



Source identification of mine water inrush based on principal component analysis and grey situation decision

Qiding Ju¹ · Youbiao Hu¹

Received: 25 May 2020 / Accepted: 23 January 2021 / Published online: 11 February 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

Coal mining is bound to destroy the underground aquifer structure, which will lead to mine water inrush disaster. An accurate and rapid identification of water inrush sources is the crux of preventing the recurrence of water inrush incidents. In this regard, 37 training water samples and 14 verification water samples were extracted from three types of aquifers in the Xieqiao coal mine, China. Na^+ , K^+ , Ca^{2+} , Mg^{2+} , Cl^- , SO_4^{2-} and HCO_3^- were used as the evaluation variables. The principal component analysis was used to eliminate the redundant ion variables in the training samples. The grey situation decision method combined with the entropy weight was used to establish the recognition model. The ion variables of the verification samples were substituted into the model calculations, and the comprehensive accuracy of the model was found to be 85.71%. The proposed method has the advantages of accuracy and speed compared to other contemporary recognition methods. The grey situation decision-making method overcomes the problem that single-factor evaluation cannot identify water inrush, and the entropy weight method can reflect the degree of difference between the variables. Based upon this recognition model, it provides a new method for recognizing water inrush sources, which would also be beneficial to prevention and control of mine water hazards.

Keywords Mine water inrush source · Principal component analysis · Grey situation decision · Linear half-order reduction function · Bayesian discriminant

Introduction

Mine water inrush is one of the extraordinarily severe coal mine accidents in China, and the main reason for these kind of accidents is the absence of in-depth research on the hydrogeological conditions of collieries. According to some statistics from China's State Administration of Coal Mine Safety, there are about 905 mines with complex hydrogeological conditions in China (Zhang et al. 2020; Wang et al. 2020; Sun et al. 2015; Xu et al. 2020). The variation characteristics of the hydrochemical parameters in the underground multi-aquifer system are a direct reflection of mine water inrush (Yin et al. 2019; Qian et al. 2017). Therefore, accurate and rapid identification of water inrush source in the multi-aquifer system is very important in protecting the lives of miners and maintaining safe production of collieries.

Mining would cause changes in the level of groundwater, temperature and hydrochemical components, which can be analyzed using various techniques, including water temperature and water level method, hydrochemical analysis and mathematical analysis. Based upon the geothermal gradient theory, aquifer temperature shows differences at certain depths. Sui et al. (2010) compared the water temperature at the water inrush point with that of the aquifer with hidden water inrush potential, which can preliminarily predict the source of mine water inrush. Lin et al. (2014) conducted a dewatering test on the aquifer, and discovered the recharge channel through the change of water level, and identified the potential source of mine water inrush. Wang and Shi (2019) used Piper, Durov, and Stiff diagrams to identify the four types of water sources. Li et al. (2016) obtained the conventional ion concentration through field sampling tests, and determined the seawater infiltration channel and water inrush source by combining these test results with multivariate statistical analysis to study the hydrochemical effect of the aquifer. Stable isotopes δD and $\delta^{18}\text{O}$ play an important role in analyzing the origin and formation of

✉ Qiding Ju
qidingju@126.com

¹ School of Earth and Environment, Anhui University of Science and Technology, Huainan 232001, China

groundwater in aquifers. They can not only determine the relationship among groundwater, precipitation and surface water, but also analyze the supply source and mixing ratio of groundwater (Chafouq et al. 2018; Yi et al. 2018; Boumaiza et al. 2020; Cao et al. 2020; Liao et al. 2020). Guan et al. (2019) used stable isotopes to verify each other with hydrogeochemical analysis to accurately identify the source of water inrush in Mingdon mine (China), and reported that, due to the high cost, most coal enterprises did not use the method often. In recent years, the mathematical statistical analysis has also developed rapidly, which is widely used in hazard identification related to water inrush activities (Liu et al. 2019; Wang et al. 2020). Based upon cluster analysis, Zhang et al. (2019a, b) established a multiple logistic regression recognition model, and identified and verified the water inrush aquifer in Qinan coal mine, China. Huang et al. (2019) established the Piper-PCA-Fisher water source recognition model, which is more accurate than the Piper diagram method or Fisher discriminant method. Based upon the principal component analysis (PCA) and BP neural network, Yang et al. (2019) proposed the water source discrimination model of mine water field monitoring system, which was applied to Lijiazui mine in Huainan (China), and exhibited the accuracy of 91%. Dong et al. (2019) combined the Fisher feature extraction and support vector machine (SVM) methods, and applied this new model to the Wuhai mining area (China). The results showed that this new combined model was more accurate and efficient in discriminating water inrush sources than the traditional SVM model. However, this method requires a large number of water sample data, and could not identify multiple water inrush sources at the same time.

The mathematical analysis method has the characteristics of simple operation, high discrimination efficiency, is objective and provides accurate results at low cost. Therefore, based upon the mathematical theory of grey situation decision-making, this paper proposes a new model for water inrush water source discrimination based on principal component analysis and entropy weight-grey situation decision method. In the model, the principal component analysis eliminates redundant variables, reduces the workload, and assigns entropy weight to each variable to characterize the degree of difference of variables. Finally, the gray situation method is used to attribute the multi-factor target evaluation to single-objective decision-making, which solves the problem that single-factor evaluation cannot reflect the water quality characteristics of water inrush sources. Since the production of Xieqiao mine in Huainan coalfield (China), there have been 24 water inrush accidents due to the influence of coal mining, which have a great impact on the safety of miners and the mine economy. In the water inrush potential areas of water trickling in the coal wall, water chemical variable information samples were extracted, and put into the

model for discrimination. This way, the predictions regarding the source of water pouring samples were made. Timely investigation and treatment for the water filling channel and water filling strength of the aquifer can effectively prevent the occurrence of water inrush accidents. This study is only aimed at quantitative analysis of hydrochemical information, and therefore, it can also be applied to other similar mine water source discrimination, thus indicating a wide range of applicability.

Hydrogeological conditions in the study area

Xieqiao coal mine (China) is located in the northeast of Yingshang county, Fuyang city, Anhui province, China and is geographically positioned as shown in Fig. 1. The terrain in the mining area is flat and belongs to the Huaihe alluvial plain. The study area belongs to a transitional climate, with obvious seasonality of hot summers and cold winters. The annual average temperature is 15.1 °C, while the annual average rainfall is 926.3 mm. Most of the rainfall falls in June, July and August, accounting for about 40% of that falls the whole year. The annual average evaporation is 1610.14 mm. The evaporation is larger than the rainfall, whereas the humidity coefficient is nearly 0.5.

The groundwater regime in the mining areas of Xieqiao consists of four subsystems, namely the loose aquifer of the Cenozoic, the coal-bearing sandstone fissure aquifer of the Permian, the limestone-karst fissure aquifer of the Carboniferous and the limestone-karst fracture aquifer of the Ordovician. The hydrogeological characteristics of the aquifer and aquifuge are shown in Fig. 1. Between the coal-bearing sandstone fissure aquifer and the loose layer pore aquifer, there is a thick clay layer covering the coal measures. Besides, the average distance between the limestone aquifers of the Taiyuan formation and the coal floor is 16.44 m. Therefore, under normal conditions, there is no direct water filling effect among the three aquifers.

Sampling and testing

The sample bottle was washed 2–3 times using water before sampling. The sample bottle should not be filled with water sample, and around 5–10 ml space was left at the top of the bottle (Huang et al. 2019; Guo et al. 2019). The water samples were maintained at a low temperature to prevent any chemical reactions (Zhang et al. 2019a, b; Zhang et al. 2016). Eight variables ($K^+ + Na^+$, Ca^{2+} , Mg^{2+} , Cl^- , SO_4^{2-} , HCO_3^- , pH and TDS) were tested. The pH value of the sample was tested using a Hanna portable pH meter within the 5 min of the collection of samples. The samples for cation analysis were acidified with nitric acid to $pH \leq 2$. The tests were conducted within 24 h after sampling at the Quality Inspection Center, Anhui University of Science and

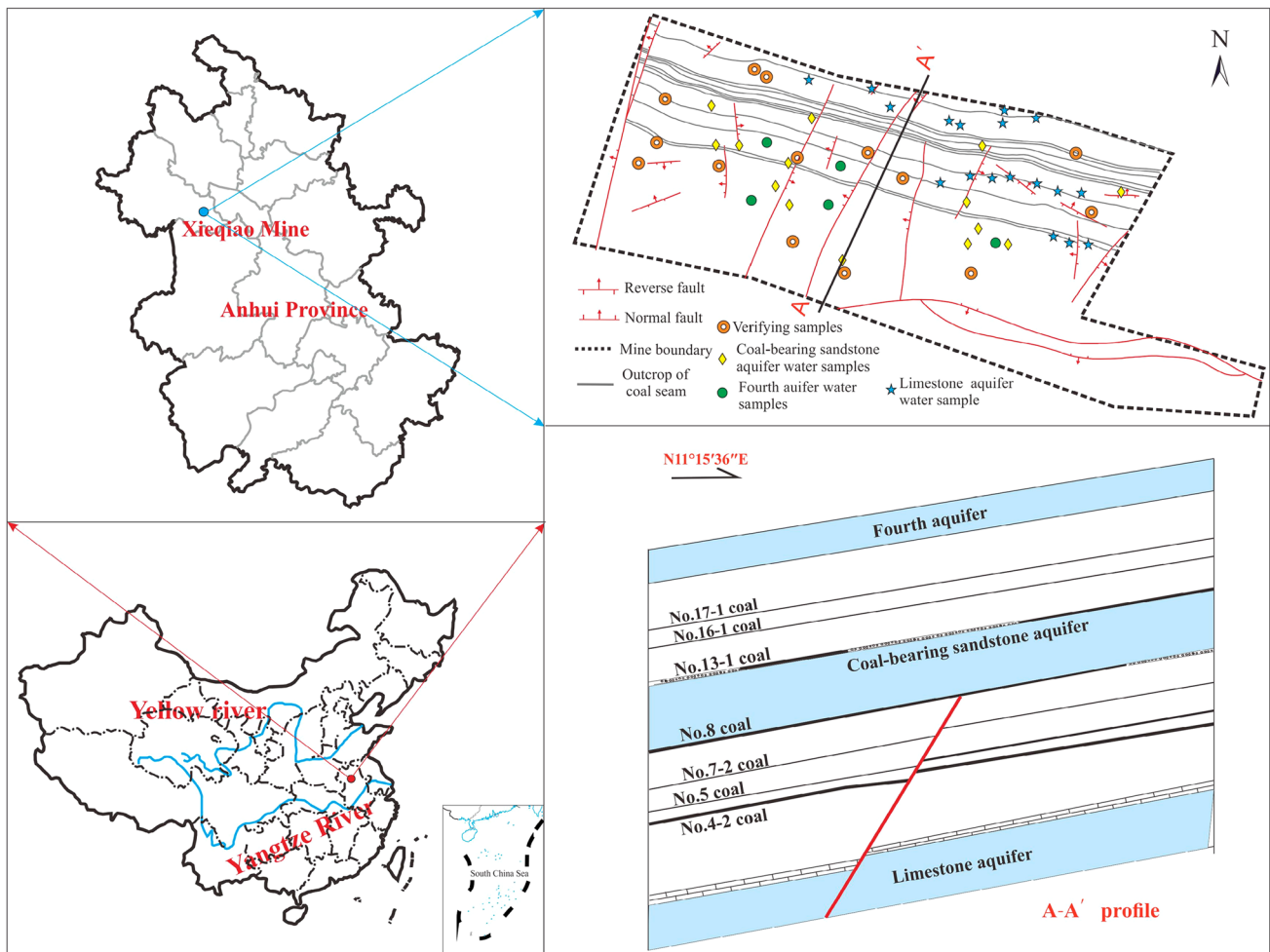


Fig. 1 Location of the study area, sampling sites and the aquifer profile

Technology, China. The Cl^- , SO_4^{2-} , and HCO_3^- tests were conducted using ion chromatography, whereas $K^+ + Na^+$, Ca^{2+} , and Mg^{2+} tests were conducted using inductively-coupled plasma mass spectrometry. In order to review the reliability of test results, the anion and cation balance was calculated to confirm that of the standard error lied within $\pm 5\%$. As shown by the results presented in Table 1 and Fig. 1, a total of 37 training water samples were collected between 2005 and 2018 in Xieqiao mine, China. The samples included five samples from the Cenozoic aquifer, 14 samples from the Permian aquifer, and 18 samples from the carboniferous aquifers. In order to show the accuracy of the model, 14 verifying samples were also collected. The Q , P , C , X_1 , X_2 , X_3 , X_4 , X_5 , and X_6 were used to represent the Cenozoic aquifer, the Permian aquifer, the Carboniferous aquifer, Na^+ , K^+ , Ca^{2+} , Mg^{2+} , Cl^- , SO_4^{2-} and HCO_3^- , respectively.

Methods

Principal component analysis

The p vectors X_1, X_2, \dots, X_p of the original data matrix X was used as a linear combination $Y = AX$. The relationship between the original and new variables is given by Eq. (1).

$$\begin{cases} Y_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ \dots \\ Y_p = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{cases}, \quad (1)$$

where $a_{i1} + a_{i2} + a_{i3} + \dots + a_{ip} = 1$; Y_i and Y_j are not related, Y_i is the maximum variance of all the linear combinations of (X_1, X_2, \dots, X_p) , and Y_2 is the combination with the largest variance among all the linear combinations of X_1, X_2, \dots, X_p that are not related to Y_1 . Moreover, the sum of the variances

Table 1 Water samples from the Xieqiao coal mines, China

N	X ₁ (mg/L)	X ₂ (mg/L)	X ₃ (mg/L)	X ₄ (mg/L)	X ₅ (mg/L)	X ₆ (mg/L)	Percent error(%)	N	X ₁ (mg/L)	X ₂ (mg/L)	X ₃ (mg/L)	X ₄ (mg/L)	X ₅ (mg/L)	X ₆ (mg/L)	Percent error(%)
<i>Q</i> (The Cenozoic aquifer)															
1	849.69	52.1	21.28	1088.74	264.18	310.46	0.07	26	795.72	3.9	4.74	759.05	432.27	156.36	3.28
2	782.12	57.97	25.12	1045.25	233.32	282.52	0.08	27	669.32	5.86	4.74	630.86	326.6	212.65	2.99
3	413.9	27.89	14.52	418.32	211.33	218.91	2.04	28	948.84	121.02	1.19	935.61	499.51	43.78	4.87
4	121.85	0.6	0.49	79.41	15.44	14.03	-0.52	29	323.39	23.42	23.71	184.74	403.45	268.97	-2.30
5	208.02	33.86	12.1	75.15	220.94	293.96	0.89	30	196.81	61.57	2.41	93.76	245.21	37.53	1.72
<i>P</i> (The Permian aquifer)															
6	749.47	15.03	11.43	712.9	9.67	738.95	2.84	31	873.51	26.03	17.02	1029.51	304.24	279.34	0.97
7	719.59	15.23	11.31	706.87	12.14	781.06	0.03	32	878.81	40.04	15.81	1052.06	336.98	261.51	0.71
8	235.93	41.71	20.51	221.44	168.12	225.16	2.26	33	211.64	30.03	3.65	165.32	275.48	12.51	1.88
9	226.83	21.47	11.86	125.67	249.76	168.87	1.76	34	266.26	38.04	19.46	167.83	428.56	81.31	0.37
10	599.3	115.19	101.33	199.71	884.15	1407.27	-2.73	35	161.51	5.61	3.4	24.82	65.32	214.79	2.33
11	136.21	55.61	19.3	59.85	214.26	140.73	3.15	36	353.02	9.62	11.67	191.45	119.11	429.58	1.91
12	911.12	10.01	7.3	1062.08	238.6	331.49	0.49	37	291.01	12.83	3.89	78	230.54	253.84	4.91
<i>V</i> (Verification samples)															
13	897.75	14.01	7.3	1064.58	226.7	306.47	0.76	p1	785.43	47.33	20.64	1027.54	228.38	260.86	0.34
14	898.1	18.02	13.38	1047.05	106.48	451.69	2.43	p2	199.07	27.05	14.59	80.12	56.39	441.15	2.56
15	1417.05	7.21	2.92	538.89	376.56	1874.53	-0.22	p3	143.61	37.85	16.94	67.63	115.27	281.45	3.40
16	1515.8	5.61	5.83	608.02	407.29	1967.29	3.71	p4	921.9	5.01	1.22	750.48	3.29	1062.97	2.28
17	1194.06	4.81	1.94	723.24	76.85	1430.31	2.73	p5	1010.64	6.21	1.22	597.33	3.29	1462.04	4.10
18	411.02	8.82	3.4	134.72	119.11	576.03	-4.91	p6	620.5	8.82	1.82	652.99	6.38	459.48	2.82
19	1009.49	7.62	4.38	823.86	1.85	1304	0.02	p7	727.17	5.21	1.7	670	117.72	579.08	1.91
<i>C</i> (The Carboniferous aquifer)															
20	900.75	47.9	20.31	1084.77	346.98	328.29	0.10	p8	439.34	15.23	4.26	106.7	349.65	596.17	0.38
21	901.9	47.9	19.82	1087.25	345.95	328.29	0.05	p9	993.86	9.42	3.28	319.05	18.32	1792.77	4.78
22	900.75	49.9	19.94	1087.25	346.36	330.73	0.06	p10	780.43	25.38	20.16	1010.39	211.33	738.04	0.54
23	976.87	303.86	2.41	1146.34	489.35	21.89	-2.84	p11	265.17	4.09	1.24	117.95	276.56	167.01	0.06
24	706.33	49.65	19.3	950.62	105.38	331.49	0.57	p12	794.7	2.07	5.04	526.03	624.39	419.05	-4.86
								p13	729.79	46.85	23.04	1133.55	336.21	287.71	-4.89
								p14	689.42	31.23	2.37	828.91	365.02	37.53	0.26

of Y_1, Y_2, \dots, Y_p is equal to the sum of the variances of X_1, X_2, \dots, X_p .

The general steps for solving the principal components are as follows.

The original variable data was standardized, and the covariance matrix Σ among the variables was calculated. The eigenvectors of the covariance matrix were $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, whereas the corresponding unit eigenvectors were T_1, T_2, \dots, T_p . The transformation matrix was given by: $A = T'$, where j is the i -th row of A and the unit feature vector T_i corresponded to the i -th largest root of Σ . In addition, the variance of the i -th principal component Y_i was equal to the i -th large characteristic root λ_i of Σ . Then, the variance contribution rate of Y_k was calculated for the k -th principal component and given by: $\eta_k = \frac{\lambda_k}{\sum_{k=1}^p \lambda_k}$. If m ($m < p$) principal components were selected, the cumulative contribution rate of principal components Y_1, Y_2, \dots, Y_m was $\xi_m = \sum_{k=1}^m \lambda_k / \sum_{k=1}^p \lambda_k$. The main component index that made the cumulative contribution rate of variance reach 75% or more was selected (Price et al. 2006; He et al. 2016; Cloutier et al. 2008).

Entropy weight

According to the information theory, information is a measure of the degree of order of the system, whereas entropy is the measure of the degree of disorder of a system. Entropy value can represent the difference in the concentration of each ion in different water samples, whereas the weight value of the corresponding ion can be obtained using the entropy weight method (Chen et al. 2019; Fausto et al. 2019; Liang et al. 2019). The calculation steps are as follows.

There were m evaluation water samples and n evaluation ions, They formed the initial matrix $R = (r_{ij})_{m \times n}$, in which r_{ij} represents the evaluation value of the j -th ion in the i -th water sample ($i = 1, 2, 3, \dots, m; j = 1, 2, 3, \dots, n$).

The proportion of p_{ij} was calculated using Eq. (2).

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^m r_{ij}} \tag{2}$$

The entropy of the j -th ion was calculated using Eq. (3).

$$e_j = -k \sum_{i=1}^m p_{ij} \ln p_{ij}, k = \frac{1}{\ln m} \tag{3}$$

The entropy weight of the j -th ion was calculated using Eq. (4).

$$\omega_j = \frac{(1 - e_j)}{\sum_{j=1}^n (1 - e_j)} \tag{4}$$

Entropy weight and grey situation decision method

The binary combination of events and countermeasures constitutes the situation. Taking an event as the core, other similar events were gathered around the core event, forming a gray event to study the countermeasures. This is the gray situation decision-making thought (Zu et al. 2018; Zhang et al. 2014). In the identification of mine water inrush, the identification index was regarded as the gray element, while the identification object was taken as an event. Different water source categories were used as the countermeasures (Li et al. 2019; Fu 2016; He and Gong 2013). The optimal situation was determined through decision analysis. The water source category corresponding to the optimal situation was the evaluation result. In general, there was no preference between different countermeasures, though different goals have different effects on optimization, and different decision makers' preferences for different goals will also lead to inconsistent goal weights. In traditional gray situation decision-making, the equal treatment of targets could not reflect the decision makers' preferences and the actual situation of the decision-making problems. Therefore, the entropy theory was applied to different indices to give weights. It improved the resolution between situations and made the decision results more accurate and reasonable. The main steps of the mathematical model are as follows.

Event a_i ($i = 1, 2, \dots, n$) and countermeasure b_j ($j = 1, 2, \dots, m$) were determined. Situation was constructed $s = (a, b)$ and the situation array was established. The p ($p = 1, 2, \dots, q$) index was given. Different situation effect measurement matrices were constructed according to different targets p , and the elements in the matrix were obtained relying on the membership function of each index. The situation effect measurement matrices were defined as given by Eq. (5).

$$D^{(p)} = (r_{ij}^{(p)} / S_{ij}^{(p)})_{m \times n} \tag{5}$$

According to the single-indicator decision effect measurement value $r_{ij}^{(p)}$ and the entropy weight of each indicator, the comprehensive effect measurement value $r_{ij}^{(\Sigma)}$ of multiple indicators was obtained. The $r_{ij}^{(\Sigma)}$ value was defined and given by Eq. (6).

$$r_{ij}^{(\Sigma)} = \sum_p \omega_p r_{ij}^{(p)} \tag{6}$$

Therefore, the comprehensive decision matrix was defined as given by Eq. (7).

$$D(\Sigma) = \begin{bmatrix} r_{11}^{(\Sigma)} & r_{12}^{(\Sigma)} & \dots & r_{1m}^{(\Sigma)} \\ r_{21}^{(\Sigma)} & r_{22}^{(\Sigma)} & \dots & r_{2m}^{(\Sigma)} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1}^{(\Sigma)} & r_{n2}^{(\Sigma)} & \dots & r_{nm}^{(\Sigma)} \end{bmatrix}. \tag{7}$$

The best situation was chosen and the decisions were made based upon the best results. If b_{j^*} was the best strategy of the comprehensive decision matrix column, the comprehensive effect measurement value $r_{ij}^{(\Sigma)}$ was defined as given by Eq. (8).

$$r_{ij}^{* (\Sigma)} = \max_j \left\{ r_{ij}^{(\Sigma)} \right\}. \tag{8}$$

If a_{j^*} was the best strategy of the comprehensive decision matrix row, the comprehensive effect measurement value $r_{ij}^{(\Sigma)}$ was defined as given by Eq. (9).

$$r_{ij}^{* (\Sigma)} = \max_i \left\{ r_{ij}^{(\Sigma)} \right\}. \tag{9}$$

Bayesian discriminant method

If n samples were taken from G matrices, each sample must belong to one of the G matrices (A_g). If p variables (x_1, x_2, \dots, x_p) were measured for each sample, then each sample can be regarded as a point in the p -dimensional space $\{R\}$. Additionally, n samples constituted a p -dimensional sample space $\{R\}$. An unknown sample $X(x_1, x_2, \dots, x_p)$ was also regarded as a point in the p -dimensional space. If it fell in the subspace with the highest probability, it could be classified as one of the G matrixes (Yan et al. 2019; Fang et al. 2020; Du et al. 2020). The Bayesian discriminant model is as follows.

There are g matrices A_g ($g = 1, 2, \dots, G$). The probability density function is given by Eq. (10).

$$f_g(x) = (2\pi)^{-\frac{p}{2}} \left| \sum -1 \right|^{\frac{1}{2}} \exp \left[-\frac{1}{2} (x - a_g)' \sum -1 (x - a_g) \right], \tag{10}$$

where $x = (x_1, x_2, \dots, x_p)'$. The parameters a_g and \sum represent the mean and covariance matrix of A_g , respectively. The prior probability q_g and parameter of A_g were known, and there was no significant difference between the matrix covariance matrix. The discriminant function is given by Eq. (11).

$$q_g f_g(x) = q_g (2\pi)^{-\frac{p}{2}} \left| \sum -1 \right|^{\frac{1}{2}} \exp \left[-\frac{1}{2} (x - a_g)' \sum -1 (x - a_g) \right], \tag{11}$$

where $g = 1, 2, \dots, G$. The multivariate linear discriminant function under Bayes criterion can be obtained and is given by Eq. (12).

$$y_g(x) = c_{0g} + c_{1g}x_1 + c_{2g}x_2 + \dots + c_{pg}x_p \tag{12}$$

Mine water inrush identification and verification

Extraction of discriminant variables and entropy weights

As shown in Fig. 2, the content of alkaline-earth metal ions (Ca^{2+} and Mg^{2+}) were significantly higher than the alkali metal ions ($\text{Na}^+ + \text{K}^+$). The main chemical type of Cenozoic, Permian and Carboniferous aquifers were $\text{Ca}\cdot\text{Na}\cdot\text{Cl}\cdot\text{HCO}_3$, $\text{Ca}\cdot\text{Mg}\cdot\text{Na}\cdot\text{Cl}\cdot\text{HCO}_3$ and $\text{Ca}\cdot\text{Na}\cdot\text{Cl}\cdot\text{SO}_4\cdot\text{HCO}_3$, respectively.

The correlation coefficient thermograph directly described the degree of correlation among the variables, as shown in Figs. 3, 4, and 5. For the Cenozoic aquifer, the correlation between the variables was large and all the variables were positively correlated. The concentrations of Cl^- and $\text{Na}^+ + \text{K}^+$, Ca^{2+} and Mg^{2+} were significantly correlated, and the corresponding correlation coefficients were 0.99 and 0.88, respectively. For the Permian aquifer, the concentrations of Ca^{2+} and Mg^{2+} were positively correlated, with the correlation coefficient of 0.96. The concentrations of Ca^{2+} , Mg^{2+} and SO_4^{2-} were positively correlated, with the correlation coefficients of 0.75 and 0.81, respectively. The concentrations of $\text{Na}^+ + \text{K}^+$ and HCO_3^- were also positively correlated, with the correlation coefficient of 0.79. For the Carboniferous aquifer, the concentrations of Cl^- and $\text{Na}^+ + \text{K}^+$ were significantly correlated, with the correlation coefficient of 0.98. The correlation coefficient between these variables was large, which will cause information overlap and affect the accuracy of the water inrush discrimination model. In order to solve these problems, the principal component analysis was used to extract the main variables as the discriminant factors of the water inrush discriminant model.

In general, the number of principal components depended on its cumulative variance ratio. When the cumulative variance ratio was greater than 80%, the number of principal components at this time can fully reflect the water chemical information of the sample. According to Kaiser criterion and the scree plot method (Fig. 6), the cumulative variance rate was found to be 85.91%. Therefore, the number of principal components was 3. As shown by the results presented in Table 2, the Principal component 1 reflected the information of 36.23% of the training samples and represented Na^+ and K^+ . Principal component 2 reflected the information of 29.15% of the training samples and represented SO_4^{2-} . Principal component 3 reflected the information of 20.53% of the training samples and represented HCO_3^- . According to the principal component score coefficients (Table 3), the relationships between the principal components P_1 , P_2 and P_3 and the original variables X_1 , X_2 , X_3 , X_4 , X_5 and X_6 were obtained, which are given by Eqs. (13)–(15).

Fig. 2 Piper diagram of all water samples from different aquifers

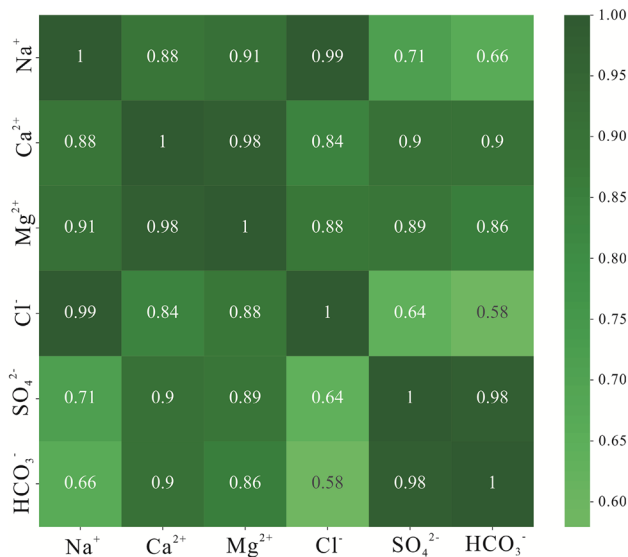
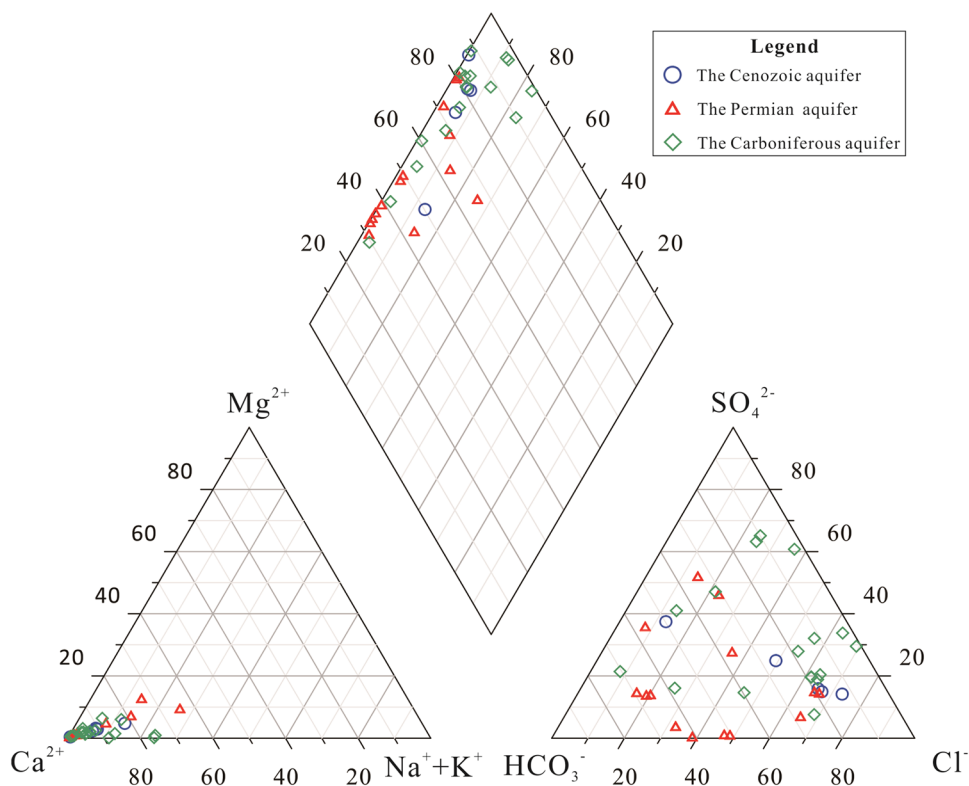


Fig. 3 Heat map of the correlation coefficient for the Cenozoic aquifer

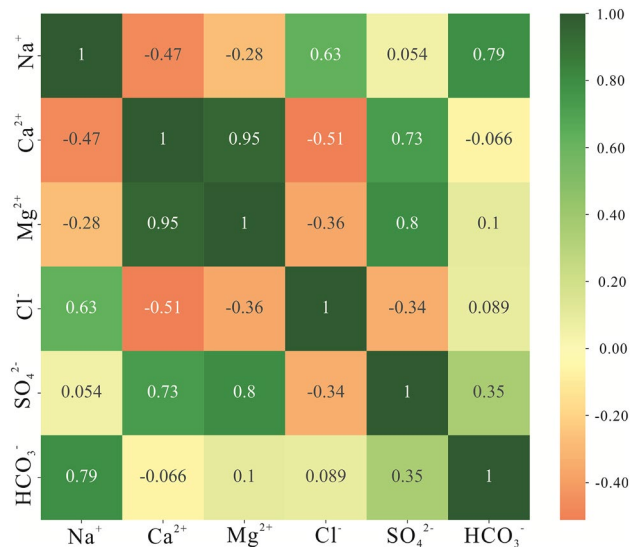


Fig. 4 Heat map of the correlation coefficient for the Permian aquifer

$$P_1 = 0.567X_1 + 0.312X_2 + 0.268X_3 + 0.471X_4 + 0.416X_5 + 0.339X_6, \tag{13}$$

$$P_2 = -0.408X_1 + 0.418X_2 + 0.490X_3 - 0.283X_4 + 0.495X_5 - 0.306X_6, \tag{14}$$

$$P_3 = -0.018X_1 - 0.462X_2 + 0.415X_3 - 0.451X_4 + 0.099X_5 + 0.633X_6. \tag{15}$$

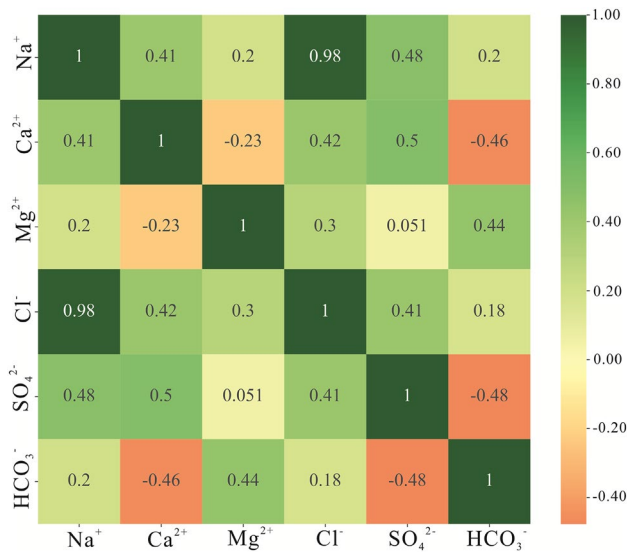


Fig. 5 Heat map of the correlation coefficient for the Carboniferous aquifer

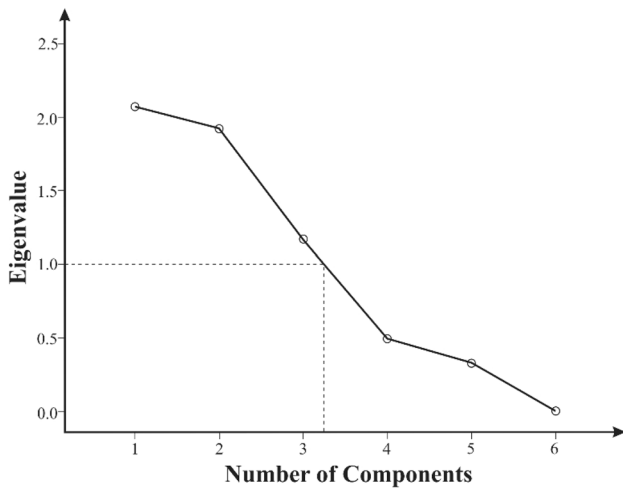


Fig. 6 Scree plot of the principal components

Table 2 Variance interpretation rate and orthogonal rotation factor loading matrix

Parameters	Principal components		
	1	2	3
X ₁	0.84	-0.54	-0.02
X ₂	0.46	0.55	-0.51
X ₃	0.40	0.65	0.46
X ₄	0.69	-0.37	-0.50
X ₅	0.61	0.65	0.11
X ₆	0.50	-0.41	0.70
Eigenvalue	2.17	1.75	1.23
% of variance explained	36.23	29.15	20.53
% of cumulative variance	36.23	65.38	85.91

The bold values reflect the contribution value of important factors

Table 3 The principal component score coefficients

Parameters	PC1	PC2	PC3
X ₁	0.57	-0.41	-0.02
X ₂	0.31	0.42	-0.46
X ₃	0.27	0.49	0.42
X ₄	0.47	-0.28	-0.45
X ₅	0.42	0.50	0.10
X ₆	0.34	-0.31	0.63

Construction of water inrush source recognition model

Substituting the ion concentration data presented in Table 1 into Eqs. (13) (14) and (15), the scores of the three principal components were obtained, and the corresponding results are presented in Table 4. According to the characteristics of ionic components of the water samples presented in Table 1, the ionic concentration values of the same aquifer were more discrete, and some data have abnormal values, as shown in Fig. 7. Therefore, Huber’s M-estimated value should be used instead of the average to reflect the concentration trend to obtain the ionic index classification values corresponding to the three water sources. Taking Huber’s M estimator of P_1, P_2 and P_3 as the optimal value, the ranking values of the game set $B = \{b_1, b_2, b_3\}$ are presented in Table 5. According to the entropy weight theory, the weights of P_1, P_2 and P_3 were also calculated and are presented in Table 6.

According to the grading criteria of the game set presented in Table 5, the membership function was calculated using the linear half-order function method, and the linear half-order function was the half-step function. Finally, the membership function graphs of P_1, P_2 , and P_3 were obtained, as shown in Fig. 8.

According to Fig. 8, the membership function of each variable is as follows.

- (1) Membership of variables P_1 .

The membership function of P_1 belonging to the Cenozoic aquifer was defined as given by Eq. (16).

$$f_{P_1}^Q = \begin{cases} 1 & x < 693.3621 \\ \frac{933.0847-x}{239.7226} & 693.3621 \leq x \leq 933.0847 \\ 0 & x > 933.0847 \end{cases} \quad (16)$$

The membership function of P_1 belonging to the Permian aquifer was defined as given by Eq. (17).

Table 4 The principal component score

<i>N</i>	<i>P</i> ₁	<i>P</i> ₂	<i>P</i> ₃	<i>n</i>	<i>P</i> ₁	<i>P</i> ₂	<i>P</i> ₃
I (The Cenozoic aquifer)				26.00	1043.60	−369.00	−214.66
1	1231.18	−586.24	−298.66	27.00	887.33	−349.93	−130.30
2	1152.98	−548.77	−299.69	28.00	1239.00	−366.32	−417.26
3	606.17	−230.66	−43.40	29.00	542.81	−45.38	120.12
4	117.95	−68.29	−27.68	30.00	290.26	30.03	−25.27
5	358.53	−66.64	159.73	31.00	1213.64	−562.87	−277.88
II (The Permian aquifer)				32.00	1238.89	−544.47	−303.16
6	1022.47	−716.63	131.70	33.00	326.95	13.81	−55.54
7	1018.01	−714.39	161.74	34.00	452.79	56.61	3.92
8	402.67	−117.03	44.36	35.00	205.80	−102.30	127.15
9	358.67	−41.33	65.88	36.00	491.37	−260.90	191.47
10	1341.07	−196.33	866.58	37.00	388.55	−97.08	138.76
11	264.68	23.20	63.18	IV (Verification samples)			
12	1233.04	−647.32	−263.35	<i>V</i> ₁	379.01	−53.43	73.75
13	1214.47	−639.13	−283.11	<i>V</i> ₂	335.75	−192.51	238.74
14	1208.48	−733.60	−194.49	<i>V</i> ₃	272.85	−82.67	146.08
15	1851.33	−1113.28	953.35	<i>V</i> ₄	1239.10	−909.12	316.54
16	1984.32	−1185.52	984.12	<i>V</i> ₅	1352.80	−1023.71	636.08
17	1535.66	−1088.21	564.10	<i>V</i> ₆	820.64	−570.48	−17.36
18	544.66	−317.74	305.64	<i>V</i> ₇	974.75	−601.88	61.40
19	1405.98	−1037.44	434.47	<i>V</i> ₈	652.44	−210.37	350.73
III (The Carboniferous aquifer)				<i>V</i> ₉	1332.07	−1029.71	972.07
20	1297.16	−572.66	−276.78	<i>V</i> ₁₀	1269.20	−704.65	15.34
21	1298.41	−574.58	−278.23	<i>V</i> ₁₁	1132.60	−547.58	−302.92
22	1299.42	−573.76	−277.50	<i>V</i> ₁₂	1101.67	−288.73	76.70
23	1399.89	−358.51	−611.57	<i>V</i> ₁₃	1028.10	−419.43	−270.66
24	1024.68	−575.76	−235.89	<i>V</i> ₁₄	955.94	−332.01	−339.70
25	1015.28	−550.47	70.74				

$$f_{P_1}^P = \begin{cases} 0 & x < 933.0847 \\ \frac{x-933.0847}{255.5420} & 933.0847 \leq x \leq 1158.6267 \\ 1 & x > 1158.6267 \end{cases} \quad (17)$$

The membership function of *P*₁ belonging to the Carboniferous aquifer was defined as given by Eq. (18).

$$f_{P_1}^C = \begin{cases} 0 & x < 693.3621 \\ \frac{x-693.3621}{\frac{239.7226}{1158.6267-x}} & 693.3621 \leq x \leq 933.0847 \\ \frac{1158.6267-x}{225.5420} & 933.0847 \leq x \leq 1158.6267 \\ 0 & x > 1158.6267 \end{cases} \quad (18)$$

(2) Membership of variables *P*₂.

The membership function of *P*₂ belonging to the Cenozoic aquifer was defined as given by Eq. (19).

$$f_{P_2}^Q = \begin{cases} 0 & x < -337.9810 \\ \frac{x+337.9810}{70.1005} & -337.9810 \leq x \leq -267.8805 \\ 1 & x > -267.8805 \end{cases} \quad (19)$$

The membership function of *P*₂ belonging to the Permian aquifer was defined as given by Eq. (20).

$$f_{P_2}^P = \begin{cases} 1 & x < -619.1755 \\ \frac{x+337.9180}{-281.1945} & -619.1755 \leq x \leq -337.9810 \\ 0 & x > -337.9810 \end{cases} \quad (20)$$

The membership function of *P*₂ belonging to the Carboniferous aquifer was defined as given by Eq. (21).

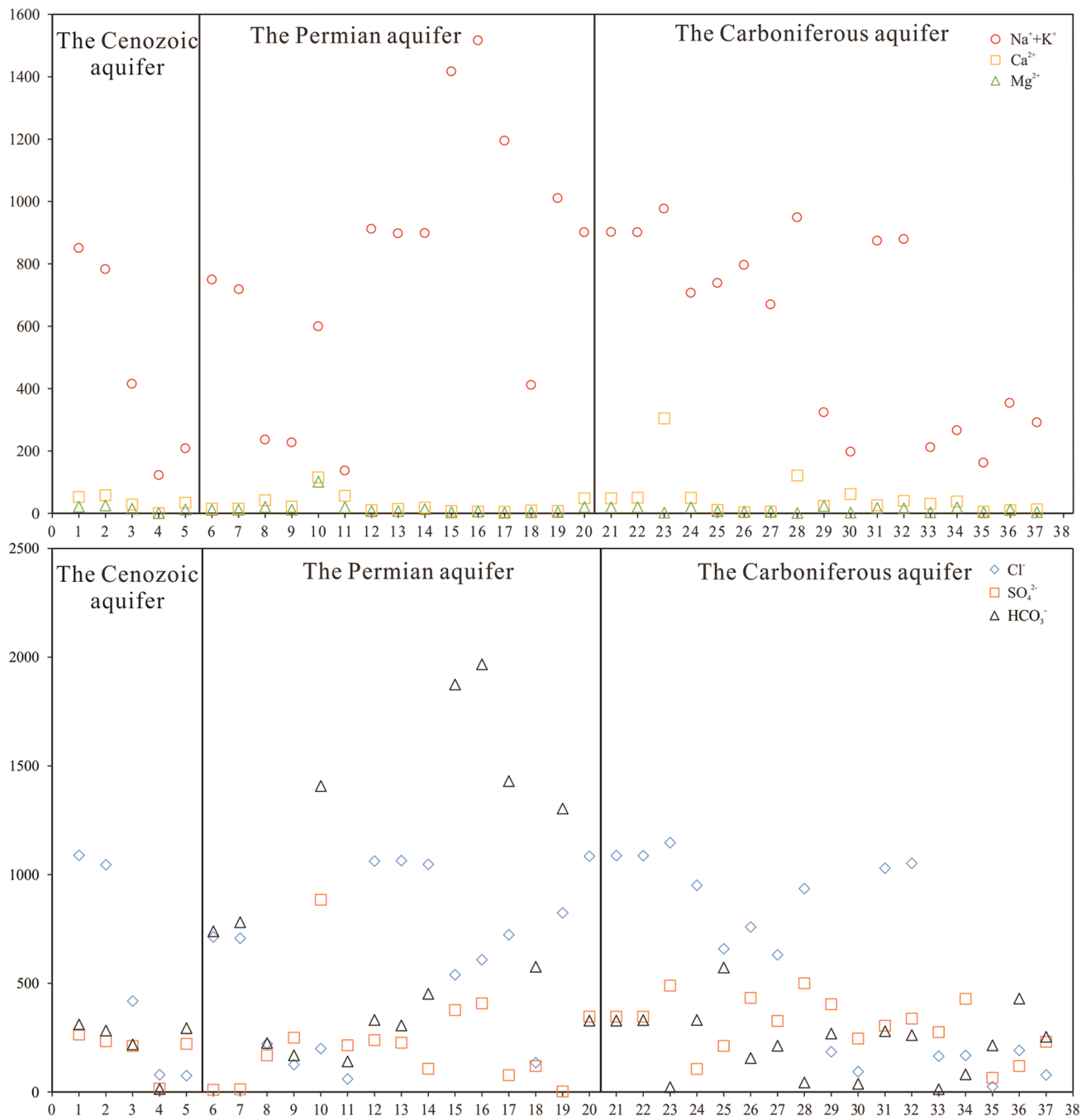


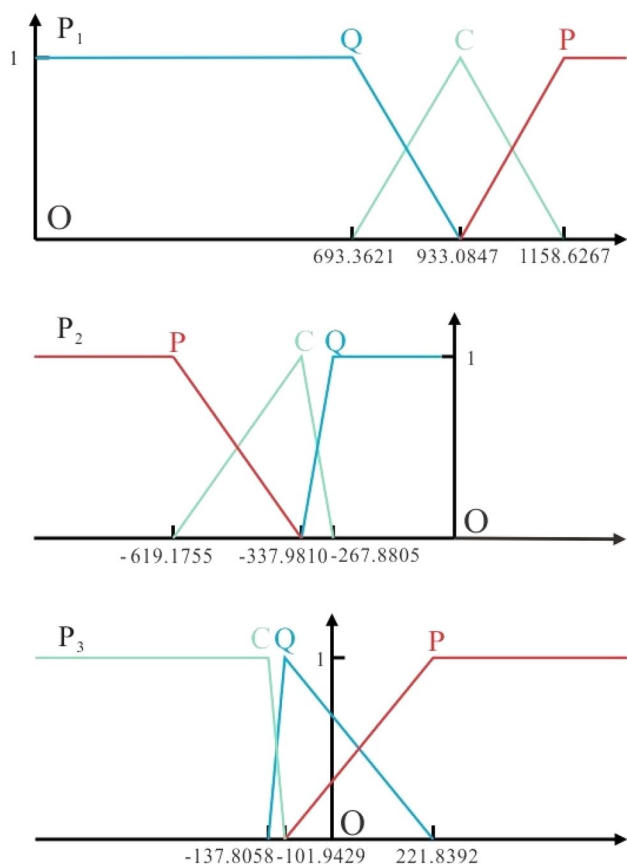
Fig. 7 Scatter plots of anion and cation water samples

Table 5 Rating standard table of countermeasures set

Aquifers	P_1	P_2	P_3
Cenozoic	693.3621	-267.8805	-101.9429
Permian	1158.6267	-619.1755	221.8392
Carboniferous	933.0847	-337.9810	-137.8058

Table 6 Entropy weight table of principal components

Principal components	P_1	P_2	P_3
Weights	0.179	0.385	0.436



$$f_{P_2}^C = \begin{cases} 0 & x < -619.1755 \\ \frac{x+619.1755}{\frac{281.1945}{x+267.8805} - 70.1005} & -619.1755 \leq x \leq -337.9810 \\ 0 & x > -337.9810 \end{cases} \quad (21)$$

(3) Membership of variables P_3 .

The membership function of P_3 belonging to the Cenozoic aquifer was defined as given by Eq. (22).

$$f_{P_3}^Q = \begin{cases} 0 & x < -137.8058 \\ \frac{x+137.8058}{\frac{35.8629}{221.8392-x} - 323.7821} & -137.8058 \leq x \leq -101.9429 \\ 0 & x > 221.8392 \end{cases} \quad (22)$$

The membership function of P_3 belonging to the Permian aquifer was defined as given by Eq. (23).

$$f_{P_3}^P = \begin{cases} 0 & x < -101.9429 \\ \frac{x+101.9429}{323.7821} & -101.9429 \leq x \leq 221.8392 \\ 1 & x > 221.8392 \end{cases} \quad (23)$$

The membership function of P_3 belonging to the Carboniferous aquifer was defined as given by Eq. (24).

$$f_{P_3}^C = \begin{cases} 1 & x < -137.8058 \\ \frac{x+101.9429}{-35.8629} & -137.8058 \leq x \leq -101.9429 \\ 0 & -101.9429 \end{cases} \quad (24)$$

Fig. 8 Membership function of principal components

Table 7 Results of discriminating water samples

n	Membership of P_1			Membership of P_2			Membership of P_3			Comprehensive membership			Discriminant results	
	Q_{P_1}	P_{P_1}	C_{P_1}	Q_{P_2}	P_{P_2}	C_{P_2}	Q_{P_3}	P_{P_3}	C_{P_3}	Q_S	P_S	C_S	Discriminant	Actual
V_1	1	0	0	1	0	0	0.46	0.54	0	0.76	0.24	0	1	1
V_2	1	0	0	1	0	0	0	1	0	0.56	0.44	0	1	1
V_3	1	0	0	1	0	0	0.23	0.77	0	0.67	0.33	0	1	1
V_4	0	1	0	0	1	0	0	1	0	s0	1	0	2	2
V_5	0	1	0	0	1	0	0	1	0	0	1	0	2	2
V_6	0.47	0	0.53	0	0.83	0.17	0.74	0.26	0	0.41	0.43	0.16	2	2
V_7	0	0.18	0.82	0	0.94	0.06	0.50	0.50	0	0.22	0.61	0.17	2	2
V_8	1	0	0	1	0	0	0	1	0	0.56	0.44	0	1	2
V_9	0	1	0	0	1	0	0	1	0	0	1	0	2	2
V_{10}	0	1	0	0	1	0	0.64	0.36	0	0.28	0.72	0	2	2
V_{11}	0	1	0.12	0	0.75	0.25	0	1	0	0	0.90	0.12	2	2
V_{12}	0	1	0.25	0.70	0	0.30	0.45	0.55	0	0.47	0.42	0.16	1	3
V_{13}	0	1	0	1	0	0	0	0	1	0.39	0.18	0.44	3	3
V_{14}	0	0.10	0.90	0.09	0	0.91	0	0	1	0.03	0.02	0.95	3	3

The bold values represent the maximum membership in grey situation decision-making method

Verification of water inrush source recognition model

The values of the principal components P_1 , P_2 and P_3 of the 14 validation samples presented in Table 1 were substituted into the corresponding membership functions (16)–(24) to obtain the degrees of membership of P_1 , P_2 and P_3 . According to the membership function and the maximum membership principle in fuzzy mathematics, the results are presented in Table 7. The recognition accuracy of water samples from the Cenozoic aquifer was 100%. The recognition accuracy of water samples from the Permian aquifer was 87.50%, whereas the recognition accuracy of water samples from the Carboniferous aquifer was 66.67%. The comprehensive accuracy of the model was 85.71%.

Discussion

The grey situation decision-making method is used to obtain the optimal situation in an environment with known and unknown factors (Zu et al. 2018; Zhang et al. 2014). It was applied to the identification of water inrush resources with many multi-factor variables, which are attributed to a single target for decision discrimination. It solves the problem that a single factor cannot identify the source category. However, traditional multi-objective grey situation decision-making methods do not reflect the decision makers' preferences and the actual situation of decision-making when dealing with decision-making goals (Li et al. 2019; Fu 2016; He and Gong 2013). By assigning entropy weight to each variable, which can reflect the degree of difference between the variables, the grey situation decision-making method is further complemented and improved. In order to evaluate the accuracy of this method, it has been compared with the Bayesian discriminant method.

On the basis of extracting the principal components, the classification function coefficients of P_1 , P_2 and P_3 are obtained using the Bayesian discriminant method, as shown by the results presented in Table 8. The linear discriminant functions of three kinds of aquifers are derived and given by Eqs. (25)–(27).

Table 8 Discriminant coefficient of classification function

Principal components	Q	P	C
P_1	0.006	0.006	0.009
P_2	0.005	0.002	0.008
P_3	-0.001	0.003	-0.001
Constants	-2.452	-4.146	-3.613

Table 9 Bayesian discriminant results

Verification samples	y_Q	y_P	y_C	Actual	Bayesian
V_1	1.91	0.65	2.50	1	3
V_2	1.64	1.80	2.37	1	3
V_3	1.37	2.24	1.96	1	2
V_4	0.12	2.42	0.05	2	2
V_5	0.09	3.83	0.26	2	2
V_6	0.36	0.42	0.77	2	3
V_7	0.33	0.68	0.28	2	2
V_8	0.06	0.40	0.23	2	2
V_9	0.58	4.70	0.83	2	2
V_{10}	2.64	2.12	3.92	2	3
V_{11}	0.52	1.76	0.70	2	2
V_{12}	1.62	2.11	2.16	3	3
V_{13}	1.89	0.37	2.56	3	3
V_{14}	1.96	0.09	2.67	3	3

The bold values represent the maximum membership in Bayesian discriminant

$$y_Q = 0.006x_{p_1} + 0.005x_{p_2} - 0.001x_{p_3} - 2.452, \tag{25}$$

$$y_P = 0.006x_{p_1} + 0.002x_{p_2} + 0.003x_{p_3} - 4.146, \tag{26}$$

$$y_C = 0.009x_{p_1} + 0.008x_{p_2} - 0.001x_{p_3} - 3.613. \tag{27}$$

Substituting the data presented in Table 4 into the function Eqs. (16), (17) and (18), the discrimination results were obtained and are presented in Table 9. The recognition accuracy of water samples from the Cenozoic aquifer was 0%. The recognition accuracy of water samples from the Permian aquifer was 75%. The recognition accuracy of water samples from Carboniferous aquifer was 100%, whereas the comprehensive accuracy of the model was 64.29%. Therefore, the new discriminant model proposed in this paper has higher accuracy than the traditional Bayesian discriminant method.

Due to the small number of Cenozoic samples, the accuracy of the Cenozoic samples discrimination was low using the Grey Situation Decision method and the Bayesian discriminant method. Therefore, the model was based on a certain number of water samples. More water samples should be collected to improve the accuracy of the model. In addition, this model should consider the impact of temperature, hydrogeological conditions and human activities on the aquifer to further improve its applicability.

Conclusion

Principal component analysis was used to eliminate the internal correlation between ionic variables. Six ions were combined into three principal components P_1 , P_2 and P_3 , which comprehensively reflected the information of water chemistry. It greatly reduces the number of variables and the calculation of the model.

The recognition accuracy of water samples from the fourth aquifer was 100%. The recognition accuracy of water samples from the coal-bearing sandstone aquifer was 87.50%. The recognition accuracy of water samples from the limestone-karst fissure aquifer in the Carboniferous was 66.67%. The comprehensive accuracy of the model was 85.71%. The entropy weight-grey situation decision model provides a new method for the identification of water inrush, and has important theoretical guiding significance for mine water prevention work.

Acknowledgements We gratefully acknowledge the support provided by the National Key Research and Development Program of China (No. 2017YFC0804101), the Natural Science Foundation of Anhui Province (No. 2008085QD191). The authors also sincerely thank editors and reviewers for reviewing the manuscript.

References

- Boumaiza L, Chesnaux R, Drias T, Walter J, Huneau F, Garel E, Stumpp C (2020) Identifying groundwater degradation sources in a Mediterranean coastal area experiencing significant multi-origin stresses. *Sci Total Environ* 746:141203–141203
- Cao T, Han D, Song X, Trolle D (2020) Subsurface hydrological processes and groundwater residence time in a coastal alluvium aquifer: evidence from environmental tracers combined with hydrochemistry. *Sci Total Environ* 743:140684
- Chafouq D, El Mandour A, Elgettafi M, Himi M, Chouikri I, Casas A (2018) Hydrochemical and isotopic characterization of groundwater in the Ghis-Nekor plain (northern Morocco). *J Afr Earth Sc* 139:1–13
- Chen JB, Wang YJ, Li FY, Liu ZC (2019) Aquatic ecosystem health assessment of a typical sub-basin of the Liao River based on entropy weights and a fuzzy comprehensive evaluation method. *Sci Rep* 9:14045
- Cloutier V, Lefebvre R, Therrien R, Savard MM (2008) Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *J Hydrol* 353(3–4):294–313
- Dong D, Chen Z, Lin G, Li X, Zhang R, Ji Y (2019) Combining the fisher feature extraction and support vector machine methods to identify the water inrush source: a case study of the Wuhai Mining Area. *Mine Water Environ* 38(4):855–862
- Du PF, Huang DH, Ning DH, Xu JJ (2020) Application of Bayesian model and discriminant function analysis to the estimation of sediment source contributions. *Int J Sediment Res* 34(34):0–577
- Fang S, Zhang Z, Wang Z, Pan H, Du T (2020) Principal Slip Zone determination in the Wenchuan earthquake Fault Scientific Drilling project-hole 1: considering the Bayesian discriminant function. *Acta Geophysica* 68(6):1595–1607
- Fausto C, Edmundas KZ, Dalia S, Abbas M (2019) Assessment of concentrated solar power (CSP) technologies based on a modified intuitionistic fuzzy topsis and trigonometric entropy weights. *Technol Forecast Soc Change* 140:0–258
- Fu S (2016) Three-parameter interval grey number multi-attribute decision making method based on information entropy. *Math Comput Appl* 21:17
- Guan Z, Jia Z, Zhao Z, You Q (2019) Identification of inrush water recharge sources using hydrochemistry and stable isotopes: a case study of Mindong No. 1 coal mine in north-east Inner Mongolia, China. *J Earth Syst Sci* 128(7):200
- Guo X, Zuo R, Wang J, Meng L, Teng Y, Shi R, Ding F (2019) Hydrogeochemical evolution of interaction between surface water and groundwater affected by exploitation. *Ground Water* 57(3):430–442
- He Y, Gong Z (2013) China's regional rainstorm floods disaster evaluation based on grey incidence multiple-attribute decision model. *Nat Hazards* 71(2):1125–1144
- He CY, Zhou MR, Yan PC (2016) Application of the identification of mine water inrush with LIF spectrometry and KNN algorithm combined with PCA. *Spectrosc Spect Anal* 36(7):2234–2237
- Huang P, Wang X (2018) Piper-PCA-fisher recognition model of water inrush source: a case study of the Jiaozuo mining area. *Geofluids* 2018:1–10
- Huang P, Yang Z, Wang X, Ding F (2019) Research on Piper-PCA-Bayes-LOOCV discrimination model of water inrush source in mines. *Arab J Geosci*. <https://doi.org/10.1007/s12517-019-4500-3>
- Li G, Meng Z, Wang X, Yang J (2016) Hydrochemical prediction of mine water inrush at the Xinli Mine, China. *Mine Water Environ* 36(1):78–86
- Li Y, Zhang D, Liu B (2019) Multi-attribute decision-making method based on cosine similarity with three-parameter interval grey number. *J Grey Syst* 31(3):45–58
- Liang XB, Liang W, Zhang LB, Guo XY (2019) Risk assessment for long-distance gas pipelines in coal mine gobs based on structure entropy weight method and multi-step backward cloud transformation algorithm based on sampling with replacement. *J Clean Prod* 227:218–228
- Liao F, Wang G, Yi L, Shi Z, Cheng G, Kong Q, Liu C (2020) Identifying locations and sources of groundwater discharge into Poyang Lake (eastern China) using radium and stable isotopes (deuterium and oxygen-18). *Sci Total Environ* 740:140163
- Lin Y, Wu Y, Pan G, Qin Y, Chen G (2014) Determining and plugging the groundwater recharge channel with comprehensive approach in Siwan coal mine, North China coal basin. *Arab J Geosci* 8(9):6759–6770
- Liu G, Ma F, Liu G, Zhao H, Guo J, Cao J (2019) Application of multivariate statistical analysis to identify water sources in a coastal gold mine, Shandong, China. *Sustainability* 11(12):3345
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909
- Qian J, Tong Y, Ma L, Zhao W, Zhang R, He X (2017) Hydrochemical characteristics and groundwater source identification of a multiple aquifer system in a coal mine. *Mine Water Environ* 37(3):528–540
- Sui W, Liu J, Yang S, Chen Z, Hu Y (2010) Hydrogeological analysis and salvage of a deep coalmine after a groundwater inrush. *Environ Earth Sci* 62(4):735–749
- Sun W, Zhou W, Jiao J (2015) Hydrogeological classification and water inrush accidents in China's coal mines. *Mine Water Environ* 35(2):214–220
- Wang D, Shi L (2019) Source identification of mine water inrush: a discussion on the application of hydrochemical method. *Arab J Geosci* 12(2):58

- Wang Y, Shi L, Wang M, Liu T (2020) Hydrochemical analysis and discrimination of mine water source of the Jiaojia gold mine area, China. *Environ Earth Sci* 79(6):123
- Wu Q, Mu W, 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 Env* 78(1):469–482
- Xu Y, Chang Q, Yan X, Han W, Chang B, Bai J (2020) Analysis and lessons of a mine water inrush accident resulted from the closed mines. *Arab J Geosci* 13(14):665
- Yan BQ, Ren FH, Cai MF, Qiao C (2019) Bayesian model based on Markov chain Monte Carlo for identifying mine water sources in Submarine Gold Mining. *J Clean Prod* 253:0–120008
- Yang Y, Yue JH, Li XH, Wang DM (2019) Online discrimination system for mine water inrush source based on PCA and BP neural network. *Acta Microscopica* 28(3):444–454
- Yi P, Yang J, Wang Y, Mugwanzeal VP, Chen L, Aldahan A (2018) Detecting the leakage source of a reservoir using isotopes. *J Environ Radioact* 187:106–114
- Yin L, Ma K, Chen J, Xue Y, Wang Z, Cui B (2019) Mechanical model on water inrush assessment related to deep mining above multiple aquifers. *Mine Water Environ* 38(4):827–836
- Zhang N, Fang ZG, Liu XQ (2014) Grey situation group decision-making method based on prospect theory. *Sci World J*:703597
- Zhang XB, Li X, Gao X (2016) Hydrochemistry and coal mining activity induced karst water quality degradation in the Niangziguan karst water system, China. *Environ Sci Pollut Res Int* 23(7):6286–6299
- Zhang H, Xing H, Yao D, Liu L, Xue D, Guo F (2019a) The multiple logistic regression recognition model for mine water inrush source based on cluster analysis. *Environ Earth Sci* 78(20):612
- Zhang HT, Xu GQ, Chen XQ, Wei J, Yu ST, Yang TT (2019b) Hydrogeochemical characteristics and groundwater inrush source identification for a multi-aquifer system in a coal mine. *Acta Geologica Sinica - English Edition* 93(6):1922–1932
- Zhang J, Xu K, Reniers G, You G (2020) Statistical analysis the characteristics of extraordinarily severe coal mine accidents (ESCMAs) in China from 1950 to 2018. *Process Saf Environ Prot* 133:332–340
- Zu XH, Yang, CL, Wang HC, Wang, YY (2018) An EGR performance evaluation and decision making approach based on grey theory and grey entropy analysis. *Plos One* 13(1):e0191626

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