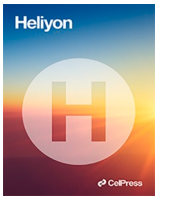
Heliyon 9 (2023) e21550

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Comprehensive model development based on Dempster–Shafer evidence theory for pollution source analysis in chemical parks

XueShan Bai [a], YongJie Yang [b], XiZhao Tian [a], Peng Wen [a,c], ZhiYuan Ma [a,*]

[a] Hebei Key Laboratory of Environment Monitoring and Protection of Geological Resources, Hebei Geological Environment Monitoring Institute, Shijiazhuang, 050021, China
[b] Hebei Solid Waste Pollution Prevention and Control Center, Shijiazhuang, 062659, China
[c] School of Geological Sciences, China University of Geosciences, Beijing, 100083, China

ARTICLE INFO

Keywords:
Evidence theory
Soil pollution
Pollution evaluation
Source analysis

ABSTRACT

Pollution source analysis is an effective method that can help chemical park managers accurately understand the characteristics and contributions of pollution sources in the park. However, as more receptor models are being used in this field, it has become difficult for managers to find the best interpretable and reasonable model among many source analysis models.

Here, we present a case study of pollution source analysis in a southern chemical park using the D–S evidence theory approach to combine the source analysis results of different receptor models for comprehensive consideration. Receptor models were used to analyse the pollution sources via positive definite matrix decomposition, principal component analysis–multiple linear regression, and Unmix models.

The results demonstrated that source analysis was greatly influenced by the uniqueness of pollutant characteristics and model receptor differences. Furthermore, incomparable analysis results and low fineness were observed. The D–S evidence theory model proposed in this study solved the above-mentioned problem to some extent and successfully extracted the four primary pollution sources in the study area, of which 45.73 % came from the metal processing industry (F1), whose primary pollutants were Cr, Ni, Zn, Cr(VI), and Cu, and 25.12 % came from the electronics manufacturing industry (F2), whose primary pollutants were Pb, Cr(VI), Cu, and Zn. 15.62 % of the contamination came from the production of chemical agents (F3), whose primary pollutant was TEHP, and 13.53 % came from the use of oil-containing auxiliary materials (F4), whose primary pollutant was TPH.

The D–S evidence theory model used in this study provides a reference for the management of chemical parks.

1. Introduction

The risk and danger of soil pollution in the chemical parks of China has increased in recent years owing to the high concentration of chemical enterprises and the continuous expansion of the construction scale, production, and storage devices in chemical parks, causing tremendous pressure on the surrounding and regional environments [1]. Furthermore, excessive accumulation of soil contaminants in chemical parks poses a substantial environmental risk [2]. Thus, accurate identification of the primary sources of soil

* Corresponding author.
E-mail address: 157698934@qq.com (Z. Ma).

https://doi.org/10.1016/j.heliyon.2023.e21550
Received 12 May 2023; Received in revised form 12 October 2023; Accepted 23 October 2023
Available online 31 October 2023
2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

contaminants by managers can aid in the development of relevant source reduction measures and make soil remediation cost-effective [3].

Globally, the methods of source resolution studies are complex and diverse and primarily include geostatistical [4], stable isotope ratio [5], receptor model [6], multivariate statistical [7], and soil profile analyses [8]. In China, greater attention is paid to general pollutants in the atmosphere and water bodies than the emission sources of soil pollutants to analyse the potential sources of pollution in chemical parks. Furthermore, no clear emission list is available. Therefore, the source-based emission list and process-based diffusion model methods are less used in soil pollution source analysis, whereas the pollutant-based receptor model method is widely used in its analysis.

The commonly used receptor models include factor analysis (FA), cluster analysis (CA), chemical mass balance, principal component analysis–multiple linear regression (PCA–MLR), positive definite matrix factor analysis (PMF), and multi-source linear analysis (Unmix) models [9]. The above-mentioned methods obtain datasets from local environmental media (receptors) affected by contaminants and then use various means to extract meaningful information to classify samples and identify the sources of contamination. The PMF model is a source analysis model and a multivariate FA tool (that evolved from FA) recommended by the U.S. Environmental Protection Agency (EPA) [10]. PCA–MLR first uses PCA to analyse the receptor and identify the primary source. Subsequently, it uses MLR to resolve the source contribution and obtain quantitative source resolution results [11]. The Unmix model uses the singular value decomposition method to reduce the dimensionality of the data space, which can determine the number, composition, and contribution of sources to each sample [12].

When considering complex soil pollution problems using receptor model analysis methods, the obtained source analysis results may differ owing to different principles or assumptions of different soil source analysis methods, which have their respective advantages and limitations in application. Therefore, a single source analysis method cannot ensure accurate results [13,14]. Some studies have used a combination of source-resolution models and selected source-resolution techniques based on the local conditions of the study area. For example, Hong Liu used three models (PCA–MLR, UNMIX, and PMF) to analyse the sources of heavy metals in the soil of a small town in Lishui City, Zhejiang Province, and compared the source analysis results with the risk assessment results to improve their reliability [15]. Some studies have combined multiple methods to improve single-source analysis models. For example, Yufeng Li used a geographically weighted regression method to improve the PCA model and constructed a robust absolute principal component score–robust geographically weighted regression receptor model to analyse the pollution sources of eight heavy metals (Cd, Hg, As, Pb, Ni, Cu, and Zn) in a city in southern China. The results revealed that the joint model had more consistent source identification and assignment [16]. Wang successfully examined the sources of heavy metal contamination in different site type soils at an e-waste treatment and disposal centre in China using a synergistic source allocation framework comprising positive matrix decomposition (PMF) and hybrid Bayesian stable isotope analysis (MixSIAR) models [17]. Agyeman applied the new decomposition of pollution factor positivity matrix (CF–PMF) receptor model based on pollution assessment to estimate the pollution level in seven cities in the Frydek-Mistek District; the root mean square and mean absolute errors of the new receptor model were significantly reduced [18].

With the gradual maturation of receptor model analysis methods, hybrid methods have become inevitable. Furthermore, they tend to diversify. However, relatively few studies have been conducted on source-resolution model integration. Based on the above considerations, an integrated source resolution model (Dempster–Shafer (D–S) evidence theory) based on evidence theory is proposed in this study. Evidence theory (DS), first proposed by Dempster, is a powerful tool for the representation and fusion of decision uncertainty information and has demonstrated good applicability in related fields [19]. The primary objectives of this study were to combine source resolution models using evidence theory to integrate and optimise the results of the PCA–MLR, PMF, and Unmix models and compensate for or eliminate the drawbacks of a single model in the source resolution process.

2. Data sources and methods

2.1. Research area

The study area is located in a chemical park in southern China, covering an area of more than 600 acres, with a production history dating back to the 1990s. The primary industrial categories include chemicals (pesticides), chemical fibres, synthetic leather, printing, hardware, and machinery. The stratigraphic distribution of the study area from top to bottom is as follows: miscellaneous fill (0–500 cm), clay (50–1050 cm), sand (600–1950 cm), and clay layers (>1700 cm). The groundwater depth of the plot during the sampling period was approximately 1.30–2.31 m. The elevation is 1.79–2.96 m. Because the plot is located near the sea outlet, the internal groundwater flow direction fluctuates greatly under the influence of tides. Therefore, seasonal differences are observed in the flow direction throughout the year. The flow direction during the sampling period was northeast to southwest. Furthermore, the surface water recharges the groundwater in the study area. This area has a long production history involving various industrial categories, and the distribution and sources of soil contaminants are complex and comprise multiple dimensions.

2.2. Sampling and chemical analysis

We conducted a survey using GPS to determine the specific location and ground elevation of the onsite sampling points and develop a sampling plan that reflects the actual situation of the site according to the production layout, material storage, and use of raw and auxiliary materials of each enterprise in the study area. The preliminary survey results and relevant technical specification requirements. A 40 m × 40 m grid was used for the distribution of points, and a 20 m × 20 m encrypted grid was used for the distribution of points in key areas and areas with obvious onsite soil contamination. A total of 967 soil samples were collected using stratified

sampling (Fig. 1). Representative soil samples were selected and sent to the ISO 9002 certified Hangzhou Zhongyi Testing Institute for further analysis. Atomic fluorescence and plasma mass spectrometry were used to determine the concentration of heavy metals, and gas chromatography–mass spectrometry was used to determine the organic content. The precision and accuracy of the analytical methods were controlled using national soil standard GSS-15 and indoor parallel samples. A total of 59 soil pollution indicators were detected in this study, and eight pollutants with high detection rates and values of total chromium (Cr), hexavalent chromium (Cr(VI)), Cu, Zn, Pb, nickel, THP, and DEHP were selected as research objects using statistical analysis. The range of pollutants exceeding the standards in the study area was obtained using the ordinary kriging interpolation method combined with the ground statistical method (Fig. 2).

2.3. Risk assessment methods

To assess pollution, various evaluation indicators, such as the pollution index (PI), Nemerow integrated pollution index, and geoaccumulation index (GI), were used in this study [20,21]. In addition, a potential ecological index was used to assess the environmental risks caused by toxic metals owing to their different toxicities [22,23]. Geographic information system techniques provide the best linear unbiased estimates of regionalised variables in a limited area and are used to analyse the spatial distribution and heterogeneity of pollution conditions [24,25]. The kriging method is essentially a linearly smoothed low-pass filter that estimates the mean of the mathematical expectations of the conditions. It completes the understanding of the spatial pattern but does not reproduce the spatial structure. Furthermore, the extreme value points are smoothed out such that some important information may be lost, which limits the application of this mainstream geostatistical technique in this study [26].

2.4. D–S evidence theory

The basic concepts of evidence theory are as follows：

(1) Identification framework

For a problem, the discriminative framework comprising all its possible outcomes is represented by Θ, and Θ is a non-empty set. The combination of all subsets in Θ is called the power set and is denoted by 2^{Θ}. The power set for the discriminative framework Θ = {A, B, C} is as follows.

$$2^{\Theta} = \{\varphi\ ,\ (A)\ ,\ (B)\ ,\ (C)\ ,\ (A、B)\ ,\ (A、C)\ ,\ (B、C)\ ,\ (A、B、C)\} \tag{1}$$

(2) Basic probability assignment (BPA)

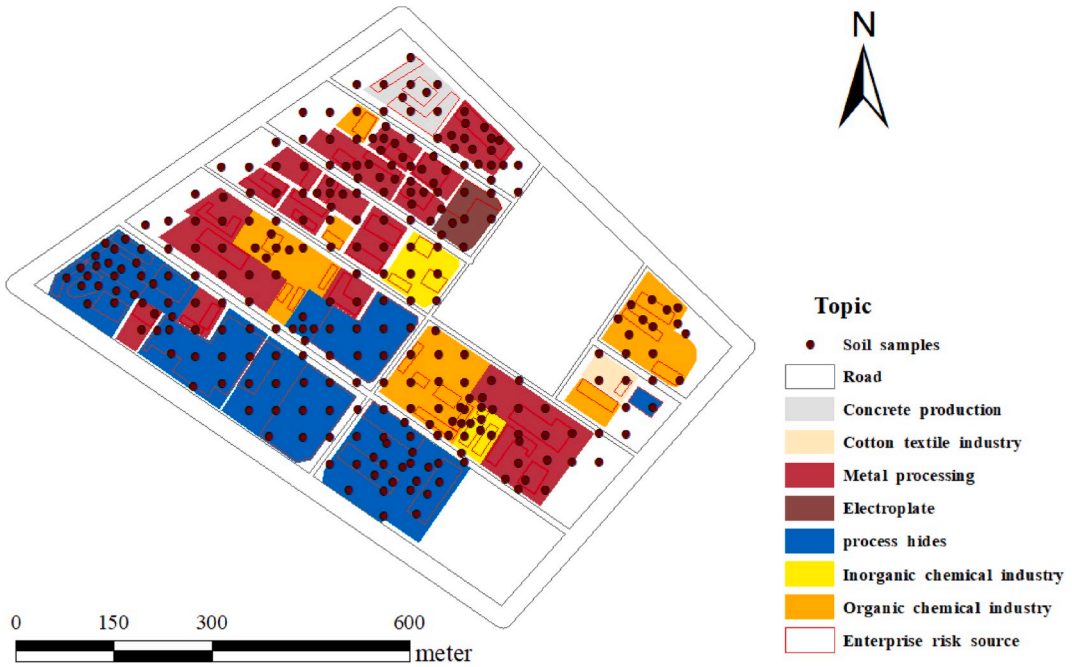


Fig. 1. Distribution of sampling sites in the study area.

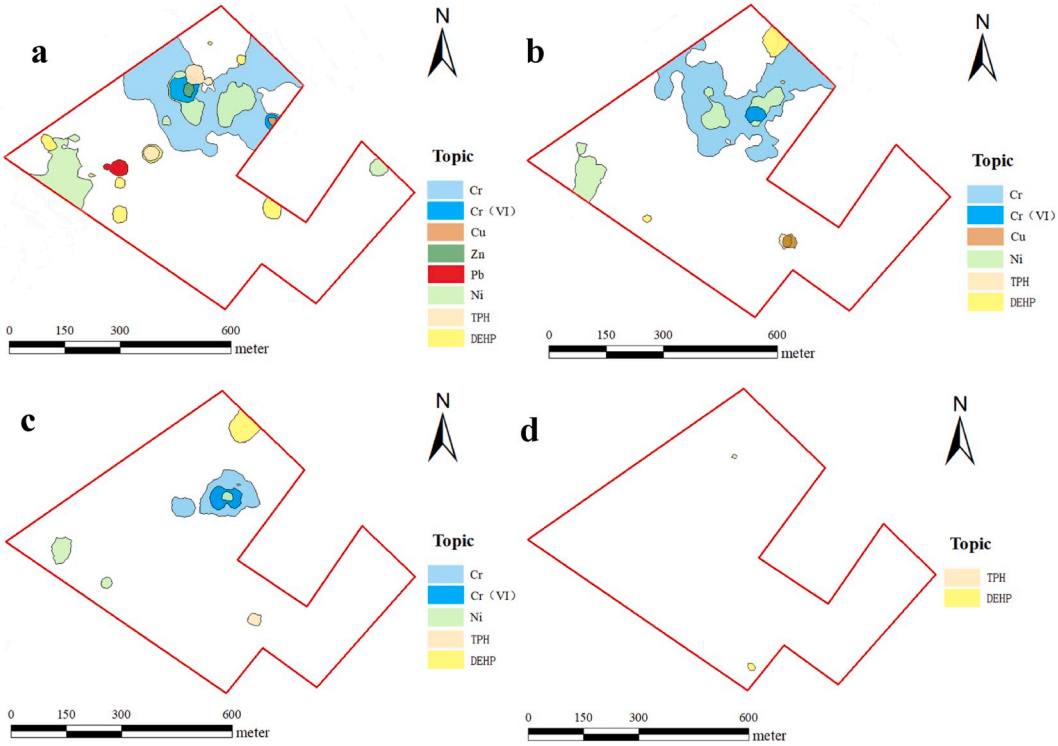


Fig. 2. Spatial distribution of excessive pollutants in the study area (5.2–3.2 m、3.2–1.2 m、1.2–0.8 m、-0.8~2.8 m) .

$$\begin{cases} \mathrm{m(A)} \geq 0 \ , \text{ for all } \mathrm{A} \in 2^{\Theta} \\ \mathrm{m}(\varphi) = 0 \\ \sum\limits_{A \subseteq 2^{\Theta}} \mathrm{m(A)} = 0 \end{cases} \tag{2}$$

If m(A) ≠ 0, then A is considered a focal element. In evidence theory, all available information about an event can be considered evidence, which can be fused based on the evidence synthesis rule originally proposed by Dempster. Suppose that m_1, m_2, … m_n are n BPAs in the discriminative framework, and the corresponding focal element is Ai (i = 1, 2, … n).

$$\begin{cases} 0 \ , \quad \mathrm{A} = \varphi \\ \dfrac{\sum\limits_{A_i \cap B_j = A} \mathrm{m}_1(A_i)\mathrm{m}_2(B_j)}{1-K} \ , A \neq \varphi \end{cases} \tag{3}$$

Then $K = \sum\limits_{\cap A_i = \varphi 1 \leq i \leq n} \prod \mathrm{m}_i(A_i)$ is the conflict coefficient. The larger the K, the more obvious the conflict between the evidence. When K = 1, the evidence is highly conflicting and Eq. (3) fails. Many improvement methods are currently available; however, only one easy-to-operate method is presented in this study to cope with the highly conflicting evidence. For a problem with m focal elements, the BPA is modified as

$$m(A_i) = \begin{cases} 10^{m(A_i)-1/m} \ , \ m(A_i) < \dfrac{1}{m} \\ 10^{m(A_i)-1/m} \ , \ m(A_i) \geq \dfrac{1}{m} \end{cases} \tag{4}$$

and normalized using the following equation.

$$m(A_i) = \frac{m(A_i)}{\sum\limits_{i=1}^{m} m(A_i)} \tag{5}$$

in this case, Eq. (3) is still available.

3. Results and discussion

3.1. Contamination assessment

Table 1 lists the basic characteristics of the eight pollutants in the study area. The concentrations of Cr, Cr (VI), Cu, Zn, Pb, Ni, THP, and DEHP ranged from ND to 14400 mg/kg, ND to 72.8 mg/kg, ND to 33300 mg/kg, ND to 22800 mg/kg, ND to 495 mg/kg, ND to 1480 mg/kg, ND to 15440 mg/kg, and ND to 314 mg/kg, respectively. The average concentrations decreased as follows: Zn (158.92 mg/kg) > Cu (155.01 mg/kg) > THP (128.08 mg/kg) > Cr (82.28 mg/kg) > Ni (53.71 mg/kg) > Pb (13.78 mg/kg) > DEHP (8.1 mg/kg) > Cr (VI) (0.2 mg/kg).

The pollutant concentrations were highly variable in the study area, and there were different degrees of exceedance; the maximum exceedance multiples ranked from the largest to the smallest as follows: Cr (96) > Cr (VI) (24.27) > THP (18.69) > Cu (16.65) > Ni (9.87) > DEHP (7.48) > Zn (6.517) > Pb (1.24). The maximum values of pollutant concentrations were deducted from the background values and then compared with the background values. The change rates of pollutant concentrations were Cu (2311.5 %) > Cr (VI) (996.26 %) > THP (592.85 %) > DEHP (474.76 %) > Cr (435.36 %) > Zn (220.36 %) > Ni (45.25 %) > Pb (25.05 %). The results revealed that the maximum value of each pollutant concentration increased significantly compared with the background value, and the largest increase was observed in Cu concentration. The coefficient of variation (CV) was positively correlated with the intervention effect of anthropogenic activities. Eight pollutants exhibited large CV values ranging from 2.3 to 16.08, which is a strong variation and indicates that the pollutants in the study area are affected by complex natural and anthropogenic sources [27].

Fig. 3 shows the status evaluation (PI values) results of the pollutants in the study area. The single-factor index evaluation results revealed that the average PI values were <1 for all pollutants (Cr: 0.64, Cr(VI): 0.07, Cu: 0.08, Zn: 0.05, Pb: 0.04, Ni: 0.36, THP: 0.16, and DEHP: 0.19) based on the national standard screening values; however, the maximum values were higher than the critical values. Among them, Cr had the highest contamination, with lightly and heavily contaminated samples, whereas Zn and Pb had the lowest contaminations, with only one heavily and lightly contaminated sample, respectively. Other contaminants, such as Cr (VI), Cu, Ni, THP, and DEHP, were present in the study area to varying degrees. To compare the pollution severity of each element, the single-factor evaluation results were ranked using the following method. According to the single-factor cumulative index evaluation method, the average pollution index was ranked from the largest to the smallest, resulting in the pollution severity ranking of each element. The study area pollution severity ranking was as follows: Cr > Ni > DEHP > THP > Cr (VI) > Cu > Zn > Pb.

Fig. 4 shows the Nemerow composite and pollution load indices evaluation results. When compared with the single-factor index evaluation results, the Merrow composite index evaluation results based on the national standard threshold and soil background values exhibited that the same sample site had a more serious degree of composite pollution. This reflects the limitations of the risk composite index evaluation. When there are more extreme values of different types of heavy metals at the calculated sample points, the Nemerow index may sometimes exaggerate or reduce the influence of certain elements in the evaluation, which makes it less sensitive to environmental quality [28,29]. Therefore, a mathematical model of pollution load index method was used for further analysis. The pollution load index method adopts 'summation and opening square' to calculate the integrated pollution degree of multiple pollutants. Therefore, it is easy to ignore the role of high pollutants, especially the more types of pollutants, and the accelerated convergence leads to considerable smoothing effect in high values, which causes uncertainty in the risk degree of multiple heavy metals in the evaluation results [30]. These results indicate that the combined pollution status of the pollutants in the study area was at an early warning level [31]. These samples were concentrated in the northwestern and northcentral parts of the study area in the first and second layers, respectively, suggesting top-down heterogeneity in the spatial distribution of pollution.

Fig. 5 shows the GI values of heavy metal pollutants compared with that of the local background. Owing to various anthropogenic activities, all heavy metal pollutants accumulated in the soil. Among them, Cu was the most enriched (62 %), accounting for moderate or above-moderate accumulation in 134 samples; Cr(VI) was the least enriched (10 %), accounting for moderate or above-moderate accumulation in 59 samples; and the remaining toxic metals were enriched to different degrees as follows: Cr: 33 %, Zn: 27 %, Pb: 22 %, and Ni: 36 %.

Table 1
The statistical characteristics and standard values of test indicators mg/kg.

Indicators	pH	Cr	Cr (VI)	Cu	Zn	Pb	Ni	THP	DEHP
Minimun	4.45	ND	ND	ND	ND	ND	ND	ND	ND
Maximun	11.92	14400	72.8	33300	22800	495	1480	15440	314
Average	8.33	82.28	0.2	155.01	158.92	13.78	53.71	128.08	8.1
Mild	8.48	ND	ND	28.4	117	ND	40	29	2.69
Standard deviation	0.62	563.52	3.18	1472.2	753.21	30.14	123.7	773.13	19.93
Variable coefficient	0.07	6.85	16.08	9.5	4.74	2.19	2.3	6.04	2.46
Back ground value	8.78	33	0.073	14.4	103	19	32	26	0.66
Standard values	6.5 ≤ pH ≤ 7.5	150	3	2000	3500	400	150	826	42
Percentage (%) higher than background value	–	35 %	11 %	71 %	63 %	30 %	65 %	54 %	78 %
Percentage (%) higher than Standard value	–	7.9 %	1 %	7 ‰	1 ‰	1 ‰	5 %	2 %	4 %
Maximum excess multiple	–	96	24.27	16.65	6.517	1.24	9.87	18.69	7.48
detection limit	–	0.4	0.016	0.1	0.1	0.1	1.0	5	0.001

Note : The standard values refer to the national standards of the People's Republic of China for soil environmental quality GB36600–2018 and GB15618-2018 for soil pH value and coefficient of variation.

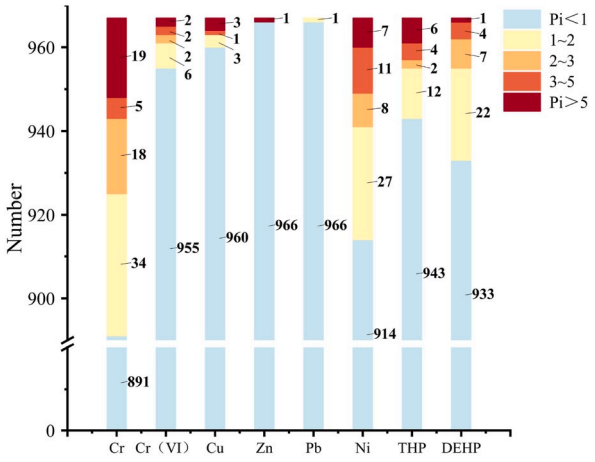


Fig. 3. Pollution index statistics of toxic pollutants in soil.

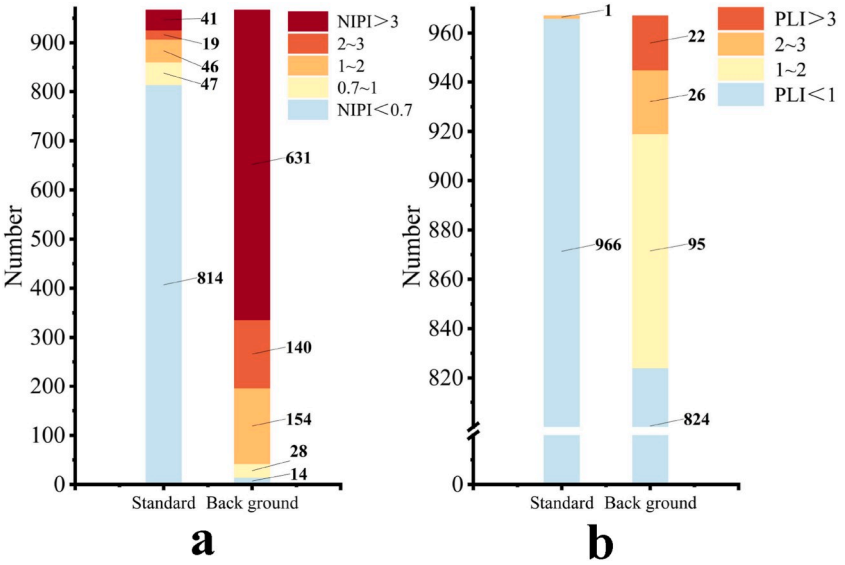


Fig. 4. Statistics of Nemero composite index and pollution load index of toxic pollutants in soils.

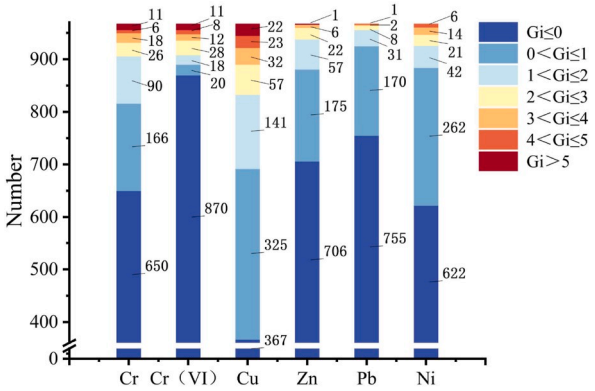


Fig. 5. Statistics of the geoaccumulation index for toxic metals in soil.

Fig. 6 shows the ER values of the toxic metals considering their toxicity. The results exhibited that the average ER values of the individual toxic metals were below 40, except for Cu (Cr: 19.7, Cu: 54.6, Zn: 1.6, Pb: 3.8, and Ni: 9.1). Although Cu is a high-risk pollutant, it is less harmful to local ecosystems owing to its relatively low toxicity. The accumulations of Cr(VI) and Ni were more severe. In particular, Cr(VI) and Ni are highly toxic elements and pose considerable ecological risks.

Based on the above-mentioned various evaluation indices results, contaminants Cr, Ni, and Cu were distributed in different soil strata in the study area. The upper soil layer was more widely distributed and had a larger exceedance multiple and variation coefficient, indicating that the degree of pollution has a serious and substantial cumulative effect on the soil. The pollutants Cr (VI), DEHP, and THP had the widest distribution and the deepest pollution depth, but their pollution severity was slightly lower. The pollutants Pb and Zn were concentrated in the surface soil of the study area and had the lowest pollution level. Therefore, the following priority order was adopted to determine the pollutants of concern in the study area: Cr, Ni, Cu, Cr(VI), DEHP, THP, Pb, and Zn. The integrated factor indices evaluation results exhibited that the northwestern and northcentral regions of the study area had the most serious integrated pollution levels. On comparing the geostatistical results of pollutant distribution and the actual production in the study area, this was revealed to be the most densely populated area for industrial activities. Pollutants are released into the surrounding environment through industrial production processes, making the distribution of contaminants in this area complex. Furthermore, multiple pollutants are superimposed to form complex pollution. Therefore, we focused on these areas in the subsequent source analysis.

3.2. Source apportionment

3.2.1. Correlation analysis

Correlation analysis is usually used as a basis for homologous identification of pollutants [32] . Table 2 shows that Zn was closely related to Cu ($R = 0.78$, $P < 0.001$), Cr(VI) with Cu and Zn ($R = 0.26$–0.34, $P < 0.001$), Pb with Zn ($R = 0.31$, $P < 0.001$), TPH with Cu and Zn ($R = 0.26$–0.29, $P < 0.001$), THP with EDHP ($R = 0.22$, $P < 0.001$), and other pollutants were weakly related to each other. Initially, the sources of pollution were judged to be correlated. However, because of the presence of many types of production industries in the study area, the sources of soil heavy metal pollution in different regions were highly variable and required further analysis.

3.2.2. PCA–MLR model

The PCA–MLR model is a pollution source analysis model that combines PCA and MLR. PCA helps reduce the dimensionality of the dataset and identify potential pollution sources, whereas MLR establishes a relationship between pollutant concentrations and potential sources. This combined approach allows for more accurate identification and quantification of pollution sources. The results of the KMO test value (0.619) and Bartlett's shape degree ($\xi = 0.000$, $p < 0.05$) analyses indicated that PCA could be applied to the pollutant concentration data in this study. After the eight pollutants were analysed using PCA, a total of four principal components (PC1, PC2, PC3, and PC4) were obtained with variance contributions of 28.71 %, 15.66 %, 13.13 %, and 13.03 %, respectively, and a cumulative variance contribution of 70.53 % (Fig. 7). In terms of factor loadings (F), PC1 was dominated by Cr(VI) ($F = 0.830$), Zn ($F = 0.819$), Ni ($F = 0.730$), and Cr ($F = 0.583$); PC2 was dominated by TPH ($F = 0.777$) and DEHP ($F = 0.756$); PC3 was dominated by Pb ($F = 0.817$); and PC4 was dominated by Cu ($F = 0.516$), Pb ($F = 0.362$), and Zn ($F = 0.325$). Cr(VI) ($F = 0.236$) dominated in PC1, PC3 and PC4, whereas organic matter dominated in PC2. These results indicated that the different heavy metals and organic pollutants in the plots were homologous. The positive loadings of both Zn and Cr(VI) were high in PC1 and PC4, indicating that both Zn and Cr(VI) had widespread sources of contamination.

Correlation CA of the eight pollutant concentrations in the study area classified them into three clusters (Fig. 7). The first cluster comprised Cr(VI), Cr, Zn, Ni, and Cu; the second cluster comprised TPH and DEHP; and the third cluster comprised only Pb. The results obtained using the multivariate statistical analysis were similar. To further quantify the primary sources of pollution and their relative contribution to the total pollution in the study area, a multiple stepwise linear regression analysis was conducted with the pollution pathway and the total concentration of pollutants as the independent and dependent variables, respectively. The regression equation was obtained as follows: $Y = 0.662F1 + 0.426F2 - 0.375F3 + 0.249F4$ ($R^2 = 0.821$, $F = 1118.516$, and $P = 0.000$), where PC3 was a

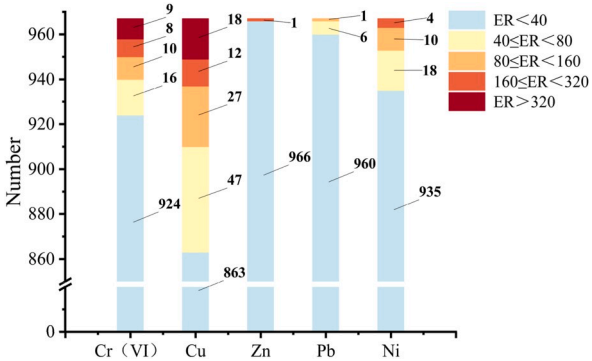


Fig. 6. Statistics of the potential ecological index for toxic metals in soil.

Table 2
Correlation and significance test of pollutant content.

	Cr	Cr (VI)	Cu	Zn	Pb	Ni	THP	DEHP
Cr	1							
Cr (VI)	0.37* *	1						
Cu	0.56* *	0.34* *	1					
Zn	0.47* *	0.26* *	0.78* *	1				
Pb	−0.48* *	−0.17* *	0.11* *	0.31* *	1			
Ni	0.52* *	0.19* *	0.45* *	0.51* *	0.12* *	1		
THP	0.14* *	0.057	0.29* *	0.26* *	0.059	−0.033	1	
DEHP	0.036	−0.043	0.059	0.067*	0.089* *	0.1* *	0.22* *	1

1 : * * : Significant correlation at 0.01 significance level , * : Significant correlation at 0.05 significance level。.

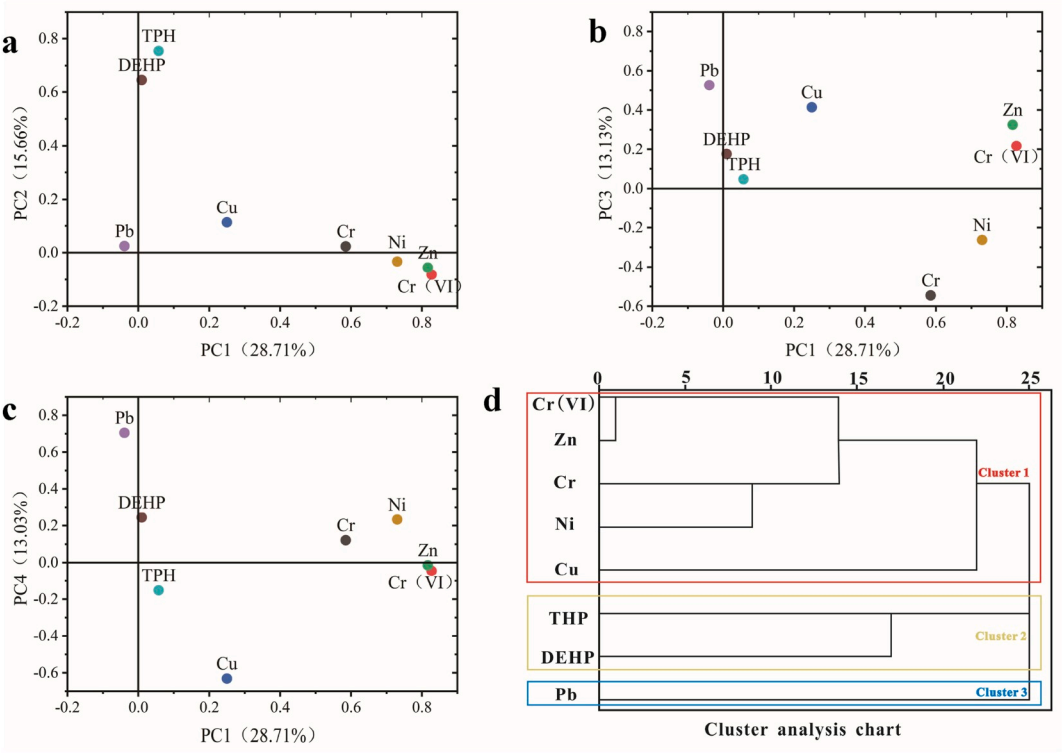


Fig. 7. PCA-MLR model pollutant component loading chart and cluster analysis chart in the study area.

negative contribution margin, which was changed to zero in subsequent analysis.

3.2.3. PMF model

The PMF model was developed using EPA PMF 5.0. The signal-to-noise ratios of all pollutants were >1.4, which can be classified as strong information. The model parameters were set, the species categories of the eight pollutants were set to 'strong' for model fitting, and the residuals were set to meet the interval of [−3, 3]. Furthermore, 3, 4, 5 and 6 factors were set as the number of potential elements, the model was run 50 times each, and over 90 % of the basic factors were recreated without interchanging. Based on these results, the PMF model was considered robust. To better interpret the results, Fpeak transformation was attempted. Because the data could not satisfy the Q-value variation of less than 5 % for the Fpeak transformation under various conditions, the pre-transformation results were used for the analysis. After comparing the Q-value changes and displacement (DISP) and bootstrap (BS) tests, the model could better explain the sources of pollution in the study area for four potential factors. The first factor (F1) explained 12.59 % of the total contamination factors in the soil and had a large TPH load (99.6 %). The second factor (F2) explained 30.73 % of the total contamination factors, including heavy loadings of Pb (97.1 %) and medium loadings of Zn (42.2 %) and Cu (35.5 %). The third factor (F3) was dominated by DEHP (97.4 %), accounting for 15.10 % of the total pollutants. The fourth element (F4) accounted for 41.58 % of the total pollutants and was dominated by Cr (99.4 %), Cu (67.3 %), Ni (62.6 %), Zn (57.7 %), and Cr(VI) (45.6 %).

3.2.4. Unmix model

The sample data were simulated using EPA UNMIX6.0 software. First, the data were processed and those with a concentration variance greater than 50 % were automatically screened using the Suggest Exclusion Tool. The results exhibited that Pb and DEHP had high concentration variances among the eight pollutants. Furthermore, all of them were included in the model because they were crucial for source analysis. The UNMIX model uses geometric 'edges' to analyse the sample data. We used the concentration sum of each pollutant metal as the base and the individual pollutant metal elements as the dependent variables to analyse the edge diagnosis results of each component, taking the concentrations of all 8 pollutants as the initial species and the concentration sum (total) as the total and standard species of the model input components, and calculated the 4-source solution that meets the requirements. The results showed that MinRsq = 0.77 and MinSig/Noise = 2.00, both of which met the model requirements. This indicated that the model could explain 77 % of the species variance and that the source resolution scheme results were reliable. We used Excel to further calculate the source contribution. Source 1 contributed more to Cr, Cr(VI), Cu, Ni, and Zn, whereas its contribution to other elements was relatively small; source 2 contributed to Pb, Zn, and Cu; source 3 contributed to DEHP and Pb; and source 4 contributed to TPH.

3.2.5. D–S evidence theory model

In this study, n models were selected for soil pollution source analysis in chemical parks. The n results constitute the identification framework. Knowing their corresponding BPA values, the relevant results can be fused into one result, which can be expressed by Eqs.

$$Y_n = f_n\,(X) \tag{6}$$

$$G = \sum_{i=1}^{n} m(Y_i) \times Y_i \tag{7}$$

It is assumed that the chosen model i supports 100 % of the results obtained using itself but not the results obtained using other models. Therefore, the BPA matrix was obtained, as shown in Table 3, and the evidence was then fused. Taking n = 3 as an example, the identification framework for model 1 is {model 1 result, model 2 result, model 3 result}, and their corresponding BPA values are {1, 0}; the identification frameworks for models 2 and 3 are {model 1 result, model 2 result, model 3 result}, and their corresponding BPA values are {0, 1}. From Eq. (3), we have K = 1, and as the evidence is highly conflicting, Eq. (3) fails. Therefore, Eqs. (4) and (5) were chosen to correct the BPA. After correction, models 1, 2, and 3 corresponded to the BPA values of {0.9901, 0.0099}, {0.0099, 0.9901}, and {0.9901, 0.0099}, respectively, and K = 0.9802; therefore, we obtained the fused identification framework {model 1 result, model 2 result, model 3} corresponding to the BPA values of {1/3, 1/3, 1/3} from Eq. (3). Furthermore, the same BPA values were obtained when none of the three results could negate each other. Thus, it can be observed from Eq. (7) that the calculated result is essentially the average of multiple results.

According to the principle of evidence theory, the BPA corresponding to the identification framework {Model 1, Model 2, Model 3} is {1/3, 1/3, 1/3} based on calculations using Eqs. (3)–(5). The D–S evidence theory model was constructed using Eq. (7), and the results are shown in Fig. 8.

3.3. Source distribution and identification

The PCA–MLR, PMF, and Unmix models had similar compositions of F1, which accounted for 47.23 % of the total pollutants in the PCA–MLR model. Furthermore, Cr(VI), Zn, Ni, and Cr had large loadings in PC1. F1 accounted for 41.58 % of the total pollutants, primarily Cr, Cu, Ni, Zn, and Cr(VI), in the PMF model, and for 48.53 % of the total pollutants, primarily Cr, Cr(VI), Ni, Cu, and Zn, in the Unmix model. According to the correlation analysis, the relationships among Cr (VI), Zn, Ni, Cr, and Cu were close; the areas of Cr (VI), Zn, Ni, Cr, and Cu exceeded the standard; and the metal processing industry distribution exhibited good agreement. Cr(VI), Zn, Ni, Cr, and Cu are characteristic elements in the metal processing industry [33–35]. The metalworking industries involved in this study include metal product, ferrous metal smelting, and rolling processing industries, whose primary products are fasteners, punching parts, rusted steel pipes, and daily hardware. The metal dust and sewage produced by the metalworking industries enter the soil through atmospheric deposition and rainwater erosion. Therefore, F1 can be regarded as a source of activity related to the metal processing industry.

Furthermore, the PCA–MLR, PMF, and Unmix models had similar compositions of F2, which explained 22.24 % of the total variation in the PCA–MLR model. There were high positive loadings of Pb, Cu, and Zn. F2 accounted for 30.73 % of the total pollutants, primarily dominated by Pb, Cu, Zn, and Ni, in the PMF model and for 22.46 % of the total pollutants, primarily dominated by Pb, Cu, and Zn, in the Unmix model. These are the primary pollutants in the electronics manufacturing industry [36–38]. The high value areas of Cu, Pb, and Zn in the study area are primarily distributed in the electronics manufacturing industry, and the manufacturing of

Table 3
The BPA matrix when model i supports 100 % of the results obtained by itself, but not the results obtained by other models.

Model	Model 1 result	Model 2 result	…	Model n result
1	1	0	0	0
2	0	1	0	0
⋮	0	0	1	0
n	0	0	0	1

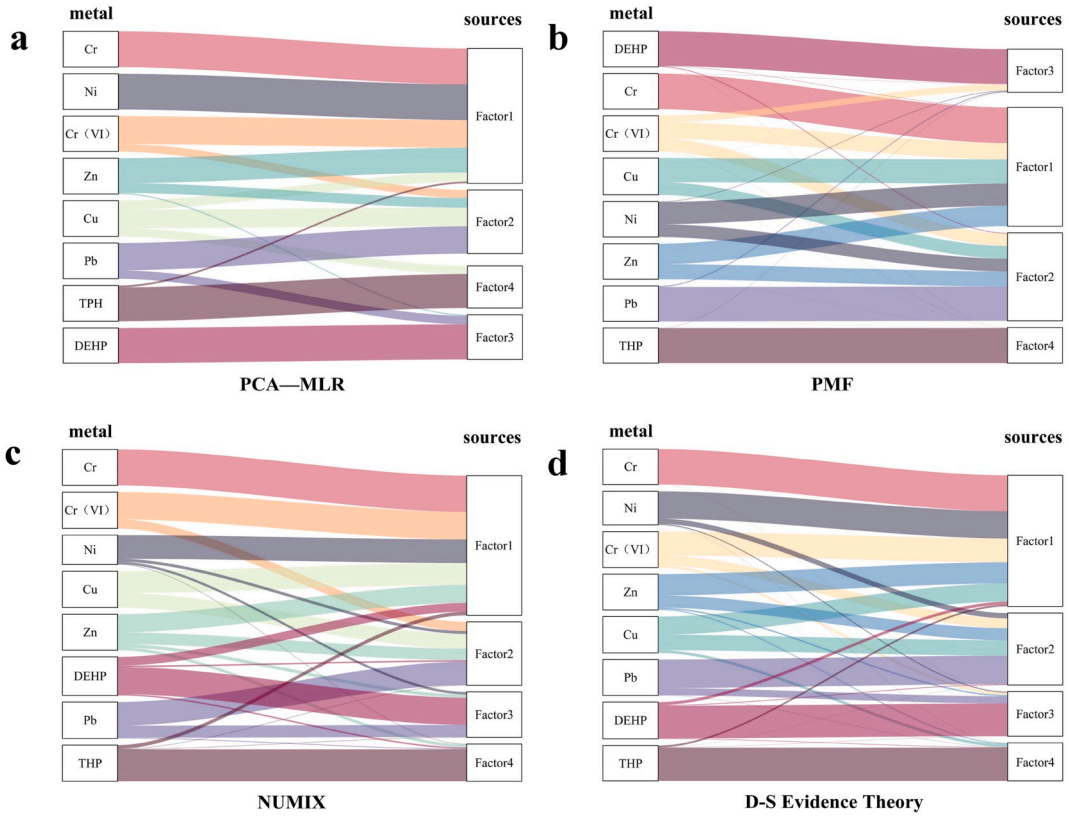


Fig. 8. Source analysis chart of pollutants in the study area.

electronic and electromechanical components and equipment is the primary source of generating Cu, Pb, and Zn in the study area. Therefore, F2 can be considered as a source of activity related to the electronics manufacturing industry.

The PCA–MLR, PMF, and Unmix models produced some divergence in the F3 and F4 components, with F3 explaining 30.55 % of the total variation in the PCA–MLR model. There were high positive loadings of TPH and DEHP. In the PMF model, F3 accounted for 15.10 % of the total pollutants, primarily dominated by DEHP, whereas F4 accounted for 12.59 % of the total pollutants, primarily dominated by TPH. In the Unmix model, F3 accounted for 16.03 % of the total pollutants, primarily dominated by DEHP, whereas F4 accounted for 12.98 % of the total pollutants, primarily dominated by TPH. The TPH and DEHP polluted areas in the study area had some compound polluted areas and also separate pollution. Some pollution sources with small contributions were not screened and treated; instead, only the primary source components were screened and identified owing to the lack of sensitivity in the PCA–MLR model [39], which identified the sources of TPH and DEHP as the same. The PMF and Unmix models are more sensitive to source components and categorize for similar source components [40]. Therefore, these models did not identify the sources of TPH and DEHP as the same. The areas with high values of TPH and DEHP in the study area are primarily distributed in synthetic leather production, chemical raw material production, and metal processing industries. The primary raw materials used in synthetic leather production enterprises are various types of resins. DEHP enters the air in gaseous form during the production process and then diffuses to pollute the locations of the enterprises and their adjacent areas [41]. Plastic steel is made during the processing of plastic steel doors and windows comprising PVC raw materials, and the production process involves high-temperature heating, causing DEHP pollution in the form of atmospheric dispersion, which contaminates the plot [42]. The primary anthropogenic sources of TPH pollution are the use of diesel, lubricating oil, and heat-transfer oil auxiliary materials in production processes [43]. To not lose the model results and better integrate the sources using the evidence theory approach, the sources of TPH and DEHP were identified separately in the PCA–MLR model. Therefore, F3 and F4 can be regarded as sources of activities related to the production of organic chemical reagents and use of oil-containing excipients, respectively.

The PCA–MLR model identified three effective principal components, whereas the PMF and Unmix models had four optimal factors with inconsistencies among the three. By identifying the different pollution sources, we discovered that the pollutants of the principal components obtained using the PCA–MLR, PMF, and Unmix models were the same, whereas the PMF and Unmix models had more refined divisions of anthropogenic pollution sources. The PCA–MLR model failed to further divide the two anthropogenic sources, chemical preparation production and oil-containing auxiliary use, into different factors during the analysis. When considered together, the pollution source compositions and contribution rates of the three models were extremely similar, indicating that the source allocation results were reliable and that the differences were primarily among the contribution rates of the individual pollutants. All

three models inevitably expanded or ignored the connections between some pollutants during the calculations; Unmix identified multiple sources of DEHP and THP, and PCA–MLR and TPH models identified a single source of DEHP and THP. Furthermore, the PMF model had Pb only from F2, and the PCA–MLR model had Ni only from F1. The D–S evidence theory model was used to determine the four sources of the eight pollutants and their specific contribution rates. In 45.73 % of the metal processing industry, the primary pollutants were Cr, Ni, Zn, Cr(VI), and Cu; in 25.12 % of the electronics manufacturing industry, the primary pollutants were Pb, Cr (VI), Cu, and Zn; in 15.42 % of the chemical preparation production industry, the primary pollutants was DEHP; and in 13.53 % of the oil-containing auxiliary materials, the primary pollutant was TPH. While retaining the primary source compositions and contribution rates of the three models, the D–S evidence theory model discriminates and calculates the source types and their contribution rates in an integrated manner and combines the distribution of pollutants in the study area, where multiple sources of pollution inevitably exist for a single pollutant in real situations. Therefore, the source analysis model results evaluated using the D–S evidence theory model are more interpretable and reasonable. In the pollution source analysis, the results obtained using several models corroborated with each other (Fig. 9).

4. Conclusions

Multiple indicators were used to evaluate the risk of soil contamination comprehensively. In addition, a D–S evidence theory model in combination with evidence theory was proposed to quantitatively assign and determine the specific sources of soil contaminants and their respective contributions. The new method was implemented in a southern chemical park and achieved satisfactory results.

The results revealed that the northwestern and northcentral regions of the study area had the highest combined contamination levels. Eight primary pollutants were identified in the following priority order: Cr, Ni, Cu, Cr(VI), DEHP, THP, Pb, and Zn. We verified the source analysis results of the PCA–MLR, PMF, and Unmix models using the receptor model for pollution source analysis. However, there were still some differences in terms of source assignment. The D-S evidence theory model reduced the limitations of the single-receptor model to a certain extent and improved the interpretability of source analysis results. Furthermore, it can reduce the decision burden of managers when facing multiple receptor models in practice. (15.42 %) > F4 using oil-containing auxiliary materials (13.53 %). The primary pollutants in F1 were Cr, Ni, Zn, and Cr (VI); in F2 were Pb, Cr(VI), Cu, and Zn; in F3 was TEHP; in and F4 was TPH.

In subsequent studies and applications, researchers should evaluate the source analysis results based on an effective quantitative evaluation system, introduce evidence weights to correct contradictory evidence, and use iterative ideas to correct the evaluation results to fully utilise the advantages of the D–S evidence theory approach.

Date availability statement

Has data associated with your study been deposited into a publicly available repository?
Data will be made available on request
All data included in this study are available upon request by contact with the corresponding author.

CRediT authorship contribution statement

XueShan Bai: Writing – review & editing, Writing – original draft, Project administration, Data curation. **YongJie Yang:** Validation, Resources, Funding acquisition, Formal analysis, Conceptualization. **XiZhao Tian:** Writing – original draft, Validation, Project administration, Methodology, Investigation. **Peng Wen:** Writing – original draft, Validation, Software, Project administration, Funding acquisition. **ZhiYuan Ma:** Writing – review & editing, Writing – original draft, Project administration, Investigation, Formal

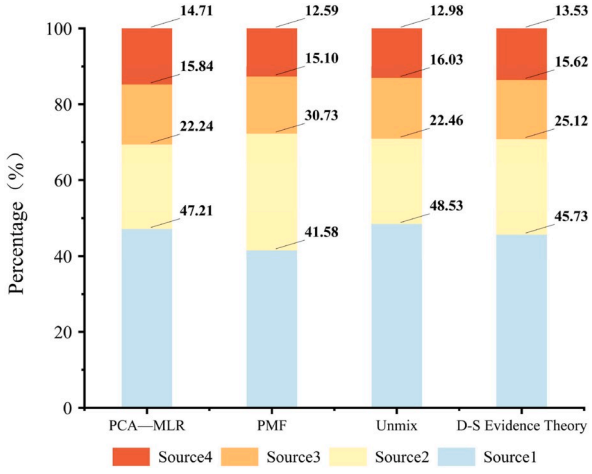


Fig. 9. Contribution of pollution sources in the study area.

analysis.

Declaration of competing interest

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

References

[1] J. Shi, D. Zhao, F. Ren, L. Huang, Spatiotemporal Variation of Soil Heavy Metals in China: the Pollution Status and Risk Assessment, 871, Science of The Total Environment, 2023.
[2] A. Parizanganeh, P. Hajisoltani, A. Zamani, Assessment of Heavy Metal Pollution in Surficial Soils Surrounding Zinc Industrial Complex in Zanjan-Iran, 2, 2010, pp. 162–166.
[3] H. Jin, P. Zhihong, Z. Jiaqing, L. Chuxuan, T. Lu, J. Jun, L. Xinghua, G. Wenyan, G. Junkang, S. Binbin, X. Shengguo, Source apportionment and quantitative risk assessment of heavy metals at an abandoned zinc smelting site based on GIS and PMF models, J. Environ. Manag. 336 (2023), 117565.
[4] C. Boente, M.T.D. Albuquerque, S. Gerassis, E. Rodríguez-Valdes, J.R. Gallego, A coupled multivariate statistics, geostatistical and machine-learning approach to address soil pollution in a prototypical Hg-mining site in a natural reserve, Chemosphere 218 (2019) 767–777.
[5] J.G. Wiederhold, A.R. Grigg, R.S. Gilli, R. Kretzschmar, Mercury stable isotope variations in soils and sediments from an industrial contamination site in Switzerland, in: EGU General Assembly Conference Abstracts, 2018, p. 6210.
[6] H. Zhou, Y. Chen, X. Yue, D. Ren, Y. Liu, K. Yang, Source apportionment of soil heavy metals: identification and hazard analysis of heavy metal sources in agricultural soils in ancient mining areas: a quantitative method based on the receptor model and risk assessment, J. Hazard Mater. 445 (2023), 130528.
[7] D. Baragano, G. Ratie, C. Sierra, V. Chrastný, M. Komarek, J.R. Gallego, Multiple pollution sources unravelled by environmental forensics techniques and multivariate statistics, J. Hazard Mater. 424 (2019), 127413.
[8] A. Martin, Y. Hassan-Loni, A. Fichtner, O. Peron, K. David, P. Chardon, An integrated approach combining soil profile, records and tree ring analysis to identify the origin of environmental contamination in a former uranium mine (Rophin, France), Sci. Total Environ. 747 (2023), 141295.
[9] X. Fei, G. Christakos, Z. Lou, R. Xiao, X. Lv, Z. Ren, Assessment and source apportionment of toxic metal in soils using integrated positive matrix factorization and Bayesian maximum entropy: a case study in Z county, southeastern China, Ecol. Indicat. 145 (2022), 109647.
[10] P. Paatero, U. Tapper, Analysis of different modes of factor analysis as least squares fit problems, Chemometrics and In telligent Laboratory Systems 18 (2) (1993) 183–194.
[11] A. Yang, Y.H. Wang, J. Hu, X.L. Liu, J. Li, Evaluation and source of heavy metal pollution in surface soil of qinghai-tibet plateau, Huanjing Kexue 41 (2) (2020) 886–894.
[12] E. Yu, H. Liu, F. Dinis, Q. Zhang, P. Jing, F. Liu, X. Ju, Contamination evaluation and source analysis of heavy metals in karst soil using UNMIX model and Pb-Cd isotopes, Int. J. Environ. Res. Publ. Health 19 (2022).
[13] Y. Huang, M. Deng, S. Wu, J. Japenga, T. Li, X. Yang, Z. He, A modified receptor model for source apportionment of heavy metal pollution in soil, J. Hazard Mater. 354 (2018) 161–169.
[14] M. Lei, K. Li, G. Guo, T. Ju, Source-specific health risks apportionment of soil potential toxicity elements combining multiple receptor models with Monte Carlo simulation, Sci. Total Environ. 817 (2022), 152899.
[15] H. Liu, S. Liao, D. Wu, G. Rukh, Z. Chen, X. Wu, L. Xiao, B. Zhong, D. Liu, Source Apportionment and Model Applicability of Heavy Metal Pollution in Farmland Soil Based on Three Receptor Models, Toxics 11 (2021).
[16] Y. Li, Z. Zhao, Y. Yuan, P. Zhu, X. Li, A. Guo, Q. Yang, Application of modified receptor model for soil heavy metal sources apportionment: a case study of an industrial city, China, Environ. Sci. Pollut. Res. Int. 26 (16) (2019) 16345–16354.
[17] Y. Wang, Y. Li, S. Yang, J. Liu, W. Zheng, J. Xu, H. Cai, X. Liu, Source apportionment of soil heavy 468 metals: a new quantitative framework coupling receptor model and stable isotopic ratios, Environ. Pollut. 26 (2022), 36174813.
[18] P.C. Agyeman, S.K. Ahado, K. John, N.M. Kebonye, Health risk assessment and the application of CF-PMF: a pollution assessment–based receptor model in an urban soil, J. Soils Sediments 21 (2021) 3117–3136.
[19] A.P. Dempster, Upper and lower probabilities induced by a multiple valued mapping, Ann. Math. Stat. 38 (2) (1967) 325–339.
[20] G. Muller, Index of geoaccumulation in sediments of the rhine river, Geol. J. 2 (1969) 108–118.
[21] N.L. Nemerow, Stream, Lake, Estuary, and Ocean Pollution, Van Nostrand Reinhold Publishing Co., New York, 1985.
[22] L. Hakanson, An ecological risk index for aquatic pollution-control - a sedimentological approach, Water Res. 14 (1980) 975–1001.
[23] H. Xie, X. Yang, J. Xu, D. Zhong, Heavy metals pollution and potential ecological health risk assessment in the Yangtze River reaches, J. Environ. Chem. Eng. 11 (2023), 109484.
[24] M. Seyedrahimi-Niaraq, H. Mahdiyanfar, A. Mokhtari, Integrating principal component analysis and U-statistics for mapping polluted areas in mining districts, J. Geochem. Explor. 234 (2022), 106924.
[25] S. Chakraborty, T. Man, L. Paulette, S. Deb, B. Li, D.C. Weindorf, M. Frazier, Rapid assessment of smelter/mining soil contamination via portable X-ray fluorescence spectrometry and indicator kriging, Geoderma 306 (2017) 108–119.
[26] A. Kolovos, A. Skupin, M. Jerrett, G. Christakos, Multi-perspective analysis and spatiotemporal mapping of air pollution monitoring data, Environ. Sci. Technol. 44 (17) (2010) 6738–6744.
[27] C.N.P.G. Arachchige, L.A. Prendergast, R.G. Staudte, Robust analogs to the coefficient of variation, J. Appl. Stat. 49 (2) (2020) 268–290.
[28] J.B. Kowalska, R. Mazurek, M. Gąsiorek, T. Zaleski, Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination-A review, Environ. Geochem. Health 40 (6) (2018) 2395–2420.
[29] B. Wang, L. Chen, Review on methods of soil quality evaluation, Soil Water Conserv. Sci. China. 4 (2) (2006) 120–126.
[30] J.L. Chen, R.Y. Li, X.J. Xie, H. Wang, J. Xu, J. Shao, J. Jian, A. Wuerman, J. Shen, Z. Yang, Distribution characteristics and pollution evaluation of heavy metals in greenbelt soils of nanjing city, Huanjing Kexue 42 (2) (2021) 909–916.
[31] L. Sang, C. Zhang, J. Yang, D. Zhu, W. Yun, Assessment of potential risk in soil and early warning analysis in four counties, northeast China, IFIP Adv. Inf. Commun. Technol. 369 (2012).
[32] M. Xiang, Y. Li, J. Yang, K. Lei, Y. Li, F. Li, D. Zheng, X. Fang, Y. Cao, Heavy metal contamination risk assessment and correlation analysis of heavy metal contents in soil and crops, Environ. Pollut. 278 (2021), 116911.
[33] R. Dragović, B. Gajić, S. Dragović, M. Đorđević, M. Đorđević, N. Mihailović, A. Onjia, Assessment of the impact of geographical factors on the spatial distribution of heavy metals in soils around the steel production facility in Smederevo (Serbia), J. Clean. Prod. 84 (2014) 550–562.
[34] Y. Gao, T. Lü, Y.K. Zhang, Source apportionment and pollution assessment of soil heavy metal pollution using PMF and RF model: a case study of a typical industrial park in Northwest China[J/OL], Environ. Sci. J. Integr. Environ. Res. 44 (6) (2023), 13227.
[35] Y. Li, Z. Ye, Y. Yu, Y. Li, J. Jiang, L. Wang, G. Wang, H. Zhang, N. Li, X. Xie, X. Cheng, K. Liu, Liu, A combined method for human health risk area identification of heavy metals in urban environments, Journal 455 of Hazardous Materials 449 (2023), 131067, 454 M.
[36] W. Wu, P. Wu, F. Yang, D.L. Sun, D.X. Zhang, Y.K. Zhou, Assessment of heavy metal pollution and human health risks in urban soils around an electronics manufacturing facility, Sci. Total Environ. 630 (2018) 53–61.
[37] S.Y. Yang, S.B. Gu, M.J. He, X.J. Tang, L.N.Q. Ma, J.M. Xu, X.M. Liu, Policy Adjustment Impacts Cd, Cu, Ni, Pb and Zn Contamination in Soils Around E-Waste Area: Concentrations, Sources and Health Risks, 741, Science of the Total Environment, 2020.

[38] L. Zhao, Y. Xu, H. Hou, Y. Shangguan, F. Li, Source identification and health risk assessment of metals in urban soils around the Tanggu chemical industrial district, Tianjin, China, Sci. Total Environ. 468–469 (2014) 654–662.
[39] H. Pan, X. Lu, K. Lei, A comprehensive analysis of heavy metals in urban road dust of Xi'an, China: contamination, source apportionment and spatial distribution, Sci. Total Environ. 609 (2017) 1361–1369.
[40] Q.Y. Guan, F.F. Wang, C.Q. Xu, N.H. Pan, J.K. Lin, R. Zhao, Y.Y. Yang, H. Luo, Source apportionment of heavy metals in agricultural soil based on PMF: a case study in Hexi Corridor, northwest China 193 (2018) 189–197.
[41] H. Chen, W. Mao, Y. Shen, W. Feng, G. Mao, T. Zhao, L. Yang, L. Yang, C. Meng, Y. Li, X. Wu, Distribution, source, and environmental risk assessment of phthalate esters (PAEs) in water, suspended particulate matter, and sediment of a typical Yangtze River Delta City, China, Environ. Sci. Pollut. Res. Int. 26 (24) (2019) 24609–24619.
[42] H. Lü, C.H. Mo, H.M. Zhao, L. Xiang, A. Katsoyiannis, Y.W. Li, Q.Y. Cai, M.H. Wong, Soil contamination and sources of phthalates and its health risk in China: a review, Environ. Res. 164 (2018) 417–429.
[43] X.Y. Liao, Z.Y. Chong, X.L. Yan, Urban industrial contaminated sites: a new issue in the field of environmental remediation in China, Environ. Sci. J. Integr. Environ. Res. 32 (3) (2011) 784–794.