International Journal of Geosciences, 2023, 14, 710-718
https://www.scirp.org/journal/ijg
ISSN Online: 2156-8367
ISSN Print: 2156-8359

Identification of Mine Water Inrush Source Based on PCA-BP Neural Network

Mingcheng Ning, Haifeng Lu

School of Earth and Environment, Anhui University of Science & Technology, Huainan, China
Email: nmcaust@163.com

How to cite this paper: Ning, M.C. and Lu, H.F. (2023) Identification of Mine Water Inrush Source Based on PCA-BP Neural Network. *International Journal of Geosciences*, **14**, 710-718.
https://doi.org/10.4236/ijg.2023.148038

Received: July 13, 2023
Accepted: August 14, 2023
Published: August 17, 2023

Abstract

It is of great significance to analyze the chemical indexes of mine water and develop a rapid identification system of water source, which can quickly and accurately distinguish the causes of water inrush and identify the source of water inrush, so as to reduce casualties and economic losses and prevent and control water inrush disasters. Taking Ca^{2+}, Mg^{2+}, Na^{+} + K^{+}, HCO_3^-, SO_4^{2-}, Cl^-, pH value and TDS as discriminant indexes, the principal component analysis method was used to reduce the dimension of data, and the identification model of mine water inrush source based on PCA-BP neural network was established. 96 sets of data of different aquifers in Panxie mining area were selected for prediction analysis, and 20 sets of randomly selected data were tested, with an accuracy rate of 95%. The model can effectively reduce data redundancy, has a high recognition rate, and can accurately and quickly identify the water source of mine water inrush.

Keywords

Mine Water Inrush, Analysis of Hydrochemical Characteristics, Principal Component Analysis (PCA), Back Propagation Neural Networks, Water Source Identification

1. Introduction

In June 2021, the Coal Industry Association issued "14th Five-Year" high-quality development guidance for coal industry' [1], pointing out that by the end of the "14th Five-Year Plan", China's energy consumption is still dominated by coal. Therefore, in order to ensure the long-term and stable development of China's economy, it is necessary to mine coal resources safely and reasonably to ensure the stable production of coal. Water inrush accident is the second largest disaster in coal mine production, but its economic loss is in the first place, and it is also an important factor restricting the development of coal safety production [2] [3]. The most important thing in the process of prediction and prevention of water inrush accidents is to determine the source of water inrush accurately and quickly, furthermore make early warning and judgment in time, and take effective measures to prevent and control water, which can greatly reduce the occurrence of water inrush accidents.

This paper takes Dingji Mine, Guqiao Mine and Pansan Mine in the middle of Panxie Mining Area in Huainan Coalfield as examples. Based on the collected mine hydrochemical data, combined with the hydrogeological and structural problems of the mine field, the mine water inrush source identification model based on PCA-BP neural network is established. which laid a foundation for the development of mine water inrush source identification system and the development of water chemistry database, and the efficient management of mine water chemical data and the rapid identification of water sources. The results have important application significance for mine water disaster prevention and control.

2. Overview of Study Area

In this paper, the central area of Panxie mining area in Huainan coalfield, located in the southern margin of Huaibei Plain, is taken as the research area. The area is dominated by Huaihe River, surrounded by small rivers such as Xifei River, Jiahe River and Nihe River, as well as two large lakes such as Huajia Lake and Jiaogang Lake and several small lakes such as coal mining subsidence water area. The hydrogeological conditions are mainly controlled by neotectonic movement and regional structure, and the groundwater in deep and shallow parts is obviously different. The main aquifers in the study area are composed of Cenozoic loose layer pore aquifer (group), Permian sandstone fissure aquifer (group), Carboniferous Taiyuan Formation limestone fissure karst aquifer, Ordovician limestone fissure karst aquifer and so on. The division of the aquifer is shown in **Figure 1**.

3. PCA-BP Neural Network Theoretical Model

3.1. PCA

PCA is a multivariate statistical method using the idea of "linear dimensionality reduction", which is used to characterize more original indicators with a few comprehensive indicators through a certain linear projection [4] [5] [6]. While retaining the vast majority of information in the original indicators, the correlation between the indicators is eliminated. The comprehensive indicators obtained by this method are called the main components [7] [8]. In specific problems, high-dimensional data is not conducive to training better parameters because of its large dispersion, while low-dimensional data can better train parameters. Therefore, through the form of dimensionality reduction, high-dimensional data indicators are mapped to low-dimensional space, and then the data with the

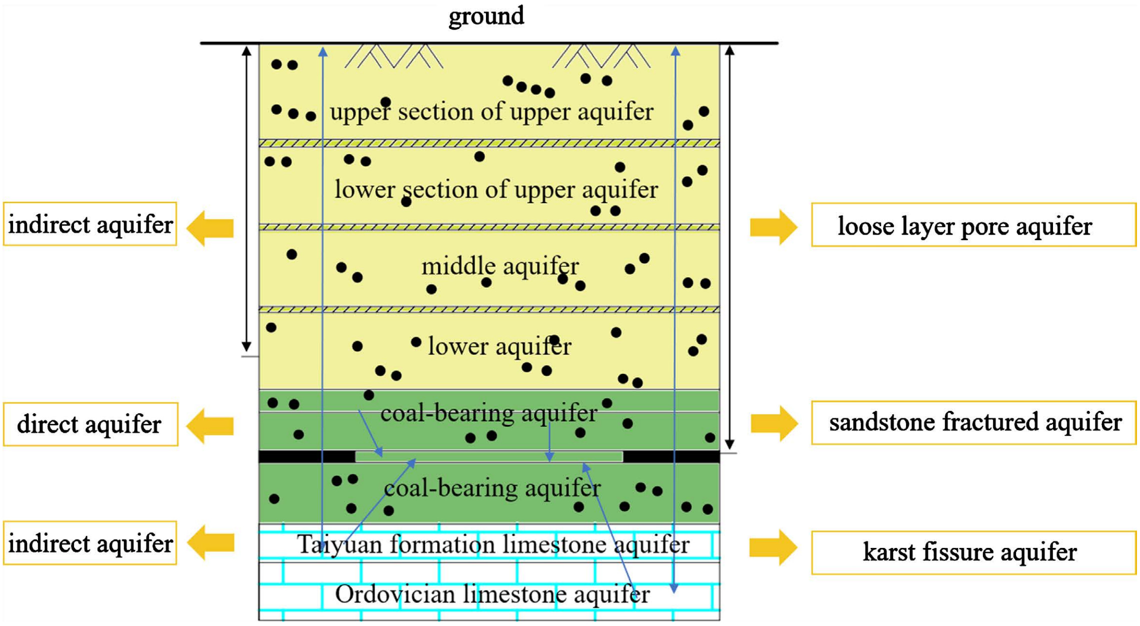

Figure 1. Division of aquifers and mine water-filled aquifers in Panxie mining area of Huainan coalfield.

largest amount of information (*i.e.*, the largest variance) is obtained in the lower dimension.

3.2. BP Neural Network

BP neural network was proposed by Rumelhart and McClelland in 1986. Today, it has become the most widely used neural network algorithm in the world. It is mainly used in function approximation, pattern recognition, data classification and data compression. BP neural network has a strong nonlinear mapping effect, which can carry out associative memory on the input information of external data, and can be effectively identified and classified, it also has the ability to optimize the calculation. These capabilities are widely used. BP neural network can be trained adaptively by sample data and learning process, and recover complete information from noise interference and incomplete information. This function makes it have important applications in image restoration, language processing and pattern recognition [9] [10].

4. Establishment and Application of PCA-BP Neural Network Model

4.1. Hydrochemical Characteristics Analysis

Because the chemical composition of groundwater in different aquifers is different, the difference of hydrochemical characteristics can be used to judge the type of mine water inrush source. The concentrations of Ca^{2+}, Mg^{2+}, K^{+}, Na^{+}, HCO_3^-, SO_4^{2-}, CO_3^{2-} and Cl^- are often selected as the discriminant indexes in the hydrochemical method of mine water inrush source identification.

In this paper, the hydrogeological and water inrush data related to Panxie Mine are analyzed [11], and it is judged that the hydrogeological type of the mining area is medium, and the mine has repeatedly occurred water inrush accidents. A total of 96 sets of water inrush data of Guqiao Mine, Dingji Mine and Pansan Mine were collected as research samples, it includes three prediction types: Cenozoic lower water (expressed by I), coal measure water (expressed by II) and Taiyuan Formation limestone water (expressed by III). A part of water inrush data in the study area is shown in Table 1.

The water chemical composition of 96 groups of data in Table 1 was statistically analyzed, and the results are shown in Table 2. The groundwater in the mining area is alkaline. From the mean value of TDS, it is judged that the groundwater in the mining area is medium mineralized water, and the TDS value of water samples in the study area is very different. On the whole, the mass concentration of the main components of cations in groundwater in the mining area is $Na^+ + K^+ > Ca^{2+} > Mg^{2+}$, The concentration of anion, the main component of anion in Cenozoic lower water and Taiyuan Formation limestone water

Table 1. Original water sample data table (mg/L, except PH).

Serial Number	Aquifer	Ca^{2+}	Mg^{2+}	$Na^+ + K^+$	HCO_3^-	SO_4^{2-}	Cl^-	PH	TDS
1	I	28.18	16.79	1498.29	327.03	308.70	1965.67	8.30	2343.00
2	I	41.40	23.64	839.06	305.40	292.17	1031.90	8.20	2390.00
3	I	35.15	17.72	857.6	329.01	29.96	1009.21	8.20	2396.00
…	…	…	…	…	…	…	…	…	…
94	III	14.90	42.41	122.14	143.08	178.59	93.38	8.82	520.00
95	III	34.24	26.23	876.66	278.43	357.59	1057.34	8.86	2560.00
96	III	10.07	1.46	270.43	26.02	21.81	393.18	9.56	720.00

Table 2. Statistical value table of hydrochemical characteristics.

Aquifer	Eigenvalue	Ca^{2+}	Mg^{2+}	$Na^+ + K^+$	HCO_3^-	SO_4^{2-}	Cl^-	PH	TDS
I	Maximum Value	71.71	46.78	1498.29	648.4	1060.7	1965.67	8.89	3290
	Minimum value	14.31	11.34	563.73	118.84	29.96	401.85	7.56	1320
	Mean Value	36.49	23.22	854.50	343.47	419.04	909.75	8.20	2352.28
	CV/%	0.33	0.39	0.20	0.42	0.55	0.31	0.04	0.16
II	Maximum Value	62.12	43.97	1937.14	3295.08	637.84	1084.42	9.91	4784
	Minimum value	0.84	0.43	33.25	5.55	0.45	27.84	7.6	240
	Mean Value	12.17	6.32	1136.38	1683.09	86.06	687.12	8.47	3149.93
	CV/%	1.32	1.88	0.57	0.70	1.57	0.56	0.05	0.49
III	Maximum Value	84.65	55.91	943.14	475.35	832.67	1074.6	11.2	3033
	Minimum value	4.43	0.6	29.34	26.02	21.81	25.05	7.3	150
	Mean Value	37.25	20.49	526.70	242.96	285.73	558.58	8.64	1518.05
	CV/%	0.61	0.72	0.62	0.51	0.67	0.78	0.10	0.63

is $Cl^- > SO_4^{2-} > HCO_3^-$. The mass concentration of the main components of anions in coal measure water is $HCO_3^- > Cl^- > SO_4^{2-}$. From the coefficient of variation, The coefficient of variation of the lower water ion concentration is less than 1, indicating that the ions in the lower water are more concentrated and the aquifer is more stable. The variation coefficients of Mg^{2+}, Ca^{2+} and SO_4^{2-} ion concentrations in coal measure water are all greater than 1, indicating that the ion distribution in coal measure water is relatively uneven, and may be related to other aquifers outside, and the aquifer is not stable. The coefficient of variation of ion concentration in Taiyuan Formation limestone water is between 0.5 and 1. Compared with lower aquifer and coal measure water, the ion dispersion in limestone water of Taiyuan Formation is poor and the uniformity is low, which may be disturbed by external aquifer.

4.2. Principal Component Analysis of Original Data Based on SPSS

Pearson correlation can reveal the similarity and source consistency of chemical composition changes in groundwater sample data [12] [13]. Based on SPSS statistical analysis software, the correlation analysis of the main ion concentration(Such as Ca^{2+}, Mg^{2+}, K^+, Na^+, HCO_3^-, SO_4^{2-}, CO_3^{2-}, Cl^-), PH value and TDS value in the water sample data of the main aquifer in the middle of Panxie mining area was carried out. The results are shown in Table 3.

It can be seen from Table 3 that the correlation coefficients of $Na^+ + K^+$ and HCO_3^-, Cl^- and TDS in water chemical indicators are 0.765, 0.775 and 0.944, respectively, which are positively correlated and closely related. The correlation coefficient between the hydrochemical indexes shows that there is a positive correlation between Cl^- and other hydrochemical indexes in the central mining area of Huainan, and there is a negative correlation between PH value and other hydrochemical indexes, and TDS is greatly affected by $Na^+ + K^+$, HCO_3^- and Cl^-.

Using SPSS to reduce the dimension of the data, the total variance in Table 4 is first obtained. According to the cumulative variance contribution rate, four

Table 3. Discriminant indicators Pearson correlation coefficient matrix.

variable	Ca^{2+}	Mg^{2+}	$Na^+ + K^+$	HCO_3^-	SO_4^{2-}	Cl^-	PH	TDS
Ca^{2+}	1	0.704	−0.151	−0.429	0.405	0.264	−0.368	−0.097
Mg^{2+}	0.704	1	−0.120	−0.405	0.404	0.308	−0.363	−0.054
$Na^+ + K^+$	−0.151	−0.120	1	0.819	−0.025	0.736	−0.304	0.952
HCO_3^-	−0.429	−0.405	0.819	1	−0.348	0.277	−0.115	0.786
SO_4^{2-}	0.405	0.404	−0.025	−0.348	1	0.062	−0.097	0.040
Cl^-	0.264	0.308	0.736	0.277	0.062	1	−0.444	0.684
PH	−0.368	−0.363	−0.304	−0.115	−0.097	−0.444	1	−0.343
TDS	−0.097	−0.054	0.952	0.786	0.040	0.684	−0.343	1

Table 4. Explains the total variance table.

Serial Number	Initial Eigenvalue			Extract Square and Load		
	Total	Variance/%	Accumulate/%	Total	Variance/%	Accumulate/%
1	3.344	41.797	41.797	3.344	41.797	41.797
2	2.566	32.070	73.867	2.566	32.070	73.867
3	0.838	10.478	84.344			
4	0.549	6.863	91.208			
5	0.352	4.395	95.603			
6	0.294	3.678	99.281			
7	0.054	0.671	99.952			
8	0.004	0.048	100.000			

principal components are determined to represent the main hydrochemical indexes in the main aquifer of the mining area, and the cumulative contribution rate is about 91.208%. According to the principal component score coefficient matrix, the correlation calculation equation between the new discriminant index Y_1, Y_2, Y_3, Y_4 and the hydrochemistry index can be obtained (the correlation coefficient matrix is multiplied by the standardized data matrix):

$$\begin{cases} Y_1 = -0.115X_1 - 0.106X_2 + 0.548X_3 + 0.456X_4 - 0.018X_5 + 0.407X_6 - 0.133X_7 + 0.534X_8 \\ Y_2 = 0.539X_1 + 0.522X_2 + 0.047X_3 - 0.248X_4 + 0.382X_5 + 0.306X_6 - 0.358X_7 + 0.074X_8 \\ Y_3 = -0.096X_1 - 0.131X_2 + 0.118X_3 - 0.191X_4 + 0.690X_5 + 0.096X_6 + 0.653X_7 + 0.108X_8 \\ Y_4 = 0.006X_1 + 0.408X_2 + 0.035X_3 - 0.161X_4 - 0.538X_5 + 0.473X_6 + 0.538X_7 - 0.060X_8 \end{cases} \quad (1)$$

4.3. Water Inrush Source Identification Based on PCA-BP Neural Network

In this study, there were 96 sets of data, 76 sets of data as training samples, and 20 sets of data as prediction samples. Before the beginning of BP neural network training, in order to eliminate the adverse effects caused by singular sample data, it is necessary to normalize the samples and limit the data to the range of [0,1] or [−1, 1]. The normalized data can accelerate the speed of gradient descent to find the optimal solution, improve the training accuracy of the network, and avoid the numerical complexity in the calculation process. The normalization results of the sample data are shown in Table 5.

The neural network structure established in this paper is an input layer-hidden layer-output layer structure. The number of nodes in the input layer is the input vector dimension, and the number of nodes in each layer is the number of neurons in each layer. The main aquifer water sample data in this paper have a total of 8 dimensions. After the principal component extraction in the previous section, the principal component score matrix is obtained. After normalization, the network input vectors Y_1, Y_2, Y_3, and Y_4 are obtained, indicating that the number of neurons in the input layer is 4.

Table 5. Table of principal component factor.

Serial Number	Y_1	Y_2	Y_3	Y_4	Aquifer
1	2.017	1.093	0.037	−1.335	I
2	0.322	0.998	−0.248	−0.302	I
3	0.488	0.174	−1.095	−0.544	I
4	0.113	0.367	−0.443	0.094	I
…	…	…	…	…	…
93	0.242	0.860	−0.373	−0.711	III
94	−2.419	−0.246	−0.514	−0.158	III
95	0.234	0.701	0.538	−0.950	III
96	−1.876	−2.293	−0.197	−0.836	III

The selection of the number of hidden layer neurons is often the most important. In this paper, the number of input layer nodes is 4, the number of output nodes is 3, and the number of hidden layer nodes is 10 selected by empirical formula. The toolbox sets 70% of the sample as the training set, 15 % of the sample as the verification set, and 15 % of the sample as the test set. From the output training error curve, it can be found that when the number of iterations is 28 and the network training error is 0.036408, the accuracy reaches the training standard. As shown in **Figure 2**, it can be found that in the fitting data completed by BP neural network training, the R value of the training set is 0.82478, the R value of the verification set is 0.91443, and the R value of the test set is 0.40698. After the training, the dynamic system simulation function predict = sim (net, C) is called in the editor to get the simulation prediction results(C is the forecast data). Comparing the simulation results with the actual results, the No.64 coal measure water in Pansan Mine was misjudged as limestone water. The overall number of misjudged samples was 1, and the accuracy of water source identification was 95%.

5. Conclusion

The principal component analysis of 96 groups of data was carried out by SPSS25.0 software to determine the principal component score coefficient, and four new discriminant indexes were obtained to replace the original hydrochemical indexes. BP neural network training and prediction is carried out by Matlab R2021b software, in which the number of misjudged samples is 1, the accuracy of water source identification is 95%, and the effect of water inrush source identification is very good. It can accurately and quickly identify the water source of mine water inrush and realize the early warning analysis of water inrush accident. However, the established PCA-BP neural network model only considers the hydrochemical characteristics. In the subsequent research, the factors affecting water level, water temperature and water inflow can be comprehensively considered to

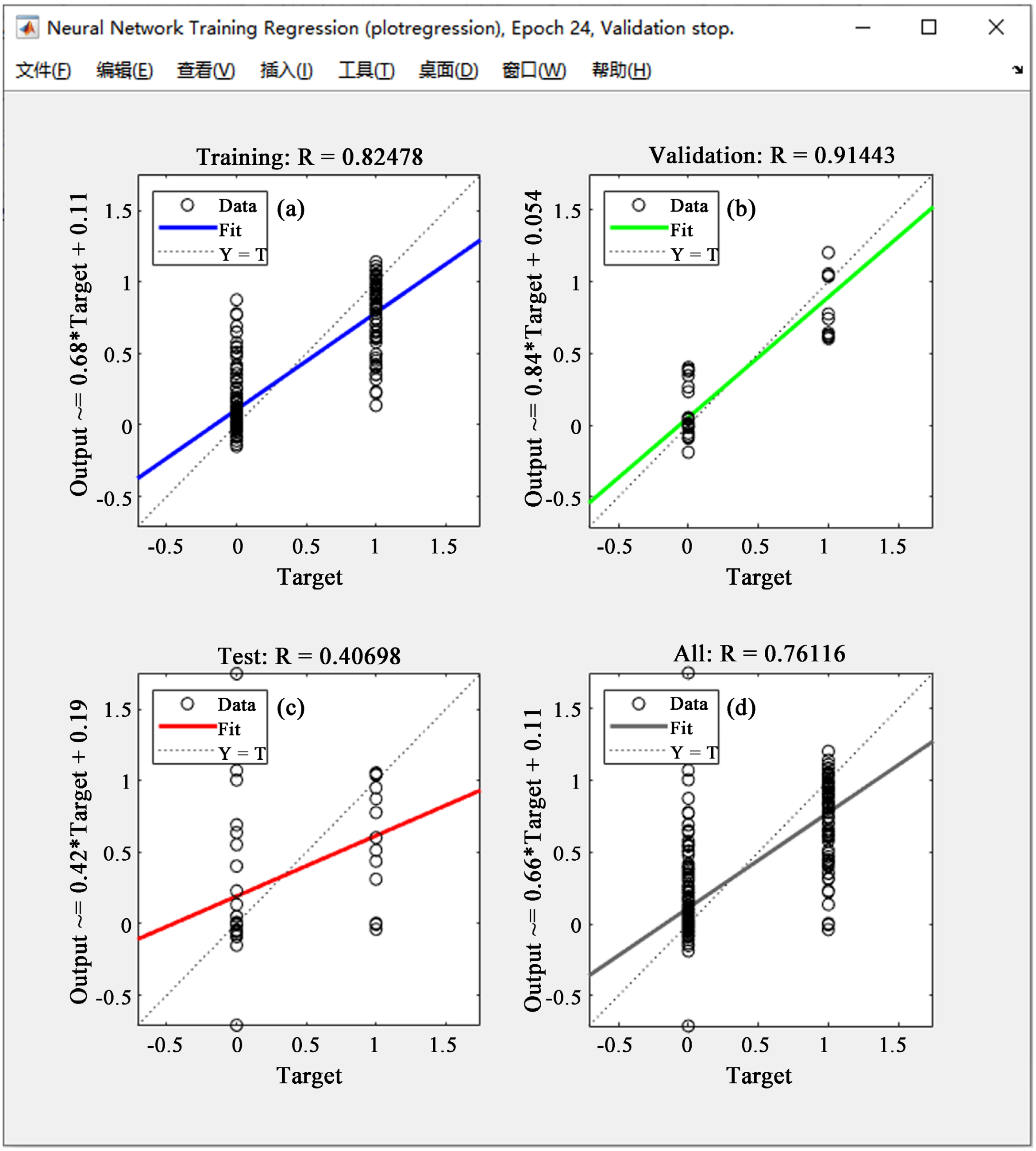

Figure 2. Sample regression analysis.

improve the water source determination model and improve the discrimination accuracy.

Acknowledgements

Part of the data in this paper is from Shaomeng Zhu, who studies at Anhui University of Science and Technology, and I express thanks to him.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] China Coal Industry Association (2021) "14th Five-Year" High-Quality Development Guidance for Coal Industry. *China Coal Association Political Research*, No. 19, 1.

[2] Yang, Y.G. and Huang, F.C. (2007) Application of Nonlinear Method in Mine Water Inrush Source Identification. *Journal of China University of Mining and Technology*, **36**, 283-286.

[3] Li, F.F., Zhang, F.B., Niu Z.R., Li, Y., Cui, J.Q. and Su B.Z. (2015) Analysis and Discrimination of Water Source of Water Drilling Hole by Set Intersection Method. *Mining Safety and Environmental Protection*, **42**, 99-102.

[4] Zhang, H., Xing, H.F., Yao, D.X., Liu, L.L., Xue, D.R. and Guo, F. (2019) The Multiple Logistic Regression Recognition Model for Mine Water Inrush Source Based on Cluster Analysis. *Environmental Earth Sciences*, **78**, Article No. 612. https://doi.org/10.1007/s12665-019-8624-2

[5] Zhai, X.R., Tian, N.C., Zhang, H.M., Wu, J.W., Ren, Z.Q. and Gong, S.L. (2016) Identification of Mine Water Recharge Relationship Based on Water Chemical Composition and Cluster Analysis. *Chinese Journal of Geological Disasters and Prevention*, **27**, 88-92.

[6] Baborowski, M., Simeonov, V. and Einax, J.W. (2012) Assessment of Water Quality in the Elbe River at Flood Water Conditions Based on Cluster Analysis, Principle Components Analysis, and Source Apportionment. *CLEAN—Soil*, *Air*, *Water*, **40**, 373-380. https://doi.org/10.1002/clen.201100085

[7] Kumar, M., Ramanathan, A. and Rao, M.S. (2006) Identification and Evaluation of Hydrogeochemical Processes in the Groundwater Environment of Delhi, India. *Environmental Geology*, **50**, 1025-1039. https://doi.org/10.1007/s00254-006-0275-4

[8] Li, X.F., Xu, J.P., Wang, Y.Q. and He, C.Z. (2004) The Establishment of Self-Adapting Algorithm of BP Neural Network and Its Application. *Systems Engineering—Theory & Practice*, **5**, 1-8.

[9] Jiang, S., Sun, Y.J., Yang, L. and Ling, C.P. (2007) Prediction of Mine Water Inflow Based on BP Neural Network Method. *China Coal Geology*, **19**, 38-40.

[10] Zhang, Y. (2015) Research on Water Quality Evaluation of Songhua River Harbin Section Based on Principal Component Analysis-BP Neural Network Method. Harbin Normal University, Harbin.

[11] Chu, T.T. (2014) Conventional Hydrochemical Analysis of Groundwater in Panxie Mining Area and Establishment of Discriminant Model. University of Science and Technology of China, Hefei.

[12] Wang Z.R., Tian, X. and Wu, X. (2018) Hydrochemical Characteristics and Quality Assessment of Shallow Groundwater and CBM Coproduced Water in the Shizhuangnan Block, Qianshui Basin, China. *Environmental Earth Sciences*, **77**, Article No. 57. https://doi.org/10.1007/s12665-017-7212-6

[13] Liu, H., Bai, H. and Tian, G.L. (2019) Spatial-Temporal Evolution Characteristics of Groundwater Hydrochemistry in Kuye River Basin, Shenmu City. *People's Yellow River*, **41**, 76-81+87.