Statistical analyses of groundwater chemistry in the Qingdong coalmine, northern Anhui province, China: implications for water–rock interaction and water source identification

Kai Chen^{1,2} · Linhua Sun^{1,2} · Jiying Xu^{1,3}

Received: 4 July 2018 / Accepted: 1 February 2021 / Published online: 17 February 2021
© The Author(s) 2021

Abstract

Hydrochemistry of groundwater is important in coal mines because it can be used for understanding water–rock interaction and inrush water source identification. In this study, major ion concentrations of groundwater samples from the loose layer aquifer (LA), coal-bearing aquifer (CA) and Taiyuan Formation limestone aquifer (TA) in the Qingdong coal mine, northern Anhui province, China, have been analyzed by a series of statistical methods for identifying the source of chemical constituents in groundwater and the source of inrush water. The results indicate that the mean concentration of the major ions in the LA were ordered as follows: $\text{HCO}_3^- > \text{SO}_4^{2-} > \text{Na}^+ + \text{K}^+ > \text{Cl}^- > \text{Ca}^{2+} > \text{Mg}^{2+} > \text{CO}_3^{2-}$, whereas average values of the CA in decreasing order are SO_4^{2-} , $\text{Na}^+ + \text{K}^+$, HCO_3^- , Cl^- , Ca^{2+} , Mg^{2+} and CO_3^{2-} , and the major ion concentrations of the TA have the following order: $\text{SO}_4^{2-} > \text{Na}^+ + \text{K}^+ > \text{Ca}^{2+} > \text{HCO}_3^- > \text{Cl}^- > \text{Mg}^{2+} > \text{CO}_3^{2-}$, and most of the samples are Na- SO_4 and Ca- SO_4 types. TDS content in water increases with aquifer depth, whereas the pH values ranged from 7.1 to 8.9, indicating a weak alkaline environment. Two sources (weathering of silicate minerals and dissolution of evaporate minerals) have been identified by principal component analysis responsible for the chemical variations of the groundwater, and their contribution ratios have been quantified by Unmix model. Moreover, based on the Q-mode cluster and discriminant analyses, the samples with known sources have been identified correctly to be 95.7% and 97.6%, respectively, and the samples with unknown sources have been determined with high probability (78–100%).

Keywords Groundwater · Hydrochemistry · Water–rock interaction · Water inrush · Coalmine

Introduction

Coal is a primary source of energy to China, and this situation will last for decades in the future. However, just because of the high production of coal, a series of disasters have occurred in the process of coal mining. Previous studies revealed that five kinds of disasters (water, fire, gas, dust and roof) have brought to mankind, huge economic losses or the death of people (Wu et al. 2013), and the reasons have

been attributed to be the complicated geological conditions, the imperfect mining technology or, the subjective neglect of the management.

In response to this situation, two types of strategies have been applied for the prevention of water inrush in coalmines: (1) the active one, which means to predict the possibility of water inrush based on the understanding of the hydro-geological conditions of the coal mine; (2) the passive one, which means to do something to minimize the loss after the water inrush (e.g., draining out of the water and recovering coal production) (Xu and Bu 2015).

Except for the man-made factors, a series of natural factors have been identified responsible for the occurrence of water inrush accident in coal mines: e.g., the quantity of water in the aquifers, the channel for water flow, the hydro-connection between aquifers and the thickness of the water resisting layer (Gui and Chen 2007). And therefore, understanding the hydrological condition of the coalmine before the water inrush is the most fundamental work for

✉ Linhua Sun
sunlinh@126.com

¹ School of Resources and Civil Engineering, Suzhou University, Anhui 234000, China

² School of Earth and Environment, Anhui University of Science and Technology, Anhui 232001, China

³ National Engineering Research Center of Coal Mine Water Hazard Controlling, Anhui 234000, China

the prevention of water inrush-related accidents in coalmines, and getting the reliable information about the source of inrush water is the most important work after water inrush.

Groundwater in the aquifers contains a series of information, including the water level, quantity, temperature, velocity, turbidity and hydrochemistry, among others (Gui and Chen 2007), which provides the possibility of the water source identification among different aquifers, and a large number of methods have been proposed (Liu et al. 2018; Shi et al. 2018; Wu et al. 2019; Wang et al. 2020). For example, the mixing ratio of water inrush sources in the Qidong coal mine, Anhui province, China, has been calculated by factor analysis, and results provide theoretical guidance for the prevention and control of water disasters (Sun and Gui 2015). Moreover, the source of inrush water had been identified by means of Q-model and principal component analysis in Zhaizhen coal mine, Shandong Province, China (Wang and Shi 2019). These studies confirmed that statistical analysis of groundwater chemistry is effective for identifying the inrush water sources in coal mines.

Roofing, gas, dust and other hazards have been fully studied in the Qingdong coal mine, northern Anhui province, China, but the source of inrush water has not received corresponding attention, which is a potential threat to coal production (Liu et al. 2017; Si et al. 2018). Therefore, in this study, forty-six groundwater samples have been collected

from the representative aquifer systems in the Qingdong coal mine, and a series of statistical methods (including principal component, cluster and discriminant analysis, as well as the analysis of the Unmix model) have been applied to their major ion concentrations, for getting the information about: (1) identifying the different types of water–rock interactions for determining the chemical compositions of the groundwater; (2) quantifying the contributions of different types of water–rock interactions and (3) establish the water source identification model for the coalmine.

Materials and methods

Study area

The coalfield in the northern Anhui province is an important energy base in China with annual output of coal over 100 million tons. The Qingdong coalmine is one of the thirty coalmines in the region. It is situated 45 km west to the Suzhou City, northern Anhui Province, China (Fig. 1), and the location of the coalmine is $116^{\circ}25'44''$ – $116^{\circ}34'44''$ (longitude) and $33^{\circ}36'34''$ – $33^{\circ}40'29''$ (latitude). The study area belongs to the monsoon warm temperate semi-humid climate, with an average temperature of 14.6°C over the years. The average annual precipitation is 811.8 mm, and

Fig. 1 Location of the Qingdong coal mine

the rainfall is mainly concentrated in July and August. The length of the mine is 3.5–6.5 km from south to north, and the width is 13 km from east to west, the total area of the mine is 51.7 km².

The aquifer in the study area from shallow to deep can be divided into four systems: loose layer (LA), coal-bearing sandstone (CA), Taiyuan Formation limestone (TA) and Ordovician limestone (OA). Details of the four aquifer systems are as follows.

LA: The coalmine is covered by the loose layer sediments, including the Tertiary and Quaternary, the average thickness of the Tertiary and Quaternary are 154 m and 92 m, respectively. The main rock types in the LA including clay, sandy clay and calcareous clay. Multi-layer composite structure is observed in the loose layer, and four secondary aquifers were further identified. In these four aquifers, the fourth aquifer is highly water-rich and can be recharged from rainfall. Therefore, the water in the fourth aquifer is a threat for the safety of coal mining due to direct contact with the coal-bearing strata.

CA: Permian is the main coal-bearing strata in the area, and the main constituent in CA including mudstone, siltstone, sandstone (mainly feldspathic quartz sandstone) and coal seams, and three secondary aquifers can be subdivided. The water in this aquifer is stored primarily in sandstone fractures. In recent years, mining activities have disrupted the integrity of the formation and altered the original runoff environment, resulting in water being able to enter the coal mining tunnel by infiltration. The water in the crack of sandstone is the main threat.

TA and OA: Comparatively, the primary rock type in the TA and OA is limestone, to a lesser extent, the mudstone and sandstone. These two aquifers are highly water-rich, with high water pressure and poor runoff conditions without human activities. Therefore, the water in the karst cave is also a threat to the safety of coal mining. In summary, the water from the fourth aquifer, the crack of the sandstone and the karst cave of the limestone threaten the safety of coal mining activities, which is similar to the Renlou coal mine in the similar hydrological unit (Sun 2013).

Sampling and analysis

Forty-six groundwater samples were collected from four representative aquifer systems in the Qingdong coal mine: 7 from the LA, 16 from the CA, and 19 from the TA. LA samples were extracted from the wells using a pump, CA samples were collected directly in the coal mining tunnel and TA samples were collected from boreholes. Other four samples (UN1 and 2 were from the boreholes of the CA, and the UN3 and 4 were from the boreholes of the TA), have been defined as unknown source samples for verifying the effect of water source identification. The contents of the

physico-chemical parameters and major ions in the water samples were measured as follows.

pH values were measured by a portable water analyzer with an accuracy of 0.01. TDS values were calculated by subtracting 0.5 times bicarbonate from the sum of all ion concentrations (Unit: mg/L). Concentrations of eight kinds of major ions (Na⁺, K⁺, Ca²⁺, Mg²⁺, Cl⁻, SO₄²⁻, HCO₃⁻ and CO₃²⁻) were analyzed in this study: Na⁺, K⁺, Ca²⁺, Mg²⁺, Cl⁻ and SO₄²⁻ were analyzed by ion chromatography, whereas the HCO₃⁻ and CO₃²⁻ were analyzed by acid–base titration. Due to the low concentrations of K⁺, Na⁺ and K⁺ were merged and presented as (Na⁺ + K⁺). All of the analyses were conducted in the Engineering and Technological Research Center of Coal Exploration, Anhui Province, China.

After all ion concentration tests are completed, the AE (analytical errors) were checked by calculating the mass balance using the following formula.

$$AE = 100 * \left(\frac{meq/L, cations - meq/L, anions}{meq/L, cations + meq/L, anions} \right)$$

With the exception of one water sample, which had an AE of 12%, the AEs of all other water samples were within –5% to +5%, indicating that the results of the analyses were reasonable and credible (Nordstrom et al. 1989).

Statistical analyses

In this study, a series of statistical methods were applied to the hydrochemical data of the groundwater: (1) For getting the information about the water–rock interaction (mainly the source of major ions), the factor analysis was applied for getting qualitative information about the source of the major ions, whereas the US EPA (US Environmental Protection Agency) Unmix model was applied for getting quantitative information. (2) For the water source identification, cluster and discriminant analyses were applied along with the method related to the source contributions generated by the Unmix model. Detailed descriptions of these statistical methods are contained in the following text.

1. Correlation analysis: a multivariate statistical analysis method that can be used to evaluate the degree of correlation between two variables at significance levels. Previous studies have confirmed that two or more elements of significant correlation usually come from similar origins or are controlled by the same processes (Sunkari and Abu 2019; Zango et al. 2019).
2. Principal component analysis: an important statistical method that is not only used to simplify and classify the data, but also is used to find the most important variables in the data. During geochemical studies, factor analy-

sis has long been used to track the sources of element (Yidana et al. 2018; Wang and Shi 2019; Gubran et al. 2019; Ramarosan et al. 2020).

- Cluster analysis: it is a multivariate method applied to find the true groups of data or stations. In clustering, the objects are grouped in order to divide the similar objects into the same class. The method has been applied to the hydrochemical studies for a long time (Tay et al. 2015; Bencer et al. 2016; Chen et al. 2020).
- Discriminant analysis: it uses linear or quadratic functions of the variables that are "best" divided into two or more predefined groups in a separate case. It is a simple and effective tool for the establishment of the intrush water source identification model (Sun and Gui 2015).
- Unmix model analysis: it is a mathematical receptor model used for quantifying the sources of contaminants (Lang and Yang 2014; Liu and Sun 2015; Jain et al. 2018; Zhang et al. 2019). The source types are identified by comparing them to measured profiles, whereas the source contributions are used to determine how much each source contributed to a sample. The basic idea of the model is $C_m = \sum X_1 \times S_1 + X_2 \times S_2 \dots X_n \times S_n$. Where C_m is the measured concentration of sample, S_n is the profile contributed by any types of sources, X_n is the portion of S_n , and $(X_n \times S_n) / \sum (X_n \times S_n)$ is the contribution of source n for the C_m . The hydrochemistry of the groundwater is also a multi-source contribution issue, the different types of water–rock interactions can be considered as sources, whereas the degrees of water–rock interaction can be considered as contributions.

Moreover, Unmix model analysis was performed by EPA Unmix software (version 6.0), and RStudio software (Team, Core R 2013) was employed to conduct the correlation analysis, principal component analysis, cluster analysis and discriminant analysis.

Results and discussion

Hydrochemistry

Major ion concentrations of the groundwater samples from the different aquifer systems are shown in Table 1 and Fig. 2. As shown in Table 1, the mean concentration of the major ions in the LA were ordered as follows: HCO_3^- (370 mg/L) > SO_4^{2-} (275 mg/L) > $\text{Na}^+ + \text{K}^+$ (223 mg/L) > Cl^- (97 mg/L) > Ca^{2+} (39 mg/L) > Mg^{2+} (38 mg/L) > CO_3^{2-} (21 mg/L), whereas average values of the CA in decreasing order are SO_4^{2-} (1000 mg/L), $\text{Na}^+ + \text{K}^+$ (647 mg/L), HCO_3^- (334 mg/L), Cl^- (184 mg/L), Ca^{2+} (46 mg/L), Mg^{2+} (23 mg/L) and CO_3^{2-} (17 mg/L), and the major ion concentrations of the TA have the following order: SO_4^{2-} (1556 mg/L) > $\text{Na}^+ + \text{K}^+$ (384 mg/L) > Ca^{2+} (321 mg/L) > HCO_3^- (258 mg/L) > Cl^- (225 mg/L) > Mg^{2+} (133 mg/L) > CO_3^{2-} (1 mg/L).

As can be seen from Fig. 2, the $\text{Na}^+ + \text{K}^+$, Cl^- and SO_4^{2-} contents of groundwater from the three aquifers were significantly different ($p < 0.05$, Fig. 2a, d and e), and the LA and CA samples had similar concentrations of Ca^{2+} , Mg^{2+} and HCO_3^- ($p < 0.05$, Fig. 2b, c and f). In general, the groundwater samples from the CA have the highest mean concentrations of $\text{Na}^+ + \text{K}^+$ (647 mg/l) compared with other two aquifer systems, whereas the groundwater samples from the TA have highest mean concentrations of Ca^{2+} (321 mg/l), Mg^{2+} (133 mg/l), Cl^- (225 mg/l), and SO_4^{2-} (1556 mg/l) and the groundwater samples from the LA have highest mean concentrations of HCO_3^- (370 mg/l) (Table 1). In common, the major ion concentrations in groundwater are closely related to the types of surrounding rock. Therefore, high content of $\text{Na}^+ + \text{K}^+$ in CA samples are associated with weathering of silicates, while high Ca^{2+} and Mg^{2+} contents in TA samples are associated with dissolution of carbonates and sulfates. Moreover, from shallow to deep, the concentration of HCO_3^- in groundwater shows a decreasing trend, while the concentrations of SO_4^{2-} and Cl^- have an increasing trend, implying poor runoff conditions for CA and TA.

Table 1 Major ion concentrations (mg/L) of the groundwater samples

Aquifer	$\text{Na}^+ + \text{K}^+$	Ca^{2+}	Mg^{2+}	Cl^-	SO_4^{2-}	HCO_3^-	CO_3^{2-}	pH	TDS
LA (n=7)	157–265 (223)	3–72 (39)	11–65 (38)	24–161 (97)	100–427 (275)	7–729 (370)	0–76 (21)	7.2–9.1 (8.1)	689–1106 (884)
CA (n=16)	326–799 (647)	11–119 (46)	5–61 (23)	83–233 (184)	381–1333 (1000)	191–525 (334)	0–98 (17)	7.7–8.9 (8.3)	1105–2649 (2081)
TA (n=19)	281–566 (384)	196–553 (321)	77–193 (133)	108–283 (225)	1255–2064 (1556)	188–317 (258)	0–14 (1)	7.1–8.5 (7.7)	2322–3524 (2751)
UN (n=4)	386–697 (534)	56–433 (208)	26–337 (144)	89–205 (145)	1115–3123 (1809)	151–293 (232)	0–14 (5)	7.7–8.3 (7.9)	2077–4557 (2948)

Fig. 2 Box plot of major ions and physico-chemical parameters contents of groundwater samples in different aquifers

As shown in Fig. 2g and h, the average TDS content in groundwater of LA, CA and TA increased from 884 mg/L, 2081 mg/L to 2751 mg/L, which means that the degree of water–rock interaction deepens with increasing aquifer depth, and the pH values ranged from 7.1 to 8.9, indicating a weak alkaline environment.

During the groundwater flows from the recharge to the discharge zone, the hydrochemical types will change from HCO_3^- , SO_4^{2-} to Cl^- (Tóth, 1999). And therefore, the classification of the hydrochemical types is important for the understanding of the hydrological evolution of the groundwater system. Classification of groundwater in this study is based on the concentration of cations and anions by using the Piper diagram, and the results are shown in Fig. 3. The results indicate that Na- SO_4 type is the most popular one, which occupies 71% of the LA, 100% of the CA and 42% of the TA samples, respectively. Moreover, the rest of the LA samples are Na- HCO_3 type (29%), and the 47% and 11% of the CA groundwater samples are classified as Ca- SO_4 and Mg- SO_4 types, respectively. Comparatively, the UN1 and UN2 groundwater samples are classified into Na- SO_4 types, whereas the UN3 and UN4 groundwater samples are classified into Mg- SO_4 and Na- SO_4 types. It can also be noticed from Fig. 3 that the UN1 and 2 samples are similar to the CA samples, whereas the UN3 and 4 samples are similar to the TA samples.

Mechanism controlling groundwater chemistry

The Gibbs diagram (Gibbs 1970) can be applied to figure out the relationship between the chemical components of groundwater and the lithology from their respective aquifer (Sunkari et al. 2020; Yu et al. 2020). The three factors controlling groundwater chemistry can be classified by the diagram: precipitation, evaporation and water–rock interaction dominance. The formulations of Gibbs are $\text{Gibbs I} = \text{Cl}^- / (\text{Cl}^- + \text{HCO}_3^-)$ and $\text{Gibbs II} = (\text{Na}^+ + \text{K}^+) / (\text{Na}^+ + \text{K}^+ + \text{Ca}^{2+})$ (unit in meq/l). In this study, Gibbs I values variation ranged from 0.05 to 0.97 and Gibbs II values variation ranged from 0.38 to 0.99. As can be seen from Fig. 4, all of the samples are located in the water–rock interaction and evaporation dominant areas, which indicated that the water–rock interaction and evaporation have played important roles in the groundwater system for controlling the groundwater chemistry. Meanwhile, from LA to TA, the TDS content in groundwater has an increasing trend, indicating that the degree of water–rock interaction increases with the increase of the aquifer depth.

Weathering of silicate minerals (feldspars), dissolution of halite and cation exchange reaction are generally considered to be the major sources of Na^+ in groundwater (Zango et al. 2019). As can be seen in Fig. 5a and b, most of the water samples in the three aquifers are above the 1:1 line,

Fig. 3 Piper diagram

which may be related to intense cation exchange reaction, because Na^+ replacement of Ca^{2+} and Mg^{2+} will lead to Na^+ enrichment in groundwater. Ca^{2+} and Mg^{2+} in groundwater are associated with dissolution of carbonate and sulfate and reverse ion exchange. As shown in Fig. 5c, the $(\text{Ca}^{2+} + \text{Mg}^{2+})/(\text{HCO}_3^- + \text{SO}_4^{2-})$ values in the groundwater of the three aquifers were different, the TA samples were close to the 1:1 line and the concentrations of $(\text{Ca}^{2+} + \text{Mg}^{2+})$ and $(\text{HCO}_3^- + \text{SO}_4^{2-})$ were significantly higher than those in the LA and CA samples, indicating the influence of carbonate and sulfate dissolution in TA. Moreover, $(\text{Ca}^{2+} + \text{Mg}^{2+})/(\text{HCO}_3^- + \text{SO}_4^{2-}) < 1$ in all samples indicates the widespread presence of cation exchange reaction in the three aquifers, especially for CA, which is also consistent with the results of Fig. 5a and b. SO_4^{2-} in groundwater is usually derived from the dissolution of sulfate minerals and the oxidation of sulfides. The high SO_4^{2-} content in CA samples relative

to Ca^{2+} and Mg^{2+} was due to the combined effects of ion-exchange reaction and oxidation of sulfides (Fig. 5c and d). In the study area, there are a large amount of pyrite in CA, and coal mining activities have led to long-term exposure of the CA to air, which provides favorable conditions for the oxidation of pyrite and then leads to the enrichment of SO_4^{2-} in groundwater ($4\text{FeS}_2 + 13\text{O}_2 + 2\text{H}_2\text{O} \rightarrow 4\text{FeSO}_4 + 2\text{H}_2\text{SO}_4 + 2\text{SO}_2\uparrow$).

Source of major ions

Correlation analysis

In this study, the Spearman correlation method was chosen, which is insensitive to outliers in the data, and the analysis results are shown in Table 2. As shown in Table 2, Ca^{2+} , Mg^{2+} , Cl^- , and SO_4^{2-} have significant positive correlations

Fig. 4 Gibbs diagrams

Fig. 5 Major ion relationship in groundwater

Table 2 Spearman correlation of the major ions in groundwater (** represent significant correlation at the level of 0.01)

	Na ⁺ + K ⁺	Ca ²⁺	Mg ²⁺	Cl ⁻	SO ₄ ²⁻	HCO ₃ ⁻
Na ⁺ + K ⁺	1.000					
Ca ²⁺	-0.265	1.000				
Mg ²⁺	-0.456**	0.903**	1.000			
Cl ⁻	0.088	0.657**	0.397**	1.000		
SO ₄ ²⁻	0.152	0.804**	0.690**	0.698**	1.000	
HCO ₃ ⁻	0.126	-0.482**	-0.437**	-0.484**	-0.564**	1.000

**Represent significant correlation at the level of 0.01

with each other, which can be explained by the dissolution of chlorides and sulfate minerals, because they can release Ca²⁺, Mg²⁺, Cl⁻ and SO₄²⁻ into the water. Furthermore, the significant negative correlation between Na⁺ + K⁺ and Mg²⁺ is usually related to the ion-exchange reaction (Table 2).

Principal component analysis

In this study, two principal components were obtained based on the initial eigenvalue higher than one, and the results was shown in Table 3 and Fig. 6. As can be seen from the table, the PC1, which accounts for 52.6% of the total variance explained, has high positive loadings of Ca²⁺, Mg²⁺, Cl⁻ and SO₄²⁻, but high negative loadings of HCO₃⁻, the second component with 23.6% of the total variance explained, has high positive loadings of Na⁺ + K⁺ and negative loadings of Mg²⁺. Based on the investigations of previous studies,

the main processes controlling the groundwater chemistry in the study area include weathering of silicate minerals, ion exchanged reaction, dissolution of carbonate, sulfate and chloride minerals (Sun and Gui 2015). Therefore, the PC2 can be explained to be the weathering of silicate minerals and ion-exchange reaction, which can lead to high concentration of Na⁺ + K⁺ in water, whereas the PC1 is related to the dissolution of chlorides and sulfate minerals, because they can release Ca²⁺, Mg²⁺, Cl⁻ and SO₄²⁻ into the water. It is also observed in Fig. 6 that the groundwater of the three aquifers has significantly different scores on PC1 and PC2 ($p < 0.05$): the LA samples have the lowest PC1 and PC2 scores, and the CA samples have the highest PC2 scores, whereas the TA samples have the highest PC1 scores. Such a phenomenon indicates that groundwater samples from different aquifer systems have undergone different types and degrees of water–rock interactions: CA-dominated by

Fig. 6 Principal component scores of groundwater samples in different aquifers

Table 3 Results of principal component analysis

Species	PC 1	PC 2
$\text{Na}^+ + \text{K}^+$	-0.190	0.909
Ca^{2+}	0.927	-0.212
Mg^{2+}	0.850	-0.398
Cl^-	0.623	0.500
SO_4^{2-}	0.875	0.273
HCO_3^-	-0.619	-0.253
Eigen value	3.155	1.418
Variance explained (%)	52.6	23.6

Table 4 Source profiles (mg/l)

Species	Source 1	Source 2	Proportion 1 (%)	Proportion 2 (%)
$\text{Na}^+ + \text{K}^+$	389	72.6	84	16
Ca^{2+}	3.24	161	2	98
Mg^{2+}	3.41	71.9	5	95
Cl^-	109	74.1	60	40
SO_4^{2-}	509	672	43	57
HCO_3^-	233	54.6	81	19

the weathering of silicates and ion-exchange reaction; TA-dominated by the dissolution of chloride and sulfate; LA-low degrees of water–rock interaction relative to the CA and TA.

The Unmix model approach

Based on the calculation of the Unmix model, two sources have been identified and the results are listed in Table 4 and shown in Fig. 7. As can be seen in Table 4, the two sources have Min Rsq = 0.98 and Min Sig/Noise = 3.72, higher than the minimum standard of the model (Min Rsq > 0.8 and Min Sig/Noise > 2), indicating that the modeling is invalid (Ai et al. 2014). As suggested in the above text, the weathering of silicate minerals, dissolution of carbonate and evaporate minerals are the main processes controlling the groundwater chemistry in the coal mining related strata (Sun and Gui 2015). Therefore, the Source 1 dominated by $\text{Na}^+ + \text{K}^+$, HCO_3^- and Cl^- should be explained to be the weathering of silicate minerals and dissolution of chloride, and the source 2 dominated by Ca^{2+} , Mg^{2+} and SO_4^{2-} can be explained to be the dissolution of evaporate minerals (sulfates). Such considerations can be demonstrated by the contributions of these sources (Fig. 7), as the samples from the LA and CA have highest contributions from the source 1 (> 60%), consistent with the assertion that this aquifer system is richer in elastic rocks (silicate minerals are the main phases) relative to the TA, which was mainly contributed by the source 2 (> 50%). As to the samples with unknown sources, the UN1 and 2 have high contributions from the source 1 (similar to the LA and CA samples), whereas the UN3 and 4 have high contributions from the source 2 (similar to the TA samples).

Fig. 7 Variations of source contributions. Samples 1–7, 8–23 and 24–42 are the LA, CA and TA samples, respectively, whereas samples and 43–46 are UN samples

Source identification of inrush water

Cluster analysis

In this study, all the samples were analyzed by Q-model cluster analysis and the "Ward" linkage and "Euclidean" distance were chosen to obtain better results (Fovell and Fovell 1993; Liu et al. 2015; Chen et al. 2020). As can be seen from Fig. 8, three groups can be identified: all the samples from the LA and two samples from CA have been classified into the Group 1, and most of the samples from the CA have been classified into the Group 2 (except for the CA4 and CA11), whereas all the TA samples have been classified into the Group 3. Therefore, only two samples (CA4 and CA11) have been classified into the wrong category, and the accuracy rate is 95.7%. It can also be noticed from the figure that the UN1 and UN2 samples are classified to be the Group 2, and the UN3 and UN4 are classified to be the Group 3, similar to the results obtained from the plots of factor scores (Fig. 6) and the variations of source contributions (Fig. 7).

Discriminant analysis

In this study, the LA, CA and TA groundwater samples are assigned to be 1, 2 and 3, respectively, whereas the UN samples are signed to be "blank." The classical method has been chosen for calculation and the results are shown in Table 5 and shown in Fig. 9. As can be seen from the table, all of the samples with known sources except for one from the

CA (18) have been classified to be the right category with high probability (92–100%), and the source-known samples have been identified correctly to be 97.6% (41/42), and the discriminant functions are as follows (unit in mg/l):

1. Function 1 = $0.001 \times (\text{Na}^+ + \text{K}^+) + 0.002 \times \text{Ca}^{2+} - 0.039 \times \text{Mg}^{2+} - 0.013 \times \text{Cl}^- - 0.002 \times \text{SO}_4^{2-} - 0.001 \times \text{HCO}_3^- + 7.120$;
2. Function 2 = $0.011 \times (\text{Na}^+ + \text{K}^+) - 0.001 \times \text{Ca}^{2+} + 0.018 \times \text{Mg}^{2+} + 0.014 \times \text{Cl}^- - 0.002 \times \text{SO}_4^{2-} + 0.001 \times \text{HCO}_3^- - 7.393$.

Based on the discrimination model established by the known samples, the source of the UN1 and 2 samples are classified to be the CA with 78% and 100% probabilities, respectively, and the source of the UN3 and 4 samples are classified to be the TA with 100% and 87% probabilities, respectively, similar to the results of cluster analysis.

Conclusions

Based on the statistical analyses of major ion concentrations of groundwater samples collected from three representative aquifer systems in the Qingdong coal mine, northern Anhui province, China, the following conclusions have been obtained: (1) The major ion concentrations of the groundwater samples from different aquifer systems are different with each other, which suggests that the types and degrees

Fig. 8 Plot of Q-mode cluster analysis

Table 5 Results of discriminant analysis

ID	Source	Predicted	Probability (%)	ID	Source	Predicted	Probability (%)
1	LA	LA	100	24	TA	TA	100
2	LA	LA	100	25	TA	TA	100
3	LA	LA	99	26	TA	TA	100
4	LA	LA	100	27	TA	TA	100
5	LA	LA	100	28	TA	TA	100
6	LA	LA	100	29	TA	TA	100
7	LA	LA	100	30	TA	TA	100
8	CA	CA	99	31	TA	TA	100
9	CA	CA	100	32	TA	TA	100
10	CA	CA	100	33	TA	TA	100
11	CA	CA	91	34	TA	TA	100
12	CA	CA	100	35	TA	TA	100
13	CA	CA	100	36	TA	TA	100
14	CA	CA	100	37	TA	TA	100
15	CA	CA	100	38	TA	TA	100
16	CA	CA	100	39	TA	TA	92
17	CA	CA	100	40	TA	TA	100
18	CA	LA	95	41	TA	TA	100
19	CA	CA	100	42	TA	TA	100
20	CA	CA	100	43	UN1	CA	78
21	CA	CA	100	44	UN2	CA	100
22	CA	CA	100	45	UN3	TA	100
23	CA	CA	92	46	UN4	TA	87

Fig. 9 Discriminant diagram

of water–rock interactions in these three aquifer systems are different. (2) Principal component analysis has identified two sources responsible for the chemical variations of the groundwater: weathering of silicate minerals and dissolution of evaporates, and their contribution ratios for the groundwater samples from different aquifer systems are different with

each other, which have been quantified by Unmix model. (3) Based on the Q-mode cluster and discriminant analyses, the samples with known sources have been identified correctly to be 95.7% and 97.6%, respectively, and the samples with unknown sources have been determined with high probability (78–100%).

Acknowledgements This work was financially supported by the Natural Science Research Project of Universities in Anhui Province (KJ2020ZD64 and KJ2020A0740) and the Scientific Foundation of Platform in Suzhou University (2016ykf02).

Compliance with ethical standards

Conflict of interest The authors have declared no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ai JC, Wang N, Yang J (2014) Source apportionment of soil heavy metals in Jiapiougou goldmine based on the UNMIX model. *Environ Sci* 35(9):3530–3536
- Bencer S, Boudoukha A, Mouni L (2016) Multivariate statistical analysis of the groundwater of Ain Djacer area (Eastern of Algeria). *Arab J Geosci* 9(4):1–10
- Chen K, Sun LH, Tang J (2020) Hydrochemical differences between river water and groundwater in Suzhou, Northern Anhui Province. *China Open Geosci* 12(1):1421–1429
- Fovell R, Fovell MY (1993) Climate zones of the conterminous United States defined using cluster analysis. *J Clim* 6(11):2103–2135
- Gibbs RJ (1970) Mechanisms controlling world water chemistry. *Science* 170(3962):1088–1090
- Gui HR, Chen LW (2007) Hydrogeochemistic evolution and discrimination of groundwater in mining district. Geological Publishing House, Beijing
- Gubran M, Ghrefat H, Zaidi F, Shehata M (2019) Integration of hydrochemical, GIS, and remote-sensing data for assessment of shallow groundwater aquifers in Wadi Nisah, Central Saudi Arabia. *Environ Earth Sci* 78(5):161
- Jain S, Sharma SK, Mandal TK, Saxena M (2018) Source apportionment of PM10 in Delhi, India using PCA/APCS, UNMIX and PMF. *Particuology* 37:107–118
- Lang YH, Yang W (2014) Source apportionment of PAHs using Unmix model for Yantai coastal surface sediments, China. *Bull Environ Contam Toxicol* 92(1):30–35
- Liu Q, Sun YJ, Xu ZM, Xu G (2018) Application of the comprehensive identification model in analyzing the source of water inrush. *Arab J Geosci* 11(9):189
- Liu LL, Wang ZP, Feng J, Zhang T (2015) Co-occurrence correlations of heavy metals in sediments revealed using network analysis. *Chemosphere* 119:1305–1313
- Liu XH, Sun LH (2015) Chemical variations of groundwater in coal bearing aquifer in northern Anhui Province, China: study based on Unmix model and major ions. *Electron J Geotech Eng* 20(15):6549–6557
- Liu Z, Yang H, Cheng W, Ni GH (2017) Stress distribution characteristic analysis and control of coal and gas outburst disaster in a pressure-relief boundary area in protective layer mining. *Arab J Geosci* 10(16):358
- Nordstrom DK, Ball JW, Donahoe RJ, Whittemore D (1989) Groundwater chemistry and water-rock interactions at Stripa. *Geochim Cosmochim Acta* 53(8):1727–1740
- Ramaroson V, Randriantsoery JR, Rajaobelison J, Fareze LP, Rakotomalala CU, Razafitsalama FA, Rasolofonirina M (2020) Nitrate contamination of groundwater in Ambohidrapeto–Antananarivo–Madagascar using hydrochemistry and multivariate analysis. *Appl Water Sci* 10(7):1–13
- Shi SS, Xie XK, Bu L, Zhou ZQ (2018) Hazard-based evaluation model of water inrush disaster sources in karst tunnels and its engineering application. *Environ Earth Sci* 77(4):141
- Si LL, Li ZH, Yang YL (2018) Coal permeability evolution with the interaction between nanopore and fracture: Its application in coal mine gas drainage for Qingdong coal mine in Huaibei coalfield, China. *J Nat Gas Sci Eng* 56:523–535
- Sun LH (2013) Hydrochemical variation during groundwater mixing: a case study with multivariate statistical approach. *Water Pract Technol* 8(3–4):399–408
- Sun LH, Gui HR (2015) Hydro-chemical evolution of groundwater and mixing between aquifers: a statistical approach based on major ions. *Appl Water Sci* 5(1):97–104
- Sunkari ED, Abu M (2019) Hydrochemistry with special reference to fluoride contamination in groundwater of the Bongo District, Upper East Region, Ghana. *Sustain Water Resour Manag* 5(4):1803–1814
- Sunkari ED, Abu M, Zango MS, Wani AML (2020) Hydrogeochemical characterization and assessment of groundwater quality in the Kwahu-Bombouaka Group of the Voltaian Supergroup. *Ghana J Afr Earth Sci* 169:103899
- Team, Core R (2013) R: A language and environment for statistical computing. Vienna, Austria. <http://www.R-project.org>
- Tay CK, Hayford E, Hodgson IO, Kortatsi BK (2015) Hydrochemical appraisal of groundwater evolution within the Lower Pra Basin, Ghana: a hierarchical cluster analysis (HCA) approach. *Environ Earth Sci* 73(7):3579–3591
- Tóth J (1999) Groundwater as a geologic agent: an overview of the causes, processes, and manifestations. *Hydrogeol J* 7(1):1–14
- Wang D, Shi LQ (2019) Source identification of mine water inrush: a discussion on the application of hydrochemical method. *Arab J Geosci* 12(2):58
- Wang Y, Shi LQ, Wang M, Liu TH (2020) Hydrochemical analysis and discrimination of mine water source of the Jiaojia gold mine area, China. *Environ Earth Sci* 79(6):1–14
- Wu Q, Mu WP, Xing Y, Qian C, Shen JJ, Wang Y, Zhao DK (2019) Source discrimination of mine water inrush using multiple methods: a case study from the Beiyangzhuang Mine, Northern China. *Bull Eng Geol Environ* 78(1):469–482
- Wu Q, Zhao SQ, Sun WJ, Cui FP, Wu C (2013) Classification of the hydrogeological type of coal mine and analysis of its characteristics in China. *J China Coal Soc* 38(6):901–905
- Xu H, Bu W (2015) Analysis of water-inrush for deep coal floor in a coal mine. *Electron J Geotech Eng* 20(10):4189–4196
- Yidana SM, Bawoyobie P, Sakyi P, Fynn OF (2018) Evolutionary analysis of groundwater flow: application of multivariate statistical analysis to hydrochemical data in the Densu Basin, Ghana. *J Afr Earth Sci* 138:167–176
- Yu H, Gui HR, Zhao HH, Wang MC, Li J, Fang HX (2020) Hydrochemical characteristics and water quality evaluation of shallow groundwater in Suxian mining area, Huaibei coalfield, China. *Int J Coal Sci Technol* 12:1–11
- Zango MS, Sunkari ED, Abu M, Lermi A (2019) Hydrogeochemical controls and human health risk assessment of groundwater fluoride and boron in the semi-arid North East region of Ghana. *J Geochem Explor* 207:106363
- Zhang J, Li RF, Zhang XY, Bai Y, Cao P, Hua P (2019) Vehicular contribution of PAHs in size dependent road dust: a source apportionment by PCA-MLR, PMF, and Unmix receptor models. *Sci Total Environ* 649:1314–1322

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.