

Hydrochemical Characteristics and Groundwater Source Identification of a Multiple Aquifer System in a Coal Mine

Jiazhong Qian¹ · Yan Tong¹ · Lei Ma¹ · Weidong Zhao¹ · Ruigang Zhang² · Xiaorui He¹

Received: 9 February 2017 / Accepted: 13 October 2017 / Published online: 23 October 2017
© Springer-Verlag GmbH Germany 2017

Abstract Hydrochemical analysis and Bayesian discrimination were used to identify groundwater sources for multiple aquifers in the Panyi Coal Mine, Anhui, China. The results showed that the Cenozoic top aquifer water were $\text{HCO}_3\text{-Na+K-Ca}$ and $\text{HCO}_3\text{-Na+K-Mg}$ types, which distinguished it from the other aquifers due to its low $\text{Na}^+ + \text{K}^+$ and Cl^- concentrations. The Cenozoic middle and Cenozoic bottom aquifer waters were mainly Cl-Na+K and $\text{SO}_4\text{-Cl-Na+K}$ types. The water types in the Permian fractured aquifer changed from Cl-Na+K to $\text{HCO}_3\text{-Cl-Na+K}$ and $\text{HCO}_3\text{-Na+K}$ moving away from the Panji anticline. In addition, HCO_3^- concentrations increased and Ca^{2+} concentrations decreased with depth in the aquifer. The Taiyuan limestone aquifers were of Cl-Na+K , $\text{SO}_4\text{-Cl-Na+K}$, and $\text{HCO}_3\text{-Cl-Na+K}$ types, and are very difficult to distinguish from the other aquifers. The precision of the Bayesian discrimination based on groundwater chemistry was 86.09%. Water chemistry indicators in the Permian fractured aquifer were moderately to highly variable and moderately to strongly correlated, spatially. The water chemistry spatial distribution indicates that the Permian fractured aquifer is recharged by the Cenozoic bottom aquifer near the Panji anticline, which reduces the accuracy of Bayesian discrimination in that area.

Keywords Water-inrush · Hydrochemistry · Geostatistics · Bayesian discrimination

Introduction

Hydrogeological conditions are extremely complex in China, which can have adverse effects on mine safety. In recent years, as the depth, intensity, and scale of coal mining have increased, groundwater disasters have become more severe and hazardous. According to an incomplete official statistical report, from 2007 to 2014, 128 water-inrush accidents caused grave losses of life and property in China (Xing et al. 2016). The revival of the Silk Road will boost coal mining, triggering more such disasters (Li et al. 2015a, 2017). Thoroughly characterizing a site's groundwater hydrochemical parameters and identifying potential inrush sources promptly and correctly are important in preventing such disasters.

Chemical characteristics of groundwater can be used to help predict potential mining hazards. Inrush sources can be identified based on the hydrochemical characteristics of the site (Waltoday et al. 2009; Wang et al. 2015). Methods of identifying inrush sources nowadays are multidisciplinary; a broad spectrum of fields can be involved, including chemistry, mathematics, computer science, GIS, and extenics. Advances have also been achieved in systematization and visualization for these approaches. For example, multivariate statistical analyses, such as cluster analysis (Güler et al. 2002; Wu et al. 2014a; Zhao et al. 2011), factor analysis (Shrestha and Kazama 2007; Wunderlin et al. 2001), principal components analysis (Bengranine and Marhaba 2003; Li et al. 2013; Ma et al. 2016), and discriminant analysis (Ma et al. 2014) have been widely used to analyze hydrochemical characteristics and identify influential factors. Recently, the fuzzy comprehensive evaluation method (Dong and Liu

✉ Lei Ma
Lei8505@hfut.edu.cn

¹ School of Resources and Environmental Engineering,
Hefei University of Technology, Hefei 230009,
People's Republic of China

² School of Civil Engineering, Hefei University of Technology,
Hefei 230009, People's Republic of China

2008; Gong and Jin 2009; Li et al. 2012), artificial neural network method (Kanti and Rao 2008; Wu and Yu 2011), analytical model of gray-related degree (Yang et al. 2010), and Bayes discrimination model (Hobbs 1997; Zhang et al. 2010) have been used to better identify water-inrush sources. Due to its costs and logistics, the collection of groundwater quality data tends to be intermittent and localized (Uddameri 2007). Bayesian statistics offers a convenient approach to obtain insights from the existing data (Roiger and Geatz 2003). Additionally, using the Bayesian approach, the influence of prior probability is considered, which improves the sensitivity and accuracy of the discrimination. Meanwhile, subjective perceptions held by decision-makers can be factored in as a prior probability and conditioned using the collected data (Schmidt 1969). In this study, a Bayesian discriminant model was implemented using MATLAB to discriminate water-inrush sources. The feasibilities and advantages of MATLAB have been validated by successful applications in other fields, such as monthly inflow forecasting in dam reservoirs (Valipour et al. 2013), land use and irrigation (Valipour 2016a, b), modelling evapotranspiration (Valipour 2015; Rezaei et al. 2016; Valipour et al. 2017) and rainfall estimation (Valipour 2016a, b).

However, it is difficult to use these models to describe overall water quality conditions due to continuous water–rock interaction (Stotler et al. 2009). Measuring the spatial variations of a wide range of indicators (i.e. chemical, physical, and biological indicators) using these models can also be challenging (Babiker et al. 2007). Geostatistical analysis can be used to estimate values in unsampled areas using variants of kriging, which was first used to quantify mineral resources, and has now been widely applied for spatial analysis in pedology, hydrology, ecology, and other fields (Ahmadi and Sedghamiz 2007; Christakos 2000; Isaaks and Srivastava 1989; JÚnez-Ferreira et al. 2016; Nikroo et al. 2010; Passarella et al. 2002; Wu et al. 2014b). Interpolation of spatial variables such as soil properties, groundwater level and ion concentrations are made by different kriging methods, including simple kriging, ordinary kriging, universal kriging, and co-kriging. The combination of Bayesian analysis and geostatistics may facilitate the intuitive understanding of the groundwater hydrochemical characteristics at a site.

Hydrochemical characterization and water inrush source identification have already been broadly studied (e.g. Ma et al. 2016; Shrestha and Kazama 2007; Wu and Zhou 2008; Wu et al. 2014a; Wu and Yu 2011). Prediction of the types of ions in groundwater, as well as their corresponding concentrations, can be very difficult due to hydrogeological uncertainty. Some (e.g. Ma et al. 2014; Wu and Yu 2011; Yang et al. 2010; Zhang et al. 2010) simply used discrimination models to identify the water-inrush sources, or compared several models to select the one that seemed to be

more accurate. The causes of misjudgments and sources of error were often overlooked. Therefore, in this study, we analyzed the causes of errors while investigating the spatial hydrogeochemical distribution.

Study Area and Settings

Study Area

The Panyi Mine is located in the Panji District, northwest of Huainan City, Anhui Province, China (Fig. 1). The total area of the mine is 54.67 km² and it is covered by a relatively flat Huai River alluvial plain. The ground elevation varies from 19 to 23 m above sea level (a.s.l.). This mine site is situated within a temperate, semi-humid climate controlled by seasonal monsoons. The annual average temperature is 15.1 °C, and the annual average precipitation is 926.30 mm. Generally, rainfall in June, July, and August accounts for about 40% of the total precipitation in a given year.

Geological Settings

The mine strata exposed by drilling were mainly Ordovician, Carboniferous, Permian, and Cenozoic. The Cenozoic strata was in unconformity contact with the underlying Paleozoic strata. The thickness of the Cenozoic strata varies from 149.00 to 369.72 m; the average thickness was 320 m. The Permian strata is the main coal-bearing strata. The Shihezi and Shanxi claystone and sandstone formations in the lower middle part of the Permian contains 27–42 coal beds.

The 121 m thick Carboniferous strata is composed of limestone, claystone, and medium-fine sandstone and contains five to ten unstable and thin coal beds and carbonaceous mudstone. The Ordovician strata is about 111.5 m thick and is composed of thick-bedded dolomite, limestone, and mudstone. The regional Ordovician has a parallel unconformable contact with the overlying Carboniferous strata.

The Panyi Mine is located in the south of the Panji anticline. The stratigraphic strike of the mine is N30°E–N60°W. The dip direction is SE–SW, and the dip angle gradually lessens from shallow to deep (20°–7°). The main geological structures in this mine are oblique trans-tensional faults.

Hydrogeological Settings

Primary aquifers that could compromise mine safety and production include a loose, sandy, porous Cenozoic aquifer group, a Permian fractured sandstone aquifer group, and a fractured and a karstified limestone aquifer of the Carboniferous Taiyuan Formation.

The Cenozoic aquifer group can be divided into three aquifers based on the sediment and the water-containing

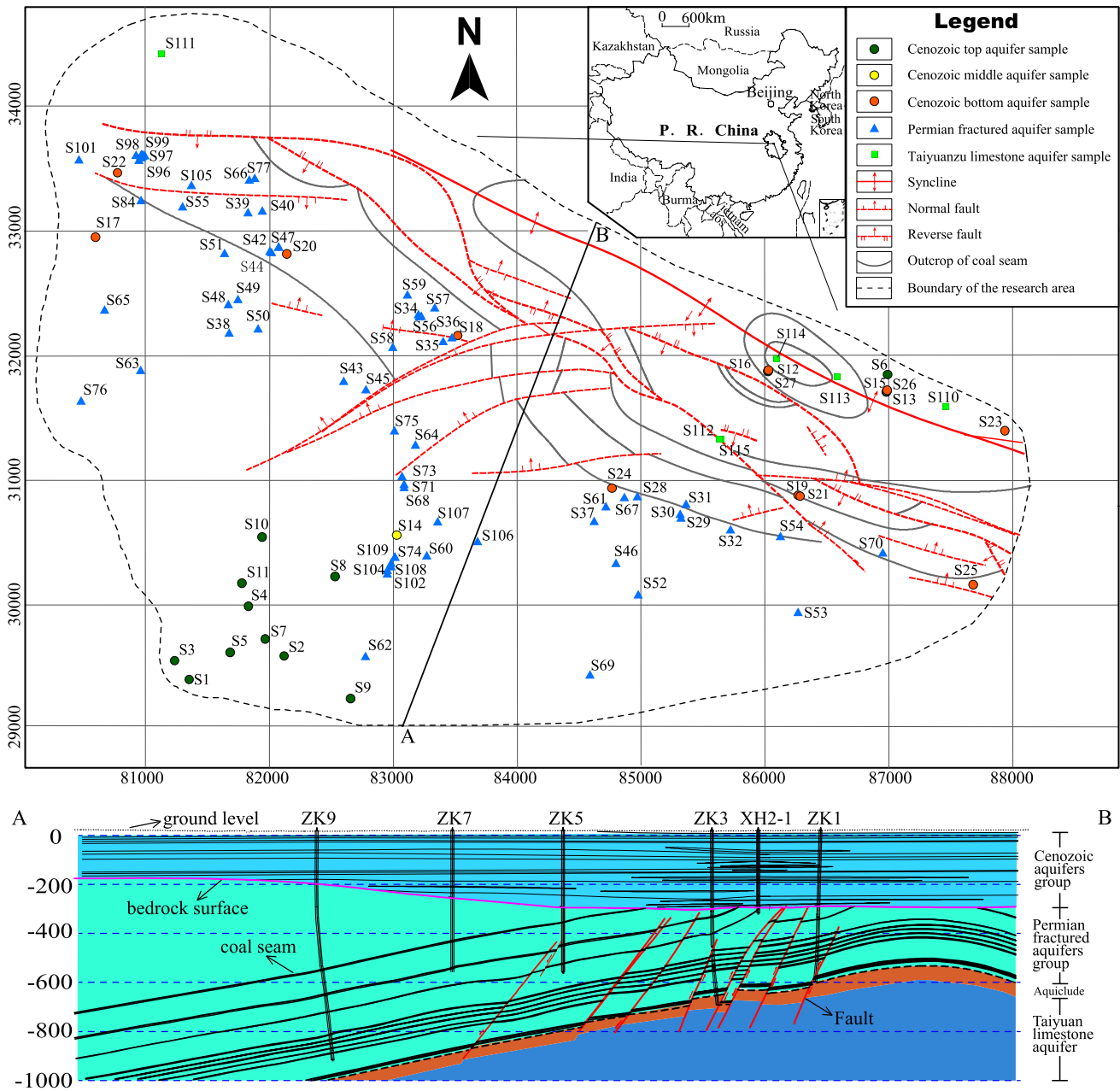


Fig. 1 Location of the study area in China, geological structure of the mine (in plan view) and the distribution of water samples (plan and sectional elevation)

condition. The lowest of the three, composed of gravel and sand layers with some clay, is 0–87.79 m thick. Its thickness increases from southeast to northwest, with an average of 62.88 m. The original water level of the aquifer was +24 m, and the natural hydraulic gradient was 1/10,000; owing to the exploitation, the groundwater level dropped to approximately –10 m, and formed a cone of depression. The specific capacity of this layer was 0.01–2.00 (L/s)/m, the hydraulic conductivity was 0.20–6.00 m/day, and the water temperature was 23–26 °C.

The middle aquifer is 40.84–170.95 m thick, with a tendency to thin from northwest to southeast. It is mainly composed of sandy clay, intercalating with silt and fine sand layers, and partly intercalating with moderately coarse sand layers. The original water level of the aquifer was about +23 m, but long-term observation data shows that its water level has been slowly descending. Its specific capacity was 0.27–1.00 (L/s)/m and its hydraulic conductivity was 2.27–4.78 m/day.

The top aquifer is 73.35–115.75 m thick, with an average thickness of 86.96 m, and thickens from east to west. Its specific capacity was 0.98–6.18 (L/s)/m and its water temperature was 16.5–19 °C. This aquifer, which is the alluvial deposit of the Huaihe River, has good water quality and is a water supply source.

The Permian fractured aquifer was distributed between the coal beds. Its lithology and thickness varies greatly. Generally, within the aquifer, fractures are not well developed. Its static water level was between 1.18 and 1.25 m and its temperature was about 24 °C.

The Taiyuan limestone aquifer had a total thickness of 140 m, including 13 limestone layers that are, in total, about 41–54 m thick. Two pumping tests showed that the original water level was 26–28 m, the specific capacity was 0.12–0.19 (L/s)/m, the hydraulic conductivity was 0.01–0.30 m/day, and the water temperature was 32–36 °C.

Materials and Methods

Water Samples

A total of 115 water samples were collected from the coal mine, of which 13 were from the Cenozoic top aquifer, three from the Cenozoic middle aquifer, 11 from the Cenozoic bottom aquifer, 82 from the Permian fractured aquifer, and six from the Taiyuan limestone aquifer (Fig. 1). Each sample was tested for the major cations (i.e. calcium (Ca²⁺), magnesium (Mg²⁺), potassium (K⁺) and sodium (Na⁺), summed together as Na⁺ + K⁺), major anions [i.e. bicarbonate (HCO₃⁻), chloride (Cl⁻), sulfate (SO₄²⁻)], and pH at the Water Quality Testing Center, Huainan Mining Industry (Group) Co., Ltd, Huainan, China. This testing center has a strict quality control system incorporating national standards. The variables were used to characterize the groundwater’s geochemical characteristics and reveal groundwater–rock interactions. The results for each aquifer are listed in Table 1.

Each water sample was tested for the major cations [i.e. calcium (Ca²⁺), magnesium (Mg²⁺), potassium (K⁺) and sodium (Na⁺), summed together as Na⁺ + K⁺], major anions [i.e. bicarbonate (HCO₃⁻), chloride (Cl⁻), sulfate (SO₄²⁻)], and pH at the Water Quality Testing Center, Huainan Mining Industry (Group) Co., Ltd, Huainan, China. This testing center has a strict quality control system incorporating national standards. The variables were selected to characterize the groundwater’s hydrogeochemical characteristics and reveal groundwater–rock interactions. The results for each aquifer are listed in Table 1.

Bayesian Discrimination

Bayesian analysis provides a way to calculate the probability of a hypothesis. It departs from the common notion of probability, which is based on the chance of observing a certain value (state) by repeated experimentation. Being more generic, Bayesian statistics can deal with subjective probabilistic notions held by decision-makers as well as objective measurements obtained via observations and experimentation (Uddameri 2007). Bayesian discrimination is based on the prior probability of a hypothesis, the probability of observing different data under a given hypothesis, and the probability of observing the observed data itself. The method combines the prior information about the unknown parameters with available information to predict subsequent findings using Bayes’ formula. Unknown parameters can then be deduced using the predicted results. The specific process is explained below:

Assume that there are *k* populations, *G*₁, *G*₂, ..., *G*_{*k*}, and their prior probabilities are *q*₁, *q*₂, ..., *q*_{*k*}, respectively. The density functions of the populations are *f*₁(*x*), *f*₂(*x*), ..., *f*_{*k*}(*x*). If not acquired in advance, values of the prior probabilities can be set as 1/*k*, which means the prior probabilities have no effects. When observing a sample *x*, we can use Bayes formula to calculate the posteriori probability of *x* coming from the *g*th population (*G*_{*g*}). The function is:

$$P\left(\frac{g}{x}\right) = \frac{q_g f_g(x)}{\sum_{i=1}^k q_i f_i(x)}, \quad g = 1, 2, \dots, k \tag{1}$$

If the following equation holds, it has the maximum probability that sample *x* comes from the *h*th population (*G*_{*h*}).

$$P\left(\frac{h}{x}\right) = \max_{1 \leq g \leq k} P(g/x) \tag{2}$$

Assuming that the population is divided into *p* classes, and is normally distributed, the *p*-dimensional normal distribution density function is:

$$f_g(x) = (2\pi)^{-\frac{p}{2}} \left| \sum^{(g)} \right|^{-\frac{1}{2}} \cdot \exp \left[-\frac{1}{2} (x - u^{(g)})' \sum^{(g)-1} (x - u^{(g)}) \right] \tag{3}$$

where *u*^(*g*), $\sum^{(g)}$ are the mean vector (*p*-dimensional) and covariance (*p*-order) of the *G*_{*g*}, respectively. Substitute Eq. (6) into Eq. (4), and find the *g* which makes P(*g*/*x*) maximum. The denominator in Eq. (4) is a constant, no matter what *g* is. Thus, the equation becomes:

$$q_g f_g(x) \xrightarrow{g} \max \tag{4}$$

After logarithmic processing and removing items unrelated to *g*, the following can then be obtained:

Table 1 Summary statistics of groundwater variables concentrations for 115 samples from five aquifers of the minfive

Aquifer ID	No. of samples	Variables	Mean	Std. deviation	Minimum	Maximum	Percentiles		
							25	50	75
1	13	Ca ²⁺	47.28	30.10	3.29	107.82	22.24	62.52	64.93
		Mg ²⁺	16.99	9.04	2.52	29.67	7.17	19.70	22.13
		Na ⁺ +K ⁺	79.56	35.57	20.92	158.62	63.68	77.94	88.97
		HCO ₃ ⁻	322.95	167.16	70.48	497.92	153.16	421.63	466.87
		Cl ⁻	28.97	24.81	6.74	95.19	11.70	23.75	34.39
		SO ₄ ⁻	28.67	48.08	1.44	153.70	4.94	9.88	19.96
		pH	7.86	0.40	7.05	8.80	7.78	7.81	8.03
2	3	Ca ²⁺	33.19	12.45	20.74	45.64	26.97	33.19	39.42
		Mg ²⁺	4.62	5.82	1.26	11.34	1.26	1.26	6.30
		Na ⁺ +K ⁺	752.45	92.33	694.31	858.91	699.22	704.12	781.52
		HCO ₃ ⁻	310.64	84.92	231.42	400.29	265.82	300.22	350.26
		Cl ⁻	683.84	116.12	601.17	816.60	617.46	633.74	725.17
		SO ₄ ⁻	467.49	212.49	268.97	691.63	355.43	441.88	566.76
		pH	8.53	0.54	8.09	9.13	8.24	8.38	8.76
3	11	Ca ²⁺	47.94	10.68	34.23	66.31	40.98	44.01	54.10
		Mg ²⁺	26.94	14.74	2.52	51.85	19.43	23.07	33.98
		Na ⁺ +K ⁺	823.11	40.54	762.31	893.19	791.42	828.69	844.75
		HCO ₃ ⁻	320.33	20.06	289.17	348.05	306.85	325.11	336.58
		Cl ⁻	999.44	78.35	844.15	1115.17	1015.90	1020.20	1026.20
		SO ₄ ⁻	327.10	89.03	204.24	557.15	297.93	305.33	318.09
		pH	8.11	0.35	7.80	8.82	7.90	8.00	8.13
4	82	Ca ²⁺	29.65	19.66	1.20	74.15	6.41	35.07	43.94
		Mg ²⁺	12.24	9.75	0.24	55.94	3.01	13.01	18.24
		Na ⁺ +K ⁺	942.20	301.30	630.39	1874.14	762.29	813.73	949.45
		HCO ₃ ⁻	695.41	804.08	162.92	3540.38	290.91	334.38	490.90
		Cl ⁻	992.02	183.95	464.04	1553.77	977.38	1027.52	1046.75
		SO ₄ ⁻	134.47	138.32	0.60	691.63	32.31	108.15	189.75
		pH	8.34	0.50	7.16	9.90	8.01	8.45	8.68
5	6	Ca ²⁺	52.97	42.84	12.63	115.80	17.05	44.46	80.54
		Mg ²⁺	21.55	11.46	7.56	36.65	11.94	24.26	27.62
		Na ⁺ +K ⁺	915.99	73.96	847.91	1055.50	872.84	900.90	920.65
		HCO ₃ ⁻	583.09	474.95	88.47	1379.00	352.64	380.59	790.09
		Cl ⁻	928.46	199.55	544.38	1061.46	911.20	1027.01	1035.48
		SO ₄ ⁻	433.99	363.35	1.65	1079.49	254.92	420.40	466.80
		pH	8.00	0.79	7.20	9.30	7.50	7.73	8.40

Aquifer ID: 1-Cenozoic top aquifer, 2-Cenozoic middle aquifer, 3-Cenozoic bottom aquifer, 4-Permian fractured aquifer, 5-Taiyuan limestone aquifer. All units of variables are mg/L except for pH

$$\begin{aligned}
 Z(g/x) &= \ln q_g - \frac{1}{2} \ln \left| \sum^g \right| - \frac{1}{2} (x - u^{(g)})' \sum^{(g)-1} (x - u^{(g)}) \\
 &= \ln q_g - \frac{1}{2} \ln \left| \sum^g \right| - \frac{1}{2} x' \sum^{(g)-1} x - \frac{1}{2} u^{(g)'} \sum^{(g)-1} u^{(g)} + x' \sum^{(g)-1} u^{(g)}
 \end{aligned}
 \tag{5}$$

The problem then becomes:

$$Z(g/x) \xrightarrow{g} \max \tag{6}$$

There are k covariance matrixes in $Z(g/x)$, and $Z(g/x)$ is a quadratic function. Solving this can be too complex so, to simplify the computational process, k covariance matrixes are assumed to be the same, namely $\sum^{(1)} = \sum^{(2)} = \dots = \sum^{(k)} = \sum$. Then, we can remove $\frac{1}{2} \ln \left| \sum^g \right|$ and $\frac{1}{2} x' \sum^{(g)-1} x$, which are unrelated to g , to obtain the following discriminant function:

$$\begin{cases} y\left(\frac{g}{x}\right) = \ln q_g - \frac{1}{2}u^{(g)'} \sum u^{(g)} + x' \sum u^{(g)} \\ y(g/x) \xrightarrow{g} \max \end{cases} \quad (7)$$

Discriminant function can also be written in polynomial form:

$$y(g/x) = \ln q_g + C_0^{(g)} + \sum_{i=1}^p C_i^{(g)} x_i \quad (8)$$

with the parameters calculated as follows:

$$C_i^{(g)} = \sum_{j=1}^p v^{ij} \mu_j^{(g)}, \quad i = 1, 2, \dots, p \quad (9)$$

$$C_0^{(g)} = -\frac{1}{2} \mu^{(g)'} \sum \mu^{(g)} = -\frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p v^{ij} \mu_i^{(g)} \mu_j^{(g)} = -\frac{1}{2} \sum_{i=1}^p C_i^{(g)} \mu_i^{(g)} \quad (10)$$

$$x = (x_1, x_2, \dots, x_p)' \quad (11)$$

$$\mu^{(g)} = (\mu_1^{(g)}, \mu_2^{(g)}, \dots, \mu_p^{(g)})' \quad (12)$$

$$\sum = (v_{ij})_{p \times p} \quad (13)$$

Geostatistical Analysis

Geostatistics is based on the theory of regionalized variables (Matheron 1971), which states that attributes within an area exhibit both random and spatially structured properties (Antunes and Albuquerque 2013; Journel and Huijbregts 1978). To understand the spatial variability in groundwater chemistry, semi-variogram analysis, which is defined as half of the variance of the increment $[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]$, was conducted to analyze spatial patterns and correlation. It can be expressed as (Li et al. 2014, 2015b):

$$\gamma(\mathbf{h}) = \frac{1}{2} E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2 \quad (14)$$

and is normally obtained using the method of moments estimator, as follows:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2 \quad (15)$$

where $N(h)$ is the number of data pairs. $Z(\mathbf{x})$ and $Z(\mathbf{x} + \mathbf{h})$ are the values at \mathbf{x} and $\mathbf{x} + \mathbf{h}$, and the two locations are separated

by vector \mathbf{h} with specified direction and distance tolerance. Some theoretical semi-variograms models, such as linear, exponential, spherical and Gaussian models, were tested for their appropriateness of fit to the sample semi-variograms.

Kriging methods are quite flexible, but within the kriging family there are varying degrees of conditions that must be met for the output to be valid. The universal kriging model is:

$$Z(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i) \quad (16)$$

where $Z(\mathbf{x}_0)$ is the estimated value at location \mathbf{x}_0 , λ_i is the kriging weight, and $Z(\mathbf{x}_i)$ is the measured value at point \mathbf{x}_i . The calculation equation for λ_i is calculated in the same manner as Ahmadi and Sedghamiz (2007) and Lee (1997).

Results and Discussion

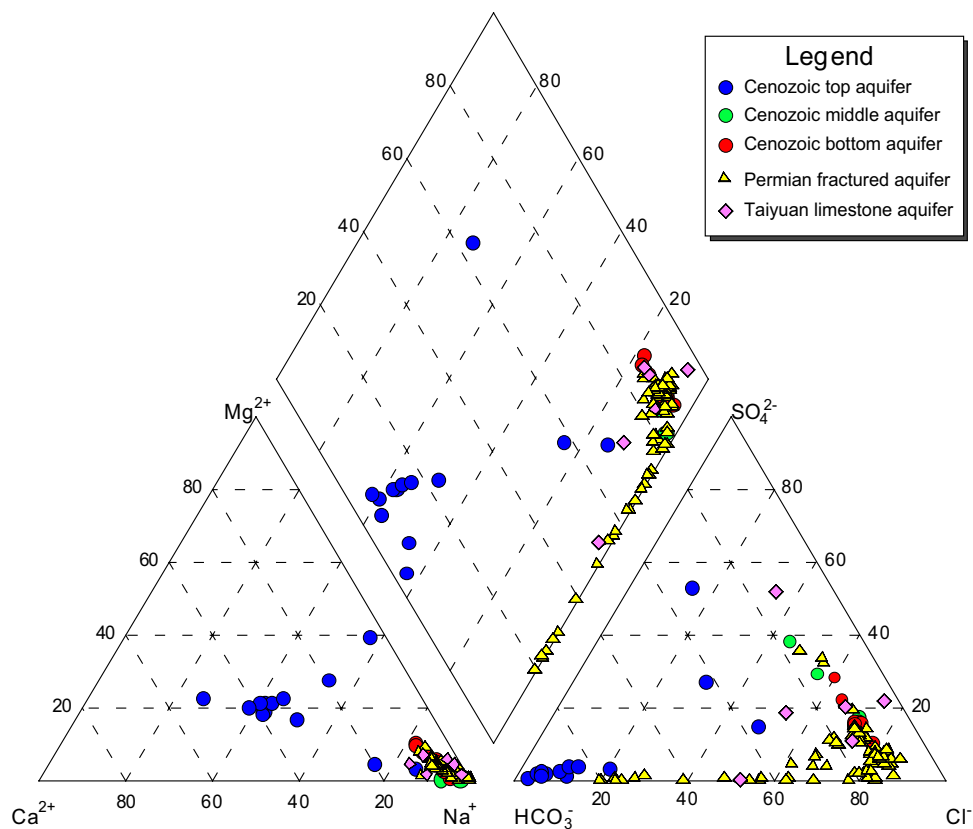
Groundwater systems are complicated and vary in space and time. Investigating the chemical characteristics of groundwater has important implications for analyzing groundwater sources, controlling groundwater disasters, and managing the use of groundwater resources. Hence, we attempted to understand the groundwater’s variability and spatial distribution of indicators as the basis for groundwater source identification by analyzing the hydrogeochemistry of the site’s aquifers.

General Chemical Characteristics of the Aquifers

A Piper trilinear diagram was constructed to evaluate water quality and analyze changing water quality trends. As seen in Fig. 2, the water quality of the Cenozoic top aquifer is very different from that of the other deeper confined aquifers. The Cenozoic top aquifers are of $\text{HCO}_3\text{-Na+K-Ca}$ and $\text{HCO}_3\text{-Na+K-Mg}$ water types. Milliequivalent percentage of HCO_3^- reaches over 80% in most of the water samples from Cenozoic top aquifer, whereas in the Cenozoic bottom aquifer, milliequivalent percentage of Cl^- takes up more than 60%. The Cenozoic top aquifer is an unconfined or weakly-confined aquifer; it directly receives low salinity recharge from precipitation and surface water. Therefore, this aquifer can be easily distinguished from the other aquifers based on its distinctive chemistry.

Most of the water samples from the Cenozoic middle and bottom aquifers are Cl-Na+K and $\text{SO}_4\text{-Cl-Na+K}$ water types (Fig. 2). This is indicative of long-term water–rock interaction and little surface water input; therefore, both are quite saline (Ma et al. 2016). According to the hydrogeological survey, the bottom gravel layer of the Cenozoic bottom aquifer is in localized contact with the outcrop of the

Fig. 2 Piper tri-linear diagram of hydrogeochemical facies of groundwater samples from the five aquifers



Permian fractured aquifer and there is a hydraulic connection between the Cenozoic bottom aquifer and the Permian fractured aquifer around the Panji anticline. Therefore, many water samples of these two aquifers overlap in the Piper trilinear diagram.

As shown in Fig. 2, water samples from the Permian fractured aquifer are all dominated by alkali metal ions ($\text{Na}^+ + \text{K}^+$) relative to the concentration of alkaline earth metal ions (Ca^{2+} , Mg^{2+}). The milliequivalent percentage of $\text{Na}^+ + \text{K}^+$ is more than 80%. Most of the water samples from this aquifer have a high Cl^- concentration, while a small number have a high HCO_3^- concentration. Therefore, the Permian fractured aquifer samples are mainly $\text{Cl}-\text{Na}+\text{K}$ and $\text{HCO}_3-\text{Cl}-\text{Na}+\text{K}$ water types, varying mostly in HCO_3^- . The shallow parts of the aquifer, which are close to the water samples from the Cenozoic bottom aquifer in the Piper trilinear diagram, have relatively low HCO_3^- concentrations. Deeper, the HCO_3^- concentration increases because of little or no impact from the Cenozoic bottom aquifer water (Ma et al. 2016).

In the Taiyuan limestone aquifer, $\text{Na}^+ + \text{K}^+$ are again the dominant cations, but the anion concentrations change greatly, forming three kinds of hydrochemical facies: $\text{Cl}-\text{Na}+\text{K}$, $\text{SO}_4-\text{Cl}-\text{Na}+\text{K}$, and $\text{HCO}_3-\text{Cl}-\text{Na}+\text{K}$. These water samples are close to those of the Cenozoic bottom aquifer in the Fig. 2. Thus, the groundwater of the

Cenozoic bottom aquifer, Taiyuan limestone aquifer and part of the Permian fractured aquifer are geochemically similar. Boxplots (Fig. 3a–f) were generated to analyze the distribution of the ion concentrations. Box edges represent the first and the third quartile with median value shown in the middle of the box. The horizontal lines at the bottom and the top of the boxplot are the minimum and the maximum values, respectively. Extremes and outliers are represented by “*” symbols and “O” symbols, respectively.

From Fig. 3c, e, the Cenozoic top aquifer has low $\text{Na}^+ + \text{K}^+$ and Cl^- concentrations and can easily be discriminated from other aquifers using these two indicators. As shown in the Fig. 3a, the Ca^{2+} interquartile ranges of the Cenozoic bottom aquifer and the Permian fractured aquifer are 2.1–3.0 and 0.3–1.3 meq/L, respectively. Figure 3b shows that the Mg^{2+} box bodies of these two aquifers are generally staggered. High HCO_3^- concentration can also be considered as characteristic of part of the Permian fractured aquifer (Fig. 3d). Thus, the Ca^{2+} , Mg^{2+} , HCO_3^- , and SO_4^{2-} could only partially distinguish the Permian fractured aquifer from the Cenozoic bottom aquifer and Taiyuan limestone aquifer. Therefore, Bayesian discrimination was employed to identify the groundwater source in the aquifers.

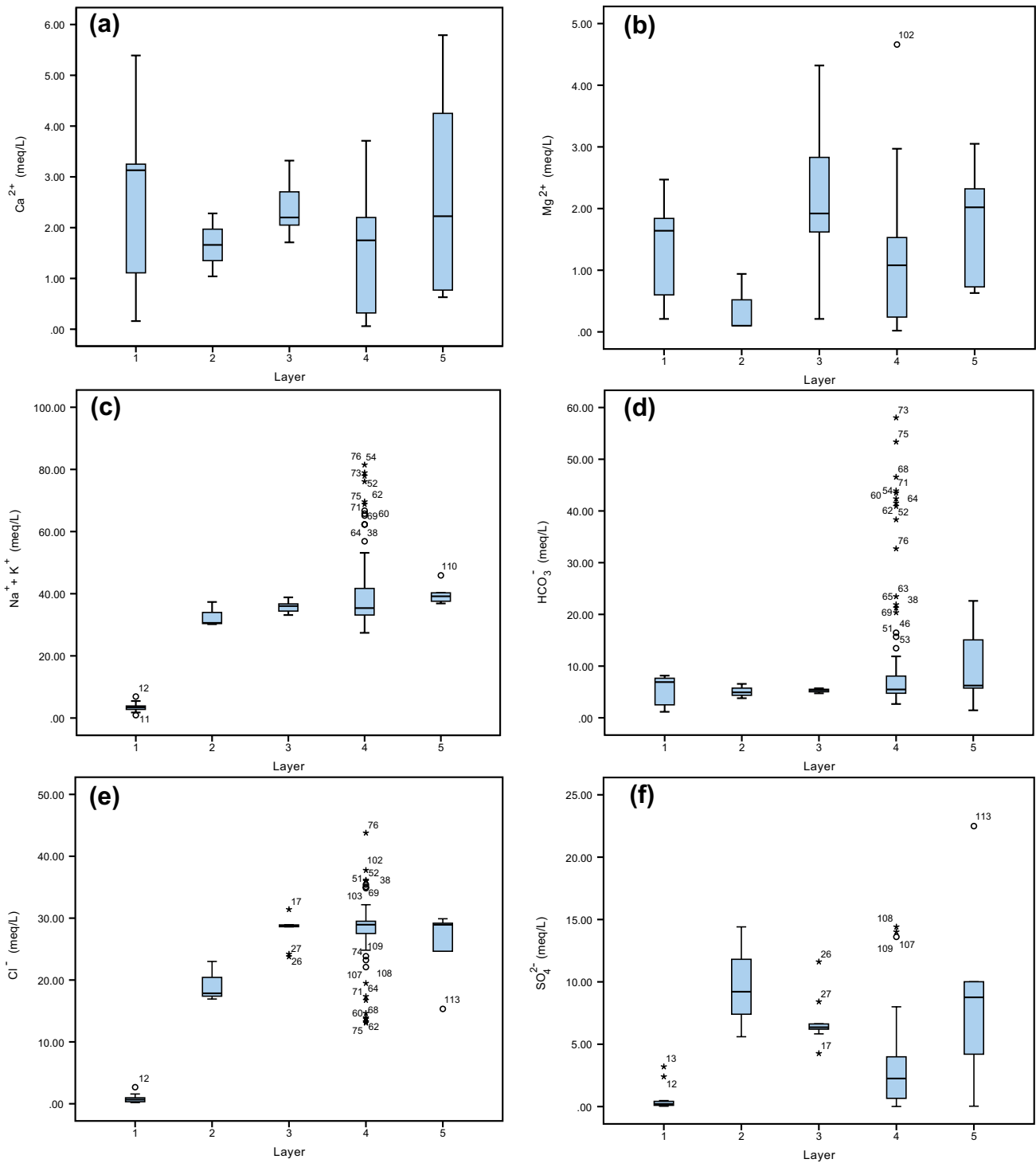


Fig. 3 Boxplots of the water samples (a–f). Aquifer ID: 1-Cenozoic top aquifer, 2-Cenozoic middle aquifer, 3-Cenozoic bottom aquifer, 4-Permian fractured aquifer, 5-Taiyuan limestone aquifer. The units are mg/L

Bayesian Discrimination Analysis

All of the samples were analyzed and discriminated by the Bayesian discrimination model, which was implemented using MATLAB. Discrimination factors used were Ca²⁺,

Mg²⁺, Na⁺ + K⁺, HCO₃⁻, Cl⁻, and SO₄²⁻. Incorrect discrimination results are shown in Table 2; the rest of the results are correct. The Bayes discrimination was 86.09% correct. All of the water samples from the Cenozoic top aquifer were correctly identified. One sample of the Cenozoic middle

Table 2 Water samples of error identification by the Bayes discrimination approach

Sample no.	Aquifer ID	Discrimination result	Water sample no.	Aquifer ID	Discrimination result
S15	2	3	S40	4	3
S18	3	4	S56	4	2
S19	3	4	S57	4	3
S20	3	4	S59	4	3
S22	3	4	S110	5	4
S26	3	2	S111	5	3
S34	4	3	S113	5	4
S36	4	3	S114	5	3

Aquifer ID: 1-Cenozoic top aquifer, 2-Cenozoic middle aquifer, 3-Cenozoic bottom aquifer, 4-Permian fractured aquifer, 5-Taiyuan limestone aquifer

aquifer was inaccurately identified as being from the bottom aquifer. Five samples from the Cenozoic bottom aquifer were inaccurately identified as being from the Permian fractured aquifer and Cenozoic middle aquifer, and six samples from the Permian fractured aquifer were incorrectly identified as from the Cenozoic bottom or middle aquifers. Four of the six samples collected from the Taiyuan limestone aquifer were incorrectly identified as the Cenozoic bottom aquifer and Permian fractured aquifer.

Combining the hydrochemical characteristics and local hydrogeology, the Cenozoic top aquifer receives recharge from the surface and meteoric water. This groundwater is characterized by low TDS and low concentrations of $\text{Na}^+ + \text{K}^+$ and Cl^- ; therefore, it can be distinguished accurately from the other waters by Bayesian discrimination.

As discussed above, long-term data reveals the hydraulic relationship between the Cenozoic middle and bottom aquifers. First, because of mining, the water level in the bottom aquifer has declined continuously, while the hydraulic gradient between the two aquifers has increased. Thus, water of the middle aquifer may leak through the locally thin and lentoid clay layer and recharge the bottom aquifer. Second, at the southern bedrock uplifting area, the bottom aquifer and its upper clay layer pinch out, resulting in direct contact between the Cenozoic middle aquifer and the bedrock. When the water level of the Cenozoic bottom aquifer is lower than the middle aquifer, water will permeate through the weathered bedrock and recharge the bottom aquifer. Thus, these aquifers may be hydraulically connected, which would explain their similar hydrochemistry and the incorrect aquifer identification for some of the water samples.

The Permian fractured aquifer is covered by the Cenozoic bottom aquifer (Fig. 1) and receives recharge from the Cenozoic bottom aquifer, as reflected in the Piper diagram (Fig. 2) where many water samples of these two aquifers overlap. The samples from the Cenozoic bottom aquifer and Permian fractured aquifer are mutually misjudged because of their chemical similarity.

In addition, the number of samples (3 and 6, respectively) from the Cenozoic middle and Taiyuan limestone aquifers resulted in an unbalanced dataset (McBain and Timusk 2011), which made it difficult to obtain accurate results using these classifications, which assumed that the dataset was balanced. Much research has been put into addressing unbalanced datasets from different perspectives, including improved algorithms and data preprocessing techniques (Duan et al. 2016). Given that the aquifers are at different depths, adopting geotemperature as a discriminant factor may be an efficient way to improve model precision.

Spatial Distribution of Groundwater Chemistry in the Permian Fractured Aquifer

From the Bayes discrimination results, samples from the Cenozoic bottom aquifer and Permian fractured aquifer sometimes are mutually misjudged. When there are hydraulic relationships in local area between the aquifers and their hydrochemical characteristics are similar, such misjudgment is inevitable. Geostatistics was therefore used to illustrate the spatial distribution characteristics of the indicators and to verify whether there is a recharge area.

Table 3 presents the statistics of the hydrochemical parameters of the Permian fractured aquifer. The coefficient of variation (C_v) reflects the dispersion of the sample point distribution: $C_v < 10\%$ is considered as low variability, $10\% < C_v < 100\%$ is considered as medium variability, and $C_v > 100\%$ is considered as high variability (Wang et al. 2001). In the Permian fractured aquifer, the C_v values of Ca^{2+} , $\text{Na}^+ + \text{K}^+$, and Cl^- are within 18.5–66.3%, presenting medium variability; the C_v of HCO_3^- is 115.6%, characterized by high variability, which is consistent with the Piper trilinear diagram. The fundamental assumption for geostatistics is that the samples follow normal distribution, suggesting that the normality test is significant before performing semi-variogram analysis. Table 3 lists the results of normality tests of the indicators and the trend analysis. According

Table 3 The statistics for the Permian fractured aquifer indicators

Indicator	C _v (%)	Sample distribution		Model distribution	Trend analysis	Interpolation model
		Skewness	Kurtosis			
Ca ²⁺	66.30	−0.07	1.81	Non-normal	First order	UK
Na ⁺ +K ⁺	32.00	1.72	4.82	Non-normal	Second order	UK
HCO ₃ [−]	115.60	2.10	6.18	Non-normal	Second order	UK
Cl [−]	18.50	−0.96	5.75	Non-normal	Second order	UK

The suitable interpolation model for each turned out to be universal kriging (UK)

to this table, universal Kriging may be a suitable interpolation model for indicators. This study simulated indicators with different fitting models for the selection of an optimal semi-variogram model. Based on the principle of the least mean prediction error, the best fitting models and the semi-variogram parameters of the models are given in Table 4. Table 4 shows that each indicator has a large range, indicating a large correlation distance. The nugget to sill ratio, C₀/(C₀+C₁), represents the proportion of the spatial heterogeneity caused by a random part in the total variation. C₀/(C₀+C₁) < 0.25 means strong spatial correlation, 0.25 ≤ C₀/(C₀+C₁) < 0.75 means medium spatial correlation, and C₀/(C₀+C₁) > 0.75 means weak spatial correlation (Guo et al. 2000). From Table 4, the Na⁺+K⁺, HCO₃[−], and Cl[−] all have strong spatial correlation, mainly due to spatial structural factors such as geologic structure, terrain, aquifer media, and soil type (Wu et al. 2014a, b). Ca²⁺ has a moderate spatial correlation, which was caused by the combined effects of structural and stochastic factors.

Figure 4a–d show the spatial distribution for Ca²⁺, Na⁺+K⁺, HCO₃[−], and Cl[−]. High concentration values of Ca²⁺ occur near the Panji anticline, while low values are found in the southern research area. The Na⁺+K⁺ and HCO₃[−] maps are similar to each other, with an increasing trend in the Panji anticline area and in the south. In general, ion concentrations vary the most perpendicular to the Panji anticline. As the space between the anticline and sampling location increases, the Ca²⁺ concentration decreases, while the Na⁺+K⁺ and HCO₃[−] concentrations increase, which suggests that the Permian fractured aquifer may be recharged by the Cenozoic bottom aquifer in areas near the anticline. The Cl[−] concentration spatial map demonstrates a complex distribution pattern. Discrepancies are found in

local areas that may have been affected by the uneven distribution of fractures or by water–rock interaction. As can be seen in the vertical distribution of the coal seam strata (Fig. 1), water sampling points for the aquifer in the southeast tend to be located deeper in the aquifer. Therefore, the high HCO₃[−] concentration and low Ca²⁺ concentration typically found in deep groundwater should be characteristic of the Permian fractured aquifer. Shallow groundwater of this aquifer had the opposite characteristics. The hydrochemistry of the shallow groundwater is similar to that of the Cenozoic bottom and middle aquifers. It is therefore likely that the shallow groundwater is recharged by these two aquifers.

In summary, the spatial distribution maps of the water chemical indicators show that the Cenozoic bottom aquifer and the Permian fractured aquifer could be hydraulically connected near the Panji anticline, which verifies the Piper diagram analysis. The spatial distribution maps also indicate the area where discrimination and classification approaches for identifying groundwater sources based on hydrogeochemistry could be of low accuracy, or even invalid. In addition, through spatial analysis, other hydrochemical indicators could be evaluated as a discrimination parameter for further research. The spatial distribution of the indicators can also provide an intuitive way to identify water sources.

Conclusions

Fractures that develop during the mining process can disturb the relative hydraulic balance of aquifers and lead to water inrush. Once an inrush occurs, the first task is to identify water-inrush sources promptly and correctly. Appropriate

Table 4 The semi-variogram parameters of hydrochemical indicators of the Permian fractured aquifer

Indicator	Semi-variogram model	Nugget (C ₀)	Partial sill (C ₁)	Sill (C ₀ +C ₁)	Range	C ₀ /(C ₀ +C ₁)
Ca ²⁺	Gaussian	57.93	93.16	151.09	841.80	38.34%
Na ⁺ +K ⁺	Exponential	0.00	61594.08	61594.08	1025.22	0
HCO ₃ [−]	Circular	0.00	502482.1	502482.1	829.04	0
Cl [−]	Exponential	36.18	30373.05	30409.23	2089.27	0.12%

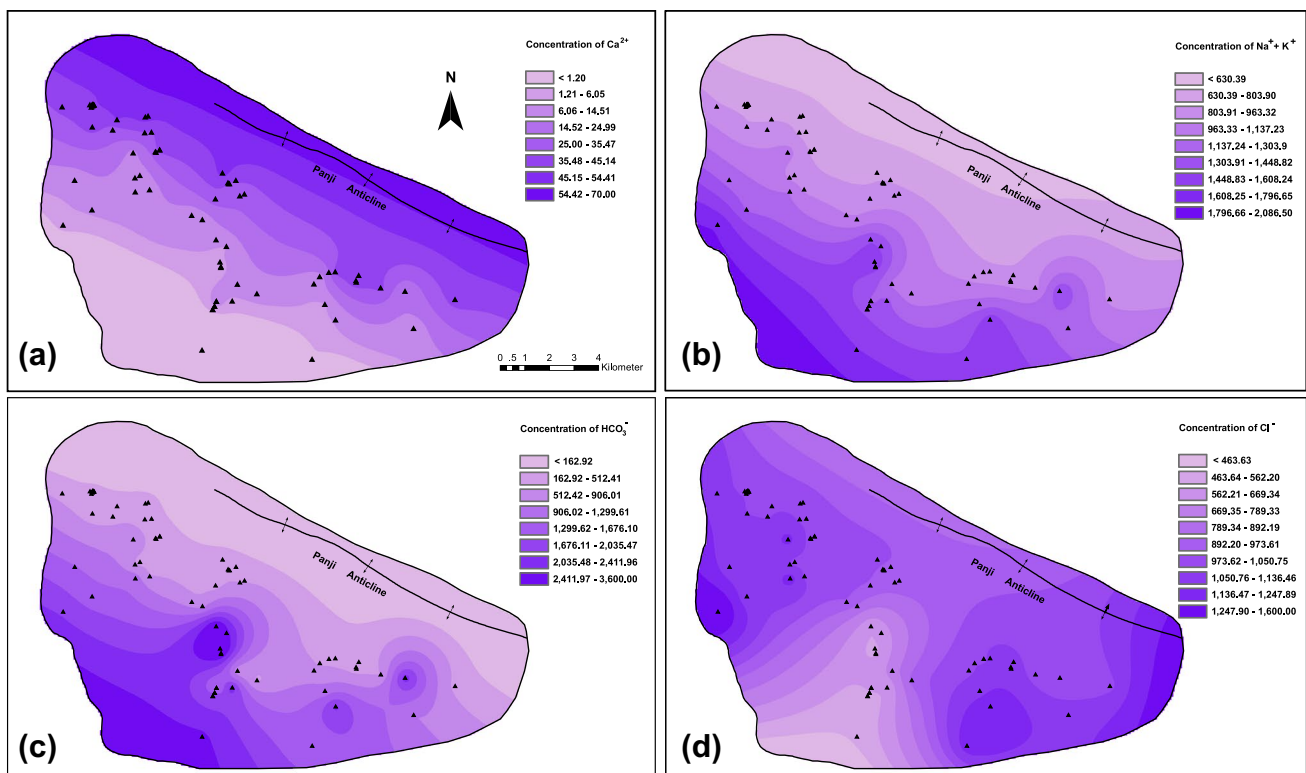


Fig. 4 Spatial distribution of groundwater chemical concentration of the Permian fractured aquifer. The units are mg/L

measures can then be taken to prevent, mitigate, and control major accidents.

To identify intrusion sources, it is necessary to first investigate what distinguishes the various water sources. Hydrochemical analysis was used in this study to understand how groundwater composition changes and to distinguish the differences and connections between different water sources. The conclusions are:

1. The Cenozoic top aquifer water is either HCO_3^- – $\text{Na} + \text{K}$ – Ca or HCO_3^- – $\text{Na} + \text{K}$ – Mg type and can easily be distinguished from water from the other aquifers. The Cenozoic middle and bottom aquifers contain water that is mainly of Cl – $\text{Na} + \text{K}$ and SO_4 – Cl – $\text{Na} + \text{K}$ types. Away from the Panji anticline, water types in the Permian fractured aquifer change gradually from Cl – $\text{Na} + \text{K}$ to HCO_3^- – Cl – $\text{Na} + \text{K}$, and HCO_3^- – $\text{Na} + \text{K}$. The Taiyuan limestone aquifer water are of Cl – $\text{Na} + \text{K}$, SO_4 – Cl – $\text{Na} + \text{K}$, and HCO_3^- – Cl – $\text{Na} + \text{K}$ types.
2. Some of the water samples from aquifers other than the Cenozoic top aquifer are similar in water quality. Based on our hydrochemical analysis of the main aquifers, we constructed a Bayesian discrimination model to potentially discriminate intrusion sources. The precision of the Bayesian model was 86.09%. Because of the hydraulic relationships between the aquifers, the hydrochemical

characteristics of these aquifers were similar around the Panji anticline, leading to erroneous identifications there.

The spatial distribution of water chemistry indicates that the Permian fractured aquifer is potentially recharged by the Cenozoic bottom aquifer near the Panji anticline. Researchers need to be cautious when discriminating water sampled from those areas.

Drawdown of the Permian fractured aquifer, intensified by mining activities and groundwater inflow, will exacerbate water leakage from other aquifers. Groundwater from different aquifers may then be mixed, causing hydrochemical variations. Therefore, further analysis of the dynamic trends, based on the chemical characteristics associated with the hydrogeological conditions, is needed to protect against groundwater damage.

This research may not only enrich the theory of mine groundwater control, but also provide technical support for the practice of disaster prevention and control in the Huainan mining area. The methods used in this paper are suitable for mines with similar geological and hydrogeological conditions in northern China. To extend the method to different regions, appropriate discrimination factors should be selected for the site and its conditions.

Acknowledgements This research was supported by the National Natural Science Foundation of China (Nos. 41602256, 41641021, and 41301537) and the Science and Technology Fund of Land and Resources of Anhui Province (2016-k-11).

References

- Ahmadi SH, Sedghamiz A (2007) Geostatistical analysis of spatial and temporal variations of groundwater level. *Environ Monit Assess* 129(1–3):277–294
- Antunes IMHR, Albuquerque MTD (2013) Using indicator Kriging for the evaluation of arsenic potential contamination in an abandoned mining area (Portugal). *Sci Total Environ* 442(15):545–552
- Babiker IS, Mohamed AAM, Hiyama T (2007) Assessing groundwater quality using GIS. *Water Resour Manag* 21(4):699–715
- Bengranine K, Marhaba TF (2003) Using principal component analysis to monitor spatial and temporal changes in water quality. *J Hazard Mater* 100(1–3):179–195
- Christakos G (2000) *Modern spatiotemporal geostatistics*. Oxford University Press, Oxford
- Dong W, Liu ZH (2008) Study on comprehensive evaluation of water resources carrying capacity based on FAHP principle. *J Arid Land Resour Environ* 10(22):6–10 (Chinese)
- Duan LX, Xie MY, Bai TB, Wang JJ (2016) A new support vector data description method of machinery fault diagnosis with unbalanced datasets. *Expert Syst Appl* 64:239–246
- Gong L, Jin CL (2009) Fuzzy comprehensive evaluation for carrying capacity of regional water resources. *Water Resour Manag* 23(12):2505–2513
- Güler C, Thyne GD, McCray JE, Turner AK (2002) Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeol J* 10(4):455–474
- Guo XD, Fu BJ, Ma KM (2000) Spatial variability of soil nutrients based on geostatistics combined with GIS: a case study in Zunghua city of Hebei province. *Chin J Appl Ecol* 11(4):557–563
- Hobbs BF (1997) Bayesian methods for analyzing climate change and water resources uncertainties. *J Environ Manag* 49(4):53–72
- Isaaks EH, Srivastava RM (1989) *An introduction to applied geostatistics*. Oxford University Press, New York
- Journel AG, Huijbregts CJ (1978) *Mining geostatistics*. Academic Press, San Diego 17: 291–293
- Júnez-Ferreira H, González J, Reyes E, Herrera GS (2016) A geostatistical methodology to evaluate the performance of groundwater quality monitoring networks using a vulnerability index. *Math Geosci* 48(1):25–44
- Kanti KM, Rao PS (2008) Prediction of bead geometry in pulsed GMA welding using back propagation neural network. *J Mater Process Technol* 200(1–3):300–305
- Lee SI (1997) Geostatistical model validation using orthonormal residuals. *KSCE J Civ Eng* 1(1):59–66
- Li S, Wang Z, Shi JB, Peng B (2012) Application of fuzzy mathematics in discriminating sources of water inrush in coal mine. *Safe Coal Mines* 43(7):136–139 (Chinese)
- Li PY, Qian H, Wu JH, Zhang YQ, Zhang HB (2013) Major ion chemistry of shallow groundwater in the Dongsheng coalfield, Ordos Basin, China. *Mine Water Environ* 32(3):195–206
- Li PY, Qian H, Howard KWF, Wu JH, Lyu X (2014) Anthropogenic pollution and variability of manganese in alluvial sediments of the Yellow River, Ningxia, northwest China. *Environ Monit Assess* 186(3):1385–1398
- Li PY, Qian H, Howard KWF, Wu JH (2015a) Building a new and sustainable “Silk Road economic belt”. *Environ Earth Sci* 74(10):7267–7270
- Li PY, Qian H, Howard KWF, Wu JH (2015b) Heavy metal contamination of Yellow River alluvial sediments, northwest China. *Environ Earth Sci* 73(7):3403–3411
- Li PY, Qian H, Zhou WF (2017) Finding harmony between the environment and humanity: an introduction to the thematic issue of the Silk Road. *Environ Earth Sci* 76(3):105. doi:10.1007/s12665-017-6428-9
- Ma R, Shi JS, Liu JC, Gui CL (2014) Combined use of multivariate statistical analysis and hydrochemical analysis for groundwater quality evolution: a case study in North China Plain. *J Earth Sci* 25(3):587–597
- Ma L, Qian JZ, Zhao WD, Curtis Z, Zhang RG (2016) Hydrogeochemical analysis of multiple aquifers in a coal mine based on non-linear PCA and GIS. *Environ Earth Sci* 75(8):1–14
- Matheron BG (1971) *The theory of regionalized variables and its applications*. Les Cahiers du Centre de Morphologie Mathématique, Ecole Nationale Supérieure des Mines, Paris
- McBain J, Timusk M (2011) Feature extraction for novelty detection as applied to fault detection in machinery. *Pattern Recogn Lett* 32(7):1054–1061. doi:10.1016/j.patrec.2011.01.019 (Source: DBLP)
- Nikroo L, Kompani-Zare M, Sepaskhah AR, Shamsi SRF (2010) Groundwater depth and elevation interpolation by kriging methods in Mohr Basin of Fars province in Iran. *Environ Monit Assess* 166(1–4):387–407
- Passarella G, Vurro M, D’Agostino V, Giuliano G, Barcelona MJ (2002) A probabilistic methodology to assess the risk of groundwater quality degradation. *Environ Monit Assess* 79(1):57–74
- Rezaei M, Valipour M, Valipour M (2016) Modelling evapotranspiration to increase the accuracy of the estimations based on the climatic parameters. *Water Conserv Sci Eng* 1(3):1–11
- Roiger RJ, Geatz MW (2003) *Data mining: a tutorial based primer*. Addison-Wesley Publishing Co., Boston
- Schmidt SA (1969) *Measuring uncertainty: an elementary introduction to Bayesian statistics*. Addison-Wesley Publishing Co., Boston
- Shrestha S, Kazama f (2007) Assessment of surface water quality using multivariate statistical technique: a case study of the Fuji river basin, Japan. *Environ Modell Softw* 22(4):464–475
- Stotler RL, Frappe SK, Ruskeeniemä T, Ahonen L, Onstott TC, Hobbs MY (2009) Hydrogeochemistry of groundwater in and below the base of thick permafrost at Lupin, Nunavut, Canada. *J Hydrol* 373(1–2):80–95
- Uddameri V (2007) Bayesian analysis of groundwater quality in a semi-arid coastal county of south Texas. *Environ Geol* 51(6):941–951
- Valipour M (2015) Temperature analysis of reference evapotranspiration models. *Meteorol Appl* 22(3):385–394
- Valipour M (2016a) How much meteorological information is necessary to achieve reliable accuracy for rainfall estimations? *Agric* 6(4):53. doi:10.3390/agriculture6040053
- Valipour M (2016b) Variations of land use and irrigation for next decades under different scenarios. *Irriga* 1:262–288. doi:10.15809/irriga.2016v1n01p262-288
- Valipour M, Banihabib ME, Behbahani SMR (2013) Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J Hydrol* 476:433–441
- Valipour M, Sefidkouhi MAG, Raeini-Sarjaz M (2017) Selecting the best model to estimate potential evapotranspiration with respect to climate change and magnitudes of extreme events. *Agric Water Manag* 180:50–60
- Waltonday K, Poeter E, Thompson A, Vaughan DJ (2009) Investigating hydraulic connections and the origin of water in a mine tunnel using stable isotopes and hydrographs. *Appl Geochem* 24(12):2266–2282

- Wang SQ, Zhu SL, Zhou CH (2001) Characteristics of spatial variability of soil thickness in China. *Geogr Res* 22(3):316–325
- Wang LH, Li GM, Dong YH, Han DM, Zhang JY (2015) Using hydrochemical and isotopic data to determine sources of recharge and groundwater evolution in an arid region: a case study in the upper-middle reaches of the Shule River basin, northwestern China. *Environ Earth Sci* 73(4):1901–1915
- Wu Y, Yu ZC (2011) Application of neural network in water source distinguishing of mine water inrush. *Ind Mine Automat* 60–62 (**Chinese**)
- Wu Q, Zhou WF (2008) Prediction of groundwater inrush into coal mines from aquifers underlying the coal seams in China: vulnerability index method and its construction. *Environ Geol* 56(2):245–254
- Wu J, Li P, Qian H, Duan Z, Zhang X (2014a) Using correlation and multivariate statistical analysis to identify hydrogeochemical processes affecting the major ion chemistry of waters: case study in Laoheba phosphorite mine in Sichuan, China. *Arab J Geosci* 7(10):3973–3982
- Wu WY, Yin SY, Liu HL, Niu Y, Bao Z (2014b) The geostatistic-based spatial distribution variations of soil salts under long-term wastewater irrigation. *Environ Monit Assess* 186(10):6747–6756
- Wunderlin DA, Diaz MP, Ame MV, Pesce SF, Hued AC, Bistoni MA (2001) Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba, Argentina). *Water Res* 35(12):2881–2894
- Xing YF, Zhang C, Wang JH (2016) Analysis and regularity of China's coal mine water inrush accident in 2007 ~ 2014. *Coal Technol* 25(7):186–188 (**Chinese**)
- Yang DD, Ma L, Qian JZ, Chen LW, Cao XC (2010) Water inrush source of Xinzhuangzi coal mine identified with gray related degree method. In: *Proc. 3rd international conf on environmental technology and knowledge transfer, UK*, pp 981–984
- Zhang CL, Qian JZ, Zhao WD, Ma L (2010) The application of Bayesian approach to discrimination of mine water-inrush source. *Coal Geol Explor* 38(4):34–37 (**Chinese**)
- Zhao J, Fu G, Lei K, Li YW (2011) Multivariate analysis of surface water quality in the Three Gorges area of China and implications for water management. *J Environ Sci* 23(9):1460–1471