# Health risk assessment of heavy metals in coal mine soils of Northwest China postprint

**Authors:** LI Yun, ZHUANG Zhong, XIA Qianrou, SHI Qingdong, ZHU Jiawei, WANG Peijuan, LI Dinghao, Yryszhan ZHAKYPBEK, Serik TURSBEKOV, XIA Qianrou

**Date:** 2025-07-17T00:00:00+00:00

## Abstract

## Abstract

Coal mining readily leads to heavy metal (HM) accumulation in soils, adversely impacting both ecological environments and human health, particularly in extremely arid and fragile regions. This study collected soil samples from the Heishan open-pit coal mine in Turpan City, northwestern China, to determine the health risks associated with heavy metals (HMs). The results demonstrate that the Positive Matrix Factorization (PMF) model apportioned soil heavy metal sources into four categories: natural sources and livestock farming (43.46%), industrial and traffic emissions (22.87%), fossil fuel combustion (10.64%), and atmospheric deposition combined with domestic pollution (23.03%). Multiple pollution assessment indices consistently revealed significant cadmium (Cd) and lead (Pb) contamination.

Monte Carlo simulation-based health risk assessment indicated that children face a 4.00% non-carcinogenic risk and a 12.00% carcinogenic risk; PMF-based health risk assessment further revealed that fossil fuel combustion contributes most substantially to health risks for both adults and children, whereas industrial and traffic sources contribute the least. Through source apportionment analysis, this study evaluated heavy metal risks in mining area soils, providing not only reliable data support for preventing and controlling soil heavy metal pollution in this arid mining region, but also establishing a theoretical foundation for future regional research.

# Full Text

## Preamble

**Journal of Arid Land (2025) 17(7): 933–957**
doi: 10.1007/s40333-025-0023-9; CSTR: 32276.14.JAL.02500239
Science Press Springer-Verlag

### Health Risk Assessment of Heavy Metals in Coal Mine Soils of Northwest China

LI Yun[1,2], ZHUANG Zhong[3], XIA Qianrou[4*], SHI Qingdong[1,2], ZHU Jiawei[1,2], WANG Peijuan[1,2], LI Dinghao[1,2], Yryszhan ZHAKYPBEK[5], Serik TURSBEKOV[5]

[1] College of Ecology and Environment, Xinjiang University, Urumqi 830046, China
[2] Key Laboratory of Oasis Ecology, Ministry of Education (Xinjiang University), Urumqi 830046, China
[3] College of Resources and Environment Sciences, China Agricultural University, Beijing 100193, China
[4] Xinjiang Academy of Ecological and Environmental Sciences, Urumqi 830011, China
[5] Department of Mine Surveying and Geodesy, Institute Mining and Metallurgical Institute named after O.A. Baikonurov, Satbayev University, Almaty 050013, Kazakhstan

**Abstract:** Coal mining activities predispose soils to heavy metal (HM) accumulation, adversely affecting ecological environments and human health, particularly in extremely arid and vulnerable regions. This study investigated soil samples from the Black Mountain Open Pit Coal Mine in Turpan City, Northwest China, to assess the health risks posed by heavy metals. The positive matrix factorization (PMF) model identified four distinct sources of soil HMs: natural and livestock sources (43.46%), industrial transportation (22.87%), fossil fuel combustion (10.64%), and atmospheric deposition combined with domestic pollution (23.03%). Multiple pollution evaluation indices consistently indicated significant cadmium (Cd) and lead (Pb) contamination. Monte Carlo simulation revealed that children faced a 4.00% probability of non-carcinogenic risk and a 12.00% probability of carcinogenic risk, while PMF-based health risk assessment demonstrated that fossil fuel combustion contributed most significantly to health risks for both adults and children, whereas industrial transportation contributed the least. This source-oriented analysis of HM risks in mining area soils provides reliable data support for pollution prevention and control in this arid mining region and establishes a theoretical foundation for subsequent regional research.

**Keywords:** arid area; soil heavy metals; positive matrix factorization; Monte Carlo simulation; health risk assessment

**Citation:** LI Yun, ZHUANG Zhong, XIA Qianrou, SHI Qingdong, ZHU Jiawei, WANG Peijuan, LI Dinghao, Yryszhan ZHAKYPBEK, Serik TURSBEKOV. 2025. Health risk assessment of heavy metals in coal mine soils of Northwest China. *Journal of Arid Land*, 17(7): 933–957. https://doi.org/10.1007/s40333-025-0023-9; https://cstr.cn/32276.14.JAL.02500239

## 1 Introduction

Mining activities frequently generate numerous ecological challenges, among which heavy metal pollution (HMP) in soils is particularly prominent (Yang et al., 2018). Heavy metals affect the development and growth of soil flora and fauna through interactions with soil microorganisms and subsequently enter the human body via the food chain, posing considerable threats to human health (Zhao et al., 2019; Liu et al., 2023). Previous studies have confirmed that soils in mining areas of southern China exhibit elevated levels of lead (Pb), cadmium (Cd), and arsenic (As) that exceed regulatory standards, endangering surrounding soils and water bodies (Cao et al., 2018; Li et al., 2019). While these studies have revealed the hazards of coal mining to soils and ecosystems, they have focused primarily on humid or sub-humid regions. In contrast, arid regions feature more fragile ecological environments characterized by limited soil moisture and sparse vegetation. As a typical arid region in China, the Xinjiang Uygur Autonomous Region has received limited research attention regarding HMP, yet may face more complex and severe contamination problems (Qi et al., 2022). Consequently, identifying the sources of HMs in and around soils in arid mining areas is essential.

Commonly recognized methods for HM source identification include positive matrix factorization (PMF) (Tian et al., 2018; Zhou et al., 2024), exploratory factor analysis (EFA) (Specht et al., 2025), chemical mass balance (CMB) modeling (Mi et al., 2023), and the Unmix model (Liao et al., 2021). Owing to its high flexibility and adaptability, the PMF model is particularly valuable for identifying soil HM sources (Xu et al., 2021) and is recommended by the US Environmental Protection Agency (EPA) (US EPA, 2015). This model can accurately analyze complex environmental data and clearly reveal the proportional contributions of different pollution sources, while also being adaptable and optimizable for various research needs (Guan et al., 2018; Anaman et al., 2022). The PMF model can thus effectively identify and analyze soil HMs in arid mining areas, providing a scientific basis for ecological environmental protection.

Various evaluation indices and methods enable the development of more targeted strategies for ecological protection in arid mining areas. Indicators such as the Nemerow integrated pollution index (NIPI) (Hu et al., 2018a; Zhou et al., 2024), geoaccumulation index (Igeo) (Zhuang et al., 2018), and enrichment factor (EF) (Wang et al., 2023) have been widely employed to assess HMP in mining soils. The human health risk assessment (HRA) model developed by the US EPA represents the most commonly used method for evaluating soil HM health risks (US EPA, 2013). However, due to uncertainties in HM concentrations, regional differences, and fluctuations in exposure parameters, traditional soil HM risk assessments may overestimate or underestimate actual risk levels (Liu et al., 2023). Since typical soil contamination involves the co-coupling of multiple HMs, single analytical methods cannot comprehensively characterize such complex pollution. Therefore, this study employed probabilistic risk estimation using Monte Carlo simulation to enhance the accuracy of the HRA model (Yang et al., 2019; Li et al., 2023) and combined PMF and HRA models to more accurately identify HM sources requiring prioritized control (Shen et al., 2024).

Turpan City in Northwest China possesses abundant mineral reserves and experiences frequent industrial manufacturing activities. However, previous studies have not adequately addressed soil HMP in and around mining areas or the associated health risks. This study investigated the concentrations and distributions of HMs in soils within and around the Black Mountain Open Pit Coal Mine in Turpan City. The research objectives were: (1) to identify soil HM sources using the PMF model; (2) to analyze pollution levels of soil HMs in and around the mining area using various pollution indices; and (3) to enhance HRA precision through Monte Carlo simulation and determine human health risks from different sources by integrating source analysis with HRA. The results may provide theoretical guidance for regional ecological protection and data support for HMP prevention and management in arid mining areas.

---

## 2.1 Study Area

The Black Mountain Open Pit Coal Mine (43°12 30 –43°16 00 N, 87°30 00 – 87°38 30 E) is located in Turpan City, Xinjiang, China, situated in a valley among the northern Tianshan Mountains, northern Yokakeng Aidai Mountains, and southern Mordelok and Black Mountains. The topography is relatively open in the east and west, with higher elevations in the north and lower elevations in the south and west. Altitudes range from 2365 to 3023 m above sea level. The study area experiences an arid climate with sparse vegetation and is highly sensitive to soil erosion while being mildly sensitive to land desertification. Gullies and ravines are relatively well-developed throughout the area and remain dry year-round, with short-term flooding occurring only during heavy rainfall events.

Total coal mine production reached $13\times10^{6}$ t/a in 2023 (National Energy Xinjiang Toksun Energy Co., Ltd., 2022). The primary ecological challenges in the area include high-intensity human activities, simple ecosystem structure, declining ecological functions, landform destruction caused by open-pit coal mining, severe damage to grassland vegetation, and exposed ground surfaces that provide sand sources for sandstorms, posing serious threats to regional ecological security (Department of Ecology and Environment of Xinjiang Uygur

Autonomous Region, 2022).

---

## 2.2 Soil Sampling and Preparation

In August 2023, 61 soil samples were collected from the open-pit mine area and surrounding regions, including discharge areas, coal mining zones, transportation roads, and village vicinities [Figure 1: see original paper]. At each sampling site, three 0–20 cm soil samples were taken, mixed, encapsulated in polyethylene bags, and numbered, with each sample weighing 1 kg.

All collected samples were transported to the laboratory for pretreatment. Soils were air-dried, and gravel, plant roots, and other debris were removed. Soils were then ground and passed through a 100-mesh nylon sieve. Each sample was divided into two portions and placed in self-sealing bags: one for measurement and the other stored as a reserve (Zhang et al., 2024b). Metal contamination of soil samples was avoided throughout the entire process.

Concentrations of cadmium (Cd), chromium (Cr), copper (Cu), iron (Fe), manganese (Mn), nickel (Ni), lead (Pb), zinc (Zn), and arsenic (As), along with soil pH, were investigated. HM concentrations were determined using inductively coupled plasma-mass spectrometry (MEE, 2023), and soil pH was measured using the potentiometric method (MEE, 2018a). HM concentrations and soil pH results are presented in Table S1.

---

### 2.3.1 Multivariate Statistical Analysis

This study employed correlation analysis to determine relationships among HMs and exploratory factor analysis (EFA) for preliminary HM classification, both performed using SPSS v.23.0 software. Specific criteria for these methods are outlined in Table 1 .

---

### 2.3.2 PMF Model

To better identify HM sources in the mining area and surrounding soils, we applied the PMF source analysis model to quantitatively analyze nine HM elements. PMF, a positive definite matrix factorization-based model developed by the US EPA for analyzing pollutant sources and contributions in complex environmental media, is now widely used for source analysis in other environmental media such as soil and water (US EPA, 2015; Yang et al., 2023). This method does not require determination of complex source compositional profiles and is straightforward to operate (Guan et al., 2018). The basic formulae are expressed in Equations 1 and 2:

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij}$$

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left( \frac{e_{ij}}{u_{ij}} \right)^2$$

where $x_{ij}$ is the concentration of the $j$th HM measured in the $i$th sample (mg/kg); $g_{ik}$ is the concentration contribution of the $k$th factor to the $i$th sample (mg/kg); $f_{kj}$ is the contribution of the $j$th element in the $k$th factor; $e_{ij}$ is the residual of the $j$th HM in the $i$th sample; $Q$ is the weighted sum of squared residuals; $n$ is the number of samples; $m$ is the number of HM species; $p$ is the number of factors contributing to HM content at the sampling point; and $u_{ij}$ is the uncertainty of the $j$th element in the $i$th sample.

Uncertainty directly affects the calculated weights of sample contaminant mass concentration and component concentration data in the PMF model input (Norris et al., 2014), and is calculated using Equations 3 and 4:

$$u_{ij} = \frac{5}{6} \times \mathrm{MDL} \quad \text{when } x_{ij} \leq \mathrm{MDL}$$

$$u_{ij} = \sqrt{(\delta \times x_{ij})^2 + (\mathrm{MDL})^2} \quad \text{when } x_{ij} > \mathrm{MDL}$$

where $u_{ij}$ is the uncertainty of the $j$th HM in the $i$th sample; $x_{ij}$ is the concentration of the $j$th HM measured in the $i$th sample; MDL is the method detection limit; and $\delta$ is the error factor of the test method. When $x_{ij} \leq$ MDL, Equation 3 was used; when $x_{ij} >$ MDL, Equation 4 was used. Error coefficients generally ranged from 0.1 to 0.6, with larger values used for unstable pollutant mass concentrations, component concentrations, or values near the detection limit. When pollutant mass concentration or component concentration data were missing, a larger error factor could be assigned (Norris et al., 2014). In this study, measured HM mass concentrations were relatively stable; therefore, the error factor was set to 0.1. MDLs for the nine HMs are shown in Table 2 .

Data processing was performed using EPA PMF v.5.0 software. Signal-to-noise (S/N) ratios for all elements in the model ranged from 8.3 to 9.0. The number of factors is a key parameter in PMF modeling when seeking optimal solutions, though no direct algorithm exists for this determination (Norris et al., 2014). The number of factors for this study was determined to be four (Karakas et al., 2017; Zhang et al., 2024a). With four factors, the model ran 20 iterations using a seed number of 8. Additionally, 77.00% of HM elements had residuals exceeding 54.00%, and 84.00% of HM elements had residuals ranging from -3 to 3. Factor selection results were tested using bootstrap (BS) and displacement (DISP) methods within the PMF model. In the BS test, mapping exceeded

80.00%, indicating that uncertainty in BS could be explained and that the number of factors was appropriate. In DISP, the absence of factor swapping for the minimum variation of $Q$ in the model indicated excellent robustness of the results.

## 2.4 Pollution Index Assessment

This study employed NIPI, Igeo, and EF to assess HM pollution in the coal mining area. Classification criteria for NIPI, Igeo, and EF are shown in Table 3 .

NIPI (Equation 5) (Nemerow, 1974; Liu et al., 2021) was calculated using the one-factor pollution index (PI) method (Hakanson, 1980) (Equation 6), which emphasizes the role of relatively contaminated HM pollutants:

$$\mathrm{NIPI} = \sqrt{\frac{P_i^2 + P_{\max}^2}{2}}$$

$$P_i = \frac{C_i}{B_i}$$

where PN is the Nemerow integrated pollution index; $P_i$ is the average value of the single-factor pollution index of HM element $i$; $P_{\max}$ is the maximum value of the single-factor pollution index of HM element $i$; $C_i$ is the measured value of HM element $i$ (mg/kg); $B_i$ is the environmental background value of soil HMs maximum allowable change in $Q$ (mg/kg); and $x_i$ is the concentration of the HM background value reference point (mg/kg).

The geoaccumulation index (Igeo) was used to quantitatively evaluate the degree of HM contamination in sediment species (Müller and Förstner, 1976), reflecting natural variation characteristics of elements and serving as an important parameter for discerning anthropogenic impacts (Equation 8):

$$I_{\text{geo}} = \log_2\left(\frac{C_i}{k \times B_i}\right)$$

where $k$ is the correction coefficient, generally considered to be 1.5 (Bourliva et al., 2018).

The enrichment factor (EF) (Equation 9) was initially employed to assess the degree of HM enrichment in sediments, representing an important indicator for quantitatively evaluating pollutant enrichment through metal element enrichment in environmental media and determining whether metals originate from natural or anthropogenic sources (Zoller et al., 1974; Li et al., 2017):

$$\mathrm{EF} = \frac{(C_i/C_r)_{\text{sample}}}{(B_i/B_r)_{\text{background}}}$$

where $C_r$ is the concentration of the selected reference element (mg/kg). This method typically selects elements that are relatively stable and less affected by human activities (such as Fe, Mn, and Al) as reference elements (Liu et al., 2010; Chen et al., 2016; Napoletano et al., 2023); in this study, Fe was used as the reference element.

---

## 2.5 Potential Ecological Risk Assessment Model

We used the potential ecological risk index (Er) method proposed by Hakanson (1980) to assess HMP risks, which is based on the toxic response of elements in the environment (Hakanson, 1980), calculated using Equation 10:

$$\mathrm{RI} = \sum_{i=1}^{m} E_i^r = \sum_{i=1}^{m} T_i^r \times \frac{C_i}{B_i}$$

where RI is the integrated potential ecological risk index; $E_i^r$ is the potential ecological risk index of the $i$th HM; and $T_i^r$ is the toxicity coefficient of the HMs, indicating the toxicity response coefficients of each HM to organisms. Corresponding toxicity coefficient values for the nine HMs were: Cd=30, As=10, Cu=Pb=Ni=5, Cr=2, and Zn=Mn=Fe=1 (Xu et al., 2008; Mei et al., 2023). Classification criteria for RI and $E_i^r$ are shown in Table 4 .

---

## 2.6 HRA Model

The HRA model developed by the US EPA was used to explore carcinogenic risk (CR) and non-carcinogenic risk (NCR) of HMs to humans through three exposure pathways (US EPA, 2013): direct ingestion, dermal exposure, and inhalation of suspended soil particles. Average daily intake of HMs for the three exposure routes was calculated using Equations 11-13:

$$\mathrm{ADE}_{\text{ing}} = \frac{\mathrm{CS} \times R_{\text{ing}} \times \mathrm{EF}' \times \mathrm{ED}}{\mathrm{BW} \times \mathrm{AT}} \times 10^{-6}$$

$$\mathrm{ADE}_{\text{inh}} = \frac{\mathrm{CS} \times R_{\text{inh}} \times \mathrm{EF}' \times \mathrm{ED}}{\mathrm{BW} \times \mathrm{AT} \times \mathrm{PEF}}$$

$$\mathrm{ADE}_{\text{der}} = \frac{\mathrm{CS} \times \mathrm{AF} \times \mathrm{SA} \times \mathrm{ABS} \times \mathrm{EF}' \times \mathrm{ED}}{\mathrm{BW} \times \mathrm{AT}} \times 10^{-6}$$

where $\mathrm{ADE_{ing}}$, $\mathrm{ADE_{inh}}$, and $\mathrm{ADE_{der}}$ are average daily exposures to HMs via direct ingestion, inhalation, and dermal exposure, respectively; CS is the soil HM concentration; $R_{\mathrm{ing}}$ is the ingestion rate (mg/d); EF' is the exposure frequency (d/a); ED is the exposure duration (a); BW is the body weight (kg); AT is the average time of exposure (d); $R_{\mathrm{inh}}$ is the inhalation rate ($m^3$/d); PEF is the particulate emission factor ($m^3$/kg); AF is the adherence factor (mg/$cm^2$); SA is the skin surface area ($cm^2$); and ABS is the absorption factor. Detailed parameter values for adults and children are presented in Table S2. NCR and CR indices were calculated using Equations 14–17:

$$\mathrm{NCR}_i = \frac{\mathrm{ADE}_i}{\mathrm{RfD}_i}$$

$$\mathrm{CR}_i = \mathrm{ADE}_i \times \mathrm{SF}_i$$

$$\mathrm{TNCR} = \mathrm{NCR_{ing}} + \mathrm{NCR_{inh}} + \mathrm{NCR_{der}}$$

$$\mathrm{TCR} = \mathrm{CR_{ing}} + \mathrm{CR_{inh}} + \mathrm{CR_{der}}$$

where $\mathrm{NCR}_i$ is the NCR index of the $i$th HM; $\mathrm{RfD}_i$ is the daily reference dose (mg/(kg · d)); $\mathrm{ADE}_i$ is the average daily exposure to the $i$th HM (mg/(kg · d)); $\mathrm{CR}_i$ is the CR index of the $i$th HM; SF is the carcinogenicity slope factor ((kg · d)/mg); TNCR is the total NCR; $\mathrm{NCR_{ing}}$, $\mathrm{NCR_{inh}}$, and $\mathrm{NCR_{der}}$ represent direct ingestion, inhalation, and dermal exposure components of $\mathrm{NCR}_i$, respectively; TCR is the total CR; and $\mathrm{CR_{ing}}$, $\mathrm{CR_{inh}}$, and $\mathrm{CR_{der}}$ represent direct ingestion, inhalation, and dermal exposure components of $\mathrm{CR}_i$, respectively. RfD and SF values are presented in Table S3. NCR $>$ 1.00 indicates the presence of non-carcinogenic health risks. CR is quantitatively evaluated using three criteria: negligible risk (CR $<$ 1.00\$×10$^{\{-6\}), acceptablerisk(1.00\times 10}${-6}\$ $\leq$ CR $<$ 1.00\$×10^{-4}\$), and significant carcinogenic risk (CR $\geq$ 1.00\$×10^{-4}\$) (Li et al., 2023). Based on HRA, we conducted Monte Carlo simulation using Crystal Ball v.3.0 software. The dataset underwent 10,000 iterations at a 95.00% confidence level to derive stable results, enabling characterization of single-factor contamination risk and composite contamination probability for the study area (Liu et al., 2023). Additionally, through the PMF model, we quantified CR and NCR posed by each source to both adults and children (Shen et al., 2024; Zhao et al., 2024).

---

## 2.7 Data Processing

Data were organized using Excel v.2016 software. Spatial analysis of the study area was conducted using ArcGIS v.10.8 software. Spatial concentrations of

HMs, spatial contribution degrees of different sources, and RI index were analyzed using the inverse distance weighting method. Exploratory factor analysis and correlation analysis were performed using SPSS v.23.0 software. HM sources were calculated using EPA PMF v.5.0 software, Monte Carlo simulation was performed using Crystal Ball v.3.0 software, and graphs were created using Origin v.2022 software.

---

### 3.1.1 HM Concentration

Soil pH values in the study area ranged from 7.05 to 9.15, with statistical results of HM concentrations presented in Table 5 . Maximum concentrations of Cd, Cu, Ni, Pb, and As were below China' s background values. The maximum Cr concentration was 1.84 times that of industrial areas in China. Except for Cr, Fe, and Ni, average and median HM values in the study area exceeded local background values. Mean values of all HMs were 1.38 (Cd), 1.26 (Cr), 1.32 (Cu), 0.88 (Fe), 1.15 (Mn), 0.84 (Ni), 1.58 (Pb), 1.00 (Zn), and 1.05 (As) times the local background values, respectively. Median values were 1.00 (Cd), 0.77 (Cr), 1.23 (Cu), 0.88 (Fe), 1.14 (Mn), 0.85 (Ni), 1.48 (Pb), 1.01 (Zn), and 1.00 (As) times the local background values, respectively. Coefficients of variation (CV) for Cd and Pb were 90.40% and 67.54%, respectively, both exceeding 50.00%. Background values used in this study were calculated from the average of three background soil samples taken near the Alagou Reservoir, which is considerably distant from the mining area (Table S4). No significant human activity was identified in the vicinity, and the ecological environment was pristine.

---

### 3.1.2 Spatial Distribution Characteristics of HMs

Spatial distribution results showed that Pb concentrations were higher in the mining area, while the other eight HMs exhibited lower concentrations [Figure 2: see original paper]. High Pb concentrations in the mining area may be related to mining activities and industrial transportation. Cadmium exhibited elevated distribution in the eastern human settlement area of the mine, suggesting anthropogenic influence. Distributions of Cr, Cu, and Ni were generally consistent, with elevated concentrations observed in the northeastern area outside the mining zone and west of the eastern human settlement area. Iron, Mn, Zn, and As all showed low concentrations within the mining area. However, Mn exhibited higher concentrations in the northeastern area outside the mining zone and relatively elevated levels east of Tonggou Village. Zinc and As demonstrated higher concentrations in the southern area outside the mining zone, with high As concentrations located at the foot of mountains and high Zn concentrations adjacent to roads. Iron displayed elevated concentrations in both northeastern and southern areas outside the mining zone. High-concentration areas of Mn and Fe were characterized by favorable ecological environments, while high-concentration areas of As were situated away from roads. Elevated Zn concentrations were likely associated with road transportation activities.

---

### 3.2.1 Multivariate Statistical Analysis

Kaiser-Meyer-Olkin (KMO) and Bartlett' s tests were performed to determine whether EFA could be appropriately conducted. Results showed that EFA was suitable (KMO = 0.7; $P < 0.001$). As shown in Figure 3a [Figure 3: see original paper], three principal components (PCs) with eigenvalues greater than 1.0 were retained. The main representative elements of PC1 were Cr, Ni, and Fe, with a variance explanation rate of 33.13%. PC2 primarily represented Cu, Mn, Zn, and As, explaining 26.17% of variance. PC3 mainly represented Cd and Pb, explaining 14.21% of variance. The rotated cumulative variance explanation rate was 73.52%.

Correlation analysis results are presented in Figure 3b. The correlation coefficient between Cr and Ni was 0.92, indicating a very strong correlation, while Cr and Fe showed a strong correlation (coefficient = 0.74). Cr and Cu exhibited a moderate correlation (coefficient = 0.54). Copper showed a very strong correlation with Mn (0.83) and strong correlations with Fe (0.70) and Ni (0.60), respectively. Iron demonstrated strong correlations with Mn (0.60) and Ni (0.75). Cadmium and Pb showed no significant correlations with other HMs. Notably, Pb displayed negative correlations with all eight other HMs, and Cd also exhibited a negative correlation with Zn.

---

### 3.2.2 PMF Model Analysis

Based on PMF model parameters and ArcGIS v.10.8 software, we obtained the contribution of each factor [Figure 3: see original paper]. Factor 1 was the greatest contributor, accounting for 43.46% of total HM concentrations. The main loading elements of Factor 1 exhibited contributions of 49.66%, 56.71%, 54.28%, 69.95%, 55.89%, and 49.71% for Cr, Cu, Fe, Mn, Ni, and Zn, respectively, and these elements were primarily enriched in the eastern part of the study area and the northeastern part of the mining area [FIGURE:3c and 3d]. Factor 2 accounted for 22.87% of total HM concentrations, with Pb as its main loading element contributing 83.14%, primarily enriched within the mining area and near southern roads [FIGURE:3e and 3f]. Factor 4 accounted for 23.03% of total HM concentrations, with As as its main loading element contributing 61.16%, primarily enriched at the foot of southern mountains outside the mining area [FIGURE:3i and 3j]. Factor 3 accounted for 10.64% of total HM concentrations, with Cd as its main loading element, explaining 74.75% of its own variance and primarily enriched in the southern human settlement zone of the study area [FIGURE:3g and 3h].

## 3.3 Pollution Index Assessment

NIPI, Igeo, and EF indices were used to analyze pollution levels of soil HMs in the mining area [Figure 4: see original paper]. The NIPI index showed that Cd and Pb were heavily polluted, with Cd' s index value of 6.86 considerably higher than other HMs. The NIPI value for Pb was 3.64. Copper (2.70) and Mn (2.52) were moderately polluted, whereas Cr (1.85), Fe (1.23), Ni (1.85), Zn (1.54), and As (1.88) were slightly polluted. Average Igeo values for all HMs were below 0, but 36.00% of Pb values ranged from 0 to 1, indicating no to moderate pollution, and 11.50% of sampling sites ranged from 1 to 2, indicating moderate pollution. For Cd, 24.59% of values indicated pollution; for Cu, 26.23%; and for Mn, 22.95%. The EF index, calculated using Fe as the reference element, showed that Cr and Ni were primarily non-enriched, while Cd, Cu, Mn, Zn, and As were mainly slightly enriched. Moderate enrichment was evident for Cd, Cu, Pb, and As, with individual sites showing significant enrichment of Cd and Pb.

## 3.4 Potential Ecological Risk Assessment

Average potential ecological risk index ($E_r$) values for HMs in the mining area and surrounding soils are shown in Figure 4d [Figure 4: see original paper]. The $E_r$ value for Cd was 42.20, indicating medium risk. Approximately 37.70% of Cd sampling sites exceeded 40.00, with a maximum value of 287.82, considered high risk. $E_r$ values for other HMs were: 0.87 (Fe), 1.00 (Zn), 1.15 (Mn), 1.29 (As), 1.58 (Cr), 4.19 (Ni), 6.61 (Cu), and 7.89 (Pb), all indicating low risk. The integrated potential ecological risk index (RI) around the mining area was 66.81, indicating a safe, low-risk state. However, RI in the eastern part of the mining area exceeded 150.00, indicating medium ecological risk, and RI in the easternmost part exceeded 300.00, indicating high ecological risk [Figure 4e: see original paper].

## 3.5.1 Conventional Health Risk Assessment

We examined CR and NCR of HMs in adults and children via three pathways (ingestion, inhalation, and dermal contact) using a conventional HRA model, with results displayed in Table 6 . Risk levels across the three pathways followed the order: ingestion > inhalation > dermal contact. For NCR, HM risk levels followed: Cd > As > Pb > Cr > Ni > Cu > Mn > Zn. The total NCR value for Cd in adults was 4.39, exceeding 1.00. For children, NCR values for Cd (3.11), Pb (1.66), and As (3.17) all exceeded 1.00. For CR, HM risk levels were ranked as Cd > Ni > As > Cr > Pb. For adults, Cd (9.09$\times 10^{-3}$) and Ni ($2.52\times 10^{-4}$) values exceeded $1.00\times 10^{-4}$; for children, Cd ($1.62\times 10^{-2}$), Ni ($4.50\times 10^{-4}$), and As ($1.22\times 10${-

4}) exceeded $1.00\times10^{-4}$. Overall, Cd posed the highest risk among HMs, with ingestion contributing significantly more risk than inhalation and dermal contact.

---

### 3.5.2 Health Risk Assessment Based on Monte Carlo Simulation

Distribution characteristics of HMs and HRA indices are displayed in Table S5. As indicated by NCR [Figure 5a: see original paper] and CR [Figure 5b: see original paper], and single HM NCR and CR values (Figs. S1 and S2; Table 6), Monte Carlo simulation-based probability results were lower than conventional risk assessment results. NCR for each HM was <1.00, total NCR for adults was <1.00, and total NCR for children exhibited a 4.00% probability of >1.00. CR values for Cd, Cr, Ni, Pb, and As were $<1.00\times10^{-4}$, and CR for Pb was $<1.00\times10^{-6}$, which can be completely disregarded. Total CR for adults was $<1.00\times10^{-4}$, within the acceptable range. Total CR for children had an 88.00% probability in the acceptable range, while 12.00% probability in the unacceptable range cannot be disregarded. This result reflects that evaluation parameters for children are more sensitive than those for adults, and children' s physiological characteristics and behavioral habits differ from adults' . Children are more likely to ingest pollutants through hand-to-mouth contact, with relatively higher HM intake per unit body weight due to physiological factors such as body weight and surface area. Additionally, children have immature immune systems and less developed metabolic functions, making them more vulnerable to higher health risks than adults.

---

### 3.5.3 Health Risk Assessment Based on PMF Model

We developed a Sankey diagram of human risks based on the PMF model [FIGURE:5c and 5d]. In Factor 1, Mn accounted for the highest proportion (20.76%) among HMs. In Factor 2, Pb accounted for the largest proportion (46.42%). In Factor 3, Cd accounted for the greatest proportion (78.77%). In Factor 4, As accounted for the largest proportion (32.34%). The contribution degree of all four sources to NCR and CR in adults and children decreased in the order: Factor 3 > Factor 1 > Factor 4 > Factor 2. NCR contributions decreased in the order of 64.49% > 24.97% > 6.24% > 4.30% for adults and 64.43% > 24.97% > 6.28% > 4.33% for children. CR contributions from high to low were 72.08% > 26.14% > 1.03% > 0.75% for both adults and children. CR results for adults and children were almost identical because Cd was the main contributing element of Factor 3, and its risk value was significantly higher than those of other HMs.

---

## 4.1 Source Allocation

Based on EFA and correlation analysis results, EFA classified HMP sources into three categories, while the PMF model further subdivided pollution sources into four distinct categories. Comparing the two analytical methods revealed the following relationships: Factor 1 explained PC1 and partially overlapped with PC2, Factors 2 and 3 collectively explained PC3, and Factor 4 accounted for the remaining variation in PC2.

The loading elements of Factor 1 (43.46%) were Cr, Cu, Fe, Mn, Ni, and Zn. The study area is surrounded by mountains on three sides, with substantial rock excavation during mining. The primary sources of Cr and Ni are soil matrices and rocks (Chen et al., 2016; Liao et al., 2021; Zhang et al., 2021a, b). Iron is the second most abundant metallic element in the Earth' s crust and showed strong correlations with both Cr and Ni in this study. Previous research has demonstrated that Mn and Cu are influenced by soil formation matrices (Guan et al., 2018; Yuanan et al., 2020). As shown in Table 1, mean and median values of Cr, Ni, and Fe were lower than background values in the study area, and high-concentration areas in the east had favorable ecological conditions. Additionally, grazing is the main human activity in the study area. Copper and Zn are present in animal feed, and animal husbandry generates large quantities of manure (Wu et al., 2010; Liang et al., 2017; Hu et al., 2018b), which naturally accumulates and spreads to soils through short-term rainfall, resulting in Cu and Zn accumulation. Therefore, Factor 1 was defined as natural and livestock sources.

The loading elements for Factor 2 (22.86%) were Pb, Fe, Cr, Ni, and Zn, with Pb as the most prominent contributor. Lead accumulation in soils originates from gasoline use in automobiles (Fei et al., 2019; Kong et al., 2021). Although leaded gasoline is no longer used in China, historical Pb contamination must be considered (Song et al., 2019). Abrasion of metal parts such as automobile tires and Zn plating also aggravates Pb, Zn, and Fe buildup in soils (Guan et al., 2018; Zheng et al., 2023). As shown in Figure 2, Pb concentrations were higher in the central and western parts of the mining area and in the southern part of the study area. Pb concentrations were not high in other areas because Pb is relatively immobile in soil. $Pb^{2+}$ has a large radius and strong polarization ability, easily combining with soil anions to form insoluble compounds (Tawinteung et al., 2005). Soil pH results (Table S1) indicated alkaline conditions in the study area. Under alkaline conditions, negative charges on clay mineral, oxide, and hydroxide surfaces increase, enhancing $Pb^{2+}$ adsorption. $Pb^{2+}$ also reacts with iron and manganese oxides and becomes immobilized on surfaces and within crystal lattices (Bradl, 2004). Simultaneously, the absence of long-term precipitation in arid areas results in low soil moisture content, further limiting Pb mobility. Zinc and Fe contents were lower inside the mine. In addition to poor mine soil conditions, a local wastewater treatment station was established inside the mine, addressing Fe and Zn pollution from mining production. Therefore, Factor 2 was defined as an industrial transportation source.

The main loading element of Factor 3 (10.64%) was Cd, with a small percentage of As. Industrial emissions are an important cause of soil Cd accumulation (Li et al., 2016; Yang et al., 2020). Fossil fuels, particularly coal, contain high amounts of Cd (Zhou et al., 2024). In mining areas, fire suppression of coal combustion layers requires burning large quantities of coal sandwiched between soil layers, generating smoke particles that diffuse throughout the mining area. Regular use of explosives for blasting during mining also causes Cd accumulation. Additionally, the mining area' s altitude exceeds 2300 m, and populated areas require heating for longer periods each year than plains regions, necessitating more coal combustion. Cadmium exhibits strong mobility in soil due to its relatively small ionic radius, allowing it to move easily between soil particle pores. Cadmium often exists in ionic states in soil and can be readily exchanged with various soil ions to form soluble complexes (Bradl, 2004). The study area' s topography is high in the west and low in the east, allowing Cd in soil to flow to low-lying areas with surface water formed by heavy rainfall, resulting in Cd accumulation in the eastern part of the study area. Mining of mineral resources can release initially stabilized HMs from ores that migrate to soils (Zhong et al., 2020). Therefore, Factor 3 was defined as fossil fuel combustion. By combining Factors 2 and 3, we generalized PC3 as an industrial source containing both industrial transport and fossil fuel combustion.

The main loading elements of Factor 4 (23.03%) were As and Zn. Coal mine blasting and mining activities moved As to surface soils, which was then blown into the dry atmosphere under strong winds, blocked by southern mountains, and settled on surfaces. Simultaneously, short-term rain erosion and summer meltwater gathering at the foot of mountains resulted in As accumulation (Luo et al., 2020; Liu et al., 2023). The existence of populated areas in the eastern part of the study area, domestic sewage, motor vehicle exhaust from human activities, and chemical fertilizer use in green spaces around residential areas caused accumulation of As, Zn, and other HMs (Vidu et al., 2020). Therefore, Factor 4 was defined as atmospheric deposition and domestic pollution.

---

## 4.2 HM Pollution Analysis

Our study found that maximum HM concentrations in the study area did not exceed Chinese standard levels (MEE, 2018b). The NIPI confirmed Cd as the most notable HM pollutant, and soil Cd content is often related to fuel combustion (Hossain Bhuiyan et al., 2021). Additionally, the three pollution indices showed Pb as the most polluted and enriched element. The mine' s annual production is 13$\times10^{6}$ t/a (National Energy Xinjiang Toksun Energy Co., Ltd., 2022), and Pb production in soils is often closely related to transport activities (Zhou et al., 2024), which could lead to Pb accumulation. High coefficients of variation for Cd and Pb also demonstrated that human activities greatly affected HM distribution in the study area. The RI showed that Cd' s potential risk was much higher than other HMs, though Cd concentration in soil was not high compared to other HMs. This phenomenon primarily reflects that Cd' s toxicity factors are much higher than those of other HMs.

---

### 4.3 Risk Assessment

HRA results showed that ingestion posed higher risks than the other two pathways for both adults and children (Table 3), consistent with previous studies (Qing et al., 2015; Zeng et al., 2019). Ingestion is the primary human exposure route to HMs (Liu et al., 2023). Monte Carlo simulation-based results generally reduced the likelihood of NCR and CR, improving accuracy. Monte Carlo simulation-based HRA uses large numbers of samples from different data types, resulting in minor errors.

Both NCR and CR were significantly higher in children than adults (Figs. S1 and S2), indicating greater children' s susceptibility to HM effects. HRA model results were not fully consistent with PMF model results. Conventional risk assessment reflected risks from three ingestion pathways but ignored other possible HM exposure routes. PMF-based HRA analyzed contributions of different sources to CR. Fossil fuel combustion sources were identified as the most critical contributor to human health risks among identified sources. This finding suggests that Cd should be designated as a priority control element within the mining area. Dry and windy conditions accelerate long-range HM diffusion; studies at the Kushk lead-zinc mine in Iran have shown that pollutants such as Pb and Zn can be wind-transported to residential areas up to 4 km away, significantly increasing population exposure risk (Mokhtari et al., 2018).

Cadmium released from coal combustion exhibits enhanced chemical stability in alkaline soils (pH 7.05-9.15), increasing its potential to threaten human health via dust inhalation and food chain transmission (Song et al., 2018). Several studies have reported high persistence and bioavailability of Cd in arid area soils, with risks of entering the human body through crop uptake significantly higher than in humid areas (Dai et al., 2017; Kubier et al., 2019). Thus, prioritizing Cd as a control element in mining areas has clear regional relevance. Second, priority should be given to Factor 1, and implementing controlled grazing practices is essential for safeguarding ecologically intact zones in the northeastern mining area. Weathering is strong in arid areas, and HM-rich mineral debris easily reaches the surface through wind or runoff. Furthermore, unreasonable grazing activities exacerbate soil disturbance, leading to dust diffusion and migration on soil surfaces. Monitoring in Northwest China' s arid areas has shown that Pb and Cd concentrations in overgrazed areas are 30.00%-50.00% higher than in nature reserves, with HM risks from grazing activities accounting for 15.00%-20.00% of NCR for children (Jiao et al., 2023).

Water quality, waste gas emissions, and noise are regularly monitored in the mining area to determine whether pollution emissions exceed standards. However, the impact of water scarcity in arid areas cannot be ignored. Limited surface runoff complicates HM dilution, resulting in high pollutant concentrations in water bodies. Groundwater concentrations of Cd and As in semi-arid areas of India exceed WHO (World Health Organization) standards, significantly increasing cancer risk (Mukherjee et al., 2020). Notably, despite China' s ban on leaded gasoline, dust emissions from transport vehicles in arid mining areas remain an important Pb pollution source, with secondary pollution from tire wear and road dust more likely to persist long-term in dry environments (Navarro et al., 2008). All pollution sources, regardless of impact magnitude, should be addressed to ensure human health risks remain within acceptable limits.

## 5 Conclusions

This study explored sources of soil HM contamination and ecological risk in the Black Mountain Open Pit Coal Mine in Turpan City, Northwest China, using Monte Carlo simulation, PMF modeling, and HRA modeling. Soil HM sources were categorized as natural and livestock, industrial transportation, fossil fuel combustion, and atmospheric deposition combined with domestic pollution. According to NIPI, Igeo, and EF indices, Cd, Pb, and As were identified as the main pollution sources. Monte Carlo simulation results showed a 4.00% probability of NCR and a 12.00% probability of CR for children, with no NCR or CR risks for adults. Fossil fuel combustion sources accounted for the largest proportion in source-based HRA, while industrial transportation sources accounted for the smallest proportion. In summary, the study area' s ecological risk is relatively low, but health risks to children cannot be ignored.

## Conflict of Interest

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

## Acknowledgements

This work was supported by the Third Xinjiang Scientific Expedition Project (2021XJKK1104), specifically the Investigation and Assessment of the Effects of Energy and Mineral Resource Development in the Turpan Basin on Regional Ecological Environment.

## Author Contributions

**Conceptualization:** LI Yun, SHI Qingdong
**Data curation:** LI Yun
**Formal analysis:** LI Yun, ZHUANG Zhong
**Funding acquisition:** XIA Qianrou, SHI Qingdong
**Investigation:** LI Yun, ZHUANG Zhong, ZHU Jiawei, WANG Peijuan, LI Dinghao, Yryszhan ZHAKYPBEK, Serik TURSBEKOV
**Methodology:** LI Yun
**Resources:** LI Yun, ZHUANG Zhong, ZHU Jiawei, WANG Peijuan, LI Dinghao, Yryszhan ZHAKYPBEK, Serik TURSBEKOV
**Software:** LI Yun
**Supervision:** XIA Qianrou, SHI Qingdong
**Validation:** LI Yun, ZHUANG Zhong, ZHU Jiawei, WANG Peijuan, LI Dinghao, Yryszhan ZHAKYPBEK, Serik TURSBEKOV
**Visualization:** LI Yun
**Writing -original draft preparation:** LI Yun
**Writing -review and editing:** ZHUANG Zhong, ZHU Jiawei, WANG Peijuan, LI Dinghao, SHI Qingdong, XIA Qianrou, Yryszhan ZHAKYPBEK, Serik TURSBEKOV

All authors approved the manuscript.

## References

Anaman R, Peng C, Jiang Z C, et al. 2022. Identifying sources and transport routes of heavy metals in soil with different land uses around a smelting site by GIS based PCA and PMF. *Science of the Total Environment*, 823: 153759, doi: 10.1016/j.scitotenv.2022.153759.

Bourliva A, Kantiranis N, Papadopoulou L, et al. 2018. Seasonal and spatial variations of magnetic susceptibility and potentially toxic elements (PTEs) in road dusts of Thessaloniki city, Greece: A one-year monitoring period. *Science of the Total Environment*, 639: 417-427.

Bradl H B. 2004. Adsorption of heavy metal ions on soils and soils constituents. *Journal of Colloid and Interface Science*, 277(1): 1-18.

Cao C C, Wang L, Li H R, et al. 2018. Temporal variation and ecological risk assessment of metals in soil nearby a Pb-Zn mine in southern China. *International Journal of Environmental Research and Public Health*, 15(5): 940, doi: 10.3390/ijerph15050940.

CEMS (China National Environmental Monitoring Centre). 1990. *Natural Background Values of Soil Elements in China*. Beijing: China Environmental Science Press. (in Chinese)

Chen H Y, Teng Y G, Lu S J, et al. 2016. Source apportionment and health risk assessment of trace metals in surface soils of Beijing metropolitan, China. *Chemosphere*, 144: 1002–1011.

Dai Y C, Nasir M, Zhang Y L, et al. 2017. Comparison of DGT with traditional methods for assessing cadmium bioavailability to *Brassica chinensis* in different soils. *Scientific Reports*, 7(1): 14206, doi: 10.1038/s41598-017-13820-3.

Department of Ecology and Environment of Xinjiang Uygur Autonomous Region. 2022. Shenhua Xinjiang Energy limited liability company Toxon Black Mountain opencast mine Environmental impact statement. [2024-06-30]. https://max.book118.com/html/2022/0629/7152025010004136.shtm.

Fei X F, Xiao R, Christakos G, et al. 2019. Comprehensive assessment and source apportionment of heavy metals in Shanghai agricultural soils with different fertility levels. *Ecological Indicators*, 106: 105508, doi: 10.1016/j.ecolind.2019.105508.

Guan Q Y, Wang F F, Xu C Q, et al. 2018. Source apportionment of heavy metals in agricultural soil based on PMF: A case study in Hexi Corridor, Northwest China. *Chemosphere*, 193: 189–197.

Hakanson L. 1980. An ecological risk index for aquatic pollution control. A sedimentological approach. *Water Research*, 14(8): 975–1001.

Hossain Bhuiyan M A, Chandra Karmaker S, Bodrud-Doza M, et al. 2021. Enrichment, sources and ecological risk mapping of heavy metals in agricultural soils of Dhaka district employing SOM, PMF and GIS methods. *Chemosphere*, 263: 128339, doi: 10.1016/j.chemosphere.2020.128339.

Hu B F, Zhao R Y, Chen S C, et al. 2018a. Heavy metal pollution delineation based on uncertainty in a coastal industrial city in the Yangtze River Delta, China. *International Journal of Environmental Research and Public Health*, 15(4): 710, doi: 10.3390/ijerph15040710.

Hu W Y, Wang H F, Dong L R, et al. 2018b. Source identification of heavy metals in peri-urban agricultural soils of southeast China: An integrated approach. *Environmental Pollution*, 237: 650–661.

Jiao Y W, Liu Y T, Wang W, et al. 2023. Heavy metal distribution characteristics, water quality evaluation, and health risk evaluation of surface water in abandoned multi-year Pyrite mine area. *Water*, 15(17): 3138, doi: 10.3390/w15173138.

Karakas F, Imamoglu I, Gedik K. 2017. Positive matrix factorization dynamics in fingerprinting: A comparative study of PMF2 and EPA-PMF3 for source apportionment of sediment polychlorinated biphenyls. *Environmental Pollution*, 220: 20–28.

Kong F J, Chen Y C, Huang L, et al. 2021. Human health risk visualization of potentially toxic elements in farmland soil: A combined method of source and probability. *Ecotoxicology and Environmental Safety*, 211: 111922, doi: 10.1016/j.ecoenv.2021.111922.

Kubier A, Wilkin R T, Pichler T. 2019. Cadmium in soils and groundwater: A review. *Applied Geochemistry*, 108: 104388, doi: 10.1016/j.apgeochem.2019.104388.

Li F, Zhang J D, Huang J H, et al. 2016. Heavy metals in road dust from Xiandao District, Changsha City, China: Characteristics, health risk assessment, and integrated source identification. *Environmental Science and Pollution Research*, 23: 13100–13113.

Li H H, Chen L J, Yu L, et al. 2017. Pollution characteristics and risk assessment of human exposure to oral bioaccessibility of heavy metals via urban street dusts from different functional areas in Chengdu, China. *Science of the Total Environment*, 586: 1186–1195.

Li Y, Ye Z, Yu Y, et al. 2023. A combined method for human health risk area identification of heavy metals in urban environments. *Journal of Hazardous Materials*, 449: 131067, doi: 10.1016/j.jhazmat.2023.131067.

Li Z R, Deblon J, Zu Y Q, et al. 2019. Geochemical baseline values determination and evaluation of heavy metal contamination in soils of Lanping Mining Valley (Yunnan Province, China). *International Journal of Environmental Research and Public Health*, 16(23): 4686, doi: 10.3390/ijerph16234686.

Liang J, Feng C T, Zeng G M, et al. 2017. Spatial distribution and source identification of heavy metals in surface soils in a typical coal mine city, Lianyuan, China. *Environmental Pollution*, 225: 681–690.

Liao S Y, Jin G Q, Khan M A, et al. 2021. The quantitative source apportionment of heavy metals in peri-urban agricultural soils with UNMIX input fluxes analysis. *Environmental Technology & Innovation*, 21: 101232, doi: 10.1016/j.eti.2020.101232.

Liu H W, Zhang Y, Yang J S, et al. 2021. Quantitative source apportionment, risk assessment and distribution of heavy metals in agricultural soils from southern Shandong Peninsula of China. *Science of the Total Environment*, 767: 144879, doi: 10.1016/j.scitotenv.2020.144879.

Liu Y, Guo H C, Yang P J. 2010. Exploring the influence of lake water chemistry on chlorophyll a: A multivariate statistical model analysis. *Ecological Modelling*, 221(5): 681–688.

Liu Z, Du Q Q, Guan Q Y, et al. 2023. A Monte Carlo simulation-based health risk assessment of heavy metals in soils of an oasis agricultural region in northwest China. *Science of the Total Environment*, 857: 159543, doi: 10.1016/j.scitotenv.2022.159543.

Luo H P, Guan Q Y, Pan N H, et al. 2020. Using composite fingerprints to quantify the potential dust source contributions in northwest China. *Science of the Total Environment*, 742: 140560, doi: 10.1016/j.scitotenv.2020.140560.

MEE (Ministry of Ecology and Environment of the People' s Republic of China). 2018a. *Soil-Determination of pH-Potentiometry*. Beijing: MEE. (in Chinese)

MEE (Ministry of Ecology and Environment of the People' s Republic of China). 2018b. *Soil Environmental Quality Risk Control Standard for Soil Contamination of Development Land*. Beijing: MEE. (in Chinese)

MEE (Ministry of Ecology and Environment of the People' s Republic of China). 2023. *Soil and Sediment-Determination of 19 Total Metal Elements-Inductively Coupled Plasma Mass Spectrometry*. Beijing: MEE. (in Chinese)

Mei W, Liu S L, Yuan Y Y, et al. 2023. Optimization of potential ecological risk index method for soil heavy metals—A case study of Chengkou County, Chongqing City. *Chinese Journal of Soil Science*, 54(2): 473–480. (in Chinese)

Mi Y Z, Zhou J, Liu M L, et al. 2023. Machine learning method for predicting cadmium concentrations in rice near an active copper smelter based on chemical mass balance. *Chemosphere*, 319: 138028, doi: 10.1016/j.chemosphere.2023.138028.

Mokhtari A R, Feiznia S, Jafari M, et al. 2018. Investigating the role of wind in the dispersion of heavy metals around mines in arid regions (a case study from Kushk Pb–Zn Mine, Bafgh, Iran). *Bulletin of Environmental Contamination and Toxicology*, 101: 124–130.

Mukherjee I, Singh U K, Singh R P, et al. 2020. Characterization of heavy metal pollution in an anthropogenically and geologically influenced semi-arid region of east India and assessment of ecological and human health risks. *Science of the Total Environment*, 705: 135801, doi: 10.1016/j.scitotenv.2019.135801.

Müller G, Förstner U. 1976. Heavy Metals in Sediments of Elbe near Stade: Changes Since 1973. *Naturwissenschaften*, 63(5): 242–243. (in German)

Napoletano P, Guezgouz N, Di Iorio E, et al. 2023. Anthropic impact on soil heavy metal contamination in riparian ecosystems of northern Algeria. *Chemosphere*, 313: 137522, doi: 10.1016/j.chemosphere.2022.137522.

National Energy Xinjiang Toksun Energy Co., Ltd. 2022. Environmental impact assessment of the 13 million tons per year production capacity verification project of Heishan Open-pit Coal Mine in Toksun County, Xinjiang Energy Co., Ltd. of National Energy Group. [2024-03-28]. https://sthjt.xinjiang.gov.cn/xjepd/gwwjhpyb/202309/b703ad647d64471193702040df62d90a.shtml.

Navarro M C, Pérez-Sirvent C, Martínez-Sánchez M J, et al. 2008. Abandoned mine sites as a source of contamination by heavy metals: A case study in a semi-arid zone. *Journal of Geochemical Exploration*, 96(2–3): 183–193.

Nemerow N L. 1974. *Scientific Stream Pollution Analysis*. Washington DC: Scripta Book Company.

Norris G, Duvall R, Brown S, et al. 2014. *EPA Positive Matrix Factorization*

*(PMF) 5.0 Fundamentals and User Guide*. Washington DC: United States Environmental Protection Agency (US EPA).

Qi R, Xue N N, Wang S Z, et al. 2022. Heavy metal(loid)s shape the soil bacterial community and functional genes of desert grassland in a gold mining area semi-arid region. *Environmental Research*, 214: 113749, doi: 10.1016/j.envres.2022.113749.

Qing X, Zong Y T, Lu S G. 2015. Assessment of heavy metal pollution and human health risk in urban soils of steel industrial city (Anshan), Liaoning, Northeast China. *Ecotoxicology and Environmental Safety*, 120: 377–385.

Shen C, Huang S F, Wang M, et al. 2024. Source-oriented health risk assessment and priority control factor analysis of heavy metals in urban soil of Shanghai. *Journal of Hazardous Materials*, 480: 135859, doi: 10.1016/j.jhazmat.2024.135859.

Song H, Liu J C, Cao Z P, et al. 2019. Analysis of disease profile, and medical burden by lead exposure from hospital information systems in China. *BMC Public Health*, 19(1): 1170, doi: 10.1186/s12889-019-7515-5.

Song S, Li Y J, Li L, et al. 2018. Arsenic and heavy metal accumulation and risk assessment in soils around mining areas: The Urad Houqi Area in arid northwest China as an example. *International Journal of Environmental Research and Public Health*, 15(11): 2410, doi: 10.3390/ijerph15112410.

Specht A J, Lindsay I C, Wells E M, et al. 2025. Spatial distribution of heavy metal contamination in soils of Fallujah, Iraq. *Exposure and Health*, 17: 31–39.

Tawinteung N, Parkpian P, DeLaune R D, et al. 2005. Evaluation of extraction procedures for removing lead from contaminated soil. *Journal of Environmental Science and Health. Part A, Toxic/Hazardous Substances & Environmental Engineering*, 40: 801–808.

Tian S H, Liang T, Li K X, et al. 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.

US EPA (United States Environmental Protection Agency). 2013. *Framework for Human Health Risk Assessment to Inform Decision Making*. Washington DC: US EPA.

US EPA (United States Environmental Protection Agency). 2015. *Positive Matrix Factorization Model for Environmental Data Analyses*. Washington DC: US EPA.

Vidu R, Matei E, Predescu A M, et al. 2020. Removal of heavy metals from wastewaters: A Challenge from current treatment methods to nanotechnology applications. *Toxics*, 8(4): 101, doi: 10.3390/toxics8040101.

Wang X Y, Liu E F, Yan M X, et al. 2023. Contamination and source apportionment of metals in urban road dust (Jinan, China) integrating the enrichment factor, receptor models (FA-NNC and PMF), local Moran' s index, Pb isotopes and source-oriented health risk. *Science of the Total Environment*, 878: 163211, doi: 10.1016/j.scitotenv.2023.163211.

Wu L J, Yue W F, Wu J, et al. 2023. Metal-mining-induced sediment pollution presents a potential ecological risk and threat to human health across China: A meta-analysis. *Journal of Environmental Management* 329: 117058, doi: 10.1016/j.jenvman.2022.117058.

Wu S, Xia X H, Lin C Y, et al. 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): 860–868.

Xu Y, Shi H D, Fei Y, et al. 2021. Identification of soil heavy metal sources in a large-scale area affected by industry. *Sustainability*, 13(2): 511, doi: 10.3390/su13020511.

Xu Z Q, Ni S J, Tuo X G, et al. 2008. Calculation of heavy metal' s toxicity coefficient in the evaluation of potential ecological risk index. *Environmental Science & Technology*, 31(2): 112–115.

Yang Q Q, Li Z Y, Lu X N, et al. 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.

Yang S Y, Zhao J, Chang S X, et al. 2019. Status assessment and probabilistic health risk modeling of metals accumulation in agriculture soils across China: A synthesis. *Environment International*, 128: 165–174.

Yang S Y, Sun L J, Sun Y F, et al. 2023. Towards an integrated health risk assessment framework of soil heavy metals pollution: Theoretical basis, conceptual model, perspectives. *Environmental Pollution*, 120596, doi: 10.1016/j.envpol.2022.120596.

Yang Y, Yang X, He M J, et al. 2020. Beyond mere pollution source identification: Determination of land covers emitting soil heavy metals by combining PCA/APCS, GeoDetector and GIS analysis. *CATENA*, 185: 104297, doi: 10.1016/j.catena.2019.104297.

Yuanan H, He K L, Sun Z H, et al. 2020. Quantitative source apportionment of heavy metal(loid)s in the agricultural soils of an industrializing region and associated model uncertainty. *Journal of Hazardous Materials*, 391: 122244, doi: 10.1016/j.jhazmat.2020.122244.

Zeng S Y, Ma Y, Yang Y J, et al. 2019. Spatial assessment of farmland soil pollution and its potential human health risks in China. *Science of the Total Environment*, 687: 642–653.

Zhang H W, Zhang F, Song J, et al. 2021a. Pollutant source, ecological and human health risks assessment of heavy metals in soils from coal mining areas in Xinjiang, China. *Environmental Research*, 202: 111702, doi:

10.1016/j.envres.2021.111702.

Zhang W H, Yan Y, Yu R L, et al. 2021b. The sources-specific health risk assessment combined with APCS/MLR model for heavy metals tea garden soils south Fujian Province, China. *CATENA*, 203: 105306, doi: 10.1016/j.catena.2021.105306.

Zhang X, Song X Y, Zhang H Y, et al. 2024a. Source apportionment and risk assessment of heavy metals in typical greenhouse vegetable soils in Shenyang, China. *Environmental Monitoring and Assessment*, 196: 72, doi: 10.1007/s10661-023-12250-1.

Zhang X J, Zhang S W, Wei X Y, et al. 2024b. Identification of sources and analysis of spatial distribution of soil heavy metals in northern China coal mining areas. *Environmental Geochemistry and Health*, 46(3): 94, doi: 10.1007/s10653-024-01877-9.

Zhao J Y, Cao C Y, Chen X, et al. 2024. Source-specific ecological risk analysis and critical source identification of heavy metal(loid)s in the soil of typical abandoned coal mining area. *Science of the Total Environment*, 947: 174506, doi: 10.1016/j.scitotenv.2024.174506.

Zhao R, Guan Q Y, Luo H P, et al. 2019. Fuzzy synthetic evaluation and health risk assessment quantification of heavy metals in Zhangye agricultural soil from the perspective of sources. *Science of the Total Environment*, 697: 134126, doi: 10.1016/j.scitotenv.2019.134126.

Zheng F, Guo X, Tang M Y, et al. 2023. Variation in pollution status, sources, and risks of soil heavy metals in regions with different levels of urbanization. *Science of the Total Environment*, 866: 161355, doi: 10.1016/j.scitotenv.2022.161355.

Zhong X, Chen Z W, Li Y Y, et al. 2020. Factors influencing heavy metal availability and risk assessment of soils at typical metal mines in Eastern China. *Journal of Hazardous Materials*, 400: 123289, doi: 10.1016/j.jhazmat.2020.123289.

Zhou H, Yue X M, Chen Y, et al. 2024. Source-specific probabilistic contamination risk and health risk assessment of soil heavy metals typical ancient mining area. *Science of the Total Environment*, 906: 167772, doi: 10.1016/j.scitotenv.2023.167772.

Zhuang Q F, Li G, Liu Z Y. 2018. Distribution, source and pollution level of heavy metals in river sediments from South China. *CATENA*, 170: 386–396.

Zoller W H, Gladney E S, Duce R A. 1974. Atmospheric concentrations and sources of trace metals at the South Pole. *Science*, 183(4121): 198–200.

## Appendix

**Table S1** Heavy metal (HM) concentration and pH value in soil samples (mg/kg)

| Sample | Cd | Cr | Cu | Fe | Mn | Ni | Pb | Zn | As | pH |
|---|---|---|---|---|---|---|---|---|---|---|
| … | … | … | … | … | … | … | … | … | … | … |

*Note: Cd, cadmium; Cr, chromium; Cu, copper; Fe, iron; Mn, manganese; Ni, nickel; Pb, lead; Zn, zinc; As, arsenic. The abbreviations are the same in the following tables and figures.*

**Table S2** Health risk assessment model parameters

| Parameter | Adults | Children | Reference |
|---|---|---|---|
| $R_{\mathrm{ing}}$ (ingestion rate) | 25,550 mg/d | $1.36\times10^{9}$ mg/d | Liu et al. (2023) |
| $R_{\mathrm{inh}}$ (inhalation rate) | … | … | … |

**Table S3** Daily reference doses (RfD) and carcinogenicity slope factor (SF) for heavy metals (HMs)

| Element | RfD (ingestion) | RfD (inhalation) | RfD (dermal) | SF (ingestion) | SF (inhalation) | SF (dermal) | Reference |
|---|---|---|---|---|---|---|---|
| Cd | $1.00\times10^{-3}$ | $1.00\times10^{-5}$ | $1.00\times10^{-5}$ | 6.10 | 6.30 | $2.00\times10^{1}$ | Yang et al. (2019) |
| … | … | … | … | … | … | … | … |

**Table S4** Three HM concentrations in soil samples from the vicinity of the Alagou reservoir

| Sample | Longitude | Latitude | Cd | Cr | Cu | Fe | Mn | Ni | Pb | Zn | As |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 87°50 35 E | 42°49 27 N | … | … | … | 31,838.24 | … | … | … | … | … |
| … | … | … | … | … | … | … | … | … | … | … | … |

**Table S5** Distribution pattern of each parameter in Monte Carlo simulation

| Parameter | Distribution | Adults | Children | Reference |
|---|---|---|---|---|
| CS (Cd) | Triangular | (0.05, 0.12, 1.24) | (26.21, 55.8, 183.89) | This study |
| … | … | … | … | … |

**Fig. S1** NCR (non-carcinogenic risk) of eight HMs based on Monte Carlo simulation. (a) Cd; (b) Cr; (c) Cu; (d) Mn; (e) Ni; (f) Pb; (g) Zn; (h) As. Dashed line indicates the mean value.

**Fig. S2** CR (carcinogenic risk) of five HMs based on Monte Carlo simulation. (a) Cd; (b) Cr; (c) Ni; (d) Pb; (e) As. Dashed line indicates the mean value.

*Note: Figure translations are in progress. See original paper for figures.*