Research Article

Application of Clustering and Stepwise Discriminant Analysis Based on Hydrochemical Characteristics in Determining the Source of Mine Water Inrush

Weitao Liu, Jie Yu, Jianjun Shen, Qiushuang Zheng, Mengke Han, Yingying Hu, and Xiangxi Meng

1Key Laboratory of Mining Disaster Prevention and Control, Shandong University of Science and Technology, Qingdao 266590, China
2College of Energy and Mining Engineering, Shandong University of Science and Technology, Qingdao 266590, China
3College of Safety and Environmental Engineering, Shandong University of Science and Technology, Qingdao 266590, China
4College of Chemical Engineering and Safety, Binzhou University, Binzhou 256600, China

Correspondence should be addressed to Jianjun Shen; shenjianjun11987@163.com

Received 2 December 2020; Revised 22 April 2021; Accepted 7 May 2021; Published 8 June 2021

Academic Editor: Francesco Frondini

Copyright © 2021 Weitao Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to explore the law of groundwater evolution, the water source connection between faults and aquifers and the main sources of mine water inrush in the deep mining area of Yangcheng Coal Mine in Jining City, 40 groups of hydrochemical samples were collected and analyzed by Piper Diagram and Durov Diagram. The results showed that the fluidity of groundwater developing to the deep became weaker, the value of total dissolved solids (TDS) became larger. So, the roof and floor of coal seam were more similar in water quality types due to the conduction of faults. Using principal component analysis (PCA) to the raw data, two principal components were extracted, and the principal component scores were used as clustering variables for hierarchical cluster analysis (HCA), 5 groups of abnormal water samples were eliminated and 3 clustering groups M1, M2 and M3 were obtained from the other water samples on the tree diagram. The results showed that the combination of HCA and hydrochemical analysis was more effective in screening water samples, and the 3 clustering groups could be qualified samples to represent 3 major aquifers (Taiyuan Formation limestone aquifer, Shanxi Formation sandstone aquifer and Ordovician limestone aquifer). Finally, taking M1, M2 and M3 as grouping variables, the discriminant functions \( f_1, f_2 \) and \( f_3 \) of the 3 aquifers were obtained, the results of stepwise discrimination analysis (SDA) showed that the discrimination model established by using 25 groups of standard water samples could discriminate the known water samples with the correct rate of 96%, 10 groups of unknown water samples collected at the fault are identified as Taiyuan Formation limestone water samples, which was consistent with the classification results of HCA, proving that the water inrush of fault DF53 was from Taiyuan Formation limestone aquifer, while the fault had little influence on Ordovician limestone aquifer.

1. Introduction

The problem of mine water inrush has been a serious limitation of the construction and development of the coal mine, if no measures are taken to prevent and control the water-abundance aquifer at the initial stage of mining, it is easy to conduct the aquifer or generate large fractures without knowing it during the mining process, causing confined water in the water-abundance aquifer to flow into the working face, stopping work or production in light cases, even causing heavy casualties in heavy cases [1]. Therefore, it is of great significance to judge the causes of water inrush in time, find out the source of mine water, and then accurately and efficiently implement measures for each aquifer and water-conducting channel to prevent and control mine water inrush.

In recent years, water source identification methods have developed rapidly. Scholars have gradually developed a single identification method into an interdisciplinary and multi-
theoretical comprehensive method to identify water sources [2–4]. Traditional water quality and hydrochemical analysis have evolved into more advanced techniques for identifying water sources. For example, the isotope method could be used to track the source of mine water and determine the connectivity of groundwater aquifers based on the composition of mine water and the ratio of measured isotope [5, 6]; The theoretical formula of nonlinear mathematics was used to calculate the correlation between sample indicators and data, and the respective discriminant models and evaluation results were obtained, such as grey system theory [7, 8], artificial neural network [9, 10], fuzzy mathematics [11, 12], which have good effects on distinguishing the source of mine water.

Multivariate statistical analysis was often combined with hydrochemical analysis, based on hydrochemical data. Factor analysis and weight were used to deal with multidimensional complex data and reflect the proportion of major ions causing groundwater pollution, the distribution of polluted ions is obtained to evaluate the pollution degree of groundwater [13]; PCA and HCA could also be combined to establish a model for analyzing and evaluating surface water quality changes and influencing factors, thereby identifying the main types of pollution sources [14]; Multivariate statistical analysis could also be used to identify the source of mine water inrush and the connection between aquifers and water-conducting channels, such as Liu et al. [15] who established a water sample database at the -375 m elevation of the Shandong coastal gold mine, used PCA and CA to identify the source of mine water, and established a Bayes discriminant model to test the sample groupings, and finally identified the faults through the workings as potential water-conducting channels. Wang et al. [16] based on the aquifer water sample data of Jiaojia Gold Mine, weighted multidimensional indexes with EWM, removed indexes with lower weights, used PCA and HCA to handle remaining 10 sample indexes, finally proved that the main source of mine water is the fault hanging wall, which provided a reference for mine water prevention. Sun and Gui [17] used PCA, CA and DA to build a model based on the water sample data from Renlou coal mine, which could understand the main aquifer from the perspective of ions, and the source of mine water could be judged by using the concentration value of main ions.

At present, mines entering the stage of deep mining are facing more complex water inrush environment, various geological structures will make different aquifers connect with each other, resulting in similar hydrochemical characteristics of collected water samples and superposition of hydrochemical data information, which makes it extremely difficult to identify water inrush sources.

For this reason, the author puts forward the method of PCA + HCA + SDA to discriminate the source of mine water, which can discriminate the mine water source with complex water source and complex structure, and the discrimination result has high accuracy, the flow chart of research methods is shown in Figure 1. At the same time, the function of HCA tree diagram is not only used for data classification, but also combined with Piper diagram to screen and eliminate data, which makes the screening of data more detailed and the result of discriminant equation more accurate. In this paper, the spatial position relationship between groundwater is fully considered, the direct optimal combination of several multivariate statistical analysis methods is studied, and finally the PCA + HCA + SDA model for distinguishing water inrush sources is proposed. On the basis of improving the correct rate of discrimination, the experimental data can be standardized, and the connection between water sources can be made clear. Provide a new solution for water source identification.

2. Study Area

2.1. Geological Conditions. Yangcheng Coal Mine is located in Guoou Town, Wenshang County, Jining City, Shandong Province, as shown in Figure 2. It belongs to temperate humid monsoon climate, with an annual average temperature of 13.5°C and an annual average precipitation of 664.7 mm, mostly concentrated in June to September. The mining area lies in the alluvial plain of the Yellow River, with flat terrain, high in the south and low in the north, with a ground elevation of +38.80 m to +41.60 m and an area of 42.63 km².

The strata in the mining area are monoclinic structure with a strike of NE and a dip of SE, with a dip angle of 11° to 25° and a local dip angle of 50°. The coal-bearing strata are mainly Carboniferous and Permian, with a total thickness of 257 m and 18 coal-bearing strata. 3 Coal seam is a thick coal seam sandwiched between sandstones of Shanxi Formation and is the main coal seam mined in the mining area, with an average thickness of 7.2 m and an inclination angle of 26° ~ 30°. The No.3 mining area, which mainly mines 3 coal seams, is taken as the main area for water sample collection, and its elevation is -870 m to -1050 m, belonging to deep mining.

Fault structures are well developed in this area. The east, south and north directions are cut off by F1, F2, F3 and F4 major faults and Wensi branch faults, respectively. The main strike of faults can be divided into NE, NW and EW directions, of which NE direction is the most developed. Geophysical exploration shows that there are 124 large and small faults in the area, all of which are normal faults. F1, Jizhuang and Yangchengba faults are large in scale, large in drop and strong in ductility, forming the main body of faults in the area. Folds in this area are distributed in the northern part of the study area, mainly including Houcang syncline, Wanglou anticline, Qulou anticline, etc. The fault structure in the study area is shown in Figure 2.

2.2. Hydrogeological Conditions. The aquifers which have great influence on the mining work of the working face can be divided into three types. (1) The first type is the sandstone aquifer of Shanxi Formation, distributed in the roof and floor of 3 coal seam, it is composed of sandstone, siltstone and mudstone, average thickness of the roof sandstone section is 31.69 m. Geophysical exploration shows that the roof sandstone section has certain static reserves. According to the production requirements, the roof sandstone water in coal seam is drained by drilling holes during mining, and the accumulated drained water reaches 95000 m³. The working face is currently less threatened
Figure 1: Flow chart of research methods.

Figure 2: Location and geological structure of the study area.
by sandstone water, but roof gushing still exists. (2) The second type is Taiyuan Formation limestone aquifer, which is 54 m away from the coal seam floor and contains 8 layers of limestone strata, which are distributed alternately with fine sandstone, medium sandstone, siltstone and mudstone. Among them, 3 layers of limestone have great influence on the working face, which are 49 m~58 m away from the coal seam and generally have a thickness of about 4 m. The water pressure of the 3 layers of limestone water is relatively high. In the horizontal boreholes of -312 m and -650 m in the mine, the water pressure is greater than 1 MPa, some reaching 3.5 MPa, and the maximum water inflow is 45 m³/h. In the initial stage of mining and infrastructure construction, the limestone water of Taiyuan Formation is easy to threaten the working face. (3) The third type is Ordovician limestone aquifer, which is mainly composed of limestone, dolomite and dolomitic limestone. It is 210 m away from the coal seam floor and belongs to the strong aquifer in the region, with a thickness of about 800 m. The unit water inflow of boreholes is 0.275 L/s m⁻¹, and the average permeability coefficient is 0.790 m/d.

The water-resisting layers in the study area are mainly Quaternary clay, mudstone and sandstone of Xiaishihezi Formation and mudstone of Benxi Formation. The water-resisting effect is affected by faults DF53 and FD228. Fault DF53 have a drop of 35 m and cuts through sandstone aquifer, coal seams, Taiyuan Formation limestone aquifer and Ordovician limestone aquifer, which have great influence on the working face of No.3 mining area. Figure 3 shows the relative positional relationship between fault DF53 and 3 main aquifers.

3. Materials and Methods

A total of 30 groups of water samples were collected from 3 main aquifers, including Shanxi Formation sandstone aquifer (I), Taiyuan Formation limestone aquifer (II) and Ordovician limestone aquifer (III). They came from several main mining faces (3308, 3310, etc.) in No. 3 mining area of Yangcheng coal mine, and 10 groups of fault water samples from 3305 working face were collected as water samples to be tested. The collected water samples were sent to the Testing and Analysis Center of Shandong coalfield geology bureau of the Fifth Exploration Team for chemical testing and a water quality testing report was obtained, in which the main ions Na⁺, K⁺, Ca²⁺, Mg²⁺, Cl⁻, SO₄²⁻ and the total hardness (TH) of the water were determined by ion chromatography. HCO₃⁻ and the total basicity (TA) of water was determined by titration with dilute sulfuric acid-methyl orange. Total dissolved solids (TDS) were obtained by filtering, drying and weighing water samples. The pH value was determined by pH tester. Finally, the ten discriminant index data of mine water were sorted and drawn into a table, as shown in Table 1.

3.1. Hydrogeochemistry. The groundwater of different aquifers usually shows differences in water chemical composition, so each aquifer has different water chemical characteristics [18, 19]. Piper Diagram and Durov Diagram were commonly used in water chemical analysis. The ion composition and proportion of each water sample could be read out on Piper Diagram.

The relative contents of cations and anions in the water sample could be read out on the lower two isosceles triangles, respectively. The upper rhombus part was used to read the total ion proportion and chemical properties of the water sample. Water samples in each aquifer will be concentrated in the rhombus part of Piper diagram, while individual water samples with composition differences would deviate from most water samples. These non-standard water samples would be eliminated, but most "standard" water samples conforming to the composition characteristics of this aquifer would be found by screening water samples in this way [20, 21].

3.2. Principal Component Analysis. PCA is a basic multivariate statistical analysis method. Its main idea is to construct a linear combination of raw variables and compress multiple groups of complex variables into simple comprehensive variables to achieve the effects of compressing data and extracting main information [20–24]. In order to avoid the influence of different data units on the calculation results, the raw data was standardized by z-scores firstly, n samples were \( x_i = (x_{i1}, x_{i2}, \cdots, x_{ip}) \), and the formula was as follows [25, 26]:

\[
\begin{align*}
Z_{ij} & = \frac{x_{ij} - \bar{x}_j}{s_j} \\
\bar{x}_j & = \frac{\sum_{i=1}^{n} x_{ij}}{n} \\
s_j^2 & = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n - 1}
\end{align*}
\]
Variables could be represented by several principal components. Considered until most of the information about the raw variables, the second principal component was not enough to represent the information much variable information as possible, when the variance of \( j \) column elements of the raw data, respectively.

\[
Z_{ij} = \text{the standardized data, and } \bar{x}_j \text{ and } s^2_j \text{ are the mean and variance of } j \text{ column elements of the raw data, respectively.}
\]

Secondly, principal components were used to reflect as much variable information as possible, when the first principal component was not enough to represent the information of the raw variables, the second principal component was considered until most of the information about the raw variables could be represented by several principal components.

According to this idea, the relationship between PCA and raw variables was [27, 28]:

\[
\begin{align*}
F_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
F_2 &= a_{21}F_1 + a_{22}F_2 + \cdots + a_{2p}F_p \\
&\quad \vdots \quad \vdots \\
F_p &= a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pp}F_p
\end{align*}
\]
On SPSS, most of the information of the raw variables needs to be expressed by the characteristic root of the principal component being greater than 1 or the cumulative contribution rate of the principal component expressing the original information being greater than 85%. SPSS is an analysis software introduced by IBM in 1984 to carry out statistical analysis and operation on data [15].

### 3.3. Hierarchical Cluster Analysis

Hierarchical Cluster Analysis (HCA) is a process of clustering samples with a higher degree of similarity into one category based on the features between them, and then re-aggregating the aggregated sub-categories according to the degree of similarity, and finally all the sub-categories are aggregated into one large category, a process that can be represented as a tree diagram where it can be clearly seen which samples are more similar to each other [29–31].

HCA defines p variables of water samples as a point in p-dimensional space, so the similarity can be expressed by the distance between classes and samples. Using different clustering methods, the clustering results may also be different, such as Single link method, Complete link method, Between-groups link method and Ward’s method. The smaller the distance, the higher the similarity between the two categories. The method used in this paper was the Between-groups link method, let the two groups be $G_p$ and $G_q$, and the expression of the distance between the two classes was as follows [4, 29]:

$$D_{pq}^2 = \frac{1}{n_p n_q} \sum_{x_i \in G_p} \sum_{x_j \in G_q} d_{ij}^2$$ (3)

Where $x_i$ and $x_j$ are the i and j samples of $G_p$ and $G_q$; $d_{ij}$ is the distance between samples of $G_p$ and $G_q$; $n_p$ and $n_q$ are the number of samples contained in the two groups.

The distance $d_{ij}$ between samples used the Squared Euclidean distance formula, which was:

$$d_{ij}^2 = \sum_{k=1}^{p} (x_{ik} - x_{jk})^2$$ (4)

Where $x_{ik}$ and $x_{jk}$ represent the k-th variable of samples $x_i$ and $x_j$.

### 3.4. Stepwise Discriminant Analysis

In a case where most of that water sample category is known, we need to introduce a discriminant function if we want to determine the type of unknown water sample or get the correct rate of discrimination. There are many variables, if all variables are introduced into the discriminant function, the calculation process will be complicated [32]. This article adopted the stepwise discriminant analysis (SDA), which would select the variables with a greater contribution to the discriminant result to introduce the discriminant function, which could greatly reduce the amount of calculation. Using Wilks’ Lambda statistics as the principle of stepwise discrimination, we selected variables that can minimize the Wilks’ Lambda of the population to enter the discriminant function for each step in the operation process. In addition, in the discrimination process, we used the size of the $F$ value to keep or delete variables. After $p$ deletions, $r$ variables were finally selected and used as variables of Bayesian discriminant function. Then, substituted the ion concentration into several discriminant functions to obtain the function value, and found the category with the largest function value, which was the final grouping of the sample [33–36].

### 4. Results and Discussion

#### 4.1. Hydrochemical Analysis

Figures 4(a), 4(b) and 4(c) showed the relative proportion of various ions in water samples and the actual concentration values of various ions and the concentration changes in 40 water samples. Combined with Figures 4(a) to 4(c) for hydrochemical analysis, the evolution process of ions in groundwater could be more clearly shown.

By observing Figures 4(a) and 4(c), it could be seen that the main water quality of Shanxi Formation sandstone water was HCO3·Cl-Na type, and HCO3− accounts for 30–40% of Shanxi Formation sandstone water. The formation of HCO3− was initially due to the hydrolysis of potassium feldspar and albite in the coal measures strata in the deep strata, which increased the Na++K+ content in the groundwater. However, Na+ and Ca2+ can undergo reduction reaction in an anoxic environment under the formation, and then SO42− could be reduced to H2S through the action of bacteria, thus increasing the proportion of HCO3−. The chemical formula of the chemical reaction is as follows:

$$2Na^+ + CaSO_4 \rightarrow Ca^{2+} + Na_2SO_4$$ (5)

$$Na_2SO_4 + 2C + 2H_2O \xrightarrow{\text{bacterial}} 2NaHCO_3 + H_2S\uparrow$$

On Figure 4(a), the points of cations in the sandstone water (I) of Shanxi Formation were concentrated, anions were dispersed, because Na+K+ accounted for more than 80% of total cations. In anions, the concentration of HCO3− and Cl− in anions had a converse changing trend: HCO3− accounts for 30–40% in water samples No.1-6, while the proportion in sandstone water samples 6-10 decreased, HCO3− accounts for 10–20%, and samples No.14-15 were also collected from aquifer I, in which HCO3− accounted for only 3–4%. But the proportion and concentration of Cl− were both increasing. Obviously, the characteristic ion HCO3− of sandstone water decreases, and the concentration of Cl− and SO42− rose sharply, the water sample type of Shanxi Formation sandstone changed from HCO3·Cl-Na type water to Cl·SO4-Na type water.

On Figure 4(b), it could be seen that the TDS of sandstone water samples were low, within 0–4000 mg/L, it belonged to medium soluble solid mine water (according to the classification standard of coal mine water promulgated by China in 2015, mine water with soluble solid content between 1000 mg/L and 6000 mg/L belonged to medium soluble solid mine water, and mine water with soluble solid content >6000 mg/L belonged to high soluble solid mine water [37]). The change of anion SO42− could be analyzed...
Figure 4: Piper Diagram (a), Durov Diagram (b) and Percentage Stacked Bar Chart (c) of 40 water samples.
from the runoff change of sandstone water, the mining disturbance caused by coal seam mining also produced fractures and fissures in the surrounding rock on the upper wall of the coal seam. Sandstone water could flow in the fissures, which accelerated the runoff of groundwater, and sulfides in coal strata undergo oxidation reaction when dissolved oxygen in water increases. The following is the chemical equation:

\[
2\text{FeS}_2 + 7\text{O}_2 + 2\text{H}_2\text{O} \rightarrow 2\text{FeSO}_4 + 2\text{H}_2\text{SO}_4
\]

\[
4\text{FeSO}_4 + 2\text{H}_2\text{SO}_4 + \text{O}_2 \rightarrow 2\text{Fe}_2(\text{SO}_4)_3 + 2\text{H}_2\text{O}
\]

\[
\text{Fe}_2(\text{SO}_4)_3 + 2\text{H}_2\text{O} \rightarrow 2\text{Fe}(	ext{OH})_3 + 3\text{H}_2\text{SO}_4
\]

Therefore, the concentration of \(\text{SO}_4^{2-}\) in groundwater increased, from less than 7% to about 15%, and the flowing mine water also dissolved \(\text{Cl}^-\) containing minerals, which made the concentration of \(\text{Cl}^-\) increased, from less than 20% to more than 30%, and the TDS of water samples had a gradual upward trend.

The anions and cations in the limestone water of Taiyuan Formation were mainly \(\text{Cl}^-\) and \(\text{Na}^+ + \text{K}^+\), it could be seen from Figure 4(c) that the proportion of \(\text{Na}^+ + \text{K}^+\) was more than 30% of the total ion content, and the content of \(\text{Cl}^-\) was more than 55% of the total ion content. The limestone water of Taiyuan Formation was a typical chloride water, and sodium chloride occupied a large proportion in the groundwater, and the water quality was Cl-Na type. The TDS of aquifer II water samples were between 3000 mg/L and 8000 mg/L, and the mine water could be divided into medium soluble solid mine water and high soluble solid mine water according to soluble solids. As could be seen from Figure 4(a), There was a phenomenon of overlapping water samples between aquifer II water samples, aquifer I water samples and unknown water samples, and the TDS or other ion contents were not completely consistent. Considering that mining operations might affect Taiyuan Formation limestone, it was speculated that there might be some channels that made Taiyuan Formation limestone water affect other aquifers.
According to Figure 4(a), water sample No.5 should be excluded. The Ordovician limestone water sample contained a lot of \( \text{SO}_4^{2-} \) and \( \text{Ca}^{2+} \) (\( \text{Ca}^{2+} \) accounted for 15% of the total ion content, and \( \text{SO}_4^{2-} \) accounted for more than 25% of the total ion content), which was an important feature of Ordovician limestone aquifer and the main difference between it and the other two aquifers (I and II). \( \text{SO}_4^{2-} \) and \( \text{Ca}^{2+} \) and \( \text{Mg}^{2+} \) originated from the dissolution of gypsum, dolomite and other halides and sulphates of calcium and magnesium in underground strata. Similarly, \( \text{Cl}^- \) and \( \text{Na}^+ + \text{K}^+ \) accounted for a higher proportion in Ordovician limestone water, and these ions came from halides of potassium and sodium which were more easily dissolved. The content of these ions increased, which made the content of TDS very high, and the TDS of No.27-28 water samples even exceeded 10000 mg/L. The III aquifer had a large buried depth, was poorly recharged by other water sources, and had a strong interaction between groundwater and rocks. The influence of DF53 fault on the third aquifer was not significant.

4.2 Spatial Distribution Characteristics of Hydrochemical Elements. Figures 5(a)–5(g) compared and analyzed various elements in aquifer and fault water of three different groundwater heads. Since uncertain water samples were obtained by drilling holes near faults, No.31-40 water samples were used to represent fault water samples, which could intuitively reflect the evolution characteristics of groundwater and the connection between aquifer and faults.

**Table 2: Correlation Coefficient Matrix between Variables.**

|          | \( \text{Na}^+ + \text{K}^+ \) | \( \text{Ca}^{2+} \) | \( \text{Mg}^{2+} \) | \( \text{Cl}^- \) | \( \text{SO}_4^{2-} \) | \( \text{HCO}_3^- \) | \( \text{TH} \) | \( \text{TA} \) | \( \text{TDS} \) | \( \text{PH} \) |
|----------|-------------------------------|-----------------------|-----------------------|-------------------|-------------------|-------------------|----------------|----------------|----------------|----------------|
| \( \text{Na}^+ + \text{K}^+ \) | 1.000                         |                       |                       |                   |                   |                   |                |                |                |                |
| \( \text{Ca}^{2+} \)       | 0.379                         | 1.000                 |                       |                   |                   |                   |                |                |                |                |
| \( \text{Mg}^{2+} \)       | 0.610                         | 0.791                 | 1.000                 |                   |                   |                   |                |                |                |                |
| \( \text{Cl}^- \)          | 0.954                         | 0.526                 | 0.760                 | 1.000             |                   |                   |                |                |                |                |
| \( \text{SO}_4^{2-} \)     | 0.242                         | 0.861                 | 0.628                 | 0.289             | 1.000             |                   |                |                |                |                |
| \( \text{HCO}_3^- \)       | -0.397                        | 0.697                 | 0.735                 | 0.533             | -0.610            | 1.000             |                |                |                |                |
| \( \text{TH} \)            | 0.479                         | 0.977                 | 0.904                 | 0.634             | 0.823             | -0.746            | 1.000           |                |                |                |
| \( \text{TA} \)            | -0.451                        | 0.685                 | 0.738                 | 0.582             | 0.585             | 0.988             | -0.739         | 1.000           |                |                |
| \( \text{TDS} \)           | 0.903                         | 0.717                 | 0.846                 | 0.951             | 0.567             | -0.634            | 0.798          | -0.668         | 1.000           |                |
| \( \text{PH} \)            | -0.659                        | 0.543                 | 0.723                 | -0.732            | -0.350            | 0.623             | -0.634         | 0.678          | -0.733         | 1.000           |
Firstly, with the increase of water head depth, the content of various ions had obvious changes, and the law of change was not completely consistent. It could be seen that Na\(^{+}\)+K\(^{+}\), Ca\(^{2+}\), Mg\(^{2+}\), Cl\(^{-}\), SO\(_4\)\(^{2-}\) all had obvious increases, resulting in a significant increase in TDS as depth increases. In deep mines, the difference of TDS in groundwater was often related to the runoff velocity of groundwater. Under the action of coal mining, the roof of coal seam was affected by fissures and boreholes, and groundwater runoff increases, which was an important factor affecting groundwater level and flow velocity. Secondly, from the changes of these ions, we could also see some characteristics of groundwater under different water heads, for example, the concentration of SO\(_4\)\(^{2-}\) was very low at -960 m water head, while the concentration increased significantly at -1120 m water head; The contents of Ca\(^{2+}\) and Mg\(^{2+}\) at -1120 m water head increased significantly, and the proportion was larger than that at other water heads; The proportion of Na\(^{+}\)+K\(^{+}\) and Cl\(^{-}\) in mine groundwater was relatively high, and their variation laws were similar, indicating that sodium chloride was one of the main components of mine groundwater. From the change of boxplots, it could be seen that some indicators were also significantly correlated, for example, with the increase of depth, the values of pH and TA both had a downward trend, because they were both related to the concentration of HCO\(_3\)\(^{-}\), HCO\(_3\)\(^{-}\) became weakly alkaline in water. With the increase of depth, the concentration of HCO\(_3\)\(^{-}\) decreases, causing the alkalinity to decrease, and pH naturally decreases, finally approaching 7, indicating that mine water is weakly alkaline.

Figures 6(a)–6(d) showed the spatial distribution characteristics of Ca\(^{2+}\), TDS, Na\(^{+}\)+K\(^{+}\) and HCO\(_3\)\(^{-}\) in the mining area. The high concentration value of Ca\(^{2+}\) was distributed in the southeast of the mining area, and the concentration value was quite different between the south and the north. The Ca\(^{2+}\) near the Yangchengba fault and Jizhuang fault was generally low, both below 660 mg/L, because the groundwater near the fault could flow freely, and Ca\(^{2+}\) in the water was easier to react with HCO\(_3\)\(^{-}\) and CO\(_2\)\(^{2-}\) to precipitate, while Ca\(^{2+}\) could be preserved in the closed groundwater with small runoff conditions. The high concentration of TDS appeared in the south of the mining area. The fault in the north of the mining area affected the runoff change of groundwater. The TDS near the fault was relatively low. For example, the TDS in most areas near the Yangchengba fault was in the range of 500–2800 mg/L, and the TDS in most areas near the Jizhuang fault was in the range of 2000–4100 mg/L, only a few areas had a sudden increase in TDS, but the range is not wide. By comparing Figures 6(b) and 6(c), it could be found that the distribution law of Na\(^{+}\)+K\(^{+}\) and TDS was similar, they all had a minimum concentration near the Yangchengba fault. The high concentration was concentrated in the south and northeast of the mining area, which was closely related to the distribution of the fault, it could be seen that the conduction state of the fault affected the groundwater flow and also affected the ion concentration distribution in the groundwater. The concentration of HCO\(_3\)\(^{-}\) was contrary to the distribution law of other ions. The highest concentration of HCO\(_3\)\(^{-}\) existed near the Yangchengba fault in the north of the mining area, while the concentration of HCO\(_3\)\(^{-}\) in the south of the mining area was getting lower and lower. This was due to the high concentration of CO\(_3\)\(^{-}\) and CO\(_2\)\(^{2-}\) in the deep stratum, and some cations would react with HCO\(_3\)\(^{-}\) and precipitate.

After the above research, it can directly reflect that the ion concentration change in the mining area not only has obvious distribution law in depth, there are also certain variation characteristics in the spatial range, this change is closely related to the geological structure in the region. The existence of faults and fissures will enable groundwater to flow freely, while runoff will change the ion concentration in groundwater and promote the interaction between water and rocks. Therefore, it is generally scientific and reliable to use the hydrochemical characteristics of groundwater to reflect the internal relations between aquifers.

4.3. Multivariate Statistical Analysis. The Kaiser-Meyer-Olkin value was 0.731 of the data in Table 1, and the value of Bartlett test was 0 (less than 0.05), there was correlation between the raw data, and the data analysis was easily affected by potential factors [38], so PCA was suitable before data analysis. In Table 2, the correlation coefficient between variables was generally large (>0.7 was marked in bold), and the correlation between variables was high. PCA can eliminate the correlation between variables [39, 40]. As can be seen from Table 2, the correlation coefficients of Na\(^{+}\)+K\(^{+}\) and Cl\(^{-}\) are 0.954, Ca\(^{2+}\) and SO\(_4\)\(^{2-}\) are 0.861, which have high positive correlation, while HCO\(_3\)\(^{-}\) is negatively correlated with other ions, indicating that HCO\(_3\)\(^{-}\) gradually decreases with the increase of other ions in water samples, which is consistent with the results of hydrochemical analysis.

PCA was performed on the original data in SPSS, and the sum of two principal components (see Table 3) was finally selected, their cumulative variance contribution rate was 85.9%, and their eigenvalues were 7.076 and 1.514, respectively, meeting the conditions that the eigenvalue was greater than 1 and the cumulative variance contribution rate was greater than 85%. F\(_1\) and F\(_2\) respectively, reflected the variance contribution rate of 70.762% and 15.138% of the original information. Finally, the obtained principal component was a linear combination of the raw 10 variables, and the score formula of the principal component was as follows:

\[
\begin{align*}
F_1 &= 0.102X_1 + 0.121X_2 + 0.131X_3 + 0.117X_4 + 0.100X_5 - 0.118X_6 + 0.131X_7 - 0.120X_8 + 0.132X_9 - 0.112X_{10} \\
F_2 &= 0.426X_1 - 0.261X_2 - 0.013X_3 + 0.354X_4 - 0.368X_5 + 0.173X_6 - 0.187X_7 + 0.126X_8 + 0.195X_9 - 0.176X_{10}
\end{align*}
\]
The principal component score was calculated by the principal component score formula, and the principal component score was saved as variables on SPSS as the basic data of HCA, and Q-type clustering analysis of HCA was carried out. Q-type clustering analysis is based on comprehensive comparison of different parameters between samples, grouping samples, and showing the relationship between samples in this way [41]. Except for the No.4 and No.5 water samples eliminated by water chemical analysis, the tree diagram of the remaining 38 groups of water samples was shown in Figure 7.

On Figure 7, the connection distance (X axis) between groups is defined as \(D_{\text{link}}/D_{\text{max}}\) the quotient of the connection distance and the maximum distance to represent the distance between samples [38, 42]. When \(D_{\text{link}}/D_{\text{max}} \leq 15\), each water sample data can be clearly divided into 3 groups, named M1, M2 and M3, respectively, from top to bottom. From the results of tree diagram grouping, it can be seen that M1, M2 and M3 correspond to aquifers II, I and III, respectively, and most water samples could be correctly divided into the corresponding aquifers, which indicated that HCA could reliably group the original water samples.

The 3 water samples of 11, 14 and 15 were different from the original grouping (marked with red rectangle in Figure 7) and were eliminated from the database. Water sample 11 had a relatively low \(\text{SO}_4^{2-}\) concentration, and \(\text{HCO}_3^-\) and Cl\(^-\) concentrations were similar to those of sandstone water, so water sample 11 was wrongly classified as I from II. Water samples 14 and 15 belonged to aquifer I, they had higher Cl\(^-\) and Na\(^+\) and \(\text{SO}_4^{2-}\), so water samples 14 and 15 were wrongly classified as aquifer II. It could be inferred that there was a hydraulic connection between the sandstone aquifer of Shanxi Formation and the limestone aquifer of Taiyuan Formation, therefore, the water samples of the two aquifers misjudged each other in a large number. From the ion level, Taiyuan Formation limestone water still exists, and a large amount of water gushing phenomenon. However, the influence of mining disturbance, the threat of water gushing from Ordovician limestone still exists, and waterproof coal pillars still need to be retained near faults to ensure mine safety.

The correct 25 groups of water samples were introduced into the discriminant function, according to the requirement of normality test, Wilks’ Lambda value was used to select variables, when the \(F\) value of the variable was greater than the specified “Enter” value of 3.84, the variable was retained, and when the \(F\) value of the variable was less than the specified “Delete” value of 2.71, the variable was deleted. Finally, 3 variables satisfying the conditions were retained, including Na\(^+\) + K\(^+\), Ca\(^{2+}\) and TA. The value of TA was highly correlated with the content of \(\text{HCO}_3^-\). In this paper, the value of TA was used to introduce the discriminant function, if without measuring TA, the concentration of \(\text{HCO}_3^-\) could also be introduced, which had little influence on the discriminant result of stepwise discriminant analysis. Taking HCA grouping M1, M2 and M3 as grouping variables, the discriminant functions of the 3 aquifers established by stepwise discriminant analysis were as follows:

\[
\begin{align*}
    f_1 &= 0.015w(\text{Na}^+ + \text{K}^+) - 0.025w(\text{Ca}^{2+}) + 0.012w(\text{TA}) - 14.286 \\
    f_2 &= -0.002w(\text{Na}^+ + \text{K}^+) + 0.011w(\text{Ca}^{2+}) + 0.063w(\text{TA}) - 13.101 \\
    f_3 &= -0.018w(\text{Na}^+ + \text{K}^+) + 0.072w(\text{Ca}^{2+}) + 0.070w(\text{TA}) - 26.365
\end{align*}
\]

In the formula, \(f_1\), \(f_2\) and \(f_3\), respectively, represent the discriminant functions of the 3 groups M1, M2 and M3, corresponding to aquifers II, I and III; \(w(\cdot)\) is that it corresponds ion concentration value.

The model was used to distinguish 25 groups of water samples, of which No.9 water sample from sandstone aquifer was misjudged as Taiyuan Formation limestone water, and the rest water samples were all accurately distinguished, with a discrimination accuracy rate of 96%. The three ion concentrations of uncertain water samples were substituted into the formula to obtain three function values \(f_1\), \(f_2\) and \(f_3\). The function with the largest function value is the grouping of water samples that were finally discriminated. Comparing the 3 groups of function values, it was found that \(f_1\) value was the largest, and 10 groups of water samples to be tested were classified as M1 by discriminant analysis (see Table 4).

The scatter graph of the canonical discriminant function (see Figure 8) was easier to show the relationship between the 3 types of water samples. Finally, the water sample to be tested was identified as limestone water of Taiyuan Formation. This result was mutually verified with the tree diagram result of HCA. On Figure 8, the water sample to be tested was closer to limestone water sample of Taiyuan Formation, and the No.9 water sample from M2 group was closer to the center point of M1, which eventually could lead to misjudgment. Through the above discrimination results, it was proved that DF53 fault was the water conduction channel of Taiyuan Formation limestone aquifer, which made the sandstone aquifer of Shanxi Formation and the water source of Taiyuan Formation limestone aquifer connected. The mixed water sources of the two aquifers made the TDS of sandstone water increase and the ionic composition change.

| Principle component | Eigenvalues | Variance | Cumulative variance % |
|---------------------|-------------|----------|-----------------------|
| 1                   | 7.076       | 70.762   | 70.762                |
| 2                   | 1.514       | 15.138   | 85.900                |

The correct 25 groups of water samples were introduced into the discriminant function, according to the requirement of normality test, Wilks’ Lambda value was used to select variables, when the \(F\) value of the variable was greater than the specified “Enter” value of 3.84, the variable was retained, and when the \(F\) value of the variable was less than the specified “Delete” value of 2.71, the variable was deleted. Finally, 3 variables satisfying the conditions were retained, including Na\(^+\) + K\(^+\), Ca\(^{2+}\) and TA. The value of TA was highly correlated with the content of \(\text{HCO}_3^-\). In this paper, the value of TA was used to introduce the discriminant function, if without measuring TA, the concentration of \(\text{HCO}_3^-\) could also be introduced, which had little influence on the discriminant result of stepwise discriminant analysis. Taking HCA grouping M1, M2 and M3 as grouping variables, the discriminant functions of the 3 aquifers established by stepwise discriminant analysis were as follows:

\[
\begin{align*}
    f_1 &= 0.015w(\text{Na}^+ + \text{K}^+) - 0.025w(\text{Ca}^{2+}) + 0.012w(\text{TA}) - 14.286 \\
    f_2 &= -0.002w(\text{Na}^+ + \text{K}^+) + 0.011w(\text{Ca}^{2+}) + 0.063w(\text{TA}) - 13.101 \\
    f_3 &= -0.018w(\text{Na}^+ + \text{K}^+) + 0.072w(\text{Ca}^{2+}) + 0.070w(\text{TA}) - 26.365
\end{align*}
\]

In the formula, \(f_1\), \(f_2\) and \(f_3\), respectively, represent the discriminant functions of the 3 groups M1, M2 and M3, corresponding to aquifers II, I and III; \(w(\cdot)\) is that it corresponds ion concentration value.

The model was used to distinguish 25 groups of water samples, of which No.9 water sample from sandstone aquifer was misjudged as Taiyuan Formation limestone water, and the rest water samples were all accurately distinguished, with a discrimination accuracy rate of 96%. The three ion concentrations of uncertain water samples were substituted into the formula to obtain three function values \(f_1\), \(f_2\) and \(f_3\). The function with the largest function value is the grouping of water samples that were finally discriminated. Comparing the 3 groups of function values, it was found that \(f_1\) value was the largest, and 10 groups of water samples to be tested were classified as M1 by discriminant analysis (see Table 4).

The scatter graph of the canonical discriminant function (see Figure 8) was easier to show the relationship between the 3 types of water samples. Finally, the water sample to be tested was identified as limestone water of Taiyuan Formation. This result was mutually verified with the tree diagram result of HCA. On Figure 8, the water sample to be tested was closer to limestone water sample of Taiyuan Formation, and the No.9 water sample from M2 group was closer to the center point of M1, which eventually could lead to misjudgment. Through the above discrimination results, it was proved that DF53 fault was the water conduction channel of Taiyuan Formation limestone aquifer, which made the sandstone aquifer of Shanxi Formation and the water source of Taiyuan Formation limestone aquifer connected. The mixed water sources of the two aquifers made the TDS of sandstone water increase and the ionic composition change.
The Ordovician limestone aquifer had little connection with other aquifers, and there was no misjudgment of water sources, which indicated that the Ordovician limestone water was relatively independent and less affected by faults.

Through the above analysis, it could be understood that the limestone water of Taiyuan Formation showed water gushing under the conduction of fault DF53, which was due to the expansion of rock fissures caused by mining, which made the originally closed aquifer became active and caused ions exchange in groundwater. The dynamic changes of groundwater were remarkable, the limestone of Taiyuan Formation and the sandstone of Shanxi Formation were located on the upper wall and the lower wall of the coal seam, respectively, under mining disturbance, Taiyuan Formation limestone water gushed out to the working face along fault DF53. As Taiyuan Formation limestone water contained a large amount of Na\(^+\) K\(^+\) and Cl\(^-\), these ions also affected the quality of sandstone water in the upper wall. Sandstone water with low TDS and high HCO\(_3\)\(^-\) content was mixed with Taiyuan Formation limestone water, which increased Cl\(^-\) concentration and TDS of mixed sandstone water samples. Combined with hydrochemical analysis and multivariate statistical analysis, it could be understood that faults and fractures were the main influencing factors of groundwater

![Hierarchical cluster analysis tree diagram.](image-url)
activity in Yangcheng Coal Mine, and groundwater movement might be aggravated by the disturbance of mining work.

5. Conclusions

Based on 40 groups of water samples collected from the working face in Yangcheng Coal Mine, this paper studied the evolution law of groundwater, analyzed the water quality characteristics of each aquifer, and screened the water samples by using hydrochemical analysis. The 10 variables were compressed by PCA, and the principal component scores were substituted into HCA to obtain the tree diagram classification results. Then the water samples were screened again, and 3 groups M1, M2 and M3 were obtained, representing Taiyuan Formation limestone water, Shanxi Formation sandstone water and Ordovician limestone water, respectively. Finally, M1, M2 and M3 were taken as grouping variables and the remaining water samples were taken as variables for stepwise discriminant analysis, and discriminant functions $f_1, f_2$ and $f_3$ were obtained. The discriminant results were compared with the tree diagram for mutual verification. The following conclusions were obtained by combining various analysis methods:

1. With the development of groundwater in Yangcheng Coal Mine to the deep, the fluidity would become weaker and TDS would increase significantly. Sandstone water of Shanxi Formation was affected by fault DF53, which could be related to limestone water of Taiyuan Formation. Na$^+$, Cl$^-$ and TDS in groundwater would increase obviously, making there were more misjudged water samples in the two aquifers.

2. Combining hydrochemical analysis and HCA tree diagram to screen water samples is more accurate than using Piper diagram only to screen water samples, and the accuracy rate of the obtained results is higher.

3. SDA retained 3 variables (Na$^+$, K$^+$, Ca$^{2+}$, TA) that had great influence on the results and introduced discriminant function, which could correctly discriminate 96% of the known water samples. Ten groups of unknown water samples were classified into Taiyuan Formation limestone water samples and verified with HCA results, proving that there was a potential connection between fault DF53 and limestone aquifer of Taiyuan Formation, while fault had little influence on Ordovician limestone aquifer.

4. The combination of PCA and HCA and SDA based on hydrochemical chemical characteristics could identify water sources in mines with fast identification speed and high accuracy, which could meet the actual requirements of mine engineering and has a guiding significance for the identification of mine water inrush.

Data Availability

The data used in this article comes from the comprehensive evaluation report on the prevention and control of water hazards in the Yangcheng Coal Mine face of Shandong Jinan Luneng Coal and Electricity Co., Ltd. I promise that the data used is true and reliable without any modification and comes entirely from the real data collected from the mine face.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant Nos. 41807211, 51874192, 51904032, 42007172), the Natural Science Foundation of Shandong Province (Grant No. ZR2019MEM084), the SDUST Research Fund (Grant No. 2018TDH102), the Open Fund of Key Laboratory of Mine Disaster Prevention and
Control (Grant No. MDPC201920) and the State Key Research and Development Program of China (Grant No. 2017YFC0804108).

References

[1] W. T. Liu, Mine Water Disaster and Prevention, Coal Industry Press, Beijing, 2016.

[2] H. J. Yang and G. C. Wang, "Summarization of methods of distinguishing sources and forecasting inflow of water inrush in coal mines," Coal Geology & Exploration, vol. 40, no. 3, pp. 48–55, 2012.

[3] Z. L. Cui, B. S. Huang, and X. Q. Long, "Research status and development trend of water source identification of mine water inrush," China Tungsten Industry, vol. 33, no. 4, pp. 42–50, 2018.

[4] Q. Liu, Y. Sun, Z. Xu, and G. Xu, "Application of the comprehensive identification model in analyzing the source of water inrush," Arabian Journal of Geosciences, vol. 11, no. 9, p. 189, 2018.

[5] C. Liu, B. Peng, and J. Qin, "Geological analysis and numerical modeling of mine discharges for the Sanshandoa gold mine in Shandong, China: 1. Geological analysis," Mine Water & the Environment, vol. 26, no. 3, pp. 160–165, 2007.

[6] H. Dou, Z. Ma, H. Cao, F. Liu, W. Hu, and T. Li, "Application of isotopic and hydro-geochemical methods in identifying sources of mine inrushwater," Mining Science and Technology, vol. 21, no. 3, pp. 319–323, 2011.

[7] L. Xiao, S. H. Tang, C. L. Zhao, T. X. Yuan, and W. Yang, "Grey-risk estimation of the water inrush from no. 9 coal floor in Guoerzhuang mine," Applied Mechanics and Materials, vol. 295-298, pp. 3019–3022, 2013.

[8] M. Qiu, L. Shi, C. Teng, and Y. Zhou, "Assessment of water inrush risk using the fuzzy Delphi analytic hierarchy process and Grey relational analysis in the Liangzhuang coal mine, China," Mine Water and the Environment, vol. 36, no. 1, pp. 39–50, 2017.

[9] J. Z. Qian, C. Lv, W. D. Zhao, and J. Pan, "Comparison of application on Elman and BP neural networks in discriminating water bursting source of coal mine," Systems Engineering-Theory & Practice, vol. 30, no. 1, pp. 145–150, 2010.

[10] L. Huang, J. Li, H. Hao, and X. Li, "Micro-seismic event detection and location in underground mines by using convolutional neural networks (CNN) and deep learning," Tunnelling and Underground Space Technology, vol. 81, pp. 265–276, 2018.

[11] X. Y. Wang, W. Zhao, X. M. Liu et al., "Identification of water inrush source from coalfield based on entropy weight-fuzzy variable set theory," Journal of China Coal Society, vol. 42, no. 9, pp. 2433–2439, 2017.

[12] X. Ding, X. Chong, Z. Bao, Y. Xue, and S. Zhang, "Fuzzy comprehensive assessment method based on the entropy weight method and its application in the water environmental safety evaluation of the Heshangshan drinking water source area, three gorges reservoir area, China," Water, vol. 9, no. 5, pp. 329–344, 2017.

[13] A. Salifu, B. Petrusovsky, K. Ghebremichael, R. Buamah, and G. Amy, "Multivariate statistical analysis for fluoride occurrence in groundwater in the northern region of Ghana," Journal of Contaminant Hydrology, vol. 140-141, pp. 34–44, 2012.

[14] Y. Wang, P. Wang, Y. Bai et al., "Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China," Journal of Hydro-Environment Research, vol. 7, no. 1, pp. 30–40, 2013.

[15] G. Liu, F. Ma, G. Liu, H. Zhao, J. Guo, and J. Cao, "Application of multivariate statistical analysis to identify water sources in a coastal gold mine, Shandong, China," Sustainability, vol. 11, no. 12, article 3345, 2019.

[16] Y. Wang, L. Q. Shi, M. Wang, and T. H. Liu, "Hydrochemical analysis and discrimination of mine water source of the Jiaojia gold mine area, China," Environmental Geology, vol. 79, no. 6, p. 123, 2020.

[17] L. H. Sun and H. R. Gui, "Establishment of water source discrimination model in coal mine by using hydrogeochemistry and statistical analysis: a case study from Renlou coal mine in northern Anhui Province, China," Journal of Coal Science & Engineering, vol. 18, no. 4, pp. 385–389, 2012.

[18] L. H. Sun, "Application of hydrochemistry for inrush water source identification in coal mine: approach based on statistical analysis," Mining Science, vol. 25, pp. 115–124, 2018.

[19] W. D. Gao, Y. D. He, and X. S. Li, "Application of hydrochemical method in determining water inrush source and inrush source in coal mines," Mining Safety & Environmental Protection, vol. 28, no. 5, pp. 44–45, 2001.

[20] P. Huang and X. Wang, "Piper-PCA-fishcer recognition model of water inrush source: a case study of the Jiaozuo mining area," Geofluids, vol. 2018, Article ID 9205025, 10 pages, 2018.

[21] P. Huang, Z. Yang, X. Wang, and F. Ding, "Research on Piper-PCA-Bayes-LOOCV discrimination model of water inrush source in mines," Arabian Journal of Geosciences, vol. 12, no. 11, p. 334, 2019.

[22] J. Y. Wang, C. P. Li, and Z. X. Li, "Risk prediction of water inrush from coal floor based on principal component clustering analysis," China Safety ence Journal, vol. 23, no. 8, pp. 120–125, 2013.

[23] G. Liu, F. Ma, G. Liu, J. Guo, X. Duan, and H. Gu, "Quantification of water sources in a coastal gold mine through an end-member mixing analysis combining multivariate statistical methods," Water, vol. 12, no. 2, p. 580, 2020.

[24] J. Zhang, L. Chen, Y. Chen et al., "Discrimination of water-inrush source and evolution analysis of hydrochemical environment under mining in Renlou coal mine, Anhui Province, China," Environmental Geology, vol. 79, no. 2, p. 61, 2020.

[25] X. Duan, F. Ma, J. Guo et al., "Source identification and quantification of seepage water in a coastal mine, in China," Water, vol. 11, no. 9, article 1862, 2019.

[26] J. T. Lu, X. B. Li, F. Q. Gong, X. R. Wang, and J. Liu, "Recognizing of mine water inrush sources based on principal components analysis and fisher discrimination analysis method," China Safety Science Journal, vol. 22, no. 7, pp. 109–115, 2012.

[27] R. K. Steinhorst and R. E. Williams, "Discrimination of groundwater sources using cluster analysis, MANOVA, canonical analysis and discriminant analysis," Water Resources Research, vol. 21, no. 8, pp. 1149–1156, 1985.

[28] G. Cünkayt, G. D. Thyne, J. E. Mccray, and K. A. Turner, "Evaluation of graphical and multivariate statistical methods for classification of water chemistry data," Hydrogeology Journal, vol. 10, no. 4, pp. 455–474, 2002.

[29] Y. J. Li, Multivariate Statistical Analysis, Beijing University of Posts and Telecommunications Press, Beijing, 2018.
[30] W. T. Zhang and W. Dong, *SPSS Statistical Analysis Advanced Tutorial*, Higher Education Press, Beijing, 2004.

[31] V. Cloutier, R. Lefebvre, R. Therrien, and M. M. Savard, "Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system," *Journal of Hydrology*, vol. 353, no. 3-4, pp. 294–313, 2008.

[32] E. E. Zhuk and E. V. Serikova, "Effectiveness of the step-by-step discriminatory analysis at choosing the informative attributes," *Automation and Remote Control*, vol. 66, no. 11, pp. 1768–1781, 2005.

[33] X. Y. Jiang and C. Q. Cheng, "Hydrochemical classification and identification of groundwater in mining region using multivariate statistical analysis," *Hydrogeology & Engineering Geology*, vol. 36, no. 4, pp. 16–20, 2009.

[34] E. Hou, Q. Wen, X. Che, W. Chen, J. Wei, and Z. Ye, "Study on recognition of mine water sources based on statistical analysis," *Arabian Journal of Geosciences*, vol. 13, no. 1, pp. 1–12, 2020.

[35] Q. Du, L. Y. Jia, and X. F. Yan, *SPSS Statistical Analysis from Getting Started to Mastering Version 2*, The People’s Posts and Telecommunications Press, Beijing, 2014.

[36] Y. T. Zhang and K. Z. Fang, *Introduction to Multivariate Statistical Analysis*, Wuhan University Press, Wuhan, 2013.

[37] Z. Z. Zhu, A. Q. Fu, Q. Y. Sun et al., "Coal mine water classification," *Chongqing Institute of Geology and Mineral Resources*, vol. GB/T 19223-2015, pp. 1–6, 2015.

[38] S. Shrestha and F. Kazama, "Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan," *Environmental Modelling & Software*, vol. 22, no. 4, pp. 464–475, 2007.

[39] H. Zhang, H. Xing, D. Yao, L. Liu, D. Xue, and F. Guo, "The multiple logistic regression recognition model for mine water inrush source based on cluster analysis," *Environmental Geology*, vol. 78, no. 20, p. 612, 2019.

[40] J. Qian, L. Wang, L. Ma, Y. H. Lu, W. Zhao, and Y. Zhang, "Multivariate statistical analysis of water chemistry in evaluating groundwater geochemical evolution and aquifer connectivity near a large coal mine, Anhui, China," *Environmental Earth Sciences*, vol. 75, no. 9, p. 747, 2016.

[41] G. A. Alther, "A simplified statistical sequence applied to routine water quality analysis: a case history," *Ground Water*, vol. 17, no. 6, pp. 556–561, 2010.

[42] K. P. Singh, A. Malik, D. Mohan, and S. Sinha, "Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study," *Water Research*, vol. 38, no. 18, pp. 3980–3992, 2004.