sources and hazardous areas of potentially toxic elements in soils, Geoderma 353 (2019) 347–358.
[72] C. Men, R. Liu, L. Xu, Q. Wang, L. Guo, Y. Miao, Z. Shen, Source-specific ecological risk analysis and critical source identification of heavy metals in road dust in Beijing, China, J. Hazard Mater. 388 (2020), 121763.
[73] F. Wang, Q. Guan, J. Tian, J. Lin, Y. Yang, L. Yang, N. Pan, Contamination characteristics, source apportionment, and health risk assessment of heavy metals in agricultural soil in the Hexi Corridor, Catena 191 (2020), 104573.