ORIGINAL PAPER

Source identification of mine water inrush: a discussion on the application of hydrochemical method

Dandan Wang[1] · Longqing Shi[2]

Abstract

This paper presents an investigation on source identification of mine water inrush using a geochemical technique. The hydrogeochemistry characteristics can reflect water–rock interaction processes; therefore, the results of hydrochemical analysis could indicate the groundwater occurrence. Although hydrochemical analysis has been reviewed in previous studies, the selection of the evaluation index and the choice of units have seldom been studied. Statistical methods, hierarchical cluster analysis (HCA) and principal component analysis (PCA), were used for analysis by SPSS 21.0. Piper, Durov, and Stiff diagrams were used to identify the four types of water sources. Four types of water samples were used to perform this research, and the major purpose of the present research is to examine the results obtained under different conditions. The results show that the situations arising from the selection of different identification indices, units, and identification methods can lead to great differences. The results are as follows: The selection of trace ions for identification indices can largely affect the discriminant results. In this study, identification results with fewer indicators are poor than results with more indicators as a whole. The unit milliequivalents per liter (mEq/L) is not useful for better identification results according to this study. The data is appropriate for PCA (the Kaiser–Meyer–Olkin measure of sampling adequacy is > 0.5, and the significance value for Bartlett's test is < 0.01), but its application to reduce dimensions cannot work under all conditions.

Keywords Source identification · Hydrochemical analysis · Hydrogeochemistry characteristics · Principal component analysis (PCA) · Mine water

Introduction

Environmental protection has gained considerable importance in the background of the negative impact of harmful environmental changes on humans. Anthropogenic activities have led to the worldwide deterioration of not only air quality but also water quality (Li et al. 2014; Purushotham et al. 2011). The carbon emission is caused by coal burning including indoor and outdoor, which accelerates the greenhouse effect. It leads to climate change on the environment as well as on forest area, coastal area, and urban area (Cetin 2016; Cetin et al. 2017; Sevik et al. 2017). In recent years, considering the severe pollution arising from the use of coal, its use has been reduced all over China, in particular in several cities. However, mining cannot be completely stopped, and the hazards encountered during mining, such as mine collapse, water inrush, discharge, and contamination, should be dealt with carefully. Among these hazards, inrush, discharge, contamination, and solution methods are closely linked to water sources.

Various methods are available for solving problems related to water sources based on element characteristics. Hydrochemical analysis is usually the first step in dealing with water problems, such as the source identification of water inrush and contamination and the assessment of water quality. Examples of these methods are fuzzy evaluation, multivariate statistical methods, distance discriminant, Bayes discriminant, and isotope tracking (Aravena et al. 1993; Dinka et al. 2015; Wang et al. 2016; Wu et al. 2017; Zhang et al. 2009). Hierarchical cluster analysis (HCA) is a powerful tool for

✉ Longqing Shi
  Cattony2002@163.com

  Dandan Wang
  wangdandan@cumt.edu.cn

[1] School of Resources and Geosciences, China University of Mining and Technology, Xuzhou 221116, China

[2] Shandong Provincial Key Laboratory of Depositional Mineralization and Sedimentary Minerals, Shandong University of Science and Technology, Qingdao 266590, China

grouping samples into significant clusters. Hence, this method was used to test water samples and check whether the samples could be grouped into specific groups. The main function of PCA is to reduce a large number of variables into a few underlying factors to explain the variability of group characteristics (Yang et al. 2016a, b).

Geochemistry characteristics of water depend on water–rock interaction processes, which are controlled by groundwater dynamics and lithology, etc. (Wolkersdorfer 2008). Therefore, chemical analyses are commonly utilized for water-related research such as water quality assessment, water-contaminant source identification, and mineral water studies. Among such analyses, hydrochemical analysis is the fundamental method because hydrochemical characteristics can reflect the occurrence conditions. The need for a greater understanding of the chemical composition of water sources, particularly for elements that are not monitored on a regular basis, is the key issue for defining their quality. Hydrochemistry is widely used for water source identification and water quality assessment, for example the use of stable isotopes and radioisotopes (Ji et al. 2017; Tallini et al. 2014), studies on the hydrogeochemical processes in groundwater (Armengol et al. 2017; Yang et al. 2016a), studies on the spatiotemporal variety of groundwater (Singh et al. 2015), and identification of pollution sources of groundwater or surface water (Datta et al. 2011; Telci and Aral 2011; Yang et al. 2016b). Piper trilinear diagrams and ion ratios are commonly used for these studies.

Mine water inrush may originate from a variety of sources such as sand water, limestone water, surface water, and stratifugic water. The coal-bearing strata belong almost completely to the Permo-Carboniferous system in North China, and with the increase in deep mining, the main controlling factors of water hazards related to limestone water become more important (Meng et al. 2012; Tripathy and Ala 2018; Wu et al. 2017). Hydrogeochemical characteristics vary with rock–water reactions such as leaching and mixing; therefore, these characteristics are directly related to the lithology (Ghesquière et al. 2015; Frape et al. 1984). If the hydrochemical method is used for water source identification, different lithologies can be identified, but water originating from similar lithologies (such as limestone water, particularly) cannot be as easily identified. Hence, the choice of the method for water source identification is crucial to mine safety.

The selection of parameters in hydrochemical analyses plays an important role in solving water-related problems. PCA can be combined with other methods to solve research problems such as water source identification (Zhang et al. 2012; Lu et al. 2012). Moreover, the unit or parameter selected differs from various studies (Howladar 2017; Huang et al. 2017; Wen et al. 2014; Zhou et al. 2010). Although different methods have been compared, no comparison of the results obtained using different units is available. Unlike the unit milliequivalents per liter (mEq/L), the unit milligrams per liter (mg/L) indicates absolute concentration. Therefore, theoretically, using milliequivalents per liter is more appropriate when solving certain problems as it can also indicate the characteristics of other elements present.

This study focuses on the research results obtained using a hydrochemical analysis method, either by itself or in combination with PCA, and the selection of parameters and units. The major objective is to study the identification accuracy of similar limestone water or similar sandstone water. Descriptive statistics, contrastive analysis, and statistical methods are used in this study to determine the influence of different methods, parameters, and units on identification results. PCA was used to reduce the number of chemical parameters, and HCA was used to test the sample grouping. The hydrochemical analysis result shows that the Piper and Stiff diagrams are useful for presenting sample characteristics, and the hydrogeochemical types are SO_4-Na·Ca, SO_4-Na, SO_4·HCO_3-Na, and HCO_3-Na. Further, other conclusions were drawn to explain the results under different conditions.

Methods and materials

Study area

The Xinwen coalfield is located in Tai'an, Shandong Province, China (Fig. 1). It lies on the axis of the Xinwen syncline; therefore, the faults are well developed here. Statistically, there are 23 medium–large faults in this area and numerous small faults. The strata in the study area are terrestrial and oceanic mutual coal-bearing deposits of North China Permo-Carboniferous type. The average thickness of the confined limestone aquifers is less than 10 m, but the Ordovician limestone is about 800 m. The limestone serves as the roof aquifer or the floor aquifer and can easily act as direct water inrush sources. Some water flow records of the Ordovician limestone over the past 4 years are presented in Table 1.

Water sampling and hydrology analysis

Limestone water was sampled from the Ordovician limestone and Xujiazhuang limestone aquifers, and sandstone water was sampled from the Lower Jurassic sandstone and the roof sandstone of the lower seam no. 3 in the Jining coalmine. The physical parameters and major ions were determined for statistical evaluation.

The water samples were collected and scattered the research aquifers based on the drilling distribution in mine area. Some physical and chemical indices such as water temperature, pH, and conductivity can be measured on the spot by sensors. And then, the filtered water samples were sent to the testing center for sample analysis. The anions were tested

Fig. 1 Geographic location of the Zhaizhen coalmine

by an ion chromatograph, and the cations were tested by an inductively coupled plasma emission spectrometer.

The Piper trilinear diagram (Piper 1944) is one of the most useful graphical representations in groundwater quality studies. This diagram helps one to understand the geochemical characteristics of groundwater. The Durov diagram improves upon the Piper trilinear diagram and incorporates total dissolved solids (TDS) and pH. The geochemical characteristics of groundwater can be clearly seen by plotting the cation and anion concentrations in the Piper trilinear diagram.

The Stiff diagram is also a graphical representation of chemical analyses (Stiff 1951). It is used to display the major

Table 1   Water flow records over 2010–2013

Sources of water inflow	Position			
	7th panel Sand water of roof and floor (m^3/min)	Central water sump Gravel or sand water (m^3/min)	No. 1 Ordovician limestone hole Ordovician limestone water (m^3/min)	No. 3 Ordovician limestone hole Ordovician limestone water (m^3/min)
January 2010	0.20	0.59	0.95	3.14
June 2010	0.19	0.64	0.83	2.85
December 2010	0.19	0.65	0.89	2.94
June 2011	0.18	0.65	52.80	176.20
December 2011	0.19	0.64	50.69	170.99
June 2012	0.18	0.63	51.72	177.36
December 2012	0.18	0.61	51.30	173.28
January 2013	10.32	36.97	52.07	174.36

ion composition of a water sample. A polygonal shape is created from four parallel horizontal axes extending on either side of a vertical zero axis. Cations and anions are plotted on each side of the zero axes. Stiff diagrams are useful in making a rapid visual comparison of different water sources. They can help in determining flow paths or showing changes in the ionic composition of a water body over space or time.

Statistical analysis

The multivariate method was used to deal with sample data and compare four types of water. To compare the water groups and different methods, PCA and HCA were used. With HCA, the samples can be grouped into significant clusters effectively based on the data of chemical parameters. Therefore, HCA was used to test the water samples in this study. PCA is useful for dimensionality reduction and was used to determine the various sources between parameters. SPSS 21.0 was used to analyze water samples and perform HCA and PCA calculations. The details of PCA are given below.

Step 1: Normalize the original dataset by computing $\mathbf{x} - \overline{\mathbf{x}}$, where $\mathbf{x}$ is a $d \times 1$-dimensional vector representing one sample data and $\overline{\mathbf{x}}$ is the $d \times 1$-dimensional mean vector of the whole dataset.

Step 2: Compute the covariance matrix of the normalized whole dataset.

Step 3: Compute the eigenvectors ($\nu_1, \nu_2, \ldots, \nu_d$) and the corresponding eigenvalues ($\lambda_1, \lambda_2, \ldots, \lambda_d$) for the covariance matrix such that $\sum \mathbf{v} = \lambda \mathbf{v}$.

Step 4: Sort the eigenvectors by decreasing the eigenvalues and choose the top k eigenvectors to obtain a $d \times k$-dimensional matrix $\mathbf{V}$ consisting of the chosen eigenvectors.

Step 5: Project the original dataset into the new subspace by computing $\mathbf{y} = \mathbf{V}'\mathbf{x}$, where $\mathbf{y}$ is the transformed $k \times 1$-dimensional sample in the new subspace.

Results

Hydrochemical characteristics

The hydrochemical characteristics of Ordovician limestone water (I), Xujiazhuang limestone water (II), Lower Jurassic sand water (III), and sand water in the roof of the lower

Table 2 Hydrochemical characteristics of limestone water

Item	Ordovician limestone water (I)					Xujiazhuang limestone water (II)				
	Mean	Max	Min	SD	CV	Mean	Max	Min	SD	CV
Na	285.26	655.30	212.03	101.33	0.36	265.81	291.27	233.37	16.29	0.06
	12.41	28.5	9.22	4.41	0.35	11.6	12.9	10.2	0.77	0.06
K	31.70	72.81	23.56	11.26	0.36	29.53	32.36	25.93	1.81	0.06
	0.81	1.86	0.60	0.28	0.35	0.76	0.83	0.66	0.05	0.06
Ca	326.67	1847.93	164.13	409.57	1.25	203.18	236.93	165.99	21.94	0.11
	16.3	92.2	8.19	20.43	1.25	10.14	11.8	8.28	1.10	0.11
Mg	53.14	79.8	4.78	19.44	0.37	56.73	84.61	26.90	13.84	0.24
	4.37	6.57	0.39	1.60	0.36	4.67	6.69	2.21	1.14	0.24
Fe	0.45	1.00	0	0.35	0.75	0.44	0.9	0	0.35	0.79
	0.016	0.036	0	0.01	0.75	0.02	0.03	0	0.01	0.79
NH_4	0.025	0.09	0	0.02	0.98	0.02	0.08	0	0.02	1.07
	0.001	0.005	0	0	0.98	0.001	0.004	0	0.001	1.15
Cl	91.55	136	71.43	19.36	0.21	94.33	140.00	76.75	1931	0.20
	2.58	3.84	2.01	0055	0.21	2.66	3.95	2.16	0.55	0.21
SO_4	1173.65	2291.79	988.57	327.01	0.28	1106.55	1211.75	1008.32	73.39	0.07
	24.42	47.7	20.6	6.80	0.28	23.03	25.2	21	1.53	0.07
HCO_3	124.60	224.92	81.28	32.09	0.26	116.17	138.88	94.83	14.37	0.12
	2.04	3.69	1.33	0.53	0.26	1.90	2.28	1.55	0.24	0.12
pH	7.59	8.40	6.00	0.68	0.89	7.34	8.30	5.35	0.868	0.12
TDS	2024.75	3663.94	1613.08	622.45	0.31	1814.68	1965.83	1667.23	101.96	0.06

The units for ion concentration are mg/L (upper) and mEq/L (lower), and for TDS, mg/L

SD standard deviation, *CV* coefficient of variation

Table 3　Hydrochemical characteristics of sandstone water

Item	Sand water of the Lower Jurassic (III)					Sand water in the roof of the lower seam no. 3 (IV)				
	Mean	Max	Min	SD	CV	Mean	Max	Min	SD	CV
Na	298.93	504.15	219.38	82.26	0.28	532.73	618.22	465.3	62.19	0.12
	13.00	21.9	9.54	3.58	0.28	23.18	26.90	20.20	2.71	0.12
K	33.21	56.02	24.38	9.14	0.28	59.192	68.692	51.70	6.91	0.12
	0.85	1.43	0.62	0.23	0.28	1.52	1.76	1.32	0.18	0.12
Ca	71.16	209.78	7.56	60.99	0.90	3.91	4.80	2.13	0.83	0.21
	3.37	10.50	0.38	3.04	0.91	0.20	0.24	0.11	0.04	0.21
Mg	15.92	47.02	5.16	12.49	0.78	1.86	3.64	0.95	0.93	0.50
	1.31	3.87	1.43	1.03	0.78	0.17	0.30	0.08	0.08	0.45
Fe	0.20	0.32	–	0.07	0.38	0.55	1.10	–	0.55	1.00
	0.007	0.09	–	0.02	2.61	0.01	0.04	–	0.02	1.41
NH_4	0.22	0.4	0.04	0.10	0.45	0.298	1.00	0.02	0.34	1.13
	0.01	0.06	0	0.01	1.04	0.02	0.06	0	0.03	1.41
Cl	39.24	63.17	18.75	8.85	0.23	60.19	71.91	41.71	10.97	0.18
	1.11	1.78	0.53	0.25	0.23	1.70	2.03	1.18	0.31	0.18
SO_4	569.30	1626.9	273.51	440.44	0.77	12.84	21.81	2.50	7.08	0.55
	11.85	33.9	5.69	9.16	0.77	8.92	52.00	0.17	19.26	2.15
HCO_3	351.51	587.18	138.16	141.61	0.40	1299.54	1617.30	1077.72	230.03	0.18
	5.76	9.62	2.26	2.32	0.40	21.32	26.5	17.70	3.75	0.18
NO_3	0.69	1.42	0.07	0.42	0.61	0.88	3.52	–	1.30	1.47
	0.01	0.06	0	0.02	1.47	0.01	0.06	–	0.02	1.47
NO_2	0.03	0.06	–	0.02	0.49	0.10	0.40	–	0.17	1.73
	0	0.03	–	0.01	1.74	0	0.03	–	0.01	1.74
pH	7.90	8.30	7.40	0.23	0.03	8.30	8.50	8.20	0.10	0.01
TDS	1532.19	3089.85	1021.13	635.17	0.41	1972.79	2273.42	1707.98	218.518	0.11

The units for ion concentration are mg/L (upper) and mEq/L (lower), and for TDS, mg/L

SD standard deviation, *CV* coefficient of variation

seam no. 3 (IV) as obtained using the descriptive statistical method are listed in Tables 2 and 3. Graphical representations of the chemical analyses of the four types of water are shown in Fig. 2, which contains the Piper, Durov, and Stiff diagrams.

The listed data and analyzed results show that the major cations of Ordovician limestone water and sand water are dominated by Na and Ca. The major anion of the two types of limestone water is dominated by SO_4. The major anions of Lower Jurassic sand water are SO_4 and HCO_3, and the leading anion of sand water in the roof of the lower seam no. 3 is HCO_3. The major cations contribute 28.53%, 25.85%, 19.81%, and 26.99% of the total dissolved solids of water types I–IV, respectively. The major anions contribute 59.18%, 60.95%, 59.53%, and 65.58% of the total dissolved solids of water types I–IV, respectively. The standard deviation and coefficient of variation reflect the dispersion degree of the data.

There is a slight difference between the two types of limestone water. Limestone is a carbonate deposition, and the

water–rock interactions are closely related to the lithology, such as dolomitic limestone or simply limestone. Comparison of the two types of sand water shows that the two differ even though they originate from sand aquifers. The differences arise depending on time, channels, temperature, etc., and these differences can affect the rock–water interactions. The percentage of sodium and the sodium adsorption ratio can be used to assess water quality.

To identify the four types of water clearly, Table 4 presents the sample analysis results: water type, conductivity, salinity hazard, sodium adsorption ratio (SAR), and exchangeable sodium ratio (ESR). Table 4 also lists the features of the four types of water: the two limestone water samples differ in water type and conductivity, and the two sandstone water samples have clearly different water types, SAR, and ESR. Thus, the two sandstone water types are easily distinguished, whereas discriminating between the two limestone water types is comparatively difficult. Limestone water and mixed water identification still remains a challenge for mine water inrush source identification.

Fig. 2 Graphical representation
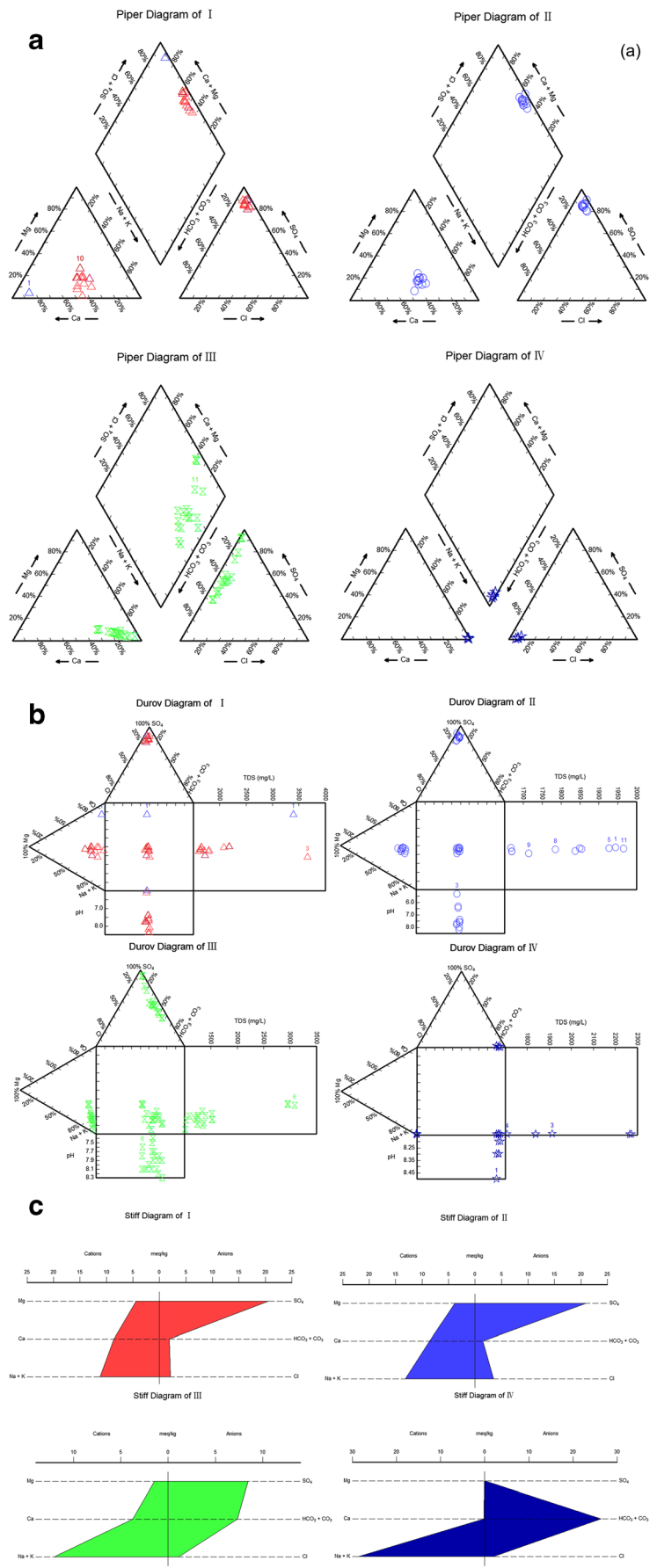of chemical analyses. **a** Piper, **b**
Durov, and **c** Stiff diagrams

Table 4 Data analysis of the four types of water

Item	I	II	III	IV
Water type	SO_4-Na·Ca	SO_4-Na	SO_4·HCO_3-Na	HCO_3-Na
Conductivity (μmhos/cm)	2689.17	2432.45	1630.97	1724.38
Salinity hazard	Very high	Very high	High	Very high
Sodium adsorption ratio (SAR)	4.31	4.26	9.71	56.5
Exchangeable sodium ratio (ESR)	0.77	0.78	4.08	69.53

The data in Tables 3 and 4 cannot provide any visual information about the water samples. The hydrochemical analysis results presented in Fig. 2 clearly show the differences between the water types. The water quality of different sandstone aquifers is different. It is also more easily distinguished from the water quality of limestone aquifers. Limestone water from different limestone aquifers can be distinguished on the basis of the Durov and Stiff diagrams.

Hierarchical cluster analysis

HCA is used to group all the water samples into several significant different clusters. As it is known that there are four types of water, HCA can be used to test the water sample data and determine whether the samples can be grouped into hydrochemical groups. Therefore, samples that are improper can be filtered before analysis to reduce the errors caused by improper data.

Figure 3 shows the dendrogram generated using HCA. There are five exceptional water samples (1, 3, 30, 41, and 43): samples 1 and 3 should belong to I, and samples 30, 41, and 43 should belong to III. Furthermore, the following information is derived: group 1 includes all six water samples of type IV, group 2 includes 17 water samples of type III, and the other three samples, namely 30, 41, and 43, are excluded. For the limestone water samples, four samples are displaced and error grouped. When the cluster standards of HCA are known, samples 1, 3, 30, 41, and 43 can be removed from the total water samples for hydrochemical analysis.

Principal component analysis

PCA is a statistical technique useful for finding patterns in data represented in high dimensions, and the main process of PCA is to find the directions that maximize the variance in the dataset (Oh and Hildreth 2016; Verma 2013). As a multivariate data analysis technique that mainly studies the inter-structure of the correlation matrices of parameters, its fundamental purpose is compressing the original data to achieve dimensionality reduction. Therefore, combinations of PCA and other mathematical methods are used to solve a variety of problems.

PCA was used to reduce the number of water chemical parameters and to choose the principal component as a new evaluation index system for hydrochemical analysis. One objective of the study is to check whether the results differ from each other when different parameters are chosen. Therefore, two parameter standards are chosen to conduct PCA. An example of one case with parameters was used to understand the process. Two components were extracted from this process. The eigenvalues were 3.97 and 1.07, and the cumulative variance explained by the two components was 84.03%, and these two components explain 66.18% and 17.85% of the total variance, respectively. The component matrix and the rotated component matrix are presented in Table 5.

The first component was correlated with the ion concentrations of Ca, Mg, and SO_4. The second component was correlated with the ion concentrations of Na + K and HCO_3. The eigenvector of the principal component was constructed based on a component matrix, and then the expression of principal components was derived.

Discussion

Identification with different parameters

In this study, two parameter groups were chosen to study the influence of parameter selection on the analysis results. One group (G_1) contained seven major ions (Na + K, Ca, Mg, Cl, SO_4, and HCO_3), and the other group (G_2) contained parameters of Na + K, Ca, Mg, Cl, SO_4, HCO_3, Fe, NH_4, TDS, and pH. Samples that excluded the four particular cases were used to carry out the contrastive analysis. The comparative results are presented in Table 6.

Table 6 shows that two types of sandstone water were correctly classified by the method using different identification indices, and six and three limestone water samples were misidentified, respectively. According to the analysis results, the return discriminant ratios of G_1 and G_2 are 88% and 94%, respectively. Further, an analysis using parameters different from G_1 and G_2 was conducted, and the result shows that the classification results are the same as those for G_1 when using Na + K, Ca, Mg, Cl, SO_4, HCO_3, pH, and TDS as the identification indices. The analysis results provide further proof that sandstone water can be identified clearly. The results also show that some significant trace ions can affect the classification to a certain extent.

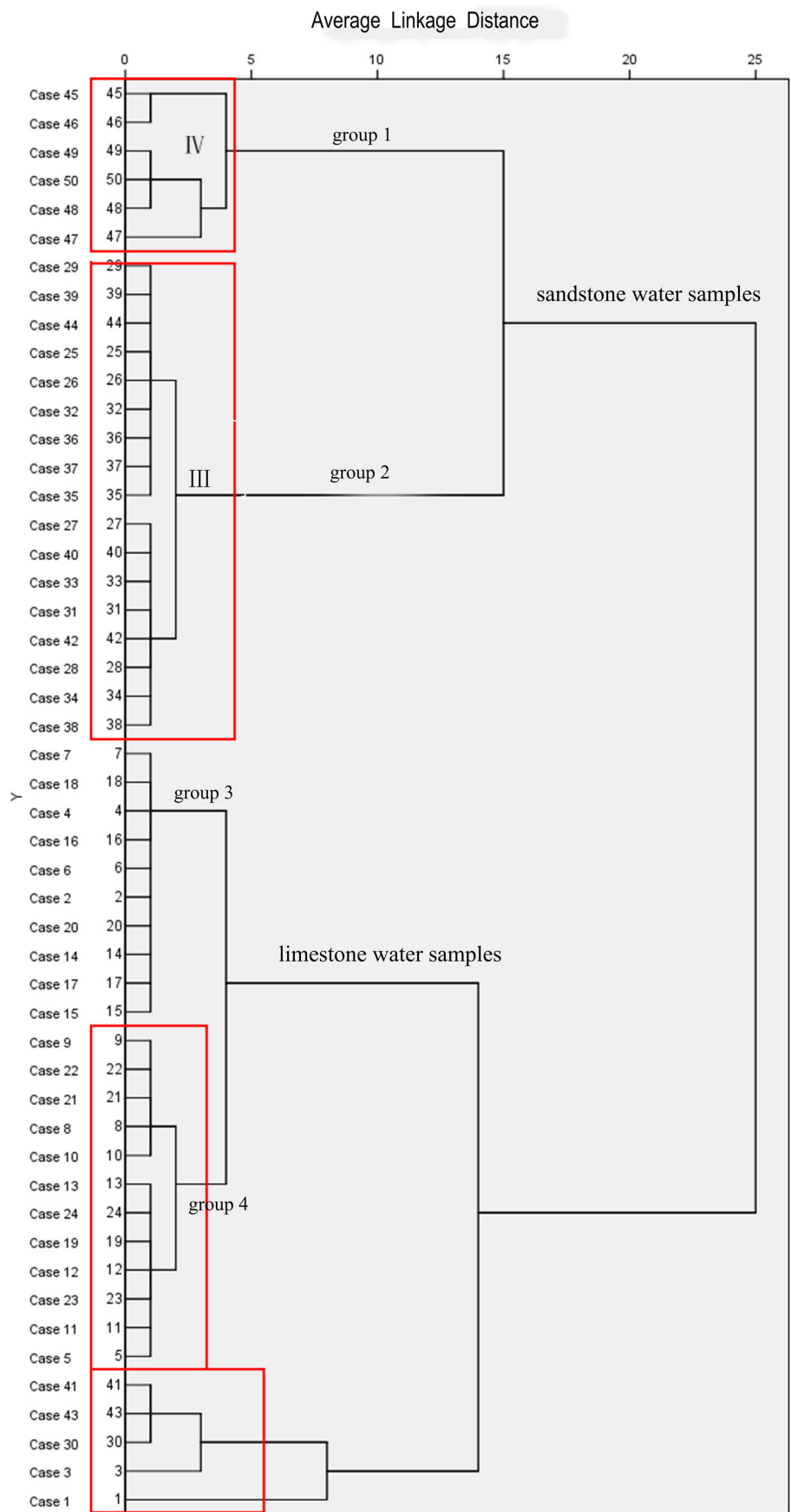**Fig. 3** Dendrogram generated from the HCA of water chemistry data

Table 5 Component matrix and varimax rotated component matrix

Variable	Component matrix		Rotated component matrix	
	1	2	1	2
Na + K	− 0.551	0.775	− 0.067	0.948
Ca	0.934	0.216	0.910	− 0.302
Mg	0.905	0.255	0.906	− 0.253
Cl	0.632	0.360	0.727	− 0.022
SO$_4$	0.912	0.168	0.866	− 0.331
HCO$_3$	− 0.863	0.448	− 0.503	0.832

Identification with different units

Milligrams per liter and milliequivalents per liter are different forms of ion concentration: the former indicates the absolute concentration of ions, and the latter is a relative concentration. Hence, the latter can better reflect the other trace ions. Two units can be used to analyze water characteristics, but no comparative analysis has been studied. One objective of this study is to check whether the units lead to different or similar research results.

According to the analysis results, using parameters in G$_1$ without TDS, the return discriminant ratios of milligrams per liter and milliequivalents per liter, which are two different forms of concentration, are 94% and 84%, respectively. Table 7 lists that two types of sandstone water were all cor-

Table 6 Comparative results with different identification indices

Actual water source	Sample ID	Return discriminant results			
		G$_1$		G$_2$	
		Misidentified sample ID	Prediction group	Misidentified sample ID	Prediction group
I	1–13	2	II*	2	II*
		5	II*		
		9	II*		
		11	II*		
II	14–24	20	I*	20	I*
		21	I*	21	I*
III	25–44	–	–	–	–
IV	45–50	–	–	–	–

En dash indicates that all samples were correctly classified

* indicates the incorrect group

Table 7 Comparative results with different units

Actual water source	Sample ID	Return discriminant results			
		Units of mg/L (G$_3$)		Units of mEq/L (G$_4$)	
		Misidentified sample ID	Prediction group	Misidentified sample ID	Prediction group
I	1–13	2	II*	2	II*
				5	II*
				11	II*
				12	II*
				13	II*
II	14–24	20	I*	18	I*
		21	I*	19	I*
				20	I*
III	25–44	–	–	–	–
IV	45–50	–	–	–	–

En dash indicates that all samples were correctly classified

* indicates the incorrect group

Table 8 Comparative results of the different methods

Actual water source	Sample ID	Return discriminant results			
		Method without PCA		Method combined with PCA	
		Misidentified sample ID	Prediction group	Misidentified sample ID	Prediction group
I	1–13	2	II*	3	II*
				6	II*
				7	II*
				12	II*
II	14–24	20	I*	16	I*
		21	I*	21	I*
III	25–44	–	–	–	–
IV	45–50	–	–	–	–

En dash indicates that all samples were correctly classified

* indicates the incorrect group

rectly classified, whereas three and eight limestone water samples were misidentified. Further, analysis with the unit milliequivalents per liter but without the parameters of Fe and NH_4 was performed. The results obtained were identical to those of G_4. In this case, the use of milliequivalents per liter appears inappropriate.

Identification with different combined methods

PCA can be combined with other mathematical methods for data preprocessing and because of its function of parameter reduction. In this study, PCA was researched to know whether it is suitable for hydrochemical analysis, and further analyzed when a combination of PCA with other methods is suitable. With regard to the return discriminant ratios of G_1, G_2, G_3, and G_4 and the correlation of the identification indices, the data in G_1 were used to present this study item. Ten identification indices (Na + K, Ca, Mg, Cl, SO_4, HCO_3, Fe, NH_4, TDS, and pH) were labeled as × 1, × 2, × 3, × 4, × 5, × 6, × 7, × 8, × 9, and × 10, respectively. Table 8 presents the comparative results of the different methods.

Table 9 presents the results of the Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity, which indicates whether the sample size used for the factor analysis was adequate. The value of KMO is 0.5. Because the significance value (p value) of Bartlett's test is 0, which is < 0.01,

Table 9 KMO and Bartlett's test

KMO measure of sampling adequacy	0.493
Bartlett's test of sphericity	
Approx. chi-square	514.579
df	45
Sig.	0.000

the value is significant, and the correlation matrix is not an identity matrix. Thus, it may be concluded that the factor model is appropriate. According to the operating steps mentioned in the section "Principal component analysis," three principal components are extracted, and the principal component expressions are derived as shown in Eqs. (1)–(3). The analysis results are presented in Table 8. The table shows that the return discriminant ratio of the combined PCA method is 88%, and that of the method without PCA is 94%.

$$X_1 = -0.18x_1 + 0.31x_2 + 0.45x_3 + 0.32x_4$$
$$+ 0.41x_5 - 0.38x_6 + 0.21x_7 - 0.27x_8$$
$$+ 0.21x_9 - 0.29x_{10} \tag{1}$$

$$X_2 = 0.60x_1 + 0.20x_2 + 0.09x_3 + 0.20x_5 + 0.25x_6$$
$$+ 0.01x_7 + 0.37x_8 + 0.60x_9 + 0.04x_{10} \tag{2}$$

$$X_3 = 0.17x_1 - 0.19x_2 + 0.08x_3 + 0.43x_4 - 0.21x_5$$
$$+ 0.37x_6 + 0.68x_7 - 0.25x_8 - 0.08x_9 + 0.17x_{10} \tag{3}$$

Limitation of the work and future study

In this paper, the identification results of water sources under different conditions are given to make a comparison. The major purpose is to study the identification accuracy of similar limestone water or similar sandstone water. Descriptive statistics, contrastive analysis, and statistical methods are used in this study to determine the influence of different methods, parameters, and units on identification results. And, the research shows that these differences can really cause changes in identification results. PCA is the most common method

used to eliminate the linear correlation between the parameters, but it was not always appropriate to analyze the data. Besides, the choice of original data and its representation are also important to the results. Different methods and discrimination index are used to do the study but seldom probe into whether it is appropriate. And, this study may bring a discussion on it.

In this study, three conditions are considered to discuss the discrimination model, but there must be some other reasons that caused the different results. Although lots of methods have been used quite mature, its application condition should make deep research all the time. Water problems account for most of the mine hazards, and its research has always been the focus. Basically, the hydrochemical analysis is necessary for dealing with such issues. Therefore, how to make the analyzing results more accurate should be considered in the future study.

Conclusions

To study the application of hydrochemical analysis in water source identification, especially for similar water, samples of limestone water from the Ordovician limestone and Xujiazhuang limestone aquifers and samples of sandstone water from the Lower Jurassic sandstone and the roof sand water of the lower seam no. 3 were obtained and analyzed. Graphical representations are used to show the hydrochemical characteristics of the four types of water. HCA was used to group several clusters and test the water samples. PCA was used to reduce the number of water chemical parameters and compare the different parameters.

According to the results of hydrology analysis, limestone water can be easily distinguished from sandstone water based on the Piper and Stiff diagrams, and the two types of sandstone water (III and IV) differ from each other more clearly than the two types of limestone water (I and II). Further, the hydrochemical types of the four types of water are SO_4-Na·Ca, SO_4-Na, SO_4·HCO_3-Na, and HCO_3-Na, respectively.

HCA is an effective method for sample grouping and was used to test the water samples. The cluster results show that most of the water samples are correctly grouped. Several cases were excluded in the analysis to reduce errors.

The three groups of the contrastive analysis show that trace ions can affect the discriminant results. Although the unit milliequivalents per liter indicate relative concentration, this form does not have useful effects on the discriminant results. Data used in this paper are appropriate for factor analysis, which is based on KMO and the significance value of Bartlett's test, but it has an inverse result compared to the original data.

References

Aravena R, Evans ML, Cherry JA (1993) Stable isotopes of oxygen and nitrogen in source identification of nitrate from septic systems. Groundwater 31(2):180–186. https://doi.org/10.1111/j.1745-6584.1993.tb01809.x

Armengol S, Manzano M, Bea SA, Martínez S (2017) Identifying and quantifying geochemical and mixing processes in the Matanza-Riachuelo Aquifer System, Argentina. Sci Total Environ 599–600:1417–1432. https://doi.org/10.1016/j.scitotenv.2017.05.046

Cetin M (2016) A change in the amount of CO_2 at the center of the examination halls: case study of Turkey. Studies on Ethno-Medicine 10(2):146–155

Cetin M, Sevik H, Saat A (2017) Indoor air quality: the samples of Safranbolu Bulak Mencilis Cave. Fresenius Environ Bull 26(10):5965–5970

Datta B, Chakrabarty D, Dhar A (2011) Identification of unknown groundwater pollution sources using classical optimization with linked simulation. J Hydro Environ Res 5(1):25–36. https://doi.org/10.1016/j.jher.2010.08.004

Dinka MO, Loiskandl W, Ndambuki JM (2015) Hydrochemical characterization of various surface water and groundwater resources available in Matahara areas, Fantalle Woreda of Oromiya region. J Hydrol: Reg Stud 3:444–456

Frape SK, Fritz P, McNutt RH (1984) Water-rock interaction and chemistry of groundwaters from the Canadian Shield. Geochim Cosmochim Acta 48(8):1617–1627. https://doi.org/10.1016/0016-7037(84)90331-4

Ghesquière O, Walter J, Chesnaux R, Rouleau A (2015) Scenarios of groundwater chemical evolution in a region of the Canadian Shield based on multivariate statistical analysis. J Hydrol 4:246–266. https://doi.org/10.1016/j.ejrh.2015.06.004

Howladar MF (2017) An assessment of surface water chemistry with its possible sources of pollution around the Barapukuria Thermal Power Plant impacted area, Dinajpur, Bangladesh. Groundwater Sustain Dev 5:38–48. https://doi.org/10.1016/j.gsd.2017.03.004

Huang PH, Wang XY, Han SM (2017) Recognition model of groundwater inrush source of coal mine: a case study on Jiaozuo coal mine in China. Arab J Geosci 10(15):323. https://doi.org/10.1007/s12517-017-3099-5

Ji XL, Xie RT, Hao Y, Lu J (2017) Quantitative identification of nitrate pollution sources and uncertainty analysis based on dual isotope approach in an agricultural watershed. Environ Pollut 229:586–594. https://doi.org/10.1016/j.envpol.2017.06.100

Li PY, Wu JH, Qian H, Lyu X, Liu H (2014) Origin and assessment of groundwater pollution and associated health risk: a case study in an industrial park, Northwest China. Environ Geochem Health 36(4):1–20. https://doi.org/10.1007/s10653-013-9590-3

Lu JT, Li XB, Gong FQ et al (2012) Recognizing of mine water inrush sources based on principal components analysis and fisher discrimination analysis method (in Chinese). China Safety Sci J 22(7):109–115

Meng ZP, Li GQ, Xie XT (2012) A geological assessment method of floor water inrush risk and its application. Eng Geol 143–144:51–60. https://doi.org/10.1016/j.enggeo.2012.06.004

Oh S, Hildreth AJ (2016) Pattern-based energy consumption analysis by chaining principle component analysis and logistic regression. In: Oh S, Hildreth AJ (eds) Analytics for smart energy management: tools and applications for sustainable manufacturing. Springer International, Cham, pp 137–177

Piper AM (1944) A graphic procedure in the geochemical interpretation of water-analyses. Eos Transactions American Geophysical Union 25(6):27–39

Purushotham D, Prakash MR, Rao AN (2011) Groundwater depletion and quality deterioration due to environmental impacts in Maheshwaram watershed of R.R. district, AP (India). Environ Earth Sci 62(8):1707–1721. https://doi.org/10.1007/s12665-010-0666-4

Sevik H, Ahmaida EA, Cetin M (2017) Chapter 31: Change of the air quality in the urban open and green spaces: Kastamonu sample. In: Koleva I, Yuksel UD, Benaabidate L (eds) Ecology, planning and design. St. Kliment Ohridski University Press, ISBN: 978-954-07-4270-0, pp 409–422

Singh A, Sharma CS, Jeyaseelan AT, Chowdary VM (2015) Spatio-temporal analysis of groundwater resources in Jalandhar District of Punjab State, India. Sustain Water Resour Manage 1(3):293–304. https://doi.org/10.1007/s40899-015-0022-7

Stiff JR (1951) The interpretation of chemical water analysis by means of patterns. J Petrol Technol 3(1951):15–17. https://doi.org/10.2118/951376-G

Tallini M, Adinolfi Falcone R, Carucci V, Falgiani A, Parisse B, Petitta M (2014) Isotope hydrology and geochemical modeling: new insights into the recharge processes and water-rock interactions of a fissured carbonate aquifer (Gran Sasso, Central Italy). Environ Earth Sci 72(12):4957–4971. https://doi.org/10.1007/s12665-014-3364-9

Telci IT, Aral MM (2011) Contaminant source location identification in river networks using water quality monitoring systems for exposure analysis. Water Qual, Expo Health 2(3):205–218. https://doi.org/10.1007/s12403-011-0039-6

Tripathy DP, Ala CK (2018) Identification of safety hazards in Indian underground coal mines. Journal of Sustainable Mining 17(4):175–183. https://doi.org/10.1016/j.jsm.2018.07.005

Verma JP (2013) Application of factor analysis: to study the factor structure among variables. In: Verma JP (ed) Data analysis in management with SPSS software. Springer Berlin, Germany, pp 359–387

Wang XY, Ji HY, Wang Q, Liu X, Huang D, Yao X, Chen G (2016) Divisions based on groundwater chemical characteristics and discrimination of water inrush sources in the Pingdingshan coalfield. Environ Earth Sci 75:872. https://doi.org/10.1007/s12665-016-5616-3

Wen T, Zhang B, Shao L (2014) Research on prediction of mine water inrush source identification——Xinzhuangzi coalfield as an example. China Saf Sci J 24(2):100–106 (in Chinese)

Wolkersdorfer C (2008) Hydrogeochemistry of mine water. Water management at abandoned flooded underground mines: fundamentals, tracer tests, modelling, water treatment. Springer Berlin Heidelberg, Berlin, pp 9–36

Wu Q, Mu WP, Xing Y, Qian C, Shen J, Wang Y, Zhao D (2017) Source discrimination of mine water inrush using multiple methods: a case study from the Beiyangzhuang Mine, Northern China. Bull Eng Geol Environ 1–14 https://doi.org/10.1007/s10064-017-1194-1

Yang QC, Li ZJ, Ma HY, Wang L, Martín JD (2016a) Identification of the hydrogeochemical processes and assessment of groundwater quality using classic integrated geochemical methods in the southeastern part of Ordos Basin, China. Environ Pollut 218:879–888. https://doi.org/10.1016/j.envpol.2016.08.017

Yang QC, Wang LC, Ma HY, Yu K, Martín JD (2016b) Hydrochemical characterization and pollution sources identification of groundwater in Salawusu aquifer system of Ordos Basin, China. Environ Pollut 216:340–349. https://doi.org/10.1016/j.envpol.2016.05.076

Zhang Y, Guo F, Meng W, Wang XQ (2009) Water quality assessment and source identification of Daliao River basin using multivariate statistical methods. Environ Monit Assess 152:105–121. https://doi.org/10.1007/s10661-008-0300-z

Zhang B, Song XF, Zhang YH, Han D, Tang C, Yu Y, Ma Y (2012) Hydrochemical characteristics and water quality assessment of surface water and groundwater in Songnen Plain, Northeast China. Water Res 46(8):2737–2748. https://doi.org/10.1016/j.watres.2012.02.033

Zhou J, Shi XZ, Wang HY (2010) Water-bursting source determination of mine based on distance discriminant analysis model. J China Coal Soc 35(2):278–282 (in Chinese)