



OPEN

Source identification and driving factor apportionment for soil potentially toxic elements via combining APCS-MLR, UNMIX, PMF and GDM

Cang Gong^{1,2}, Xiang Xia^{1✉}, Mingguo Lan³, Youchang Shi³, Haichuan Lu¹, Shunxiang Wang¹ & Ying Chen^{1✉}

The contamination and quantification of soil potentially toxic elements (PTEs) contamination sources and the determination of driving factors are the premise of soil contamination control. In our study, 788 soil samples from the National Agricultural Park in Chengdu, Sichuan Province were used to evaluate the contamination degree of soil PTEs by pollution factors and pollution load index. The source identification of soil PTEs was performed using positive matrix decomposition (PMF), edge analysis (UNMIX) and absolute principal component score-multiple line regression (APCS-MLR). The geo-detector method (GDM) was used to analysis drivers of soil PTEs pollution sources to help interpret pollution sources derived from receptor models. Result shows that soil Cu, Pb, Zn, Cr, Ni, Cd, As and Hg average content were 35.2, 32.3, 108.9, 91.9, 37.1, 0.22, 9.76 and 0.15 mg/kg in this study area. Except for As, all are higher than the corresponding soil background values in Sichuan Province. The best performance of APCS-MLR was determined by comparison, and APCS-MLR was considered as the preferred receptor model for soil PTEs source distribution in the study area. APCS-MLR results showed that 82.70% of Cu, 61.6% of Pb, 75.3% of Zn, 91.9% of Cr and 89.4% of Ni came from traffic-industrial emission sources, 60.9% of Hg came from domestic-transportation emission sources, 57.7% of Cd came from agricultural sources, and 89.5% of As came from natural sources. The GDM results showed that distance from first grade highway, population, land utilization and total potassium (TK) content were the main driving factors affecting these four sources, with *q* values of 0.064, 0.048, 0.069 and 0.058, respectively. The results can provide reference for reducing PTEs contamination in farmland soil.

Keywords Potentially toxic elements, Source identification, APCS-MLR, UNMIX, PMF, Geographic detector

After years of urbanization and industrialization, soil potentially toxic elements (PTEs) pollution caused by high-intensity and intensive human activities has become a global environmental problem¹⁻⁶. It changes the chemical and physical properties of soil, reduces the quality of soil and crops, and further endangers human health through the food chain^{1,7-9}. In China, the pollution of PTEs in soil is more prominent in some areas. According to the National soil pollution Survey¹⁰, the over-standard rates of Cd, Hg, As, Cu, Pb, Cr, Zn and Ni were 7.0%, 1.6%, 2.7%, 2.1%, 1.5%, 1.1%, 0.9% and 4.8%, respectively. Therefore, taking measures to control and repair PTEs contamination in farmland soil is the key to ensure soil environment quality and agricultural products safety.

Quantitative research on the source analysis and driving factors of soil PTEs is of great significance for effective control of soil PTEs contamination. In general, soil PTEs sources are influenced by natural factors, anthropogenic activities factors or a combination of these factors^{3,10-13}. Anthropogenic factors include mining activities, agricultural activities, industrial activities, transportation activities, etc., while natural factors include

¹Research Center of Applied Geology of China Geological Survey, Chengdu, China. ²Key Laboratory of Natural Resource Coupling Process and Effects, Beijing, China. ³Technology Innovation Center for Analysis and Detection of the Elemental Speciation and Emerging Contaminants, China Geological Survey, Kunming, China. ✉email: 360395302@qq.com; gc452@sina.com

soil parent material, elevation, hydrological conditions, etc. The combined action of these factors further leads to the complex spatial variability of PTEs in soil, which increases the difficulty of source analysis.

Quantitative source analysis of PTEs is to use the chemical and physical characteristics of contaminants in receptor model to distinguish contamination sources and quantify contribution apportionment rate of various sources. Among them, positive matrix decomposition (PMF), edge analysis (UNMIX) and absolute principal component scores–multiple line regression (APCS-MLR) were frequently used in soil PTEs source analysis^{14–16}. APCS-MLR obtains the absolute factor score and the quantitative contribution of each factor by reducing and multiple linear regression respectively¹⁵. UNMIX model automatically delete unreasonable data through the system model, does not need to set contamination sources number and uncertainty, and reduces anthropogenic factors influence¹⁶. According to the concentration of PTE in each soil sample, PMF obtained the source profile, source contribution and the corresponding uncertainty¹⁴. Due to the different algorithms used, the results of source distribution of PTEs in soil by these receptor models may be different. As these models require numerical data to satisfy the algorithm, type variables such as soil parent material and land use type cannot be calculated¹⁶. In addition, the spatial variability of data was not taken into account by these models. In fact, spatial stratification heterogeneity among variables has significant effect on the accumulation of soil PTEs.

In contrast, the geo-detector method (GDM) can well overcome the limitations of PMF, UNMIX and APCS-MLR models. GDM is an effective tool for spatial variance analysis, which can detect the nonlinear relationship among multiple explanatory variables, quantitatively determine the influence of various influencing factors on the spatial heterogeneity of soil PTEs, and solve the limitations of traditional methods in analyzing category variables^{17–19}. It has been widely used in groundwater²⁰, land use^{21,22}, ecological vulnerability^{23,24}, soil PTEs pollution^{10,25,26} and other fields. Therefore, based on measurement data, auxiliary data and spatial information, this study proposes a comprehensive method to combine receptor models APCS-MLR, UNMIX and PMF with GDM. Auxiliary data include classification variables, such as elevation, slope, aspect, land use type, soil pH, organic matter (OM), total nitrogen (TN) content, total phosphorus (TP) content, total potassium (TK) content. Spatial information includes distance from industrial land, commercial distance, railway distance, residential land distance, highway distance and river distance. This approach provides an in-depth understanding of the source analysis and driving factors identification of PTEs in soil at the town scale.

The study area is located in the heart of Chengdu Plain, Sichuan Province, which is a famous national agricultural park and Chinese chives production base. However, with the rapid expansion of urbanization, the rapid development of industrial enterprises and the widespread use of agricultural substances such as pesticides and chemical fertilizers, the soil environment of Chengdu Plain has changed to varying degrees, directly affecting the content of PTEs in agricultural products in Chengdu Plain. In this study, it was speculated that anthropogenic factors such as traffic activities, industrial activities and agricultural activities may be the main pollution sources of HMs in the soil of the study area. Combined with the spatial information and spatial analysis data of the sampling points, this hypothesis was tested by a comprehensive method, and the source analysis of PTEs in the soil was objectively explained. The purpose of this study was to (1) evaluate the level of PTEs contamination in the central soil of Chengdu Plain, Sichuan Province; (2) APCS-MLR, UNMIX and PMF models were used to assign potential contamination sources of soil PTEs; (3) GDM was used to analysis the driving factors affecting soil PTEs pollution sources. The research results revealed the potential contamination sources and corresponding driving factors of soil PTEs contamination, and provided targeted strategies for the control of soil PTEs contamination.

Materials and methods

Study area

The study area is located in the central zone of Chengdu Plain, Chengdu, Sichuan Province, about 40 km away from the urban area of Chengdu. The study area covers an area of 80 km² and a population of 7.51×10^4 . Except a small part of the area is shallow hill platform, most of the area is flat dam, and 90% of the soil is black oil-sand soil formed by impact. 3/4 of the research area is located in the water source protection area, and many major rivers such as Baitiao river and Xuyan river flow through this area. Major transportation hubs such as National highway 317, Chengguan expressway and Chengguan express railway run through the whole territory. The land use type is mainly cultivated land (about 46.5%), followed by forest land (24.1%) and residential land (13.9%), and 1.9% of industrial land is scattered in the southwest and central regions.

Sample collection and measurement

The sample sites were arranged in 1 km² grid, the density of cultivated land and forest land was 9 points km⁻², and the density of residential areas, schools, industrial land and construction land was 4 points km⁻². Sample collection will be completed in April 2021. According to the Code for Geochemical Evaluation of Land Quality (DZ/T0295-2016), 788 samples of topsoil (0–20 cm) were collected (538 samples of cultivated soil, 209 samples of forest soil and 41 samples of other land soil). The sampling locations were recorded by GPS. The sampling location is shown in Fig. 1. In order to improve sample representativeness, the "X" sampling method (1 sub-sample was collected at each of the 4 end points of "X" and the middle crossing point) was used to collect 5 sub-samples from a 20–50 m area around each sampling point and mix them into one sample. The samples were air-dried for a week, removed from debris, passed through a 10-mesh plastic sieve, and sent to the Institute of Comprehensive Utilization of Mineral Resources, Chinese Academy of Geological Sciences for analysis and testing. pH was determined by ion selective electrode method (for water extraction without carbon dioxide, the ratio of soil to water was 1/2.5), TOC by volumetric method (analysis methods for regional geochemical sample-part 27: determination of organic carbon contents by potassium dichromate volumetric method (DZ/T 0279. 27–2016)), TN by combustion infrared method, As, Hg and Se by atomic fluorescence method, Cu, Pb, Zn, Ni, Cr, Cd, TP and TK by X-ray fluorescence method and inductively coupled plasma optical /mass spectrometry. The quality of

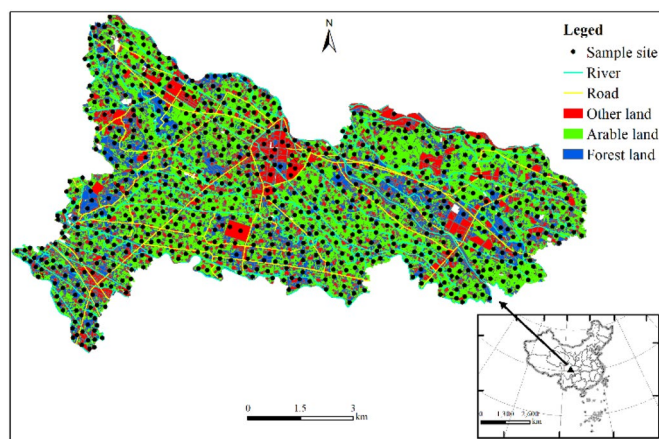


Figure 1. Distribution of sampling locations in the study area (map were generated with software ArcMap10.8 <http://www.esri.com/>).

analysis and test was controlled by means of inserting national soil standard substances (GBW07424, GBW07425, GBW07426, GBW07427 and GBW07428), repeatability inspection, anomaly inspection and blank test.

Pollution assessment

The pollution factor (PF) and pollution load index (PLI) were used to evaluate the pollution degree of soil PTEs^{10,16,21}. PF and PLI were calculated based on the Eqs. (1), (2):

$$PF = \frac{C_i}{C_b} \quad (1)$$

$$PLI = \sqrt[n]{PF_1 \times PF_2 \times \dots \times PF_n} \quad (2)$$

where C_i is the concentration of soil PTEs, C_b is the geochemical background concentration of Sichuan province. There are five classes based on PF value: ≤ 1 (uncontaminated), $1 < PF \leq 2$ (slightly contaminated), $2 < PF \leq 3$ (mild contaminated), $3 < PF \leq 5$ (moderate contaminated), > 5 (high contaminated). Four classes based on PLI value: ≤ 1 (uncontaminated), $1 < PLI \leq 2$ (moderate contaminated), $2 < PLI \leq 5$ (highly contaminated), > 5 (extremely contaminated).

Absolute principal component scores–multivariate linear regression (APCS-MLR)

The APCS-MLR model was proposed by Thurston and Spengler in 1985. Which can determine the load of PTEs to each pollution source, and calculate the average contribution of each source to soil PTEs. The detailed steps are as follows:

$$Z_{ij} = \frac{C_{ij} - \bar{C}_j}{\sigma_j} \quad (3)$$

$$(Z_0)_i = \frac{0 - \bar{C}_i}{\sigma_i} = -\frac{\bar{C}_i}{\sigma_i} \quad (4)$$

$$X_i = b_0 + \sum_{k=1}^m b_k APCS_k \quad (5)$$

where Z_{ij} is the concentration after normalization, C_{ij} is the concentrations of the i th sample of the j th PTE, \bar{C}_j is the average concentrations of j th PTE, σ_j is the standard deviation of the j th PTE, b_0 is the intercept of regression for soil PTE, b_k is the regression coefficient of the soil PTE, m is the number of factors, $APCS_k$ is the adjusted score of the k th source, $b_k APCS_k$ can be regarded as the contribution of the k th source to the soil PTE concentrations.

UNMIX model

In this model, the data space dimensionality is reduced via singular value decomposition, and then number of sources, source composition and contribution rate of sources of each sample can be estimated (USEPA, 2007). The fundamental model can be characterized as follow:

$$C_{ij} = \sum_{k=1}^m U_{ik}D_{kj} + S_{ij} \quad (6)$$

where C_{ij} is the concentrations of b th PTEs in i th sample, U_{ik} is the contribution of k th in the i th sample, D_{kj} is the concentrations of the j th PTEs from k th source, S_{ij} is the error. The source component spectrum parsed by the model needs to meet minimum system requirements that can be interpreted by the model (Min $Rsq(R^2) > 0.8$, Min Sig/Noise($S/N > 2$)).

Positive Matrix Factorization (PMF)

The PMF model decomposed the original matrix X_{ij} into two factor matrices g_{ik} and f_{kj} as well as a residual matrix e_{ij} , and it was expressed as follows (USEPA, 2014):

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

where, x_{ij} is the concentration of the j th PTE in the i th sample (mg/kg); g_{ik} is the contribution of the k th source in the i th sample; f_{kj} is the concentrations of the j th PTE from the k th source factor; and e_{ij} is the residual. The residual error matrix e_{ij} is calculated by the minimum value of the objective function Q calculated according to the Eq. (9):

$$Q = \sum_{i=1}^n \sum_{j=1}^m \left[\frac{e_{ij}}{u_{ij}} \right]^2 \quad (8)$$

where u_{ij} is the uncertainty of the j th PTEs in i th samples. The uncertainty (u) of the PTEs was calculated as follows²⁷:

When the concentration values below the detection limit:

$$x_{ij} = \frac{d_{ij}}{2}, u_{ij} = \frac{d_{ij}}{2} \quad (9)$$

When the concentration values beyond the detection limits:

$$x_{ij} = c_{ij} \quad (10)$$

$$\text{if } x_{ij} \leq 3 d_{ij}, u_{ij} = \frac{d_{ij}}{3} + 0.2 \times c_{ij} \quad (11)$$

$$\text{if } x_{ij} > 3 d_{ij}, u_{ij} = \frac{d_{ij}}{3} + 0.1 \times c_{ij} \quad (12)$$

For missing values:

$$x_{ij} = \bar{c}_{ij} u_{ij} = 4\bar{c}_{ij} \quad (13)$$

where x_{ij} is the concentration of sample species, d_{ij} is the detection limit, σ_{ij} is the x_{ij} concentration uncertainty, c_{ij} is the sample measured concentration, \bar{c}_{ij} is the measured concentrations geometric mean.

Geo-detector method (GDM)

GDM was an effective tool to analyze the spatial variance that can identify the explanatory variables affecting the dependent variable based on the assumption that explanatory variable (X) is associated with dependent variable (Y) if their spatial pattern is consistent¹⁷. The detailed calculation references¹⁷.

The dependent variables are the mean factor scores of APCS-MLR and PMF, 24 explanatory variables including TP, TK, TN, OM, pH, Se, slope (X_1), slope direction (X_2), elevation (X_3), distance from river (X_4), distance from tributary (X_5), population (X_6), land utilization (X_7), distance from super highway (X_8), distance from first grade highway (X_9), distance from secondary road (X_{10}), distance from tertiary highway (X_{11}), distance from township road (X_{12}), distance from transportation service station (X_{13}), distance from irrigation canals and ditches (X_{14}), distance from railway (X_{15}), distance from residential area (X_{16}), distance from industrial zone (X_{17}) and distance from commercial service area (X_{18}) were selected. Natural factors are represented by TP, TK, TN, OM, pH, Se, X_1 – X_5 . Anthropogenic factors are represented by X_6 – X_{18} . If the independent variable is numerical quantity, it needs to be discretized into type quantity¹⁷. The natural breakpoint method is used to divide 15 influencing factors into 10 categories. GeoDetector (<http://www.geodetector.org/>) SPSS26.0, ArcGIS10.8 and Origin2019 were used in this study. The schematic technical route of combining receptor models APCS-MLR, UNMIX and PMF with GDM was proposed in this study was showed in Fig. 2

Results and discussion

Concentrations and distribution of soil PTEs

The descriptive statistical characteristics of As, Cd, Cu, Cr, Hg, Ni, Pb and Zn concentrations in the overall soil and different land use types soils are showed in Table 1. Except for As, the concentration of total soil PTEs in

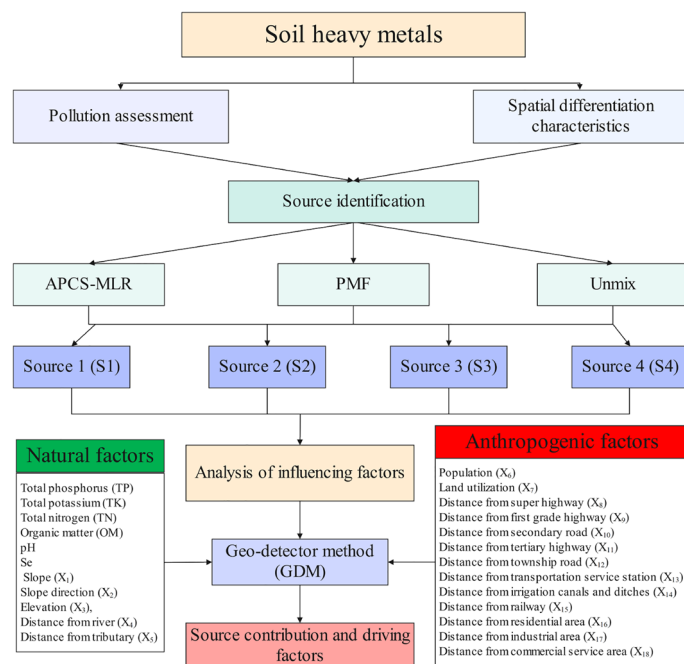


Figure 2. The schematic technical route of the comprehensive methodology.

study area was higher than the background value of Sichuan province (BSS)²⁸, and the average values of Cu, Pb, Zn, Cr, Ni, Cd and Hg in soil were 1.13, 1.05, 1.26, 1.16, 1.14, 2.78 and 2.46 of the corresponding BSS, respectively. In particular, the maximum concentrations of Cu, Zn, Cd and Hg were 19.5, 21.0, 10.5 and 14.4 times of the corresponding BSS. The average values of Cu, Pb, Zn, Cr, Ni, Cd, As and Hg in cultivated soil were 1.15, 1.04, 1.28, 1.17, 1.14, 2.78, 0.90 and 2.62 times higher than their corresponding BSS, respectively. The average values of Cu, Pb, Zn, Cr, Ni, Cd, As and Hg in forest soil were 1.09, 1.06, 1.21, 1.15, 1.13, 2.66, 0.92 and 2.46 times higher than corresponding BSS, respectively. For other land soil, mean values of Cu, Pb, Zn, Cr, Ni, Cd, As and Hg were 1.15, 1.00, 1.23, 1.19, 1.10, 3.04, 0.94 and 3.11 times higher than corresponding BSS, respectively. In this study, the average values of soil PTEs were all lower than the risk screening value of soil PTEs (GB15618-2018)¹⁰, indicating that soil PTEs did not pose a significant threat to crops and humans on the whole. However, 1.52%, 0.13%, 0.25%, 0.13% and 13.1% of soil Cu, Pb, Zn, Cr and Cd contents exceeded the risk screening values, indicating that there was a significant risk of point source pollution in the study area.

In addition, the coefficient of variation (CV) values of Cu, Hg and Zn were all more than 50%, especially the CV values of Hg in different land use types were all above 50%, indicating that there were great spatial differences of these soils PTEs²⁹.

Figure 3 shows the spatial distribution of soil PTEs in the study area. It can be seen that the high value areas of Cu, Ni and Zn were generally distributed in the central and western regions, and the low content areas mainly appear in the western regions. The high content areas of Hg and Pb were mainly concentrated in the middle of study area. The high value areas of Cd were sporadically distributed in the central and eastern regions, and the eastern regions were generally higher than the central and western regions. The high value areas of As were mainly concentrated in the western regions, and the western regions were obviously higher than the central and eastern regions. The high value area of Cr was scattered in the whole study area. It can be seen that there is a significant difference in the spatial distribution of soil PTEs content on the town scale.

Soil PTEs contamination level

In order to better understand the level of soil PTEs pollution, we calculated the PF and PLI values of PTEs in total soils and different land use patterns soils in the study area (Fig. 4). The average PF values of PTEs in both the total soils and forest soils were in descending order: Cd > Hg > Zn > Cr > Ni > Cu > Pb > As, Cd > Hg > Zn > Cr > Cu > Ni > Pb > As in cultivated soils, and Hg > Cd > Zn > Cr > Cu > Ni > Pb > As in other soils. The average PF values of Hg and Cd in total soils and different use type soils were the highest, among which Hg and Cd in total soils, cultivated soils and forest soils were at 2–3 mild pollution levels, while in other use type soils were at 3–5 moderate pollution levels, indicating that soils were polluted by Hg and Cd to a certain extent under different land use types. The average PF values of Zn, Cr, Ni and Cu and Pb in the total soils and different utilization type soils were 1–2, which belong to slightly contaminated. The average PF value of As in all soils were less than 1, which belongs to the uncontaminated. The PLI value is calculated from the PFs of a single PTE to evaluate the pollution status of multiple PTEs. PLI values showed that total soils (1.34), cultivated soils (1.34), forest soils (1.32) and other soils (1.36) were moderately polluted. Overall, soils were contaminated with PTEs to varying degrees regardless of land use type. Overall, PF and PLI revealed that human activities influence the accumulation

		Cu	Pb	Zn	Cr	Ni	Cd	As	Hg
Total soils (n = 788)	Minimum (mg/kg)	18.5	19.8	55.0	61.7	23.6	0.082	4.17	0.022
	Maximum (mg/kg)	607	90.3	1820	264	56.6	0.83	18.0	0.88
	Median (mg/kg)	34.4	32.3	106	92.3	37.3	0.21	9.43	0.13
	Average mean (mg/kg)	35.2	32.3	108.9	91.9	37.1	0.22	9.76	0.15
	Standard deviation	21.3	5.01	63.5	10.1	3.75	0.07	2.09	0.09
	CV (%)	60.5	15.5	58.3	11.0	10.1	31.1	21.4	58.4
	Geometric mean (mg/kg)	34.2	31.9	105.9	91.5	36.9	0.21	9.55	0.14
Cultivated soils (n = 538)	Minimum (mg/kg)	18.5	19.8	55.0	61.7	26.7	0.082	4.17	0.030
	Maximum (mg/kg)	607	60.2	1820	133	56.6	0.83	18.0	0.71
	Median (mg/kg)	34.5	32.4	107	92.3	37.5	0.21	9.32	0.13
	Average mean (mg/kg)	35.7	32.2	110.6	92.1	37.3	0.22	9.67	0.16
	Standard deviation	25.2	4.55	76.1	7.94	3.67	0.07	2.02	0.08
	CV (%)	70.7	14.1	68.8	8.63	9.86	31.3	20.9	54.4
	Geometric mean (mg/kg)	34.4	31.9	107	91.7	37.1	0.21	9.47	0.14
Forest soils (n = 209)	Minimum (mg/kg)	19.0	20.4	59.8	64.0	23.6	0.092	4.93	0.038
	Maximum (mg/kg)	102	90.3	170	108	49.2	0.44	17.8	0.88
	Median (mg/kg)	34.4	32.4	106	92.5	36.9	0.21	9.52	0.13
	Average mean (mg/kg)	33.9	32.6	105	91.2	36.9	0.21	9.91	0.15
	Standard deviation	6.78	6.01	14.6	8.13	3.80	0.06	2.29	0.090
	CV (%)	20.0	18.4	13.9	8.92	10.3	28.6	23.1	61.2
	Geometric mean (mg/kg)	33.4	32.2	104	90.8	36.7	0.21	9.67	0.13
Other land soils (n = 41)	Minimum (mg/kg)	20.0	22.7	71.6	75.2	28.2	0.11	7.09	0.022
	Maximum (mg/kg)	96.6	46.0	152	264	45.4	0.47	14.8	0.60
	Median (mg/kg)	34.3	30.5	105	89.5	36.0	0.23	10.2	0.12
	Average mean (mg/kg)	35.8	30.9	106	93.7	35.9	0.24	10.2	0.19
	Standard deviation	11.0	4.8	16.6	27.9	4.21	0.09	1.80	0.14
	CV (%)	30.7	15.6	15.7	29.8	11.7	36.6	17.7	77.3
	Geometric mean (mg/kg)	34.7	30.6	104	91.6	35.7	0.23	10.0	0.14
BSS		31.1	30.9	86.5	79	32.6	0.079	10.8	0.061
RSV		50	90	200	150	70	0.3	40	1.8

Table 1. Descriptive statistics of soil PTEs in the study area. CV Coefficient of variation, BSC Background value of soil in Sichuan province²⁸, RSV Risk screening values (GB15618-2018)¹⁰.

of soil PTEs. Human activities strongly affected Cd and Hg in the soil of the study area, followed by Zn, Cr, Ni, Cu and Pb, and had little effect on As. To understand the contribution of anthropogenic and natural activities to soil PTEs accumulation, a receptor model should be used to further quantify source resolution.

Source apportionment of PTEs in soils

Evaluation of source apportionment of different receptor models

Figure 5a–c, exhibits the composition of factors as analyzed by the APCS-MLR, UNMIX and PMF receptor models. The accuracies of the three models were evaluated by using the determination coefficient (R^2) and their respective diagnostic parameters. The R^2 values of each soil PTEs ranged from 0.72 to 0.98 in the APCS-MLR model, 0.80–1.00 in the UNMIX model, and 0.53–1.00 in the PMF model. As far as R^2 was concerned, the Unmix model has the highest reliability, followed by APCS-MLR and PMF. In terms of their respective diagnostic parameters, the KMO and Bartlett test results of the APCS-MLRU model show that the KMO value was 0.771, meeting the requirement of >0.6 , and the P value was 0.000, meeting the requirement of $P < 0.001$ ³⁰, which was suitable for APCS-MLR analysis (Fig. 4a). Previous studies have shown that APCS-MLR model source apportionment has good repeatability and produces relatively accurate results^{31–33}. For the PMF model, the S/N of 8 kinds of soil PTEs were all greater than 7, indicating that the accuracy of the apportionment results was high (Fig. 4c), however, the R^2 of Cr, Pb and Cu were 0.53, 0.59 and 0.63 respectively, all less than 0.7, which indicated that the fitness of these PTEs was not ideal. Previous studies have shown that the PMF model was extremely sensitive to outliers, and reasonable results can't be obtained without eliminating outliers¹⁵. For the UNMIX model, the R^2 of each soil PTEs was very high, and Min R^2 was 0.94 and Min S/N was 2.10, which meets the requirement Min $R^2 > 0.8$ and Min S/N > 2.0 , respectively. UNMIX model can compensate for the loss of some elements contribution and is an effective tool for soil PTEs source analysis apportionment^{16,34,35}. Although the Unmix model identified three sources for all PTEs in this study, the proportion of each PTEs in the three sources was very close (Fig. 4b), which failed to effectively separate and identify the sources of each PTEs. This may be due to the relatively small change of soil PTEs in the small-scale areas of this study.

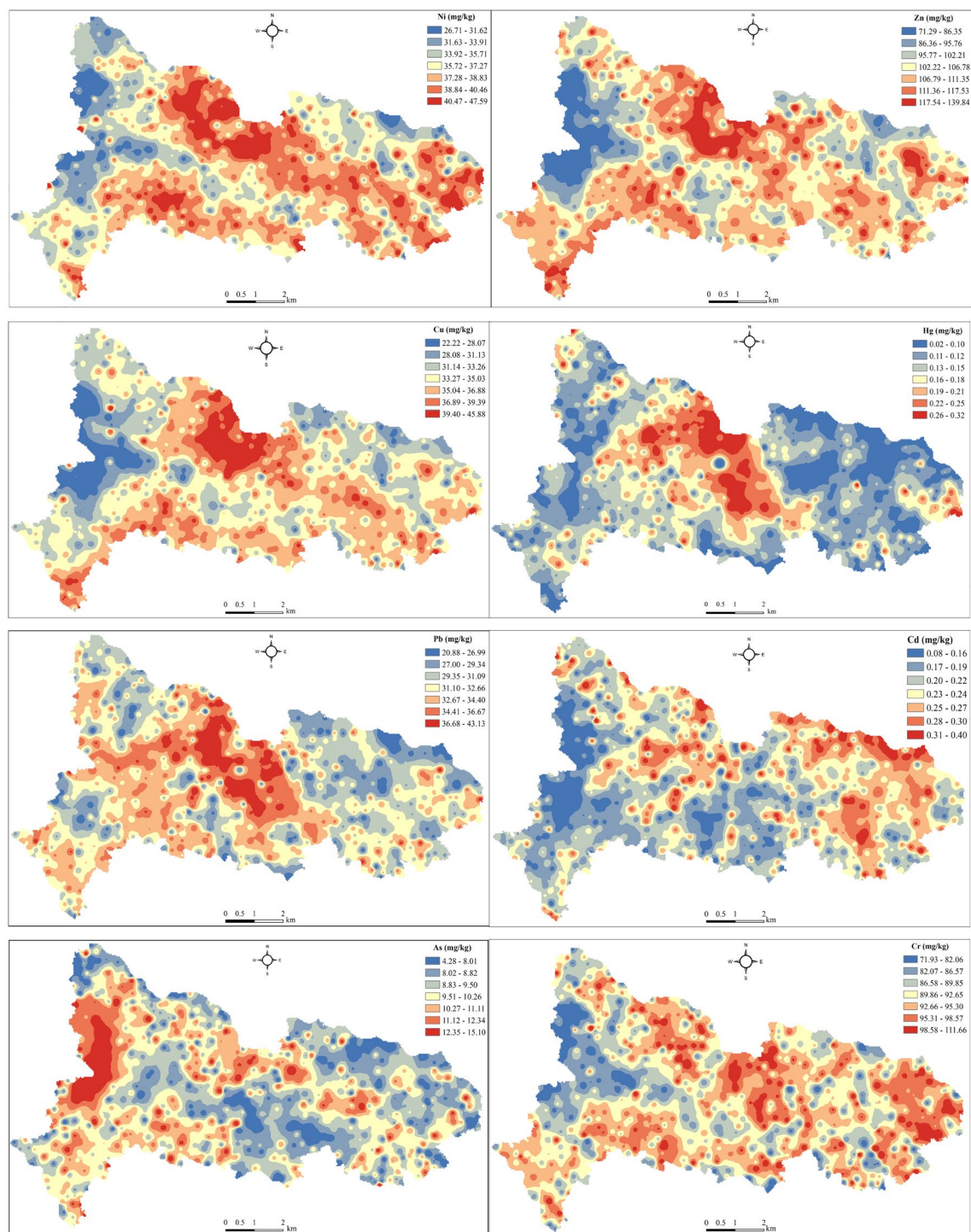


Figure 3. Spatial distribution of the soil PTEs (map were generated with software ArcMap10.8 <http://www.esri.com/>).

Pearson correlation analysis was further used to explain the relationship between the factors obtained from the three models. As shown in Fig. 6, the same factors produced by the APCS-MLR model and the model PMF have a strong correlation. For example, the correlation coefficients of source 1 (S1), source 2 (S2), source 3 (S3) and source 4 (S4) in APCS-MLR and PMF were all greater than 0.80, reaching a significant level of 0.05. However, most of the factors produced by UNMIX have no correlation except the individual factors of APCS-MLR and PMF. In general, APCS-MLR model was considered to be the preferred model for soil PTEs quantitative source apportionment in this study area.

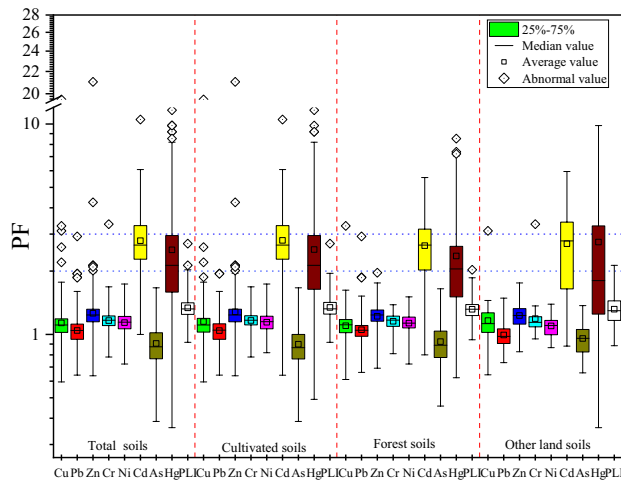


Figure 4. Pollution index of PTEs under total soil and different land utilization soil.

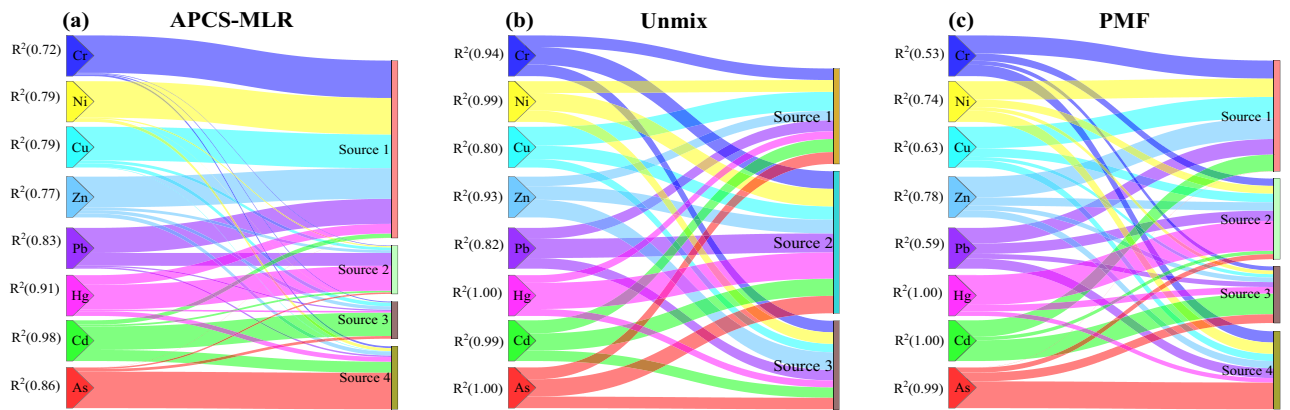


Figure 5. Contribution of each factor for PTEs in different receptor models.

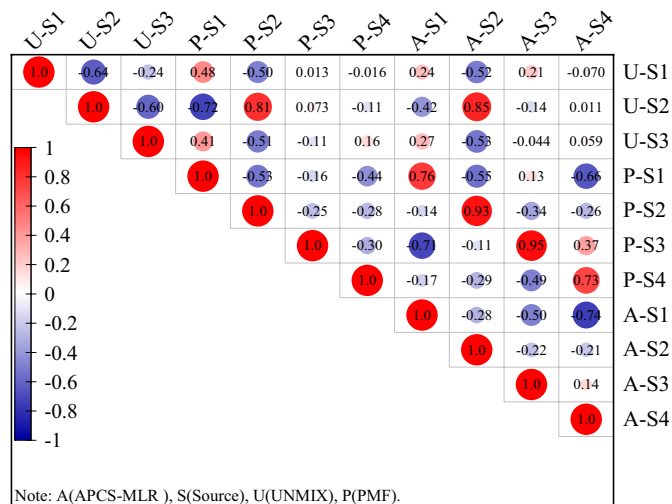


Figure 6. Pearson correlation showing contributions of factors derived from APCS-MLR, UNMIX and PMF.

Comparison of source apportionment results for different receptor models

In source apportionment, high load soil PTEs with certain factors can be served as typical PTEs. As shown in (Fig. 5a,c), S1 of APCS-MLR and PMF apportionment were mainly dominated by Cu, Pb, Zn, Cr and Ni, with contribution rates of 31.7–82.7%, 38.4–61.6%, 50.8–75.3%, 44.4–91.9% and 44.9–89.4%, respectively. S2 was mainly related to Hg, and the contribution rate of two models were 60.9–73.1%. In addition, Pb show moderate load with S2 in the two models, Cr and Ni show moderate load with S2 in PMF models. S3 of the two models were mainly related Cd, and the contribution rates was 50.9–57.7%. S4 has a significant correlation with As in two models, and the contribution rate was between 65.3 and 9.5%. Meanwhile, Cd shows moderate load with S4 in the APCS-MLR model, Pb, Cr and Ni show moderate load with S4 in the PMF model. As shown in Fig. 5b, S1 of UNMIX apportionment was mainly dominated by Cu. S2 was mainly related to Pb, Cr, Ni, Cd, As and Hg. S3 was mainly related to Zn. It is obvious that the results of UNMIX producing were quite different from those of APCS-MLR and PMF.

Driving factors affecting pollution sources by GDM

Quantify the influence intensity of each influence factor on soil PTEs source apportionment factors by using the q value obtained from GDM. The maximum q value indicates that the influencing factor was the dominant explanatory variable of soil PTEs source apportionment factors^{18,36,37}. As shown in Fig. 7, S1 was dominated by X_9 ($q=0.064$, $P<0.01$), followed by X_8 ($q=0.039$, $P<0.01$), X_{11} ($q=0.064$, $P<0.01$) and X_{17} ($q=0.055$, $P<0.01$), indicating that S1 has a significant spatial correlation with X_8 , X_9 , X_{11} and X_{17} , indicating that S1 represented the anthropogenic source of soil pollution. X_6 , X_8 , X_9 , X_{11} and X_{15} had significant spatial correlation with S2, with q values of 0.048 ($P<0.01$), 0.035 ($P<0.05$), 0.040 ($P<0.05$), 0.041 ($P<0.01$) and 0.048 ($P<0.01$), respectively. X_6 and X_7 have very significant spatial correlation with S3 ($P<0.01$), and q values were 0.058 and 0.069, respectively. X_8 , X_9 , X_{11} and X_{15} have significant spatial correlation with S3 ($P<0.05$), which means that S3 represents the anthropogenic source of soil pollution. TK ($q=0.058$, $P<0.01$), OM ($q=0.058$, $P<0.01$) and X_3 ($q=0.046$, $P<0.01$) had very significant spatial correlation with S4, indicating that S4 represented the natural source of soil pollution.

Interpretation of sources obtained by multiple methods

S1 interpretation

According to the multi-receptor model (Fig. 4), S1 contributes the most to Cu (82.7%), Pb (61.6%), Zn (75.3%), Cr (91.9%) and Ni (89.4%) in soil, followed by Hg (23.6%), Cd (10.6%) and As (0.3%). The results showed that S1 had the greatest contribution to Cu, Pb, Zn, Cr and Ni in soil, followed by Hg. In general, Cu, Pb and Zn may come from traffic emissions caused by motor vehicle engine, brake pads, galvanized parts and tires wear^{5,38}. Previous studies have shown that Pb mainly comes from traffic emissions, such as leaded gasoline emissions, engine brake wear and catalytic combustion³⁹, as well as Zn pollution caused by tire wear or galvanized parts corrosion⁴⁰. It was believed that copper oxide catalysts were often used in engine combustion systems to control carbon monoxide emissions from automobile exhaust⁴¹. In addition, industrial activities, including hardware industry, metallurgical industry and electronics industry, will further increase the accumulation of soil Cd, Zn, Cu, Cr and Ni⁴². It was found that Cr in soil is related to industrial production activities such as wear of metal parts and chromium-plated parts, iron and steel production and metallurgical manufacturing⁴³. Cr metal particles were produced in the operation and production process of iron and steel plants, hardware market, logistics base and automobile and parts industry, especially Cr elements will be released in different degrees in all aspects of cement production, which will lead to the enrichment of Cr in the surrounding soil⁴⁴.

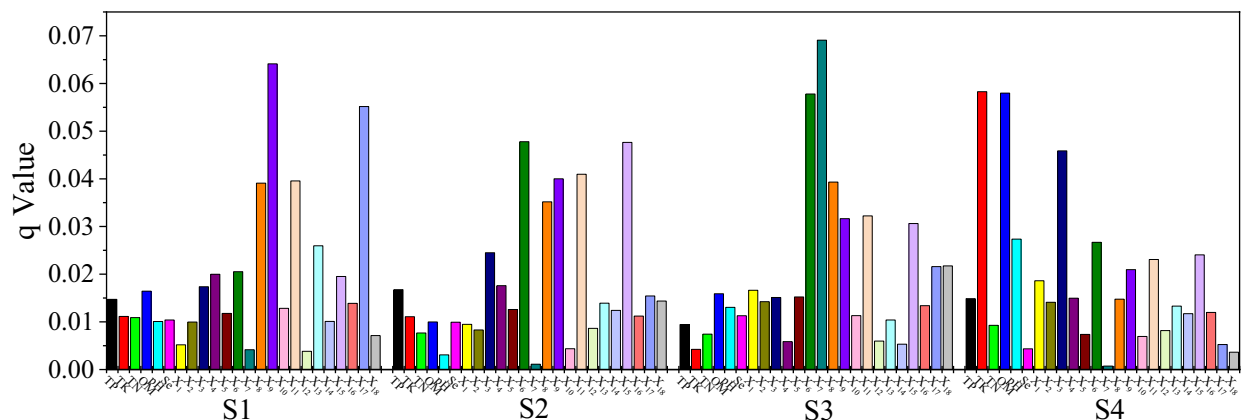


Figure 7. GDM results of 24 influence factors on 4 apportionment sources. X_1 : Slope; X_2 : Slope; X_3 : Elevation; X_4 : Distance from river; X_5 : Distance from tributary; X_6 : Population; X_7 : Land utilization; X_8 : Distance from super highway; X_9 : Distance from first grade highway; X_{10} : Distance from secondary road; X_{11} : Distance from tertiary highway; X_{12} : Distance from township road; X_{13} : Distance from transportation service station; X_{14} : Distance from irrigation canals and ditches; X_{15} : Distance from railway; X_{16} : Distance from residential area; X_{17} : Distance from industrial zone; X_{18} : Distance from commercial service area.

Based on the results of GDM, the spatial correlation between S1 with X_8 , X_9 , X_{11} and X_{17} were the highest. There are a large number of roads distributed in the study area, and the traffic and traffic flow are very busy. With the rapid increase of the number of motor vehicles in Chengdu, the traffic flow, road dust density and tail gas emission also increase sharply. potentially toxic element particles generated by road traffic activities are eventually enriched in the surrounding soil through atmospheric sedimentation and air dust adsorption. In addition, there were factories and enterprises in the study area, such as building materials decoration, plastic production, cement production, concrete manufacturing, red brick firing and printing, etc. The potentially toxic element particles and raw materials that may be scattered in these shops and workshops, as well as the PTEs carried by the “Three wastes” produced by industrial activities, were enriched into the soil by atmospheric sedimentation, rain water erosion and infiltration, causing the accumulation of soil PTEs. Overall, S1 represented traffic-industrial source.

S2 interpretation

S2 was dominantly explained by Hg (60.9%), followed by Pb (31.5%), whereas low loading were observed for Cu (7.1%), Zn (6.9%), Cr (1.4%), Ni (3.0%), Cd (4.5%) and As (3.6%). In this study, the high CV value (Table 1) and PF value (Fig. 3) of Hg indicate that human activities strongly influence the accumulation of soil Hg. Relevant research shows that Hg was related to coal combustion^{45,46}, waste gas emission^{47,48} and mineral mining, and these industrial activities will increase the emission of Hg into the atmosphere. In addition, some studies show that traffic factors have a great effect on the accumulation of Hg⁴⁹, and the exhaust gas of gas companies will also emit Hg into the atmosphere⁵⁰. Hg in soil was mostly from the dry and wet deposition of the atmosphere⁵¹. It has been reported that, railway transportation as a serious source of organic and inorganic pollution^{52,53}.

GDM shows that X_6 and X_{15} were the highest driver of S2 (Fig. 7), followed by X_8 , X_9 , X_{11} . It means that human domestic emissions and traffic activities have the greatest impact on S2. PTEs produced by human production, domestic and transportation activities can enter the soil through atmospheric deposition, and then transfer to other places through runoff. In addition, as shown in Fig. 3, the hot spots of Hg's spatial distribution appear in the central area with heavy traffic and dense population. Therefore, S2 can represent domestic-transport emissions source.

S3 interpretation

The content of Cd (57.7%) was the highest in S3, and the content of other PTEs were less than 10% in S3. Under land utilization, the PF value of soil Cd in the study area was the highest (Fig. 4), reached the level of mild and moderate pollution. Previous studies have shown that Cd has been introduced into soil with the agricultural application of phosphate fertilizers, organic fertilizers and pesticides⁵⁴. In addition, a large amount of Cd was discharged into the environmental medium through waste water, waste gas and waste residue, and then into the soil⁵⁵.

The results of GMD showed that there was a significant spatial correlation between S3 with X_6 and X_7 , indicating that agriculture activity was the main driving factor of S3. It has been reported that phosphorus fertilizer was the main source of Cd in French farmland soil, accounting for more than 1/2 of the input of Cd in French agricultural soil⁵⁶. The input flux of Cd through fertilization in Songnen plain soil in China ranges from 0.001 to 0.255 mg/m²/a⁵⁷, and the input flux of Cd through insecticides (oxazepine, chlorpyrifos and carbendazam) in Zhangshi Irrigation Area of Shenyang was 2.0×10^{-6} mg/m²/a⁵⁸. According to our field investigation, in order to increase yield, people increased the use of pesticides, chemical fertilizers and organic fertilizers, thus bringing Cd into the soil and increasing the accumulation of soil Cd. Therefore, S3 can represent the agricultural source.

S4 interpretation

The explanation of S4 was mainly As (89.5%). In this study, the average value of As in the soil of the whole region was 0.90 times of its corresponding background value, indicating that the impact of human activities was limited (Table 1). In addition, the mean PF value of As showed relatively low pollution (Fig. 4), which confirmed that the concentration of soil As in this study area was dependent on natural sources. Previous studies have reported that As in soil may be an indicator of soil parent material⁵⁹. However, anthropogenic input of As includes agricultural activities, such as the application of organic and chemical fertilizers and sewage irrigation⁶⁰, and industrial emissions^{61,62}.

Combined with GMD results (Fig. 6), S4 had the highest spatial correlation with TK ($q=0.058$, $P<0.01$), OM ($q=0.058$, $P<0.01$) and X_3 ($q=0.046$, $P<0.05$), and TK, OM and X_3 represents the natural influencing factors. Therefore, S4 can represent the natural source.

Significance of multi-source interpretation

On the one hand, since soil PTEs comes from both natural sources and human activities such as agriculture, transportation, and industry, PF, PLI, geoaccumulation index and other methods can only quantify the pollution level of soil PTEs, but can't quantify the contribution of soil PTEs. On the other hand, the source of soil PTEs in a given region is not fixed and will change over time, and the emission intensity of different sources will also be different. In this case, the receptor model is very useful for analysis the source of soil PTEs, especially in the absence of detailed information about the source of emissions. However, use receptor models to identify source of soil PTEs must be interpreted based on previous research experience. However, the combination of receptor models (APCS-MLR, PMF, UNMIX) and GMD can solve this puzzle and explain the origin of soil PTEs more effectively.

In order to further explain the source factors from the receptor model, this study considers the influencing factors such as soil TP, TK, TN, OM, pH, elevation, slope, and distance from the sampling point to the river as natural pollution sources; the influencing factors such as the distance from the sampling point to industrial

land and road as industrial pollution sources and traffic pollution sources; the distance from the sampling point to the residential land and the number of population were considered as domestic pollution sources; land use was considered as agricultural pollution sources. GMD was used to identify the main influencing factors of soil PTEs source factors. GMD results can effectively define the source factors obtained by the receptor model, which specific pollution sources were natural, industrial pollution, traffic, domestic, agricultural and other pollution sources, or were affected by multiple pollution sources together. Overall, this integrated approach provides a deep understanding of the multiple sources of soil PTEs pollution and a better definition of pollution factors.

Conclusions

With the rapid development of industry and agriculture, PTEs pollution in farmland soil has become increasingly serious. In this study, receptor model and GDM were combined to quantitatively analyze the driving factors of soil PTEs in the central zone of Chengdu Plain, Sichuan Province. Except As, the PTEs content in soil was higher than the corresponding soil background value in Sichuan Province. APCS-MLR model can fit the observed and predicted PTEs values of most soils in the study area well, and was an ideal receptor model. The APCS-MLR model was used to identify four potential pollution sources, including traffic-industrial pollution source S1 (Cu (82.7%), Pb (61.6%), Zn (75.3%), Cr (91.9%) and Ni (89.4%)), domestic-transportation pollution source S2 (Hg (60.9%)), agricultural sources S3 (Cd (57.7%)) and natural sources S4 (As (89.5%)). GDM results showed that Distance from first grade highway, population, land utilization, and TK content contributed the most to S1, S2, S3, and S4, which helped explain the source factors derived from the receptor model.

In fact, soil potentially toxic element accumulation was a complex process, which was affected by many environmental factors. In the next research, more environmental factors affecting soil PTEs accumulation should be considered, such as precipitation, Gross Domestic Product (GDP), fertilizer amount, soil parent material, etc., and the spatial correlation between the influencing factors and the source factors should be deeply analyzed. At the same time, a comprehensive study and analysis of inventory data such as traffic emissions, fertilizer consumption, industrial emissions, and atmospheric subsidence were carried out to verify the results of the combination of receptor model and GDM.

Data availability

The authors declare that all data supporting the findings of this study are available within the article.

Received: 28 January 2024; Accepted: 2 April 2024

Published online: 13 May 2024

References

- Liu, H. W. *et al.* Quantitative source apportionment, risk assessment and distribution of heavy metals in agricultural soils from southern Shandong Peninsula of China. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2020.144879> (2021).
- Tepanosyan, G., Sahakyan, L., Belyaeva, O., Asmaryan, S. & Saghatelian, A. Continuous impact of mining activities on soil heavy metals levels and human health. *Sci. Total Environ.* **639**, 900–909 (2018).
- Ding, Q., Cheng, G., Wang, Y. & Zhuang, D. Effects of natural factors on the spatial distribution of heavy metals in soils surrounding mining regions. *Sci. Total Environ.* **578**, 577–585 (2017).
- Wei, B. G. & Yang, L. S. A review of heavy metal contaminations in urban soils, urban road dusts and agricultural soils from China. *Microchem. J.* **94**, 99–107. <https://doi.org/10.1016/j.microc.2009.09.014> (2010).
- Yu, T. *et al.* Research progress in current status of soil heavy metal pollution and analysis technology. *Geol. China* **48**, 460–476 (2021).
- Chen, D. L. *et al.* Accumulation and source apportionment of soil heavy metals in molybdenum-lead-zinc polymetallic ore concentration area of Luanchuan. *Rock Miner. Anal.* **42**, 839–851. <https://doi.org/10.15898/j.ykcs.202208090147> (2023).
- Yang, A., Wang, Y., Hu, J., Liu, X. & Li, J. Evaluation and source of heavy metal pollution in surface soil of Qinghai-Tibet plateau. *Environ. Sci.* **41**, 886–894. <https://doi.org/10.13227/j.hjcx.201907195> (2020).
- Wong, C. S. C., Li, X. D. & Thornton, I. Urban environmental geochemistry of trace metals. *Environ. Pollut.* **142**, 1–16. <https://doi.org/10.1016/j.envpol.2005.09.004> (2006).
- Qin, Y. L., Zhang, F. G., Xue, S. D., Ma, T. & Yu, L. S. Heavy metal pollution and source contributions in agricultural soils developed from Karst landform in the southwestern region of China. *Toxics* <https://doi.org/10.3390/toxics10100568> (2022).
- Gong, C. *et al.* Ecological and human health risk assessment of heavy metal(loid)s in agricultural soil in hotbed chives hometown of Tangchang. *Southwest China Sci. Rep.* <https://doi.org/10.1038/s41598-022-11397-0> (2022).
- Lazar, G. C. B., Statescu, F. & Toma, D. Study of heavy metal dynamics in soil. *Environ. Eng. Manag. J.* **19**, 359–367 (2020).
- Ren, Y. *et al.* Research progress on distribution, harm and control technology of heavy metals in soil. *Geol. China* <https://doi.org/10.12029/20230320001> (2023).
- Li, W. M. *et al.* Evaluation and source of heavy metal pollution in surface soils in typical alpine agricultural areas of Qinghai province. *Rock Miner. Anal.* **42**, 598–615. <https://doi.org/10.15898/j.ykcs.202209170174> (2023).
- Wei, R. F., Meng, Z. R., Zerizghi, T., Luo, J. & Guo, Q. J. A comprehensive method of source apportionment and ecological risk assessment of soil heavy metals: A case study in Qi Qingyuan city, China. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2023.163555> (2023).
- Ma, J. W. *et al.* Source apportionment and model applicability of heavy metal pollution in farmland soil based on three receptor models. *Toxics* <https://doi.org/10.3390/toxics11030265> (2023).
- Guo, G. H., Li, K., Zhang, D. G. & Lei, M. 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.* <https://doi.org/10.1016/j.scitotenv.2022.154721> (2022).
- Wang, J. & Xu, C. Geodetector: principle and prospective. *Acta. Geographica. Sinica.* **72**, 116–134 (2017).
- Wang, J.-F. *et al.* Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. *Int. J. Geogr. Inf. Sci.* **24**(1), 107–127 (2010).
- Cao, F., Ge, Y. & Wang, J.-F. Optimal discretization for geographical detectors-based risk assessment. *GIScience Remote Sens.* **50**(1), 78–92 (2013).
- Zeng, J. Q. *et al.* A practical method for identifying key factors in the distribution and formation of heavy metal pollution at a smelting site. *J. Environ. Sci.* **127**, 552–563. <https://doi.org/10.1016/j.jes.2022.06.026> (2023).

21. Zhao, Y. H. *et al.* Quantitative analysis of factors influencing spatial distribution of soil erosion based on geo-detector model under diverse geomorphological types. *Land* <https://doi.org/10.3390/land10060604> (2021).
22. Wang, H., Wan, Q. N., Huang, W. & Niu, J. Q. Spatial heterogeneity characteristics and driving mechanism of land use change in Henan province, China. *Geocarto Int.* <https://doi.org/10.1080/10106049.2023.2271442> (2023).
23. Ruan, W. Q., Li, Y. Q., Zhang, S. N. & Liu, C. H. Evaluation and drive mechanism of tourism ecological security based on the DPSIR-DEA model. *Tour. Manag.* **75**, 609–625. <https://doi.org/10.1016/j.tourman.2019.06.021> (2019).
24. Gu, H. L., Huan, C. Y. & Yang, F. J. Spatiotemporal dynamics of ecological vulnerability and its influencing factors in shenyang city of China: Based on SRP model. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph20021525> (2023).
25. Howard, J. L., Orlicki, K. M. & LeTarte, S. M. Evaluation of some proximal sensing methods for mapping soils in urbanized terrain, Detroit, Michigan, USA. *Catena* **143**, 145–158. <https://doi.org/10.1016/j.catena.2016.03.011> (2016).
26. Guo, G. H., Li, K. & Lei, M. Accumulation, environmental risk characteristics and associated driving mechanisms of potential toxicity elements in roadside soils across China. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2022.155342> (2022).
27. Zhang, J. Z. *et al.* Trace elements in PM_{2.5} in Shandong Province: Source identification and health risk assessment. *Sci. Total Environ.* **621**, 558–577. <https://doi.org/10.1016/j.scitotenv.2017.11.292> (2018).
28. *Element Background Values of China Soils*. China National Environmental Monitoring Center edn, (China Environment Publishing Group, 1990).
29. Manta, D. S., Angelone, M., Bellanca, A., Neri, R. & Sprovieri, M. Heavy metals in urban soils: a case study from the city of Palermo (Sicily). *Italy. Sci. Total Environ.* **300**, 229–243. [https://doi.org/10.1016/s0048-9697\(02\)00273-5](https://doi.org/10.1016/s0048-9697(02)00273-5) (2002).
30. Jung, J. *et al.* Environmental forensic approach towards unraveling contamination sources with receptor models: A case study in Nakdong River South Korea. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2023.164554> (2023).
31. Zhang, M. *et al.* Quantitative source identification and apportionment of heavy metals under two different land use types: Comparison of two receptor models APCS-MLR and PMF. *Environ. Sci. Pollut. Res.* **27**, 42996–43010. <https://doi.org/10.1007/s11356-020-10234-z> (2020).
32. Qu, M. K., Wang, Y., Huang, B. & Zhao, Y. C. Source apportionment of soil heavy metals using robust absolute principal component scores-robust geographically weighted regression (RAPCS-RGWR) receptor model. *Sci. Total Environ.* **626**, 203–210. <https://doi.org/10.1016/j.scitotenv.2018.01.070> (2018).
33. Lv, J. S. Multivariate receptor models and robust geostatistics to estimate source apportionment of heavy metals in soils. *Environ. Pollut.* **244**, 72–83. <https://doi.org/10.1016/j.envpol.2018.09.147> (2019).
34. Liao, S. Y. *et al.* The quantitative source apportionment of heavy metals in peri-urban agricultural soils with UNMIX and input fluxes analysis. *Environ. Technol. Innov.* <https://doi.org/10.1016/j.eti.2020.101232> (2021).
35. Jin, Z. & Lv, J. S. Integrated receptor models and multivariate geostatistical simulation for source apportionment of potentially toxic elements in soils. *Catena* <https://doi.org/10.1016/j.catena.2020.104638> (2020).
36. Yang, Y., Yang, X., He, M. & Christakos, G. Beyond mere pollution source identification: Determination of land covers emitting soil heavy metals by combining PCA/APCS. *GeoDetector and GIS anal. Catena* **185**, 104297 (2020).
37. Qiao, P., Yang, S., Lei, M., Chen, T. & Dong, N. Quantitative analysis of the factors influencing spatial distribution of soil heavy metals based on geographical detector. *Sci. Total Environ.* **664**, 392–413 (2019).
38. Luo, F., Ba, J. J., Su, C. T., Pan, X. D. & Yang, Y. Contaminant assessment and sources analysis of heavy metals in soils from the upper reaches of the Wushui river. *Rock Miner. Anal.* **38**, 195–203. <https://doi.org/10.15898/j.cnki.11-2131/td.201806040069> (2019).
39. Cai, L. M., Wang, Q. S., Wen, H. H., Luo, J. & Wang, S. Heavy metals in agricultural soils from a typical township in Guangdong Province, China: Occurrences and spatial distribution. *Ecotox. Environ. Safe.* **168**, 184–191. <https://doi.org/10.1016/j.ecoenv.2018.10.092> (2019).
40. Councell, T. B., Duckenfield, K. U., Landa, E. R. & Callender, E. Tire-wear particles as a source of zinc to the environment. *Environ. Sci. Technol.* **38**, 4206–4214. <https://doi.org/10.1021/es034631f> (2004).
41. Dey, S. & Dhal, G. C. Controlling carbon monoxide emissions from automobile vehicle exhaust using copper oxide catalysts in a catalytic converter. *Mater. Today Chem.* <https://doi.org/10.1016/j.mtchem.2020.100282> (2020).
42. Wang, H. Z., Cai, L. M., Wang, Q. S., Hu, G. C. & Chen, L. G. A comprehensive exploration of risk assessment and source quantification of potentially toxic elements in road dust: A case study from a large Cu smelter in central China. *Catena* <https://doi.org/10.1016/j.catena.2020.104930> (2021).
43. Boudia, H., Vassalo, L., Hadjel, M., Prudent, P. & Boudenne, J. L. Spatial contamination and health risks of heavy metal(loid)s in surface soils from a petrochemical complex in the north-eastern region of Algeria. *Int. J. Environ. Sci. Technol.* **16**, 4707–4718. <https://doi.org/10.1007/s13762-018-02195-1> (2019).
44. Li, H. Y. *et al.* Pollution characteristics, key toxic components and sources of PM_{2.5}-bound heavy metals in coking polluted area of Jiexiu. *Shanxi. China Environ. Sci.* **43**, 1528–1538. <https://doi.org/10.19674/j.cnki.issn1000-6923.20221123.009> (2023).
45. Altaf, A. R., Adewuyi, Y. G., Teng, H. P., Gang, L. & Abid, F. Elemental mercury (Hg⁰) removal from coal syngas using magnetic tea-biochar: Experimental and theoretical insights. *J. Environ. Sci.* **122**, 150–161. <https://doi.org/10.1016/j.jes.2021.09.033> (2022).
46. Huang, J. L. *et al.* Health risk assessment of heavy metal(loid)s in park soils of the largest megacity in China by using Monte Carlo simulation coupled with Positive matrix factorization model. *J. Hazard. Mater.* <https://doi.org/10.1016/j.jhazmat.2021.125629> (2021).
47. Wan, Z. X. *et al.* Removal of mercury from flue gas using coal gasification slag. *Fuel Process. Technol.* <https://doi.org/10.1016/j.fuproc.2022.107258> (2022).
48. Wu, J., Li, J., Teng, Y. G., Chen, H. Y. & Wang, Y. Y. A partition computing-based positive matrix factorization (PC-PMF) approach for the source apportionment of agricultural soil heavy metal contents and associated health risks. *J. Hazard. Mater.* <https://doi.org/10.1016/j.jhazmat.2019.121766> (2020).
49. Mokhtarzadeh, Z., Keshavarzi, B., Moore, F., Marsan, F. A. & Padoan, E. Potentially toxic elements in the Middle East oldest oil refinery zone soils: source apportionment, speciation, bioaccessibility and human health risk assessment. *Environ. Sci. Pollut. Res.* **27**, 40573–40591. <https://doi.org/10.1007/s11356-020-09895-7> (2020).
50. Su, C. H. *et al.* Heavy metals in soils from intense industrial areas in south China: Spatial distribution, source apportionment, and risk assessment. *Front. Environ. Sci.* <https://doi.org/10.3389/fenvs.2022.820536> (2022).
51. Liu, R. H., Wang, Q. C., Lu, X. G., Fang, F. M. & Wang, Y. Distribution and speciation of mercury in the peat bog of Xiaoxing'an Mountain, northeastern China. *Environ. Pollut.* **124**, 39–46. [https://doi.org/10.1016/s0269-7491\(02\)00432-3](https://doi.org/10.1016/s0269-7491(02)00432-3) (2003).
52. Wilkomirski, B., Sudnik-Wójcikowska, B., Galera, H., Wierzbicka, M. & Malawska, M. Railway transportation as a serious source of organic and inorganic pollution. *Water Air Soil Pollut.* **218**, 333–345. <https://doi.org/10.1007/s11270-010-0645-0> (2011).
53. Malawska, M. & Wilkomirski, B. An analysis of soil and plant (*Taraxacum officinale*) contamination with heavy metals and polycyclic aromatic hydrocarbons (PAHs) in the area of the railway junction Ilawa Glowna. *Poland. Water Air and Soil Pollut.* **127**, 339–349. <https://doi.org/10.1023/a:1005236016074> (2001).
54. Hu, Y. N., Cheng, H. F. & Tao, S. The challenges and solutions for cadmium-contaminated rice in China: A critical review. *Environ. Int.* **92–93**, 515–532. <https://doi.org/10.1016/j.envint.2016.04.042> (2016).
55. 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.* <https://doi.org/10.1016/j.scitotenv.2019.134466> (2020).
56. Carne, G. *et al.* Mass balance approach to assess the impact of cadmium decrease in mineral phosphate fertilizers on health risk: The case-study of French agricultural soils. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2020.143374> (2021).

57. Xia, X. Q. *et al.* Soil heavy metal concentrations and their typical input and output fluxes on the southern Song-nen Plain, Heilongjiang province. *China. J. Geochem. Explor.* **139**, 85–96. <https://doi.org/10.1016/j.gexplo.2013.06.008> (2014).
58. Fu, Y. H., Li, F. M., Guo, S. H. & Zhao, M. Y. Cadmium concentration and its typical input and output fluxes in agricultural soil downstream of a heavy metal sewage irrigation area. *J. Hazard. Mater.* <https://doi.org/10.1016/j.jhazmat.2021.125203> (2021).
59. Wu, Q. M. *et al.* Spatial distribution, ecological risk and sources of heavy metals in soils from a typical economic development area Southeastern China. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2021.146557> (2021).
60. Liu, J., Zhang, X. H., Tran, H., Wang, D. Q. & Zhu, Y. N. Heavy metal contamination and risk assessment in water, paddy soil, and rice around an electroplating plant. *Environ. Sci. Pollut. Res.* **18**, 1623–1632. <https://doi.org/10.1007/s11356-011-0523-3> (2011).
61. Yang, B., Zhang, G. L., Deng, W. & Ma, J. In *2nd International Conference on Chemical, Material and Metallurgical Engineering (ICCMME 2012)*. 3239–3243 (2013).
62. Meharg, A. A. & Meharg, C. The pedosphere as a sink, source, and record of anthropogenic and natural arsenic atmospheric deposition. *Environ. Sci. Technol.* **55**, 7757–7769. <https://doi.org/10.1021/acs.est.1c00460> (2021).

Author contributions

Cang Gong: conceptualization, methodology, formal analysis, data curation, writing—original draft, visualization. Mingguo Lan and Haichuan Lu: methodology, software, formal analysis, investigation, data curation. Youchang Shi and Shunxiang Wang: conceptualization, writing—review and editing, supervision. Xiang Xia: conceptualization, investigation, supervision, writing—review and editing. Ying Chen: conceptualization, investigation, supervision, writing—review and editing. All authors have approval to the final version of the manuscript.

Funding

This research was supported by Open Foundation of the Key Laboratory of Natural Resource Coupling Process and Effects (No. 2023KFKTB011), Geological Survey Project of China Geological Survey (DD20243098).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.X. or Y.C.

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 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2024