Pol. J. Environ. Stud. Vol. XX, No. X (XXXX), 1-22
DOI: *10.15244/pjoes/231368* ONLINE PUBLICATION DATE: 2026-09-29

Original Research

Analysis of Groundwater Chemistry and Water Source Characteristics in the Banji Mining Area Based on a Dual Model Combining APCS-MLR and PMF

Qingfa Liao[1, 2], Qiding Ju[3]*, Bing Li[1, 2], Yanxi Zhang[1], Biao Xu[1, 2]

[1]National Engineering Laboratory for Protection of Coal Mine Eco-environment, Huainan Mining (Group) Co., Ltd, Huainan 232000, China
[2]Ping An Coal Mining Engineering Technology Research Institute Co. Ltd., Huainan, Anhui 232001, China
[3]School of Earth and Environment, Anhui University of Science and Technology, Huainan, Anhui 232001, China

Received: 07 July 2026
Accepted: 20 August 2026

Abstract

Quantitative hydrochemical source apportionment underpins precise mine water inrush prevention and inter-aquifer pollution control. Ninety-two groundwater samples from four aquifers in the Anhui Banji Mining Area were tested for six major ions (Ca^{2+}, Mg^{2+}, Na^{+}+K^{+}, Cl^{-}, SO_4^{2-}, HCO_3^{-}). Piper diagrams and Pearson correlation analysis identified dominant HCO_3-Ca·Mg hydrochemical facies, with some Quaternary pore water evolving into the Cl·SO_4-Na type due to evaporite leaching. Parallel Absolute Principal Component Scores-Multiple Linear Regression (APCS-MLR) and Positive Matrix Factorization (PMF) models were used for multi-source contribution quantification and cross-verification. APCS-MLR extracted three factors (83.99% cumulative variance) representing evaporite leaching, carbonate dissolution and silicate weathering, but produced unassigned residuals and negative contributions due to its unconstrained linear algorithm. Validated by the Root Mean Square Error (RMSE) elbow rule and Bootstrap test, PMF selected three optimal factors with stronger end-member indication, all non-negative contributions summing to 100% per ion. Both models agreed on key water-rock reactions, while PMF performed better in end-member recognition and robustness. This dual-model mutual verification improves the reliability of single-method results, reveals hydrochemical evolution driven by structures and mining disturbances, and offers technical support for water inrush source identification in Huainan coal mines.

Keywords: groundwater hydrochemistry, source apportionment, Absolute Principal Component Scores-Multiple Linear Regression (APCS-MLR), Positive Matrix Factorization (PMF), mine water inrush hazard

*e-mail: qdju@aust.edu.cn
Tel.: +86-18055491803
0009-0008-5742-8557

Introduction

Accurately identifying the types of water sources responsible for mine water inrushes and determining the contribution rate of each source represent core technical challenges in the precise prevention and control of mine water hazards [1, 2]. Utilizing multivariate statistical models to quantitatively trace the origins of groundwater sources is currently the mainstream and highly efficient approach in mine hydrogeological research [3]. Similar engineering and hydrogeological problems prevail in deep underground engineering projects in China: excavation-induced damage to aquicludes and activation of tectonic water-conducting structures are the fundamental triggers for water and gas inrush disasters, which has been fully verified in source-tracing research on high-temperature water-gas outbursts in deep tunnels of Southwest China [4]. In large-scale coal mining areas with interconnected multi-aquifer systems, the spatio-temporal evolution characteristics of multiple groundwater bodies directly govern the recharge composition of mine inflow water, and quantitative apportionment of water source contribution ratios is an indispensable technical prerequisite for realizing safe coal mining and scientific water-preserved mining layout [5].

The Huainan Coalfield and its surrounding mining areas are located in the Huai River Basin. The roof and floor strata of coal-bearing formations generally contain multiple groundwater systems, including Quaternary pore water, sandstone fissure water, and karst water from the Taiyuan Formation and the Ordovician System [6]. Although these aquifers differ significantly in burial depth, lithological composition, and recharge, discharge, and drainage conditions, they generally exhibit varying degrees of hydraulic connectivity with one another [7-10]. The hydraulic connectivity of deep karst aquifers in the study area is essentially controlled by the void structure development and permeability evolution of broken limestone strata, which provides the geological conditions for cross-formational groundwater migration after mining disturbance [11]. During mining operations, once the water-blocking layers in the roof and floor are breached or tectonic water-conducting channels are opened, deep karst water can easily penetrate into the mine. This not only causes water inrush (or sudden water outburst) hazards that seriously threaten mine safety but also leads to cross-layer contamination of karst aquifers [12, 13]. Therefore, accurately identifying the source types of mine water inflows (or sudden water outbursts) and their respective contribution ratios is a key prerequisite for formulating targeted water control plans, ensuring mine safety, and promoting green coal mining [14-16].

Traditional methods for identifying groundwater sources – represented by the Piper three-line diagram, the ion ratio coefficient method, and correlation analysis – can relatively intuitively reveal the chemical types of groundwater and the intrinsic relationships among ions, and are the most widely used qualitative analytical tools in hydrogeochemical research [17-19]. However, these methods are essentially qualitative or semi-quantitative diagnostic tools; they struggle to provide definitive quantitative results regarding the contribution ratios of various water sources under complex conditions involving multiple aquifers and multicomponent mixtures [20, 21]. This limitation is particularly pronounced in cases where mine water inflows (or outbursts) are often supplied by a mixture of groundwater from multiple layers [22, 23]. To address this shortcoming, source-receptor models have gradually been introduced into the field of groundwater chemistry research. Among these, the Absolute Principal Component Scores-Multiple Linear Regression (APCS-MLR) model extracts latent common factors of water chemistry components via principal component analysis and then uses multiple linear regression to quantify the contribution of each factor to ion concentrations. While its calculation process is straightforward and its physical interpretation intuitive, its unconstrained linear regression mechanism cannot avoid statistical artifacts such as negative contributions and unrecognized sources [24-27]. In contrast, the Positive Matrix Factorization (PMF) model performs factor decomposition on the receptor concentration matrix using non-negative constraints and sample uncertainty weighting. Theoretically, this model better aligns with the actual physical process of non-negative mixing of groundwater ionic components; however, determining the number of factors and interpreting their physical significance still relies on the researcher's professional judgment [28-30].

A review of existing relevant research reveals that there are still two notable gaps in current studies on groundwater source analysis: First, most studies use only one of the two models – APCS-MLR or PMF – to conduct quantitative calculations. A single model cannot overcome its inherent algorithmic limitations, and since the analysis results lack a cross-validation step, their reliability is difficult to effectively verify. Second, given the unique engineering context of the Huainan Coalfield – where multiple aquifers are hydraulically interconnected and mine water inrushes are predominantly caused by mixed recharge from multiple groundwater layers – there is a lack of systematic comparative studies examining the consistency of end-member identification, patterns of deviations in contribution rates, and applicability for on-site water control using both models on the same dataset. Consequently, it is difficult to provide clear guidance on model selection for identifying the actual sources of water inrushes in mining areas [31]. Against this research backdrop, the combined application of these two receptor models – which complement each other in their underlying mechanisms – allows for mutual validation of results to offset the shortcomings of each algorithm. Furthermore, a quantitative evaluation of their respective strengths and weaknesses under coal mine hydrogeological conditions provides clear complementary value and practical significance.

In light of this, this paper takes the Banji Mining Area as its subject of study. A total of 92 groundwater samples were systematically collected from four aquifers – the Fourth Aquifer, sandstone aquifer, Taiyuan limestone aquifer, and Ordovician limestone aquifer. The concentrations of six major hydrochemical ions – Ca^{2+}, Mg^{2+}, $Na^{+}+K^{+}$, Cl^{-}, SO_4^{2-}, and HCO_3^{-} – were measured. First, Piper's three-line diagram and Pearson correlation analysis are used to characterize the hydrochemical types of groundwater in each aquifer and the intrinsic relationships among ion components, thereby preliminarily revealing the dominant processes controlling hydrochemical evolution. Subsequently, APCS-MLR and PMF models are constructed separately to perform quantitative source apportionment of groundwater chemical components, identifying the hydrogeochemical end-members corresponding to each model and their contribution rates. Based on this, a systematic comparison of the two receptor models is carried out across three dimensions: consistency in end-member identification, causes of differences in contribution rate values, and applicability to mine engineering.

This paper does not merely conduct repetitive calculations with the two models; its core innovations lie in three areas. First, cross-validation is realized by virtue of the algorithmic differences between the two models, which reduces errors derived from the linear hypothesis of the APCS-MLR model and the subjectivity in selecting the number of factors for the PMF model. This practice improves the reliability of multi-source partitioning and addresses the lack of verification in single-model analysis. Second, this work quantitatively defines the applicable scope of the two models for discriminating mixed multi-aquifer water inrush in the Huainan Coalfield, fills the gap in regional comparative case studies of these two models, and offers a reference for model selection in analogous coal mines. Third, a low-cost technical framework relying only on conventional constant ions to quantitatively calculate water source contribution rates is established. This method is suitable for emergency water inrush tracing scenarios with limited underground sampling and testing conditions. Ultimately, a complete discrimination framework for mine water inflow and sudden water inrush sources is proposed for similar mining areas, providing a theoretical foundation and technical support for mine water hazard prevention and the rational utilization of groundwater resources.

Materials and Methods

Study Area Overview and Data Sources

Overview of the Study Area

The Banji Coalfield is located at the junction of Yingshang County, Fuyang City, Anhui Province, and Lixin County, Bozhou City, Anhui Province. It is 25 kilometers north of Lixin County and 30 kilometers south of Yingshang County. Its geographic coordinates range from 116°09′ to 116°30′ east longitude and 32°51′ to 32°56′ north latitude. The entire mining area extends approximately 6 km from east to west and 4 to 7 km from north to south, covering an area of 33.6 km^2.

The study area has a warm-temperate semi-humid monsoon climate, characterized by cold winters and hot summers, with distinct seasonal boundaries. Specifically, spring and autumn are windy, rainfall is concentrated primarily in summer, the snowfall season typically lasts 2 to 3 months, and the soil freezing depth generally ranges from 4 to 12 cm. The long-term average precipitation is 908.6 mm; precipitation is unevenly distributed throughout the year, with June through September accounting for more than 60% of the annual total, constituting the primary recharge period for regional groundwater.

The study area generally features a gentle slope that is higher in the west than in the east and higher in the south than in the north. The terrain is open with minimal undulation, which facilitates the infiltration of atmospheric precipitation to replenish groundwater. The vertical hydrogeological structure of the mining area is clearly stratified; from top to bottom, it consists of the Quaternary loose aquifer (Quaternary Aquifer), the sandstone aquifer (Sandstone Aquifer), the Taiyuan Formation limestone aquifer (Taiyuan Limestone Aquifer), and the Ordovician limestone aquifer (Ordovician Limestone Aquifer). The Quaternary loose aquifer is particularly thick, with a stable impermeable clay layer in its lower section that effectively blocks the vertical infiltration of atmospheric precipitation and surface water. The sandstone aquifer serves as the direct water-supply aquifer for the mine; its water-bearing capacity is unevenly distributed, and the development of fractures is significantly controlled by tectonic structures. The Taiyuan Formation and Ordovician limestone aquifers are thick, exhibit well-developed karst features, and possess high water-bearing capacity; they constitute the region's primary aquifer units and represent significant potential sources of water for deep mine water inrushes. Groundwater in the study area generally flows from northwest to southeast, primarily receiving lateral runoff recharge from the surrounding area. Locally, there is overflow through loose layers and concealed recharge along fault zones. Discharge is primarily through artificial extraction and mine drainage. The sampling points in the study area are shown in Fig. 1.

Data Sources and Hydrochemical Statistical Characteristics

A total of 92 water samples were collected during this sampling campaign, including 16 samples from Quaternary, 26 from Sandstone, 28 from Taiyuan, and

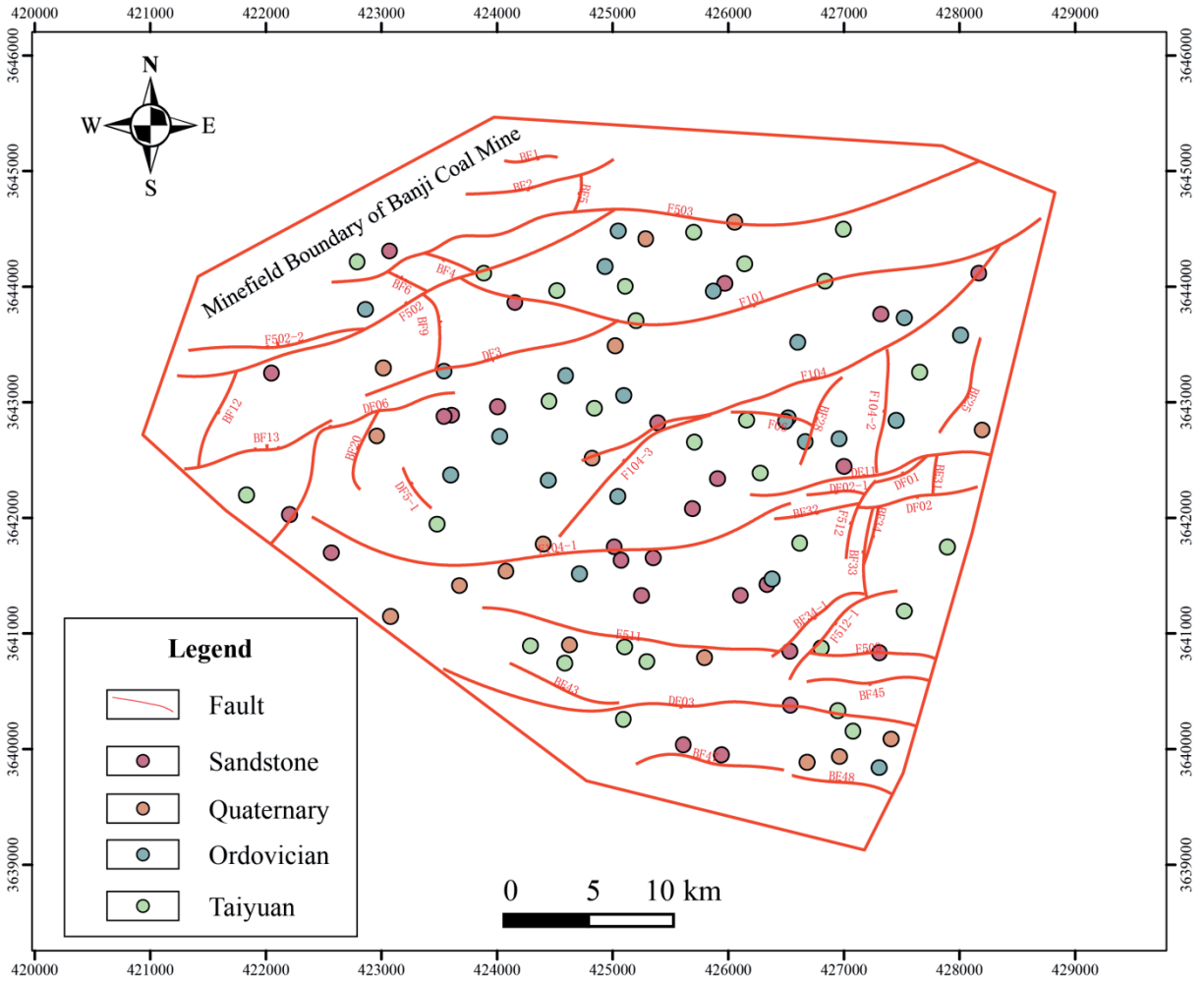

Fig. 1. Sampling points in the study area.

22 from Ordovician. Six ion concentrations, including Ca^{2+}, Mg^{2+}, Na^{+}+K^{+}, Cl^{-}, SO_4^{2-}, and HCO_3^{-}, were measured in milligrams per liter, and the statistical results are presented in Table 1.

In terms of mean values, there are significant differences in the hydrochemical composition among the four aquifer types. The HCO_3^{-} concentrations in the Fourth Aquifer, Taihui Aquifer, and Aohui Aquifer are generally higher, reflecting the widespread development of carbonate karst filtration; the average Cl^{-} concentrations in the Sandstone Aquifer and the Fourth Aquifer are significantly higher, indicating ion enrichment resulting from rock salt dissolution.

A comparison of the coefficients of variation reveals that the degree of data dispersion varies among different ions. The coefficient of variation for SO_4^{2-} is generally high across all four aquifer types, indicating that sulfate concentrations are subject to strong spatial fluctuations due to the influence of local sulfide oxidation and runoff conditions; in contrast, the coefficient of variation for HCO_3^{-} is generally low, suggesting that this component is relatively stable.

Overall, the ionic composition of the four aquifer types has formed distinct hydrochemical fingerprints in terms of both mean values and fluctuation ranges, allowing for clear differentiation among them and providing a solid data foundation for subsequent multivariate statistical analysis and source apportionment modeling.

Hydrochemical Characteristics of Groundwater

Identification of Hydrochemical Types

Hydrochemical types are a comprehensive reflection of aquifer lithology, hydrogeochemical processes, and runoff conditions. In this study, the Piper tri-line plot (as shown in Fig. 2) was used to classify the four types of aquifers in the study area, visually illustrating the differences in ionic composition and distribution among the various water bodies.

The cation triangle plot shows that the sample points are generally concentrated in the lower half of the diagram, near the side of the line connecting Ca^{2+} and Na^{+}+K^{+}. Mg^{2+} concentrations are generally low, with Mg^{2+} accounting for no more than 20% at most sample points, indicating that Mg^{2+} does not play a dominant role in the cation composition. However, differences can

Table 1. Statistical results of hydrochemical component concentrations in aquifer groundwater.

Sources	Statistical Indicators	Ca^{2+}	Mg^{2+}	$Na^{+}+K^{+}$	Cl^{-}	SO_4^{2-}	HCO_3^{-}
Quaternary	Maximum	50.20	25.55	946.25	1063.50	282.42	555.53
	Minimum	8.00	5.83	103.90	6.81	39.51	111.06
	Mean	28.17	11.98	523.15	542.07	152.35	369.30
	Standard Deviation	10.51	6.34	358.94	519.77	91.70	138.86
	Coefficient of Variation	0.37	0.53	0.69	0.96	0.60	0.38
Taiyuan	Maximum	55.33	24.37	841.52	1020.96	207.56	1356.47
	Minimum	2.70	1.83	29.26	7.06	9.65	2.99
	Mean	14.07	7.21	693.23	751.39	82.29	486.79
	Standard Deviation	9.89	5.46	205.01	292.28	50.56	273.83
	Coefficient of Variation	0.70	0.76	0.30	0.39	0.61	0.56
Ordovician	Maximum	35.04	94.26	987.57	1128.20	186.84	1095.92
	Minimum	2.90	1.94	494.23	328.69	1.38	50.34
	Mean	17.10	11.61	641.54	608.30	58.60	564.20
	Standard Deviation	10.35	18.92	120.83	251.91	65.05	233.49
	Coefficient of Variation	0.61	1.63	0.19	0.41	1.11	0.41
Sandstone	Maximum	43.69	22.12	981.50	1159.20	406.64	723.55
	Minimum	7.94	1.95	60.46	3.41	2.55	22.18
	Mean	22.52	11.15	614.73	654.02	99.42	312.30
	Standard Deviation	10.16	5.54	197.27	286.08	98.13	162.26
	Coefficient of Variation	0.45	0.50	0.32	0.44	0.99	0.52

still be observed among the four aquifer types: samples from the "Four Aquifers" group are relatively dispersed, with some clearly leaning toward the $Na^{+}+K^{+}$ side. This is typically associated with shallow pore water burial, strong evaporation and concentration effects, and cation exchange. In contrast, samples from sandstone, Ordovician marl, and Tai marl are more clustered on the side closer to Ca^{2+}, reflecting a Ca^{2+}-dominated profile. This is closely related to the dissolution of carbonate minerals (calcite, dolomite) and calcium feldspar in sandstone.

From the anion triangle diagram, the distribution range of the sample points is slightly broader than that of the cation diagram, extending from the HCO_3^{-} side all the way to the side enriched in Cl^{-} and SO_4^{2-}. This indicates that the anion types in the study area are not uniform and that there is a gradual transition from the HCO_3^{-} type to the $Cl^{-}{\cdot}SO_4^{2-}$ type. The Aogui water sampling points are closer to the HCO_3^{-} vertex, indicating that this type of water is dominated by HCO_3^{-}, corresponding to a relatively active carbonate dissolution runoff process, with rapid groundwater circulation and a short residence time. In contrast, samples from sandstone and Tai Gray formations are predominantly located in the Cl^{-} and SO_4^{2-}-enriched zones, suggesting that these water bodies have longer residence times and deeper hydrogeological processes, and may also involve sulfate contributions resulting from pyrite oxidation or gypsum dissolution.

Based on the comprehensive hydrochemical classification shown in the central diamond diagram, the vast majority of samples are concentrated in the lower-right quadrant – that is, near the HCO_3^{-} vertex and the $Ca^{2+}+Mg^{2+}$ side. This indicates that groundwater in the study area is predominantly of the HCO_3-Ca·Mg type, characterized by low mineralization and relatively shallow circulation pathways. However, a few sampling points, such as Aquifer 4, are clearly shifted toward the upper-left region enriched in $Cl^{-}+SO_4^{2-}$. This indicates that the hydrochemical type of these water bodies has shifted toward the $Cl{\cdot}SO_4$-Na type or some hybrid type, which may reflect the presence of local evaporation-concentration effects or a certain degree of mixing with deep mineralized water.

Analysis of Ionic Correlation Characteristics

To preliminarily reveal the intrinsic relationships among the groundwater chemical components in each aquifer and to identify the dominant processes that may control hydrochemical evolution, this study conducts a Pearson correlation analysis on six major ions across

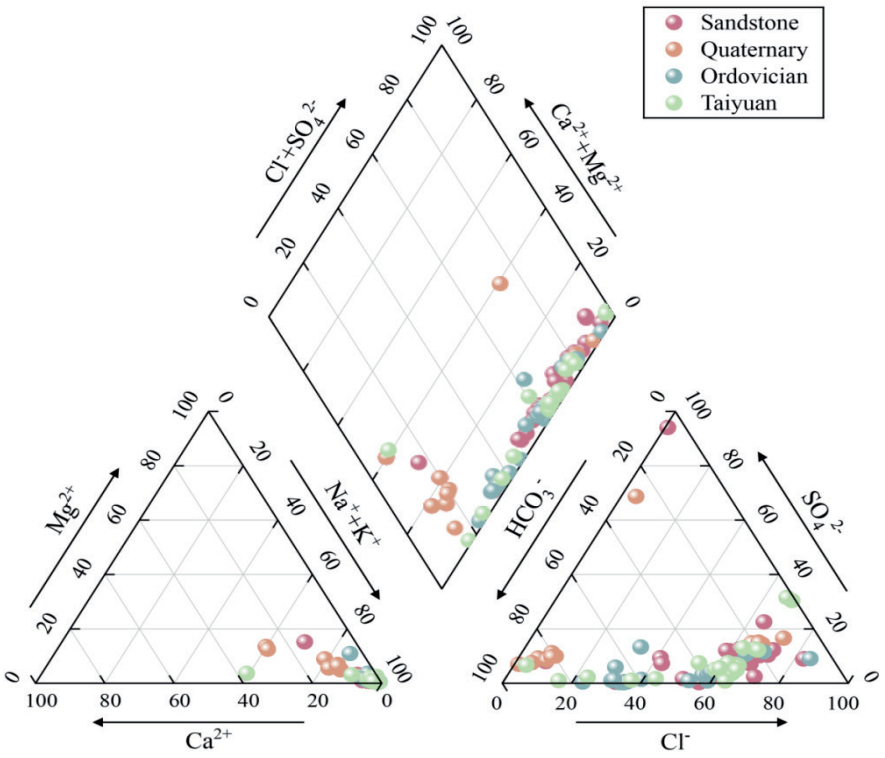

Fig. 2. Piper diagrams for the aquifers and mixed water in the study area.

the four aquifer types, with the results presented in Fig. 3.

In the Fourth Aquifer, Na^++K^+ and Cl^- exhibit an extremely strong positive correlation, indicating a common origin that mainly derives from the dissolution of halite minerals in the strata. In the Ordovician limestone aquifer, Na^++K^+ and Cl^- likewise show a high positive correlation, while Ca^{2+} is generally positively associated with the remaining ions, reflecting a hydrochemical signature dominated by overall carbonate rock leaching. In the sandstone aquifer, only Na^++K^+ and Cl^- display a relatively strong positive correlation, with an r value of 0.87, while the correlations among the other ion pairs are generally weak, suggesting that the formation mechanism of the hydrochemical components in this layer is comparatively simple and mainly governed by a straightforward dissolution and mixing process, without an evident pattern of coordinated multi-component differentiation. In addition, SO_4^{2-} and HCO_3^- show a negative correlation, reflecting that sulfide oxidation alters the acid-base conditions of groundwater and thereby inhibits carbonate dissolution. The ionic correlation pattern in the Taiyuan limestone aquifer is broadly consistent with that of the Ordovician limestone aquifer, with a relatively strong correlation between cations and Cl^-, while sulfate maintains only a weak negative correlation with HCO_3^-.

Overall, a strong positive correlation between Na^++K^+ and Cl^- is consistently observed across all four aquifer types; however, the strength of the correlation between HCO_3^- and other ions (particularly SO_4^{2-} and Cl^-) varies significantly among the different aquifers. This further confirms that the four groundwater types possess mutually independent hydrochemical fingerprints, thereby meeting the requirements for subsequent quantitative analysis of multiple water sources.

The Theoretical Basis of APCS-MLR and PMF Models

APCS-MLR Model

This model corrects the negative values inherent in conventional principal component analysis by adjusting scores using absolute principal components. By combining this approach with multiple linear regression, it enables the qualitative identification of pollution sources and the quantitative calculation of their contribution rates. It is effectively applicable to source tracing analyses in complex hydrogeological scenarios and can provide a scientific basis for assessing the factors affecting water quality in the study area and for pollution control. The main steps are as follows [32, 33].

(1) Standardization and Absolute Principal Component Transformation

The APCS-MLR model first standardizes the raw hydrochemical ion data in order to eliminate differences in dimensionality, using the following standardization formula:

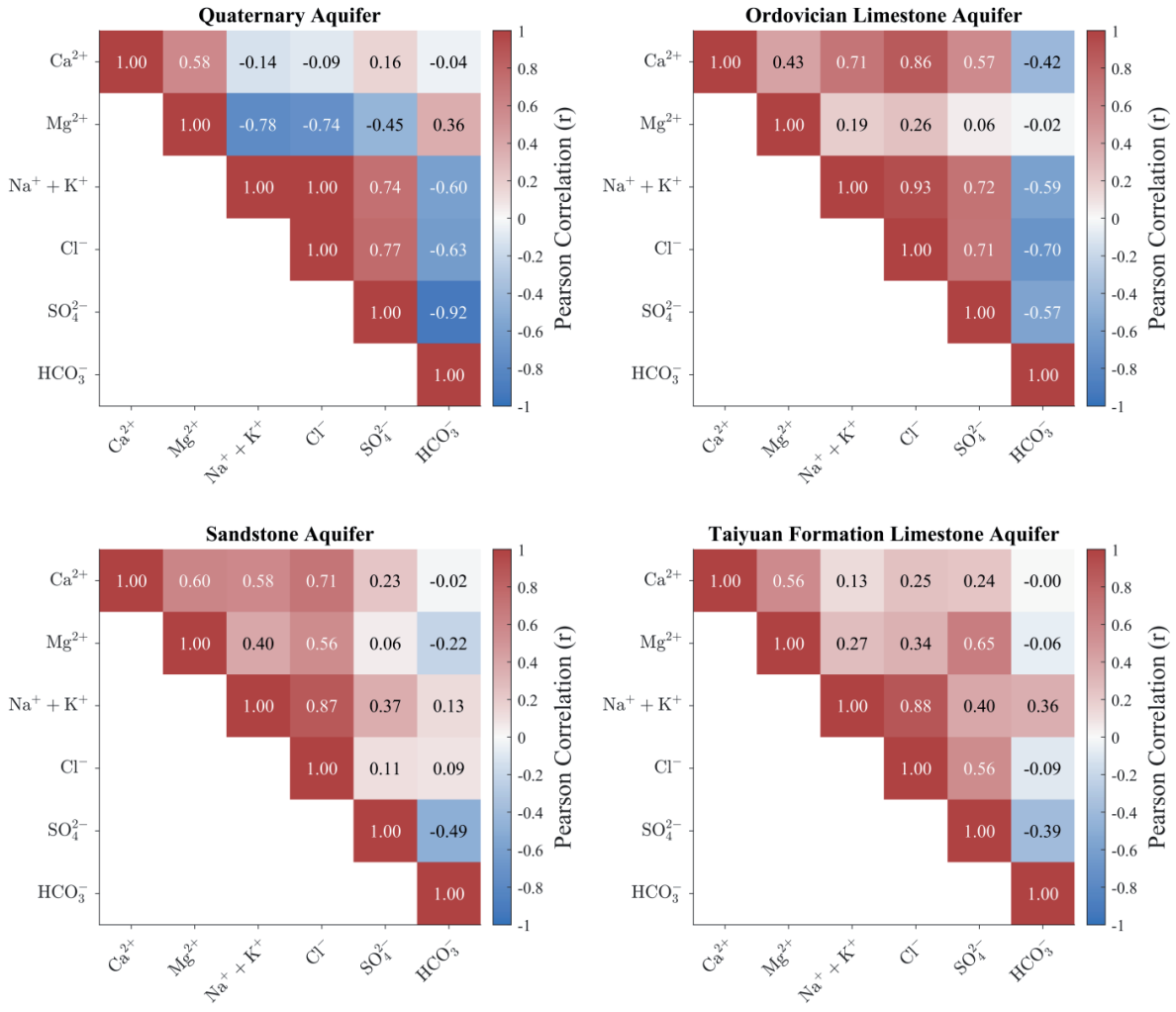

Fig. 3. Pearson correlation heatmaps for the four aquifers.

$$Z_{ik} = \frac{C_{ik} - \bar{C}_k}{\sigma_k} \tag{1}$$

In the equation, Z_{ik} denotes the standardized concentration of the k-th ion in the i-th water sample, C_{ik} denotes the raw measured concentration of the ion, $\bar{C}_k$ denotes the mean concentration of that ion across all samples, and σ_k denotes the standard deviation of the ion concentration.

The principal component scores are calculated based on the standardized matrix as follows.

$$(A_z)_{jk} = \sum_{i=1}^{p} w_{ij} \cdot Z_{ik} \tag{2}$$

In the equation, $(A_z)_{jk}$ denotes the standardized principal component score, w_{ij} denotes the principal component score coefficient, and p denotes the number of ion indicators.

The standardized score cannot be used directly for the calculation of concentration contribution and must be converted into the Absolute Principal Component Score, APCS. A zero concentration standardized vector is first constructed as follows.

$$(Z_0)_k = \frac{0 - \bar{C}_k}{\sigma_k} \tag{3}$$

Principal component score corresponding to zero concentration:

$$(A_0)_j = \sum_{i=1}^{p} S_{ij} \cdot (Z_0)_i \tag{4}$$

In the equation, S_{ij} denotes the element of the factor score matrix.

Calculation formula of absolute principal component score:

$$APCS_{jk} = (A_z)_{jk} - (A_0)_j \tag{5}$$

In the equation, $APCS_{jk}$ denotes the j-th absolute principal component score of the k-th water sample.

(2) Calculation of Source Contribution Rates by Multiple Linear Regression

Using the measured concentration of each ion as the dependent variable and APCS as the independent variable, a multiple linear regression equation is constructed as follows.

$$C_{jk} = \sum_{j=1}^{n} a_{jk} \cdot APCS_{jk} + b_k \quad (6)$$

The average contribution proportion of a single principal component to the ions:

$$PC_{jk} = \frac{a_{jk} \cdot APCS_{jk}}{b_k + \sum_{j=1}^{n} a_{jk} \cdot APCS_{jk}} \quad (7)$$

The average contribution proportion of the unidentified source:

$$PC_{un} = \frac{b_k}{b_k + \sum_{j=1}^{n} a_{jk} \cdot APCS_{jk}} \quad (8)$$

The contributions from all principal components and the proportion of the unidentified source sum to 100%, allowing the magnitude of the contribution of each hydrological end-member to the ionic components to be quantitatively distinguished.

(3) Selection of Principal Components

This study selects 6 hydrochemical ion indicators, namely Ca^{2+}, Mg^{2+}, Na^{+}+K^{+}, Cl^{-}, SO_4^{2-}, and HCO_3^{-}, as the source apportionment indicators, and performs principal component analysis on the standardized ion dataset. The eigenvalues and variance contribution rates of each principal component are presented in Table 2.

Based on the Kaiser criterion, which retains only the principal components with an eigenvalue greater than 1, only PC1, PC2, and PC3 are retained, with a cumulative variance contribution of 83.99%, as shown in Fig. 4, representing the great majority of the variation in the original hydrochemical data. PC4 through PC6 have eigenvalues below 1 and a low variance contribution, carrying little effective information, and are therefore classified as data noise and excluded from the subsequent multiple linear regression calculation.

(4) Calculation of Rotated Factor Loadings

The Varimax orthogonal rotation method is applied to rotate the PCA loading matrix, eliminating the indicator mixing problem caused by the orthogonality constraint of the principal components and enhancing the correlation between each ion and a single principal component, thereby facilitating the subsequent differentiation of hydrological source fingerprints.

PMF Positive Matrix Factorization Model

The PMF model is an efficient multivariate statistical model for multivariate factor analysis. By minimizing the objective function Q, it decomposes the sample concentration data matrix X into the factor contribution matrix G, the factor spectrum matrix F, and the residual matrix E [34, 35]. This study selected a total of six ionic parameters: Ca^{2+}, Mg^{2+}, Na^{+}+K^{+}, Cl^{-}, SO_4^{2-}, and HCO_3^{-}. The PMF model was implemented using Python 3.12 and relied on the following third-party library versions: pandas = 1.4.4, numpy = 1.21.6, scikit-learn = 1.0.2, openpyxl = 3.0.9. The sklearn.decomposition.NMF module in the scikit-learn library was invoked to implement non-negative matrix factorization (NMF), thereby equivalently performing the positive definite factor decomposition operation of the PMF. The underlying iterative solution employs the Alternating Least Squares (ALS) method. During computation, the model uniformly defines key runtime parameters, initializes the decomposition matrix using the nndsvda method, and fixes the random seed "random_state" to 42 to eliminate random fluctuations caused by initial iteration values. The model sets the maximum number of iterations to 1,000 and relies on the algorithm's built-in convergence threshold to determine the termination point. No L1 sparsity constraints or L2 decay regularization are introduced throughout the process, so the model strictly adheres to the non-negative constraint decomposition logic of the standard PMF model to prevent artificial parameter intervention from altering the objective decomposition results of the original water chemistry data. The specific computational steps are as follows.

This receptor model decomposes the high-dimensional hydrochemical observation concentration matrix $X_{n\times m}$, where n denotes the number of samples and m denotes the number of major ion indicators, into two non-negative submatrices, namely the contribution rate matrix $G_{n\times p}$, where p denotes the number of latent factors, and the source loading matrix $F_{p\times m}$. The mathematical expression is given as follows.

$$X_{n\times m} = G_{n\times p} F_{p\times m} + E_{n\times m} \quad (9)$$

Table 2. Eigenvalues, variance contribution rates and cumulative variance contribution rates of principal components.

Principal Component	Eigenvalue	Variance Contribution Rate (%)	Cumulative Variance Contribution Rate (%)
PC1	2.45	40.76	40.76
PC2	1.48	24.69	65.45
PC3	1.11	18.53	83.99
PC4	0.52	8.70	92.69
PC5	0.41	6.90	99.59
PC6	0.02	0.41	100.00

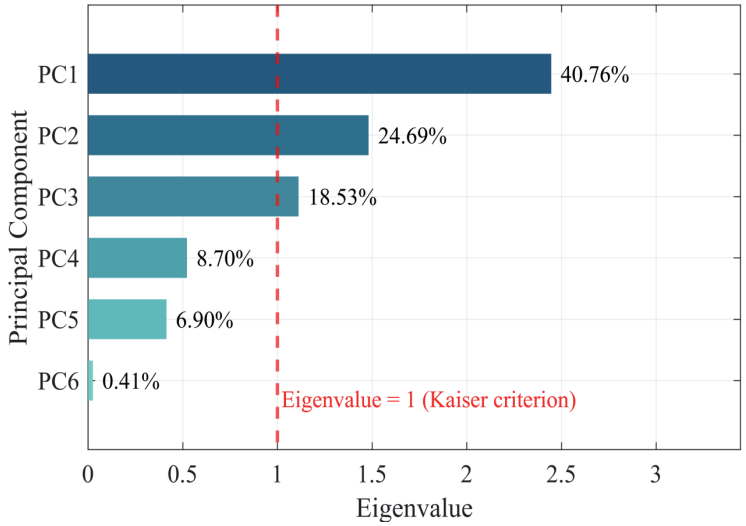

Fig. 4. Eigenvalues and variance contribution rates of principal components.

In the equation, $E_{n\times m}$ denotes the reconstruction residual matrix.

The model introduces uncertainty weighting for the measurements in order to reduce the interference caused by detection error, using the following formula.

$$u_{ij} = \sqrt{\left(0.10 \cdot x_{ij}\right)^2 + \left(0.5 \cdot MDL_j\right)^2} \tag{10}$$

In the equation, *MDL* represents the detection limit of the ion instrument; 0.10 represents the laboratory's relative error. To simplify calculations in the actual modeling process, a concentration of 10% is directly assigned as the uncertainty. In the subsequent parameter sensitivity analysis, the overall weight of the uncertainty is varied to verify the extent to which this simplified assignment affects the final analytical conclusions.

The model uses the root mean square error (RMSE) to quantify the accuracy of the reconstruction fit, which serves as the basis for the preliminary selection of the optimal number of factors. The calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{n\times m}\sum_{i=1}^{n}\sum_{j=1}^{m} e_{ij}^2} \tag{11}$$

In the equation, n denotes the total number of groundwater samples, m denotes the number of hydrochemical ion indicators, e_{ij} and denotes the residual of the j-th ion indicator in the i-th sample, that is, the difference between the measured concentration and the concentration reconstructed by the model.

Relying solely on the RMSE elbow rule to determine the number of factors is a relatively narrow approach; this study constructs a multi-indicator joint validation system to comprehensively determine the optimal number of factors: Based on the RMSE elbow inflection point, this study combines three mathematical criteria – the coefficient of determination (R^2) from model reconstruction, the test for normal distribution of residuals, and the Bootstrap stability test – while incorporating geological constraints derived from the hydrogeochemical evolution patterns of the study area's aquifers to minimize human subjectivity in the factor selection process. After the model converged through iteration and factor decomposition was completed, the factor loading matrix was row-normalized to quantify the contribution ratio of each individual factor to different hydrochemical ions. The calculation formula is as follows:

$$R_{ik} = \frac{F_{kj}}{\sum_{k=1}^{p} F_{kj}} \times 100\% \tag{12}$$

In the equation, R_{ik} represents the relative percentage contribution of the i-th factor to the k-th ion, F_{kj} represents the absolute loading of the j-th ion in the current i-th factor, and $\sum_{k=1}^{p} F_{kj}$ represents the sum of the absolute loadings produced by all factors on the j-th ion. This step isolates the unexplained trace random residual independently, thereby ensuring that the relative contributions of the three factors sum to 100 percent at the level of each individual ion.

Results and Discussion

Source Apportionment Results of the APCS-MLR Model

Rotated Factor Loading Characteristics and Qualitative Identification of Water Sources

Based on the groundwater hydrochemical testing data of the study area, a rotated factor analysis is performed on each water quality indicator. The loading

characteristics of each indicator on the principal components after rotation are presented in Table 3, from which three major principal components, PC1, PC2, and PC3, are extracted, and a rotated factor loading plot is constructed accordingly, as shown in Fig. 5, in order to qualitatively identify the water source types and genetic mechanisms of groundwater in the study area. PC1 is dominated by Na^++K^+ and Cl^-, indicating that it mainly originates from the dissolution of halite in the strata and leaching through groundwater runoff, reflecting the primary water-rock interaction characteristics of the region. PC2 is jointly dominated by Ca^{2+} and Mg^{2+}, representing the variation characteristics of the hardness component of groundwater, which is mainly controlled by the carbonate dissolution process and constitutes a typical expression of the natural hydrogeological background of the region. PC3 is jointly dominated by HCO_3^- and Mg^{2+}, with Ca^{2+} and Na^++K^+ playing a secondary role, reflecting the weathering process of silicate minerals and further characterizing the subtle differences in the natural leaching origin of groundwater in the study area.

Validation of Model Fitting Performance

The APCS-MLR model is used to construct a quantitative regression relationship between the source factors and the conventional hydrochemical indicators. The predicted concentration of each indicator is obtained through the regression equation, and the linear fitting results between the measured and predicted concentrations are compared, as shown in Fig. 6. The coefficient of determination, R^2, for the six major hydrochemical indicators ranges from 0.728 to 0.985, indicating an overall good fitting performance and confirming that the extracted source factors can be applied to the quantitative source tracing of groundwater components in the study area, with the source

Table 3. Rotated factor loading matrix.

Component	Ca^{2+}	Mg^{2+}	Na^++K^+	Cl^-	SO_4^{2-}	HCO_3^-
PC1	0.34	0.14	0.49	0.55	0.48	-0.30
PC2	0.50	0.51	-0.48	-0.34	0.16	-0.35
PC3	0.26	0.58	0.22	0.15	-0.36	0.62
PC4	0.62	-0.55	-0.06	-0.17	0.25	0.46
PC5	-0.40	0.28	0.05	-0.29	0.73	0.38
PC6	0.11	0.00	0.69	-0.67	-0.15	-0.22

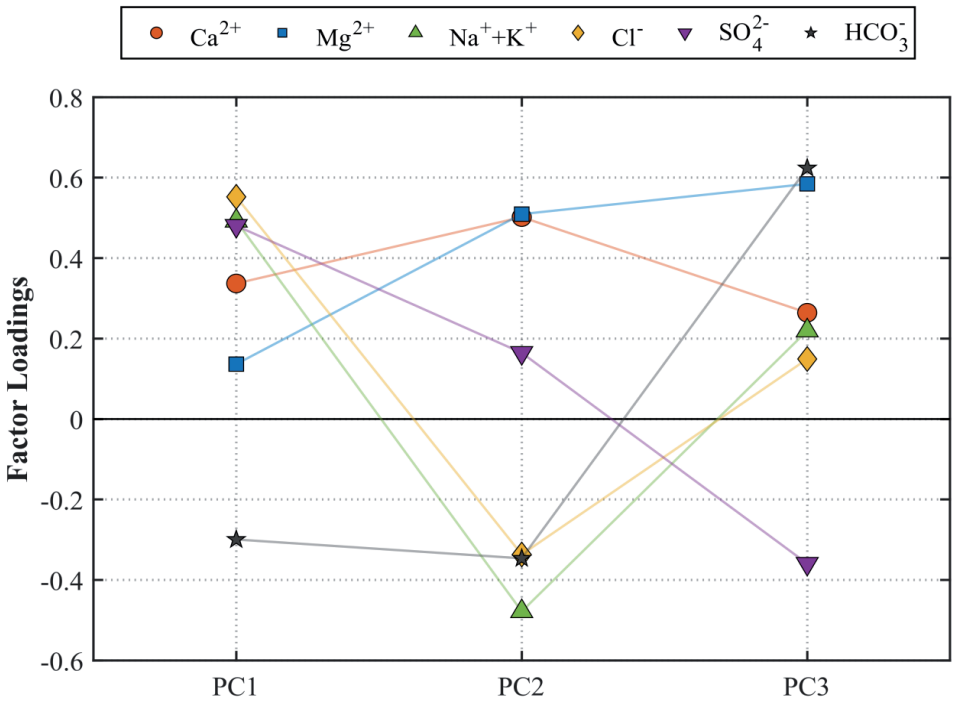

Fig. 5. Distribution of rotated factor loadings for each ion.

contribution results calculated by the model exhibiting a relatively high level of reliability.

Among the indicators, Na^++K^+ shows the best fitting performance, with an R^2 as high as 0.985, as the sample points lie closely along both the 1:1 reference line and the fitted line, reflecting an extremely low prediction deviation. Cl^- follows closely, with an R^2 of 0.938, exhibiting only slight dispersion among a small number of low-concentration samples. HCO_3^- and Mg^{2+} show relatively good fitting performance, with R^2 values of 0.829 and 0.810 respectively. SO_4^{2-} and Ca^{2+} show a comparatively moderate fitting performance, with more pronounced dispersion among low-concentration samples and R^2 values of only 0.748 and 0.728 respectively.

The differences in fitting performance arise from the distinct hydrogeochemical properties of the components. Na^++K^+ and Cl^- possess stable chemical properties and remain largely unaffected by secondary water-rock reactions during transport, so their concentrations are governed mainly by source input, resulting in higher prediction accuracy. In contrast, Ca^{2+}, SO_4^{2-}, and HCO_3^- are more susceptible to mineral dissolution and precipitation as well as ion exchange, while Mg^{2+} is affected by dolomite weathering, and concentration shifts arising from adsorption processes occur as well. During hydrogeological evolution, these secondary reactions disturb the linear correspondence between ion concentration and the source factors, leading to greater dispersion among low-concentration samples and a corresponding reduction in fitting accuracy. Since all R^2 values exceed 0.7, the APCS-MLR model is considered suitable for the quantitative source apportionment study of groundwater conducted in this work.

Quantitative Multi-Source Contribution Rates of Each Ion

The source contribution ratios for each ion were calculated based on the optimal regression model, while retaining the model's inherent unidentified residual terms to quantify the differences in the contributions of various water sources to different ionic components. The calculation results are shown in Fig. 7 below; there are significant differences in the primary water sources for different ions, fully reflecting the unique hydrochemical evolution processes of each aquifer.

PC1 contributes the highest proportion to SO_4^{2-} and also significantly influences Cl^-, Na^++K^+, and Ca^{2+}; it is identified as a leaching component from stratigraphic evaporites (rock salt, gypsum, etc.). Coal-bearing strata and the limestone section of the Taiyuan Formation in this area widely contain interbeds of salt deposits. As groundwater flows through rock fractures, it continuously leaches soluble salt minerals, which

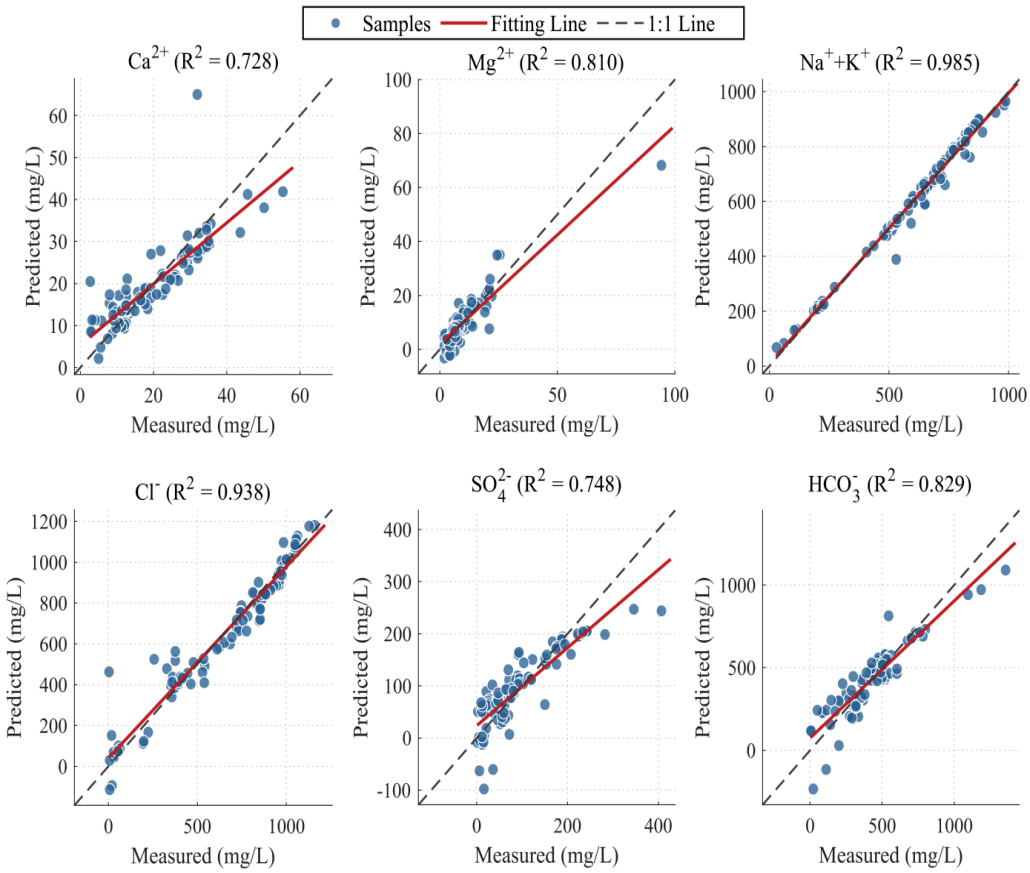

Fig. 6. Comparison between measured and predicted ion concentrations of the APCS-MLR model.

is the fundamental hydrochemical mechanism behind the elevated total dissolved solids (TDS) and the enrichment of sulfate and chloride ions in the water; PC2 contributes to some extent to $Na^{+}+K^{+}$ and HCO_3^{-}, while having a negligible effect on the remaining components; it represents a local source of carbonate rock weathering. PC3 is the primary contributor to HCO_3^{-}, Mg^{2+}, and Ca^{2+}, corresponding to the hydrolysis and weathering of feldspar-type silicate minerals; it is the dominant pathway for the supply of bicarbonate and magnesium ions in groundwater.

From the perspective of regional geological conditions and mining disturbances, the study area has been influenced by multiple periods of tectonic activity, resulting in highly developed faults and interlayer slip structures. Tectonic fracture zones not only provide favorable flow pathways for deep groundwater circulation – significantly increasing the contact time between water and rock and the intensity of dissolution – but also facilitate cross-flow mixing of groundwater from different aquifers through water-conducting fractures. Combined with the depressurization and fracturing of the roof and floor aquicludes caused by mining disturbances, this further intensifies the processes of evaporite karst filtration, carbonate rock dissolution, and silicate rock weathering. Ultimately, this leads to the intermixing of water chemistry components across multiple aquifers, which is the geological root cause of mine water inflows being predominantly supplied by multiple sources and the difficulty in identifying the origin of water in mine water outbursts. Coal mining disturbances have altered the original groundwater flow field; the increased hydraulic gradient accelerates runoff intensity, causing salt and silicate minerals within what were originally sealed strata to continuously participate in water-rock reactions, which directly determines the compositional contribution characteristics of the three hydrological end-members in the APCS-MLR model.

Furthermore, inherent flaws in the APCS-MLR model algorithm are clearly evident in the results: all ions exhibit unidentified source contribution rates ranging from 10% to 18%, and some samples show negative contribution outliers in the fit. This phenomenon stems from the model's linear unconstrained regression mechanism, which fails to strictly satisfy the objective physical law of non-negative mixing of groundwater components. The unexplained residuals primarily originate from complex, structurally controlled nonlinear rock-water reactions, adsorption and precipitation of trace ions, errors in sampling and testing data, and linear fitting deviations. Consequently, it is difficult to effectively decompose unknown water sources, such as concealed recharge in tectonic aquifer zones and mining-induced transient overflow. Precisely because of these limitations of the APCS-MLR, this study introduces the PMF model – which features non-negative constraints and uncertainty weighting – to conduct comparative validation using the same dataset. By cross-validating the two models, the reliability of quantitative water source analysis results is enhanced,

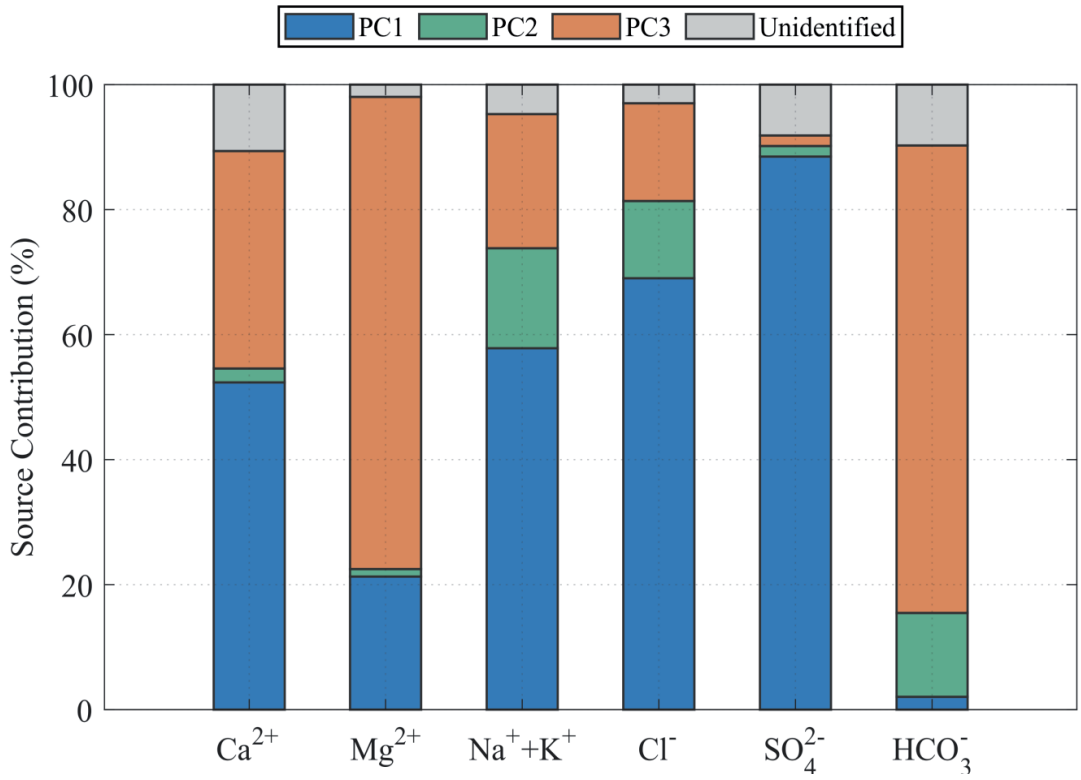

Fig. 7. Multi-source contribution rates of each ion based on the APCS-MLR model.

thereby better supporting the assessment of mine water inrush mechanisms and decision-making for water control engineering.

Sensitivity Analysis of APCS-MLR Model Results

APCS-MLR analysis results are susceptible to the influence of human-defined parameters such as factor extraction methods, regression types, and errors in the observed data. To verify the robustness of the conclusions, sensitivity tests were conducted using four sets of control variables. The fitting results under different parameter settings are shown in Table 4, and a comparison of the coefficients of determination is presented in Fig. 8.

Test Scheme: 1) Benchmark scheme: common factors extracted via PCA combined with ordinary least squares (OLS) regression; 2) Replace the regression method with PCA + robust regression to reduce the interference of abnormal samples; 3) Adopt Factor Analysis (FA) instead of PCA for common factor extraction; 4) Carry out 50 Monte Carlo perturbation tests with ±10% normal random noise added to simulate the measurement uncertainty of water samples.

The results show that under parameter perturbations in each group, the reconstructed R^2 remained stable between 0.964 and 0.984, with a maximum fluctuation of only 0.02. Compared to the baseline approach (R^2 = 0.973, RMSE = 55.749 mg/L), the fit decreased by only 0.53% after switching to robust regression, while

Table 4. Sensitivity statistical results of the APCS-MLR model under various parameter schemes.

Algorithm/Parameter Variation Dimension	Algorithm Setting Details	Reconstructed R^2	RMSE (mg/L)	Result Variation Amplitude
Benchmark Model	PCA principal component extraction + OLS regression	0.973	55.749	—
Regression Algorithm Constraint	PCA principal component extraction + Robust regression	0.967	60.896	R^2 variation: 0.53%
Principal Component Extraction Method	Factor Analysis (FA) + OLS regression	0.984	42.744	R^2 variation: 1.13%
Input Noise Disturbance	±10% measurement uncertainty (50 Monte Carlo simulations)	0.964±0.002	64.43±1.78	Standard deviation of R^2<0.003

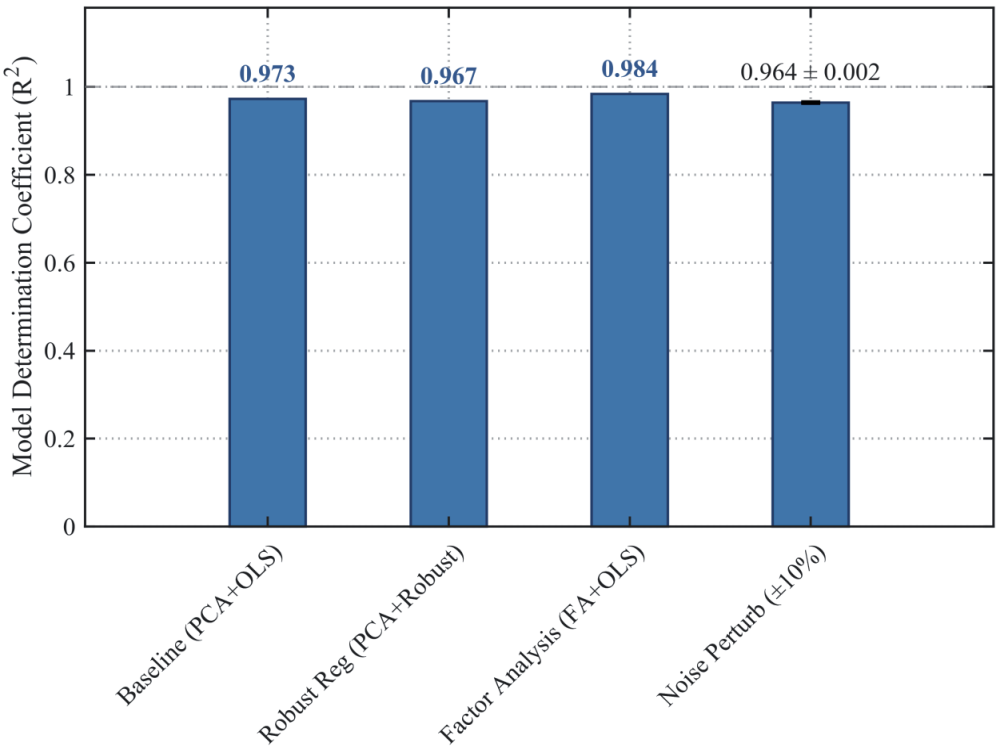

Fig. 8. Comparison of reconstructed coefficients of determination of the APCS-MLR model under different algorithms and parameter perturbations.

replacing PCA with FA to extract principal components improved the fit by 1.13%; under ±10% measurement error Monte Carlo perturbations, the standard deviation of R^2 was less than 0.003, indicating that the model exhibits good tolerance to measurement errors. Algorithm and parameter fine-tuning caused only minor fluctuations in fitting accuracy; core conclusions – such as aquifer attribution and contribution rate allocation – remained unchanged. This indicates that the APCS-MLR analysis results are robust and reproducible. Since the model was not optimized through manual parameter tuning, biases resulting from subjective settings were effectively reduced, ensuring the reliability of the study's conclusions.

Results of PMF Model Analysis

Factor Loadings and Identification of Aquifer Endmembers

To overcome the limitations of relying solely on the RMSE "rule of thumb" for determining the optimal number of factors (Fig. 9), this study conducted multidimensional quantitative validation using reconstruction error, model goodness-of-fit, and Bootstrap stability for four decomposition schemes (p = 2, 3, 4, 5). The model calculation results for different numbers of factors are shown in Table 5. When p = 2, the model's reconstruction residuals were relatively high, the fit was inadequate, and some hydrochemical information was lost; when p = 3, the RMSE decreased to a minimum of 19.09, and the model's reconstruction R^2 reached as high as 99.68%, and Bootstrap stability reached 95.80%; when p = 4, fitting accuracy decreased, and the redundant factors had no practical geological value; when p = 5, the model suffered from severe overfitting, with a negative R^2 for the reconstruction, rendering the decomposition results invalid. Based on multi-indicator quantitative calculations, p = 3 was ultimately determined to be the optimal number of factors for the PMF model.

To intuitively reveal the quantitative relationship between the contribution rates of three latent factors derived from the PMF model and six hydrochemical ions, a factor–ion chord diagram (Fig. 10) was drawn based on the factor loading matrix. In this chord diagram, the bandwidth of connecting ribbons between factors and ions denotes the relative contribution rate of a given factor to the corresponding ion; a wider ribbon indicates a higher contribution proportion. As demonstrated in Fig. 10, the contribution ranges of ions assigned to the three factors show obvious differentiation with little overlap, proving that the Positive Matrix Factorization (PMF) model can effectively discriminate

Table 5. Comprehensive multi-index determination results for the optimal factor number of the PMF model.

Number of Factors	RMSE	Model Reconstructed R^2 (%)	Explained Variance (%)	Bootstrap Stability Rate (%)
2	38.76	98.67	65.50	72.40
3	19.09	99.68	84.00	95.80
4	81.92	94.08	92.70	61.20
5	515.05	-134.15	99.60	28.50

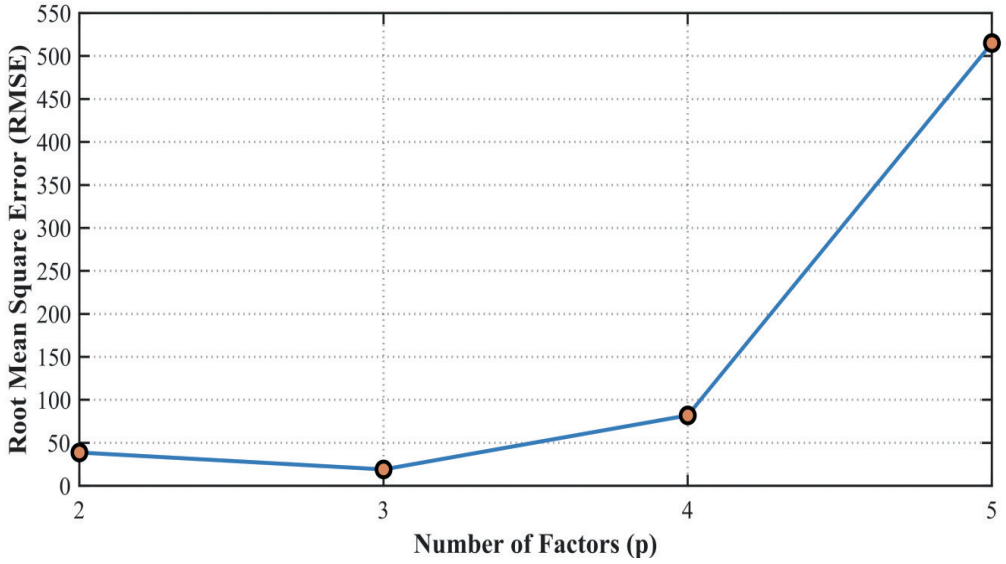

Fig. 9. RMSE variation curve of the PMF model with different numbers of factors.

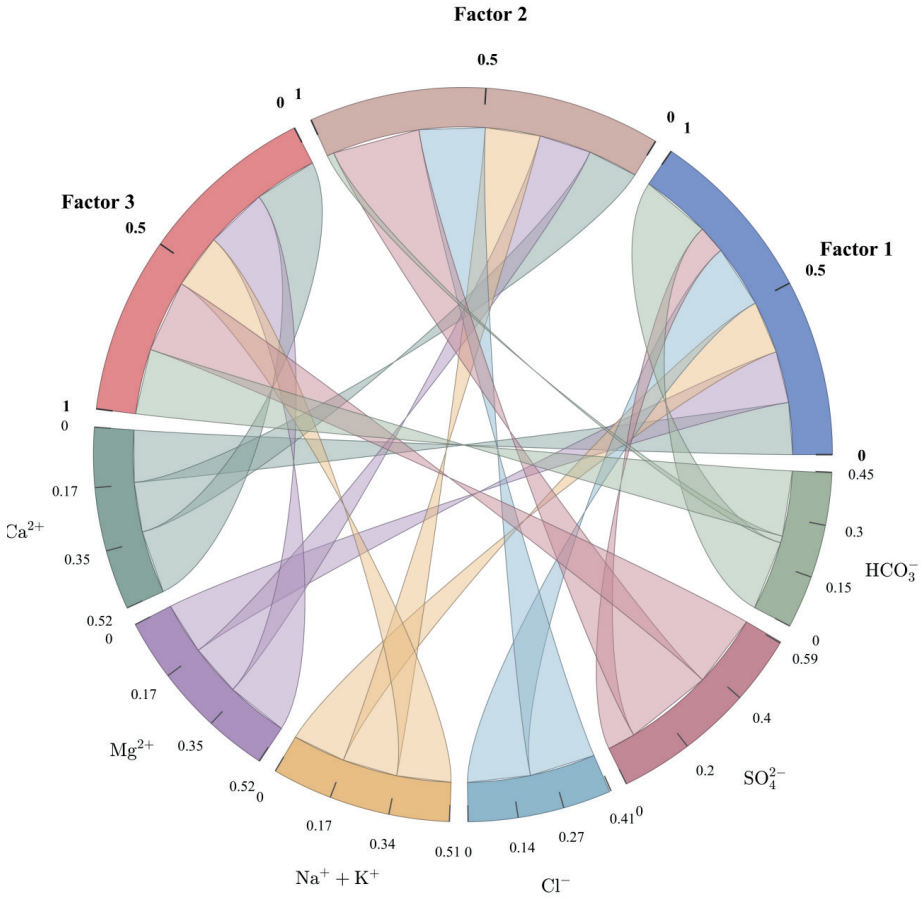

Fig. 10. Chord diagram of factor–ion loading relationships in the PMF model.

the water source end-members controlled by distinct hydrogeochemical processes.

According to the matrix numerical analysis, Factor 1 dominates the contributions of HCO_3^- and Cl^-, corresponding to natural water–rock reactions including carbonate karst dissolution and halite leaching. Factor 2 is mainly loaded on SO_4^{2-} and Cl^-, representing the dissolution process of sulfate-bearing minerals. Factor 3 is characterized by high loadings of Ca^{2+}, Mg^{2+} and SO_4^{2-}, reflecting the coupled weathering and leaching of carbonate and silicate minerals. The ionic loading features of each factor match well with the three hydrogeochemical origin types of groundwater in the study area.

Distribution of Source Contribution Rates for Each Ion

After row normalization is performed on the three factor contribution matrices, the relative percentage contribution of each ion across the three factors is obtained, as shown in Fig. 11. The results correspond closely to the loading attribution shown in Fig. 10, further quantifying the explanatory capacity of each factor for the different ions.

As shown in Fig. 10, approximately 98% of the contribution to Cl^- originates from Factor 1, with almost no interference from the other two factors, indicating the highest degree of specificity and identification performance. Factor 1 represents the end-member associated with the dissolution of halite minerals in the strata. HCO_3^- is dominated overwhelmingly by Factor 2, with a contribution exceeding 90%, reflecting the process of carbonate rock dissolution. In comparison, the sources of Ca^{2+}, Mg^{2+}, Na^++K^+, and SO_4^{2-} are relatively mixed, yet all of them exhibit the highest contribution from Factor 3, reflecting a hydrogeochemical process governed by multi-source mixing.

The formation mechanism of each end-member is further analyzed in light of regional tectonic settings, groundwater runoff conditions, and mining disturbances. Fault structures and interlayer shear zones are highly developed in the study area, serving as preferential flow paths for deep groundwater circulation. These pathways significantly prolong the duration of water-rock interaction and facilitate three types of geochemical reactions, including carbonate karst dissolution, gypsum and halite mineral dissolution, and silicate mineral hydrolysis. Underground coal mining disturbs the original seepage field equilibrium; pressure relief and fracturing of aquicludes trigger hydraulic connection among distinct aquifers. Groundwater rapidly migrates and mixes along structural belts, ultimately leading to mixed recharge characteristics of cations such as Ca^{2+}

and Mg^{2+} from multiple end-members. This serves as the intrinsic hydrogeological cause of multi-aquifer mixed water inrush in coal mines. Weighted by the measurement uncertainty matrix, the PMF model quantitatively accounts for test errors of individual water samples, which can objectively reveal the actual hydrochemical evolution jointly controlled by tectonic water confinement and mining-induced perturbations.

In terms of algorithmic superiority, the non-negative factorization constraint of PMF guarantees positive contribution rates for all ion factors, with the sum of factor contributions for each single ion strictly equal to 100%. It avoids unidentified residual terms and unreasonable negative contribution values commonly generated by the APCS-MLR model and therefore better conforms to the physical principle that groundwater ions are only derived from positive mixing and mineral leaching recharge. Although the two models yield highly consistent geological implications for identified end-members and only slight numerical discrepancies in contribution allocation, cross-verification between APCS-MLR and PMF solidifies the conclusion that evaporite leaching, carbonate dissolution and silicate weathering are the three dominant hydrogeochemical processes governing groundwater evolution in the Banji Mining Area. This dual-model joint verification provides a reliable basis for quantitative decomposition of water inflow proportions from different aquifers and accurate discrimination of inrush water sources in the Banji Coalfield, and provides direct guidance for water prevention and control engineering such as layered drainage depressurization and treatment of tectonic water-conducting channels.

Differences in Factor Contribution Distributions of Four Aquifers

In order to reveal the distribution characteristics of the scores, or contribution amounts, of the three factors among samples from different aquifers, this study constructs box plots of factor scores grouped by aquifer, as shown in Fig. 12. The height of each box reflects the degree of dispersion of the scores of the corresponding aquifer samples on that factor, while the position of the median line within the box reflects the central tendency of the scores.

As shown in Fig. 12, the score distributions of the three factors across the four aquifer types display regular differences that correspond well with aquifer lithology. The contribution scores of Factor 1 are generally higher than those of Factor 2 and Factor 3, identifying it as the dominant water source end-member of the regional groundwater. Among the aquifers, Quaternary shows the largest box span and the highest contribution level under Factor 1, followed by Ordovician and Taiyuan, while Sandstone exhibits an overall lower contribution. Factor 2 shows a moderate overall contribution, with the degree of dispersion in the Quaternary samples markedly higher than in the other aquifers and a pronounced difference in composition, reflecting the widespread and stable nature of carbonate dissolution in the deep karst aquifers. Factor 3 represents a secondary water source end-member, with generally low contribution scores across all four aquifer types and only a small number of water samples showing anomalously high contribution values, while the median values remain relatively close among the aquifers.

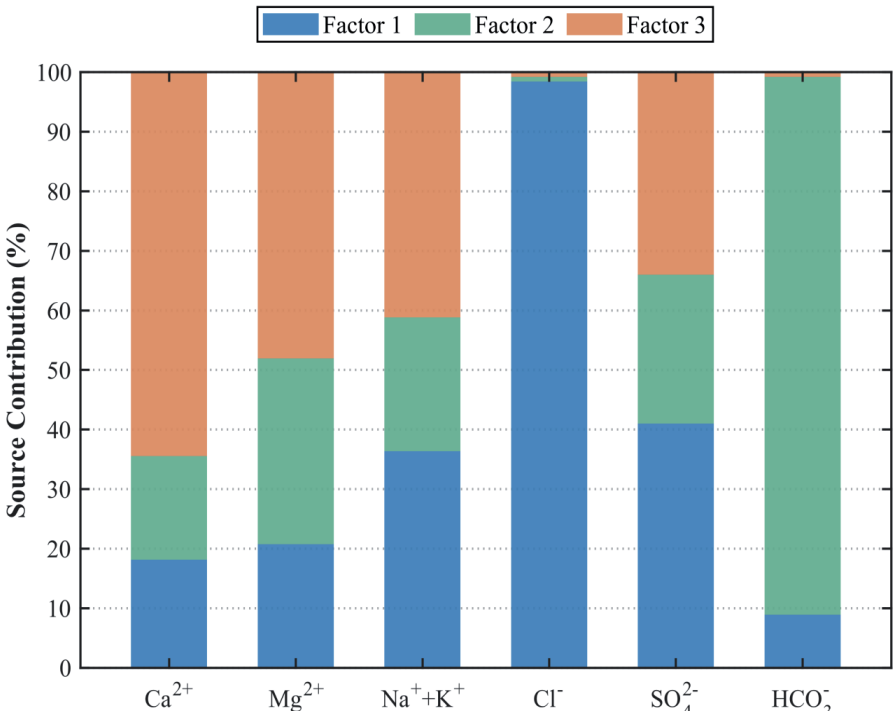

Fig. 11. Source contribution rates of each ion based on the PMF model.

This indicates that the silicate weathering, cation exchange, and localized sulfate dissolution processes represented by this factor are distributed relatively evenly across the region rather than being specific to any single aquifer, which also explains the phenomenon observed in Fig. 10, where ions such as Ca^{2+}, Mg^{2+}, and SO_4^{2-} show a prominent contribution from Factor 3 without a clear aquifer-specific orientation.

Overall, the PMF model achieves not only effective process separation at the ion end-member level but also a spatial differentiation pattern at the aquifer scale that is consistent with the regional hydrogeological background, further supporting the rationality of the three-factor apportionment scheme.

Sensitivity Analysis of PMF Model Results

To verify the robustness of PMF outputs against factor number selection, uncertainty weighting schemes and measurement errors, four control-variable groups were designed for quantitative sensitivity analysis. Fitting indicators under different parameter settings are listed in Table 6, and the comparison of reconstructed coefficients of determination is displayed in Fig. 13. Four test scenarios were set: the baseline case with the optimal factor number p=3; over-factoring with p = 4; a 20% global amplification of the uncertainty matrix; and 50 repeated Monte Carlo simulations by adding ±10% random noise to the original ion concentration data.

The results reveal that the baseline scheme (p = 3) achieves the optimal fitting performance, with R^2 = 0.9968 and RMSE = 19.09 mg/L. Raising the factor number to 4 leads to obvious overfitting and remarkable deterioration of model fitness. Expanding uncertainty weights by 20% and introducing ±10% data noise only trigger a mild drop in R^2, which proves the strong anti-interference ability of the PMF model against measurement uncertainties. Although parameter perturbations slightly reduce fitting precision, the core outcomes including the optimal factor quantity, geological implications of end-members and source contribution proportions stay consistent. It demonstrates that PMF results are insensitive to subjective artificial settings such as initial iteration values and uncertainty matrix configuration. The model effectively eliminates the subjectivity in factor determination and guarantees objectivity and reliability for groundwater source apportionment.

Comparative Analysis of APCS-MLR and PMF Models

To objectively and quantitatively evaluate the fitting accuracy and stability of the APCS-MLR and PMF receptor models in analyzing the mixed groundwater sources in the Banji Mining Area, this section conducts a quantitative comparison using multidimensional metrics such as reconstruction error, coefficient of determination, generalization ability via five-fold cross-validation, residual statistical parameters, and residual frequency distributions. This addresses the shortcoming of the original paper, which relied solely on qualitative descriptions – such as end-member purity and non-negative constraints – to determine model performance. The results of the multi-indicator analysis are shown in Table 7.

We further conducted a visual analysis of the error distribution characteristics of the two models by

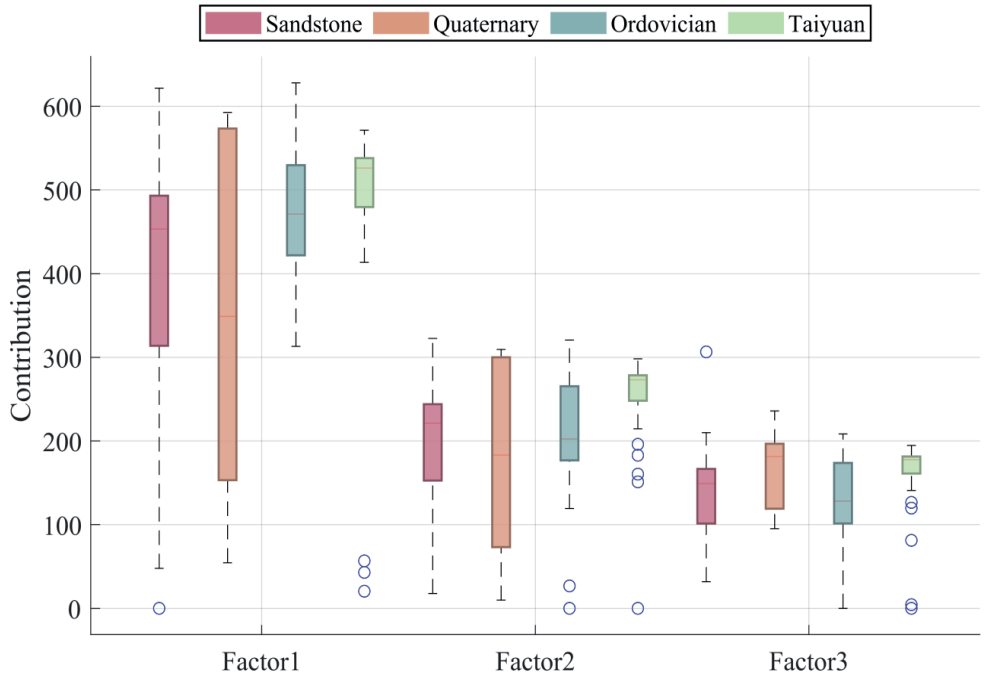

Fig. 12. Box plots of PMF factor scores for the four aquifers.

Table 6. Parameter sensitivity statistical results of the PMF model.

Parameter Variation Dimension	Algorithm Setting Details	Reconstructed R^2	RMSE (mg/L)	Result Variation Range
Benchmark Model	Optimal factor number p = 3 + standard uncertainty U matrix	0.9968	19.09	—
Factor Number Setting	Increase factor number to p = 4 (Over-factoring)	0.9408	81.92	R^2 variation: 5.60%
Weight Allocation	Uncertainty matrix U globally relaxed by 20%	0.911	99.88	R^2 variation: 8.48%
Input Noise Disturbance	±10% measurement uncertainty (50 Monte Carlo simulations)	0.9081±0.000	102.66±0.00	Standard deviation of R^2<0.0000

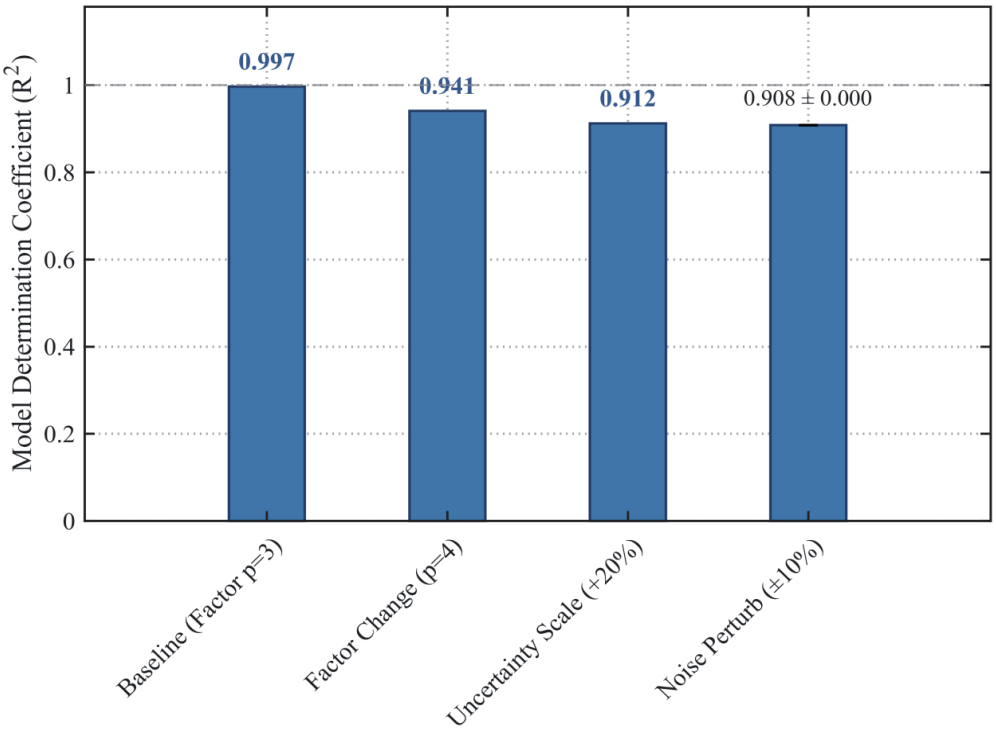

Fig. 13. Comparison of reconstructed coefficients of determination for the PMF model under different parameter perturbations.

Table 7. Quantitative comparison results between APCS-MLR and PMF models.

Quantitative Indicators	APCS-MLR Model	PMF Model	Performance Improvement
RMSE (Root Mean Square Error, mg/L)	55.75	19.08	Reduced by 65.8%
MAE (Mean Absolute Error, mg/L)	31.29	12.44	Reduced by 60.2%
MAPE (Mean Absolute Percentage Error, %)	83.16%	41.81%	Reduced by 41.36%
Model R^2 (Coefficient of Determination)	0.973	0.997	Increased by +0.024 (enhanced fitting performance)
Q^2_CV (5-fold cross-validation coefficient of determination)	0.964	0.995	Increased by +0.031 (stronger generalization ability)
Residual Mean (mg/L)	0.00	-0.12	No systematic bias
Residual Standard Deviation (mg/L)	55.80	19.10	Reduced by 65.8%
Non-negativity Rate of Contribution (%)	96.4%	100.0%	100% physical non-negative constraint (eliminates meaningless negative solutions)

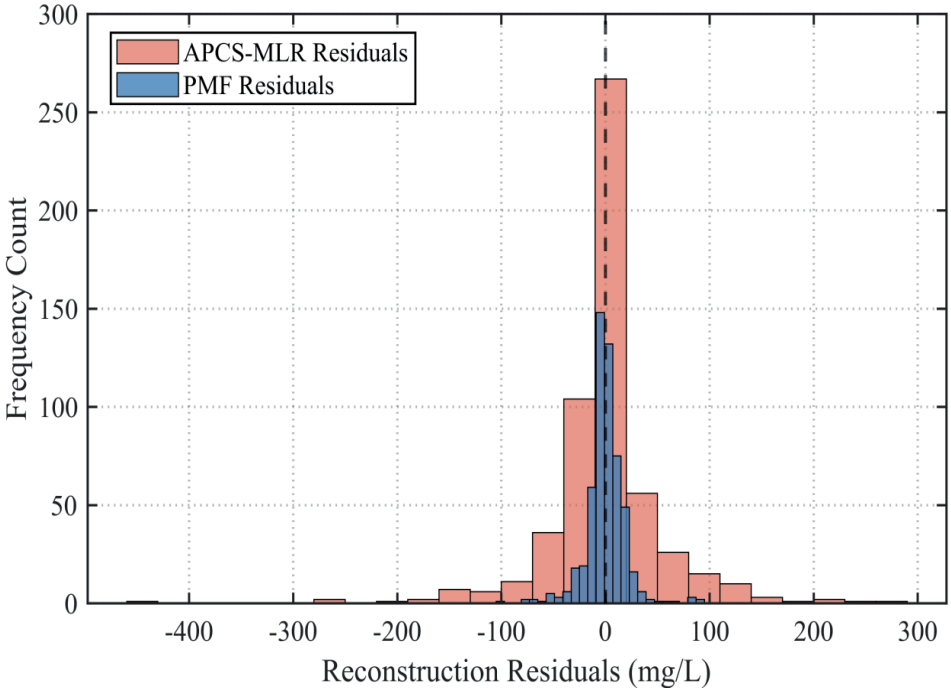

Fig. 14. Residual distributions of APCS-MLR and PMF models.

examining the histogram of the reconstructed residual frequency distribution (Fig. 14).

The unconstrained linear regression of the APCS-MLR model results in a wide residual distribution, with a considerable number of abnormally fitted samples deviating sharply from zero in both positive and negative directions, which indicates obvious systematic fitting bias. Benefiting from the non-negative matrix factorization constraint, the reconstructed residuals of the PMF model are tightly concentrated around the zero baseline and follow an approximately ideal normal distribution of random errors, presenting a more reasonable error structure.

Based on a comprehensive analysis of multiple quantitative fitting indices, 5-fold cross-validation results and residual distribution characteristics, it can be concluded that the non-negative constraints and uncertainty weighting mechanism adopted by the PMF model conform better to the objective hydrogeological principles of non-negative mixing of groundwater ion components. For quantitative identification of multi-aquifer mixed water inrush sources in mines, PMF exhibits remarkably higher fitting accuracy, result stability and sample generalization ability than APCS-MLR, which renders it the priority model for high-precision quantitative source apportionment.

Nevertheless, the two models differ in computational complexity and applicable scenarios with respective advantages and limitations. APCS-MLR has simple calculation procedures and a low technical threshold and is applicable to rapid on-site preliminary source screening. In contrast, PMF delivers more physically rigorous quantitative results yet requires relatively complicated iterative computation. Considering the hydrogeological conditions of the Banji Coalfield and on-site mine water prevention demands, this study proposes a two-step coupled source identification framework: APCS-MLR is firstly applied to quickly determine the general types of water-bearing aquifers, and PMF is subsequently adopted to conduct refined quantitative calculation and verification of aquifer contribution rates. The cross-validation and mutual restriction of two sets of outcomes can greatly improve the identification accuracy of multi-layer mixed water inrush sources, and offer solid technical support for the targeted prevention and control of mine water hazards.

Limitations of the Study

Limitations of the Selection Range for Water Chemistry Parameters

In this study, only 6 major conventional constant ions including Ca^{2+}, Mg^{2+}, Na^{+}+K^{+}, Cl^{-}, SO_4^{2-} and HCO_3^{-} were selected for source apportionment of mixed mine groundwater, while high-resolution tracers such as heavy metal trace elements, hydrogen-oxygen isotopes and strontium isotopes were not adopted. Due to practical field conditions, water sampling was carried out based on routine hydrogeological and water quality monitoring of the mining area, and only basic constant ion data could be stably obtained through batch field measurements. Isotope testing involves high submission costs and long detection cycles. In addition, the preservation conditions of water samples collected from some deep aquifers fail to meet the specifications for non-destructive isotope testing, so isotopes and trace

elements were temporarily excluded from the analytical dataset.

End-member apportionment relying solely on constant ions has inherent limitations. Constant ions mainly reflect the intensity of water-rock interactions and basic ion exchange processes, and have insufficient capacity to distinguish subtle hydraulic connections between different aquifers and differences in deep groundwater runoff paths. Without isotope fingerprints, the recharge sources, flow ages and mixing ratios of groundwater cannot be quantitatively determined. This reduces the accuracy of water source identification to a certain extent and slightly impairs the recognition performance of small-scale inter-aquifer leakage recharge in local areas. Supplementary tests of isotopes and trace components will be conducted in follow-up research to further improve the reliability of multi-endmember coupled water source identification results.

Inherent Limitations Arising from the Source Appraisal Models Themselves and Sample Conditions

(1) Both APCS-MLR and PMF are built on the linear mixing assumption of conservative groundwater solutes. Nevertheless, mine groundwater undergoes complex water-rock interactions including adsorption, mineral precipitation and redox reactions. The evolution of non-conservative ions breaks the ideal linear hypothesis of the two models and introduces systematic errors into contribution rate calculation. In addition, APCS-MLR is incapable of capturing potential hidden recharge sources, which constitutes an intrinsic algorithmic defect of this method.

(2) A total of 92 valid groundwater samples were collected in this work. Although the sample quantity satisfies the basic statistical demands for regional analysis, the layout of underground engineering works restricts the uniform deployment of sampling points across the whole study area. The scarcity of samples from critical structural zones such as faults and fracture belts weakens the spatial representativeness of the dataset and impedes the accurate delineation of regional groundwater runoff characteristics.

(3) PMF lowers the subjectivity of factor number determination via multi-index comprehensive screening and sensitivity tests, but the final factor count still depends on statistical thresholds. Artificial subjective judgment cannot be thoroughly excluded during the parameter setting process.

(4) The end-member features and corresponding parameters of the two models are calibrated according to the local hydrogeological conditions of the Banji Mining Area. Given the disparities in aquifer lithology, hydrochemical composition and groundwater flow regimes among different coalfields, the established model schemes cannot be directly extrapolated to other mining areas. Model recalibration and end-member parameter readjustment are required before practical application in analogous mines.

Conclusions

Based on 92 water samples collected from the four aquifer types in the Banji Mining Area, namely Quaternary, Sandstone, Taiyuan, and Ordovician, this study systematically investigates the hydrochemical characteristics and water source composition of groundwater through the combined use of hydrochemical diagram analysis, correlation analysis, and the two receptor models APCS-MLR and PMF. The main conclusions are as follows.

(1) Groundwater in the study area is predominantly of the HCO_3-Ca·Mg type, reflecting the chemical characteristics of shallow-circulation water dominated by carbonate rock dissolution; at some sampling points in the Quaternary aquifer, the water shows a marked tendency toward the Cl·SO_4-Na type, indicating the influence of local evaporation and concentration or the mixing of deep mineralized water. Pearson correlation analysis showed that a strong positive correlation between Na^++K^+ and Cl^- consistently existed across all four aquifer types, indicating that rock salt dissolution is a widespread process; however, the strength of the correlation between HCO_3^- and other ions varied significantly among the different aquifers, suggesting that the four groundwater types possess distinct and distinguishable hydrochemical fingerprints, thereby providing a data foundation for subsequent quantitative source analysis.

(2) The APCS-MLR model extracted three principal components, accounting for a cumulative variance contribution of 83.99%, corresponding to the rock salt dissolution component (PC1), the carbonate rock dissolution component (PC2), and the silicate weathering component (PC3), respectively. The dominant water sources for each ion differed significantly, and the model generally provided a good fit between the measured and predicted values for the six ions. However, due to the limitations of the unconstrained linear regression mechanism, the model retained an unidentifiable source contribution of 10%–18% for all ions, reflecting the inherent limitations of this model in characterizing non-negative mixing processes in groundwater.

(3) Using a combination of the RMSE elbow method, goodness-of-fit, and Bootstrap stability, the PMF model determined that the optimal number of factors is 3. These factors correspond to rock salt dissolution (98% contribution from Cl^-), carbonate rock dissolution (90% contribution from HCO_3^-), and a mixed-process end-member, respectively. This model exhibits greater factor specificity; the non-negative constraint ensures that all contribution rates are positive and sum to 100%, with no unidentifiable residuals. The spatial distribution of factor scores aligns well with the aquifers' lithology and hydrogeological conditions, validating the reliability of the interpretation results.

(4) Both models reached consistent conclusions regarding the dominant hydrogeochemical processes, with only slight differences in the numerical values of

the contribution rates. APCS-MLR is computationally efficient and suitable for rapid source tracing during emergency water inrush incidents; PMF's physical mechanisms better align with groundwater flow patterns, yielding more robust quantitative results. The approach of cross-validating the two models using the same dataset effectively mitigates algorithmic biases inherent in a single model and provides a reliable technical reference for identifying the sources of water inrush and implementing precise prevention and control measures for water-related hazards in multi-aquifer mines within the Huainan Coalfield.

Acknowledgments

This paper is supported by the Natural Science Research Project of Anhui Universities (2024AH050391); Open Fund of Engineering Research Center for Mine Water Hazard Detection and Early Warning in Coal Industry (2023-CIERC-01).

AI Usage Disclosure Statement

DeepSeek V4.1 Flash was used for language polishing and grammatical revision of this manuscript. All AI-modified texts were carefully reviewed, revised and validated by all authors. The authors take full responsibility for all content in the paper.

Conflict of Interest

The authors declare no conflict of interest.

References

1. CUI M.K., QIAO W., XU K.Q., FENG P.C. Study on the evolution of groundwater environment and formation potential of acid mine drainage under long-term mining in sulfur-rich coal mine. Process Safety and Environmental Protection, **213**, 109007, **2026**.
2. LIANG J., ZHAO W., LIU S., HANG Y., CUI H., LI C., XIAO L. Application of entropy weight-variable fuzzy set theory in water-inrush source identification of multiple aquifers in deep coal mine. Energy Exploration & Exploitation, **44** (3), 1159, **2026**.
3. KOU B.C., WEN T.X. Identification model of mine water inrush source based on XGBoost and SHAP. Scientific Reports, **15** (1), **2025**.
4. ZHANG G.H., XIE Z.Z., ZHANG B.W., JIAO Y.Y., ZOU J.P., WU J.Q., XIONG F., TANG Z.C. Sources of high-temperature water and gas inrush during tunnel excavation: A case of Bangfu tunnel in Southwest China. Journal of Rock Mechanics and Geotechnical Engineering, **16** (12), 5027, **2024**.
5. FENG P., QIAO W., HUANG Y., WANG Q., CAI J., WU H. Spatio-Temporal Evolution of Multiple Water Bodies and a Water Conservation Mining Strategy in the Xinjiang Coalfield: A Case Study of the Yushuquan and Yongxin Mines. Mine Water and the Environment, **44** (3), 694, **2025**.
6. JU Q., HU Y., LIU Q. Hydrogeochemical impacts of coal mining on multi-aquifer systems: A case study in the Linhuan Mining District, China. Ecotoxicology and Environmental Safety, **308**, 119492, **2025**.
7. CABRERA-SILLAS Y.G., GONZÁLEZ-ACEVEDO Z.I., GARCÍA-ZARATE M.A., RUIZ-AGUILAR D. Integrated geophysics and water geochemistry characterization of groundwater salinization in the La Misión coastal aquifer, Baja California, Mexico. Journal of South American Earth Sciences, **181**, 106167, **2026**.
8. YAN J., ANAMUL H., LIU J., FENG S. Spatial variation of fluoride and sub-basin driving factors in the surface water of Huai River Basin. Applied Water Science, **16** (6), **2026**.
9. SAHOO M.M., SWAIN J.B. Hydrogeochemical investigation and fluoride levels in lakes: a comprehensive health risk assessment. Environmental Geochemistry and Health, **47** (7), **2025**.
10. LU C., YIN H., CHENG W., DONG F., ZHANG W., LI B., JIA Y. Hydrochemical evolution and response mechanisms in multiple aquifers affected by coal mining activities: Insights from machine learning and inverse geochemical modeling. Applied Geochemistry, **198**, 106683, **2026**.
11. XUE S., LI Z., FENG G., ZHANG Y., GAO J., ZHANG R., LIU Y. Visualization characterization of void structure evolution in broken limestone and its influence on permeability. Fuel, **405** (Part A), 136412, **2026**.
12. CAO Z., GUO F., WANG W., DU F., LI Z., ZHANG S., WANG Q., ZHAI Y. Fault-Induced Floor Water Inrush in Confined Aquifers under Mining Stress: Mechanisms and Prevention Technologies – A State-of-the-Art Review. Fluid Dynamics & Materials Processing, **21** (10), 2419, **2025**.
13. WANG G., YIN S., CAO M., ZHONG C., HU S., YANG C., WU M. Water hazard prevention technology for confined mining beneath dual extremely thin aquicludes in roof and floor. Scientific Reports, **16** (1), **2026**.
14. TONG X., ZHENG X., JIN Y., DONG B., LIU Q., LI Y. Prevention and control strategy of coal mine water inrush accident based on case-driven and Bow-tie-Bayesian model. Energy, **320**, **2025**.
15. WEN J., GU S., SU P., LI J. Study on the two-phase coupling migration mechanism of deceleration aggregate and water in coal mine water inrush channel. Scientific Reports, **15** (1), **2025**.
16. MA H. Characteristics and Hazards Prevention of Bed Separation Water Inrush: A Case Study of the Cuimu Coal Mine, China. Water, **18** (7), 813, **2026**.
17. RAVI M., KUMAR M.D., THOLKAPIYAN M., VENKATRAMAN V. Integrated Hydrogeochemical and seasonal assessment of groundwater using ionic ratios, gibbs and piper methods in Madurai North Taluk. Environmental Earth Sciences, **85** (9), 219, **2026**.
18. GHOSH S., JHA M.K. Hydrogeochemical characterization of groundwater and critical assessment of its quality in a coastal basin. Environment, Development and Sustainability, **27** (1), 765, **2025**.
19. DHEERAJ V.P., SINGH C.S., ALAM A., SONKAR A.K. Hydrogeochemical quality investigation of groundwater resource using multivariate statistical methods, water quality indices (WQIs), and health risk assessment in Korba Coalfield Region, India. Stochastic Environmental Research and Risk Assessment, **39** (3), 937, **2025**.

20. CAO H., QIAN J., CHEN H., LIU C., GAO S., LYU M., DONG W., HU C. Managed aquifer recharge and extraction effects on groundwater level and quality dynamics in a typical temperate semi-arid fissured karst system: a multi-method quantitative study. Hydrology and Earth System Sciences, **29** (20), 5213, **2025**.
21. LI Y., GE X., DU Y., DING H., JI X., ZHOU C., CHEN Y., CUI D., LU A. Spatial distribution and genesis of iron, manganese, arsenic, and nitrogen in groundwater of typical areas. Science China Earth Sciences, **68** (9), 2797, **2025**.
22. ZHAO P., ZHANG H., YANG X., WANG H., ADNAN M., WU Y., LIU Y., GUO L., QIN H.B., XIAO B., XUE S. Mechanisms of toxic metal(loid) migration in the slag-soil-groundwater continuum: Geochemical evidence from an abandoned industrial complex. Journal of Cleaner Production, **523**, 146404, **2025**.
23. XIA Y., SU H., WANG W., LI S., LI Z. Layered control of stream-groundwater interactions: Insights into regional hydrological connectivity in China's Loess Plateau. Journal of Hydrology, **654**, 132807, **2025**.
24. WANG M., WANG X., XIONG Y., ZHANG X. Quantitative source apportionment of heavy metals and metalloid in urban lake sediments using PCA-APCS-MLR receptor model: a case study of Qianhu Lake, China. Environmental Geochemistry and Health, **48** (9), **2026**.
25. CAO Z., QIAN H., GAO Y., LI K., LI Y., SHI X., LI S., ZHAO W., YANG S., TIAN P., WU P., MA Y. Ecological risk assessment and source identification of heavy metals in the sediments of the Danjiang River Basin: A quantitative method combining multivariate analysis and the APCS-MLR model. Ecological Indicators, **175**, 113518, **2025**.
26. MI J., LU J., LIU T., LI Z., SHI X., LV C., JIA Y., WU Z. Activated sludge in municipal wastewater treatment plant serves as a potentially source of microplastics: Source apportionment based on APCS-MLR and PMF receptor models. Journal of Hazardous Materials, **503**, 141072, **2026**.
27. ZHOU Z., CHEN M., LI Y. Source apportionment of groundwater contamination using integrated PCA-APCS-MLR modeling: A case study of an industrial zone in Southwest China. Environmental Research, **292**, 123628, **2026**.
28. JAHAN A., RAI N., KHAN M.U. Groundwater Contamination Characterization and Source Apportionment of Heavy Metals and Associated Source-Specific Health Risk Appraisal Using Monte Carlo Simulation Coupled with PCA-MLR and PMF Models in the Middle Gangetic Basin, India. Exposure and Health, **17** (3), 775, **2025**.
29. FENG J.W., ZHANG M., PING J.H., LI H., LIU J.Q. Scale effects and time-lag responses of groundwater hydrochemistry to land use change: A spatiotemporal analysis of a rapidly urbanizing typical agricultural area in the Lower Yellow River alluvial plain. Journal of Hydrology: Regional Studies, **66**, 103630, **2026**.
30. LI W., YANG B., HUANG R., YANG J., CUI J., LU X., XIONG J. Health risk assessment of potentially toxic elements in a Tibetan Plateau Wetland, China, using APCS-MLR, PMF, and Monte Carlo simulation. Journal of Soils and Sediments, **25**, 3905, **2025**.
31. YIN H., IKARD S.J., RUCKER D.F., BROOKS S.C., DAI Z., CARROLL K.C. Imaging Hyporheic Exchange by Integrating Deep Learning and Physics-Informed Inversion of Time-Lapse Self-Potential Data. Geophysical Research Letters, **52** (21), **2025**.
32. WU Y., ZHANG Q., LUO Y., JIN K., HE Q., LU Y. Spatial and temporal distribution characteristics and source apportionment of biogenic elements using APCS-MLR model in the main inlet tributary of Danjiangkou Reservoir. Environmental Science and Pollution Research, **32** (7), 3729, **2025**.
33. KEESARI T., KUMARI B., DAUJI S., ROY A., MAITRA A., GUPTA S., PANDA B., PANT D., SHARMA D.A. Geochemical study of trace metals and health risk assessment in Indian geothermal aquifer using health indices, MCS and APCS-MLR. Scientific Reports, **16**, **2026**.
34. OFFROY M., AYADI A., GOVOHETCHAN L., AYOUB J., HANCEWICZ T.M., DUPONCHEL L., MARCHETTI M. Enhanced Raman hyperspectral imaging using RS-NMF: a novel Regularized Sparse Non-negative Matrix Factorization for spectral unmixing. Chemometrics and Intelligent Laboratory Systems, **269**, 105602, **2026**.
35. CAO J., GUO J., SHI H., CHEN Y., SHEN Z., MA H., HOU D. Sources apportionments of heavy metal(loid) s in soils based on positive matrix factorization (PMF) and lead isotope fingerprinting. Journal of Environmental Management, **395**, **2025**.