WILEY | Hindawi

Research Article
Identification of Mine Water Inrush Source Based on PCA-FDA: Xiandewang Coal Mine Case

Bo Li [ID],[1,2] Qiang Wu,[3] and Zijie Liu[1]

[1]*Key Laboratory of Karst Georesources and Environment, Ministry of Education, Guizhou University, Guiyang, Guizhou 550025, China*
[2]*Key Laboratory of Karst Environment and Geological Disaster Prevention and Control of Guizhou Province, Guizhou University, Guiyang, Guizhou 550025, China*
[3]*National Engineering Research Center of Coal Mine Water Hazard Controlling, China University of Mining and Technology (Beijing), Beijing 100083, China*

Correspondence should be addressed to Bo Li; libo1512@163.com

When mine water inrush accidents occur, timely and accurately identifying the water inrush source plays an important role in determining the cause of water inrush and making a solution to a disaster. According to the differences of water chemical composition in each water sources of mine, eight kinds of indicators of water chemical composition were selected as sample variables for water inrush source identification. On this basis, an identification model of water inrush source was established by using principal component analysis (PCA) and Fisher discriminant analysis (FDA) combined. The model was used to identify the water inrush source of 14 groups of training samples and 12 groups of samples to be judged in different water sources of the Xiandewang coal mine, and it was compared with the results of the conventional identification model which used the FDA method. Results of this study showed that having processed data by using the PCA method can effectively eliminate the effects of information superposition between sample indicators, and the identification accuracy of mine water inrush source was significantly increased. Related study in this paper can provide some basis and reference for the study of mine water inrush source identification technology.

1. Introduction

Coal is a kind of primary energy. Various kinds of disaster accidents often occur in the process of coal mining, and water inrush is one of the factors that cause serious accidents in coal mines [1–3]. Many water inrush cases show that when a water inrush accident occurs in a coal mine, timely and accurately identifying the source of water inrush can find causes of the occurrence quickly and make solutions to the water inrush timely, which is important for the prevention and control of water inrush disasters in a coal mine [4–7]. People often use groundwater chemistry, isotope, water temperature, water level, and other indicators to identify the water inrush source. Experience has shown that the water chemistry analysis method is a more effective method among them [8–10]. The basis of using the water chemistry analysis method to identify the water inrush source is that the groundwater of different aquifers has different water chemical composition. These components which can be used to distinguish the characteristics of groundwater in different aquifers are called "standard components", and "standard components" used more extensively are conventional components like Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, HCO^{3-}, alkalinity, acidity, hardness, TDS, and pH. Based on the analysis of the water chemical composition of aquifers, different researchers have proposed different methods for water source identification, and examples below are representative [11] used the fuzzy comprehensive evaluation method and the cluster analysis method to identify the water inrush sources of the Mindong No. 1 mine [12] used the maximum likelihood method to identify the potential water sources of the Sanshandao Gold Mine based on the hydrogeochemical

and isotopic analyses [13] used the distance discrimination method to identify mine inrush water sources, and the results were verified through the grey relational analysis method. Liu et al. (2015) established a water source identification model based on the BP neural network theory and randomly selected water samples collected during mine excavation to predict water source identification. These methods above have a positive effect on enriching the water inrush source identification technology, but they did not take information superposition between identification indicators of water chemical into consideration, which caused problems like low precision of classification and long response time. To solve these problems, this paper introduced the principal component analysis (PCA) method into the water inrush source identification technology, refined the water chemical indicator data of different water sources, converted multiple related indicator variables into a new independent one by linear combination, and eliminated the effects caused by information superposition between indicators so that characteristics of different water sources can be described more effectively. On this basis, the Fisher discrimination analysis (FDA) method was combined to establish a water source identification analysis model. By using this model, the water inrush source of the typical coal mine was identified and the results of identification were good.

2. Methods of Mine Water Inrush Source Identification

2.1. Principal Component Analysis (PCA). As a statistical method, PCA is aimed at converting a set of potentially correlated variables to a new set of variables that are linearly uncorrelated by means of orthogonal transformation; and the new variables obtained through transformation are known as principal components and they are capable of keeping the original information to be revealed unchanged in the aspect of expressing information. Data processing based on PCA plays a part in effectively eliminating the correlation of high-dimensional data, realizes data dimension reduction, and simplifies data structures [14, 15]. A mathematical model of PCA is expressed as follows.

P variables $(X_1, X_2, \cdots, X_p)$ of a raw data matrix X form a linear combination denoted as $Y = AX$, namely,

$$\begin{cases} Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots a_{1p}X_p, \\ Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots a_{2p}X_p, \\ \vdots \\ Y_p = a_{p1}X_1 + a_{p2}X_2 + \cdots a_{pp}X_p, \end{cases} \tag{1}$$

where $a_{i1} + a_{i2} + \cdots + a_{ip} = 1$, Y_i is uncorrelated with Y_j ($i \neq j$; $i, j = 1, 2, \cdots, p$), Y_1 has the maximum variance if compared with all linear combinations of $X_1, X_2, \cdots, X_p$; among all linear combinations of $X_1, X_2, \cdots, X_p$ that are uncorrelated with Y_1, Y_2 has the maximum variance; Y_P has the maximum variance in comparison with those of all linear combinations correlated with none of $Y_1, Y_2, \cdots, Y_{p-1}$; and the sum of

variances of $Y_1, Y_2, \cdots, Y_p$ is equal to that of variances of $X_1, X_2, \cdots, X_p$.

Steps of figuring out its principal components are generally as follows:

(1) The original variable data are normalized, which is followed by calculations of a covariance matrix $\sum$ for all variables

(2) Eigenvectors of the covariance matrix figured out can be ranked as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ and the corresponding unit eigenvectors are $T_1, T_2, \cdots$, and T_p. In the event of conversion matrix $A = T^{'}$, row i of A represents an eigenvalue T_i in the ith place of $\sum$; and a variance of ith principal component Y_i is just an eigenvalue λ_i in i^{th} place of $\sum$ as well

(3) The variance contribution rate of kth principal component Y_k is denoted by $\eta_k = (\lambda_k / \sum_{k=1}^{p} \lambda_k) = 1$. If m ($m < p$) principal components are adopted, the cumulative variance contribution rate of principal components $Y_1, Y_2, \cdots$, and Y_m is expressed in $\xi_m = (\sum_{k=1}^{m} \lambda_k / \sum_{k=1}^{p} \lambda_k)$

(4) Determination of the number of principal components depends on the cumulative variance contribution rate in general. Usually, the fact that the cumulative variance contribution rate reaches at least 80% indicates that the following requirement can be satisfied: sample information of the first m principal components extracted contains most of the information about primary samples

2.2. Fisher Discrimination Analysis (FDA). FDA serves as a multivariable statistical analysis method that uses eigenvalues of a research object to identify its type. Basic thoughts of FDA can be described as follows. Dimensionality reduction for a multidimensional data is achieved through a projection so as to simplify the corresponding problem and determine a discrimination function based on a principle of maximum interclass distance and minimum within-class distance [16–19]. A mathematical model of FDA is expressed as follows.

It is assumed that there are n ensembles $G_1, G_2, \cdots$, and G_n ($i = 1, 2, \cdots, n$); and their corresponding mean vectors and covariance matrixes are $U^{(1)}, U^{(2)}, \cdots, U^{(n)}$ and $V^{(1)}, V^{(2)}, \cdots, V^{(n)}$, respectively. In case that samples with a size of m_i are taken from the ensemble G_i, that is,

$$X_a^i = \left(x_{a_1}^i, x_{a_2}^i, \cdots, x_{a_p}^i \right)^T, \quad (a = 1, 2, \cdots, \text{m} ; i = 1, 2, \cdots, n), \tag{2}$$

then, $U^T X_a^i = (U^{(1)} x_{a_1}^i + U^{(2)} x_{a_2}^i + \cdots + U^{(p)} x_{a_p}^i)^T$ ($i = 1, 2, \cdots, n$) stands for a projection of sample X_a^i on a space axis, which can be denoted as follows:

$$\begin{cases} \overset{-i}{X} = \dfrac{1}{m_i} \sum_{i=1}^{m_i} X_a^i, \\[2mm] \bar{X} = \dfrac{1}{m} \sum_{i=1}^{n} \sum_{a=1}^{m_i} X_a^i, \ n = \sum_{i=1}^{n} m_i, \end{cases} \tag{3}$$

where X^{-i} and $\bar{X}$ are mean values of the selected samples and the total sample, respectively; in which case, intragroup deviation e for a set of samples is

$$\begin{aligned} e &= \sum_{i=1}^{n} \sum_{a=1}^{m_i} \left(U^T X_a^i - U^T \overset{-i}{X} \right)^2 \\ &= U^T \left\{ \sum_{i=1}^{n} S_i \right\} U \triangleq U^T W U \left(W = \left\{ \sum_{i=1}^{n} S_i \right\} \right), \end{aligned} \tag{4}$$

where S_i refers to a sample difference of sample T_i (sample size $= m_i$) projection X_a^i on the space axis; and W is a "total within-class scatter" matrix. Between-group deviation b of the sample is

$$\begin{cases} b = U^T \left[m_i \sum_{i=1}^{n} \left(\overset{-i}{X} - \bar{X} \right) \left(\overset{-i}{X} - \bar{X} \right)^T \right] U \triangleq U^T B U, \\[2mm] B = \left[m_i \sum_{i=1}^{n} \left(\overset{-i}{X} - \bar{X} \right) \left(\overset{-i}{X} - \bar{X} \right)^T \right], \end{cases} \tag{5}$$

where B is a "within-class scatter" matrix of the sample. In order to discriminate it from the total sample under the circumstance that a discrimination function is adopted, ϕ can be expressed as

$$\phi = \frac{b}{e} = \frac{U^T B U}{U^T W U}. \tag{6}$$

If $M = U^T B U - \lambda(U^T W U - 1)$ and its partial differential is calculated, then

$$\frac{\partial M}{\partial U} = 2 B U - 2 \lambda W U \triangleq 0. \tag{7}$$

In Equation (7), λ is its eigenvalue. Through simplification, the following equation is acquired.

$$\left(W^{-1} B - \lambda I \right) U \triangleq 0. \tag{8}$$

In the above equation, U is an eigenvector to which the maximum eigenvalue λ corresponds; and I represents the ratio of a within-sample sum-of-squared difference for the total sample to a sum-of-squared difference between adjacent samples. As can be observed from the above equation, both the maximum eigenvalue and eigenvector U of $W^{-1} B$ can be obtained, thereby figuring out its discrimination function.

2.3. Procedures of Mine Water Inrush Source Identification Based on PCA–FDA. The discriminating idea of using the PCA and PDA methods to identify the water source of mine water inrush is shown in Figure S1. The discriminating process is as follows:

(1) Water sample data are normalized

(2) A correlation matrix of normalized data is figured out to analyze and identify correlation of variables

(3) PCA is conducted for correlated variables for the purpose of reducing the dimensionality of such variables

(4) FDA is utilized to identify training samples, determine differences between their actual values and discriminant values, and verify accuracy of the model

(5) FDA model is adopted to identify samples to be discriminated

3. Project Case

3.1. Study Area

3.1.1. Physical Geography. Under the jurisdiction of Baita Town in Shahe City, the Xiandewang coal mine is located at the southwest 35 km away from Xingtai City of Hebei Province, China. The location is shown in Figure 1. Being high in the west and low in the east, it is undulating in terrain and has maximum and minimum elevations of +339.6 m and +194.10 m, respectively. As this region belongs to a semiarid warm temperate continental monsoon climate, precipitation mainly takes place from July to September each year. For the past 10 years, annual precipitation ranges between 351.5 mm and 800 mm, generating an average annual precipitation of 507.74 mm. Moreover, its geographical coordinates are $114°11'15'' \sim 114°15'00''$ E and $36°48'45'' \sim 36°55'00''$ N.

3.1.2. Hydrogeology. According to hydrogeologic prospecting data of the mining area, aquifers threatening safe mining include an Ordovician limestone karst fractured aquifer, a Permian sandstone fractured aquifer, and Daqing and Yeqing Carboniferous limestone karst fractured aquifers. Due to differences in chemical components, structures, lithological associations, and fracture development of various rock strata, their water-bearing characteristics and water yield properties are also significantly different. Once a fault is encountered or any damage is caused to roof-floor strata in the process of coal mining, it is much likely for underground water in an aquifer to burst into a mine and thus result in water inrush. For this reason, an investigation on mine water inrush source identification for Xiandewang coal mine is beneficial for water inrush control there.

4. Indicators for Source Identification

Considering that each aquifer contains diversified water chemical compositions, it is infeasible to adopt chemical constituents of a type of water as indicators for source identification. Taking into account the groundwater detection of coal mines, a brief water quality analysis method is generally used,
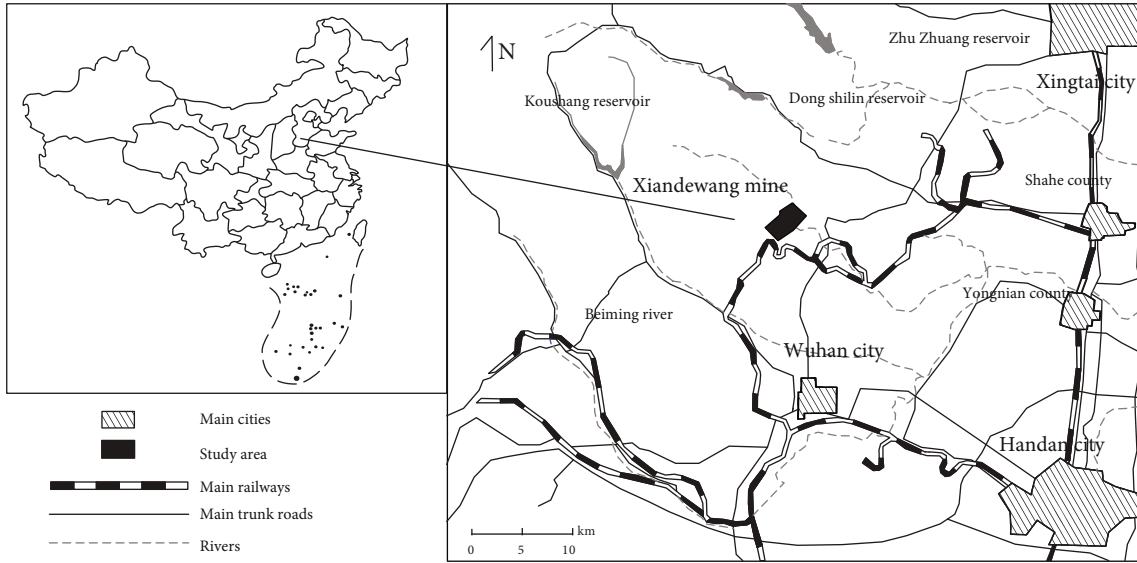
FIGURE 1: Traffic location.

TABLE 1: Training sample data (unit: mg/l).

| Sample number | Indicators of water chemical composition | | | | | | | | Aquifer |
	Hardness	PH	Na$^+$+K$^+$	Ca^{2+}	Mg^{2+}	Cl$^-$	SO$_4$$^{2-}$	HCO$_3$$^-$	
1	199.71	7.36	11.95	62.32	10.70	14.89	106.19	108.59	I
2	182.19	7.90	16.32	52.71	12.28	11.35	30.46	200.72	I
3	224.73	7.53	15.63	64.13	15.69	13.83	16.46	261.11	I
4	271.78	7.50	15.40	79.56	17.75	14.18	54.33	277.59	III
5	246.25	7.83	14.25	70.74	16.90	19.15	42.39	244.64	III
6	278.29	7.59	46.21	68.34	26.14	18.08	79.30	331.88	III
7	252.26	7.68	24.60	64.33	22.25	12.41	56.39	281.25	III
8	204.21	8.05	164.38	37.88	26.63	34.39	39.10	578.36	III
9	6.42	7.60	145.53	24.65	12.89	72.69	9.47	393.50	II
10	11.44	7.60	79.78	63.53	11.07	34.39	26.75	366.05	II
11	11.51	8.77	299.10	2.81	1.09	85.10	130.48	444.75	II
12	0.45	10.50	323.24	2.20	0.00	323.24	0.00	2.20	IV
13	10.51	9.30	307.15	2.40	0.00	67.37	40.75	440.48	IV
14	10.51	9.30	307.15	2.40	0.00	67.37	40.75	440.48	IV

and, refer to relevant literature [20–22], contents of 8 chemical constituents such as hardness (A_1), pH value (A_2), Na$^+$+K$^+$ (A_3), Ca^{2+} (A_4), Mg^{2+} (A_5), Cl$^-$ (A_6), SO^{4-} (A_7), and HCO$_3$$^-$ (A_8) were selected as identification indicators of a mine water inrush source discrimination model. From underground water monitoring data of the Xiandewang coal mine over the years, hydrochemical data of 14 water samples taken from 4 different aquifers were used as training samples. To be specific, the aquifers consist of an Ordovician limestone karst fractured aquifer (I), a Permian sandstone fractured aquifer (II), a Carboniferous Daqing limestone karst fractured aquifer (III) and a Carboniferous Yeqing limestone karst fractured aquifer (IV). As for sample data, please refer to Table 1.

4.1. Data Processing Based on PCA. Data of training samples in Table 1 were firstly normalized, and their normalized values were equal to a difference between the actual value and the minimum value divided by a difference of the maximum and the minimum values. Table 2 presents the normalized data. Subsequently, the normalized data were processed based on PCA and the correlation coefficient matrix for hydrochemical constituents of various water sources is shown in Table 3. It can be observed from Table 3 that such 8 hydrochemical constituents are clearly correlated with each other. For example, a correlation coefficient of Ca^{2+} and the total hardness is 0.983, while that of Na$^+$+Ks$^+$ and Cl$^-$ is equal to 0.919; in addition, information about sample indicators significantly overlaps, so that it is

TABLE 2: Normalized data.

Sample number	Indicators of water chemical composition								Actual result
	Hardness	PH	Na$^+$+K$^+$	Ca^{2+}	Mg^{2+}	Cl$^-$	SO$_4^{2-}$	HCO$_3^-$	
1	0.1514	0.1917	0.0000	0.1626	0.1236	0.0480	0.0939	0.0000	I
2	0.1377	0.4167	0.0148	0.1365	0.1418	0.0000	0.0204	0.2393	I
3	0.1710	0.2625	0.0125	0.1675	0.1812	0.0336	0.0068	0.3962	I
4	0.1879	0.3875	0.0078	0.1854	0.1952	0.1058	0.0320	0.3534	III
5	0.2130	0.2875	0.1161	0.1789	0.3019	0.0913	0.0678	0.5800	III
6	0.1926	0.3250	0.0429	0.1680	0.2570	0.0144	0.0455	0.4485	III
7	0.8337	0.0000	0.2399	0.7983	0.0939	0.1394	0.8434	0.6878	III
8	1.0000	0.1667	0.1628	1.0000	1.0000	0.1586	1.0000	0.6656	III
9	0.0000	0.2917	0.4525	0.0604	0.1489	0.8317	0.0000	0.7401	II
10	0.0039	0.2917	0.2298	0.1659	0.1279	0.3124	0.0168	0.6688	II
11	0.0040	0.7792	0.9727	0.0011	0.0126	1.0000	0.1175	0.8732	II
12	0.3236	0.2500	0.4393	0.2213	0.0000	0.3509	0.2425	1.0000	IV
13	0.0032	1.0000	1.0000	0.0000	0.0000	0.7596	0.0304	0.8621	IV
14	0.0032	1.0000	1.0000	0.0000	0.0000	0.7596	0.0304	0.8621	IV

TABLE 3: Correlation coefficient matrix of chemical constituents.

Indicator	Hardness	PH	Na$^+$+K$^+$	Ca^{2+}	Mg^{2+}	Cl$^-$	SO$_4^{2-}$	HCO$_3^-$
Hardness	1.000							
PH	-0.584	1.000						
Na$^+$+K$^+$	-0.340	0.828	1.000					
Ca^{2+}	0.983	-0.609	-0.363	1.000				
Mg^{2+}	0.690	-0.408	-0.410	0.730	1.000			
Cl$^-$	-0.428	0.700	0.919	-0.417	-0.377	1.000		
SO$_4^{2-}$	0.962	-0.448	-0.120	0.961	0.631	-0.199	1.000	
HCO$_3^-$	0.019	0.404	0.759	-0.006	-0.156	0.710	0.178	1.000

inevitable to affect accuracy of the corresponding mine water inrush source identification model when data about such 8 water samples are utilized to identify sources of water inrush. Consequently, misjudgments may be made. Therefore, data of training samples were processed based on PCA by virtue of the abovementioned PCA mathematical model. In this way, a cumulative contribution rate diagram of all indicators can be acquired. In Figure 2, it is clear that the first 4 indicators include 97.37% of the raw data information content and thus can effectively summarize the information of raw data.

In order to reconstruct the new characteristics of the original data, the eight indicators (A_1, A_2, A_3, A_4, A_5, A_6, A_7, and A_8) that had a certain correlation were recombined into a new set of independent indicators (Y_1, Y_2, Y_3, and Y_4) to replace the original indicators; according to the PCA calculation, a PCA matrix can be obtained, as shown in Table 4.

Each value in the principal component analysis matrix in Table 4 indicates the degree of influence of the four extracted principal components $Y_1 - Y_4$ on the original eight indicators $A_1 - A_8$; higher values in the matrix indicate a closer relationship between the indicator and the principal components, and at the same time, the values in the matrix are the coefficients of the factor expressions of the indicators. According to Table 4, the extracted principal components $Y_1 - Y_4$ and the original indicators are expressed as

$$Y_1 = 0.873A_1 - 0.826A_2 - 0.742A_3 + 0.888A_4 + 0.749A_5 \\ - 0.750A_6 + 0.748A_7 - 0.387A_8,$$

$$Y_2 = 0.453A_1 + 0.262A_2 + 0.650A_3 + 0.442A_4 + 0.239A_5 \\ + 0.571A_6 + 0.634A_7 + 0.812A_8,$$

$$Y_3 = -0.064A_1 + 0.392A_2 + 0.045A_3 - 0.030A_4 + 0.557A_5 \\ + 0.029A_6 - 0.066A_7 - 0.242A_8,$$

$$Y_4 = 0.137A_1 + 0.265A_2 + 0.134A_3 + 0.084A_4 - 0.268A_5 \\ - 0.130A_6 + 0.162A_7 - 0.271A_8.$$

$$(9)$$

4.2. Analysis of Source Identification. The four principal components Y_1, Y_2, Y_3, and Y_4 gained through PCA were selected as identification indicators; then, water sources
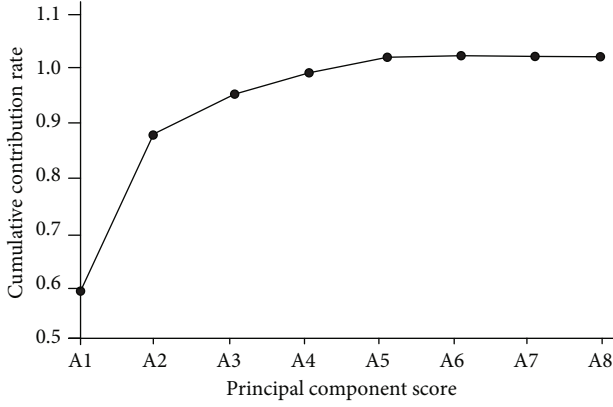
FIGURE 2: Cumulative contribution rates of all principal components.

TABLE 4: PCA matrix.

Parameter	Y_1	Y_2	Y_3	Y_4
A_1	0.873	0.453	-0.064	0.137
A_2	-0.826	0.262	0.392	0.265
A_3	-0.742	0.650	0.045	0.134
A_4	0.888	0.442	-0.030	0.084
A_5	0.749	0.239	0.557	-0.268
A_6	-0.750	0.571	0.029	-0.130
A_7	0.748	0.634	-0.066	0.162
A_8	-0.387	0.812	-0.242	-0.271

TABLE 5: Central values of discrimination functions in different groups.

Type	Discrimination function		
	F_1	F_2	F_3
I	3.164	-0.959	-0.036
II	-2.261	-0.165	0.499
III	1.406	0.813	-0.034
IV	-3.248	-0.232	-0.406

could be classified into 4 types in consistency with differences in aquifers. On the assumption that the covariance matrices in the group are equal, the coefficients of the discriminant formula are calculated according to the formula (6) in the FDA principle mentioned above. The principle of determining the coefficient is to maximize the distance difference between the various types and minimize the distance within the type. The following discriminant function is obtained as follows:

$$F_1 = 1.554Y_1 - 3.819Y_2 + 0.606Y_3 + 2.953Y_4,$$
$$F_2 = 3.077Y_1 + 1.417Y_2 + 0.273Y_3 - 3.424Y_4, \quad (10)$$
$$F_3 = 3.606Y_1 + 2.875Y_2 + 2.464Y_3 - 4.597Y_4.$$

Central values of such 3 discrimination functions in 4 groups are presented in Table 5. Taking the first discrimi-

TABLE 6: Identification results of training samples.

Sample number	PCA-FDA	FDA	Actual type	F_1	F_2	F_3
1	I	I	I	4.764	-1.720	-0.562
2	I	I	I	2.932	-0.918	0.023
3	I	III	I	0.433	-0.238	1.796
4	I	I	III	2.115	-0.322	0.387
5	III	III	III	0.391	0.429	0.869
6	III	III	III	1.448	0.000	0.628
7	III	III	III	1.258	1.136	-2.619
8	III	III	III	1.820	2.824	0.566
9	II	II	II	-2.153	0.046	0.966
10	II	II	II	-0.983	0.161	0.829
11	IV	II	II	-3.647	-0.703	-0.299
12	IV	III	IV	-2.867	0.983	-0.618
13	IV	IV	IV	-3.438	-0.839	-0.301
14	IV	IV	IV	-3.438	-0.839	-0.301

nant, for example, its central values for type I water source (Ordovician limestone), type II water source (Permian sandstone water), type III water source (Carboniferous Daqing limestone water), and type IV water source (Carboniferous Yeqing limestone water) are 3.164, -2.261, 1.406, and -3.248, respectively. The three discriminant functions calculate the coordinates of each water sample in each dimension. By comparing the distance between the water sample to be judged and the center value of the four types of water source groups, the water source group to which the sample belongs is determined.

4.3. Validity Check for Source Identification. For the sake of validity check for a PCA and FDA-based water source identification model, all training samples given in Table 1 were substituted in the established identification model one by one for reverse identification. Regarding reverse identification results, they have been presented in Table 6. It can be observed from this table that misidentification is incurred for water samples 4 and 11 with identification accuracy at 12/14 = 85.7%. In terms of the conventional Fisher water source identification model, water samples 3, 4, and 12 are subjected to misidentification as well and their accuracy of identification is only up to 11/14 = 78.6%. Comparison of identification results between such 2 models signifies that the PCA and FDA-based water source identification model is rather reliable and can meet source identification requirements of mine water inrush better.

To further check the accuracy of model, the verified PCA and FDA-based water source identification model was utilized to identify 12 samples taken from the Xiandewang coal mine. As for relevant identification samples and results, they have been shown in Table 7. It can be known from this table that except for water sample 4 with misidentification, identification results of other water samples are consistent with their actual outcomes, generating identification accuracy at 11/12 = 91.7%. Therefore, such a model can be used to

TABLE 7: Samples to be identified and identification results.

Sample number	Indicator of water chemical composition								Actual result	PCA-FDA	FDA
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8			
1	225.73	7.82	19.31	66.33	14.59	14.89	29.22	256.23	I	I	I
2	228.73	7.79	42.53	69.34	13.5	9.57	34.16	323.95	I	I	II
3	227.23	7.20	20.00	66.70	14.72	16.31	23.05	269.66	I	I	I
4	31.53	8.34	172.88	7.82	2.92	89.65	71.62	265.54	III	II	II
5	1062.59	6.90	53.47	306.61	72.11	15.96	832.26	352.02	III	III	III
6	600.12	7.00	12.54	179.16	37.09	16.31	367.97	313.58	III	III	III
7	16.02	8.50	220.70	3.61	1.70	89.35	29.64	413.63	II	II	II
8	12.01	8.00	197.49	3.01	1.09	78.36	25.11	373.37	II	II	I
9	7.51	7.80	274.27	2.00	0.61	120.2	22.23	502.1	II	II	II
10	100.66	8.40	268.85	65.45	0.00	23.54	135.65	507.28	IV	IV	IV
11	0.45	10.50	323.24	2.20	0.00	70.20	44.04	285.52	IV	IV	IV
12	0.39	9.40	299.79	2.20	0.00	71.98	4.94	579.18	IV	IV	IV

classify water samples to be identified effectively. Additionally, the conventional Fisher identification model was also employed for water source identification and it results in identification accuracy of 9/12 = 75%. To sum up, the PCA and FDA-based water source identification model performs better in identifying water samples.

5. Conclusion

According to the characteristics of mine water inrush source, the water source identification analysis model based on PCA-FDA and the conventional water source identification analysis model based on FDA were used to identify the water inrush source of the 14 groups of training samples and 12 groups of samples to be judged, respectively, and the identification accuracies were 85.7% and 78.6%, 91.7% and 75% correspondingly. The result shows that after processing data with the PCA method, the identification accuracy of mine water inrush source has increased greatly, compared with using the FDA method only.

In the process of water source identification, the eight indicators in the original data are reduced to four principal components; the PCA method can project high-dimensional data into a low-dimensional space, complete the process of dimensionality reduction of the data, and accurately characterize the water chemistry of each water source highly precisely with less independent identification indicators. This method greatly reduces the number of input-influencing factors when building the identification model and eliminates the effects caused by information superposition between each identification indicator in the identifying process. The FDA method can gather samples of the same kind together, distinguish the different samples, and achieve the analysis of mine water inrush identification. Combining these two methods can establish an accurate water source identification model with minimum characteristic information, simplify data structure, shorten analysis time, and improve analysis precision, which is a more effective analysis method for identifying water inrush sources.

In this case, when the data of the water sample was returned to the identification model, there were a few misjudgments because the number of training samples was not enough relatively. In order to increase the prediction accuracy of the model, in subsequent research, hydrochemical data should be collected in large quantities, comprehensive database of the mine hydrochemistry should be established, the training of the model should be enhanced, and the identification accuracy of the identification model should be improved.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Supplementary Materials

The technology idea of water inrush water identification. *(Supplementary Materials)*

References

[1] M. F. Howladar, "Coal mining impacts on water environs around the Barapukuria coal mining area, Dinajpur, Bangladesh," *Environment and Earth Science*, vol. 70, no. 1, pp. 215–226, 2013.

[2] B. Li and Q. Wu, "Catastrophic evolution of water inrush from a water-rich fault in front of roadway development: a case study of the Hongcai coal mine," *Mine Water and the Environment*, vol. 38, no. 2, pp. 421–430, 2019.

[3] C. F. Santos and Z. T. Bieniawski, "Floor design in underground coal mines," *Rock Mechanics and Rock Engineering*, vol. 22, no. 4, pp. 249–271, 1989.

[4] H. Gu, F. Ma, J. Guo, K. Li, and R. Lu, "Assessment of water sources and mixing of groundwater in a coastal mine: the Sanshandao gold mine, China," *Mine Water and the Environment*, vol. 37, no. 3, pp. 351–365, 2018.

[5] B. Li, Q. Wu, X. Q. Duan, and M. Y. Chen, "Risk analysis model of water inrush through the seam floor based on set pair analysis," *Mine Water and the Environment*, vol. 37, no. 2, pp. 281–287, 2018.

[6] G. Q. Li, Z. P. Meng, Z. Q. Wang, and J. Yang, "Hydrochemical prediction of mine water inrush at the Xinli Mine, China," *Mine Water and the Environment*, vol. 36, no. 1, pp. 78–86, 2017.

[7] J. S. Wu, S. D. Xu, R. Zhou, and Y. Qin, "Scenario analysis of mine water inrush hazard using Bayesian networks," *Safety Science*, vol. 89, pp. 231–239, 2016.

[8] I. D. Clark and K. G. Raven, "Sources and circulation of water and arsenic in the Giant Mine, Yellowknife, NWT, Canada," *Isotopes in Environmental and Health Studies*, vol. 40, no. 2, pp. 115–128, 2004.

[9] 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.

[10] N. A. Rapantová, Ś. Krzeszowski, A. Grmela, and C. Wolkersdorfer, "Quantitative assessment of mine water sources based on the general mixing equation and multivariate statistics," *Mine Water and the Environment*, vol. 31, no. 4, pp. 252–265, 2012.

[11] Z. L. Guan, Z. F. Jia, Z. Q. Zhao, and Q. Y. You, "Identification of inrush water recharge sources using hydrochemistry and stable isotopes: a case study of Mindong No. 1 coal mine in north-east Inner Mongolia, China," *Systems Science*, vol. 128, no. 7, p. 200, 2019.

[12] X. L. Duan, F. S. Ma, H. J. Zhao et al., "Determining mine water sources and mixing ratios affected by mining in a coastal gold mine, in China," *Environment and Earth Science*, vol. 78, no. 10, p. 299, 2019.

[13] J. Zhou, X. Z. Shi, and H. Y. Wang, "Water-bursting source determination of mine based on distance discriminant analysis model," *Journal of China Coal Society*, vol. 35, no. 2, pp. 278–282, 2010.

[14] I. M. Farnham, K. J. Stetzenbach, A. K. Singh, and K. H. Johannesson, "Deciphering groundwater flow systems in Oasis Valley, Nevada, using trace element chemistry, multivariate statistics, and geographical information system," *Mathematical Geology*, vol. 32, no. 8, pp. 943–968, 2000.

[15] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification," *International Journal of Remote Sensing*, vol. 40, no. 18, pp. 7190–7220, 2019.

[16] H. J. Chen, X. B. Li, A. H. Liu, and S. Q. Peng, "Identifying of mine water in rush sources by Fisher discriminant analysis method," *Journal of Central South University*, vol. 40, no. 4, pp. 1114–1120, 2009.

[17] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*, pp. 41–48, Madison, WI, USA, August 1999.

[18] P. C. Yan, M. R. Zhou, Q. M. Liu, R. Wang, and J. Liu, "Research on the source identification of mine water inrush based on LIF technology and PLS-DA algorithm," *Guang pu xue yu guang pu fen xi= Guang pu*, vol. 36, no. 9, pp. 2858–2862, 2016.

[19] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.

[20] X. Liu, L. W. Chen, M. L. Lin, and S. D. Li, "Fisher recognition analysis for coal mining inrush water source under mining-induced disturbance and inversion of groundwater recharge relation," *Hydrogeology & Engineering Geology*, vol. 40, no. 4, pp. 36–43, 2013.

[21] L. H. Sun, "Statistical analysis of hydrochemistry of groundwater and its implications for water source identification: a case study," *Arabian Journal of Geosciences*, vol. 7, no. 9, pp. 3417–3425, 2014.

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