



Research article

Statistical analyses of hydrochemistry in multi-aquifers of the Pansan coalmine, Huainan coalfield, China: implications for water-rock interaction and hydraulic connection

Kai Chen^{a,b}, Qimeng Liu^{a,*}, Tingting Yang^{a,b}, Qiding Ju^{a,b}, Yu Feng^a^a School of Earth and Environment, Anhui University of Science & Technology, Huainan 232001, China^b State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mine, Huainan 232001, China

ARTICLE INFO

Keywords:

Hydrochemistry
Water-rock interaction
K-means cluster analysis
Receptor model
Pansan coal mine
China

ABSTRACT

Understanding the groundwater hydrogeochemical processes and aquifer hydraulic connections are essential for effective prevention of water inrush in concealed coal mines. In this study, 40 groundwater samples were collected from the loose layer aquifer (LA), coal measure aquifer (CA), and limestone aquifer (LA) in the Pansan coal mine, Huainan coalfield, China, and the major ion concentrations were analyzed by bivariate diagrams ($\text{Na}^+ + \text{K}^+ - \text{Cl}^-$ versus $\text{Ca}^{2+} + \text{Mg}^{2+} - \text{SO}_4^{2-} - \text{HCO}_3^-$ and CAI-I versus CAI-II), multivariate statistical methods, and receptor model in order to identify the water-rock interactions and aquifer hydraulic connections. Piper diagram showed that groundwater in LA and TA was dominated by the Na-Cl type, while groundwater in CA was mainly of the Na-HCO₃ type. Based on the results of bivariate diagrams and PCA/FA, weathering of silicate minerals and cation exchange (source 1), sulfate dissolution (source 2) and chloride dissolution (source 3) were the main processes controlling the groundwater chemistry. Unmix model revealed that the mean contribution of source 1 to CA samples was 74%, while LA and TA samples have higher contributions from evaporite dissolution (source 2 and source 3) relative to CA samples. Moreover, both clustering analysis methods (Q-type hierarchical and K-means cluster) confirmed the existence of a hydraulic connection between LA and TA in the northeastern part of the study area. It is concluded that the application of multivariate statistical analysis to interpret groundwater chemistry can provide useful guidance to prevent water inrush in coal mines.

1. Introduction

Coal is the main source of energy supply in China [1, 2]. As reported by National Bureau of Statistics of China, coal accounts for 56.8% of total energy consumption in 2020. In recent years, inrush water has become an increasing threat to coal production because of the gradual increase in mining depth [3]. Most practices have proven that quickly and accurately identifying the source of the inrush water and then implementing countermeasures is the key to minimize the loss [4]. Differences in the residence time of groundwater and the type of surrounding rocks lead to different characteristics of groundwater chemistry in different aquifers, which can be used to determine the source of inrush water in coal mines [5, 6, 7]. In addition, the evaluation of the hydraulic connection between different aquifers can be targeted to take measures to prevent the occurrence of inrush water accidents [8]. Therefore, understanding the

water-rock interactions and aquifer hydraulic connections are important for the safe production of concealed coal mines.

The relationships between major ions in groundwater and principal component analysis/factor analysis can effectively reveal hydrogeochemical processes in aquifers, but it cannot quantify the contributions of these processes, which limits the full understanding of hydrochemical evolution [9, 10, 11, 12, 13]. Receptor models, typified by the Unmix model, have been widely used for source identification of heavy metals and polycyclic aromatic hydrocarbons in soil, and PM_{2.5} and PM₁₀ in atmospheric particulate matter [14, 15, 16]. Huang et al. (2010) introduced the Unmix model for the first time to quantify the contributions of identified pollution sources to each pollutant and each monitoring site in the Qiantang River, China [17]. Recently, Nagaraju et al. (2018) used the Unmix model to evaluate the contributions of weathering of felsic and mafic silicates to groundwater chemistry in the

* Corresponding author.

E-mail address: qmliu@aust.edu.cn (Q. Liu).

Andhra Pradesh region of India [18], Chen et al. (2022) quantified the influence of geological factors and anthropogenic activities on groundwater in northern Anhui Province, China, and Subba Rao et al. (2022) successfully determined the sources of Ca^{2+} , Mg^{2+} , Na^+ , K^+ , HCO_3^- , Cl^- , NO_3^- , SO_4^{2-} , and F^- in groundwater in the Telangana region, southern India based on the model, which confirmed the validity and reliability of the method [19, 20].

Cluster analysis is a traditional statistical method that provides a simple and efficient way of regrouping samples and has been often used in studies of the hydraulic connections of different water bodies [21]. Teng et al. (2018) used cluster analysis to assess the hydrochemical response mechanism of exchange between surface water and groundwater in the Songhua River basin, northeast China [22], Ghesquière et al. (2015) identified the hydraulic connection between granular and bedrock aquifers in a region of the Canadian Shield [23], and Zhang et al. (2020) revealed the recharge process of confined groundwater in limestone aquifer to overlying sandstone aquifer under the influence of karst collapse columns by cluster analysis, these studies were important for the management of regional water resources [24]. However, different clustering algorithms usually generate different results for the same dataset; therefore, two algorithms (including Q-type hierarchical cluster and K-means cluster) were employed in this study for comparative analysis of the hydraulic connections of multiple aquifer systems to reduce the one-sidedness of the analysis.

Huainan coalfield is the largest coal production base in eastern China [25]. Decades of mining activities in this area have altered the original underground stress field, damaged the structure of the rock formations, and affected the natural circulation of groundwater, which resulted in the frequent occurrence of inrush water accidents [26, 27, 28]. The Pansan coal mine is an important production mine in Huainan coal field, with a capacity of 5 million tons of coal per year and a design mining depth of 900 m, which is threatened by inrush water from several aquifers. However, no study has so far been conducted to comprehensive assess the groundwater chemistry of multiple aquifers in the coal mine. Given

the consideration above, the objectives of this study are to: (1) reveal the groundwater chemistry characteristics and hydrochemical facies in the main aquifers; (2) identify understand the hydrogeochemical processes controlling the major ion concentrations and quantify their contributions and (3) identify the aquifer hydraulic connections in this area. The results of the study are expected to help understand the hydrogeochemical evolution pattern of this coal mine and other coal mines with similar hydrogeological conditions in the Huainan coalfield, as well as to provide useful guidance for the prevention of inrush water.

2. Materials and methods

2.1. Study area

Pansan coal mine ($116^\circ 41' 45''$ – $116^\circ 48' 45''$ E, $32^\circ 47' 30''$ – $32^\circ 52' 30''$ N) is located in north central Anhui Province, China, and is one of the main production mines in the Huainan coalfield (Figure 1a). The mine covers an area of approximately 54.3 km^2 , the resource reserves are 771.287 million tons, the production capacity is 5 million tons/year. The study area belongs to the cold-temperate temperate and humid climate, with an average annual temperature of 15.1°C and an average annual rainfall of 1104.7 mm, and most of rainfall is usually concentrated in June to August. The area belongs to the Huai River alluvial plain, with elevations between 19.5 m and 23.5m.

The revealed strata in this area from new to old are the Quaternary, Neoproterozoic, Paleocene, Permian, Carboniferous and Ordovician, the coal-bearing strata are Permian and Carboniferous, with an average thickness of 1300 m. Statistics show that the Pansan coal mine contains 34 coal seams, the total thickness of recoverable coal seams is 27.66 m, and the 13-1, 11-2 and 8 seams being exploited (Figure 1c). The main aquifer systems in the coal mine from shallow to deep can be classified as the loose layer aquifer system (LA), the coal measure aquifer system (CA) and the Taiyuan limestone aquifer system (TA), with detailed information of each aquifer as described follows:

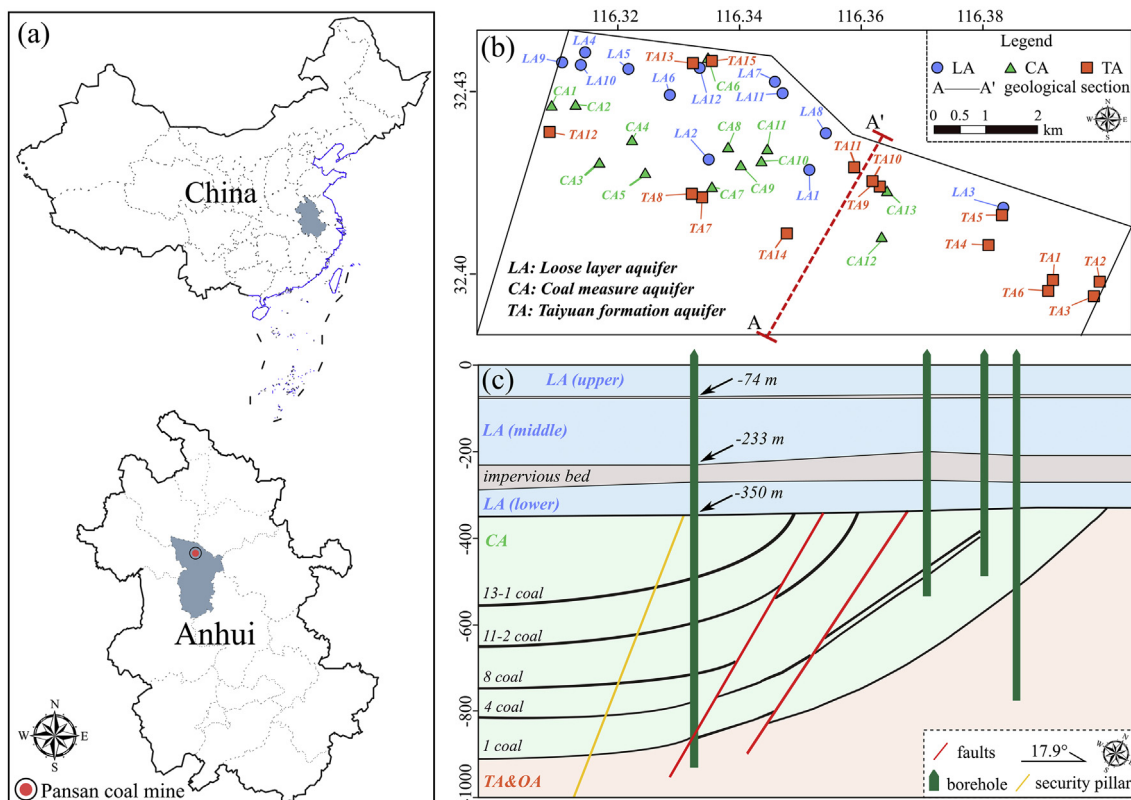


Figure 1. (a) Location of the study area; (b) Distribution of groundwater samples; (c) Hydrogeological profile description.

LA: According to the borehole data, the LA thickness ranges from 187 to 484 m, and is characterized by a gradual development from south-east to north-west. The aquifer can be further divided into three secondary aquifers: upper, middle and lower. The upper and central loose aquifers usually have limited impact on coal production due to the presence of impervious beds (Figure 1c), but the lower aquifer is one of the main threats to coal mining as it overlies the coal bearing strata. The average thickness of the lower aquifer is 57.7 m, and the lithology is dominated by mudstone, sandstone, siltstone and interbedded gravels, among which the gravels are mainly quartzite, quartz sandstone, magmatic rocks and a few chert gravels. Because of the aquiclude, groundwater recharge in LA is limited and is mainly discharged through infiltration.

CA: The average thickness of CA in the Pansan coal mine is 848.9 m, and the surrounding rocks consist of sandy mudstone, siltstone, fine sandstone and coarse sandstone. Due to the influence of tectonic extrusion, the aquifer is rich in water and fissures are developed, which is the direct recharge source of roadway. In addition, the security pillar was designed in the aquifer to block the recharge of the upper LA (Figure 1c). Groundwater in CA is mainly static storage, and it can discharge downward through sandstone fractures.

TA: The average thickness of the TA is 118.3 m. It consists of dark grey limestone, mudstone, and fine and medium sandstone, which contains 12–13 limestone layers, interspersed with 5–10 layers of unstable thin coal seams and carbonaceous mudstone. The formation is rich in brachiopods, corals, sea lily stems and tadpole fossils. The aquifer usually has no direct impact on coal mining, but because TA has strong water pressure and is located at the bottom of 1 coal seam (Figure 1c), it may cause serious inrush water accidents if coal mining activities cause fault activation. TA can be recharged by the Ordovician limestone aquifer, and the anthropogenic pumping is the main discharge pathway.

2.2. Sample collection and analysis

Considering the distribution characteristics of the aquifer systems and the mining conditions, 40 groundwater samples were collected in August 2020: 12 from LA (lower), 13 from CA, and 15 from TA. LA and TA samples were obtained through borehole pumping, and CA samples were collected directly in the roadway (Figure 1b). The clean polyethylene bottles (2L) were washed three times and then filled with water samples, sealed and sent to the laboratory to be filtered through 0.45 µm membrane and stored at 4 °C environment. All testing procedures were completed within one week. The concentrations of seven water quality parameters were tested, including Ca²⁺, Mg²⁺, Na⁺, Cl⁻, SO₄²⁻, HCO₃⁻ and CO₃²⁻; HCO₃⁻ and CO₃²⁻ were tested by titration of 0.05 mol/L HCl with methyl orange and phenolphthalein as indicators, and other parameters were tested by ion chromatography (ICS-600 for cations and ICS-900 for anions, Thermo Fischer Scientific, USA). All

analyses were performed in triplicate to ensure the accuracy of the analysis and conducted in the State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mine, Anhui Province, China.

2.3. Statistical analysis methods

Basic statistics and one-way ANOVA analyses were conducted using Excel 2016 and Origin (version 8.0). Piper diagram for classifying groundwater chemical types was generated by AqQA software (version 1.5), and R (4.1.2 environment, <https://www.r-project.org/>) platform was employed to conduct cluster analysis and factor analysis using factoextra and igraph packages. Moreover, unmix model analysis was performed by environmental protection agency (EPA) unmix software 6.0.

2.3.1. Cluster analysis

Cluster analysis is an unsupervised learning tool that can be used to classify samples into different groups by different algorithms and has been widely used in different fields due to its simple and intuitive properties [29, 30]. The results obtained by different algorithms were different. To reduce the error of a single evaluation method, hierarchical clustering (Q-type) and non-hierarchical clustering (K-means) were chosen in this study to compare and analyze the hydraulic connections of different aquifer systems.

Q-type hierarchical cluster analysis: It starts by calculating the distances between different samples, merging the closest samples into the same class each time until a hierarchical nested clustering tree is finally generated [31]. To obtain a more appropriate clustering effect, the Ward linkage method and the Euclidean distance were chosen in this study.

K-means cluster analysis: Different from Q-type cluster analysis, it assigns each sample to its nearest cluster center by pre-inputting K objects as initial cluster centers, and then repeats the process by iterative operations after all samples are assigned until no sample was reassigned to a different cluster center [32].

2.3.2. Principal component analysis/factor analysis

Principal component analysis (PCA) is a traditional dimensionality reduction method based on an orthogonal transformation that can combine multiple original variables into a few uncorrelated principal components (PCs). Factor analysis (FA) further polarizes the variable loadings on the obtained PCs to generate new “PCs”, called varifactors (VFs) that can better explain the potential information of the dataset [33, 34]. According to previous studies, the corresponding loadings of the variables on the VF are defined as weak, moderate and strong at 0.3–0.5, 0.5–0.75 and >0.75, respectively [30]. In studies of environmental science, variables with strong loadings on the same VF usually have same or similar sources.

Table 1. Basic statistics of major ion concentrations in groundwater of three aquifer systems.

Aquifer	Statistics	Parameters (unit: mg/L)							
		Na ⁺	Ca ²⁺	Mg ²⁺	Cl ⁻	HCO ₃ ⁻	SO ₄ ²⁻	CO ₃ ²⁻	TDS
LA (n = 12)	Range	591–923	17–76	11–44	684–1037	184–343	86–596	6–43	1772–2576
	Mean	(745)	(41)	(26)	(884)	(263)	(327)	(18)	(2172)
	SD	90	17	12	142	50	147	10	216
CA (n = 13)	Range	799–1751	3–37	0–16	181–1000	12–3088	6–1691	42–294	1932–4233
	Mean	(1264)	(10)	(3)	(530)	(1899)	(195)	(105)	(3056)
	SD	279	9	4	243	794	441	65	644
TA (n = 15)	Range	29–1146	10–85	1–52	32–1448	26–475	22–595	12–56	350–3389
	Mean	(469)	(47)	(23)	(522)	(209)	(274)	(24)	(1463)
	SD	384	23	18	502	123	193	13	1088

Note: SD = standard deviation; TDS = total dissolved solids.

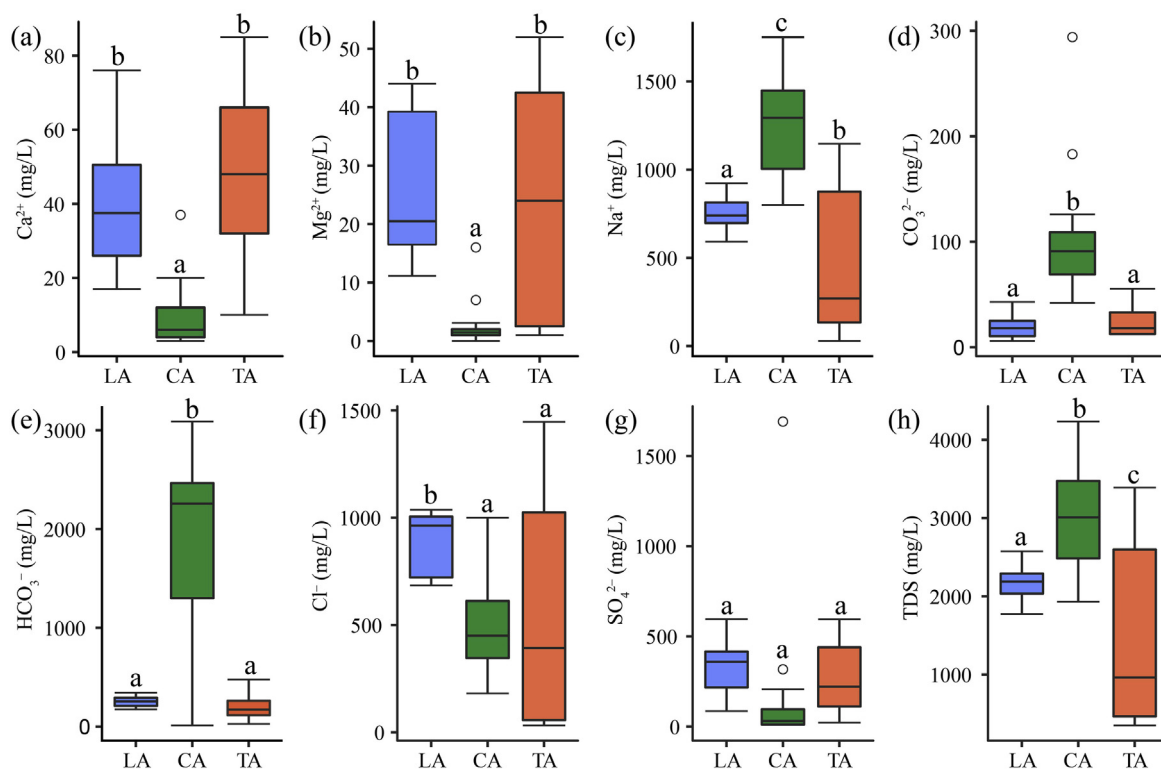


Figure 2. Comparison of the distribution characteristics of major ion concentrations in groundwater of different aquifers (a, b and c represent the statistical differences at $p < 0.05$ level).

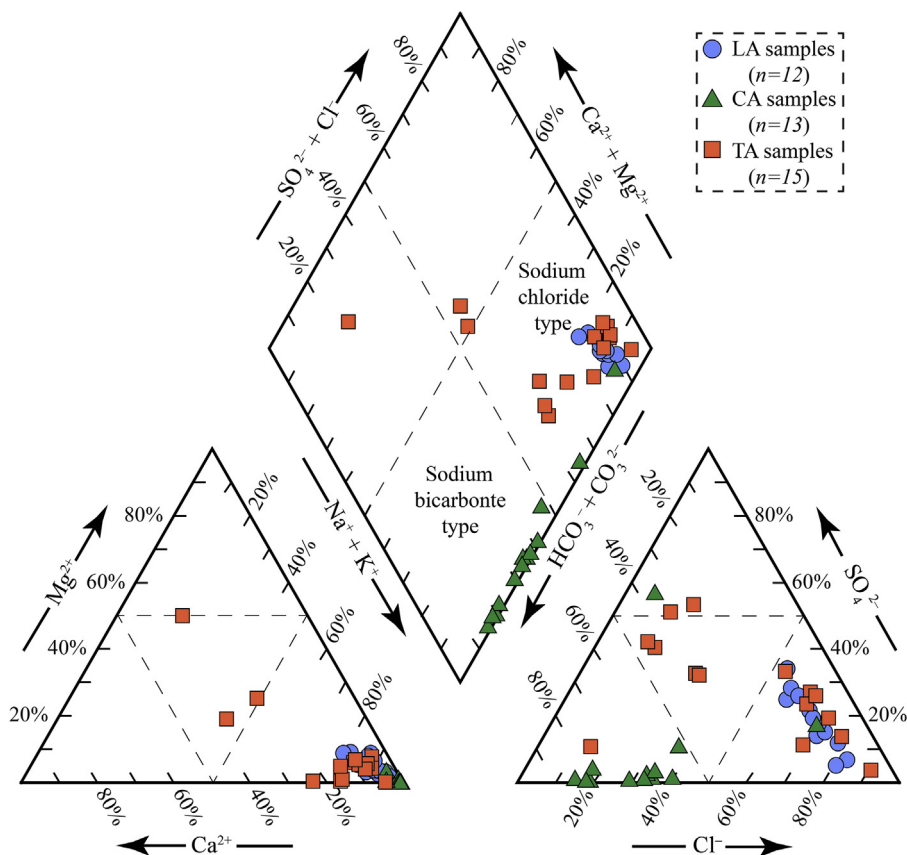


Figure 3. Chemical analyses of groundwater in the Pansan coal mine.

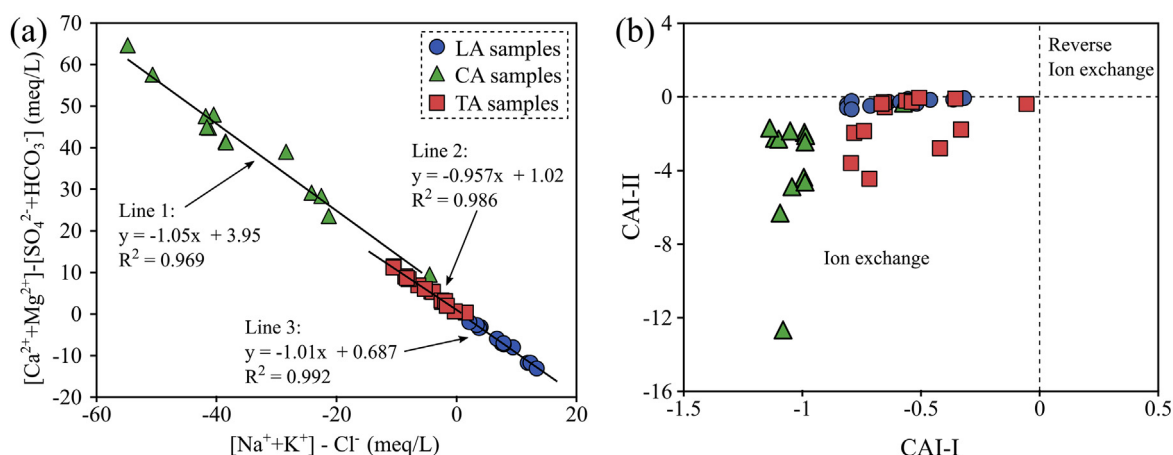


Figure 4. (a) Relationship between $([Na^+ + K^+] - Cl^-)$ versus $([Ca^{2+} + Mg^{2+}] - [SO_4^{2-} + HCO_3^-])$ and (b) Plot of CAI-I versus CAI-II of groundwater samples.

2.3.3. Unmix model analysis

The Unmix model, which uses geometric techniques with self-modelling resolution to enforce non-negative constraints on source contributions, has long been used in source apportionment studies of atmospheric particulate matter, soil heavy elements and polycyclic aromatic hydrocarbons [14, 16, 35]. In recent years, the contributions of different types of water-rock interactions to the major ion concentrations in groundwater in the Huabei coalfield in eastern China and the Wanaparthi region in southern India has been successfully quantified, which confirmed the reliability of the model for hydrochemical studies [36]. Based on the assumption that the concentrations of the parameters are linear combinations of different sources, the physical basis of the model can be expressed as follow [37]:

$$C_{ij} = \sum_{k=1}^N a_{ik} * S_{kj} \quad (1)$$

In Eq. (1), C_{ij} is the measured concentration of i_{th} ion in j_{th} sample, a_{ik} is the dimensionless mass fraction, S_{kj} represents the total amount of mass from k_{th} source in j_{th} sample, and is obtained by iterative loop operation of the model, and $a_{ik} * S_{kj}$ represents the contribution of k_{th} source to C_{ij} .

3. Results and discussion

3.1. Major ion concentrations

The statistical results of groundwater parameters of the Pansan coal mine are listed in Table 1, and the distribution characteristics of these parameters in the three aquifers are compared in Figure 2. As shown in Table 1, the mean concentrations of Mg^{2+} , Cl^- and SO_4^{2-} in LA are higher than CA and TA, and the major ion concentrations in LA have the

following order: Cl^- (884 ± 142 mg/L) > Na^+ (745 ± 90 mg/L) > SO_4^{2-} (327 ± 147 mg/L) > HCO_3^- (263 ± 50 mg/L) > Ca^{2+} (41 ± 17 mg/L) > Mg^{2+} (26 ± 12 mg/L) > CO_3^{2-} (18 ± 10 mg/L). The CA samples have lower mean concentrations of Ca^{2+} and Mg^{2+} but higher concentrations of Na^+ , HCO_3^- and CO_3^{2-} than the LA and TA samples, and their mean concentrations were HCO_3^- (1899 ± 794 mg/L) > Na^+ (1264 ± 279 mg/L) > Cl^- (530 ± 243 mg/L) > SO_4^{2-} (195 ± 441 mg/L) > CO_3^{2-} (105 ± 65 mg/L) > Ca^{2+} (10 ± 9 mg/L) > Mg^{2+} (3 ± 4 mg/L). Comparatively, the TA samples have lower concentrations of Na^+ , Cl^- and HCO_3^- but higher Ca^{2+} concentration relative to LA and CA samples. The major ion concentrations of TA samples in a descending order were: Cl^- (522 ± 502 mg/L) > Na^+ (469 ± 384 mg/L) > SO_4^{2-} (274 ± 193 mg/L) > HCO_3^- (209 ± 123 mg/L) > Ca^{2+} (47 ± 23 mg/L) > Mg^{2+} (32 ± 18 mg/L) > CO_3^{2-} (24 ± 13 mg/L). Moreover, the TDS contents in groundwater from the three aquifers have the following order: CA (3056 ± 644 mg/L) > LA (2172 ± 216 mg/L) > TA (1463 ± 1088 mg/L).

The results of the one-way ANOVA based on the least significant different (LSD) method showed that the concentrations of Ca^{2+} , Mg^{2+} , CO_3^{2-} and HCO_3^- in LA and TA samples were significantly different from those in CA samples ($p < 0.05$), whereas the Cl^- concentration in CA and

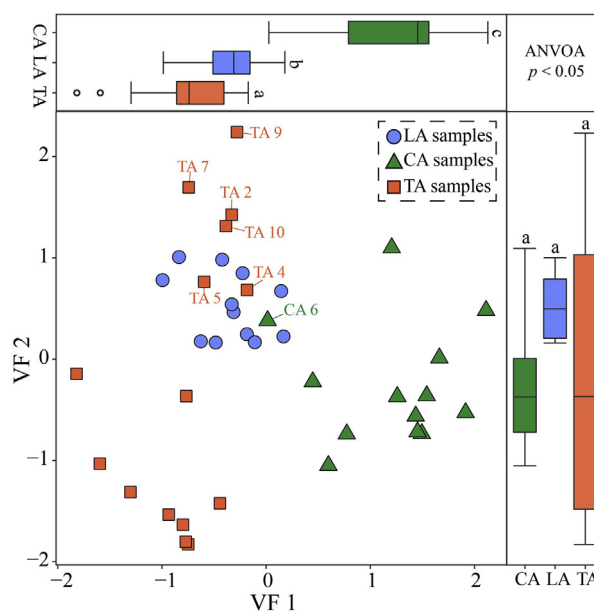


Figure 5. Factor scores of groundwater samples in different aquifers of Pansan coal mine.

Table 2. Results of PCA/FA including factor loading matrix, eigenvalue, percentage of variance and cumulative percentage variance.

Parameters	VF1	VF2
HCO_3^-	0.894	-0.157
Na^+	0.887	0.447
Ca^{2+}	-0.748	0.482
Mg^{2+}	-0.631	0.621
Cl^-	0.078	0.846
SO_4^{2-}	-0.131	0.573
Eigenvalues	2.57	1.89
% of variance	42.8%	31.5%
Cumulative %	72.3%	

Table 3. Source contributions to major ion concentrations in groundwater calculated with Unmix model.

Species	R ²	Source profiles (Unit: mg/L)			Proportions (Unit: %)		
		source 1	source 2	source 3	source 1	source 2	source 3
HCO ₃ ⁻	0.96	765	129	0	86%	14%	0%
Na ⁺	0.92	427	294	88	53%	36%	1%
SO ₄ ²⁻	0.99	0	338	0	0%	100%	0%
Cl ⁻	0.49	178	79.6	377	27%	13%	60%
Ca ²⁺	0.76	0.89	3.52	27.2	3%	11%	86%
Mg ²⁺	0.92	0.02	0	19.2	0%	0%	100%

TA samples were significantly different from those in LA (Fig. 2a, b, d, e and f). In addition, Na⁺ and TDS concentrations showed significant differences in the LA, CA and TA samples (Fig. 2c and h), in contrast, SO₄²⁻ concentrations showed no significant differences in groundwater from the three aquifers (Figure 2g), which implied similarities and differences in water chemistry control mechanisms in different aquifers.

3.2. Hydrochemical types

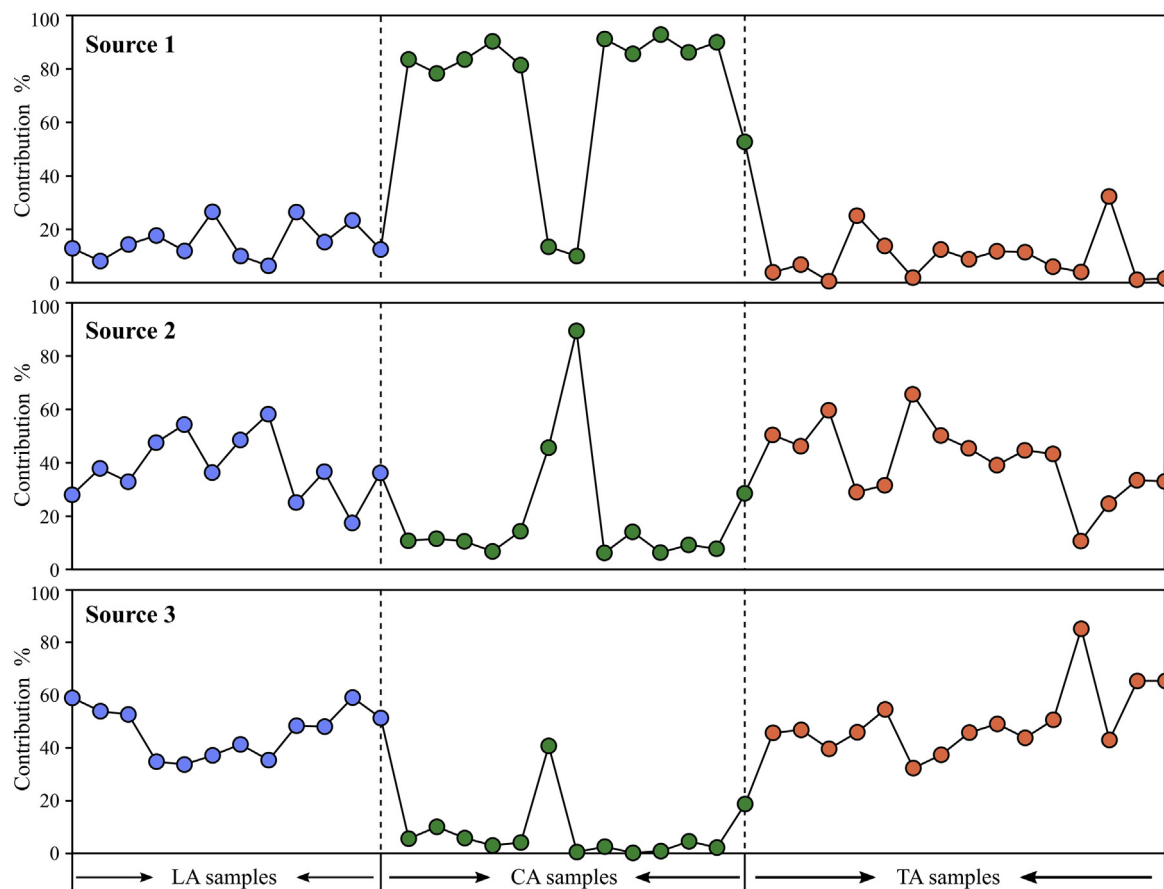
The dominant anion and cation concentrations in groundwater are exhibited in the Piper diagram [38]. As shown in the figure, Na⁺ is the dominant cation in groundwater except three TA samples, whereas the dominant anions in groundwater of the three aquifers are different. All of the LA samples are located in the Cl⁻ dominant area, and therefore classified as Na–Cl water type. Na–HCO₃ is the most dominant water type in CA groundwater samples, followed by Na–Cl and Na–SO₄ types. In comparison, the distribution of TA samples in Figure 3 is more dispersed relative to LA and CA samples, and more than 53% of the groundwater

samples from TA are of Na–Cl water type, ~33% of Na–HCO₃ water type and only ~13% of Na–SO₄ water type.

3.3. Source of major ions

3.3.1. Relationship between major ion concentrations

The relationship between [(Na⁺ + K⁺) – Cl] and [(Ca²⁺ + Mg²⁺) – (SO₄²⁻ + HCO₃⁻)] in groundwater can be used to identify the occurrence of cation exchange [39, 40]. Generally, when cation exchange is the important controlling factor for groundwater chemistry, the relationship should be linear with a slope of –1 [41, 42, 43, 44, 45]. As presented in Figure 4a, the relationships of [(Na⁺ + K⁺) – Cl] versus [(Ca²⁺ + Mg²⁺) – (SO₄²⁻ + HCO₃⁻)] for LA, CA and TA samples can be expressed as: $y = -1.01x + 0.687$ ($R^2 = 0.992$), $y = -0.957x + 1.02$ ($R^2 = 0.986$) and $y = -1.05x + 3.95$ ($R^2 = 0.969$), respectively. This confirms that cation exchange plays an important role in the evolution of groundwater chemistry in the aquifers. Moreover, the chloro-alkaline indices (CAI-I and CAI-II) can further determine the direction of cation exchange in

**Figure 6.** Source contributions to groundwater samples from different aquifers.

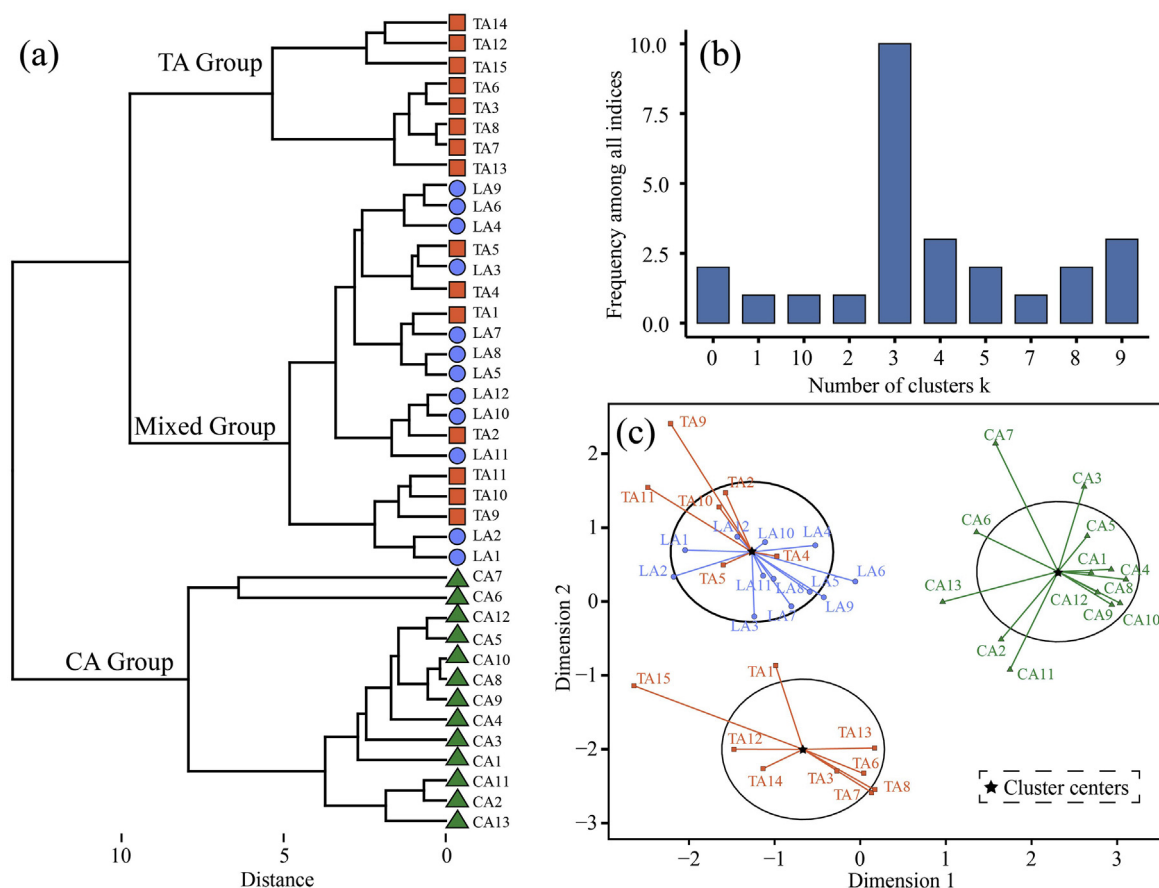


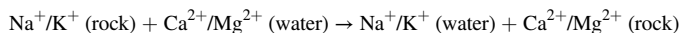
Figure 7. Cluster analysis results. (a) Results of Q-type hierarchical clustering analysis; (b) Optimal number of cluster centers; (c) Results of K-means cluster analysis.

groundwater. Typically, positive or negative values of CAI-I and CAI-II represent the replacement of Ca^{2+} and Mg^{2+} in water by Na^+ and K^+ in rock or the reverse. The formulas for CAI-I and CAI-II are (Unit: meq/L):

$$CA-I = (Cl^- - Na^+ - K^+) / Cl^-$$

$$CA-II = (Cl^- - Na^+ - K^+) / (SO_4^{2-} + HCO_3^- + CO_3^{2-} + NO_3^-)$$

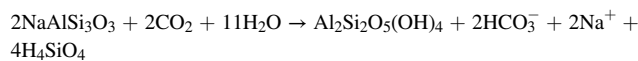
As shown in Figure 4b, all samples were located in the third quadrant and most of the CA samples had lower CAI-I and CAI-II values relative to CA and TA samples, indicating that cation exchange resulted in increased concentrations of Na^+ and decreased concentrations of Ca^{2+} and Mg^{2+} concentrations in groundwater, and that this phenomenon was more significant in CA.



3.3.2. PCA/FA

After maximum variance rotation, the first two VFs were extracted according to the Kaiser's criteria (eigenvalue higher than 1), with a cumulative variable interpretation of 72.3%, which can be used to reveal the hydrochemical processes. The loadings of each parameter in the two VFs are summarized in Table 2 and the factor scores of each groundwater samples are presented in Fig. 5. The first VF (VF1) showed strong positive loadings on HCO_3^- (0.894) and Na^+ (0.887), and moderate negative positive loading on Ca^{2+} (-0.748) and Mg^{2+} (-0.631), accounting for 42.8% of the total variance. High concentrations of HCO_3^- and Na^+ in the groundwater of the mine are usually associated with the weathering of

silicate minerals in the aquifer, with representative chemical equation as follows [24]:



In addition, strong cation exchange can also lead to increase in Na^+ concentration while decreasing in Ca^{2+}/Mg^{2+} concentration in water (as proved in 3.3.2). Therefore, VF1 could be considered as the combined effect of weathering of silicate minerals and cation exchange. The second VF (VF2) explained 31.5% of the total variance and had a strong positive loading on Cl^- (0.846), moderate positive loading on Mg^{2+} (0.603) and SO_4^{2-} (0.573). Considering the mineral composition of the strata in the study area, VF2 could be interpreted as evaporite dissolution, because the dissolution of chloride and sulfate minerals would simultaneously release these ions into water.

As shown in Figure 5, the VF2 scores of the samples have no significant difference at the 0.05 level, but the VF1 scores were significantly different, reflecting that the types and degrees of water-rock interaction were different in the three aquifer systems. Most of the samples from CA have high VF1 scores (>0, except for CA6), followed by LA samples (mean = -0.35) and TA samples (mean = -0.78), suggesting that groundwater in CA was mainly influenced by weathering of silicate minerals and cation exchange relative to groundwater in LA and TA. In comparison, several TA samples (including TA2, 4, 5, 7, 9 and 10) and all of the LA samples have VF2 scores higher than 0, which indicates that they were influenced primarily by evaporite dissolution. Moreover, the vother TA samples have both the lowest VF1 and VF2 scores indicating that they were affected by limited water-rock interactions (Figure 5).

3.3.3. Unmix model analysis

In this study, three sources (source 1, source 2 and source 3) were obtained, and the values of the Min Rsq and the Min Sig./Noise were 0.86 and 2.14, respectively, which satisfy the requirements for employing the Unmix model (Min Rsq >0.8 and Min Sig./Noise >2). In addition, the mean R^2 between the observed and predicted concentrations of parameters was 0.84, showing a good relationship between the observed and predicted values. Thus, the Unmix model was suitable for source apportionment. Table 3 summarizes the average contributions of different sources to each major ion in groundwater. HCO_3^- and Na^+ were primarily influenced by source 1, with high contribution ratios of 86% and 53%, respectively. Source 3 had the highest contribution to SO_4^{2-} , and the contribution of source 3 was 60% for Cl^- , 86% for Ca^{2+} and 100% for Mg^{2+} . Combined with the results of PCA/FA above, the main controlling factors of groundwater chemistry in the three aquifers were weathering of silicate minerals, cation exchange, and evaporite dissolution. Therefore, source 1 can be interpreted as weathering of silicate minerals and cation exchange, whereas source 2 and source 3 can represent dissolution of sulfate minerals and chloride minerals, respectively. Furthermore, the contributions of three sources to each groundwater sample are presented in Figure 6. As shown in Figure 6, the mean contribution of source 1 to the CA samples (74%) was significantly higher than that of the LA samples (16%) and TA samples (10%), which can be explained by the presence of more sandstones in this stratum (e.g., feldspathic quartz sandstone). On the contrary, evaporite dissolution, represented by source 2 and source 3, contributed similarly to LA and TA samples, and the mean contribution was higher than that of CA samples.

3.4. Hydraulic connections between different aquifers

As shown in Figure 7a, within distance of 10, all of the groundwater samples were divided into three groups: the first group (TA Group), contained 8 (accounting for 53%) TA samples, the second group (Mixed Group) represented 7 TA samples (TA1, 2, 4, 5, 9, 10 and 11) and all of the LA samples, while the CA samples were contained in third group (CA Group). Therefore, the CA samples can be clearly distinguished from other aquifer samples, but it is difficult to distinguish TA and LA samples. Statistics showed that 39% of the indices recommended that the most suitable clustering effect would be obtained when k was set to 3, which also corresponds to the actual hydrogeological conditions in the study area (Figure 7b). Therefore, the $k = 3$ was selected for K-means cluster analysis in this study, and the results were visualized in Figure 7c. As shown in Figure 7c, the 40 water samples were divided into three groups around three clustering centers: (Group-I, Group-II, and Group-III). All of the CA samples belonged to Group-I, Group-II contained 9 TA samples, and Group-III included all LA samples and the other TA samples (TA 2, 4, 5, 9, 10, and 11). Both clustering methods show that some of the TA samples and LA samples were mixed, indicating that there is a hydraulic connection between them. Combined with Fig. 1b&c, it can be seen that the TA samples mixed with LA were located in the northeastern part of the mine area, and the direct contact between TA and LA in this area may be the main reason for the hydraulic connection between the two aquifers.

4. Conclusion

In this study, the bivariate diagrams, multivariate statistical methods and unmix model were used to understand the groundwater hydrogeochemical processes and aquifer hydraulic connections in the Pansan coal mine, Huanan coalfield, eastern China. The statistics shows that groundwater in loose layer aquifer has high Mg^{2+} , Cl^- and SO_4^{2-} concentrations, and groundwater from coal measure aquifer has high concentrations of Na^+ , HCO_3^- and CO_3^{2-} , whereas groundwater from limestone aquifer has the highest Ca^{2+} concentration relative to other aquifers. Piper diagram shows that groundwater in loose layer and limestone aquifers is dominated by the Na-Cl type, while groundwater in

coal measure aquifer is mainly of the Na- HCO_3 type. The results of bivariate diagrams ($\text{Na}^+ + \text{K}^+ - \text{Cl}^-$ versus $\text{Ca}^{2+} + \text{Mg}^{2+} - \text{SO}_4^{2-} - \text{HCO}_3^-$ and CAI-I versus CAI-II) and PCA/FA reveal that the weathering of silicate minerals, cation exchange, and dissolution of chloride and sulfate minerals are the dominate process controlling the groundwater chemistry. Unmix model evaluates the contribution rate of the source 1 (weathering of silicate minerals and cation exchange), source 2 (dissolution of sulfate minerals) and source 3 (dissolution of chloride minerals) for each major ion and each groundwater sample. Moreover, both clustering analysis methods (Q-type hierarchical cluster and K-means cluster) confirm the existence of a hydraulic connection between LA and TA in the northeastern part of the study area. This study shows the effectiveness of multivariate statistical analysis in interpreting hydrogeochemical processes and hydraulic connections in aquifers, and provides useful guidance for prevention of water inrush in the Pansan coal mine and other coal mines in the Huanan coalfield.

Declarations

Author contribution statement

Kai Chen: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Qimeng Liu: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Tingting Yang: Performed the experiments; Wrote the paper.

Qiding Ju: Contributed reagents, materials, analysis tools or data.

Yu Feng: Performed the experiments.

Funding statement

Dr. Kai Chen was supported by Innovation Project of Anhui University of Science & Technology [2022CX1001], 2021 Graduate Science Research Projects in Anhui Higher Education Institutions [YJS20210375 & YJS20210378], Natural Science Foundation of Anhui Province [1908085ME145].

Data availability statement

Data included in article/supp. material/referenced in article.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

- [1] L. Yuan, Scientific conception of precision coal mining, *J. China Coal Soc.* 42 (2017) 1–7.
- [2] L. Yuan, Research progress of mining response and disaster prevention and control in deep coal mines, *J. China Coal Soc.* 46 (2021) 716–725.
- [3] J. Zhang, L. Chen, X. Hou, J. Li, X. Ren, M. Lin, M. Zhang, Y. Wang, Y. Tian, Effects of multi-factors on the spatiotemporal variations of deep confined groundwater in coal mining regions, North China, *Sci. Total Environ.* 823 (2022) 153741.
- [4] K. Chen, L. Sun, J. Xu, Statistical analyses of groundwater chemistry in the Qingdong coalmine, northern Anhui province, China: implications for water–rock interaction and water source identification, *Appl. Water Sci.* (2021) 50.
- [5] J. Qian, Y. Tong, L. Ma, W. Zhao, R. Zhang, X. He, Hydrochemical characteristics and groundwater source identification of a multiple aquifer system in a coal mine, *Mine Water Environ.* 37 (2018) 528–540.
- [6] C. Song, L. Yao, J. Gao, C. Hua, Q. Ni, Identification model of water inrush source based on statistical analysis in Fengyu minefield, Northwest China, *Arabian J. Geosci.* 14 (2021) 518.
- [7] Y. Chen, L. Tang, S. Zhu, Comprehensive study on identification of water inrush sources from deep mining roadway, *Environ. Sci. Pollut. Res.* 29 (2022) 19608–19623.

- [8] Y. Chen, S. Zzhu, S. Xiao, Discussion on controlling factors of hydrogeochemistry and hydraulic connections of groundwater in different mining districts, *Nat. Hazards* 99 (2019) 689–704.
- [9] M. Jampani, R. Liedl, S. Hülsmann, S. Sonkamble, P. Amerasinghe, Hydrogeochemical and mixing processes controlling groundwater chemistry in a wastewater irrigated agricultural system of India, *Chemosphere* 239 (2020) 124741.
- [10] A. Barakat, M. El Baghdadi, J. Rais, B. Aghezzaf, M. Slassi, Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques, *Int. Soil Water Conserv. Res.* 4 (2016) 284–292.
- [11] X. Cao, Y. Lu, C. Wang, M. Zhang, J. Yuan, A. Zhang, S. Song, Y. Baninla, K. Khan, Y. Wang, Hydrogeochemistry and quality of surface water and groundwater in the drinking water source area of an urbanizing region, *Ecotoxicol. Environ. Saf.* 186 (2019) 109628.
- [12] Y. Zhai, F. Zheng, X. Zhao, X. Xia, Y. Teng, Identification of hydrochemical genesis and screening of typical groundwater pollutants impacting human health: a case study in Northeast China, *Environ. Pollut.* 252 (2019) 1202–1215.
- [13] K. Nakagawa, H. Amano, H. Asakura, R. Berndtsson, Spatial trends of nitrate pollution and groundwater chemistry in Shimabara, Nagasaki, Japan, *Environ. Earth Sci.* 75 (2016) 234.
- [14] J. Xu, J.-Y. Guo, G.-R. Liu, G.-L. Shi, C.-S. Guo, Y. Zhang, Y.-C. Feng, Historical trends of concentrations, source contributions and toxicities for PAHs in dated sediment cores from five lakes in western China, *Sci. Total Environ.* 470–471 (2014) 519–526.
- [15] A.E. Peter, S.M. Shiva Nagendra, I.M. Nambi, Comprehensive analysis of inhalable toxic particulate emissions from an old municipal solid waste dumpsite and neighborhood health risks, *Atmos. Pollut. Res.* 9 (2018) 1021–1031.
- [16] V. Murari, N. Singh, R. Ranjan, R.S. Singh, T. Banerjee, Source apportionment and health risk assessment of airborne particulates over central Indo-Gangetic Plain, *Chemosphere* 257 (2020) 127145.
- [17] F. Huang, X. Wang, L. Lou, Z. Zhou, J. Wu, Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques, *Water Res.* 44 (2010) 1562–1572.
- [18] A. Nagaraju, E. Balaji, L.H. Sun, A. Thejaswi, Processes controlling groundwater chemistry from Mulakalacheruvu area, Chittoor District, Andhra Pradesh, South India: a statistical approach based on hydrochemistry, *J. Geol. Soc. India* 91 (2018) 425–430.
- [19] K. Chen, Q. Liu, W. Peng, X. Liu, Source apportionment and natural background levels of major ions in shallow groundwater using multivariate statistical method: a case study in Huaibei Plain, China, *J. Environ. Manag.* 301 (2022) 113806.
- [20] N. Subba Rao, A. Dinakar, L. Sun, Estimation of groundwater pollution levels and specific ionic sources in the groundwater, using a comprehensive approach of geochemical ratios, pollution index of groundwater, unmix model and land use/land cover – a case study, *J. Contam. Hydrol.* 248 (2022) 103990.
- [21] K. Chen, L. Sun, J. Tang, Hydrochemical differences between river water and groundwater in Suzhou, northern Anhui province, China, *Open Geosci.* 12 (2020) 1421–1429.
- [22] Y. Teng, B. Hu, J. Zheng, J. Wang, Y. Zhai, C. Zhu, Water quality responses to the interaction between surface water and groundwater along the Songhua River, NE China, *Hydrol. J.* 26 (2018) 1591–1607.
- [23] O. Ghesquière, J. Walter, R. Chesnaux, A. Rouleau, Scenarios of groundwater chemical evolution in a region of the Canadian Shield based on multivariate statistical analysis, *J. Hydrol.: Reg. Stud.* 4 (2015) 246–266.
- [24] H. Zhang, G. Xu, X. Chen, A. Mabaire, J. Zhou, Y. Zhang, G. Zhang, L. Zhu, Groundwater hydrogeochemical processes and the connectivity of multilayer aquifers in a coal mine with karst collapse columns, *Mine Water Environ.* 39 (2020) 356–368.
- [25] J. Chen, G. Liu, H. Li, B. Wu, Mineralogical and geochemical responses of coal to igneous intrusion in the Pansan Coal Mine of the Huainan coalfield, Anhui, China, *Int. J. Coal Geol.* 124 (2014) 11–35.
- [26] J. Qian, L. Wang, L. Ma, Y. Lu, W. Zhao, Y. Zhang, Multivariate statistical analysis of water chemistry in evaluating groundwater geochemical evolution and aquifer connectivity near a large coal mine, Anhui, China, *Environ. Earth Sci.* 75 (2016) 747.
- [27] X. Chen, C. Jiang, L. Zheng, L. Zhang, X. Fu, S. Chen, Y. Chen, J. Hu, Evaluating the genesis and dominant processes of groundwater salinization by using hydrochemistry and multiple isotopes in a mining city, *Environ. Pollut.* 283 (2021) 117381.
- [28] G. Wang, J. Wu, X. Zhai, H. Zhang, Y. Bi, Temporal Variation of Hydrogeochemical Characteristics and Processes of Aquifers in the Liuqiao Coal Mine, *Energy Exploration & Exploitation*. (2022) 014459872210886.
- [29] R.K. Blashfield, M.S. Aldenderfer, The Literature on cluster Analysis, *Multivariate Behav. Res.* 13 (1978) 271–295.
- [30] L. Liu, Z. Tang, M. Kong, X. Chen, C. Zhou, K. Huang, Z. Wang, Tracing the potential pollution sources of the coastal water in Hong Kong with statistical models combining APCS-MLR, *J. Environ. Manag.* 245 (2019) 143–150.
- [31] C.C. Bridges, Hierarchical cluster Analysis, *Psychol. Rep.* 18 (1966) 851–854.
- [32] T.M. Kodinariya, P.R. Makwana, Review on determining number of cluster in K-means clustering, *Int. J.* 1 (2013) 90–95.
- [33] M. Haji Gholizadeh, A.M. Melesse, L. Reddi, Water quality assessment and apportionment of pollution sources using APCS-MLR and PMF receptor modeling techniques in three major rivers of South Florida, *Sci. Total Environ.* (2016) 566–567, 1552–1567.
- [34] H. Zhang, S. Cheng, H. Li, K. Fu, Y. Xu, Groundwater pollution source identification and apportionment using PMF and PCA-APCA-MLR receptor models in a typical mixed land-use area in Southwestern China, *Sci. Total Environ.* 741 (2020) 140383.
- [35] M.S. Gulgundi, A. Shetty, Source apportionment of groundwater pollution using Unmix and positive matrix factorization, *Environ. Process.* 6 (2019) 457–473.
- [36] N. Subba Rao, B. Sunitha, L. Sun, B. Deepthi Spandana, M. Chaudhary, Mechanisms controlling groundwater chemistry and assessment of potential health risk: a case study from South India, *Geochemistry* 80 (2020) 125568.
- [37] R.C. Henry, Multivariate receptor modeling by N-dimensional edge detection, *Chemometr. Intell. Lab. Syst.* 65 (2003) 179–189.
- [38] A.M. Piper, A graphic procedure in the geochemical interpretation of water-analyses, *Trans. AGU.* 25 (1944) 914.
- [39] J. Zhang, L. Chen, X. Hou, X. Ren, J. Li, Y. Chen, Hydrogeochemical processes of carboniferous limestone groundwater in the Yangzhuang coal mine, Huaibei coalfield, China, *Mine Water Environ.* 41 (2022) 504–517.
- [40] M. Zhang, L. Chen, D. Yao, X. Hou, J. Zhang, H. Qin, X. Ren, X. Zheng, Hydrogeochemical Processes and Inverse Modeling for a Multilayer Aquifer System in the Yuaner Coal Mine, Huaibei Coalfield, China, *Mine Water Environ.* 2022.
- [41] P. Liu, N. Hoth, C. Drebenstedt, Y. Sun, Z. Xu, Hydro-geochemical paths of multilayer groundwater system in coal mining regions — using multivariate statistics and geochemical modeling approaches, *Sci. Total Environ.* (2017) 601–602, 1–14.
- [42] J. Liu, Y. Peng, C. Li, Z. Gao, S. Chen, A characterization of groundwater fluoride, influencing factors and risk to human health in the southwest plain of Shandong Province, North China, *Ecotoxicol. Environ. Saf.* 207 (2021) 111512.
- [43] K. Hua, J. Xiao, S. Li, Z. Li, Analysis of hydrochemical characteristics and their controlling factors in the Fen River of China, *Sustain. Cities Soc.* 52 (2020) 101827.
- [44] Z. Gao, C. Han, Y. Xu, Z. Zhao, Z. Luo, J. Liu, Assessment of the water quality of groundwater in Bohai Rim and the controlling factors—a case study of northern Shandong Peninsula, north China, *Environ. Pollut.* 285 (2021) 117482.
- [45] L. Fang, A. Zhou, X. Li, J. Zhou, G. Pan, N. He, Response of antimony and arsenic in karst aquifers and groundwater geochemistry to the influence of mine activities at the world's largest antimony mine, central China, *J. Hydrol.* 603 (2021) 127131.