Journal of Environmental Management 410 (2026) 130146

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Source apportionment and health risk prioritization of groundwater contamination in a coal-mining area: An integrated SOM-PMF-Monte Carlo framework

Zheng Yin [a], Shiqi Zhang [a,b], Xiao Fu [a,*], Ran Sun [a,b], Ran Lv [a,b], Meng Yuan [a,b], Ziyuan Gao [c], Gang Wu [a,b]

[a] State Key Laboratory of Regional and Urban Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, 100085, China
[b] University of Chinese Academy of Sciences, Beijing, 100049, China
[c] China Medical University, China Medical University-Queen's University Belfast Joint College, Shenyang, 110122, China

ARTICLE INFO

Keywords:
Groundwater contamination
Source apportionment
Health risk prioritization
Arid mining region
Environmental management

ABSTRACT

Groundwater quality in arid/semi-arid coal-mining areas is increasingly threatened by intensive human activities. However, quantitatively linking specific pollution sources to health risks and translating this into risk-based management priorities remains a critical challenge. This study addresses this gap by proposing an integrated framework combining self-organizing maps (SOM), positive matrix factorization (PMF), and Monte Carlo simulation to systematically decipher groundwater contamination sources and prioritize associated health risks, using the Huolingol mining area in Inner Mongolia, China, as a representative case. SOM analysis recognized three hydrochemical clusters: Cluster 1 (Na-HCO_3/Ca-Na-HCO_3 type), influenced by cation exchange and silicate weathering under alkaline conditions, exhibited elevated F^- and NH_4^+-N concentrations. Cluster 2 (Ca-Na-SO_4/Ca-Na-HCO_3 type) was characterized by elevated TDS, TH, SO_4^{2-}, and NO_3^-, indicating significant anthropogenic impact. Cluster 3 (Ca-HCO_3 type) represented the natural background with minimal contamination. Source apportionment using PMF revealed that human-driven composite pollution is the main cause of water quality evolution, with the mixed source of coal mine drainage and industrial/domestic wastewater contributing the highest proportion (22.42%), highlighting the overwhelming impact of mining and industrial activities. Monte Carlo simulation showed children faced a 33.9% probability of exceeding the safe health threshold (HI > 1), 1.75 times higher than adults (19.4%). F^- and SO_4^{2-} were the main risk drivers. Source-specific analysis indicated geogenic fluoride (Factor 1) contributes ~48% of total health risk, followed by organic degradation (~26%) and anthropogenic sulfate (~23%). This framework quantitatively links sources to health risks and provides actionable guidance for groundwater management in water-stressed mining regions.

1. Introduction

Groundwater is often the primary water source in arid and semi-arid regions. It supports drinking, irrigation, and industrial use, playing a key role in socioeconomic stability and ecological security (Foster and Chilton, 2003; Ghazavi et al., 2012; Qu et al., 2023; Wu et al., 2021). However, with the large-scale exploitation of coal resources and the rapid advancement of regional industrialization, groundwater systems in mining areas and their surrounding regions are now facing unprecedented environmental pressures (Ju et al., 2023; Younger, 2000). Intensive human activities—including mine drainage, wastewater discharge, and agricultural practices—have introduced multiple contaminants into aquifers. These activities have significantly altered groundwater quality and posed persistent threats to environmental safety and human health (Banks et al., 1997; Su et al., 2026; Yan et al., 2025). Against this background, systematically revealing the chemical composition and evolution patterns of groundwater, quantitatively identifying the sources of groundwater contamination and scientifically assessing the associated human health risks have become critical scientific issues. These efforts are urgently needed to balance coal mining with sustainable water use and support ecological restoration and water management policies in arid mining areas (Cui et al., 2025).

* Corresponding author.
E-mail address: xiaofu@rcees.ac.cn (X. Fu).

https://doi.org/10.1016/j.jenvman.2026.130146
Received 4 April 2026; Received in revised form 20 May 2026; Accepted 3 June 2026
Available online 6 June 2026

Traditional groundwater quality assessments mainly rely on hydrochemical facies analysis (e.g., Piper diagrams, Gibbs plots) and ionic ratio calculations. These methods effectively identify major geochemical processes but lack quantitative source identification (Yan et al., 2025). In recent years, machine learning methods have demonstrated substantial value across diverse domains, including resource market analysis and environmental management (Jin and Xu, 2025a, 2025b), as well as building energy prediction and carbon emission modeling (Naghipour, 2026;Naghipour et al., 2026a;Naghipour et al., 2026b; Naghipour et al., 2026c;Naghipour and Naghipour, 2025a;Naghipour and Naghipour, 2025b). Among these methods, self-organizing maps (SOM) are particularly well-suited for analyzing high-dimensional environmental data, enabling objective classification of groundwater samples based on multi-parameter similarity and overcoming the subjectivity of traditional graphical methods (Santos et al., 2020; Zhong et al., 2022). Comprehensive water quality evaluation models (e.g., Entropy weight-technique for order of preference by similarity to ideal solution (EW-TOPSIS)) achieve ranking of groundwater quality by integrating multi-indicator weighting and approximation to the ideal solution, providing a scientific basis for water resource management (Dehghan Rahimabadi et al., 2023; Wang et al., 2023). Receptor models such as Positive Matrix Factorization (PMF) have emerged as powerful tools for apportioning contamination sources by decomposing chemical data into factor contributions and profiles (Cui et al., 2025; Li et al., 2021). Meanwhile, probabilistic risk assessment frameworks incorporating Monte Carlo simulation have gained traction for their ability to address parameter uncertainties and provide more realistic characterizations of human health risks (Chen et al., 2023; Wu et al., 2024). In the broader context mentioned above, the integration of unsupervised learning (e.g., SOM) with receptor modeling (PMF) and health risk assessment (Monte Carlo simulation) offers a promising avenue for evidence-based environmental governance in resource-dependent regions. By linking contamination sources to health outcomes, such integrated frameworks can guide the equitable allocation of limited resources for pollution control and public health protection. This aligns with key principles of sustainable resource management and circular economy frameworks (Jin and Xu, 2024; Xu and Zhang, 2023).

The Huolingol mining area in Inner Mongolia, China, represents a typical coal-resource-based industrial region, with an annual output of 35 million tons of coal and 2.7 million tons of primary aluminum (Huolingol Municipal People's Government, 2024). Located in an arid/semi-arid climate zone with mean annual precipitation of only 354.7 mm and evaporation of 1732.7 mm, the region relies almost exclusively on groundwater for domestic, agricultural, and industrial water supply (Wang, 2015). However, intensive coal mining and associated industrial activities have exerted mounting pressures on the groundwater system. Previous investigations have reported elevated concentrations of SO_4^{2-}, NO_3^-, F^-, and NH_4^+-N in local groundwater, with some parameters exceeding the Class III standards of the Chinese Groundwater Quality Standard (GB/T 14,848-2017) (Wang, 2015; Zhang, 2016). Despite these observations, a comprehensive understanding of the contamination sources, their quantitative contributions, and the associated human health risks remains lacking. In particular, no study has systematically linked the spatial heterogeneity of groundwater chemistry to specific pollution sources and further translated this knowledge into health risk characterization and management prioritization.

To address this gap, the present study developed an integrated framework combining self-organizing maps (SOM), positive matrix factorization (PMF), and Monte Carlo simulation to: (i) classify groundwater samples based on multi-parameter hydrochemical signatures; (ii) quantitatively apportion contamination sources and their contributions; (iii) characterize probabilistic health risks for different populations and identify priority pollutants; and (iv) quantify each source's contribution to the health risk for targeted management. The novelty of this study lies in the end-to-end integration of SOM, PMF, and Monte Carlo to generate management-oriented outputs (priority pollutants, vulnerable populations), rather than in the individual methods themselves. The Huolingol mining area was selected as a representative case due to its typical arid climate, intensive coal-mining activities, and documented evidence of groundwater contamination. This study provides not only a scientific basis for local groundwater management but also a methodologically transferable framework for other water-stressed mining regions facing similar contamination challenges.

2. Study area

2.1. Study area

The study focused on the Huolingol mining area, a critical coal-producing region in Northeast China. This area encompasses Huolingol City and the adjacent coal-mining districts in northern Jarud Banner (Fig. 1). Its core, Huolingol City (coordinates: 45°16′-45°46′ N, 118°17′46″-119°46′12″ E; total area: 585 km^2), is situated on the western side of the Huolin River Basin, along the western margin of the southern Greater Khingan Mountains. The terrain exhibits an elongated northeast-southwest orientation, characterized by surrounding low to medium mountains, with ridged plateaus and broad valley depressions in the central part. Elevation ranges from 850 to 1291 m above sea level, generally descending from the southeast to the northwest. The region experiences a typical mid-temperate semi-arid continental monsoon climate, characterized by dry and windy springs, long and severely cold winters, cool and short summers, and autumns with early frost. Mean annual temperature is approximately 0.84 °C. Annual precipitation averages 354.7 mm, with about 80% concentrated in July–September. In contrast, annual evaporation is considerably higher, reaching 1732.7 mm. Major land use types include cultivated land, grassland, forestland, built-up areas, water bodies, wetland, and bare land (Wang, 2015).

2.2. Geology and hydrogeology

The stratigraphy belongs to the Greater Khingan–Yanshan division, specifically the Boketu–Erenhot subdistrict. Exposed strata are relatively simple, mainly including Mesozoic Jurassic formations (Hongqi (J_1h), Tuchengzi (J_2t), Manketouebo (J_3mk), Manitou(J_3mn), Baiyingaolao (J_3b)), Cretaceous Huolin River Formation (K_1d), and the Cenozoic Quaternary System (Q). Jurassic strata are widely distributed and comprise clastic sedimentary rocks (such as feldspathic quartz siltstone), volcanic rocks, and pyroclastic rocks (e.g., andesite, rhyolite, and tuff). The Cretaceous Huolin River Formation (K_1d) consists of coal-bearing clastic rocks, with lithologies including sandstone, mudstone, and conglomerate. The Quaternary and Neogene systems are mainly composed of loose sand-gravel layers and basalt.

The groundwater system is divided into three fourth-level hydrogeological units: Sharhure, Ulgai River, and Dundnuur. Based on occurrence and hydraulic properties, aquifers are classified into five systems: (1) the Quaternary unconsolidated porous and bedrock weathered-fractured porous phreatic aquifer system (Q + J); (2) the Quaternary unconsolidated porous and clastic rock porous-fractured phreatic aquifer system (Q + K); (3) the clastic rock porous-fractured phreatic aquifer system (K); (4) the bedrock fractured phreatic aquifer system (J); and (5) the clastic rock porous-fractured confined aquifer system (K). The two Quaternary unconsolidated porous-fractured aquifer systems are the most widespread and have the best water yield (max single-well yield 1000–3000 m^3/d). They are shallow (<30 m) and serve as the main water-extraction targets.

Groundwater recharge mainly originates from atmospheric precipitation infiltration. In the surrounding low to medium mountain areas with exposed bedrock, precipitation rapidly infiltrates through rock fractures to form bedrock fissure water, which discharges into valleys and basins. Natural groundwater flow varies among units: Sharhure

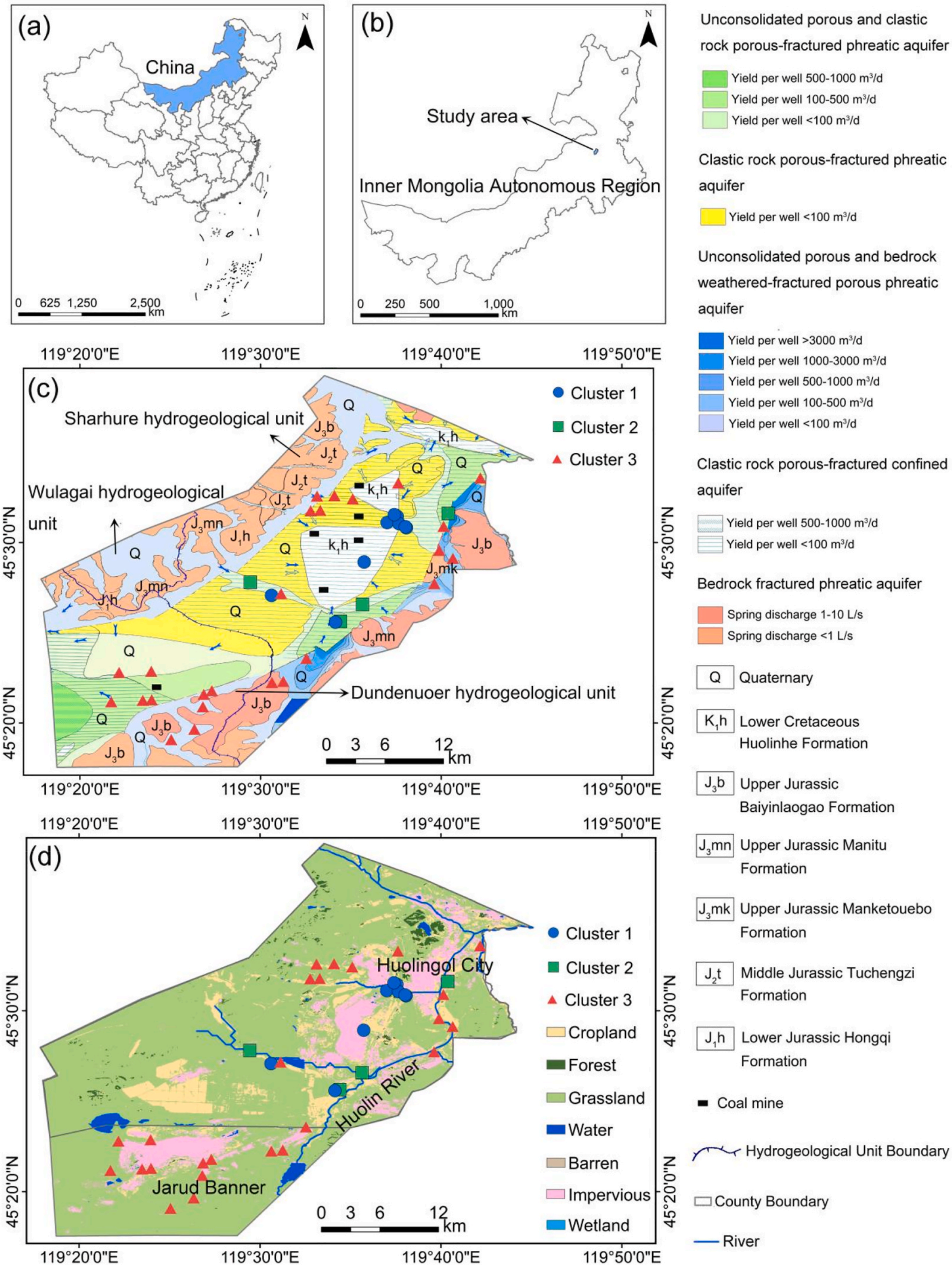

Fig. 1. (a) Location of Inner Mongolia in China, (b) location of study area in Inner Mongolia, (c) hydrogeological map and groundwater sampling points in the study area, (d) land use map. Well depths of the sampling points range from 20 to 120 m; detailed location coordinates and well depth information for each sample are provided in Table S1.

flows southwest–northeast; Dundnuur converges south–north (or southeast–northwest). In the unconsolidated and clastic rock aquifers within river valleys and basins, groundwater also receives lateral runoff recharge from the aforementioned mountainous areas and is influenced by seasonal floods and surface water infiltration. However, intensive human activities (groundwater extraction, open-pit mine dewatering) have formed a regional groundwater cone of depression in the Sharhure unit. This has significantly altered the natural flow field, causing groundwater to converge toward extraction and dewatering centers. Artificial extraction and mine dewatering have become the main groundwater discharge pathways (Wang, 2015).

The study area has 14 coal mining districts with a total annual capacity of about 35 million tons. Most are open-pit mines, with coal hosted in the fourth member of the Lower Cretaceous Huolin River Formation. The coal is high-quality aged lignite, low-sulfur ($<0.5\%$) and low-phosphorus ($<0.02\%$), with average calorific value $>$ 3100 kcal/kg.

3. Materials and methods

3.1. Sampling and analysis

A total of 38 groundwater samples were collected from the Huolingol mining area between May 12 and 29, 2025, with 25 samples located in Huolingol City and the other 13 situated in the northern part of the adjacent Jarud Banner (Fig. 1c and d). All samples were phreatic water obtained from agricultural irrigation wells, domestic drinking water wells, and coal mine monitoring wells, with well depths ranging from 20 m to 120 m. The water table in unconfined aquifers ranges from about 2 to 25 m (Wang, 2015), ensuring all wells fully penetrate the saturated zone. Sampling points were evenly distributed, ensuring good regional representativeness. Before sampling, each well was purged for approximately 5 min to eliminate the influence of stagnant water in the well. Samples were filtered through 0.45 μm membranes and stored in pre-cleaned polyethylene bottles rinsed 2∼3 times with local groundwater. Triplicate samples were collected at each site. Sample preservation, transport, and pretreatment strictly followed national specifications. All samples were ultimately sent to Inner Mongolia Xin'antai Testing and Evaluation Technology Co., Ltd. for physicochemical analysis.

Analyzed parameters included pH, major cations (Mg^{2+}, Ca^{2+}, Na^+, K^+), major anions (HCO_3^-, Cl^-, SO_4^{2-}), permanganate index (COD_{Mn}), ammonia nitrogen (NH_4^+-N), nitrate (NO_3^-), total dissolved solids (TDS), total hardness (TH), and fluoride (F^-). pH was measured on-site using a portable pH meter. COD_{Mn} was determined by the acidic potassium permanganate titration method. NH_4^+-N was analyzed using ultraviolet-visible spectrophotometry (instrument model UV-9600, serial number 03005). TDS was measured gravimetrically. TH was determined via EDTA-2Na titration. Cations were analyzed by atomic absorption spectrophotometry (instrument model WFX-120B, serial number 03006). HCO_3^- was measured by acid-base indicator titration. The remaining anions (Cl^-, SO_4^{2-}, NO_3^-, F^-) were analyzed using ion chromatography (instrument model CIC-100, serial number 03117). Sulfate sulfur and oxygen isotopes ($\delta^{34}S$-SO_4^{2-} and $\delta^{18}O$-SO_4^{2-}) were analyzed using a stable isotope ratio mass spectrometer (DeltaV Plus) after sulfate precipitation as $BaSO_4$. Analytical precision was ±0.2‰ for $\delta^{34}S$ and ±0.3‰ for $\delta^{18}O$. Ion charge balance errors were below 5% for all samples, indicating reliable data. Heavy metals (As, Cd, Pb, Zn, Fe, Mn, Hg) were mostly below detection limits or far below drinking water standards. They were excluded due to negligible health risks.

3.2. Self-organizing map (SOM)

Groundwater samples were classified using a hybrid approach combining Self-Organizing Map (SOM) and K-means clustering. SOM projects high-dimensional hydrochemical data onto a two-dimensional neuron grid, preserving topological relationships while enabling nonlinear dimensionality reduction (Qu et al., 2022; Appukuttan et al., 2025). The optimal map size was determined based on the quantization error (QE) and topographic error (TE). K-means clustering was then applied to the SOM neuron weights, with the optimal number of clusters determined by the silhouette coefficient and Davies-Bouldin Index (Mao et al., 2021). The analysis was implemented in R.

3.3. Spatial interpolation of hydrochemical parameters

Spatial distribution maps of 14 hydrochemical parameters were generated using inverse distance weighting (IDW, power = 2, four-sector neighborhood) in ArcGIS 10.8 Geostatistical Analyst. Leave-one-out cross-validation (LOOCV) was conducted for each parameter to evaluate performance. Cross-validation results (mean error ME, root mean square error RMSE) are summarized in Table S2.

3.4. Entropy weight-TOPSIS model

Groundwater quality was assessed using the entropy weight-technique for order of preference by similarity to ideal solution (EW-TOPSIS) method, which combines objective entropy weighting with TOPSIS ranking (Chen, 2021). To anchor the evaluation to regulatory benchmarks, virtual standard water samples derived from the Class I-IV thresholds of the Chinese Groundwater Quality Standard (GB/T 14, 848-2017) were incorporated into the decision matrix (detailed thresholds are listed in Table S3, and the corresponding virtual sample parameters are provided in Table S4). Entropy weights were calculated as detailed in the Supplementary Material (Table S5). This approach shifts the evaluation from purely relative comparison to standard-based classification.

3.5. Positive matrix factorization (PMF) model

Source apportionment was performed using Positive Matrix Factorization (PMF, EPA PMF 5.0; Norris et al., 2014), a receptor model that decomposes a measured data matrix into factor contributions and profiles under non-negativity constraints (Paatero and Tapper, 1994). The model minimizes an objective function Q based on the measured concentrations and their associated uncertainties, which were calculated from the minimum detection limit and relative standard deviation (detailed equations provided in Supplementary Material). Input data included 12 chemical parameters (TDS, TH, COD_{Mn}, NH_4^+-N, NO_3^-, Cl^-, HCO_3^-, F^-, Na^+, K^+, Ca^{2+}, Mg^{2+}) from all 38 sampling sites. Models were run with 4–6 factors and multiple random seeds. The optimal solution was selected based on residual analysis, Q value, and geochemical interpretability.

3.6. Monte Carlo-based human health risk prioritization

Non-carcinogenic health risks from F^-, NO_3^-, and SO_4^{2-} were assessed following the USEPA exposure model (USEPA, 2004), considering ingestion and dermal contact pathways for adults and children. The inclusion of SO_4^{2-} as a health risk driver is justified by its potential health effects. High SO_4^{2-} in drinking water may cause gastrointestinal effects, especially in children (Chien et al., 1968; USEPA, 2004). A reference dose (RfD) of 8.2 mg/kg/day was adopted from Su et al. (2026).

Monte Carlo simulation (10,000 iterations) was performed using Crystal Ball® software to address parameter uncertainties. Input parameters were assigned probabilistic distributions based on published literature (Chen et al., 2023; Su et al., 2026; Yan et al., 2025): ingestion rate (IR) and body weight (BW) followed normal distributions (children: IR 1.20 ± 0.12 L/d, BW 26.8 ± 2.68 kg; adults: IR 2.50 ± 0.25 L/d, BW 70.0 ± 7.00 kg); exposure frequency (365 d/y), duration (6 y for children, 30 y for adults), and averaging time (2190 d for children, 10,950 d for adults) were point estimates. All exposure parameters were assumed independent, which is conservative and widely adopted in similar assessments (Chen et al., 2023; Su et al., 2026). Parameters and RfD are summarized in Table S13. In each iteration, hazard quotients (HQs) were calculated for each contaminant and summed to obtain a probabilistic distribution of the hazard index (HI) (Eqs. S20–S23). Sensitivity analysis identified the most influential parameters.

Deterministic HI values were also calculated using mean exposure parameters and the measured contaminant concentrations at each sampling point. To link PMF sources to health risks, factor contributions to HI were estimated from the PMF contribution matrix and source profiles (Tables S11 and S12), following the method described by (Chen et al., 2023) (Eqs. S24–S26). Spatial distribution of deterministic HI values was generated using IDW interpolation (power = 2, four-sector neighborhood; Section 3.3), with LOOCV using adult HI data.

4. Results

4.1. Hydrochemical classification

Considering the quantization error (QE = 0.99) and topological error (TE = 0.104), the SOM topology was ultimately defined as 5×5, with a learning rate of 0.1 and 1000 training iterations. Silhouette coefficient analysis indicated that the peak value (0.28) was achieved when the number of clusters was set to 3. Consequently, the groundwater samples were classified into three clusters using the self-organizing map combined with K-means clustering: Cluster 1 comprises 9 samples, Cluster 2 comprises 4 samples, and Cluster 3 comprises 25 samples (Fig. 2).

Spatially, the three clusters exhibited distinct hydrogeological characteristics (Fig. 1c): Cluster 1 primarily occurred in clastic rock porous-fractured phreatic aquifers, distributed around the depression cone formed by dewatering from open-pit coal mines, representing a typical slow-flow or stagnant runoff zone. Cluster 2 was concentrated in unconsolidated porous and clastic rock porous-fractured phreatic aquifers, mainly located in river confluence zones, corresponding to a typical groundwater stagnation zone. Cluster 3 predominantly occurred in unconsolidated porous and bedrock weathered-fractured porous phreatic aquifers, distributed in the peripheral areas of the sampling network and adjacent to bedrock fractured aquifers, reflecting the hydrogeochemical characteristics of recharge or near-recharge zones.

Piper diagram analysis (Fig. 3) revealed clear hydrochemical differences. Cluster 1 was dominated by Na-HCO_3 and mixed Ca-Na-HCO_3 types, with cations primarily dominated by Na^+ (though Ca^{2+} and Na^+ are comparable in some samples) and anions centered on HCO_3^-. Cluster 2 consisted of mixed Ca-Na-SO_4 and Ca-Na-HCO_3 types, with both Ca^{2+} and Na^+ co-dominating the cation composition and anions including both SO_4^{2-} and HCO_3^-. Cluster 3 exhibited a typical Ca-HCO_3 type, with cations strongly enriched in Ca^{2+} and anions dominated by HCO_3^-. SOM integrates multi-indicator data (pH, TDS, TH, major ions), while Piper diagrams only use major ions. Thus, samples in the same SOM cluster may show different Piper facies, highlighting the advantage of SOM for comprehensive characterization.

4.2. *Contaminant characteristics and spatial distribution*

Fig. 4 presents box plots of hydrochemical variables for the three groundwater clusters; Fig. 5 shows their spatial distributions. LOOCV confirms reliable IDW interpolation: the mean errors (ME) relative to mean concentrations are small for most parameters (e.g., TDS: 3.62%, TH: 0.3%, Na^+: 0.15%), indicating negligible systematic bias for these analytes. Cl^- and NH_4^+-N exhibit slightly higher relative ME values (16.5% and 18.1%, respectively), which is expected given their high spatial heterogeneity (Table S2). For these two parameters, a modest positive bias (overestimation) exists, but the interpolation remains useful for visualizing spatial trends. The root mean square errors (RMSE) vary across parameters, reflecting differences in spatial variability and concentration ranges; parameters with higher spatial heterogeneity exhibit larger RMSE, as expected. These validation metrics collectively support the use of IDW interpolation for visualizing spatial trends in this dataset.

Cluster 1 (central-northeastern slow-flow zones) had significantly higher NH_4^+-N, F^-, HCO_3^-, Na^+, and K^+ than the other clusters (Fig. 4). The elevated Na^+, K^+, and HCO_3^- reflect intense silicate weathering and cation exchange under sluggish flow, creating an alkaline, Na-rich environment that promotes F^- desorption and inhibits its precipitation. Consequently, this cluster showed the highest F^- concentrations among all samples, with 44% exceeding the Class III standard (1.0 mg/L), indicating a geogenic fluoride contamination issue exacerbated by natural hydrodynamic conditions. Spatially, high F^- was concentrated in the central-north (Fig. 5h), consistent with Cluster 1 distribution. Widespread NH_4^+-N exceedances (67% above the Class III standard of 0.50 mg/L) suggest anthropogenic inputs (domestic wastewater, agricultural non-point sources).

Cluster 2 (central-eastern stagnation zones) had the highest concentrations of TDS, TH, COD_{Mn}, NO_3^-, SO_4^{2-}, Ca^{2+} and Mg^{2+} (Fig. 4). Mean TDS (1009.25 ± 246.71 mg/L) and TH (611.75 ± 135.32 mg/L) both exceeded the Class III standards, classifying this groundwater as highly mineralized and hard. SO_4^{2-} emerged as a key contaminant in this cluster, with all samples showing elevated sulfate, probably related to anthropogenic acid inputs from coal mine drainage and industrial wastewater. Correspondingly, the concentrations of TDS, TH, and SO_4^{2-} in the central-eastern part of the study area were significantly higher

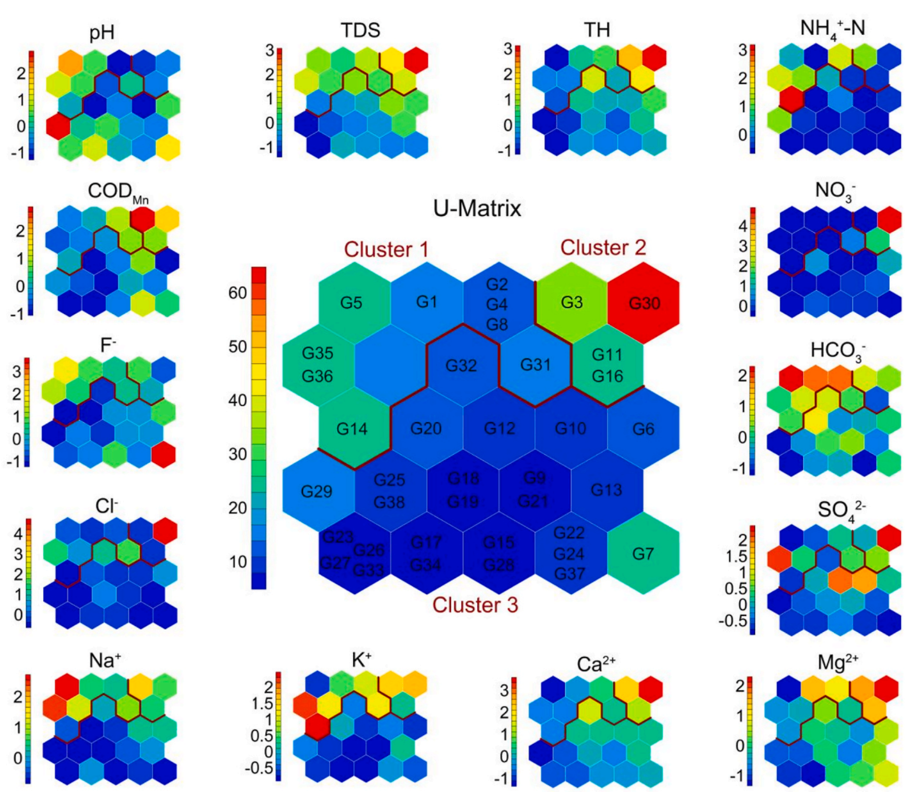

Fig. 2. Self-organizing maps of 14 variables and the corresponding U-Matrix plot.

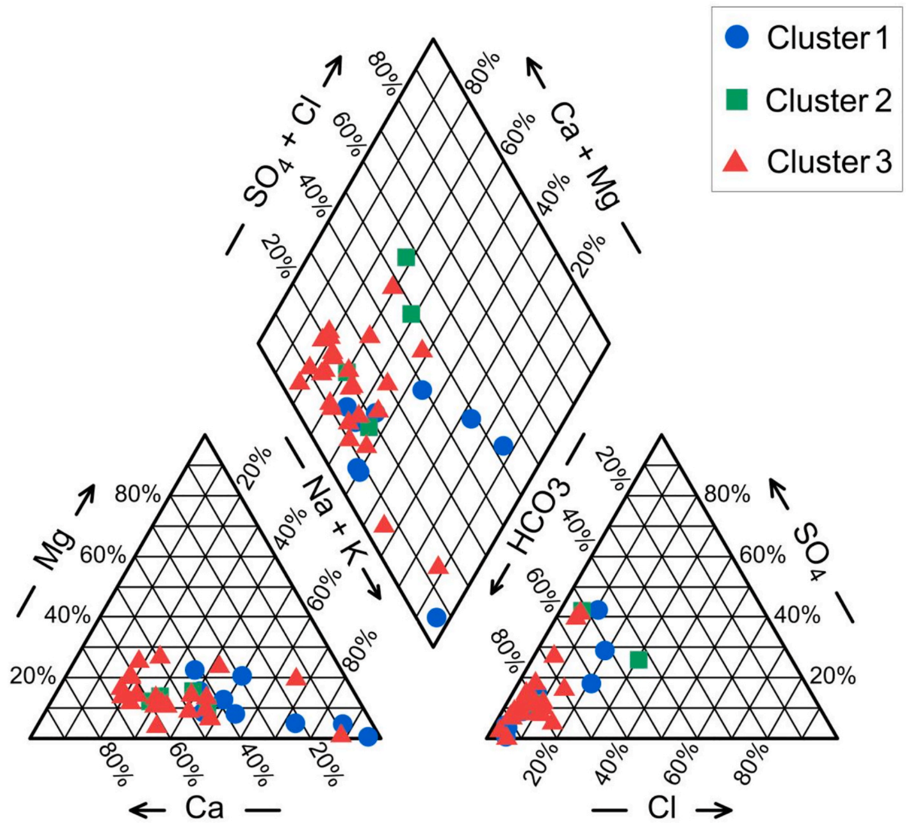

Fig. 3. Piper diagram of hydrochemical types for three groundwater clusters.

than those in other regions (Fig. 5b, c, i). Nitrate contamination was also prevalent: 50% of samples exceeded the Class III standard for NO_3^-, with high-value areas aligning closely with villages characterized by intensive agricultural cultivation (Fig. 5f). Stagnant conditions slow pollutant migration and promote accumulation of NO_3^- and SO_4^{2-}, making this area a mixed contamination hotspot.

Cluster 3 (eastern, western, southern peripheral recharge zones) had the lowest concentrations across all parameters (Fig. 4). All samples met the Class III standards for all measured indicators, reflecting minimal anthropogenic influence. This area is far from mining, industrial, and agricultural activities, with rapid groundwater renewal that limits contaminant accumulation. Cluster 3 represents the natural background and serves as a reference for contamination assessment.

In summary, groundwater contamination in the study area exhibits clear spatial heterogeneity: geogenic fluoride contamination dominates in Cluster 1 (slow-flow zone), while anthropogenically sourced sulfate and nitrate contamination prevails in Cluster 2 (stagnation zone adjacent to industrial and agricultural activities). Cluster 3 retains natural background quality. These patterns provide a spatial basis for subsequent source apportionment and health risk assessment.

4.3. Groundwater quality evaluation

Based on the EW-TOPSIS model calculation, the five-level classification thresholds for groundwater quality were determined using the relative closeness (Ci) of four virtual groundwater samples (Table S6). The specific classification is as follows: Ci $\geq$ 0.8347 corresponds to "Class I (Excellent)"; 0.7424 $\leq$ Ci $<$ 0.8347 to "Class II (Good)"; 0.6028 $\leq$ Ci $<$ 0.7424 to "Class III (Moderate)"; 0.1934 $\leq$ Ci $<$ 0.6028 to "Class IV (Poor)"; and Ci $<$ 0.1934 to "Class V (Very Poor)". The quality categories and Ci values of groundwater samples for each cluster are presented in Fig. S1 and Table S6. The evaluation results indicate that 84.2% of the groundwater samples in the study area were rated as "Moderate" or higher, reflecting generally good overall water quality that was suitable or could be made suitable for drinking after appropriate treatment. The remaining 15.8% of the samples were classified as "Poor" (Class IV), all of which fell within Cluster 1 and Cluster 2. In contrast, all samples in Cluster 3 attained "Moderate" or higher quality categories, demonstrating significantly better water quality compared to Cluster 1 and Cluster 2.

5. Discussions

5.1. Hydrochemical processes determining major ions

Gibbs diagrams (Gibbs, 1970) were used to identify dominant controls (evaporation, rock weathering, precipitation) (Fig. 6a and b). Most samples fell in the rock weathering zone, indicating that water-rock interactions controlled hydrochemical composition. A few Cluster 1 and 3 samples deviated toward the right, suggesting cation exchange or minor anthropogenic inputs. Gaillardet diagrams (Gaillardet et al., 1999) (Fig. 6c and d) identified dominant weathering minerals. All samples clustered near the silicate endmember, especially Clusters 1 and 3. This indicates that silicate mineral weathering was the primary process controlling groundwater chemistry, consistent with the regional geology of widespread silicate rocks (volcaniclastic, clastic rocks).

To further identify the sources and controlling processes of groundwater chemical components, a series of ionic ratio diagrams were employed (Fig. 6e–h).

The Na^+/Cl^- ratio diagram (Fig. 6e) shows that most samples have Na^+/Cl^- ratios well above 1 (1.5–22.7), ruling out halite dissolution as the primary sodium source. Cluster 1 (around mine dewatering cone) had higher ratios, showing strong Na^+ enrichment. To further distinguish whether Na^+ enrichment is attributed to silicate weathering or cation exchange, the chloro-alkaline index (CAI-I) was quantitatively evaluated (Fig. S2). The results showed that Cluster 1 had strongly

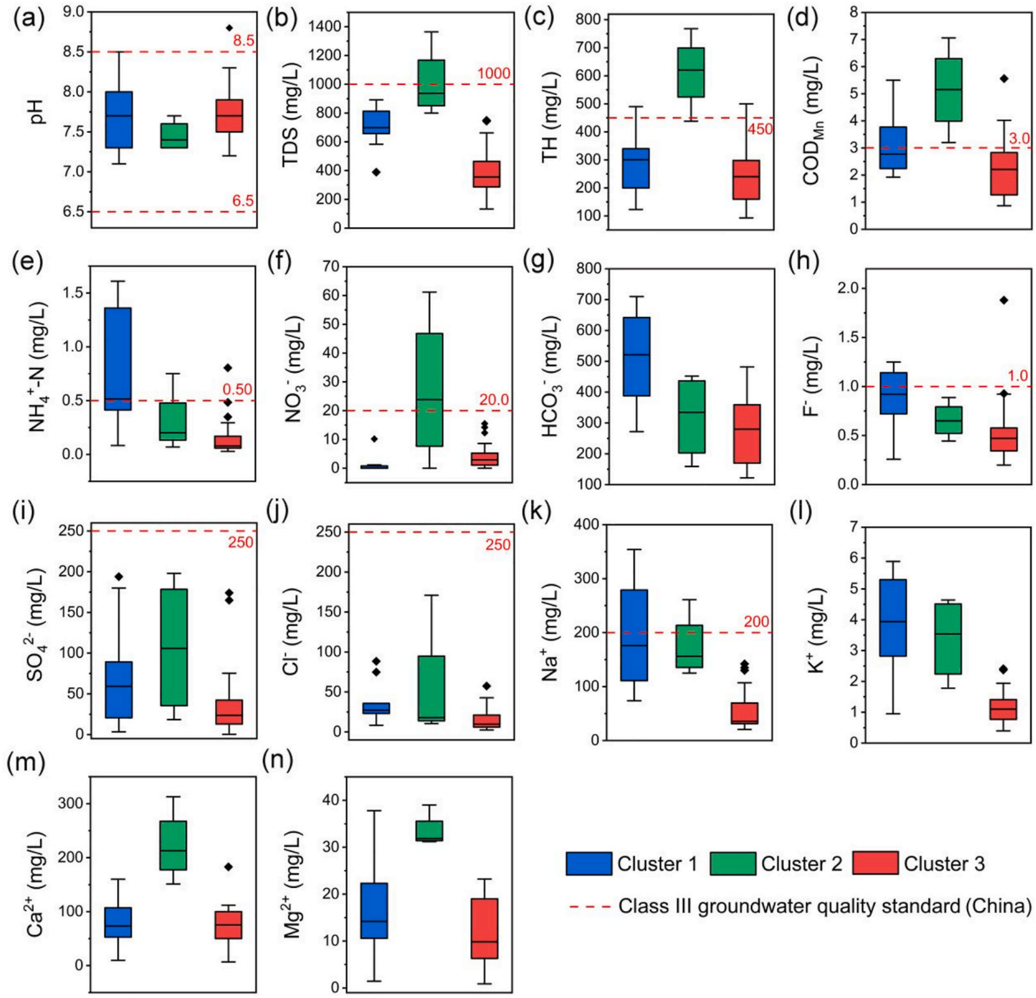

Fig. 4. (a)-(j) Box plots of hydrochemical variables for three groundwater clusters and their permissible limits. Note: The permissible limits refer to the Class III standards of the "Groundwater Quality Standards" (GB/T 14,848-2017).

negative CAI-I, confirming intense forward Na^+-Ca^{2+} cation exchange (Fan et al., 2024). This caused Ca^{2+} and Mg^{2+} depletion and Na^+ enrichment, creating favorable conditions for F^- enrichment (high pH, Na-rich environment promotes F^- desorption). This was related to the flat hydraulic gradient and sluggish flow around the dewatering cone (Zhang, 2016). Clusters 2 and 3 had weaker cation exchange, with excess Na^+ mainly from silicate weathering (albite in pyroclastic rocks).

The Ca^{2+}/SO_4^{2-} ratio diagram (Fig. 6f) shows that most samples have $Ca^{2+}/SO_4^{2-} > 1$ and deviate from the 1:1 line typical of gypsum dissolution, indicating gypsum dissolution was not dominant. PHREEQC saturation indices confirmed all samples undersaturated with gypsum (SI < 0) (Table S7), further ruling out gypsum dissolution as a significant SO_4^{2-} source. This finding aligned with previous studies reporting that the strata in this region were almost devoid of gypsum (Wang, 2015; Zhang, 2016). Most samples exhibited relative excess of Ca^{2+}, which may originated from silicate weathering and/or the dissolution of calcium-silicate minerals driven by external acidic inputs (Zhang et al., 2023). A few samples in Cluster 1 showed elevated SO_4^{2-} content. Given the limited distribution of sulfide minerals in the study area (Wang, 2015; Zhang, 2016), these SO_4^{2-} were likely derived from anthropogenic activities, such as coal mine drainage or industrial wastewater discharge (Yang et al., 2024).

The Ca^{2+}/HCO_3^- ratio diagram (Fig. 6g) further revealed distinct patterns of water-rock interactions. Most Cluster 3 and some Cluster 1 samples aligned with the 1:1 line. With no carbonate strata (Wang, 2015; Zhang, 2016), this reflects silicate weathering (e.g., anorthite) (Raju et al., 2009). Some Cluster 1 and 3 samples plotted above the line ($Ca^{2+}/HCO_3^- < 1$), indicating Ca^{2+} adsorption via cation exchange. All Cluster 2 samples plotted below the line ($Ca^{2+}/HCO_3^- > 1$). Combined with the weak influence of gypsum dissolution (Fig. 6f), this pattern points to external acidic inputs (e.g., anthropogenic sulfuric acid) dissolving calcium-bearing silicate minerals. This process releases substantial Ca^{2+} and Mg^{2+} while introducing SO_4^{2-} without a corresponding increase in HCO_3^- (Li et al., 2008; Nordstrom et al., 2015), causing high hardness and SO_4^{2-} in Cluster 2.

Dual isotopes ($\delta^{34}S$-SO_4^{2-} and $\delta^{18}O$-SO_4^{2-}) were used to trace sulfate sources in groundwater across the three clusters (Fig. 6h). The isotopic ranges for different sulfate sources (coal mine drainage, atmospheric deposition, fertilizer, sewage, and evaporite dissolution) were adapted from literature (Cui et al., 2025; Su et al., 2026). Two Cluster 2 samples fell in the coal mine drainage field, indicating sulfate from oxidation of sulfur-bearing compounds in coal seams mobilized by mining activities; the other two fell in the sewage field. Cluster 3 samples clustered near atmospheric deposition, with some overlapping the fertilizer and coal mine drainage fields. Cluster 1 samples spanned mine drainage to higher isotopic values, indicating mixing of mining sulfate with agricultural/domestic sources.

Nitrate sources were further distinguished using the Cl^- vs. NO_3^-/Cl^- diagram (Fig. S3). Cluster 1 had low NO_3^-/Cl^- and low Cl^-, showing no significant nitrate pollution. In Cluster 2, a "low Cl^-, high NO_3^-/Cl^-" signature indicated agricultural fertilizer pollution, while a "high Cl^-, moderate NO_3^-/Cl^-" pattern suggested sewage/manure contamination

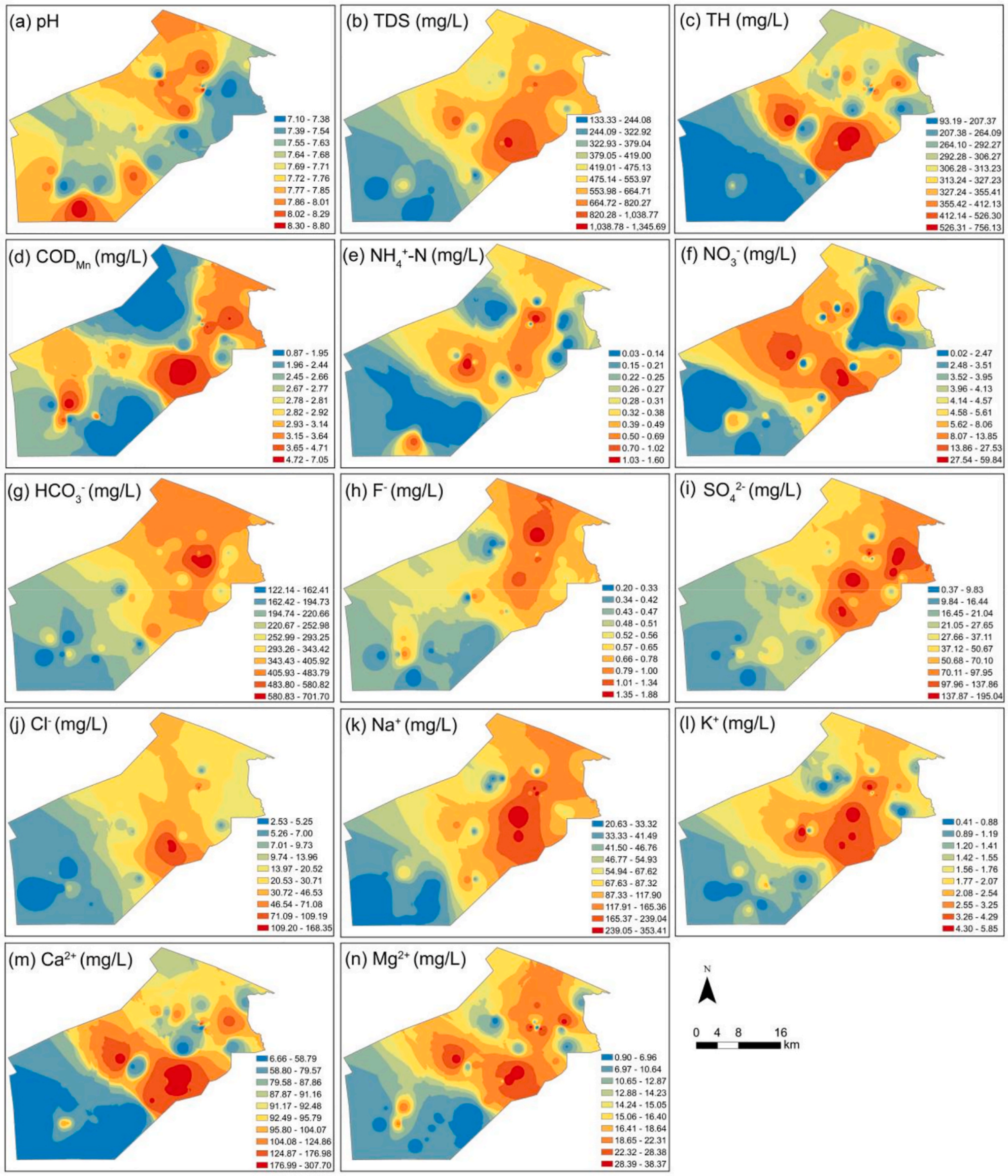

Fig. 5. (a)-(n) Spatial distribution of the hydrochemical variables of groundwater in the study area.

(Su et al., 2026). Cluster 3 had low ratios, reflecting natural background with minor anthropogenic influence.

Integrating the above graphical analyses, the dominant hydrogeochemical processes governing the three groundwater clusters in the study area could be clearly delineated: Cluster 1 was primarily dominated by intense forward cation exchange, with silicate weathering providing the material basis. This process led to severe depletion of Ca^{2+} and Mg^{2+}, significant enrichment of Na^+, and the formation of typical Na-HCO_3 type water; Cluster 2 was mainly driven by anthropogenic pollution, exhibiting characteristics typical of human-altered systems. External acidic inputs promoted strong dissolution of silicate minerals, releasing large amounts of cations such as Ca^{2+}, while introducing substantial SO_4^{2-} and Cl^-, collectively forming a highly mineralized mixed-pollution water body; Cluster 3 represented the regional natural background hydrochemical signature, controlled primarily by silicate weathering of Jurassic volcaniclastic rocks. It exhibited low mineralization and reflected the favorable hydrogeochemical environment of recharge or near-recharge zones. These process insights support source identification and risk assessment of priority pollutants (F^- in Cluster 1; SO_4^{2-}/NO_3^- in Cluster 2).

5.2. *Source identification and apportionment*

PMF analysis identified six sources as the optimal solution. The model fit well: residuals mostly within -3 to 3, Q(robust)/Q(true) $= 0.80$. R^2 for 12 parameters ranged from 0.35 to 0.99 (mean 0.85) (Table S8), showing good agreement between predicted and measured values. Robustness was verified by bootstrap (100 runs, $R \geq 0.6$) and DISP analyses. All six factors were stably reproduced, with mapping rates 68–100% (Table S9). DISP showed no factor swaps at dQmax $= 4, 5, 15$, indicating no rotational ambiguity (Table S10). These diagnostics support the reliability of the six-factor solution, although Factor 4 should be interpreted cautiously due to a lower mapping rate. Factor profiles and contributions are presented in Table S11 and Fig. 7.

Factor 1 made prominent contributions to Na^+ (33.26%), K^+ (43.19%), and F^- (49.68%). This pattern is consistent with intense forward cation exchange (Na^+-Ca^{2+}), leading to significant Na^+ enrichment and corresponding depletion of Ca^{2+} and Mg^{2+}. Weathering of fluorine-bearing silicate minerals (e.g., biotite, amphibole) in Jurassic volcaniclastic rocks, enhanced by prolonged water-rock interaction and cation exchange, was the main F^- source. The consistent

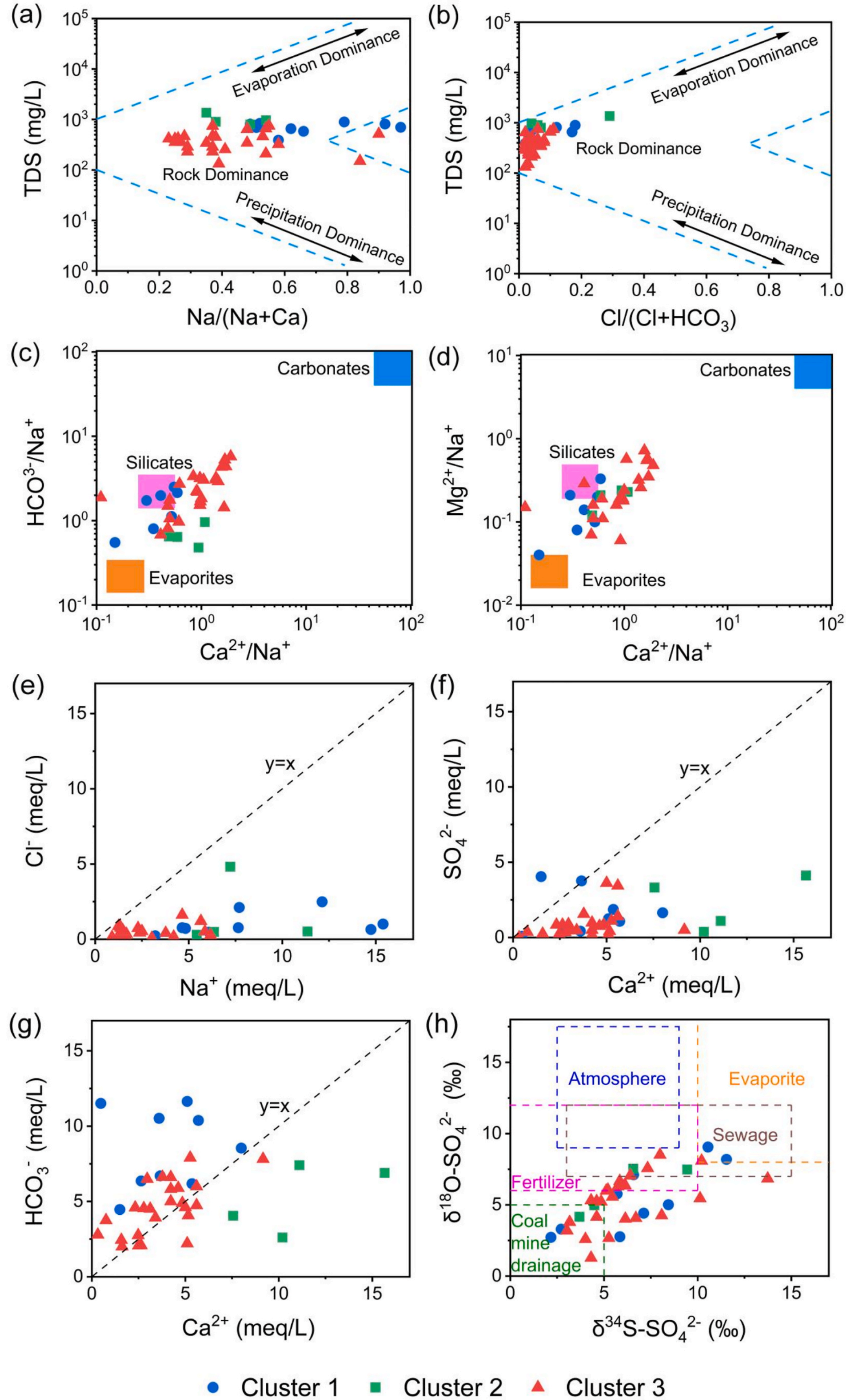

Fig. 6. Hydrochemical processes controlling groundwater chemistry. (a-b) Gibbs diagrams, (c-d) Gaillardet diagrams, and (e-h) ionic ratio correlation diagrams: (e) Cl^- vs. Na^+, (f) SO_4^{2-} vs. Ca^{2+}, (g) HCO_3^-vs. Ca^{2+}, (h) $\delta^{18}O$-SO_4^{2-}vs. $\delta^{34}S$-SO_4^{2-}.

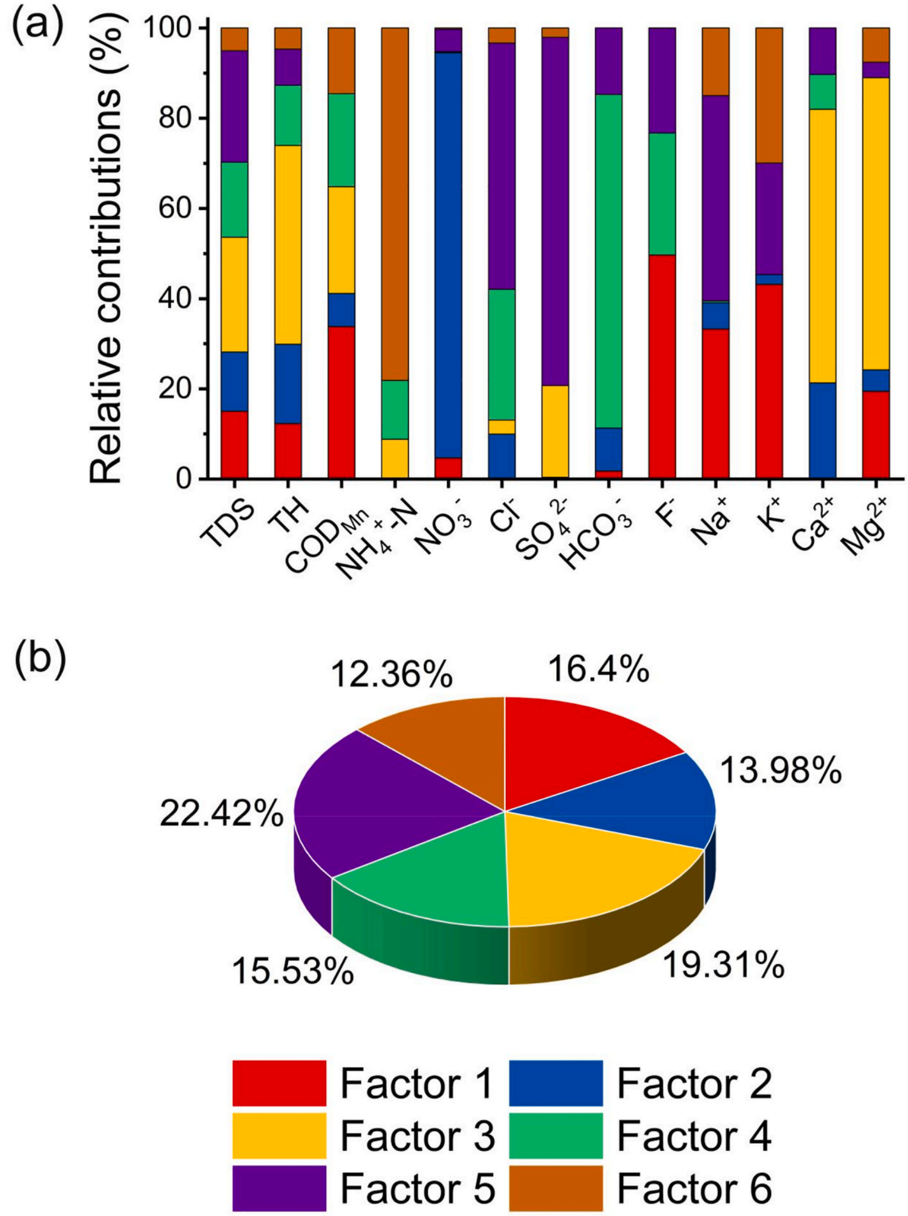

Fig. 7. The PMF analytic results (a) Contribution rate of the six factors to each chemical parameters. (b) Proportion of the six factors in source apportionment (Factor 1: cation exchange-silicate weathering; Factor 2: agricultural fertilizers; Factor 3: human-enhanced silicate dissolution; Factor 4: organic matter degradation-carbonate; Factor 5: mixed coal mine drainage and industrial/domestic wastewater; Factor 6: domestic sewage and livestock manure).

undersaturation of all samples with respect to fluorite (CaF_2; Table S7) indicated that F^- concentrations were not limited by fluorite solubility, further supporting this interpretation. Accordingly, Factor 1 was identified as a cation exchange-silicate weathering source contributing to both major ion evolution and fluoride enrichment.

Factor 2 showed an exceptionally high contribution to NO_3^- (89.86%), far exceeding other indicators. NO_3^- enrichment in groundwater is typically linked to agricultural fertilizers, domestic sewage and manure discharge, or light industrial wastewater (Xu et al., 2025). However, organic nitrogen sources would also elevate COD_{Mn} (Hou et al., 2022). In Factor 2, the contribution to COD_{Mn} was only 7.36%, ruling out significant organic inputs. Thus, the NO_3^- in Factor 2 primarily originated from inorganic agricultural fertilizers. This factor was identified as an agricultural fertilizer source.

Factor 3 showed high contributions to Mg^{2+} (64.77%), Ca^{2+} (60.68%), and SO_4^{2-} (20.33%). This factor represents anthropogenically enhanced silicate weathering rather than a natural geochemical process. Exogenous H_2SO_4 from coal mine drainage (oxidation of sulfur compounds) accelerates dissolution of Ca-Mg silicates, releasing Ca^{2+}, Mg^{2+}, SO_4^{2-} (Li et al., 2008; Nordstrom et al., 2015). This interpretation is supported by SO_4^{2-} isotopes (Fig. 6h): samples with high Factor 3 contributions (all belonging to Cluster 2, Table S12) fell in the mine drainage field. This confirms that the acidity driving silicate dissolution originates from coal mining activities rather than natural processes. Thus, Factor 3 was therefore identified as a human-enhanced silicate mineral dissolution source.

Factor 4 was dominated by HCO_3^- (73.99%), with moderate contributions to COD_{Mn} (20.59%), Cl^- (28.99%), and F^- (27.04%), but low contributions to major cations such as Na^+, Ca^{2+}, and Mg^{2+}. This pattern does not represent typical silicate or carbonate dissolution. Instead, microbial degradation of organic matter in the vadose zone or shallow aquifer generated CO_2, driving carbonate equilibrium toward HCO_3^- formation (Cangemi et al., 2021; Yamanaka, 2012). The concurrent presence of COD_{Mn}, Cl^-, and F^- further supported inputs from mild domestic sewage or soil organic matter (Zhou et al., 2026). Factor 4 was thus identified as an organic matter degradation-carbonate source.

Factor 5 showed significant contributions to SO_4^{2-} (77.23%), Cl^- (54.58%), and Na^+ (45.52%). Sulfate dual isotope data (Fig. 6h) indicated that SO_4^{2-} mainly from coal mine drainage or wastewater, while Na^+ and Cl^- co-enrichment also reflected mixed domestic/industrial wastewater inputs (Mu et al., 2024; Yan et al., 2025). High Factor 5 samples were distributed in mining, industrial, agricultural areas (Table S12). Factor 5 was thus identified as a mixed source of coal mine drainage and industrial/domestic wastewater.

Factor 6 showed the highest contribution to NH_4^+-N (78.15%), followed by K^+ (29.95%) and Na^+ (14.99%). These ions are common in domestic sewage and livestock manure: NH_4^+-N derives from livestock farming and nitrogen-based fertilizers, while K^+ and Na^+ come from domestic wastewater and livestock excreta (Jiang et al., 2024; Li et al., 2020). Inadequate sewage and manure management allowed pollutant infiltration. Consequently, Factor 6 was determined as a mixed source of domestic sewage and livestock manure.

The six factor contributions were normalized (Fig. 7b), quantifying each source's impact on groundwater contamination. The results showed that the most significant influence was attributed to the mixed pollution source of coal mine drainage and industrial/domestic wastewater (Factor 5, 22.42%), followed by silicate mineral dissolution (Factor 3, 19.31%), cation exchange-silicate weathering (Factor 1, 16.40%), organic matter degradation-carbonate (Factor 4, 15.53%), agricultural fertilizers (Factor 2, 13.98%), and domestic sewage and livestock manure (Factor 6, 12.36%). This ranking shows that anthropogenic-driven composite contamination dominated groundwater quality evolution. Mining and industrial activities exerted the strongest pressure. Human-enhanced silicate dissolution ranked second, confirming strong

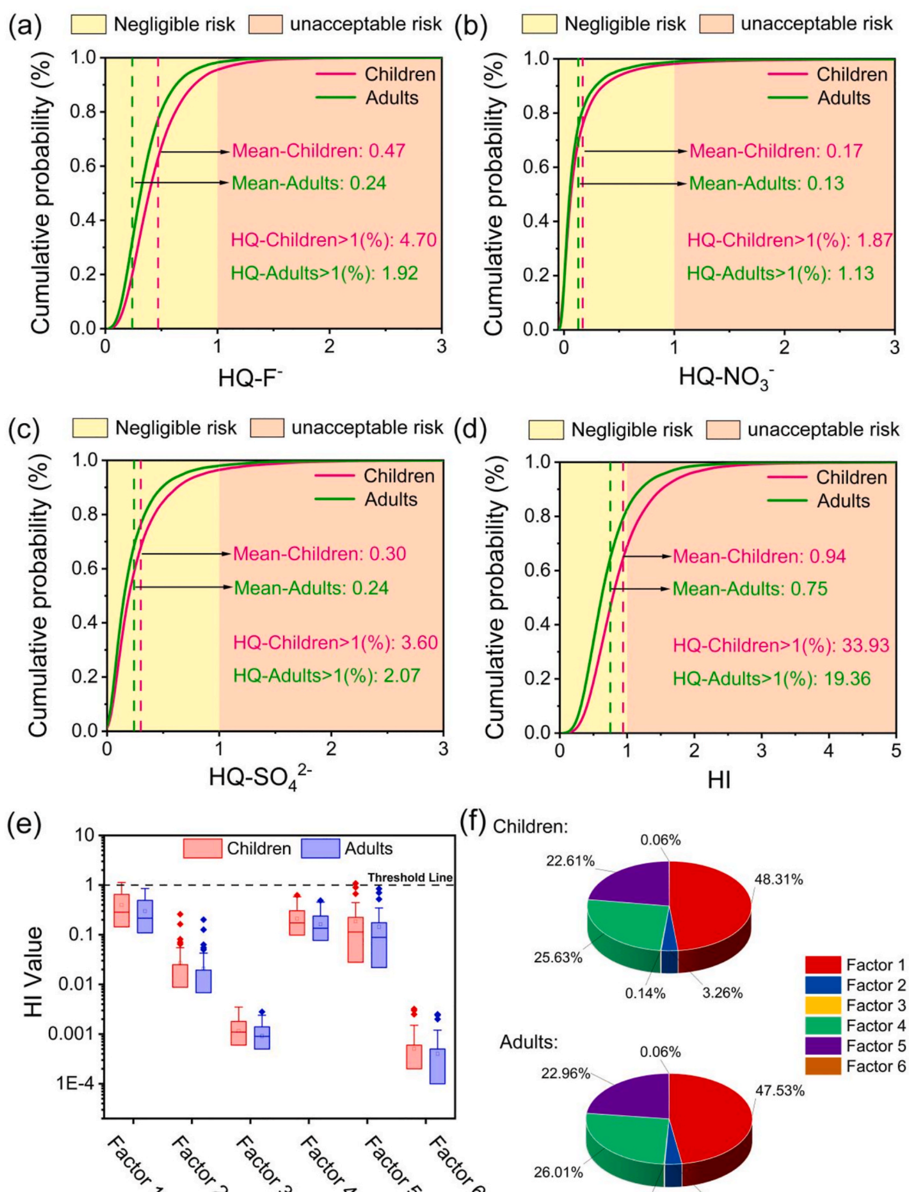

Fig. 8. Human health risk prioritization and source-specific risk contribution. Cumulative distribution functions of HQ for (a) F^-, (b) NO_3^-, (c) SO_4^{2-} and (d) HI. (e) Box plots showing the distribution of HI values attributed to each PMF-resolved factor. (f) Percentage contribution of each factor to the total HI for Children and Adults. (HQ: hazard quotient; HI: hazard index; Factor 1: cation exchange-silicate weathering; Factor 2: agricultural fertilizers; Factor 3: human-enhanced silicate dissolution; Factor 4: organic matter degradation-carbonate; Factor 5: mixed coal mine drainage and industrial/domestic wastewater; Factor 6: domestic sewage and livestock manure).

human alteration of natural water-rock interactions. In summary, the six PMF-resolved factors fall into three categories: (i) natural background (Factor 1, 16.40%); (ii) human-enhanced geochemical reactions (Factors 3 and 4, 34.84%); and (iii) direct pollutant inputs (Factors 2, 5, and 6, 48.76%). This classification reveals that direct human inputs and enhanced reactions together accounted for the vast majority of contamination, while natural background played a minor role.

5.3. Human health risk prioritization

5.3.1. Probabilistic risk characterization

Monte Carlo simulation method (10,000 random iterations) was used to assess non-carcinogenic health risks from F^-, SO_4^{2-}, and NO_3^- via ingestion and dermal contact for adults and children. The results showed that the non-carcinogenic risks associated with these three pollutants differed significantly across population groups (Fig. 8a–d). Regarding the hazard quotient (HQ) for individual pollutants, the probabilities of HQ exceeding the safety threshold (>1) in children were 4.70% for F^-, 3.60% for SO_4^{2-}, and 1.87% for NO_3^-, whereas the corresponding exceedance probabilities in adults were 1.92%, 2.07% and 1.13%, respectively. These findings indicated that F^- and SO_4^{2-} were priority pollutants in local groundwater, and children faced higher health risks from all three pollutants compared to adults. The simulated hazard index (HI) showed that children had a 33.93% probability of exceeding the safe threshold (HI > 1), 1.75 times higher than adults (19.36%). Therefore, children should be a key protection target in regional groundwater health risk management.

5.3.2. Source-specific health risk contributions

Factor contributions to HI were quantified for children and adults (Fig. 8e and f). Factor 1 (cation exchange-silicate weathering) had the highest median HI (Fig. 8e), with several child samples exceeding HI > 1, confirming it as the dominant risk source. Factor 4 (organic matter degradation-carbonate) and Factor 5 (mixed coal mine drainage and industrial/domestic) followed, with Factor 5 also showing HI > 1 in some samples. Factor 1 alone accounted for approximately 48% of total HI for both groups (Fig. 8f). Factors 4 and 5 contributed approximately 26% and 23%, respectively. This shows that geogenic F^- enrichment (Factor 1) was the primary health risk driver, followed by organic matter degradation (Factor 4) and anthropogenic SO_4^{2-} inputs (Factor 5). Higher HI in children across all factors further confirms their vulnerability. This quantitative source-risk linkage provides a scientific basis for prioritizing control measures.

5.3.3. Spatial distribution and sensitivity analysis of health risks

Spatially, high HI areas were concentrated in residential zones southeast of the central open-pit coal mine (Fig. S4 a and b), coinciding with the regions of Cluster 2 and partial Cluster 1. Deterministic risk estimates showed that 36.65% of children and 26.31% of adults exceeded the safe threshold (HI > 1). Probabilistic risks from Monte Carlo simulation (33.93% for children, 19.36% for adults) were lower than these deterministic values, as probabilistic simulation incorporates parameter variability and extreme values, more comprehensively reflecting real-world exposure scenarios (Guyonnet et al., 1999; Yuan et al., 2024). IDW interpolation for HI was validated by LOOCV: ME = 0.022, RMSE = 0.4815 (Table S2). Near-zero ME confirms no systematic bias. Moderate RMSE stems from the small sample size (n = 38), but high-risk zone delineation is supported by measured high HI values in Cluster 2 and partial Cluster 1. This highlights the need to prioritize child health protection in high-risk areas.

Sensitivity analysis identified F^-, SO_4^{2-}, and NO_3^- concentrations as the most influential parameters driving HI uncertainty (Fig. S4 c and d). For children, sensitivity indices were 56.51% (F^-), 52.28% (SO_4^{2-}), and 37.13% (NO_3^-), with a similar pattern in adults (56.42%, 53.31%, and 36.15%, respectively). This consistent ranking ($F^- > SO_4^{2-} > NO_3^-$) indicates that controlling F^- and SO_4^{2-} would be the most effective strategy for reducing health risks in this region.

5.4. Integrated source-risk assessment and management implications

By integrating PMF source apportionment with health risk assessment, this study provides a quantitative basis for linking specific pollution sources to health outcomes, enabling more targeted management strategies. Based on these source-specific risk contributions, the following interventions are recommended in order of priority.

(1) Top priority: geogenic fluoride enrichment (Factor 1, ~48% of total HI). For Cluster 1 (central-north around mine dewatering cone), fluoride is naturally released from weathering of F-bearing minerals. Management should focus on alternative water supplies or point-of-use treatment (e.g., reverse osmosis), considering local economic realities and community acceptance. Optimizing well placement and monitoring groundwater levels would also help prevent further fluoride mobilization.
(2) Second priority: organic matter degradation (Factor 4, ~26% of total HI). This source contributes substantially to HI via fluoride. Unlike geogenic fluoride, the fluoride associated with this source is likely mobilized from geological materials by degradation of organic matter from domestic/agricultural sources. Management should focus on reducing organic loading into aquifers through improved rural sanitation and wastewater collection.
(3) Third priority: anthropogenic sulfate inputs (Factor 5, ~23% of total HI). For Cluster 2 (southeastern residential zones), the dominant source is mixed coal mine drainage and industrial/domestic wastewater. Management should prioritize enhanced treatment of mine drainage, upgrading industrial wastewater collection, and strict discharge regulation.
(4) Cross-cutting priority: child health protection. Children face 1.75× higher health risk than adults (33.9% vs 19.4%). Targeted interventions in high-risk areas (Cluster 2 and partial Cluster 1) should include point-of-use treatment in schools, regular health surveillance, and public health advisories. Although nitrate (Factor 2) and sewage (Factor 6) contribute minimally (~3% and <0.1%), they remain relevant for local agricultural and sanitation management.

5.5. Limitations and future research directions

This study has several limitations to be addressed in future work. First, the analysis is based on a single sampling campaign (May 2025). While this represents a stable hydrogeological period, seasonal variations (especially for NO_3^- affected by agriculture and recharge) cannot be captured. Detailed screen and sampling depths were not recorded, limiting vertical water quality characterization. Future studies should include multi-seasonal monitoring, record detailed well construction data, and follow standard protocols requiring parameter stabilization before sample collection to resolve temporal and vertical dynamics and validate source apportionment stability. Furthermore, the health risk assessment has two limitations: the additive HI model does not account for potential synergistic or antagonistic effects among co-occurring contaminants (F^-, NO_3^-, SO_4^{2-}), and dermal exposure parameters rely on default USEPA values lacking region-specific data. Future studies should explore mixture toxicology models and develop region-specific exposure parameters to refine risk estimates. Third, the PMF receptor model provides a static snapshot of source contributions, whereas the groundwater system is dynamic (preferential flow paths, drawdown cones, seasonal recharge). Future studies should integrate numerical groundwater flow and transport modeling (e.g., MODFLOW) with receptor models to resolve temporal dynamics and simulate source contribution evolution under changing hydrogeological conditions. Additional future work should include: compound-specific isotope analysis (e.g., $\delta^{15}N\text{-}NO_3$) to better constrain nitrate sources; inclusion of organic contaminants (e.g., PAHs) for more comprehensive risk assessment; and development of probabilistic risk-cost optimization frameworks to support cost-effective remediation.

6. Conclusions

This study developed an integrated framework combining SOM, PMF, and Monte Carlo simulation to link groundwater contamination sources with health risks in a typical arid coal-mining area. Application to the Huolingol mining area yielded three main findings.

(1) Three groundwater clusters were identified with distinct contamination patterns: Cluster 1 (Na-HCO_3 type, slow-flow zones): elevated F^- and NH_4^+-N, with F^- enrichment controlled by cation exchange and silicate weathering under alkaline conditions. Cluster 2 (Ca-Na-SO_4 type, stagnation zones near industry/agriculture): high TDS, TH, SO_4^{2-}, NO_3^-, indicating strong anthropogenic impact. Cluster 3 (Ca-HCO_3 type): natural background quality.
(2) PMF source apportionment quantified the dominance of anthropogenic sources. Six contamination sources were identified, with mixed coal mine drainage and industrial/domestic wastewater contributing the most (22.42%), followed by human-enhanced silicate dissolution (19.31%). Together they accounted for over 40% of total contamination, highlighting the overwhelming impact of mining and industrial activities.
(3) Health risk assessment identified children were the most vulnerable (33.9% HI > 1 vs. 19.4% for adults), with F^- and SO_4^{2-} as primary risk drivers. High-risk areas coincided with Cluster 2 and partial Cluster 1. Source-specific analysis showed that geogenic fluoride (Factor 1) contributed ~48% of total HI, followed by organic matter degradation (Factor 4, ~26%) and anthropogenic sulfate (Factor 5, ~23%).

This framework quantitatively links pollution sources to health risks and can be applied to identify priority pollutants and vulnerable populations in similar mining regions. It provides evidence-based prioritization for equitable resource allocation and delivers actionable guidance for policymakers in regions where mining livelihoods coexist with water scarcity and public health vulnerabilities.

CRediT authorship contribution statement

Zheng Yin: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Shiqi Zhang:** Investigation. **Xiao Fu:** Conceptualization, Funding acquisition, Writing – review & editing. **Ran Sun:** Investigation, Methodology. **Ran Lv:** Investigation. **Meng Yuan:** Investigation. **Ziyuan Gao:** Investigation. **Gang Wu:** Funding acquisition, Supervision.

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.

Acknowledgments

This research was supported by the National Key Research and Development Program of China (Grant No. 2022YFF1303201).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2026.130146.

Data availability

Data will be made available on request.

References

Appukuttan, A., Aju, C.D., Reghunath, R., Srinivas, R., Krishnan, K.A., Arya, A., 2025. Exploring hydrochemical drivers of drinking water quality in a tropical river basin using self-organizing maps and explainable AI. Water Res. 284, 123884.
Banks, D., Younger, P.L., Arnesen, R.-T., Iversen, E.R., Banks, S.B., 1997. Mine-water chemistry: the good, the bad and the ugly. Environ. Geol. 32 (3), 157–174.
Cangemi, M., Ranaldi, M., Carapezza, M.L., Vinci, M., Madonia, P., 2021. Preliminary geochemical characterization of the Mts. Simbruini karst aquifer (Central Italy). Ann. Geophys. 64 (6).
Chen, P., 2021. Effects of the entropy weight on TOPSIS. Expert Syst. Appl. 168, 114186.
Chen, K., Liu, Q., Yang, T., Ju, Q., Hou, X., Gao, W., Jiang, S., 2023. Groundwater pollution source identification and health risk assessment in the North Anhui Plain, eastern China: insights from positive matrix factorization and Monte Carlo simulation. Sci. Total Environ. 895, 165186.
Chien, L., Robertson, H., Gerrard, J.W., 1968. Infantile gastroenteritis due to water with high sulfate content. Can. Med. Assoc. J. 99 (3), 102–104.
Cui, H., Duan, L., Pan, H., Liu, T., 2025. Geochemical pattern, quality and driving forces of multi-layer groundwater in a high-capacity mining area basin: a comprehensive analysis based on the interweaving of multiple factors. J. Hydrol. 660, 133376.
Dehghan Rahimabadi, P., Behnia, M., Nasabpour Molaei, S., Khosravi, H., Azarnivand, H., 2023. Assessment of groundwater resources potential using Improved Water Quality Index (ImpWQI) and entropy-weighted TOPSIS model. Sustain. Water Resour. Manag. 10 (1), 7.
Fan, W., Zhou, J., Zheng, J., Guo, Y., Hu, L., Shan, R., 2024. Hydrochemical characteristics, control factors and health risk assessment of groundwater in typical arid region Hotan Area, Chinese Xinjiang. Environ. Pollut. 363, 125301.
Foster, S.S.D., Chilton, P.J., 2003. Groundwater: the processes and global significance of aquifer degradation. Philos. Trans. R. Soc. B Biol. Sci. 358 (1440), 1957–1972.
Gaillardet, J., Dupré, B., Louvat, P., Allègre, C.J., 1999. Global silicate weathering and CO_2 consumption rates deduced from the chemistry of large rivers. Chem. Geol. 159 (1–4), 3–30.
Ghazavi, R., Vali, A.B., Eslamian, S., 2012. Impact of flood spreading on groundwater level variation and groundwater quality in an arid environment. Water Resour. Manag. 26 (6), 1651–1663.
Gibbs, R.J., 1970. Mechanisms controlling world water chemistry. Science 170 (3962), 1088–1090.
Guyonnet, D., Côme, B., Perrochet, P., Parriaux, A., 1999. Comparing two methods for addressing uncertainty in risk assessments. J. Environ. Eng. 125 (7), 660–666.
Hou, W.L., Chen, X., Wu, J.H., Zhang, C., Yu, J.H., Bai, J., Chen, T.T., 2022. Sources and spatiotemporal variations of nitrogen and phosphorus in Liaodong Bay, China. Mar. Pollut. Bull. 185.
Huolingol Municipal People's Government, 2024. Huolingol municipal government work report, 2024). Retrieved September 26, 2025, from. http://www.hlgls.gov.cn/zwgk/gzbg/202408/t20240807_872969.html.
Jiang, X., Wang, H., Wu, P., Lei, Y., Deng, L., Wang, W., 2024. Enhanced narH gene expression contributing to nitrite accumulation in simultaneous nitrification and denitrification under Na^+ stress instead of K^+ stress. Chem. Eng. J. 479, 147637.
Jin, B., Xu, X., 2024. Machine learning-based scrap steel price forecasting for the Northeast Chinese market. Int. J. Empir. Econ. 3 (4), 2450011.
Jin, B., Xu, X., 2025a. Machine learning WTI crude oil price predictions. J. Int. Commer. Econ. Policy. 16 (1), 2550004.
Jin, B., Xu, X., 2025b. Predicting open interest in thermal coal futures using machine learning. Miner. Econ. 38, 795–809.
Ju, Q., Hu, Y., Liu, Q., Chai, H., Chen, K., Zhang, H., Wu, Y., 2023. Source apportionment and ecological health risks assessment from major ions, metalloids and trace elements in multi-aquifer groundwater near the Sunan mine area, Eastern China. Sci. Total Environ. 860, 160454.
Li, Q., Zhang, H., Guo, S., Fu, K., Liao, L., Xu, Y., Cheng, S., 2020. Groundwater pollution source apportionment using principal component analysis in a multiple land-use area in southwestern China. Environ. Sci. Pollut. Res. 27 (9), 9000–9011.
Li, S.-L., Calmels, D., Han, G., Gaillardet, J., Liu, C.-Q., 2008. Sulfuric acid as an agent of carbonate weathering constrained by $\delta^{13}C_{DIC}$: Examples feom Southwest China. Earth Planet Sci. Lett. 270 (3), 189–199.
Li, W., Wu, J., Zhou, C., Nsabimana, A., 2021. Groundwater pollution source identification and apportionment using PMF and PCA-APCS-MLR receptor models in Tongchuan City, China. Arch. Environ. Contam. Toxicol. 81 (3), 397–413.
Mao, H., Wang, G., Rao, Z., Liao, F., Shi, Z., Huang, X., Chen, X., Yang, Y., 2021. Deciphering spatial pattern of groundwater chemistry and nitrogen pollution in Poyang Lake Basin (eastern China) using self-organizing map and multivariate statistics. J. Clean. Prod. 329, 129697.
Mu, D., Li, P., De Baets, B., Li, D., Li, Z., He, S., 2024. A multi-perspective exploration of the salinization mechanisms of groundwater in the Guanzhong Basin, China. Sci. Total Environ. 957, 177421.
Naghipour, P., 2026. Metaheuristic model tuning for energy forecasting in low-income housing in the MollaZeynal region of Tabriz city: linking comfort and carbon. Next Res 4, 101265.
Naghipour, P., Naghipour, A., Bakirova, T., 2026c. Life-cycle assessment and multi-objective optimization of natural-insulated envelopes across Iranian climates. Build. Eng. 4 (1).

Naghipour, P., Naghipour, A., Bakirova, T., Ghiyasi, H., Faramarzi, F.S.G., Faramarzi, F. S.G., 2026a. Decarbonizing precast concrete building components: Cradle-to-site carbon modeling and optimization, explainable machine learning, and a transportation efficiency index. Build. Eng. 4 (2).
Naghipour, P., Naghipour, A., Bakirova, T., Soltani Gerd Faramarzi, F., Soltani Gerd Faramarzi, F., 2026b. Adobe versus concrete: passive energy analysis in residential buildings in the hot and arid climate of Kashan city. Build. Eng. 4 (1).
Naghipour, P., Naghipour, A., 2025a. Evaluating heating energy consumption in residential buildings using hybrid machine learning models: the case of parsabad city. Next Res 2 (3).
Naghipour, P., Naghipour, A., 2025b. Energy performance analysis of residential buildings in Bandar Anzali: influence of orientation and aspect ratio. Next Sustain 5, 100140.
Nordstrom, D.K., Blowes, D.W., Ptacek, C.J., 2015. Hydrogeochemistry and microbiology of mine drainage: an update. Appl. Geochem. 57, 3–16.
Norris, G., Duvall, R., Brown, S., Bai, S., 2014. EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide. U.S. Environmental Protection Agency, Washington, DC (EPA/600/R-14/108).
Paatero, P., Tapper, U., 1994. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5 (2), 111–126.
Qu, S., Duan, L., Mao, H., Wang, C., Liang, X., Luo, A., Huang, L., Yu, R., Miao, P., Zhao, Y., 2023. Hydrochemical and isotopic fingerprints of groundwater origin and evolution in the Urangulan River basin, China's Loess Plateau. Sci. Total Environ. 866, 161377.
Qu, S., Duan, L., Shi, Z., Liang, X., Lv, S., Wang, G., Liu, T., Yu, R., 2022. Hydrochemical assessments and driving forces of groundwater quality and potential health risks of sulfate in a coalfield, northern Ordos Basin, China. Sci. Total Environ. 835, 155519.
Raju, N.J., Ram, P., Dey, S., 2009. Groundwater quality in the lower Varuna River Basin, Varanasi District, Uttar Pradesh. J. Geol. Soc. India 73 (2), 178–192.
Santos, M.R., Roisenberg, A., Iwashita, F., Roisenberg, M., 2020. Hydrogeochemical spatialization and controls of the Serra Geral Aquifer System in southern Brazil: a regional approach by self-organizing maps and k-means clustering. J. Hydrol. 591, 125602.
Su, H., Zhang, Y., Cao, W., Li, Y., Zhang, H., Li, Z., Ma, L., 2026. Spatiotemporal characteristics, drivers, sources, and health risks of nitrate and sulfate in groundwater on the Chinese Loess Plateau. Water Res. 288, 124531.
USEPA, 2004, US Environment Protection Agency, Washington DC. Risk Assessment Guidance for Superfund Volume I: Human Health Evaluation Muanual (Part E, Supplemental Guidance for Dermal Risk Assessment).
Wang, S., 2015. Groundwater Quality Assessment of Holingol [D]. Jilin University.
Wang, X., Xiao, C., Yang, W., Liang, X., Zhang, L., Zhang, J., 2023. Analysis of the quality, source identification and apportionment of the groundwater in a typical arid and semi-arid region. J. Hydrol. 625, 130169.
Wu, C., Fang, C., Wu, X., Zhu, G., Zhang, Y., 2021. Hydrogeochemical characterization and quality assessment of groundwater using self-organizing maps in the Hangjinqi gasfield area, Ordos Basin, NW China. Geosci. Front. 12 (2), 781–790.
Wu, X.C., Yu, J.H., Yang, S.M., Zhang, Y.H., Hu, Q.L., Xu, X.J., Wang, Y., Wang, Y.S., Luo, H., Xie, Z., 2024. Hydrogeochemistry, water quality, and health risk analysis of phreatic groundwater in the urban area of Yibin City, Southwestern China. Water 16 (24).
Xu, D.H., Li, T., Lin, Y.Z., Chen, T.F., 2025. Source apportionment of nitrate in groundwater based on correlation monitoring indicators, in Liaodong Bay. Earth Sci. Front. 32 (4), 376–387.
Xu, X., Zhang, Y., 2023. Wholesale food price index forecasts with the neural network. Int. J. Comput. Intell. Appl. 22 (4), 2350024.
Yamanaka, M., 2012. Contributions of C3/C4 organic materials and carbonate rock to dissolved inorganic carbon in a karst groundwater system on Miyakojima Island, southwestern Japan. J. Hydrol. 412–413, 151–169.
Yan, Y., Zhang, Y., Sun, Z., Xie, Z., Yao, R., Chen, S., Uddin, M.G., Pu, Y., Yang, C., Wang, Y., Wang, Y., 2025. Using unsupervised machine learning and positive matrix factorization models to drive groundwater chemistry and associated health risks in a coal-mining rural region. J. Hydrol. 661, 133691.
Yang, L., Tang, Y., Sun, H., He, L., Li, R., 2024. Hydrochemical characteristics of abandoned coal mines derived acid mine drainage in a typical karst basin (Wuma river basin, Guizhou China). Heliyon 10 (11), e31963.
Younger, P.L., 2000. Predicting temporal changes in total iron concentrations in groundwaters flowing from abandoned deep mines: a first approximation. J. Contam. Hydrol. 44 (1), 47–69.
Yuan, B., Liu, H.P., Du, P., Chen, J., Zhang, Y.H., Zhang, H., 2024. Health risk assessment for an arsenic-contaminated site based on Monte Carlo Simulation and parameters optimization. Environ. Sci. 45 (2), 1049–1057.
Zhang, S., 2016. Groundwater Hydrogeochemical Characteristics and Spatial Distribution Research of Huolinguole [D]. Inner Mongolia University.
Zhang, L.Z., Liang, X.J., Xiao, C.L., Yang, W.F., Zhang, J., Wang, X.K., 2023. Hydrochemical characteristics and the impact of human activities on groundwater in a semi-arid plain: a case study of western Jilin Province, Northeast China. Environ. Sci. Pollut. Res. 30 (51), 110204–110219.
Zhong, C., Wang, H., Yang, Q., 2022. Hydrochemical interpretation of groundwater in Yinchuan basin using self-organizing maps and hierarchical clustering. Chemosphere 309, 136787.
Zhou, B., Xu, X., Wang, X., Tian, W., Peng, J., Lei, K., 2026. Integrating random forest and isotopic tracers to optimize PMF-based source apportionment of watershed pollution. J. Contam. Hydrol. 277, 104838.