Environ Geochem Health (2023) 45:6353–6370
https://doi.org/10.1007/s10653-023-01587-8

ORIGINAL PAPER

Distribution, source investigation, and risk assessment of topsoil heavy metals in areas with intensive anthropogenic activities using the positive matrix factorization (PMF) model coupled with self-organizing map (SOM)

Wanjun Jiang · Lishan Meng · Futian Liu · Yizhi Sheng · Sheming Chen · Jilong Yang · Hairu Mao · Jing Zhang · Zhuo Zhang · Hang Ning

Received: 26 October 2022 / Accepted: 21 April 2023 / Published online: 13 June 2023

Abstract Over the past decade, heavy metal (HMs) contamination in soil environments has become severe worldwide. However, their resulting ecological and health risks remained elusive across a variety of soil ecosystems due to the complicated distributions and sources. This study investigated the HMs (Cr, As, Cu, Pb, Zn, Ni, Cd, and Hg) in areas with multi-mineral resources and intensive agricultural activities to study their distribution and source apportionment using a positive matrix factorization (PMF) model coupled with self-organizing map (SOM). The potential ecological and health risks were assessed in terms of distinct sources of HMs. The results disclosed that the spatial distribution of HM contaminations in the topsoil was region-dependent, primarily located in areas with high population intensity. The geo-accumulation index (I_{geo}) and enrichment factor (EF) values collectively displayed that the topsoils were severely contaminated by Hg, Cu, and Pb, particularly in residential farmland areas. The comprehensive analysis combined with PMF and SOM identified both geogenic and anthropogenic sources of HMs including natural, agricultural, mining, and mixed sources (caused by multi-anthropogenic factors), accounting for 24.9%, 22.6%, 45.9%, and 6.6% contribution rates, respectively. The potential ecological risk was predominantly due to the enrichment of Hg, followed by Cd. The non-carcinogenic risks were mostly below the acceptable risk level, while the potential carcinogenic health risks caused by As and Cr should be paid prime attention to, particularly for children. In addition to the 40% geogenic sources, agricultural activities contributed to 30% of the non-carcinogenic risk, whereas mining activities contributed to nearly half of the carcinogenic health risks.

Keywords Heavy metals · Source apportionment · Ecological-health risk · Human activities

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10653-023-01587-8.

W. Jiang · L. Meng · F. Liu · S. Chen (✉) · J. Yang · J. Zhang · Z. Zhang · H. Ning
Tianjin Center, China Geological Survey, Tianjin 300170, China
e-mail: tjdzdczx2013@126.com

W. Jiang · L. Meng · F. Liu · S. Chen · J. Yang · J. Zhang · Z. Zhang · H. Ning
Center of Geoscience Innovation, North China, Tianjin 300170, China

Y. Sheng (✉)
Center for Geomicrobiology and Biogeochemistry Research, State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Beijing 100083, China
e-mail: shengyz@cugb.edu.cn

H. Mao
School of Water Resources & Environment, China University of Geosciences, Beijing 100083, China

Introduction

Soil, the core element connecting the atmosphere, hydrosphere, biosphere, and lithosphere, provides the natural support and biological barrier of basically all ecosystems on the earth's surface. The soil quality is inextricably linked to humans and environments (Lehmann et al., 2020). In recent years, due to the rapid expansion of cities, contamination of soil by heavy metals (HMs) has become increasingly severe worldwide (Árvay et al., 2017; Li et al., 2014). The accumulation of HMs not only affects the ecological environment, but also biological entitles, i.e., animals and plants. HMs contaminants directly or indirectly threaten human health by introducing metal elements that are essential for human health into the body, through various environmental exposure pathways such as drinking water and food chain (Huang et al., 2021; Jiang et al., 2022a; Yang et al., 2018).

In general, HMs such as Cd, Cr, Hg, Pb, Cu, Zn, Ni, and metalloid arsenic (As) are derived from both geogenic and anthropogenic sources (Kebonye et al., 2021). They are released naturally through weathering of geological parent materials and/or from various human activities including agriculture, transportation, mining and smelting, and wastes rich in metal residues (Li et al., 2021; Ungureanu et al., 2017). HMs are prioritized as contaminants under control due to their potential toxicity, persistence, and irreversibility (Lin et al., 2018; Man et al., 2010). HMs in soils would cause a threat to the health of human beings when they are converted from solid form to either ionic moieties or organometallic moieties (Madrid et al., 2002). To improve the soil environmental quality and protect human health, it is necessary to differentiate the natural and anthropogenic sources of HMs, determine the quantitative contribution, and assess the impact of HMs exposure on the ecological environment and human health (Cheng et al., 2020; Huang et al., 2021).

The past studies have been conducted on the distribution, source apportionment, and risk assessment of topsoil HMs under various human activities. However, most of the research focused only on human activity or the same receptor medium, such as the agricultural (Rodriguez et al., 2008; Zhang et al., 2020), urban parks (Gu et al., 2017; Liu et al., 2020; Wang et al., 2019), industrial park (Li et al., 2021; Long et al., 2021), and mining and smelting area (Árvay et al., 2017; Tian et al., 2018; Xiao et al., 2017). The distribution, the inter-relationship between heavy metals, and quantitative analysis of sources and health risk assessment of topsoil HMs under the influence of mixed anthropogenic activities (e.g., long-term mining and smelting of polymetallic ores and intensive agricultural activities) remained poorly understood.

Multivariate statistical analyses such as Principal Component Analysis (PCA), Chemical Mass Balance (CMB), and Cluster Analysis (CA) have been widely employed to reveal the distribution and possible sources of HMs (Davis et al., 2009; Facchinelli et al., 2001; Li et al., 2004). However, these methods are limited in identifying multiple sources simultaneously, obtaining the nonlinear correlations, and classifying different variables and samples, especially for high-dimensional and complex datasets (Astel et al., 2007; Kim et al., 2020). Compared with the conventional models, the positive matrix factorization (PMF) was developed to cope with uncertainties and error propagation problems (Huang et al., 2021; Zanotti et al., 2019). Recently, PMF has been widely applied to identify pollution sources and apportion contributions in various environmental media (Brinkman et al., 2006; Zhang et al., 2018). The success of its application is affected by the error of sample data values, model structure, and parameter representation, and the conclusion would rely on the interpretation of background data in the study area (Huang et al., 2021; Li et al., 2020; Mao et al., 2023). In addition, to accurately explain the relationship between variables, owing to the complexity and uncertainty of soil heavy metal occurrences, it is imperative to reduce and classify the data. To overcome the shortcomings of traditional classification methods, self-organizing map (SOM), which can reveal local relationships between variables and classify nonlinearity and dispersed data (Jiang et al., 2022a; Kebonye et al., 2021; Lee et al., 2019), was increasingly adopted in data classification of related research in the field of earth environmental science (Agyeman et al., 2022a; Nakagawa et al., 2020; Qu et al., 2021). Moreover, SOM supports the technique of using reference vectors to provide an informational picture of the data, which clearly shows the interdependencies between variables (Pearce et al., 2011; Wang et al., 2020; Zhu et al., 2020).

In the current study, primary topsoil HMs (Cu, Pb, Zn, Cr, Ni, Cd, As, Hg, and Mn) in areas with typical

multi-mineral resources and intensive agricultural activities were analyzed to evaluate the pollution level of topsoil HMs. Combining with PMF and SOM as well as Pearson correlation analysis, the distribution and potential contaminant sources were investigated. The potential ecological-health risks were evaluated. This study provides a strategy and scientific basis for preventing, controlling, and remediating the HMs contaminations in areas undergoing mining and agricultural activities.

Study area

Mountainous, low mountainous, and plain areas make up the majority of the physiognomy in the study area. In terms of topography and land-use type, the study area is divided into mountainous, mining, agricultural, and residential areas. The region with strong agricultural activity (i.e., farmland zones) is primarily located at the upper piedmont sloping plain in the southwest, followed by the riverbed or riparian ribbon alluvial plain in the southern or eastern areas, or the valleys of the northwestern mountain (Fig. 1). There are also many agricultural activities in the central or northern region even though some farmlands are distributed in the valleys and mountains. Additionally, large-scale vegetable planting bases, modern agricultural industrial parks, and fruit industrial parks are found. There are abundant coal, iron, and gold ore resources, along with lead and zinc, phosphorite, and graphite mineral resources, while the other mineral resources are relatively small in size or distribution of ore spots (Jiang et al., 2022b).

The coal resources are primarily located in the southwest and east mining areas of the study area. Iron mines are mainly located in the south and northwest area, followed by the central area, where hematite, magnetite, and titanium-magnet minerals of large scale are developed. The non-ferrous mineral resources (mainly lead–zinc ore, with a few copper and molybdenum ore) controlled by the metallogenic geological conditions are distributed in the southeast. The gold ores (49 deposits) are the mineral resources with the broadest distribution area and the largest mining scale in this region (around 1500 km^2). Additionally, silver ore resources in the study area are less abundant, and most silver deposits are formed by the symbiotic or associated combination of Ag and other beneficial ore-forming elements (e.g., Pb and Zn).

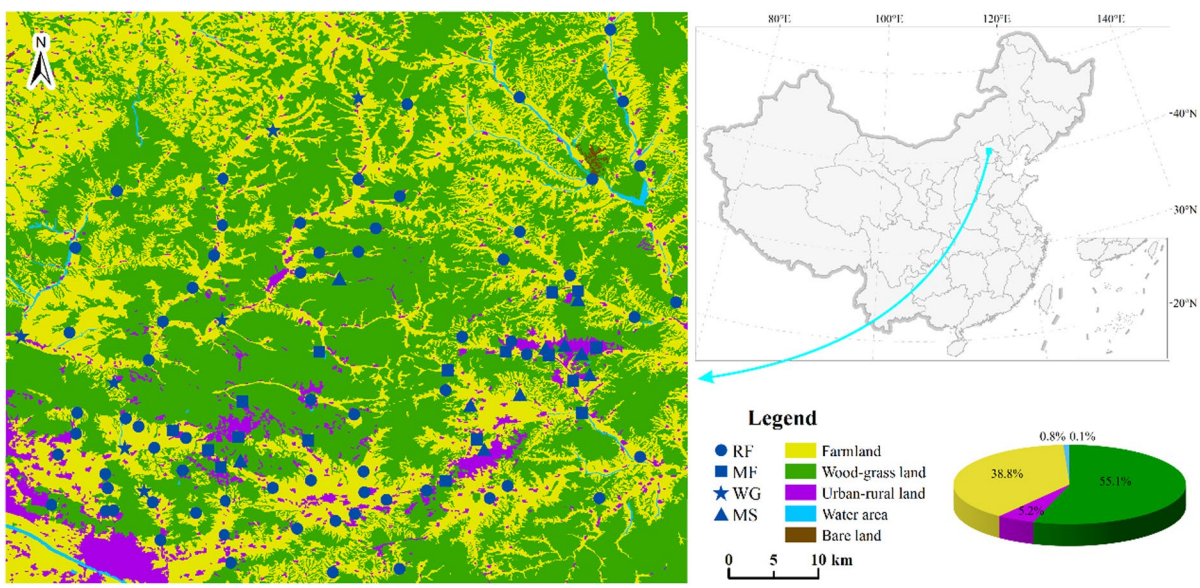

Fig. 1 Sampling location of HMs in topsoil samples under different land-use types in the study area

Materials and methods

Soil sampling and analyses

A total of 101 topsoil samples (from 0 to 20 cm) were collected vertically by soil drills, including 67 residential farmland soils (RF), 17 mining farmland soils (MF), and 7 woodland–grassland soils (WG) (Fig. 1). In addition to the soil samples, 10 mining or tailing/slag waste samples (MS) were collected in the mining factory, smelting plant, and tailings pond (Fig. 1). Each sample is taken by a clean sampler to reduce cross-contamination and stored in polyethylene bags and sent to the laboratory for analysis. All samples were naturally dried at 20 °C until constant weight after removing the plant residual roots, gravel, and other debris, and then crushed and passed through a 2-mm sieve.

The samples of HMs were digested in sterile dry Teflon crucibles by a concentrated acid mixture (HNO_3-HF-$HClO_4$) and a quantified aqua regia solution (HCl-HNO_3), and then placed on the heating plate in a fume cupboard. The electric heating plate was heated to 260 ℃ until the liquid evaporates completely. Then, 5 ml of HNO_3 was added to the samples and held at 130 ℃ in a drying baker to dissolve the residue. When crucibles cooled, the samples were removed to a 50-ml volumetric flask and diluted with ultrapure water for testing. The concentrations of Cu, Pb, Ni, and Cd were determined by Inductively Coupled Plasma Mass Spectrometry (ICP-MS) (Thermo X series II, Thermo Fisher Scientific, USA), and those of As and Hg were analyzed by Atomic Fluorescence Spectrometer (AFS, HJ-680–2013, China). The concentrations of Zn, Cr, and Mn were measured by inductively coupled plasma atomic emission spectrometry (ICP-AES) (iCAP6300, Thermo Scientific, Waltham, USA). The analyses and quality assurance/quality control (QA/QC) for HMs were determined by duplicate samples, blanks, and standard substances (GSS-24) from the Centre of National Standard Reference Material of China. The recoveries of substrate samples and the standard substances ranged from 88 to 106% and 98% to 103%, respectively. The detection limits (MDLs) of Cu, Pb, Zn, Cr, Ni, Cd, As, and Hg were 0.6, 0.5, 0.03, 0.2, 0.6, 0.05, 0.2, and 0.005 mg/kg, respectively. The relative standard deviation of the duplicate samples was less than 5%. pH was measured by glass electrode method using a pH meter (Thermo, USA).

Evaluation method of pollution levels

Geochemical methods including geo-accumulation index (I_{geo}) and enrichment factor (EF) can provide relative levels of soil heavy metal contamination.

Geo-accumulation index (I_{geo})

I_{geo} is an intuitive geochemical standard for quantitative evaluation of heavy metal contamination levels which takes into account the influence of background values caused by natural geological processes as well as exogenous HMs produced by human activities:

$$I_{geo} = \log_2[C_i/(K \cdot B_i)] \quad (1)$$

where C_i represents the measured concentration of heavy metal i (mg/kg), B_i denotes the geochemical background value of corresponding element i (mg/kg), and K is the correction coefficient, generally 1.5 (Chen et al., 2019; Long et al., 2021).

Enrichment factor (EF)

EF is usually used to evaluate the enrichment degree of metal elements in topsoil and can also help to distinguish an anthropogenic source from natural sources:

$$\mathrm{EF} = \frac{(C_i/C_{ref})\mathrm{sample}}{(C_i/C_{ref})\mathrm{background}} \quad (2)$$

where C_i represents the concentration of element i, and C_{ref} is the reference element for normalization (mg/kg). Mn is adopted as the reference element because it is more stable in the earth's crust than other metals (Cheng et al., 2018; Wu et al., 2010). The classification of the I_{geo} and EF is shown in Table S1.

Self-organizing map (SOM)

The self-organizing map (SOM) is an unsupervised competitive learning neural network method. Similar sample points in high-dimensional space could be mapped to the neighboring neurons in the two-dimensional output layer (2D) after the systematic analysis of nonlinear complex data (Haselbeck et al., 2019;

Mao et al., 2021). Additionally, the global and local relationships between variables are both displayed in the unified distance matrices (U-matrix) and median distance matrices (D-matrix) of SOM (Lee et al., 2019; Mao et al., 2021), as described by Jiang et al. (2022a) and Kim et al. (2020). The calculations and visualization of SOM were executed using MATLAB software, and a more detailed description is shown in Text S1.

Positive matrix factorization model (PMF)

In this study, the PMF receptor model 5.0 developed by the USEPA (2014) was used to identify the sources of HMs in topsoil. The PMF model decomposed the original matrix through several calculations and obtained the optimal matrix G and F to minimize the objective function Q:

$$X_{ij} = \sum_{k=1}^{p} G_{ik} F_{kj} + E_{ij} \tag{3}$$

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{E_{ij}}{U_{ij}} \right)^2 \tag{4}$$

where i, j, and k represent the number of samples, elements, and different pollution sources, respectively. X_{ij} is the concentration of the jth chemical component of the ith sample (mg/kg), F_{kj} is the contribution concentration of the jth chemical component of the source k (mg/kg), G_{ik} is the contribution of the source k to the ith sample, and E_{ij} is the residual matrix.

$$\begin{aligned} &\text{If } c \le \text{MDL}, U_{ij} = \text{MDL}, \text{ else, } U_{ij} \\ &\quad = \sqrt{(\text{Error fraction} \times C)^2 + (0.5 \times \text{MDL})^2} \end{aligned} \tag{5}$$

where U_{ij} is the uncertainty of the jth chemical composition of the ith sample. MDL and error fraction are the method detection limit and the percentage of measurement uncertainty, respectively. And C is the concentration of the elements.

Potential ecological-health risks assessment

Potential ecological risks assessment

The potential ecological risk index was developed to evaluate the potential impact of contaminations on the ecological environment according to the toxicity of HMs and the response of the environment:

$$\text{RI} = \sum_{i=1}^{n} E_r^i = \sum_{i=1}^{n} T_r^i \times (C_m^i / C_R^i) \tag{6}$$

where RI is the comprehensive potential ecological risk index for all HMs; E_r^i represents the potential ecological risk index of single heavy metal i; T_r^i stands for the biological toxic response coefficient of corresponding heavy metal i, and the values of element increase were 1(Zn), 1(Mn), 2(Cr), 5(Cu), 5(Ni), 5(Pb), 10(As), 30(Cd) and 40(Hg) (Agyeman et al., 2022b; Hakanson, 1980; Jiang et al., 2020; Li et al., 2021; Zuo et al., 2022); C_M^i and C_R^i are the measured concentration and the reference value of the heavy metal i, respectively (mg/kg). The assessment standard for E_R^i and RI is presented in Table S2.

Human health risks assessment (HRA)

In general, direct oral ingestion, dermal contact, and inhalation absorption are three major pathways for human long-term exposure to HMs in topsoil (Chen et al., 2015; Wang et al., 2019). The HRA model proposed by the United States Environmental Protection Agency (USEPA, 2011) was employed to quantitatively evaluate the potential health risk of hazardous substances to the human body.

Exposure doses through oral ingestion, dermal contact, and inhalation can be calculated as follows:

$$\text{ADI}_{\text{ing}-i} = C_i \times \frac{\text{Ing}R \times \text{EF} \times \text{ED}}{\text{BW} \times \text{AT}} \times \text{CF} \tag{7}$$

$$\text{ADI}_{\text{dermal - }i} = C_i \times \frac{\text{SA} \times \text{AF} \times \text{ABS} \times \text{EF} \times \text{ED}}{\text{BW} \times \text{AT}} \times \text{CF} \tag{8}$$

$$\text{ADI}_{\text{inh}-i} = C_i \times \frac{\text{Inh}R \times \text{EF} \times \text{ED}}{\text{PEF} \times \text{BW} \times \text{AT}} \tag{9}$$

where ADI_{ing}, $\text{ADI}_{\text{dermal}}$, and ADI_{inh} represent the average daily intake from soil ingestion, dermal, and inhalation absorption, respectively (mg/kg·day). The significance and exact values of each parameter are shown in Table S3.

Non-carcinogenic risk assessment

Non-carcinogenic risk assessment of a single element in metal-contaminated soil is usually characterized by the hazard quotient (HQ). However, the hazard index (HI) represented the accumulative/total risk value of various non-carcinogenic indicators from all exposure pathways:

$$\mathrm{HI} = \sum \mathrm{HQ}_i = \sum \frac{\mathrm{ADI}_i}{\mathrm{RfD}_i} \tag{10}$$

where the RfD_i is the corresponding reference toxicity threshold dose of element i [mg/(kg·day)]. For non-carcinogenic risk, if HI > 1, there will be obvious potential adverse health effects for the exposed individual. Otherwise, it is considered to be an acceptable level (Gu et al., 2017; Zhang et al., 2020).

Carcinogenic risk assessment

Generally, carcinogenic risk (CR) refers to the probability of developing any type of cancer over an individual lifetime due to carcinogenic exposure in the environment. Similarly, total carcinogenic risks (TCR) are obtained by summing the individual cancer risks across all carcinogens and/or exposure pathways, as follows:

$$\mathrm{CR} = \mathrm{ADI} \times \mathrm{SF} \tag{11}$$

$$\mathrm{TCR} = \sum \mathrm{CR}_i = \sum \mathrm{ADI}_i \times \mathrm{SF}_i \tag{12}$$

where SF_i represents the cancer slope factor of the element under different exposure pathways [(kg·day)/mg], and the values of SF and RfD from the literature are shown in Table S4. Generally, the tolerable level of carcinogenic risk ranges from 1×10^{-6} to 1×10^{-4}, whereas TCR less than 1×10^{-6} is considered a negligible risk. If the values of TCR exceed 1×10^{-4}, the level of carcinogenic risk is high and poses potential threat to human (Chen et al., 2015; Li et al., 2014). The values of RfD and SF for different HMs are shown in Table S4.

Source-oriented HRA

The contribution of different sources to health risk was quantified by combining the PMF with HRA model. The concentration C_i of different HMs in Eqs. (7)–(9) was replaced by C_{ki}^n, as follows:

$$C_{ki}^n = C_i \times C_k^n \tag{13}$$

After that, the source-oriented HRA was repeated using Eqs. (10)–(12). Where C_k^n represents calculated contribution rate of heavy metal i at sample k originating from identified source n.

Results and discussion

Characteristics and distributions of HMs

The geochemical background values (GBVs) of soil environmental quality and the risk control standard values (RSVs) for soil contamination of agricultural land (GB 15,618—2018) were employed for contrastive analysis of the content and distribution characteristics of HMs. The HMs in the topsoil samples vary greatly and were dependent upon land-use types (Table 1). Although the contents of HMs such as Cu, Cr, Pb, and Cd in a few mining waste residues were above the RSVs of corresponding soil quality standards, the HMs contents in most soil samples were lower (Table 1). However, with the exception of 74.26% Cr and 64.36% As, the contents of Cu, Pb, Zn, Ni, Cd, and Hg in most samples were indeed greater than the GBVs (Table 1). Specifically, the exceeding rates of Pb (86.14%) and Cu (83.17%) were the highest, followed by Zn, Ni, Cd, and Hg (> 50%), implying that Cu and Pb may be the main contaminants.

Understanding the spatial distribution of HMs is helpful to uncover the source apportionment of HMs (Huang et al., 2021; Jin et al., 2019). The spatial distribution of investigated HMs contents exhibited regional differences (Fig. S1), which were likely affected by both natural and human causes. Samples with Cu contents exceeding the RSVs were mainly distributed in the east and southeast of the study area (Fig. S1a). Samples with high Cu contents (1.5–twofold of GBVs) were found in the south and southwest of the study area. In the same areas, there were samples with high Cr and Ni contents (exceeding the RSVs or twofold of GBVs) (Fig. S1d and S1e). Most samples also had high levels of Pb and Hg (exceeding twofold of GBVs) (Fig. S1b and S1h), and high Zn

Table 1 Heavy metal content statistics of soil samples in the study area (mg/kg)[a]

Element	Cu	Pb	Zn	Cr	Ni	Cd	As	Hg	Mn	pH
RF (n = 67)										
Min	13.1	14.1	31.5	25.7	11.5	0.08	0.55	0.01	420	7.13
Mean	32.84	38.56	78.96	61.59	29.45	0.19	9.05	0.1	733.13	8.92
Max	302	297	254	349	94.6	1.15	28.3	1.34	1600	9.74
Q_1 (25%)	23	24.2	62.1	47.8	21.9	0.12	6.76	0.02	660	8.64
Median	27.2	26.9	71.1	58.7	29.5	0.15	8.55	0.03	700	8.92
Q_3 (75%)	32.4	31.3	79.7	65.8	31.8	0.2	9.95	0.07	780	9.32
RSD	1.06	1.10	0.45	0.61	0.38	0.81	0.43	2.14	0.23	0.06
MF (n = 17)										
Min	23.8	3.6	37.2	42.4	27.6	0.07	0.7	0.01	600	8.65
Mean	45.91	33.69	69.68	123.25	53.73	0.14	7.15	0.08	860	9.16
Max	119	108	113	448	142	0.27	14.8	0.32	1200	9.77
Q_1 (25%)	29	23.7	56.4	60.9	29.8	0.11	2.75	0.01	690	8.93
Median	36.6	26.5	69.7	68.8	32.4	0.13	8.27	0.03	820	9.05
Q_3 (75%)	51.4	39.8	74.9	107	69.8	0.17	10.2	0.05	940	9.52
RSD	0.59	0.83	0.30	0.96	0.66	0.36	0.59	1.37	0.24	0.04
WG (n = 7)										
Min	20.4	19.8	57.6	42.8	21.9	0.09	6.48	0.01	530	7.03
Mean	24.96	26.73	61.64	53.01	26.56	0.12	7.76	0.06	648.57	9.12
Max	29.8	32.6	64.9	75.1	35.4	0.15	9.38	0.2	760	9.71
Q_1 (25%)	22.9	21.5	59	42.9	22.4	0.11	6.5	0.03	580	9.16
Median	23.4	27	62.6	49.6	26.5	0.12	7.44	0.04	650	9.52
Q_3 (75%)	27.8	30.7	63.7	59.2	30.1	0.14	9.13	0.06	740	9.62
RSD	0.13	0.17	0.04	0.22	0.18	0.17	0.16	1.06	0.13	0.10
MS (n = 10)										
Min	7.83	2.44	28	11	5.68	0.071	0.12	0.0047	460	7.95
Mean	227.83	57.86	372.84	206.48	66.18	2.49	6.65	0.01	1058	9.38
Max	1030	417	3170	1060	202	23.9	48	0.031	1900	9.76
Q_1 (25%)	44.6	5.11	40.4	45.4	17.8	0.082	0.75	0.006	960	9.39
Median	74	11.925	48.15	68.15	56.3	0.105	1.87	0.00945	1000	9.485
Q_3 (75%)	326	39.8	93	232	98.5	0.16	4.84	0.014	1200	9.65
RSD[a]	1.42	2.21	2.64	1.54	0.92	3.02	2.20	0.75	0.38	0.06
BV[b]	22.19	20.68	67.34	68.33	28.47	0.13	9.15	0.024	618. 46	8.47
RSV[c]	100	170	300	250	190	0.6	20	1		-

RF: Residential farmland soil; MF: Mining farmland soil; WG: Woodland and grassland soil; MS: Mining waste or tailing/slag waste soil; n: Number of observations; and Q1 and Q3 represent the upper and lower quartile, respectively

[a]Relative standard deviation

[b]Background values (Wei et al., 2019)

[c]Risk screening values from soil environmental quality risk control standard for soil contamination of agricultural land (GB 15,618—2018) (Ministry of Ecology and Environment of P.R. China 2018)

and Cd concentrations were observed in a few samples (Fig. S1c and S1f). Comparatively, irrespective of a few mining residues and farmland samples, As was much less abundant (Fig. S1d and S1g). In contrast, the content of HMs in the samples in the north mountain valleys of the study area had relatively low levels of HMs (Fig. S1).

On the whole, the distribution Cu, Pb, Zn, Ni, Cd, and Hg displayed a similar pattern, i.e., the piedmont sloping plain and valley in the south of the study area, in contrast with the lower values in mountain valleys in the northwest/northeast. Because the south area had intensive human activities, this result implied that their sources were likely associated with human activities. Relative to the topsoils, the maximum contents of HMs in the mining waste or tailing/slag waste samples (MS) were higher than the GBVs and even several times greater than the RSVs (Table 1). The wastes in the mine tailings pond and concentrators could pose a threat to the surrounding pristine topsoils, likely resulting in potential health risks to residents in the mining area.

Pollution assessment of HMs

The contamination of topsoil samples ranged from the uncontaminated level to the extremely contaminated level, as determined by the classification of I_{geo} (Fig. 2a, Table S1). Despite As and Ni, other HMs mostly exhibited levels of heavily contaminated or higher. In the cases of Zn, Cr, and Cd, the heavily contaminated samples appeared in the MS samples in the mining area. Generally, the contamination of HMs in topsoil displayed the order of Hg > Cu > Pb > Cd > Ni > Zn > Cr > As (Fig. S1). Distinct anthropogenic activity intensity such as industrial production, mining, and agricultural activities could have contributed to the soil contaminations (Chen et al., 2015, 2019; Marrugo-Negrete et al., 2017; Wang et al., 2019). Grassland (WG) samples appeared to be the least contaminated by HMs, i.e., the I_{geo} values were relatively lower than other land-use types. For the residential farmland (RF) samples, the I_{geo} values of Zn, Cd, Hg, Pb, and As were the highest, while the highest I_{geo} values for Cu, Ni, and Cr were observed in mine wastes (MS).

The EF of HMs ranged from the minimal enrichment level to extremely high enrichment (Table S1, Fig. 2b). Consistently, a majority of samples had significant enrichment of Cu, Zn, Cd, and Hg, whereas only a few samples had significant enrichment of Ni, Pb, As, and Cr. Analogous to the I_{geo}, Hg was considered to be the most enriching parameter, and moderate enrichment or above levels accounted for 25.74% of total samples (Fig. S3). In addition, the Pb and Cu samples at the similar level independently accounted for 14.85% and 9.9% of total samples, thereby they were also considered the prime contaminants in topsoils. Comparatively, other HMs were mostly at the level of minimal enrichment or even lower. Among different land-use types, despite the MS samples, EF values of Zn, Cd, Hg, Pb, and As in RF samples were higher than the other two land-use types, also in line with the I_{geo} result. In contrast, the EF values of Cu, Ni, and Cr were the highest in MF samples located

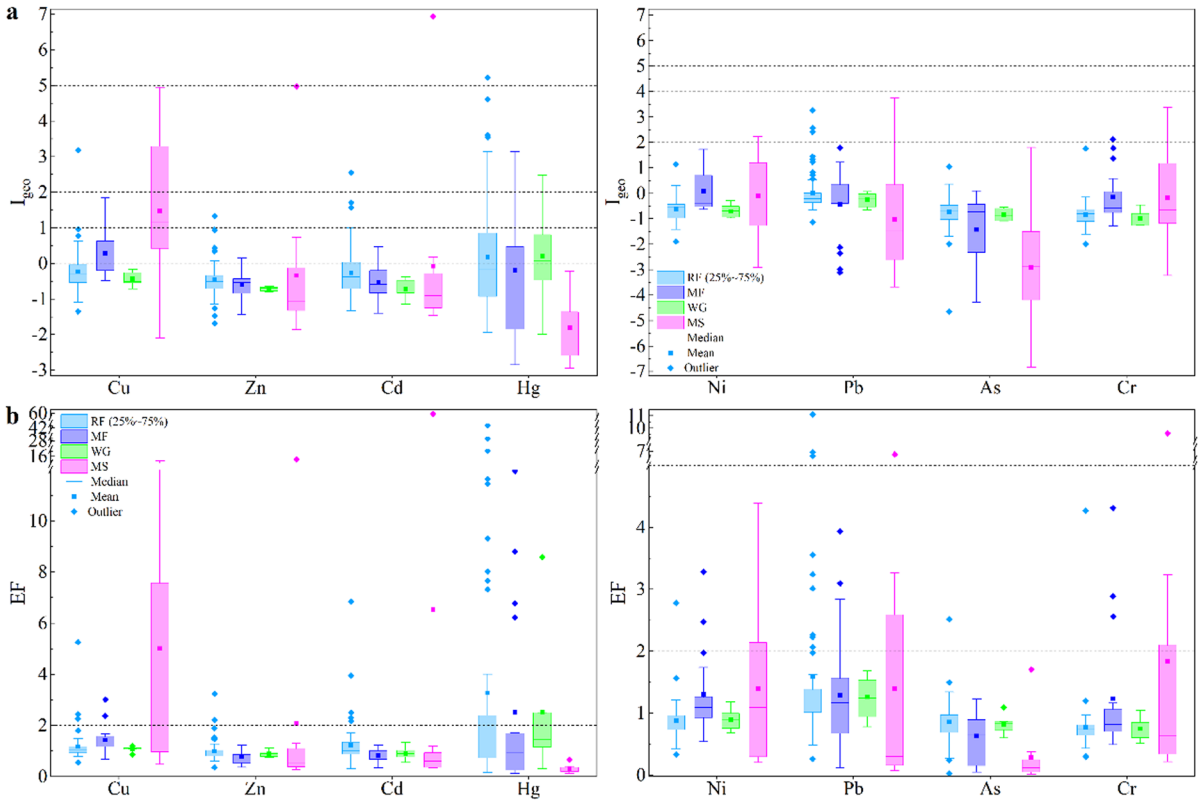

Fig. 2 Box diagrams of **a** I_{geo} and **b** EF in different sample types

in mining farmland areas (Fig. 2b), and the high EF values of Hg in both WG and RF samples were also observed. The above results illustrated that the influence of anthropogenic activities on the accumulation of HMs such as Hg, Cu, and Pb in the study area should be given prime attention.

Source investigation of HMs

Self-organizing map (SOM) for source identification

Totally 54 (6×9) neurons and 4 clusters were selected according to Heuristic rules and the Davies–Bouldin index (Fig. 3a). The samples in the same cluster likely had similar distributions of HMs; thus, these samples might have similar underlying sources. Each U-matrix map represented an index value obtained after dimension reduction, as marked by shades of blue to yellow (Lee et al., 2019; Zhu et al., 2020). The neurons with high values were shown in yellow, while the neurons with low values were indicated in blue (Fig. 3b). Afterward, informative and qualitative relations among the parameters including HMs and sample types were intuitively shown by comparing SOM graphs according to the color gradient.

The color change gradients of Cu, Cr, Ni, and Mn were similar, indicating that these HMs have positive inter-correlations (Fig. 3b). Similarly, Hg, Cd, Pb, and Zn had nearly consistent color change gradients, hinting that the possible factors affecting their contents and distributions might be close. In contrast, As had a unique color change gradient, indicating that the factors affecting its distribution and source were different from those of other HMs. The D-matrix of SOM was then divided into three clusters, each of which likely represented different sources of HMs.

Cluster 1, characterized by the As (other metals did not fall within this area), was located on the left side of the D-matrix map in the SOM results. It was composed of approximately 75% topsoil samples in the study area. However, except for a few samples, the content of As in the topsoil samples was low (Fig. S1g) as also evidenced by the low I_{geo} and EF values (Fig. 2). These topsoil samples were widely distributed in the study area (Fig. S4), suggesting that the source of As was possibly

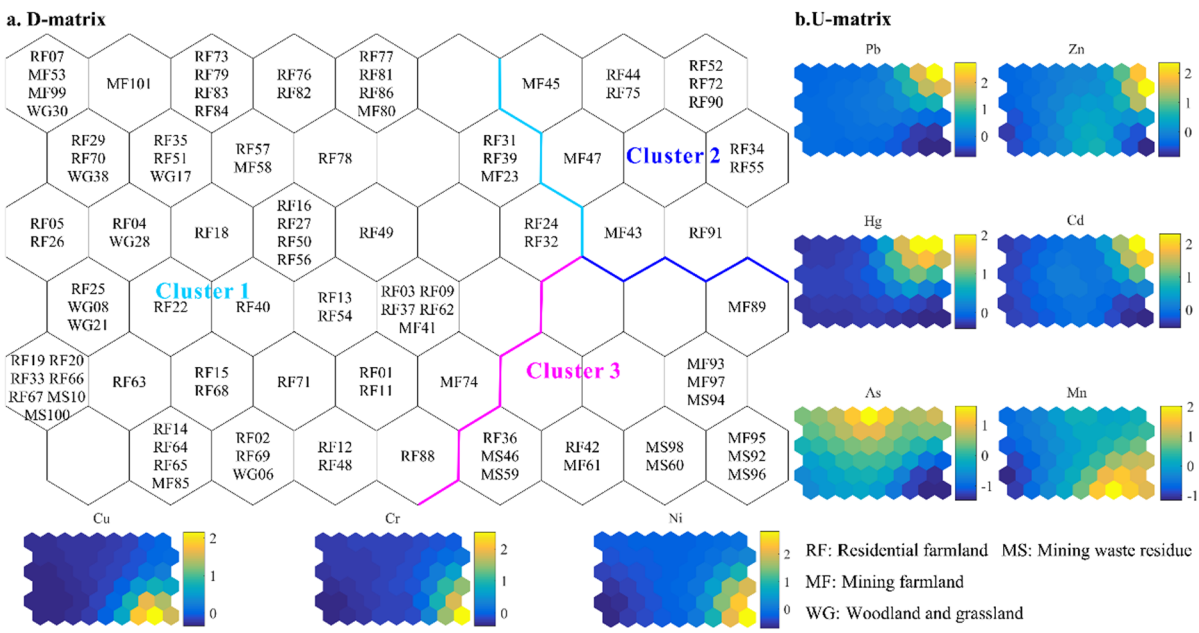

Fig. 3 **a** The SOM matrix map of topsoil samples: the clustering pattern in the SOM, different colors represent different clusters, and the number in a hexagon denotes the sample number. **b** The SOM visualization of corresponding variables. The blue and yellow colors correspond to low and high values, respectively, which can detect the correlation between variables

affected by geogenic sources, e.g., the weathering of rocks or soil parent materials.

Cluster 2, characterized by Hg, Cd, Pb, and Zn, was located at the upper-right corner of the D-matrix. Totally 8 RF samples and 3 MF samples were scattered in the upper part of the alluvial plain in the southwestern mountain of the study area, and some of them were close to the MS samples (Fig. S4). Based on the results of I_{geo} and EF (Fig. 2), the distribution and sources of these HMs (cluster 2) would have been severely affected by anthropogenic activities, such as mining and agricultural activities.

Cluster 3, characterized by Cu, Cr, Ni, and Mn, was located in the lower-right part of the D-matrix, including 7 MS, 5 MF, and 2 RF samples. The samples were mainly located near the mining region in the southeastern mountainous of the study area (Fig. S4), suggesting that these HMs in topsoil were potentially affected by mining activities.

Pearson correlation and PMF model

Comprehensive analysis using Pearson linear correlation and the PMF 5.0 model was employed to investigate the potential sources of HMs (Fig. 4). Pearson correlation analysis was employed to initially determine the correlations in the pairwise comparisons of HMs in topsoil, revealing a potential common source or geochemical characteristics (Cheng et al., 2020; Huang et al., 2021; Zuo et al., 2022). Then, the factor contributions and the correlation coefficients of variables were correlated to verify the results of the PMF model (Huang et al., 2021; Mao et al., 2023; Zuo et al., 2022). Nine variables and 101 samples were selected to investigate the source of HMs and the amount of their contribution utilizing the PMF model. Considering the results of SOM, 3–6 factors were examined with 20 base runs in random seed mode to find the "optimal solution" (Cheng et al., 2020; Zhang et al., 2018). Four factors (Fig. 4a) were determined based on the minimum and stable objective function Q, which was applied to ensure residual matrix E (Cheng et al., 2020; Salim et al., 2019; Zanotti et al., 2019). The signal-to-noise ratios (S/N) of all HMs were larger than 4, defined as "strong," and the most regression coefficients (R^2) between the observations and predictions were larger than 0.6. It indicated that the PMF was applicable to locate the source, and the results were reliable (Jiang et al.,

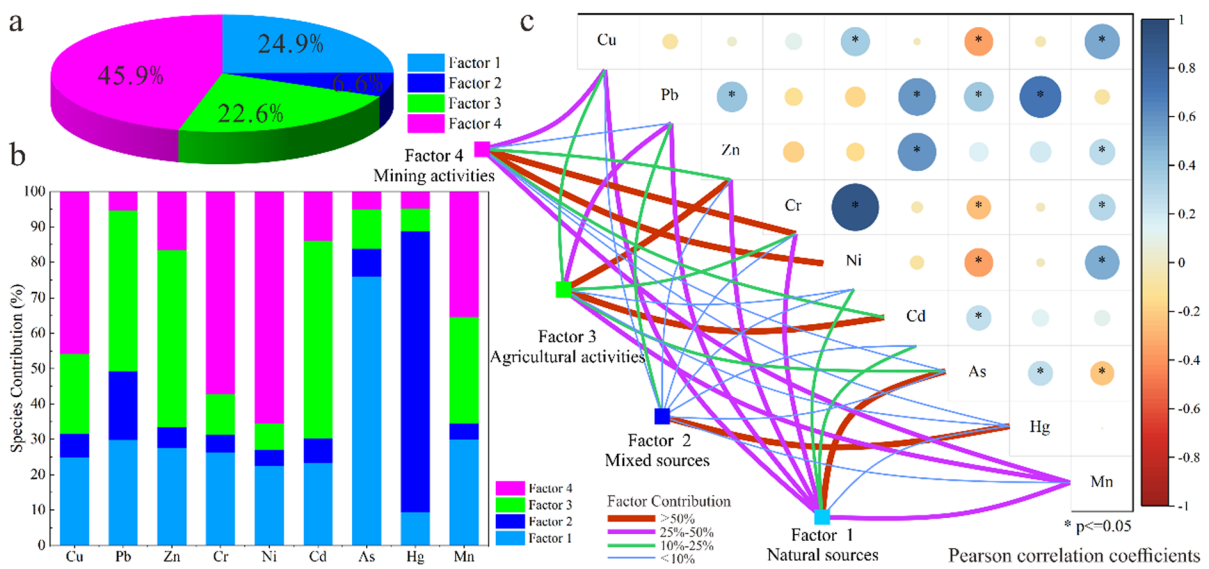

Fig. 4 Source apportionment of HMs. **a** The contribution of each factor in the results of PMF model. **b** Factor profiles of HMs in different types of topsoil derived from PMF model. **c** The Pearson correlations between HMs and the relationships with different sources using PMF model. The correlation coefficient was represented by color gradient in pairwise comparisons of HMs. Factor contribution in PMF model was related to each HMs

2020; Mao et al., 2023; Norris et al., 2014; Salim et al., 2019).

Factor 1, accounting for 24.9% of the contribution rate (Fig. 4a), was mainly explained by As (75.91%, Fig. 6b), followed by Cu, Pb, Zn, Cr, and Mn (below 30%) (Fig. 4c). As illustrated by I_{geo}, EF, and SOM results, the source of As could be recognized as geogenic. As and Cr are widely distributed in the earth's crust (Facchinelli et al., 2001; Šajn et al., 2011) and might be associated with natural sources (Jin et al., 2019; Zhang et al., 2018). Although the sources of Cu, Pb, Zn, and Cr were related to intense anthropogenic activities in many environments (Cai et al., 2019; Marrugo-Negrete et al., 2017), the metallogenic geological background conditions of the nonferrous mineral resources (mainly lead–zinc ore, with a few copper ore) in the study area could contribute to their enrichment. Thus, Factor 1 could represent the effect of some natural sources (e.g., the weathering of rocks or soil parent materials). However, we cannot exclude the possibility that a few topsoil samples were influenced by both anthropogenic and geogenic inputs.

Factor 2, with 6.6% of the total contribution rate, was mainly associated with Hg (79.35%, Fig. 4b), followed by Pb (Fig. 4c). The previous studies have shown that Hg in the topsoils could harbor diverse sources, such as the mining and smelting of gold or mercury mines (Árvay et al., 2017; Csavina et al., 2012), copper ore smelting (Tomiyasu et al., 2017; Xiao et al., 2017), tailpipe or fossil fuel burning emissions (Jiang et al., 2006; Pacyna et al., 2010), and mercury pesticide or sludge fertilizer (Dong et al., 2017; Huang et al., 2021). The mine wastes (MS) had the lowest I_{geo} and EF values for Hg, indicating that mining activities were not the main source of Hg in topsoil. Because Hg was significantly correlated with Pb (Fig. 4c), the sources of Hg and Pb might be similar. As a major marker to identify traffic sources, Pb particles emitted from vehicle exhaust will contaminate the soil with dust (Duzgoren-Aydin et al., 2004; Fei et al., 2022; Kadi, 2009; Sun et al., 2019). Moreover, the spatial distribution showed that samples with high Hg content were mainly distributed in areas with the high population density and frequent anthropogenic activities (Fig. S1h). Therefore, Factor 2 might be interpreted as mixed sources caused by multi-anthropogenic factors.

Factor 3, accounting for a 22.6% contribution rate, was mainly characterized by Cd and Zn (~50%, Fig. 4b), followed by Pb, Mn, Cu, and Cr (Fig. 4c). Cd exists in phosphate fertilizer (Hu et al., 2018; Nan et al., 2002; Nicholson et al., 2003), normally referred to the hallmark of agricultural practice (Baltas et al., 2020; Sun et al., 2013). The Zn and Cu were found in almost all agricultural inputs, including fertilizers, pesticides, fungicides, and manures (Li et al., 2006; Marrugo-Negrete et al., 2017). As such, agricultural activities would be one of the main sources of Cd, Zn, and Cu (Hu et al., 2018; Lu et al., 2012; Nogueirol et al., 2010). As observed by the I_{geo} and EF results (Fig. 2), some farmland soils, e.g., RF samples, were replete with Cd and Zn. Meanwhile, the spatial distribution of Cd and Zn in topsoil in the study area also exerted similar distributions (Fig. S1c and S1f), and the contents of Cd and Zn were significantly correlated (Fig. 4c), thereby Factor 3 might be tightly related to agricultural activities.

Factor 4 with a contribution rate of 45.9% (Fig. 4a) was primarily characterized by Ni and Cr (~60%, Fig. 4b), followed by Cu and Mn (Fig. 4c). Generally, the earth's crust parent materials and pedogenesis were considered to be the main sources of Cr, Ni, and Mn (Liu et al., 2020; Micó et al., 2006; Wang et al., 2019). However, several studies have indicated that Cr and Ni in topsoils were derived from industrial activities, including perennial mining, ore smelting, coal consumption, and steel production (Fei et al., 2022; Li et al., 2014; Yang et al., 2018). Meanwhile, iron ore, chromite, gold, and other mining activities were also potential sources of Cr and Ni in soils (Luo et al., 2010; Xue et al., 2000). The Cu, which contributed 45.86% in Factor 4, may be related to the smelting and processing of metals (Liu et al., 2020). Moreover, Cr and Ni were significantly correlated ($p < 0.05$), and Cu was also correlated with Ni. Furthermore, Cr, Ni, Cu, and Mn were negatively correlated with As that was initially determined as a geogenic product (Fig. 4c). Because the MF and MS samples located in the mining areas had high I_{geo} and EF values of Cr, Ni, and Cu (Fig. 2; Fig. S1), this factor would be assigned to mining activities.

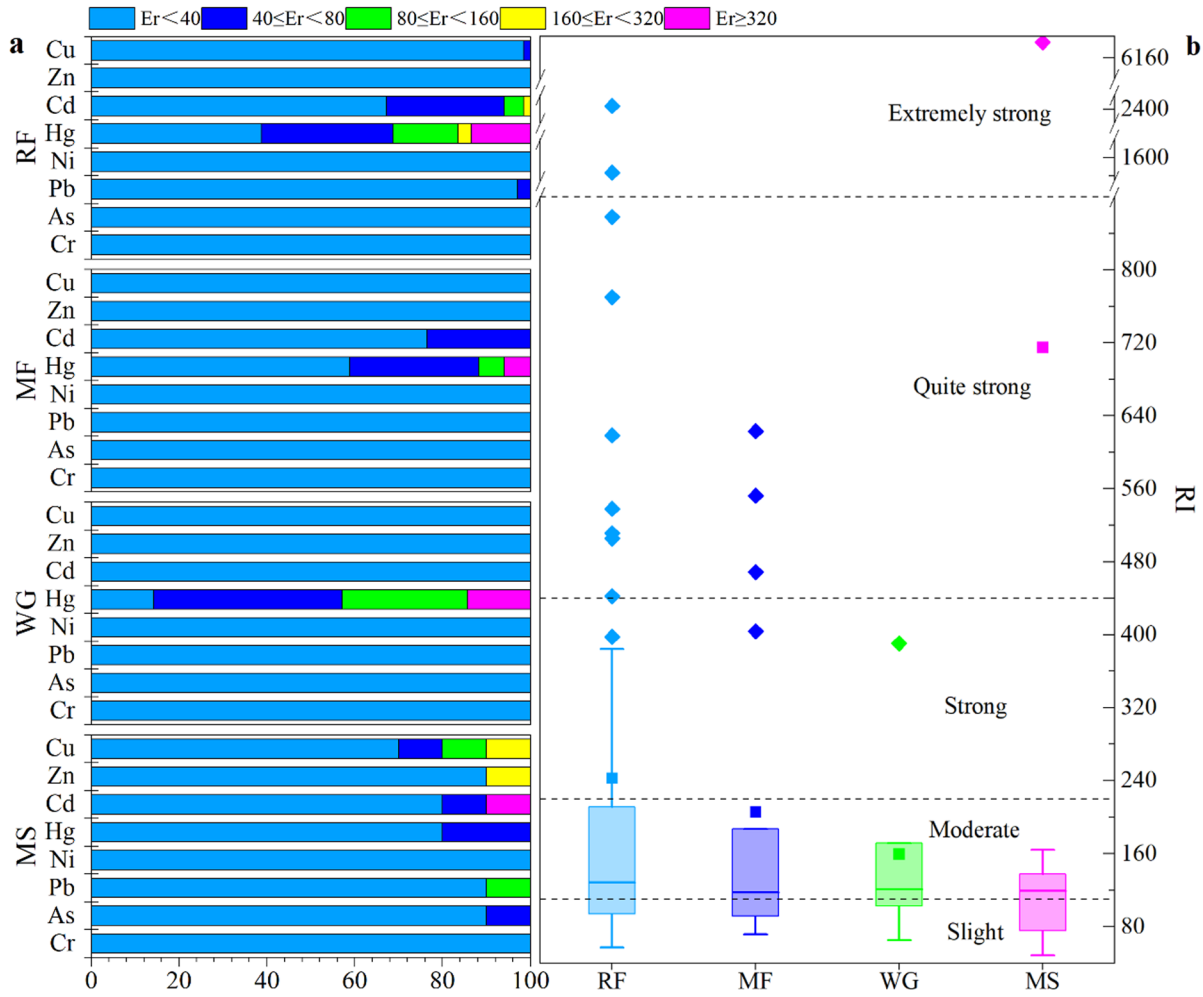

Fig. 5 Statistical analysis of Er and RI for HMs in topsoil samples

Potential ecological risks assessment

The above analyses suggested that the HMs contamination potentially posed threats to local human health. The Er values of Hg and Cd in the RF and MF were relatively higher than those of other HMs (Fig. 5a). Approximately 61.19% of Hg and 32.84% of Cd in the RF samples were at the moderate-risk level or higher, and 2.99% of Pb and 1.49% of Cu samples were at the same level. In addition, 41.18% of Hg and 23.53% of Cd in the MF samples were at a moderate-risk level or higher. Among them, the very-high-risk level (Er > 320) was found in both MF and RF samples. The Er values of Hg in the WG samples at the moderate-risk level or higher were up to 85.71%. To sum up, the potential ecological risk in the study area was mainly caused by Hg, followed by Cd.

For the total ecological risks, 62.69% of RF, 52.94% of MF, 71.43% of WG, and 50% of MS samples presented a moderate-risk level or higher according to the classification of RI (Fig. 5b, Table S2). In particular, 7 samples in RF and 3 samples in MF were at quite a strong risk level, and even 2 RF samples and 1 MS sample showed extremely strong risk. Hence, the potential ecological risks of HMs decreased in the order of RF > MF > WG > MS. Due to the high ecotoxicity, HMs can still pose high ecological risks even at a low contamination level (Chen et al., 2019; Huang et al., 2021; Li et al., 2021). According to the

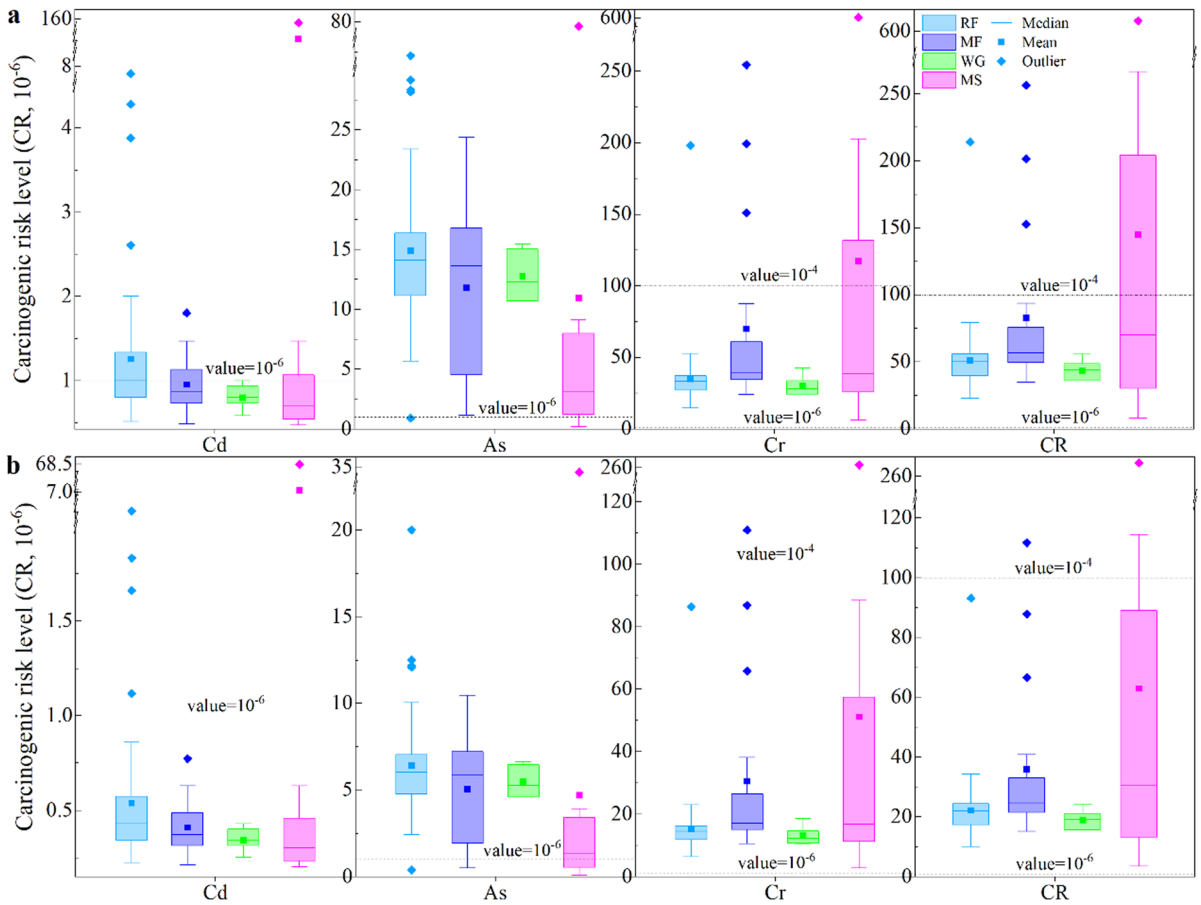

Fig. 6 Box diagram of carcinogenic risk value of HMs in topsoil samples: **a** and **b** represent the carcinogenic risk for children and adults of each land-use type

Er values of HMs, potential ecological risks were mainly caused by Hg and Cd.

Probabilistic health risks assessment

Except for two Pb samples for children, the maximum non-carcinogenic hazard quotient (HQ) of HMs (Cu, Pb, Zn, Ni, Hg, and Mn) in the samples for both children and adults in each land-use type showed an acceptable risk level (< 1) (Fig. S5). Additionally, the non-carcinogenic hazard index (HI) values of only 3 topsoil samples for children were greater than the unacceptable risk level (> 1) (Fig. S5a). Moreover, all samples for adults posed negligible health risks (Fig. S5b). Nonetheless, the HI values for children were higher than that for adults, ranging from 0.364 to 0.488, close to the unacceptable risk level (Table S5), in line with the previous studies (Huang et al., 2021; Long et al., 2021). As such, the non-carcinogenic risk for children generated by Pb, particularly in the area with frequent anthropogenic activities should be paid close attention.

The potential carcinogenic hazards to both children and adults caused by Cd, As, and Cr had high levels of risk ($CR > 10^{-6}$) (Fig. 6). The overall CR followed the order: Cr > As > Cd. Cr had the highest potential carcinogenic health risk level for both children and adults, especially in MF samples (Fig. 6). Cr (VI), classified as class A carcinogen (Park et al. 2004; WHO 2017), is a known source of carcinogens and toxicity and has been identified as a cause of gastric, liver, lung, oral, kidney, and urinary cancers (Tseng et al., 2018; Huang et al., 2021; Zhang et al., 2021). Meanwhile, the carcinogenic risk level caused

by As was the second highest, while the level of Cd was the lowest (was tolerable or close to acceptable, CR < 10^{-6}) irrespective of some RF samples. As can lead to chronic arsenic poisoning (WHO, 2017, 2020). Although the contaminations of Cr, As, and Cd were insignificant (Fig. 2), these HMs still pose significant carcinogenic hazards to both children and adults. The Pb and Ni were also classified as carcinogenic indicators by the IARC (2023) and WHO, but their carcinogenic risks were negligible (CR < 10^{-6}).

The health risks of HMs under different land-use types are shown in Table S5. The overall total carcinogenic risk level (TCR) of the land-use type followed the order: MS > MF > RF > WG, suggesting that the mining activities could result in health risks to the surrounding soils. The TCR of the HMs for children was greater than the negligible level (< 10^{-6}), and 8 topsoil samples exceeded the unacceptable risk values. The TCR of the HMs for adults was lower than that for children. The results of source-oriented HRA revealed that natural sources, mining activities, and natural sources should collectively cause significant impacts on the production of human health risks (Fig. 7). In addition to the natural sources, more proportion of agricultural activities contributed to the non-carcinogenic risk, whereas mining activities were more relevant to the carcinogenic risk. Consequently, when analyzing the comprehensive health risk assessment, especially for children, the potential carcinogenic health risks caused by exposure to Cr, As, and Cd should not be ignored. Therefore, targeted risk management and remediation strategies and actions must be implemented to avoid long-lasting health impacts on local ecology and residents (Zhang et al., 2021; Zhang et al., 2022).

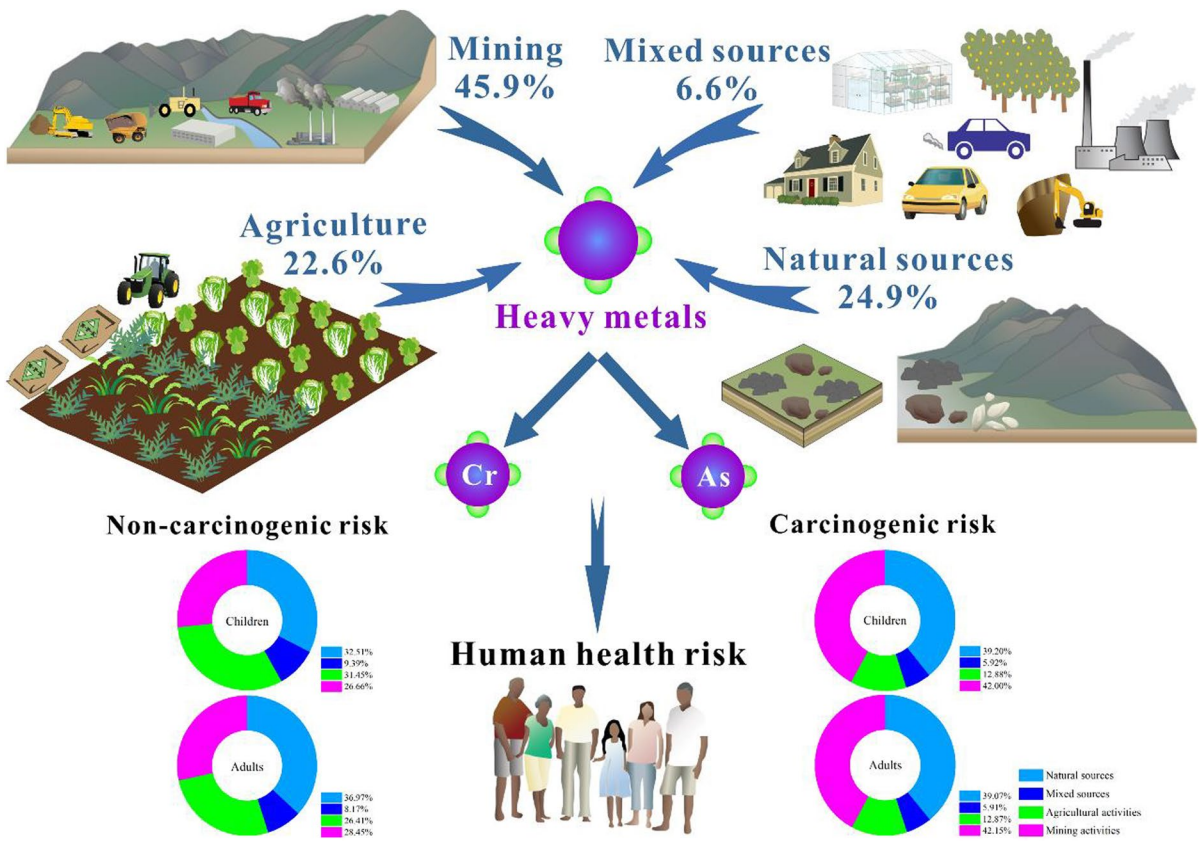

Fig. 7 Conceptual scheme of PMF results (illustrating the different sources of contamination) and the corresponding health risks (non-carcinogenic risk and carcinogenic risk)

Conclusion

The spatial distribution of the investigated HMs contents in the topsoil exhibited regional differences. The topsoil samples in the study area were largely contaminated with Hg and Cu as a result of the influences of anthropogenic activities. Comprehensive analysis combining with SOM method, PMF model, and Pearson correlation analysis suggested that the natural, mixed, agricultural, and mining sources were the pivotal driving forces for the accumulation of HMs, whereas mining activities accounted for the highest proportion among the four sources due to the long-term mining history. Owing to the high ecotoxicity, HMs can pose high ecological risks even at the low contamination level. The potential ecological risk in the study area was mainly caused by Hg, followed by Cd. As compared to the limited non-carcinogenic risk caused by agriculture activities, the potential carcinogenic health risks caused by mining activities should be monitored.

Acknowledgements Not applicable.

Author contributions Wanjun Jiang: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing-review& editing, Funding acquisition. Lishan Meng: Conceptualization, Investigation. Futian Liu: Investigation, Funding. Yizhi Sheng: Conceptualization, Methodology, Writing - original draft, Writing-review& editing, Supervision. Sheming Chen: Conceptualization, Writing - original draft, Funding, Investigation, Supervision. Jilong Yang: Investigation. Hairu Mao: Methodology. Jing Zhang: Investigation. Zhuo Zhang: Investigation. Hang Ning: Investigation.

Funding This work was financially supported by grants from the National Natural Science Foundation of China (41702262, 42102298 and 41972196) and China Geological Survey Program (DD20230431, DD20230426).

Availability of data and materials Not applicable.

Declarations

Competing interests The authors declare that there are no competing interests.

References

Agyeman, P. C., Kebonye, N. M., John, K., Borůvka, L., Vašát, R., & Fajemisim, O. (2022a). Prediction of nickel concentration in peri-urban and urban soils using hybridized empirical Bayesian kriging and support vector machine regression. *Scientific Report-UK, 12*(1), 1–16.

Agyeman, P. C., Kingsley, J. O. H. N., Kebonye, N. M., Ofori, S., Borůvka, L., Vašát, R., & Kočárek, M. (2022b). Ecological risk source distribution, uncertainty analysis, and application of geographically weighted regression cokriging for prediction of potentially toxic elements in agricultural soils. *Process Safety and Environmental, 164*, 729–746.

Árvay, J., Demková, L., Hauptvogl, M., Michalko, M., Bajčan, D., Stanovič, R., Tomáš, J., Hrstková, M., & Trebichalský, P. (2017). Assessment of environmental and health risks in former polymetallic ore mining and smelting area, Slovakia, spatial distribution and accumulation of mercury in four different ecosystems. *Ecotoxicology and Environmental Safety, 144*, 236–244.

Astel, A., Tsakovski, S., Barbieri, P., & Simeonov, V. (2007). Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Research, 41*(19), 4566–4578.

Baltas, H., Sirin, M., Gökbayrak, E., & Ozcelik, A. E. (2020). A case study on pollution and a human health risk assessment of heavy metals in agricultural soils around Sinop province Turkey. *Chemosphere, 241*, 125015.

Brinkman, G., Vance, G., Hannigan, M. P., & Milford, J. B. (2006). Use of synthetic data to evaluate positive matrix factorization as a source apportionment tool for PM2.5 exposure data. *Environmental Science Technology, 40*(6), 1892–1901.

Cai, L. M., Wang, Q. S., Wen, H. H., Luo, J., & Wang, S. (2019). Heavy metals in agricultural soils from a typical township in Guangdong Province, China, Occurrences and spatial distribution. *Ecotoxicology Environmental Safety, 168*, 184–191.

Chen, H. Y., Teng, Y. G., Lu, S. J., Wang, Y. Y., & Wang, J. S. (2015). Contamination features and health risk of soil heavy metals in China. *Science of the Total Environment, 512–513*, 143–153.

Chen, R., Chen, H., Song, L., Yao, Z., Meng, F., & Teng, Y. (2019). Characterization and source apportionment of heavy metals in the sediments of Lake Tai (China) and its surrounding soils. *Science of the Total Environment, 694*, 445–486.

Cheng, W., Lei, S., Bian, Z., Zhao, Y., Li, Y., & Gan, Y. (2020). Geographic distribution of heavy metals and identification of their sources in soils near large, open-pit coal mines using Positive Matrix Factorization. *Journal of Hazardous Materials, 387*, 121666.

Cheng, Z., Chen, L. J., Li, H. H., Lin, J. B., Yang, Z. B., Yang, Y. X., Xu, X. X., Xian, J. R., Shao, J. R., & Zhu, X. M. (2018). Characteristics and health risk assessment of heavy metals exposure via household dust from the urban area in Chengdu China. *Science Total Environmental, 619–620*, 621–629.

CMEE. (2018). *Ministry of Ecology and Environment of P.R. China. Soil environmental quality risk control standard for soil contamination of agricultural land (GB15618–2018)*. Beijing: China Environmental Science Press, 1–7 (in Chinese)

Csavina, J., Field, J., Taylor, M. P., Gao, S., Landázuri, A., Betterton, E. A., & Sáez, A. E. (2012). A review on the importance of metals and metalloids in atmospheric dust and aerosol from mining operations. *Science of the Total Environment, 433*, 58–73.

Davis, H. T., Aelion, C. M., McDermott, S., & Lawsonc, A. B. (2009). Identifying natural and anthropogenic sources of metals in urban and rural soils using GIS-based data, PCA, and spatial interpolation. *Environmental Pollution, 157*, 2378–2385.

Dong, H., Lin, Z., & Wan, X. (2017). Risk assessment for the mercury polluted site near a pesticide plant in Changsha, Hunan, China. *Chemosphere, 169*, 333–341.

Duzgoren-Aydin, N. S., Li, X. D., & Wong, S. C. (2004). Lead contamination and isotope signatures in the urban environment. *Environmental International, 30*, 209–217.

Facchinelli, A., Sacchi, E., & Mallen, L. (2001). Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. *Environmental Pollution, 114*(3), 313–324.

Fei, X. F., Lou, Z. H., Xiao, R., Ren, Z. Q., & Lv, X. N. (2022). Source analysis and source-oriented risk assessment of heavy metal pollution in agricultural soils of different cultivated land qualities. *Journal of Cleaner Production, 341*, 130942.

Gu, Y. G., Lin, Q., & Gao, Y. P. (2017). Metals in exposed-lawn soils from 18 urban parks and its human health implications in southern China's largest city Guangzhou. *Journal of Cleaner Production, 163*, 164–171.

Hakanson, L. (1980). An ecological risk index for aquatic pollution control A Sedimentological Approach. *Water Research, 14*, 975–1001.

Haselbeck, V., Kordilla, J., Krause, F., & Sauter, M. (2019). Self-organizing maps for the identification of groundwater salinity sources based on hydrochemical data. *Journal of Hydrology, 576*, 610–619.

Hu, W., Wang, H., Dong, L., Huang, B., Borggaard, O. K., Hansen, H. C. B., He, Y., & Holm, P. E. (2018). Source identification of heavy metals in peri-urban agricultural soils of southeast china: An integrated approach. *Environmental Pollution, 237*, 650.

Huang, J. L., Wu, Y. Y., Sun, J. X., Li, X., Geng, X. L., Zhao, M. L., Sun, T., & Fan, Z. Q. (2021). Health risk assessment of heavy metal(loid)s in park soils of the largest megacity in China by using Monte Carlo simulation coupled with Positive matrix factorization model. *Journal of Hazardous Materials, 415*, 125629.

IARC. (2023). *Agents classified by the IARC monographs* (Vol. 1–133). International Agency for Research on Cancer. https://monographs.iarc.who.int/agents-classified-by-the-iarc/. Accessed 5 May 2023.

Jiang, G. B., Shi, J. B., & Feng, X. B. (2006). Mercury pollution in China. *Environmental Science and Technology, 40*, 3672–3678.

Jiang, H. H., Cai, L. M., Wen, H. H., Hu, G. C., Chen, L. G., & Luo, J. (2020). An integrated approach to quantifying ecological and human health risks from different sources of soil heavy metals. *Science of the Total Environment, 701*, 134466.

Jiang, W. J., Liu, H. W., Sheng, Y. Z., Ma, Z., Zhang, J., Liu, F. T., Chen, S. M., Meng, Q. H., & Bai, Y. N. (2022). Distribution, source apportionment, and health risk assessment of heavy metals in groundwater in a multi-mineral resource area North China. *Exposure and Health*. 14, 807–827.

Jiang, W., Sheng, Y., Liu, H., Ma, Z., Song, Y., Liu, F., & Chen, S. (2022b). Groundwater quality assessment and hydrogeochemical processes in typical watersheds in Zhangjiakou region, northern China. *Environmental Science and Pollution Research., 29*(3), 3521–3539.

Jin, Y. L., O'Connor, D., Ok, Y. S., Tsang, D. C. W., Liu, A., & Hou, D. Y. (2019). Assessment of sources of heavy metals in soil and dust at children's playgrounds in Beijing using GIS and multivariate statistical analysis. *Environment International, 124*, 320–328.

Kadi, M. W. (2009). Soil pollution hazardous to environment, a case study on the chemical composition and correlation to automobile traffic of the roadside soil of Jeddah city, Saudi Arabia. *Journal of Hazardous Materials, 168*, 1280–1283.

Kebonye, N. M., Eze, P. N., John, K., Gholizadeh, A., Dajčl, J., Drábek, O., Němeček, K., & Borůvka, L. (2021). Self-organizing map artificial neural networks and sequential Gaussian simulation technique for mapping potentially toxic element hotspots in polluted mining soils. *Journal of Geochemical Exploration, 222*, 106680.

Kim, K. H., Yun, S. T., Yu, S. Y., Choi, B. Y., Kim, M. J., & Lee, K. J. (2020). Geochemical pattern recognitions of deep thermal groundwater in South Korea using self-organizing map: Identified pathways of geochemical reaction and mixing. *Journal of Hydrology, 589*, 125202.

Lee, K. J., Yun, S. T., Yu, S., Kim, K. H., Lee, J. H., & Lee, S. H. (2019). The combined use of self-organizing map technique and fuzzy c-means clustering to evaluate urban groundwater quality in Seoul metropolitan city, South Korea. *Journal of Hydrology, 569*, 685–697.

Lehmann, J., Bossio, D. A., Kögel-Knabner, I., & Rillig, M. C. (2020). The concept and future prospects of soil health. *Nature Reviews Earth & Environment, 1*(10), 544–553.

Li, J. T., Qiu, J. W., Wang, X. W., Zhong, Y., Lan, C. Y., & Shu, W. S. (2006). Cadmium contamination in orchard soils and fruit trees and its potential health risk in Guangzhou China. *Environmental Pollution, 143*, 159–165.

Li, J., Wang, G. C., Liu, F., Cui, L. F., & Jiao, Y. J. (2021). Supplementary material source apportionment and ecological-health risks assessment of heavy metals in topsoils near a factory, central China. *Expo Health, 13*, 79–92.

Li, X. D., Lee, S. L., Wong, S. C., Shi, W., & Thornton, I. (2004). The study of metal contamination in urban soils of Hong Kong using a GIS-based approach. *Environmental Pollution, 129*, 113–124.

Li, Y., Yuan, Y., Sun, C., Sun, T., Liu, X., Li, J., et al. (2020). Heavy metals in soil of an urban industrial zone in a metropolis: Risk assessment and source apportionment. *Stochastic Environmental Research and Risk Assessment, 34*, 435–446.

Li, Z. Y., Ma, Z. W., Kuijp, T. J., Yuan, Z. W., & Huang, L. (2014). A review of soil heavy metal pollution from mines in China, pollution and health risk assessment. *Science of the Total Environment, 468–469*, 843–853.

Lin, Y., Ma, J., Zhang, Z. D., Zhu, Y. F., Hou, H., Zhao, L., Sun, Z. J., Xue, W. J., & Shi, H. D. (2018). Linkage

between human population and trace elements in soils of the Pearl River Delta: Implications for source identification and risk assessment. *Science of the Total Environment, 610–611*, 944–950.

Liu, L. L., Liu, Q. Y., Ma, J., Wu, H. W., Qu, Y. J., Gong, Y. W., Yang, S. H., An, Y. F., & Zhou, Y. Z. (2020). Heavy metal(loid)s in the topsoil of urban parks in Beijing, China, concentrations, potential sources, and risk assessment. *Environmental Pollution, 260*, 114083.

Long, Z. J., Zhu, H., Bing, H. J., Tian, X., Wang, Z. G., Wang, X. F., & Wu, Y. H. (2021). Contamination, sources and health risk of heavy metals in soil and dust from different functional areas in an industrial city of Panzhihua City Southwest China. *Journal of Hazardous Materials, 420*, 126638.

Lu, A. X., Wang, J. H., Qin, X. Y., Wang, K. Y., Han, P., & Zhang, S. Z. (2012). Multivariate and geostatistical analyses of the spatial distribution and origin of heavy metals in the agricultural soils in Shunyi, Beijing, China. *Science of the Total Environment, 425*, 66–74.

Luo, L., Ma, Y. B., Zhang, S. Z., Wei, D. P., & Zhu, Y. G. (2009). An inventory of trace element inputs to agricultural soils in China. *Journal of Environmental Management, 90*, 2524–2530.

Luo, W., Lu, Y. L., Zhang, Y., Fu, W. Y., Wang, B., Jiao, W. T., Wang, G., Tong, X. J., & Giesy, J. P. (2010). Watershed-scale assessment of arsenic and metal contamination in the surface soils surrounding Miyun Reservoir, Beijing China. *Journal of Environmental Management, 91*, 2599–2607.

Madrid, L., Diaz-Barrientos, E., & Madrid, F. (2002). Distribution of heavy metal contents of urban soils in parks of Seville. *Chemosphere, 49*, 1301–1308.

Man, Y. B., Sun, X. L., Zhao, Y. G., Lopez, B. N., Chung, S. S., Wu, S. C., Cheung, K. C., & Wong, M. H. (2010). Health risk assessment of abandoned agricultural soils based on heavy metal contents in Hong Kong: The world's most populated city. *Environment International, 36*, 570–576.

Mao, H. R., Wang, G. C., Liao, F., Shi, Z. M., Zhang, H. Y., Chen, X. L., Qiao, Z. Y., Li, B., & Bai, Y. F. (2023). Spatial variability of source contributions to nitrate in regional groundwater based on the positive matrix factorization and Bayesian model. *Journal of Hazardous Materials, 445*, 130569.

Mao, H. R., Wang, G. C., Rao, Z., Liao, F., Shi, Z. M., Huang, X. J., Chen, X. L., & Yang, Y. (2021). Deciphering spatial pattern of groundwater chemistry and nitrogen pollution in Poyang Lake Basin (eastern China) using self-organizing map and multivariate statistics. *Journal of Cleaner Production, 329*, 129697.

Marrugo-Negrete, J., Pinedo-Hernandez, J., & Díez, S. (2017). Assessment of heavy metal pollution, spatial distribution and origin in agricultural soils along the Sinú River Basin Colombia. *Environmental Research, 154*, 380–388.

Nakagawa, K., Yu, Z., Berndtsson, R., & Hosono, T. (2020). Temporal characteristics of groundwater chemistry affected by the 2016 Kumamoto earthquake using self-organizing maps. *Journal of Hydrology, 582*, 124519.

Nan, Z. R., Li, J. J., Zhang, J. M., & Cheng, G. D. (2002). Cadmium and zinc interactions and their transfer in soil crop system under actual field conditions. *Science of the Total Environment, 285*, 187–195.

Nicholson, F. A., Smith, S. R., Alloway, B. J., Carlton-Smith, C., & Chambers, B. J. (2003). An inventory of heavy metals inputs to agricultural soils in England and Wales. *Science of the Total Environment, 311*, 205–219.

Nogueirol, R. C., Alleoni, L. R. F., Nachtigall, G. R., & de Melo, G. W. (2010). Sequential extraction and availability of copper in Cu fungicide-amended vineyard soils from Southern Brazil. *Journal of Hazardous Materials, 181*, 931–937.

Norris, G.A., Duvall, R., Brown, S., Bai, S., 2014. EPA PositiveMatrix Factorization (PMF) 5.0 Fundamentals and User Guide.

Pacyna, E. G., Pacyna, J. M., Sundseth, K., Munthe, J., Kindbom, K., Wilson, S., Steenhuisen, F., & Maxson, P. (2010). Global emission of mercury to the atmosphere from anthropogenic sources in 2005 and projections to 2020. *Atmospheric Environment, 44*, 2487–2499.

Park, R. M., Bena, J. F., Stayner, L. T., Smith, R. J., Gibb, H. J., & Lees, P.S. (2004). Hexavalent chromium and lung cancer in the chromate industry: A quantitative risk assessment. *Risk Analysis: An International Journal*, *24*(5), 1099–1108.

Pearce, A. R., Rizzo, D. M., & Mouser, P. J. (2011). Subsurface characterization of groundwater contaminated by landfill leachate using microbial community profile data and a nonparametric decision-making process. *Water Resources Research, 47*, W06511.

Qu, S., Shi, Z. M., Liang, X. Y., Wang, G. C., & Han, J. Q. (2021). Multiple factors control groundwater chemistry and quality of multi-layer groundwater system in Northwest China coalfield-Using self-organizing maps (SOM). *Journal of Geochemical Exploration*, *227*, 106795.

Rodriguez, J. A., Nanos, N., Grau, J. M., Gil, L., & Lopez-Arias, M. (2008). Multiscale analysis of heavy metal contents in Spanish agricultural topsoils. *Chemosphere, 70*, 1085–1096.

Šajn, R., Halamić, J., Peh, Z., Galović, L., & Alijagić, J. (2011). Assessment of the natural and anthropogenic sources of chemical elements in alluvial soils from the Drava River using multivariate statistical methods. *Journal of Geochemical Exploration, 110*(3), 278–289.

Salim, I., Sajjad, R. U., Paule-Mercado, M. C., Memon, S. A., Lee, B. Y., Sukhbaatar, C., & Lee, C. H. (2019). Comparison of two receptor models PCA-MLR and PMF for source identification and apportionment of pollution carried by runoff from catchment and sub-watershed areas with mixed land cover in South Korea. *Science of the Total Environment, 663*, 764–775.

Sun, C. Y., Liu, J. S., Wang, Y., Sun, L. Q., & Yu, H. W. (2013). Multivariate and geostatistical analyses of the spatial distribution and sources of heavy metals in agricultural soil in Dehui, Northeast China. *Chemosphere, 92*, 517–523.

Sun, L., Guo, D., Liu, K., Meng, H., Zheng, Y., Yuan, F., & Zhu, G. (2019). Levels, sources, and spatial distribution of heavy metals in soils from a typical coal industrial city of Tangshan, China. *CATENA, 175*, 101–109.

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.

Tomiyasu, T., Kodamatani, H., Imura, R., Matsuyama, A., Miyamoto, J., Akagi, H., Kocman, D., & Kotnik, Jo.ž., Fajon, V., Horvat, M.,. (2017). The dynamics of mercury near Idrija mercury mine, Slovenia, horizontal and vertical distributions of total, methyl, and ethyl mercury concentrations in soils. *Chemosphere, 184*, 244–252.

Tseng, C. H., Lei, C., & Chen, Y. C. (2018). Evaluating the health costs of oral hexavalent chromium exposure from water pollution: A case study in Taiwan. *Journal of Cleaner Production, 172*, 819–826.

Ungureanu, T., Iancu, G. O., Pintilei, M., & Chicoș, M. M. (2017). Spatial distribution and geochemistry of heavy metals in soils, a case study from the NE area of Vaslui county Romania. *Journal of Geochemical Exploration, 176*, 20–32.

USEPA. (2011). *Exposure factors handbook*, Final ed. Washington DC: U.S. Environment Protection Agency.

USEPA. (2014). *EPA positive matrix factorization (PMF) 5.0 fundamentals and user guide*. Washington DC: U.S. Environment Protection Agency.

Wang, S., Cai, L. M., Wen, H. H., Luo, J., Wang, Q. S., & Liu, X. (2019). Spatial distribution and source apportionment of heavy metals in soil from a typical county-level city of Guangdong Province, China. *Science of the Total Environment, 655*, 92–101.

Wang, Z., Xiao, J., Wang, L., Liang, T., Guo, Q., Guan, Y., & Rinklebe, J. (2020). Elucidating the differentiation of soil heavy metals under different land uses with geographically weighted regression and self-organizing map. *Environmental Pollution, 260*, 114065.

Wei, J., Zheng, X. G., Zhang, Y. X., Wang, C. L., & Liu, J. T. (2019). Geochemical characteristic of soil in Zhangxuan grape production areas of Heibei province. *Chinese Journal of Agricultural Resources and Regional Planning., 40*(4), 65–69. (in Chinese).

WHO. (2017). *Guidelines for drinking-water quality: Fourth edition incorporating first Addendum* (4th Edn + 1st Add). World Health Organization. https://apps.who.int/iris/handle/10665/254637.

WHO. (2020). *Lead poisoning and health*. World Health Organization.

Wu, S., Xia, X. H., Lin, C. Y., Chen, X., & Zhou, C. H. (2010). Levels of arsenic and heavy metals in the rural soils of Beijing and their changes over the last two decades (1985–2008). *Journal of Hazardous Materials, 179*(1–3), 860–867.

Xiao, R., Wang, S., Li, R., Wang, J. J., & Zhang, Z. (2017). Soil heavy metal contamination and health risks associated with artisanal gold mining in Tongguan, Shaanxi, China. *Ecotoxicology and Environmental Safety, 141*, 17–24.

Xue, H. B., Sigg, L., & Gächter, R. (2000). Transport of Cu, Zn and Cd in a small agricultural catchment. *Water Research, 34*, 2558–2568.

Yang, Q. Q., Li, Z. Y., Lu, X. N., Duan, Q. N., Huang, L., & Jun, B. J. (2018). A review of soil heavy metal pollution from industrial and agricultural regions in China: Pollution and risk assessment. *Science of the Total Environment, 642*, 690–700.

Zanotti, C., Rotiroti, M., Fumagalli, L., Fumagalli, G. A., Canonaco, F., Stefenelli, G., Prévôt, A. S. H., Leoni, B., & Bonomi, T. (2019). Groundwater and surface water quality characterization through positive matrix factorization combined with GIS approach. *Water Research, 159*(1), 122–134.

Zhang, B., Liu, J., Sheng, Y., Shi, J., & Dong, H. (2021). Disentangling microbial syntrophic mechanisms for hexavalent chromium reduction in autotrophic biosystems. *Environmental Science & Technology., 55*(9), 6340–6351.

Zhang, H., Li, A., Wei, Y., Miao, Q., Xu, W., Zhao, B., Guo, Y., Sheng, Y., & Yang, Y. (2022). Development of a new methodology for multifaceted assessment, analysis, and characterization of soil contamination. *Journal of Hazardous Materials., 6*, 129542.

Zhang, R., Chen, T., Zhang, Y., Hou, Y., & Chang, Q. (2020). Health risk assessment of heavy metals in agricultural soils and identification of main influencing factors in a typical industrial park in northwest China. *Chemosphere, 252*, 126591.

Zhang, X., Wei, S., Sun, Q., Wadood, S. A., & Guo, B. (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.

Zhu, G., Wu, X., Ge, J., Liu, F., Zhao, W., & Wu, C. (2020). Influence of mining activities on groundwater hydrochemistry and heavy metal migration using a self-organizing map (SOM). *Journal of Cleaner Production, 257*, 120664.

Zuo, L., Lu, X. W., Fan, P., Wang, L. Q., Yu, B., Lei, K., Yang, Y. F., & Chen, Y. R. (2022). Concentrations, sources and ecological -health risks of potentially toxic elements in finer road dust from a megacity in north China. *Journal of Cleaner Production, 358*, 132036.