



OPEN

A combined model method was used to identify the main influencing factors of soil heavy metal pollution sources in Qian river, China

Yuru Li¹, Zihua Wang¹, Lei Yu Liu¹, Yani Geng¹✉ & Jun Zhang^{1,2}✉

The issue of soil heavy metal contamination has garnered significant global attention, with the identification of heavy metal sources and their driving factors being crucial for the prevention and management of soil heavy metal pollution. This study introduces a comprehensive source-driver model integrating the Positive Matrix Factorization (PMF) model, Geographically Weighted Regression (GWR), and Geo-detector Model (GDM). Inductively Coupled Plasma Mass Spectrometry (ICP-MS) was employed to quantify the concentrations of eight heavy metals in the soils of the lower reaches of the Qian River. The findings revealed that: (1) The average concentrations of the 8 heavy metals did not exceed the risk screening thresholds for soil environmental quality. Specifically, the mean concentrations of Ni, Zn, As, and Cd were 42.16, 102.07, 18.23, and 0.32 mg kg⁻¹, respectively, which are 1.46, 1.47, 3.7, and 6.17 times higher than the background values for soil in Shaanxi Province, with Cd exhibiting a coefficient of variation of 0.58. This high degree of variation is attributed to anthropogenic activities. Spatially, each heavy metal was more heavily concentrated in the southeastern and northwestern regions of the study area. (2) The results of PMF model showed that soil heavy metals in the study area mainly came from nature, industry, agriculture, and traffic, and the contribution of each source was 19.12%, 23.42%, 36.85% and 20.61%. Notably, agricultural sources emerged as the predominant contributors to soil pollution in the region. (3) GDM and GWR results showed that distance from village, soil type and elevation were the main drivers of soil heavy metal pollution sources in the study area. This study provides a reference for the analysis of soil heavy metal sources, and the results can provide a theoretical basis for the prevention and control of soil heavy metal pollution in the study area.

Keywords Soil heavy metals, PMF model, Geo-detector, Geographically weighted regression

Soil is an indispensable resource for human survival and development and a fundamental component of the ecosystem¹, playing an important role in maintaining food production and ecosystem stability². However, with the acceleration of industrialization and urbanization, the environmental problems caused by soil pollution have become more and more serious, and have become a global environmental problem that needs to be solved urgently³⁻⁵. Soil pollutants come from a wide variety of sources and can be broadly categorized into three major groups: inorganic pollutants, organic pollutants and harmful microorganisms⁶. Among them, inorganic pollution of heavy metals is the most harmful, because of its irreversible, persistent and difficult to recover and other characteristics⁷. Trace amounts of heavy metals such as nickel, iron and zinc are essential to the human body, but excessive amounts of Cd, Cr, Pb and As are seriously harmful to living organisms. The National Soil Pollution Survey Bulletin shows that the overall state of the soil environment is less than optimistic, with the exceedance rates of As, Cd, Cr, Cu, Hg, Ni, Pb and Zn being 2.7%, 7.0%, 1.1%, 2.1%, 1.6%, 4.8%, 1.5% and 0.9%, respectively⁸. Soil pollution caused by excessive accumulation of heavy metals not only damages the ecological

¹Shaanxi Provincial Key Laboratory of Disaster Monitoring and Mechanism Simulation, Baoji College of Arts and Science, Baoji 721013, China. ²Key Laboratory of Subsurface Hydrology and Ecological Effect in Arid Region of Ministry of Education, Chang'an University, Xi'an 710064, China. ✉email: gyn1977@126.com; zhangjun1190@126.com

environment, but also poses a threat to human health through the food chain⁹. It can enter the body through inhalation, ingestion or dermal contact, causing cancer and bone, cardiovascular and cerebrovascular diseases¹⁰. Therefore, it is of great significance for the prevention and control of soil heavy metal pollution to effectively identify the sources of soil heavy metals and determine the driving factors of soil heavy metal pollution.

To address the problem of contamination by heavy metals in agricultural soils, a number of scholars have done extensive research on its pollution characteristics¹¹, source analysis¹², and influencing factors¹³. As for source analysis, a variety of source analysis approaches have been applied to analyze and explore the origins of heavy metal contamination in soil, such as absolute principal component analysis/multiple linear regression (APCS/MLR)¹⁴, chemical mass balance (CMB)^{15,16}, positive matrix factorization (PMF)¹⁷ and the UNMIX model¹⁸. Both the UNMIX model and the PMF model are common quantitative source resolution models. The UNMIX model is based on the multiple The UNMIX model is based on the principle of linear regression, which uses the linear relationship between the concentrations of chemical components in the receptor to identify the pollution sources and their contributions, and requires the input of some possible source compositional spectra in advance, and optimizes the model parameters through iterative calculations, so that the output of the model's source compositional spectra and the source contributions are consistent with the actual situation¹⁹. Among them, PMF, a quantitative source resolution model, has the benefit of not requiring precise source component spectral information and having a superior ability to deal with missing and inaccurate data²⁰. Additionally, the role that different heavy metals play in each source can be quantitatively examined by imposing non-negative restrictions to the factor decomposition matrix so as to the resultant spectrum of source components and contribution rates do not have negative values²¹. It has been frequently used to analyze the contribution of pollutants from the source, such as water bodies and soils²². Therefore, in this research, the quantitative PMF data were combined with GIS analysis to spatially display the research area's pollutant sources' distribution.

However, receptor models (e.g., PMF models, APCS/MLR and UNMIX models, etc.) are purely mathematical-statistical models that not only lack the expression of spatial heterogeneity, but also have some common limitations²³. First, the source factors obtained from receptor models are based on existing research and expert experience, which can only broadly categorize the final results and tend to lead to subjectivity in source interpretation. Second, due to the variability of data, instability of model structure, and sensitivity of model parameters in the source interpretation study of the receptor model, the effects of environmental variables such as elevation, soil type, and land-use type on the accumulation of heavy metals in soil could not be analyzed, and it was difficult to formulate corresponding control and remedial measures²⁴. Researchers have already studied the sources of heavy metals by combining the geo-environmental factors of heavy metal pollution spreading or aggregation with the results of source analysis together^{25–27}. Soil heavy metals are affected by both natural and human activities²⁸, and traditional statistical models are unable to accurately quantify the interaction between sources of soil heavy metals because they need to assume linearity and be affected by multivariate covariance when calculating the interaction^{29,30}. In contrast, the theoretical basis of Geo-detector modeling (GDM) is based on the spatial heterogeneity of pollutants³¹, which can effectively address the limitations of traditional methods in analyzing categorical variables³². Therefore, many scholars have adopted GDM to study the spatial differentiation of soil heavy metal pollution³³. However, GDM can only realize the quantitative expression of global impact factors, and it can't quantitatively analyze the impact factors of heavy metal pollution locally, and it lacks the spatial quantitative expression of pollutants or sources³⁴. In contrast, Geographically Weighted Regression modeling (GWR), as a local spatial regression technique³⁵, considers the existence of spatial heterogeneity among spatial variables³⁶. The spatial coordinates of the sample data are considered in the analysis, which can express the heterogeneity and continuity of the spatial variables more intuitively³⁷. The combination of the GDM model and the GWR model for quantitative analysis can both eliminate the problem of multifactor multiple covariance and quantitatively spatially visualize the impact factors³⁸. Therefore, in this study, the PMF model, GDM and GWR were combined to construct a new combined model, which provides a new pathway method for the source allocation of soil heavy metals and the identification of driving factors.

The study area is located in the urban-rural zone of Baoji City, and in recent years, with the rapid development of industry and agriculture in Baoji City, the land at the edge of the city has been heavily exploited and utilized. The downstream area of the Qian River has a relatively flat topography on both sides of the rivers due to river alluviation, which is suitable for industrialized development and agricultural farming production³⁹. We hypothesized that agricultural activities, industrial production, and transportation emissions may be the main sources of soil heavy metal contamination, and selected a variety of natural and anthropogenic factors to validate this hypothesis and ultimately explore the main driving mechanisms of each source. The novelty of this study lies in the construction of an integrated model to quantitatively analyze the sources of soil heavy metals in watersheds, taking into account the spatial information of the sampling sites and the characteristics of the surrounding environment, identifying the spatial relationship between the sources of soil heavy metals and the environmental factors, and revealing how a variety of environmental factors jointly affect the accumulation of soil heavy metal sources, which is conducive to the objective explanation of the source allocation of soil pollution sources.

Therefore, this paper analyzes eight heavy metal elements (Cr, Mn, Ni, Cu, Zn, As, Pb and Cd) in farmland soils downstream of the Qian Rivers, firstly, using the PMF method in combination with GIS analysis to explore the main sources of soil heavy metal contamination, the amount of contamination contribution, and the characteristics of spatial distribution. Secondly, using GDM and GWR, the spatial differentiation and regression analysis of each influence factor of the pollution sources were carried out to reveal the influence of each factor and interaction on the spatial distribution of the pollution sources. Finally, the PMF source analysis results, GDM and GWR analysis results are combined to quantify the driving factors of pollution sources. The objectives of this study are: (1) to analyze the level and spatial distribution characteristics of heavy metal contamination in agricultural soils in the small-scale area of the lower reaches of the Qian River. (2) Quantify the sources of heavy

metal pollution using the PMF model and analyze its spatial analysis pattern. (3) Identify the driving factors of soil heavy metal pollution sources using GDM and GWR. The results of the study revealed the potential sources of soil heavy metal pollution and their driving factors, which can provide a theoretical basis and practical basis for the prevention and control of soil heavy metal pollution in farmland in the Qian River Basin.

Materials and methods

Study area

The Qian River originate in the Liupan Mountain area in Gansu Province, one of the principal tributaries of the Wei River, the lower portions are the primary irrigation-irrigated agricultural areas. It flows into Wei River, near the village of Didian in Chencang District, Baoji. The lower portions of the Qian River, which have a watershed area of 222.736 km², were chosen as the study region for this article. Average annual temperature of 11.3 °C, average annual precipitation of 634.6 mm, a population density of 673 persons per km², a DEM of 507–600 m, the study region has a warm-temperate continental monsoon climate. The lower river has wider terraces, more tributaries, fertile land, and better irrigation conditions, making it an agricultural area. There is a lot of traffic and population in the study region, including the Yinkun and Baofeng Expressways and the Longqian Railway. Across the river's western bank are Linghua Village, Nanwan Village, Zhangjiazhai Village, Minjiaya Village, and Didian Village. Across the river's eastern bank are Changqing Village, Qian River Bus Station, Shaanxi Dongling Smelting Company, and Fengxiang Railway Station.

Soil sampling and chemical analysis

Sampling was conducted from June 5 to 8, 2019, in conjunction with local conditions, with a total of 28 sampling points. In actual sampling, GPS was used to locate the sampling points, and four samples of topsoil (0–20 cm) were collected from the area of 20–50 m around each sampling point using the plum-type and checkerboard sampling methods, and were combined into a single sample of approximately 1 kg. The distribution of sampling points is shown in Fig. 1. The soil samples were thoroughly mixed and picked to remove impurities such as stones and roots, dried naturally, ground and passed through a 100-mesh nylon sieve. The sieved soil samples were weighed 0.1000 g and digested by HClO₄-HNO₃-HF-H₂O₂ method. Inductively coupled plasma mass spectrometer (ICP-MS, NexION350X, PE, US) was used to determine the contents of Cd, As, Cu, Pb, Zn, Cr, Mn and Ni in the samples, which was repeated three times, and the recoveries were controlled to be in the range of 90–105% using the GSS-25 as the quality control sample, and the agents used were all of superior purity. The detection limits of Cd, As, Cu, Pb, Zn, Cr, Mn and Ni were 0.08, 0.48, 0.06, 0.07, 0.63, 0.14, 0.17 and 0.43 mg·kg⁻¹, respectively.

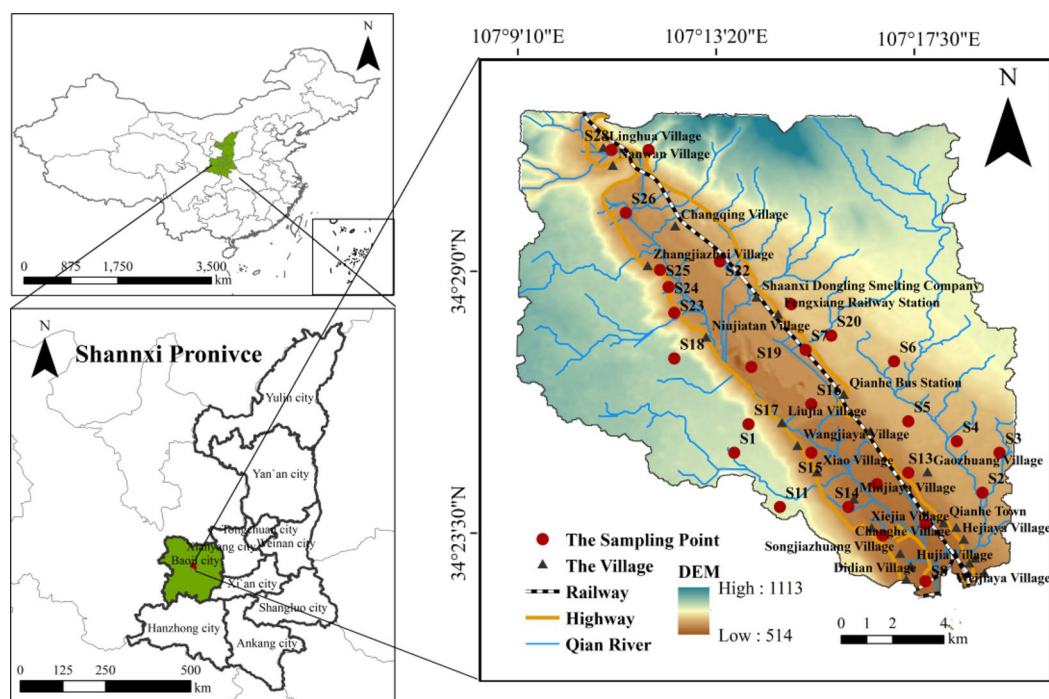


Fig. 1. Distribution of sampling locations in the study area. This map was created using ArcGIS Desktop 10.8 (<https://www.esri.com/>). The vector data for administrative boundaries were obtained from the 1:1,000,000 public edition of the basic geographic information dataset (2021), available through the National Catalogue Service for Geographic Information of China (<https://www.webmap.cn/>). DEM data is derived from remotely sensed imagery from the Geospatial Data Cloud (<http://www.gscloud.cn/>).

Influencing factors

This study makes reference to the current soil pollution results to be able to determine the primary factors influencing the soil heavy metals in the study region^{40,41}. Six anthropogenic factors: land use type, GDP per capita, energy consumption, distance to road, distance to river, and distance to village. Seven natural factors: Digital Elevation Model (DEM), temperature, precipitation, soil type, vegetation cover, Sand, and Clay—were chosen. Digital Elevation Model (DEM) is a model that enables the digital simulation of surface topography from limited topographic elevation data. By sampling and processing elevation data from the Earth's surface, a set of ordered arrays of values are generated to represent the elevation of different places. It can be accessed through Geospatial Data Cloud (<http://www.gscloud.cn/>). Soil texture and soil type data were obtained from the Chinese soil dataset based on the harmonized world soil database (harmonized world soil database, HWSD) of the National Tibetan Plateau Scientific Data Center. Vector data from the National Geographic Information Resources Catalogue Service System (<http://www.webmap.cn/>) and the rest of the data from the Resource Environment Data Cloud Platform (<http://www.resdc.cn/>). Preprocessing data accordingly in ArcGIS 10.8 and ENVI 5.3. The impact factors were processed in ArcGIS 10.8 as 200 m x 200 m grid data and the corresponding grid center values were extracted. Using the natural breakpoint approach, the following variables were categorized: temperature, precipitation, land use type, Vegetation Cover, and DEM into 5 groups. Soil type, Clay, and Sand were all classified into 5 categories according to sand, powder, and clay content. GDP per capita and energy consumption were classified into 5 categories according to the K-mean method. Distance factor is created with the help of MultipleRing Buffer analysis of ArcGIS toolbox for roads, rivers and villages at different distances to analyze the effect of different distances on the accumulation of heavy metals. The classification results are shown in Fig. 2.

Research methods

Source apportionment

The PMF model is a straightforward and efficient source analysis technique for spectroscopy of unknown source composition, which is mostly used for source analysis of heavy metals in soil, sediment, and atmospheric particulate matter. The model begins by decomposing the elemental concentration matrix into a factor score matrix (g), a factor loading matrix (f), and a residual matrix (e). The formula can be written as:

$$X_{ij} = \sum_{k=1}^p g_{ik} f_{kj} + e_{ij} \quad (1)$$

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

$$U = \sqrt{(EF \times C)^2 + (MDL \times 00.5)^2} \quad (3)$$

$$U = \frac{5}{6} \times MDL \quad (4)$$

Where: C is the measured concentration of heavy metal elements ($\text{mg}\cdot\text{kg}^{-1}$); MDL is the method detection limit of the elemental measuring instrument. During the operation of the model, it is necessary to repeat the calculation and select the optimal result according to the signal-to-noise ratio S/N (Signal to Noise) and Q value. The contribution rate of pollution sources is determined based on prior knowledge of heavy metal sources^{42,43}.

Geo-detector model

Geo-detector are tools for detecting and quantitatively analyzing interactions between influencing factors, based on the theory of spatial differentiation and using spatial statistical methods. The model consists of four sub-models: factor detector, ecological detector, risk detector, and interaction detector⁴⁴, of which the factor detector is the core component. The formula can be written as:

$$Y_i = \beta_0 (\mu_i, \nu_i) + \sum_k \beta_k (\mu_i, \nu_i) X_{ik} + \varepsilon_i \quad (5)$$

Geographically weighted regression model

GWR is a local spatial regression model that can quantitatively respond to the values of spatial auto-correlation as well as the spatial heterogeneity of each parameter in different regions⁴⁵. The formula can be written as:

$$Y_i = \beta_0 (\mu_i, \nu_i) + \sum_k \beta_k (\mu_i, \nu_i) X_{ik} + \varepsilon_i \quad (6)$$

Results and discussion

Descriptive statistics of agricultural soil heavy metal content

The descriptive statistics of heavy metals in agricultural soils in the study area are analyzed in Table 1. The order of the mean content of the eight heavy metals is: Mn > Zn > Ni > Cr > Cu > As > Pb > Cd. The soil in the research region has a pH of 7.1 to 7.8, with the screening value of pH > 7.5 in the Risk Control Standard for Soil Pollution

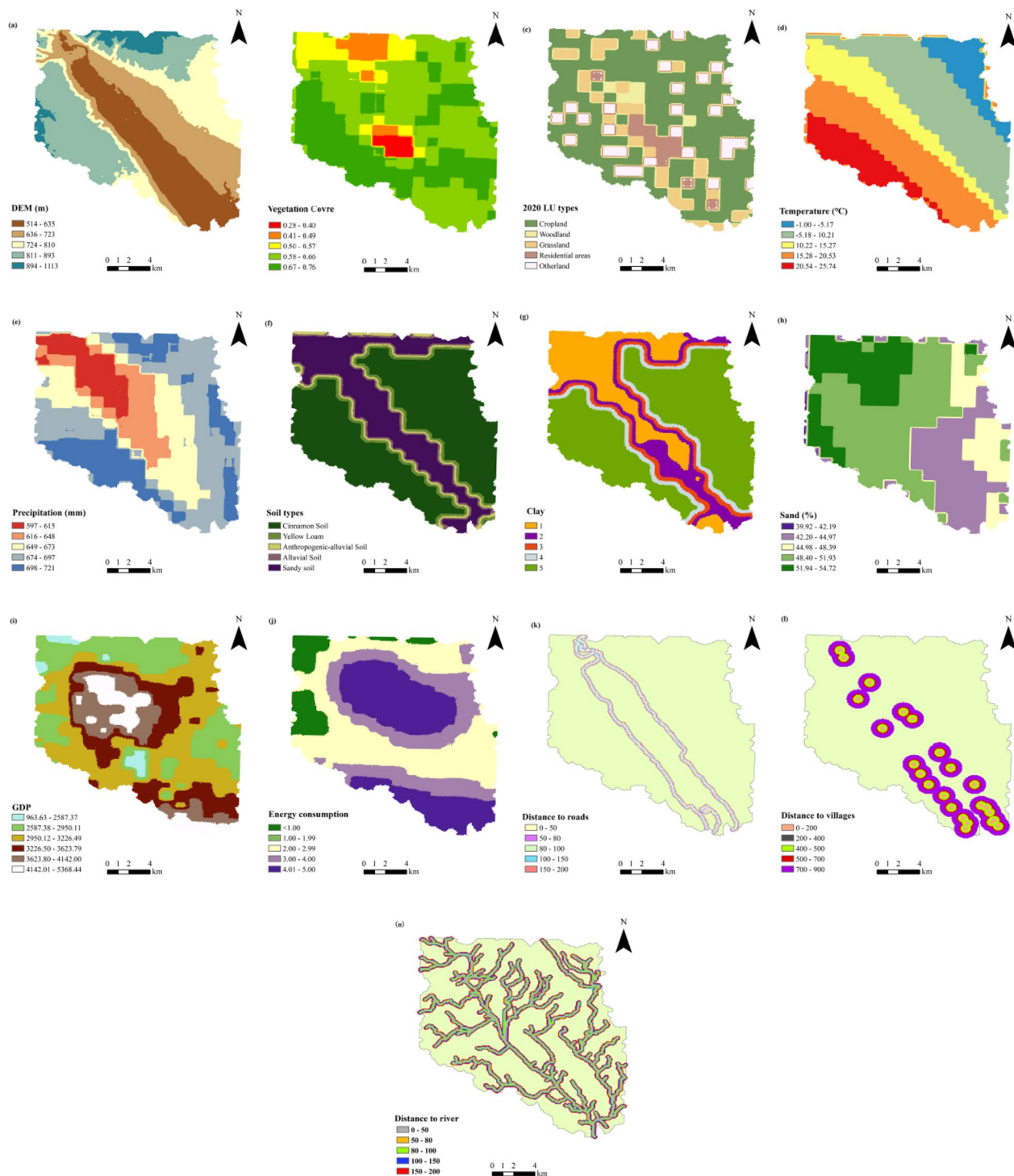


Fig. 2. The spatial distribution of the influence factors. Maps were drawn by authors, using ArcGIS 10.8(<https://www.esri.com/>).

on Agricultural Land (GB 15618-2018) as the evaluation standard, and each heavy metal's average value is below the risk control value. However, the maximum value for elemental As was $41.53 \text{ mg}\cdot\text{kg}^{-1}$, this is 1.66 times the value of the risk screening. The average content values of the elements Ni, Zn, As and Cd were 42.16, 102.07, 18.23 and $0.32 \text{ mg}\cdot\text{kg}^{-1}$, which were 1.46, 1.47, 3.7 and 6.17 times higher than the background values of the soils in Shaanxi Province. Shows that the research area's soil has been pollution to varied degrees, which could be dangerous for people's health. This finding is in line with Ning's study⁴⁶. Among the eight heavy metals, the highest value of coefficient of variation (CV) was 0.58 for Cd element. Suggests that the distribution of elemental Cd in the study area is more dispersed and may be influenced by other environmental factors²⁰.

Elements	Max	Min	Mean	SD	CV	Soil background values in Shaanxi Province	(GB 15618-2018) Screening value
Cr	51.07	3.084	22.61	8.81	0.39	62.5	250
Mn	493.78	135.81	255.76	65.37	0.26	557	–
Ni	67.02	16.56	42.16	8.45	0.21	28.8	190
Cu	33.60	4.43	19.78	5.98	0.30	21.4	100
Zn	199.57	20.91	102.07	44.08	0.43	69.4	300
As	41.53	5.48	18.23	9.02	0.49	11.2	25
Cd	0.58	0.08	0.32	0.18	0.58	0.094	0.6
Pb	13.69	1.43	10.01	2.49	0.25	21.4	170

Table 1. Descriptive statistics of soil heavy metal concentrations ($\text{mg}\cdot\text{kg}^{-1}$). “–” indicates that no data are available for the relevant standard.

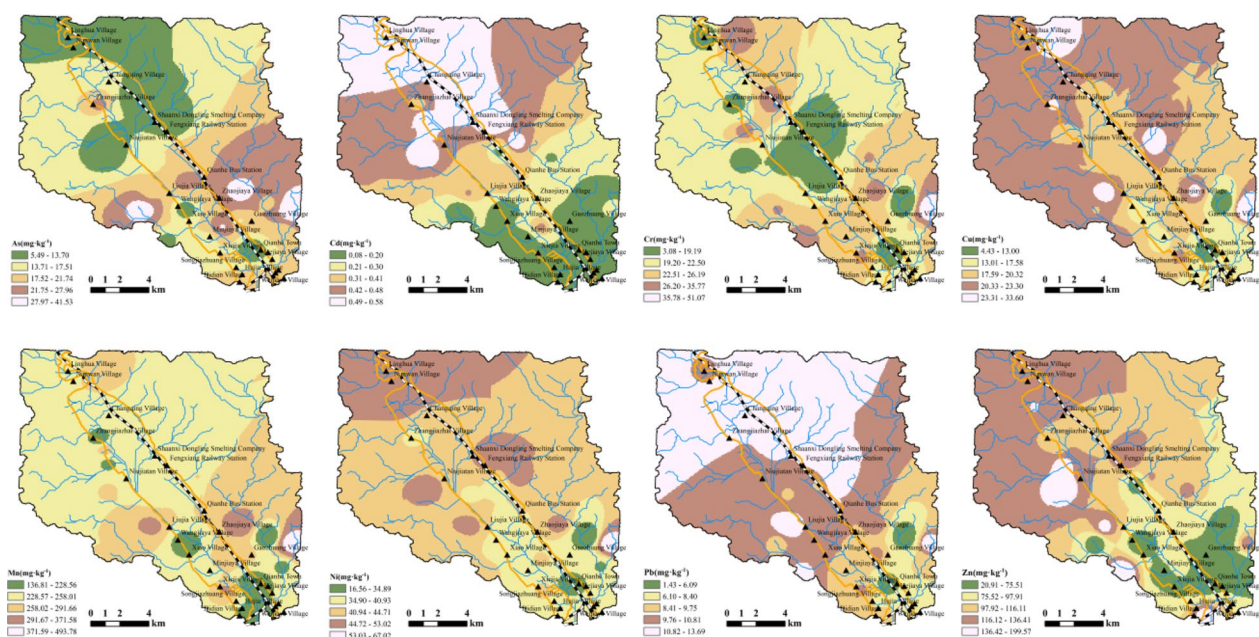


Fig. 3. Spatial distribution of heavy metal concentrations in the study area. Maps were drawn by authors, using ArcGIS 10.8 (<https://www.esri.com/>).

Spatial distribution of heavy metals in agricultural soils

The spatial characteristics of soil heavy metal content values in the study area are shown in Fig. 3. The spatial distribution of each heavy metal content showed significant variance. The southeastern region of the research area is where As, Cr, and Mn element high-value locations are primarily concentrated. Higher Cd content is mainly concentrated in the intersection areas of Fengxiang railway station, Longqian railway trunk line, and Baofeng highway. The majority of the high Pb locations are concentrated in the areas around Nanwan, Liujia village, and Fengxiang railway station. In contrast, in the areas around Fengxiang railway station and Linghua village, higher values of Ni. Cu has a more sporadic geographical distribution; the high-value sites are primarily located in Nanwan and Zhangjiazhai villages, etc.

Sources of heavy metals in agricultural soils

The operating principle of PMF is based on the matrix decomposition of the least iterative squares method⁴⁷. After setting 3–6 factors respectively for multiple operations, determine that when the number of factors is 4, the predicted value is close to the measured value. Except for Cr and Cu, where the fitted curve R^2 is 0.673 and 0.528, other elements have fitted curve R^2 values larger than 0.8. It shows that the PMF model runs better overall and the number of factors selected can explain the data information better.

Figure 4a represents the spectrum of source components in soil resolved by the PMF model. Ni, Mn, and Zn all made a large contribution to factor 1, contributing 76%, 40%, and 39%, respectively. Studies have shown that the element Mn is commonly used as a marker for natural sources and is mainly influenced by crustal activity and soil erosion⁴⁷. The lower parts of the Qian River have higher terrain on both sides, wider river valley terraces, and Mn elements build up in the soils downstream due to river erosion. Ni is an element of the Fe group that readily binds to oxides in the soil, and the formation process is closely linked to the soil-forming material. Studies have also confirmed that the geological backdrop mostly controls the amount of Ni in soil^{22,48}. The coefficients

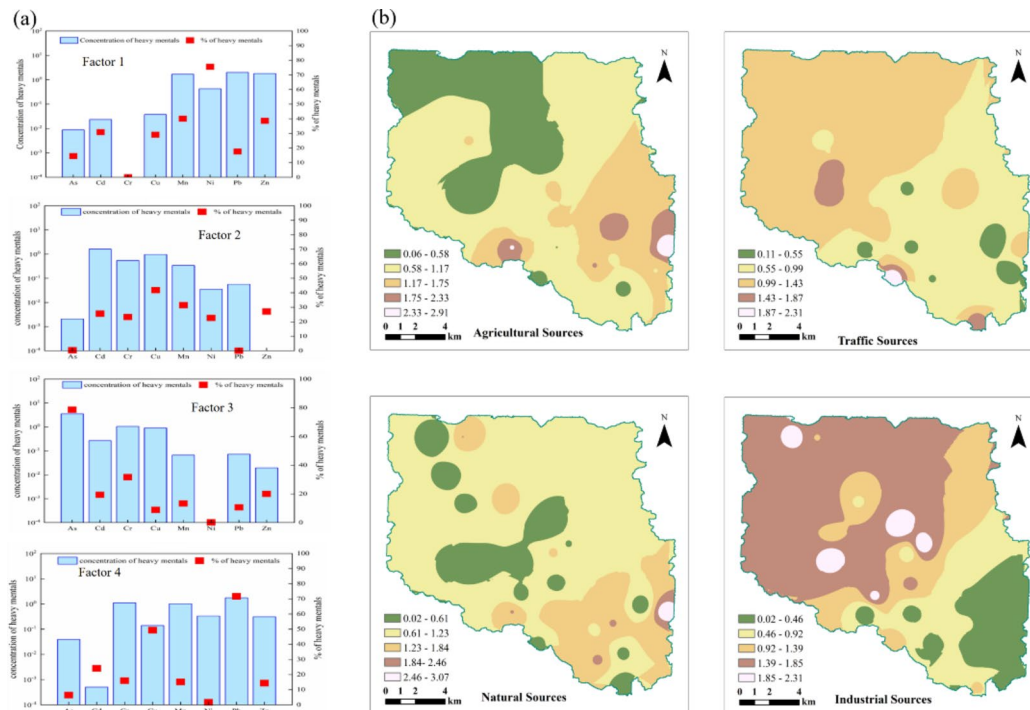


Fig. 4. Analysis of soil heavy metal pollution source results. (a) Is the composition spectrum of pollution sources resolved by the PMF model; (b) is the spatial distribution of heavy metal pollution sources. Maps were drawn by authors, using ArcGIS 10.8 (<https://www.esri.com/>).

Interactive results	Explanatory power
Nonlinear attenuation	$q(x_1 \cap x_2) < \text{Min}[q(x_1), q(x_2)]$
Single-factor attenuation	$\text{Min}[q(x_1), q(x_2)] < q(x_1 \cap x_2) < \text{Max}[q(x_1), q(x_2)]$
Mutually independent	$q(x_1 \cap x_2) = q(x_1) + q(x_2)$
Two-factor enhancement	$q(x_1 \cap x_2) > \text{Max}[q(x_1), q(x_2)]$
Nonlinear enhancement	$q(x_1 \cap x_2) > q(x_1) + q(x_2)$

Table 2. Classification of interaction results.

of variation for Ni, Mn, and Zn were 0.21, 0.26, 0.43 respectively, as indicated in Table 2. The coefficients of variation are all small, indicating that the elements are less affected by human disturbance. Furthermore, it has been demonstrated that secondary Zn enrichment by rainfall runoff and other factors can occur in riparian soils⁴⁹. Therefore, factor 1 can represent natural sources such as rainwater washout and weathering of parent rock minerals.

Factor 2 contributes more to Cu and Cd with 42% and 36% respectively. Elemental Cd is considered to be an industrial characteristic element⁵⁰, and very high levels of Cd are contained in waste slag and dust. A significant coal-fired smelter was also discovered to be present downstream of the research region, according to the survey. Furthermore, Table 2 illustrates that the coefficient of variation for Cd is 0.58, which could be impacted by regional sources of contamination. Cu elements may come from industrial production processes⁵¹, where industrial and mining operations generate enormous amounts of copper-rich dust, which through air deposition attaches to the soil and builds up over time. Additionally, throughout the mining process, a lot of tailings and residues are left on the surface of the ground, which makes it simple for copper to diffuse and migrate into the surrounding region and enrich the soil with copper through rainfall leaching⁵². Therefore, Factor 2 is primarily an industrial source affected by coal combustion.

Factor 3 greatly contributes to the occurrence of As and Cr at 79% and 32%, respectively. Studies show that a substantial source of arsenic contaminate in soil is the widespread use of insecticides and herbicides in plantation⁵³. Furthermore, the application of As supplements in chemical fertilizers and animal waste may result in soil As contamination⁵⁴. The research area's main agricultural activities are found on both banks of the lower Qian River. For farmers, peripheral agriculture is a substantial source of revenue. Large amounts of pesticides and fertilizers are added to the soil to increase crop yields. As a result, As and Cr residues build up in the soil. Therefore, it can be judged that factor 3 is mainly influenced by agricultural sources.

Factor 4 contributes 72% to Pb. Research has demonstrated that the burning of leaded gasoline, brake eroding, and catalytic combustion result in a notable increase in elemental Pb enrichment⁵⁵, resulting in the most serious contamination of Pb in soils on both sides of traffic arteries. It was discovered that there are a lot of important roads close to the research region with a lot of traffic and frequent vehicle access, which could cause a lot of Pb to build up in the soil. So, it may be concluded that Factor 4 is primarily affected by traffic sources.

In summary, there are four primary sources of heavy metals in agricultural soils in the Qian River watershed, namely, natural sources, industrial sources, agricultural sources and traffic sources. Soil heavy metal source contribution ratios from PMF analysis were 19.12%, 23.42%, 36.85%, and 20.61%, respectively. Expression of the spatial distribution of the contribution ratios of the four sources of heavy metals in soil Fig. 4b. There is a lot of variation in how different sources are distributed in space. The research area's agricultural sources are primarily focused in the south-east, with a significant distribution of farmland cultivation regions close to villages. Traffic sources are mainly concentrated in the area where the main Long qian railway line and highway intersects. Natural sources are dispersed. Industrial sources are mainly concentrated in the vicinity of Liujiawan village and Linghua village, with industrial plants and mines such as coal-fired power plants.

Influence factor analysis

Factor detection

Factor identification clarifies the explanatory capacity of several independent factors on the degree of effect of the dependent variable. The dominating factor's explanatory power on the dependent variable increases with its q value. Factors ranked differently in their degree of influence on heavy metals, demonstrating the heterogeneity of the heavy metals' spatial variation process³¹. Four pollution sources were analyzed as independent variables, and 13 factors were investigated by factor detection for their influence on the sources of pollution. Although the explanatory strength of the parameters for each of the four pollution sources varies slightly, as Fig. 5 illustrates, in general, soil type, distance from the village, and DEM all exhibit substantial explanatory power for the geographical distribution of pollution sources. The first dominant factor for industrial sources (IS) was distance from villages (0.683), followed by DEM (0.170) and then soil type (0.162). The first dominant factor for natural sources (NS) was the distance to villages (0.348), followed by GDP (0.081) and then the night light index (0.079). The first dominant factor for traffic sources (TS) was distance to the village (0.507), followed by DEM (0.227) and then soil type (0.200). The first dominant factor for agricultural sources (AS) was distance to villages (0.628), followed by soil type (0.175) and then Vegetation Cover (0.172). The study area is located in the urban-rural area of Baoji City, where residents' self-built houses and industrial and mining enterprises are staggered. Furthermore, the study area's geography makes agricultural cultivation unsuitable for large-scale development, and the majority of the small patches of reclaimed farmland are found in the areas next to villages. Additionally, to facilitate people's mobility, roads are constructed close to mining and industrial facilities as well as settlements. As a result, distance to villages becomes the primary driver for each source.

Overall, distance to village, soil type and DEM had strong explanatory power for four sources of pollution. On the one hand, compared to comparable locations with greater heights and more undulating topography, the flat land between the river and the mountains considerably reduces the expense of establishing villages. Creating a few conducive factors for economic expansion in order to promote the growth of the transportation, industrial, and agricultural sectors. In contrast, due to barriers like mountains and rivers, industrial site selection, agricultural land reclamation, and transit route development frequently concentrate on settlements in the studied region. Making distance to villages has a greater impact on industrial, agricultural, and transport sources. Industry, agriculture and, transport can also be a reflection of the economic strength of a region.

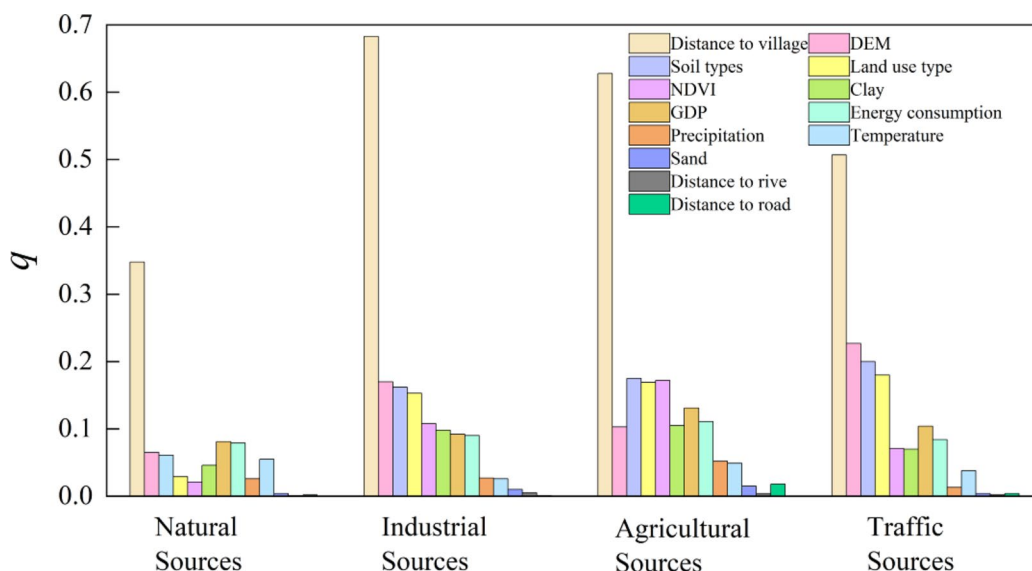


Fig. 5. Detection (q) of geographical environment factors.

Geographical variables have been demonstrated to significantly influence the economic development of 35.77% of China's counties, according to studies⁵⁶. Of the counties with significant impact, 29.39% experienced positive effects while 70.61% were negatively affected by the economic development of the region. The average elevation of the county is significantly and negatively correlated with economic development, also indicating that topographical factors have a greater impact on economic development⁵⁷. Therefore, the Government's decision-making authorities should take into account the impact of different factors on economic development and make reasonable development plans, so as to promote the coordinated development of the local economy.

Interaction detection

Interaction detectors can be used to investigate the interaction effects of different influencing factors on the spatial distribution of pollution sources in the study area. If the q value of the interaction between factors is greater than the q value of a single factor and shows bilinear enhancement and non-linear enhancement, it means that the influence of the interaction between environmental impact factors is higher than the influence of a single factor³¹. In this paper, the top 4 influencing factors were mainly selected to be analyzed, as shown in Table 3. For industrial, transport, and agricultural sources, the strongest interaction values for distance to village \cap precipitation were 0.794, 0.670, and 0.794, respectively. The interaction value of the distance to village \cap temperature was the second highest among industrial, transport, and natural sources, with 0.790, 0.663, and 0.785, respectively. In natural sources, distance to village \cap temperature and distance to village \cap Vegetation Cover were the top two interaction values with 0.486 and 0.465, respectively. Suggests that the interplay between natural and anthropogenic factors contributes more to the source of pollution. Overall, the factor interaction of distance to village with precipitation, temperature, and DEM was more than the explanatory capacity of a single pollution source component. It has been demonstrated that the features of soil heavy metal contamination distribution is more significantly impacted by the interplay of environmental and human variables, and that this interaction can more clearly expose the mechanisms underlying the spatial variability of soil heavy metals^{34,58}.

The interaction detector analysis' findings indicated that the two-factor's explanatory power was somewhat more than the one-factor's, which is in line with what other researchers have discovered⁵⁹. Factors such as precipitation, temperature, GDP, and DEM had low explanatory power for pollution sources as a single factor, but higher strength of interpretation than one-factor in the interaction in combination with distance to villages. Reflecting the combined impact of human activities on the process of changes in the spatial distribution of sources of pollution. Human activity has changed the characteristics of how soil heavy metals are distributed in their natural condition, creating a new geographical distribution pattern⁴⁹. Compound effects of factors hold a crucial position in the accumulation and migration of soil heavy metal element in the study area. Under the joint action of natural, industrial, agricultural, and traffic sources, the distribution of the high value areas of heavy metals in farmland soils in the study area varies greatly in the east-west direction.

GWR analysis of dominant factors

For local spatial regression analysis, the GWR model was further used to investigate the district variability of the prevailing factors in direction and intensity. The factors were also quantitatively spatially visualized and their spatial regression coefficients were analyzed. The larger the coefficient, the greater the influence of the dominant factor on the source of pollution somewhere in space⁶⁰. By analyzing the influencing factors of industrial, natural, transport, and agricultural sources, the top three influencing factors in terms of explanatory power were screened and analyzed in GWR.

Negative regression coefficients in the dominant factor of NS occupied a majority of the sample points in the research area, indicating that distance to villages (Fig. 6a), GDP (Fig. 6b), and energy consumption index (Fig. 6c) are negatively inhibiting natural sources. The negative high values of all three impact factors were fragmented and focused in the lower part of the research area near Gaozhuang village, Qian River town, and Didian village. The sub-negative highs are weakening as they continue to move towards the north-west of the study area, suggesting that the impact of the three influences on NS is decreasing from the southeast to the northwestern part of the research area. The strongest explanatory power of the IS impact factor is the distance to the village (Fig. 6d), with regression coefficients of $(-1.36$ to $1.91)$. Regression coefficients exhibit significant spatial differentiation, primarily displaying a positive influence. This was followed by DEM (Fig. 6e) and soil type (Fig. 6f), with regression coefficients of $(-2.6122$ to $2.04)$ and $(-1.68$ to $2.16)$ respectively. The spatial feature of the regression coefficients shows that most of the areas with high values of the three impact factors are located in the central part of the study area, where the Qian River Industrial Park is located, including smelters, railway

Interaction detection	Interaction factor 1 $q(x_1 \cap x_2)$	Interaction factor 2 $q(x_1 \cap x_2)$	Interaction factor 3 $q(x_1 \cap x_2)$	Interaction factor 4 $q(x_1 \cap x_2)$
IS	Distance to village \cap precipitation (0.794)	Distance to village \cap temperature (0.790)	Distance to village \cap DEM (0.771)	Distance to village \cap GDP (0.747)
NS	Distance to village \cap temperature (0.486)	Distance to village \cap Vegetation Cover (0.465)	Distance to village \cap precipitation (0.453)	Distance to village \cap GDP (0.449)
TS	Distance to village \cap precipitation (0.670)	Distance to village \cap temperature (0.663)	Distance to village \cap GDP (0.654)	Distance to village \cap DEM (0.615)
AS	Distance to village \cap precipitation (0.794)	Distance to village \cap temperature (0.785)	Distance to village \cap GDP (0.735)	Distance to village \cap DEM (0.724)

Table 3. Interaction effects on soil heavy metals.

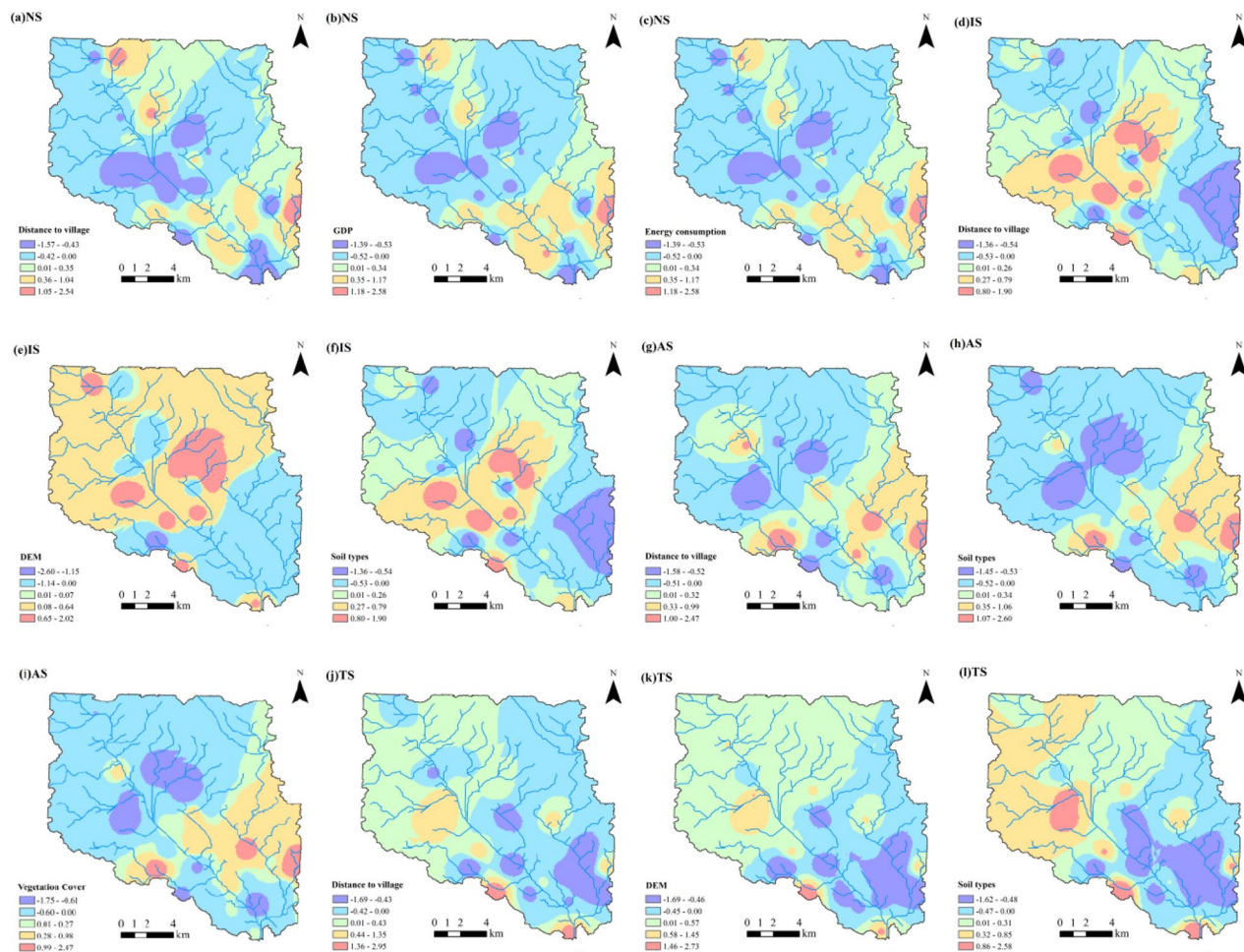


Fig. 6. Regression coefficients of leading factors. Maps were drawn by authors, using ArcGIS 10.8 (<https://www.esri.com/>).

stations, and metal manufacturing industries. By comparing with the DEM classification in Fig. 2, the highest regression coefficients are in the DEM 635–723 m interval. These areas are relatively low in altitude, flat terrain, densely with roads, and have a high concentration of factories. The impact of anthropogenic activities is more pronounced and is responsible for the concentration of pollution from heavy metal industrial sources in the study area.

The three influencing factors in the AS, distance to the village (Fig. 6g), soil type (Fig. 6h), and Vegetation Cover (Fig. 6i), have a predominantly positive contribution to the AS. The locations with high values of distance to villages are primarily dispersed in the southern portion of the research area, which is densely populated with villages, flat terrain, and more farmland distribution, according to the spatial distribution of the regression coefficients. By comparing with the soil type classification in Fig. 2, the areas of high values of soil type regression coefficients are mainly located in the sandy soil and brown soil intervals. The southern region is where the Qian River and the Wei River meet, where sediment brought in by the rivers of the upstream region gathers, making the soil deep and fertile and benefiting from the development of agricultural production. The three influencing factors in the TS, distance to villages (Fig. 6j), DEM (Fig. 6k), and soil type (Fig. 6l) have a clear regional differentiation in their influence on the traffic sources. The regression coefficients are sporadically distributed with alternating positive and negative high values. In the vicinity of Xiao Village, Songjia Village, and Niujiatan Village, positive and high values of DEM regression coefficients are primarily distributed, which positively contributes to the impacts of traffic sources.

Source-oriented impact factor analysis

Combining the PMF model, GDM and, GWR model influence factor analysis results, the relationship between environmental factors, soil heavy metals and, pollution sources was plotted (Fig. 7). For heavy metal elements, DEM, land use type, and soil type have a considerable explanatory power. The research location is in a river valley, with a river floodplain in the middle and mountains on each side that range in elevation from 514 to 1113 m. Elevation has been demonstrated to influence flood events and river flow direction, which in turn influences the distribution pattern of heavy metals⁶¹. Prolonged fluvial alluviation brings in organic matter inputs, and the increase in organic matter enhances the complexation of heavy metals, that accumulate in Soil on both

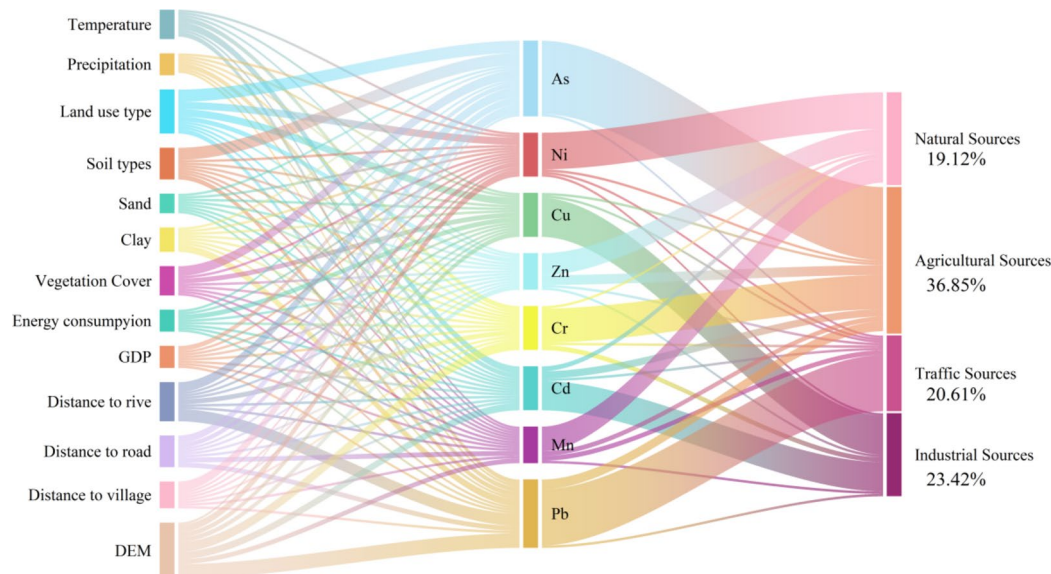


Fig. 7. Relationship between heavy metals, pollution sources, and impact factors.

sides of the river⁵⁸. Weathering, soil erosion, and crystalline activity increase the buildup of certain heavy metal components in riverbed soils downstream⁶². Soil formed by natural weathering is the product of rock weathering and erosion and deposition, the factors affecting natural soil formation are mainly parent material, climate, topography, biology, and time⁶³, and its formation and evolution are a long process. The mixture of sediment, gravel, and soil in the upstream area brought about by gravity and river alluvium cannot be immediately converted into arable soil, so environmental factors, such as different land use types and soil types, have a greater impact on the distribution of heavy metals in soil. Comparison of the results of heavy metal source analyses reveals both consistency and discrepancy. Consistency, such as DEM, is also a significant environmental driver of heavy metal pollution sources, and there is a correlation. Discrepancy, such as distance to villages was the primary environmental factor in all four source analyses, but had a lesser impact on heavy metal distribution. This is mostly due to the fact that heavy metal distribution characteristics in soil have been modified by human activity, forming new spatial distribution patterns. Villages were first established in the flatter topography of the study area, followed by a continuous process of land reclamation and development of agriculture, industry, and transport. Make the interaction of anthropogenic and natural factors the primary motivating force behind the many sources of pollution.

Discussion

The sources of heavy metals in agricultural soils are influenced by a combination of natural factors (topography, climate, soil matrices, etc.) and human activities (pesticide and fertilizer application, industrial emissions, traffic exhaust, etc.). In this study, we analyzed the heavy metal contents of the Qian River and found that the contents of Ni, Zn, As and Cd were higher than the soil background values in Shaanxi Province, and we preliminarily concluded that there may be some degree of contamination of these four elements by human activities. From the spatial distribution of the eight heavy metal elements, the spatial distribution of As, Cr and Mn is similar, and the high value areas are all in the southeastern part of the study area, which is densely populated with villages, many transportation routes, and a wide area of arable land, which is affected by agricultural and transportation activities. The spatial distribution of Cd, Cu, Pb, Zn and Ni is similar, and their high value areas are all located in the northern part of the study area, where factories such as the Evergreen Industrial Park and the Dongling Smelter are located, which are more affected by industrial emissions. In order to clarify the sources of heavy metals in the soils of the study area, we further analyzed the sources of the soils in the study area using the PMF model. We found that Ni, Mn and Zn in the study area may be derived from natural sources of rock mineral weathering and rainfall washout, and Shun's study also showed that the heavy metals Mn and Ni in the soil of northwestern farmland were derived from natural sources⁶⁴; Cu and Cd are influenced by industrial parks and non-ferrous metal smelters in the study area, and may come from industrial sources of coal burning activities, and Shao's study also showed that smelting and electroplating companies are the main source of soil Cd⁴³; As and Cr are influenced by pesticide and fertilizer application from agricultural sources, which is consistent with Lin's study⁶⁵; Pb comes from the source of traffic activities, which is also related to the dense road network and high traffic flow in the study area, and Chao et al. also showed that a large amount of Pb is deposited in the area in the lower Yellow River drinking water source area under the influence of vehicle exhaust emissions and friction between car tires and the ground⁶⁶.

In the GDM model, distance from the village was the main driver for the four sources, with soil type and elevation also affecting them to varying degrees. In this paper, our study area is located in the urban-rural area, where rural self-built houses are intricately interspersed with urban residential buildings and factories and

enterprises, and the road network is dense, and where most of the farmland is distributed closer to the rural settlements due to the topography of the area. Also some of the factory sites will be located far away from the city and for ease of transportation, these factories will be built in more accessible peri-urban areas, which may be the distance from the villages to be the main driver of heavy metal contamination of the soil in the study area. The topography of the study area is low in the middle and high in the surroundings, which may carry heavy metals from high to low places under the action of running water scouring, and in the areas with low elevation and close to the residential areas, the agricultural activities are more frequent, the use of agricultural inputs such as fertilizers and pesticides is higher, and the industrial activities and the traffic flow are also higher, which may lead to the accumulation of heavy metals. The results of the GWR model also showed that distance from the village was the main driver for the four sources of pollution. However, we can find from Fig. 6 that the effect of distance from the village on the four sources is different, with distance from the village mainly negatively inhibiting the natural sources, while positively promoting most of the agricultural, transportation, and industrial sources, probably because the greater the distance from the village, the degree of disturbance to the natural environment from human activities is gradually reduced. And the closer the distance to the village, the more the regional ecological environment is affected by the combination of factors such as anthropogenic disturbances, land use changes and socio-economic factors.

For the analysis of soil heavy metal sources, the performance of different models varies. In this study, we used PMF and GIS to identify the source, contribution and spatial distribution of soil heavy metals, but this judgment of pollution sources are based on the experience of previous studies, which can lead to the subjectivity of source analysis. The accumulation of heavy metals in soil is affected by a combination of factors, such as soil type, soil-forming parent material, land-use practices and other environmental factors, and the spatially stratified heterogeneity of these factors also affects the accumulation of heavy metals. In recent years, some scholars have introduced environmental factors as covariates through GDM and machine learning to identify the main influencing factors of heavy metals in soil, so as to objectively explain and validate the results of source analysis. For example, Sun combined GDM and PMF to more rationally explain the source of soil heavy metals in Lijiang City⁶⁷. Guo identified the sources and influencing factors of soil heavy metals in eastern Yunnan by combining the receptor model, SOM and GDM, and found that distance from roads, lithology, distance to industrial areas and land use contributed most to soil heavy metal pollution in the study area³. Bi predicted the concentration of arable soil and identified the main sources of contamination through machine learning techniques and bivariate localized Moran's indices, and found that the performance of the random forest model was superior⁶⁸. Zhang identified the sources of soil HMs based on random forest models, artificial neural networks, support vector machines, and a wide range of relevant soil properties and land use data⁶⁹. GDM and GWR are based on the theory of spatial hierarchical heterogeneity and explain the degree of spatial variance in the outcome variable by quantifying a set of independent variables and their interactions, which enables the identification of key drivers affecting the outcome variable, but they also suffer from a high degree of data dependence, and differences in the method of discretization of the independent variables and spatial scales can affect the accuracy of the results. Machine learning learns from massive amounts of data, mines the information in it, and builds relevant models for prediction and evaluation. The internal logic of its algorithms is relatively complex, and the quality of the data affects the accuracy and interpretability of the models, leading to predictions that are difficult to decipher and untrustworthy. Therefore, in the future, we can try to combine machine learning models such as receptor modeling, GDM and random forest to better understand the source and spatial distribution of soil heavy metals.

Conclusion

In this study, the PMF receptor model was combined with GDM and GWR to synthesize pollution sources and drivers to quantitatively resolve the sources of soil heavy metals and identify the main drivers of heavy metal pollution. The results showed that:

- (1) The contents of Ni, Zn, As and Cd in farmland soils in the lower reaches of the Qian River are 1.46, 1.47, 3.7 and 6.17 times higher than the background values for soils in Shaanxi Province, and there is a certain amount of contamination, especially Cd and As, which have a high degree of variability (0.58, 0.49) and are highly influenced by human activities. The average value of heavy metal content did not exceed the risk screening value of heavy metals in Chinese soils, but the individual sampling points of As element were higher than the risk screening value of soil contamination, indicating that the overall risk of heavy metal pollution in soils in the study area is controllable, but attention should be paid to special areas.
- (2) The spatial distribution map of heavy metals shows that the high values of As, Cr and Mn are all located in the southeastern part of the study area, which is due to the accumulation caused by agricultural activities, traffic pollution and geological background in the southeastern part of the area; the high values of Cd, Cu, Pb, Zn and Ni are all located in the northern part of the study, which is probably due to the concentration of industrial activities in the area, coupled with some topography and traffic activities, which lead to the accumulation of heavy metals.
- (3) The results of PMF model showed that the sources of soil heavy metals in the study area were natural (19.12%), industrial (23.42%), agricultural (36.85%) and transportation (20.61%) sources, Ni, Mn and Zn were mainly from natural sources, Cd and Cu were mainly from industrial sources, As and Cr were controlled by the agricultural activities were mainly from agricultural sources, Pb was mainly from the transportation sources; of which agricultural sources are the main source of soil heavy metal pollution in the study area.
- (4) The GDM results showed that distance from village, soil type, elevation and land use type were the main drivers of the four pollution sources. The explanatory power of the GDM factor interaction was higher than that of the single factor. The interaction of spatial distribution of heavy metal pollution sources was mainly

the interaction of distance to village with natural factors such as precipitation, temperature and elevation, etc. The GWR results showed that distance to village, as the most important driver of the four pollution sources, had its high values mainly in the central and southern parts of the study area.

Although this study revealed the sources and influencing factors of heavy metals in agricultural soils in the Qian River watershed through PMF combined with GDM and GWR, there are still limitations. For example, the distribution of sampling points may not fully capture the spatial distribution of the entire watershed. Secondly, there is some subjectivity in the choice of influencing factors, and the accumulation of heavy metals in soil is influenced by various environmental factors. In the future, additions can be considered in terms of source-diffuser-receptor relationships and environmental factors affecting the process of heavy metal accumulation in soils, such as prevailing wind direction, data from the inventory of industrial pollutant emissions, and the inventory of data on the application of agricultural fertilizers. Methods such as receptor-based modeling and machine learning models have advantages in heavy metal source resolution, concentration prediction, and environmental covariate assessment, but are also overly dependent on data accuracy. Therefore, the joint application of multiple methods to compensate for the deficiencies of different source analysis methods better improves the accuracy of soil heavy metal source analysis.

Data availability

All the data are included in the article.

Received: 29 September 2024; Accepted: 15 April 2025

Published online: 23 April 2025

References

- Zhang, X. et al. Spatial distribution of metal pollution of soils of Chinese provincial capital cities. *Sci. Total Environ.* **643**, 1502–1513 (2018).
- Cui, W. et al. Characterization and source apportionment of heavy metal pollution in soil around red mud disposal sites using absolute principal component scores-multiple linear regression and positive matrix factorization models. *Environ. Geochem. Health.* **46** (12), 492 (2024).
- Guo, G. et al. Quantitative source apportionment and associated driving factor identification for soil potential toxicity elements via combining receptor models, SOM, and geo-detector method. *Sci. Total Environ.* **830**, 154721–154721 (2022).
- Kouchou, A. et al. Evaluation of the environmental and human health risk related to metallic contamination in agricultural soils in the mediterranean semi-arid area (Saiss plain, Morocco). *Environ. Earth Sci.* **79**, 131 (2020).
- Wang, J. et al. Source apportionment of heavy metal and their health risks in soil-dust fall plant system near by a typical non-ferrous metal mining area of Tongling, Eastern China. *Environ. Pollut.* **254**, 113089 (2019).
- Liang, J., Wu, H. B. & Wang, X. X. Distribution characteristics and health risk assessment of heavy metals and PAHs in the soils of green spaces in Shanghai, China. *Environ. Monit. Assess.* **191** (6), 345 (2019).
- Liang, J. et al. Spatial distribution and source identification of heavy metals in surface soils in a typical coal mine City, Lianyuan, China. *Environ. Pollut.* **225**, 681–690 (2017).
- MEPRC (Ministry of Environmental Protection of the People's Republic of China). *Bulletin of National Soil Pollution Survey (in Chinese)* (2014).
- Yang, L. et al. Metal(loid)s Spatial distribution, accumulation, and potential health risk assessment in soil-wheat systems near a Pb/Zn smelter in Henan Province, central China. *Int. J. Environ. Res. Public Health.* **19** (5), 2527 (2022).
- Yu, J. et al. Major influencing factors identification and probabilistic health risk assessment of soil potentially toxic elements pollution in coal and metal mines across China: A systematic review. *Ecotoxicol. Environ. Saf.* **274**, 116231 (2024).
- Hu, B. et al. Current status, Spatial features, health risks, and potential driving factors of soil heavy metal pollution in China at Province level. *Environ. Pollut.* **266**, 1–21 (2020).
- Yang, Q. et al. A review of soil heavy metal pollution from industrial and agricultural regions in China: pollution and risk assessment. *Sci. Total Environ.* **643**, 1502–1513 (2018).
- Bernardino, C. A. R. et al. Metal accumulation in roadside soils of Rio de Janeiro, Brazil: impact of traffic volume, road age, and urbanization level. *Environ. Monit. Assess.* **191** (3), 156 (2019).
- Jin, Z. & Lv, J. Integrated receptor models and multivariate Geostatistical simulation for source apportionment of potentially toxic elements in soils. *Catena* **194**, 104638 (2020).
- Srivastava, A., Sengupta, B. & Dutta, S. A. Source apportionment of ambient VOCs in Delhi City. *Sci. Total Environ.* **343**, 207–220 (2005).
- Wei, H. W. et al. Spatio-Temporal characteristics and source apportionment of water pollutants in upper reaches of Maotiao river, Southwest of China, from 2003 to 2015. *J. Environ. Inf.* **37** (2), 93–106 (2021).
- Zerizghi, T. et al. An integrated approach to quantify ecological and human health risks of soil heavy metal contamination around coal mining area. *Sci. Total Environ.* **814**, 152653 (2022).
- Liao, S. et al. The quantitative source apportionment of heavy metals in peri-urban agricultural soils with UNMIX and input fluxes analysis. *Environ. Technol. Innov.* **21**, 101232 (2021).
- Chen, Z. et al. Combination of UNMIX, PMF model and Pb-Zn-Cu isotopic compositions for quantitative source apportionment of heavy metals in suburban agricultural soils. *Ecotoxicol. Environ. Saf.* **234**, 113369 (2022).
- Fei, X. et al. Contamination assessment and source apportionment of heavy metals in agricultural soil through the synthesis of PMF and Geog-Detector models. *Sci. Total Environ.* **747**, 141293 (2020).
- Al-Dabbous, A. N. & Kumar, P. Source apportionment of airborne nanoparticles in a middle Eastern City using positive matrix factorization. *Environ. Sci. Process. Impacts.* **17**, 802–812 (2015).
- Hu, B. et al. Spatio-temporal variation and source changes of potentially toxic elements in soil on a typical plain of the Yangtze river delta, China (2002–2012). *J. Environ. Manag.* **271**, 110943 (2020).
- Guan, Q. et al. Source apportionment of heavy metals in farmland soil of Wuwei, China: comparison of three receptor models. *J. Clean. Prod.* **237**, 117792 (2019).
- Wang, G. et al. Traffic-related trace elements in soils along six highway segments on the Tibetan plateau: influence factors and Spatial variation. *Sci. Total Environ.* **581–582**, 811–821 (2017).
- Zhao, Y. et al. Cadmium source identification in soils and high-risk regions predicted by geographical detector method. *Environ. Pollut.* **263**, 114338 (2020).
- Azizi, K., Ayoubi, S. & Nabiollahi, K. Predicting heavy metal contents by applying machine learning approaches and environmental covariates in West of Iran. *J. Geochem. Explor.* **233**, 1 (2022).

27. Zhang, H. Z. et al. Identifying driving factors and their interacting effects on sources of heavy metal in farmland soils with geodetector and multi-source data. *Huanjing Kexue*. **44** (4), 2177–2191 (2023).
28. Nanos, N. & Martín, J. A. R. Multiscale analysis of heavy metal contents in soils: Spatial variability in the Duero river basin (Spain). *Geoderma*. **189**, 554–562 (2012).
29. Wang, S. et al. Concentrations, Spatial distribution, sources and environmental health risks of potentially toxic elements in urban road dust across China. *Sci. Total Environ.* **805**, 150266 (2022).
30. Li, P. et al. County-scale temporal-spatial distribution and variability tendency of heavy metals in arable soils influenced by policy adjustment during the last decade: a case study of Changxing, China. *Environ. Sci. Pollut. Res. Int.* **22**, 17937–17947 (2015).
31. Gao, J. B. & Wang, H. Temporal analysis on quantitative attribution of karst soil erosion: a case study of a peak-cluster depression basin in Southwest China. *Catena* **172**, 369–377 (2019).
32. Ma, C. et al. Relationships between heavy metal concentrations in soils and reclamation history in the reclaimed coastal area of Chongming Dongtan of the Yangtze river estuary, China. *J. Soils Sediments*. **15**, 139–152 (2015).
33. Zhou, C. S., Chen, J. & Wang, S. J. Examining the effects of socioeconomic development on fine particulate matter (PM_{2.5}) in China's cities using Spatial regression and the geographical detector technique. *Sci. Total Environ.* **619**, 436–445 (2018).
34. Tao, H. et al. Quantifying influences of interacting anthropogenic-natural factors on trace element accumulation and pollution risk in karst soil. *Sci. Total Environ.* **721**, 137770 (2020).
35. McMillen, D. P. Geographically weighted regression: the analysis of spatially varying relationships. *Am. J. Agric. Econ.* **86**, 554–556 (2003).
36. Tu, J. Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. *Appl. Geogr.* **31**, 376–392 (2011).
37. Qu, M. K. et al. Spatially nonstationary relationships between copper accumulation in rice grain and some related soil properties in paddy fields at a regional scale. *Soil. Sci. Soc. Am. J.* **78**, 1765–1774 (2014).
38. Bhowmik, A. K. et al. Mapping human health risks from exposure to trace metal contamination of drinking water sources in Pakistan. *Sci. Total Environ.* **538**, 306–316 (2015).
39. Zhang, J., Gao, Y. & Yang, N. Ecological risk and source analysis of soil heavy metals pollution in the river irrigation area from Baoji, China. *PLoS ONE*. **16** (8), e0253294 (2021).
40. Rennert, T. & Rinklebe, J. Modelling the potential mobility of cd, Cu, Ni, Pb and Zn in mollic fluvisols. *Environ. Geochem. Health*. **39**, 1291–1304 (2017).
41. Hossain Bhuiyan, M. A. et al. Enrichment, sources and ecological risk mapping of heavy metals in agricultural soils of Dhaka district employing SOM, PMF, and GIS methods. *Chemosphere* **263**, 128339 (2021).
42. Shen, C. et al. Distribution characteristics, source analysis and ecological risk assessment of heavy metals in the typical industry reclaimed soil. *Huanjing Kexue*. **45** (3), 1769–1780 (2024).
43. Shao, F. et al. Sources apportionments of heavy metal(loid)s in the farmland soils close to industrial parks: integrated application of positive matrix factorization (PMF) and cadmium isotopic fractionation. *Sci. Total Environ.* **924**, 171598–171598 (2024).
44. Wang, J. F. & Xu, C. D. Geo-detector: principle and prospective. *Acta Geo-graph.* **1**, 1 (2017).
45. Tu, J. & Xia, Z. G. Examining spatially varying relationships between land use and water quality using geographically weighted regression I: model design and evaluation. *Sci. Total Environ.* **407**, 358–378 (2008).
46. Ning, W. J. et al. Evaluating a sampling regime for estimating the levels of contamination and the sources of elements in soils collected from a rapidly industrialized town in Guangdong Province, China. *Arch. Environ. Contam. Toxicol.* **82**, 403–415 (2022).
47. Marchand, C. et al. The partitioning of transitional metals (Fe, Mn, Ni, Cr) in Mangrove sediments downstream of a ferrallitized ultramafic watershed (New Caledonia). *Chem. Geol.* **300**, 70–80 (2012).
48. Jiang, Y. et al. Source apportionment and health risk assessment of heavy metals in soil for a Township in Jiangsu Province, China. *Chemosphere* **168**, 1658–1668 (2017).
49. Liao, X. et al. Exploring the database of soil environment survey using geo-self-organizing-map: a pilot study. *J. Geogr. Sci.* **29** (10), 1610–1624 (2019).
50. Zhou, X. Y. & Wang, X. R. Impact of industrial activities on heavy metal contamination in soils in three major urban agglomerations of China. *J. Clean. Prod.* **230**, 1–10 (2019).
51. Qiao, Y. et al. Distribution and geochemical speciation of heavy metals in sediments from coastal area suffered rapid urbanization, a case study of Shantou Bay, China. *Mar. Pollut. Bull.* **68**, 140–146 (2013).
52. Fujita, M. et al. Heavy metal contamination of coastal lagoon sediments: Fongafale Islet, Funafuti Atoll, Tuvalu. *Chemosphere* **95**, 628–634 (2014).
53. Jiang, H. H. et al. An integrated approach to quantifying ecological and human health risks from different sources of soil heavy metals. *Sci. Total Environ.* **701**, 134466 (2020).
54. Wang, L. H. et al. Arsenic accumulation, distribution, and source analysis of rice in a typical growing area in North. *China Ecotoxicol. Environ. Saf.* **167**, 429–434 (2019).
55. Cai, L. M. et al. Heavy metals in agricultural soils from a typical Township in Guangdong Province, China: occurrences and Spatial distribution. *Ecotoxicol. Environ. Saf.* **168**, 184–191 (2019).
56. Huang, Z. et al. Non-stationarity of influencing factors of China's County economic development base on a multiscale geographically weighted regression model. *ISPRS Int. J. Geo-Inf.* **12**, 109 (2023).
57. Shen, T. et al. On hedonic price of second-hand houses in Beijing based on multi-scale geographically weighted regression: scale law of Spatial heterogeneity. *Econ. Geogr.* **40**, 75–83 (2020).
58. Proshad, R. et al. Potential toxic metals (PTMs) contamination in agricultural soils and foodstuffs with associated source identification and model uncertainty. *Sci. Total Environ.* **789**, 147962 (2021).
59. Qiao, P. et al. Quantitative analysis of the factors influencing Spatial distribution of soil heavy metals based on geographical detector. *Sci. Total Environ.* **664**, 392–413 (2019).
60. Dziauddin, M. F., Powe, N. & Alvanides, S. Estimating the effects of light rail transit (LRT) system on residential property values using geographically weighted regression (GWR). *Appl. Spat. Anal. Policy*. **8**, 1–25 (2015).
61. Wilson, M. J. The importance of parent material in soil classification: a review in a historical context. *Catena* **182**, 104131 (2019).
62. Dong, B. et al. Multiple methods for the identification of heavy metal sources in cropland soils from a resource-based region. *Sci. Total Environ.* **651**, 3127–3138 (2019).
63. Mico, C. et al. Assessing heavy metal sources in agricultural soils of an European mediterranean area by multivariate analysis. *Chemosphere* **65**, 863–872 (2006).
64. Shun, C. et al. Heavy metal pollution status, children health risk assessment and source apportionment in farmland soils in a typical polluted area, Northwest China. *Stoch. Environ. Res. Risk Assess.* **38** (6), 2383–2395 (2024).
65. Lin, C. et al. Risk assessment and sources of heavy metals in farmland soils of yellow river irrigation area of Ningxia. *Huanjing Kexue*. **44** (1), 356–366 (2023).
66. ChaoYue, P. et al. Soil heavy metal contamination, sources, and health risk of typical drinking water sources in the suspended reach of the lower yellow river. *Huanjing Kexue*. **44** (12), 6710–6719 (2023).
67. Sun, S. J. et al. Source and influencing factor analysis of soil heavy metals based on PMF model and geodetector. *Huan Jing Ke Xue*. **45** (9), 5474–5484 (2024).
68. Bi, Z. et al. Machine learning-driven source identification and ecological risk prediction of heavy metal pollution in cultivated soils. *J. Hazard. Mater.* **476**, 135109 (2024).

69. Zhang, H. et al. Machine learning-based source identification and Spatial prediction of heavy metals in soil in a rapid urbanization area, Eastern China. *J. Clean. Prod.* **273**, 122858 (2022).

Author contributions

Yuru Li: Data curation, writing, review and editing. Zihua Wang: Collected and analyzed data. Leiyu Liu: Formal analysis, writing, review and editing. Yani Geng: Conceptualization, project administration, resources, supervision, validation. Jun Zhang: Conceptualization, investigation, methodology, validation, resources, writing, review and editing.

Funding

The authors acknowledge the financial support of the National Natural Science Foundation of China (NO.41771215), Key Laboratory of Groundwater Hydrology and Ecological Effects of Ministry of Education (Chang'an University) Open Fund Project (NO.300102290504), Research Project of Major Theoretical and Practical Problems in Shaanxi Social Sciences (NO.2020Z042), Project of Shaanxi Key Laboratory of Disaster Monitoring and Mechanism Simulation (NO.20JS012), Baoji University of Arts and Sciences Graduate Innovation Research Project (YJSCX24ZD08) funded.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.G. or J.Z.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025