METHODOLOGY

Construction and application of mine water inflow prediction model based on multi-factor weighted regression: Wulunshan Coal Mine case

Bo Li[1,2] · Huang Wu[1] · Pu Liu[2] · Juan Fan[3] · Tao Li[4]

Abstract

Accurate prediction of the amount of water inflows in mines is of great significance for safe production in mining. To improve the accuracy of the prediction, based on the analysis of the hydrological and geological conditions of the mine, the main factors affecting the water inflows in mines were determined. Using entropy method, the weight values of the factors affecting the water inflows in mines were calculated, and the non-linear regression fitting between the water inflows and various factors was carried out using multiple regression theory and MATLAB function programming. Combining with the factor weights determined by the entropy method, a weighted non-linear regression prediction model for water inflows in mines was established. The model not only takes into account the fact that the water inflows in mines are affected by multiple factors, but also reflects the characteristic that the importance of factors is different. By comparing with the multiple linear regression prediction model and the measured water inflows, it is proved that the weighted non-linear regression prediction model for water inflows in mines can overcome the defects of existing methods, minimize the prediction error caused by low degree of hydrological and geological exploration, and improve the prediction accuracy.

Keywords Mine water inflow · Multivariate nonlinear regression · Multiple factors · Entropy weighting method · Weights

Introduction

Coal is the most abundant and most widely distributed fossil fuel on the earth. As an important power for the rapid development of the global economy, coal resources are increasingly needed (Wang et al. 2015a, b). Meanwhile, the hydrogeological problems facing coal mining become increasingly prominent. Mine water inflow refers to the quantity of surface water or groundwater pouring into a well lane system within unit time through fissures, faults or other channels in the process

Communicated by: H. Babaie

✉ Bo Li
libo1512@163.com

1   Key Laboratory of Karst Georesources and Environment, Ministry of Education, Guizhou University, Guiyang 550025, Guizhou, China

2   College of Resource and Environmental Engineering, Guizhou University, Guiyang 550025, Guizhou, China

3   School of Geology and Environment, Xi'an University of Science and Technology, Xian 710054, Shanxi, China

4   School of Mines and Civil Engineering, Liupanshui Normal University, Liupanshui 553004, China

of mining (Xu and Gong 2011; Dong et al. 2021). In order to ensure the production safety, mine water inflow prediction is one of the priorities in water hazards prevention and control in both the prospecting stage and the mine construction and production stage (Wu et al. 2013; Singh and Atkins 1985). Mine water inflow is directly related to the rationality of coal mining scheme and drainage capacity designs, and more importantly, determines whether coal mining is safe (Li et al. 2022a; Hu and Zhao 2021; Polak et al. 2016). In order to make the prediction results of mine water inflow more consistent with reality, many researchers have conducted a lot of studies on mine water inflow prediction methods, establishing two types of prediction methods: uncertainty analysis methods and deterministic mathematical models. The uncertainty methods mainly include correlation analysis (Qiu et al. 2020), support vector machine (Li et al. 2010), neural network (Zuo et al. 2011) and grey system theory (Wang et al. 2015a, b; Ma and Bai 2015; Xu et al. 2012). The deterministic methods mainly include numerical method (Li et al. 2015; Wu et al. 2019; Bai et al. 2021; Krukovska and Vynohradov 2019; Bouw and Morton 1987), and analytical method (Hou 2012; Li et al. 2014). Miladinović et al. (2015) used a linear correlation regression model to prediction and correction on the mine water inflow of the Štavalj Coal Mine in southwestern Serbia. Wei et al. (2011)

built a water inflow prediction model based on support vector machines, combined with the practical demands to predict the water inflow at the new working face, and the results were verified during the mining process of the new working face. Shao et al. (2014) established a non-linear artificial neural network prediction model and predicted the normal mine water inflow during mine operation, and the results were consistent with the actual predicted data. Ma et al. (2020) established two exponentially weighted moving average modified gray water inflow models optimized by particle swarm optimization, and obtained water inflow prediction equations based on actual data from Buliangou coal mine. Guo et al. (2009) established a three-dimensional numerical model of COSFLOW and simulated the water inflow of two mines in Australia. Singh et al. (2012) used the SEEP/W finite element software package to predict the water inflow from surface mining excavation, and compared the prediction results with the analytical solution for verification. Li et al. (2021) combined Monte Carlo methods and FLAC3D to generate a discrete fracture network, established an optimized water inflow prediction model based on the fluid–solid coupling method, and proposed an analytical formula for water inflow prediction. Chen et al. (2015) used the big water well method and a three-dimensional numerical model to predict the water inflow from the roof sandstone aquifer in coal seam mining, and compared and analyzed the results of the prediction methods. Zhang et al. (2017) used the no. 2 coal seam in the Pingdingshan No. 10 coal mine as an example, three methods (analogue, big well, and numerical simulation) were used to forecast mine water inflow and their performance. The existing mine water inrush prediction methods generally emphasize a single factor and do not consider multiple factors that affect the occurrence and control of mine water inrush. They also do not integrate the mechanisms, positions and progress of mine water inrush into the mining engineering process, resulting in a prediction of the mine water inrush as a static quantity instead of a dynamic changing quantity that changes with the mining engineering. In recent years, with the complexity of global mining conditions, especially the increasing depth of mining, the complexity of mine hydrological and geological conditions often exceeds the empirical range, making some methods no longer applicable. Based on the analysis of typical coal mine hydrological and geological conditions, this paper determines the main factors affecting the mine water inrush. Using multivariate regression theory and MATLAB function programming, a multivariate nonlinear regression fitting of the mine water inrush and various factors was carried out, combined with the factor weight determined by the entropy value method, and a weighted mine water inrush multivariate nonlinear regression prediction model was constructed. The established prediction model compensates for the flaw of previous prediction methods that did not consider the differences in the importance of each factor and can minimize the prediction error caused by low survey level or lack of hydrological and geological parameters.

Overview of the research area

Physical geography

Wulunshan Coal Mine is located in Shuguang Town, Nayong County, Guizhou Province, China. Its geographic coordinates are $105°16'01'' \sim 105°20'35''$E and $26°34'59'' \sim 26°40'15''$N. The mine lot is 9.6 km long from south to north, and 4.6 km wide on average, covering an area of 44.02km^2 (the geographic location is shown in Fig. 1). The Wulunshan Coal Mine is characterized by a plateau-middle mountainous topography, with an elevation ranging from 1,500 m to 2,000 m and a relative height difference of 300 m to 500 m. The terrain is overall higher in the northwest and lower in the southeast. The dominant landform includes river valleys and gullies. The main rivers within the mine area include the Shuigong River, the Sancha River, and the gullies on either side, which are distributed in a branching pattern and belong to the Wujiang River System.

Geological condition

(1) Strata and coal seam

The disclosed strata of the mine lot are sorted from old to new: Middle Permian Maokou Formation (P_2m), Upper Permian Emeishan Basalt Formation ($P_3\beta$), Upper Permian Longtan Formation (P_3l), Upper Permian Changxing Formation (P_3c), Lower Triassic Feixianguan Formation (T_1f), Lower Triassic Yongningzhen Formation (T_1yn) and Quaternary (Q). The coal-containing strata are Longtan Formation, composed of siltstone, fine sandstone, bioclastic limestone, mudstone, and coalbed. The potential mining strata are No.3, No.5^{-3}, No.6^{-3}, No.8, and No.33, and the major mining strata are No.3 and No.8 (see Fig. 2).

(2) Geological structure

Wulunshan Coal Mine is located in the south section of Jiaga anticline and the west wing of the Shuigonghe syncline. The whole is a monoclinic structure, superposition of secondary anticlines and synclines. The stratum towards is $130 \sim 160°$, leaning to the northeast. The shallow dip angle is sharp, $25 \sim 40°$; the deep dip angle gradually becomes gentle, $5 \sim 20°$. The faults within the mining area are mainly high angle normal faults with an inclination of $66 \sim 80°$ and fault throw of about $10 \sim 25$ m (see Fig. 3).

Hydrogeological condition

The main groundwater source in the studied area is rainfall, and the groundwater level is controlled by terrain and rainfall. The water can be divided into carbonate karst water,
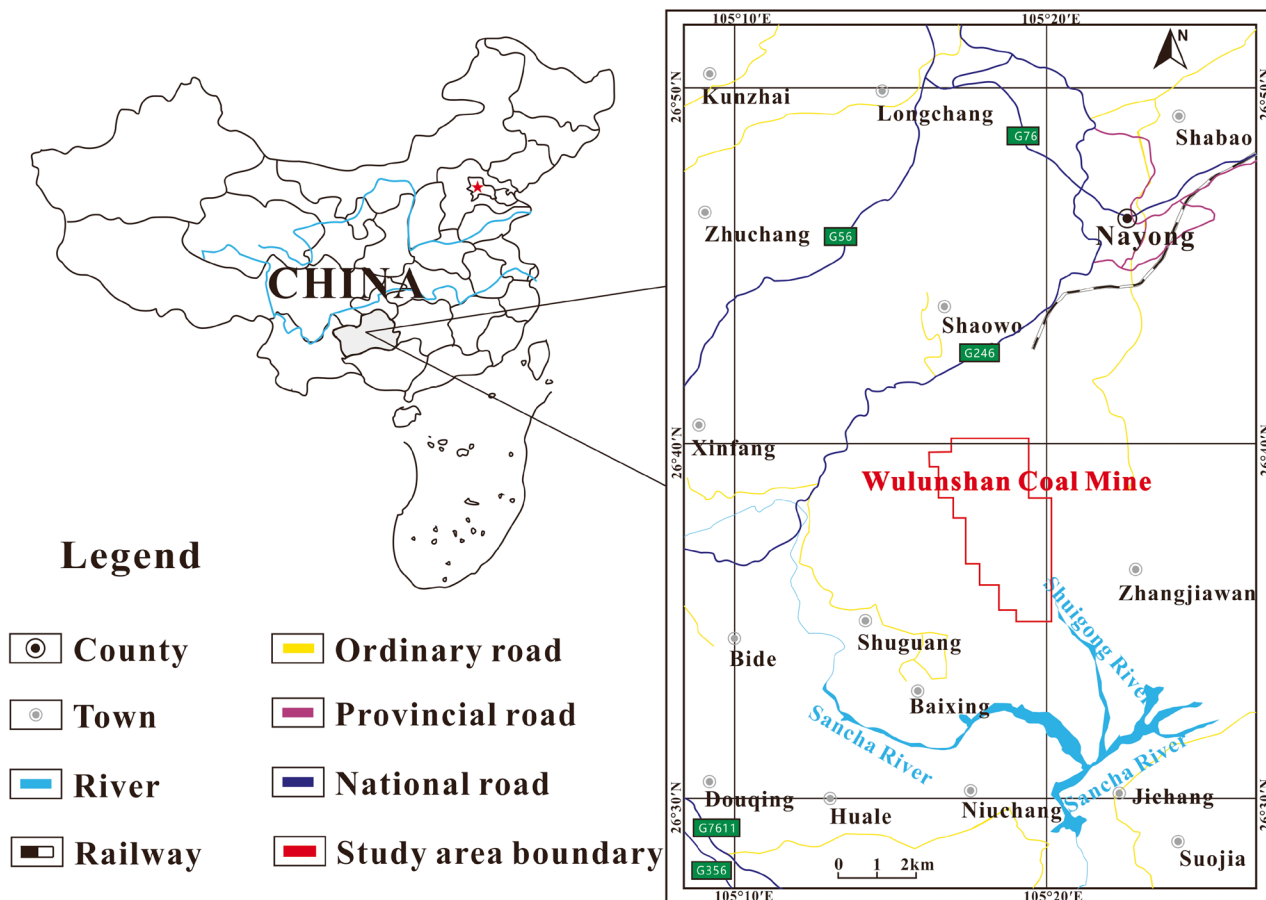
Fig. 1 Location of Wulunshan Coal Mine

clasolite fissure water and loose rock pore water. No.2 (T_1f^2) and No.4 (T_1f^4) sections of Feixianguan Formation and Maokou Formation (P_2m) are thick limestone strata, which feature evident dissolution, developed karst caves, dissolution pores and grikes, and strong water yield capacity. As a result of the Zk1 drilling water pumping test, the unit water inflow is 0.052L/(s m), the water temperature is $13 \sim 16℃$, and the pH is $6.8 \sim 7.74$. The rest of the clasolite strata feature thin limestone, underdeveloped karst and weak water yield capacity. The scattered Quaternary (Q) residual diluvial loose rock formations feature strong water permeability and moderate water yield capacity. The direct sources of water filling for coal mining are the fissured aquifers of Longtan Formation (P_3l) and Changxing Formation (P_3c), and the indirect sources of water filling are the karst water aquifers of Feixianguan Formation No.2 section (T_1f^2) and Maokou Formation (P_2m). The hydrogeological profiles are shown in Fig. 4. In mines where the coal seams are surrounded by impermeable rock formations, water inflow can occur through fractures and faults in the rock, which can serve as pathways for water to enter the mine. If the water is

under high pressure, it can cause increased stress on the surrounding rock, leading to the development of new fractures or the expansion of existing ones. This can increase the risk of rock falls and other types of rock instability, which can compromise the safety of the mine and its workers.

Model data for training and validation

In this study, the data used for model training and validation was collected from Wulunshan Coal Mine. The data used in this study was collected from 65 time points between 2013 and 2017 (Table S1). Of these 65 time points, 50 were used for training and 15 were used for validation. The data types included in this study are precipitation, aquifer thickness, mining area, mining depth, mining thickness, driving footage, and water inflow. These data types were selected for their potential impact on water inflow into the mine and their contribution to the overall understanding of the system. The collection of this data allowed for the creation of models that can be used to make predictions about water inflow into the mine.

Fig. 2 Bar graph of the relationship between strata and coal seam

Chronostratigraphic			Stratum code	Lithology columnar	Thickness(m) Min~Max / Average	Lithology characteristics	Remarks
System	Series	Formation					
Quaternary			Q		0~10	Loam, sandy-loam	Aquifer
Triassic	Lower Triassic Series	Yongningzhen Formation	T_1yn		49.00	Gray limestone, argillaceous limestone	Aquiclude
		Feixianguan Formation	T_1f^6		89.72~112.31 / 101.00	Fuchsia argillaceous siltstone	Aquiclude
			T_1f^4		57.1~70.59 / 65.28	Gray limestone, siltstone	Aquifer
			T_1f^3		163.00~186.83 / 170.94	Grey-purple siltstone, marl	Aquiclude
			T_1f^2		45.59~83.03 / 66.04	Gray limestone, bioclastic limestone	Aquifer
			T_1f^1		156.29~212.62 / 194.29	Grey-green argillaceous siltstone, mudstone	Aquiclude
Permian	Upper Permian Series	Changxing Formation	P_3c		21.80~86.37 / 42.86	Dark gray fine sandstone, mudstone	Aquiclude
		Longtan Formation	P_3l		310.00~426.00 / 344.00	Gray, gray-black fine sandstone, siltstone, mudstone, bioclastic limestone Included No. 3, No. 5^{-3}, No. 6^{-3}, No. 8, and No. 33 coal seams	Aquiclude
		Emeishan Basalt Formation	$P_3\beta$		94.16~272.16 / 180.00	Dark gray basalt	Aquiclude
	Middle Permian Series	Maokou Formation	P_2m		439.02~445.67 / 440.00	Light gray limestone, bioclastic limestone	Aquifer

Correlation analysis and weight determination of the influencing factors

Mine water inflow is closely related to geological structure, engineering, hydrogeological condition, etc. Moreover the relationships between water inflow and the influencing factors often present a highly nonlinear complex relationship (Qiu et al. 2017; Liu et al. 2018; Li and Zhou 2006; Shi et al. 2017). Predecessors have done a lot of related research on this, representative as Wu et al. (2017) used the vulnerability index method, which couples GIS with the analytic hierarchy process, to evaluate the water inrush risk of the Gushuyuan coal mine No. 15 seam. Li et al. (2022b) used grey relational analysis and analytic hierarchy process to establish an evaluation
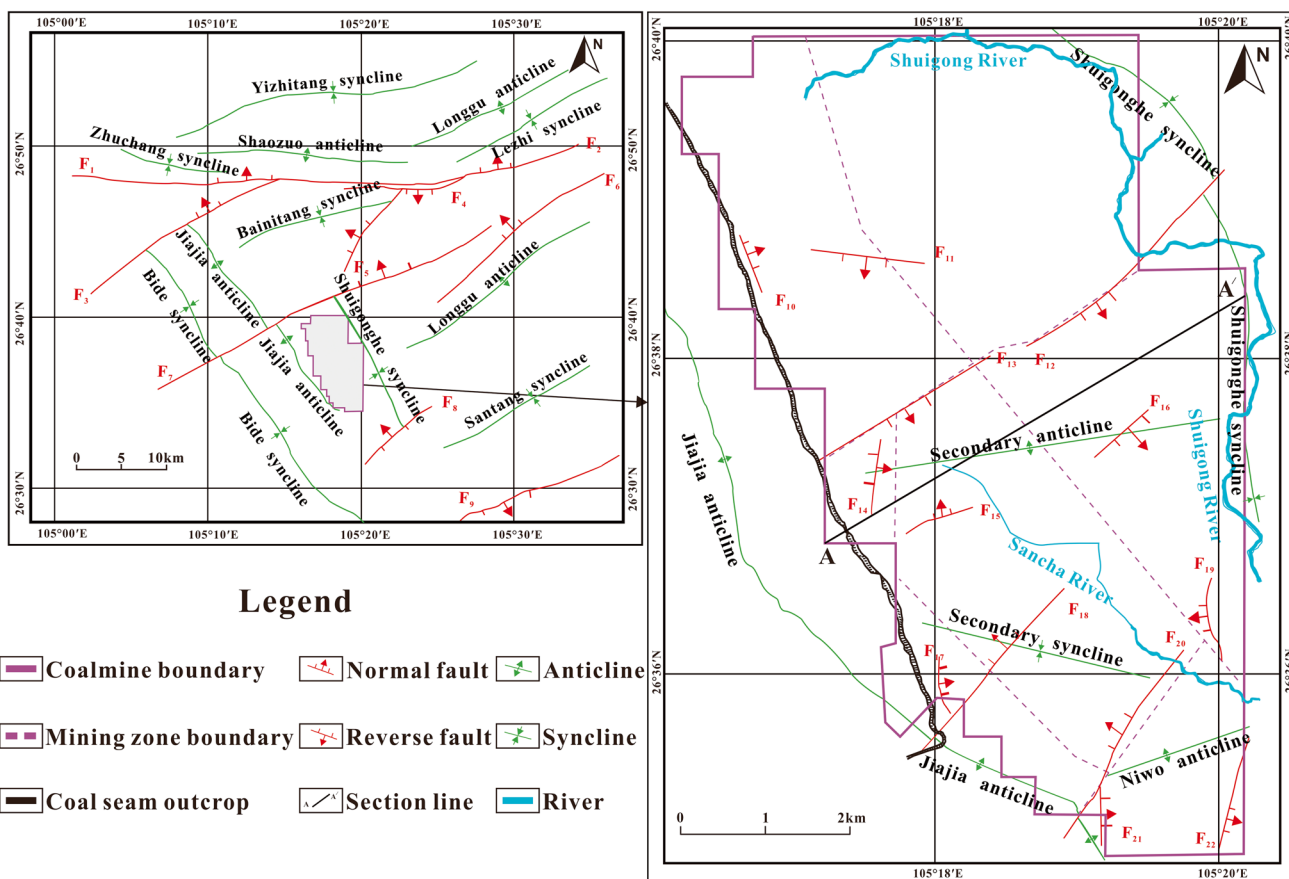
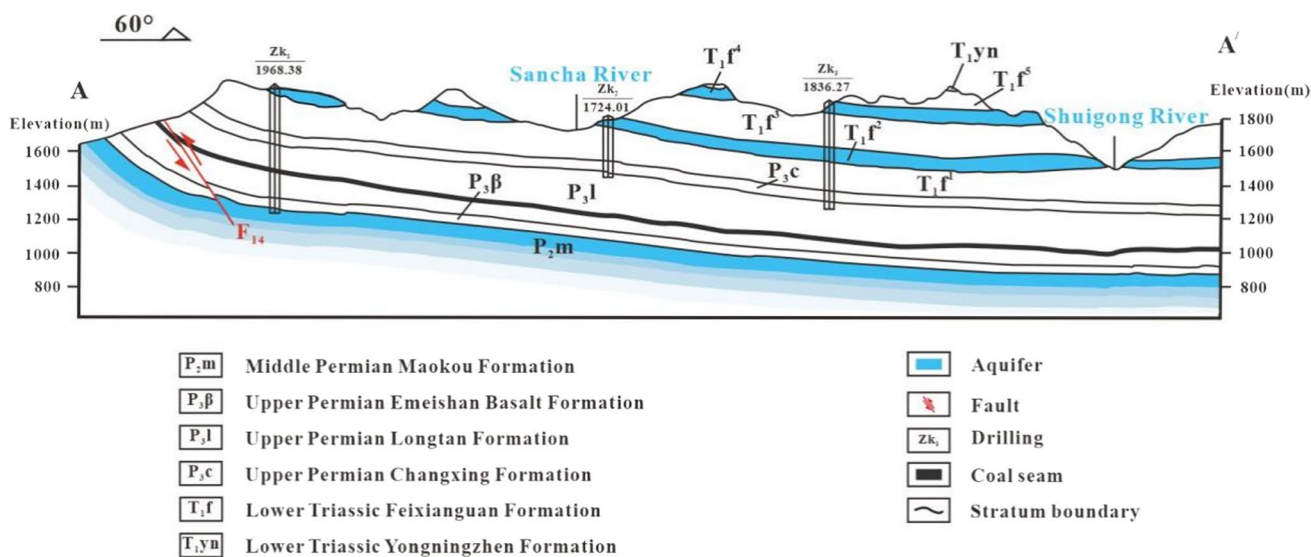**Fig. 3** Distribution of geological structure

Fig. 4 Hydrogeological profiles

model for water inrush from the coal floor. They applied it to the typical working face of Yuzhou coalfield in north China to demonstrate the evaluation process. This paper, based on the results of previous studies and the hydrogeological and geological conditions of Wulunshan Coal Mine, determines the influencing factors of water inflow are precipitation, aquifer thickness, mining area, mining depth, mining thickness and driving footage. Besides, the author collected 65 groups of monthly water inflow data from the Wulunshan Coal Mine from July 2012 to November 2017, 50 groups of which serve as training samples for the prediction model, and 15 of which are used for testing the prediction results of the prediction model, as shown in Table S1 (supplementary material).

Correlation analysis theory

In the early twentieth century, the British statistician Pearson put forward a coefficient for calculating the linear correlation between two variables, called Pearson's correlation coefficient (Gross 1975; Katsaounis 2004). The coefficient is hereby used to analyze the correlation between mine water inflow and each influencing factor, as shown in Eq. (1) (Fiorillo and Doglioni 2010; Liu et al. 2019a, b).

$$r = \frac{\sum_{i=1}^n (x_i - \frac{1}{n}\sum_{i=1}^n x_i)(y_i - \frac{1}{n}\sum_{i=1}^n y_i)}{\sqrt{\sum_{i=1}^n (x_i - \frac{1}{n}\sum_{i=1}^n x_i)^2}\sqrt{\sum_{i=1}^n (y_i - \frac{1}{n}\sum_{i=1}^n y_i)^2}}, r^2 \leq 1 \tag{1}$$

Where x_i is mine water inflow; y_i is a factor affecting mine water inflow; r is the correlation coefficient.

Based on 65 groups of measured data about the mine water inflow of Wulunshan Coal Mine from July 2012 to November 2017, a correlation analysis is conducted between mine water inflow and each influencing factor. The results are shown in Table 1.

According to Table 1, the correlation coefficients between mine water inflow and influencing factors are between -0.76 and 0.61. Mine water inflow is positively correlated to precipitation, aquifer thickness, mining area, mining depth and mining thickness, and negatively correlated to driving footage.

Entropy method theory

The entropy method is an objective assignment method. Each influencing factor is weighted by judging the dispersion degree of an influencing factor and the degree of deviation between data and the characteristics of the data itself (Xue et al. 2021; Xu et al. 2020). There are three main steps to determine the weight of an influencing factor using the entropy method:

The raw data matrix is normalized. Suppose the original data matrix of m influencing factors and n months is $A = [a_{ij}]_{m \times n}$, standardize it to get Eq. (2) for influence factors positively correlated with mine water inflow, and Eq. (3) for influence factors negatively correlated with mine water inflow:

$$r_{ij} = \frac{a_{ij} - \min_j\{a_{ij}\}}{\max_j\{a_{ij}\} - \min_j\{a_{ij}\}} \tag{2}$$

$$r_{ij} = \frac{\max_j\{a_{ij}\} - a_{ij}}{\max_j\{a_{ij}\} - \min_j\{a_{ij}\}} \tag{3}$$

Where $\min_j\{a_{ij}\}$ is the minimum value of the ith influencing factor; $\max_j\{a_{ij}\}$ is the maximum value of the ith influencing factor.

Information entropy is defined. The information entropy of the ith influencing factor is as shown in Eq. (4).

$$E_i = -\frac{1}{\ln n}\sum_{j=1}^n \left(\frac{r_{ij}}{\sum_{j=1}^n r_{ij}}\right)\ln\left(\frac{r_{ij}}{\sum_{j=1}^n r_{ij}}\right) \tag{4}$$

Where n is the total number of months; r_{ij} is the standard value of the ith influencing factor in the jth month.

The weight is defined. The weight of the ith influencing factor is as shown in Eq. (5).

$$Wi = \frac{1 - Ei}{m - \sum_{i=1}^m Ei} \tag{5}$$

Table 1 Correlation coefficient matrix of influencing factors

Indicators	Precipitaion	Aquifer thickness	Mining area	Mining depth	Mining thickness	Driving footage	Water inflow
Precipitation	1.00						
Aquifer thickness	0.61	1.00					
Mining area	0.30	0.28	1.00				
Mining depth	0.27	0.07	0.09	1.00			
Mining thickness	-0.66	-0.76	-0.35	-0.12	1.00		
Driving footage	-0.05	-0.04	-0.06	-0.05	-0.01	1.00	
Water inflow	0.21	0.09	0.15	0.21	0.04	-0.04	1.00

Table 2 The weights of influencing factors

Factors	Precipitation	Aquifer thickness	Mining area	Mining depth	Mining thickness	Driving footage
W_i	0.36	0.14	0.27	0.06	0.13	0.04

where $0 \leq w_i \leq 1$ and $\sum_{i=1}^{m} w_i = 1$; m is the total number of influence factors; E_i is the information entropy of the ith influence factor.

The weights of factors influencing are determined using entropy method, as shown in Table 2.

It can be concluded from Table 2 that the influencing factors of mine water inflow for Wulunshan Coal Mine can be sorted by weight: precipitation > mining area > aquifer thickness > mining thickness > mining depth > driving footage.

Building of a weighted multiple nonlinear regression prediction model for mine water inflow

Building of multiple linear regression prediction model

Multiple regression analysis studies the relationship between one dependent variable and multiple independent variables based on the given values of multiple explanatory variables (Cohen 1968; Liu et al. 2019a, b). The functional expression of multiple regression analysis can be either linear or nonlinear depending on the causal relationship between the independent or dependent variable (Ouedraogo et al. 2019). MATLAB function programming is used to realize multiple linear fitting between water inflow and precipitation, aquifer thickness, mining area, mining depth, mining thickness and driving footage. The fitting parameters are shown in Table 3.

According to Table 3, the multiple linear regression coefficients between water inflow and influencing factors are 0.05, 0.04, 1.21, 0.02, 12.43, and -0.01. The equation of the

Table 3 Fitting parameters for multiple linear regression

Model	B	Std.Error	t	Sig
(Constant)	-23.15	26.37	-0.88	0.39
P	0.05	0.07	0.76	0.45
M	0.04	0.04	1.00	0.32
A	1.21	0.89	1.36	0.18
D	0.02	0.02	1.01	0.32
T	12.43	7.72	1.61	0.12
L	-0.01	0.01	-0.78	0.44

water inflow multiple linear regression prediction model is therefore obtained, as shown in Eq. (6).

$$Q = 0.05P + 0.04M + 1.21A + 0.02D + 12.43T - 0.01L - 23.15 \tag{6}$$

Where Q is water inflow(m³/h); P is precipitation(mm); M is aquifer thickness;(m) A is mining area(10³m²); D is mining depth(m); T is mining thickness(m); L is driving footage(m).

Building of weighted multiple nonlinear regression prediction model

A scatter plot between influencing factors and water inflow shows a highly nonlinear relationship. MATLAB functional programming fitting is used to determine the unary nonlinear fitting curves between water inflow and influence factors. As shown in Figs. 5, 6, 7, 8, 9 and 10.

Based on the MATLAB function fitting curves, the influencing factor function fitting indicators are Sum of Squares due to Error (SSE), Coefficient of Determination (R^2), Adjusted Coefficient of Determination (Adjusted R^2), and Root Mean Square Error (RMSE). The function fitting indicators are shown in Table 4.

The unary nonlinear regression function relational expressions between water inflow and influencing factors are determined by function fitting indicators and function fitting curves. As shown in Eq. (7) to Eq. (12).

$$Q = a_1 P^3 + a_2 P^2 + a_3 P + a_4 \tag{7}$$
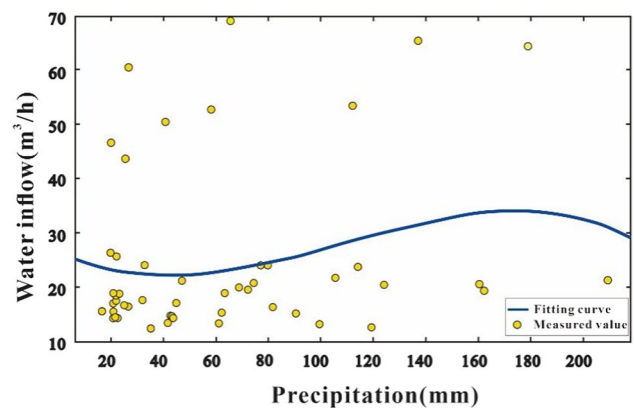
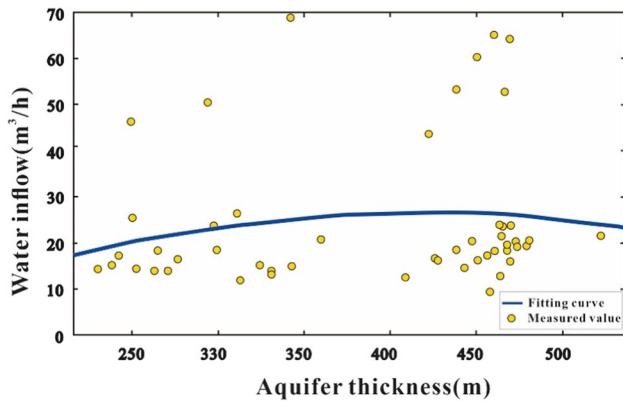**Fig. 5** Unary nonlinear fitting curve between precipitation and water inflow

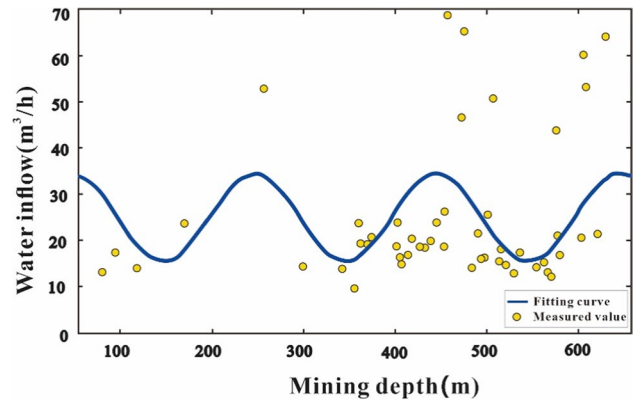**Fig. 6** Unary nonlinear fitting curve between aquifer thickness and water inflow

Fig. 9 Unary nonlinear fitting curve between mining depth and water inflow
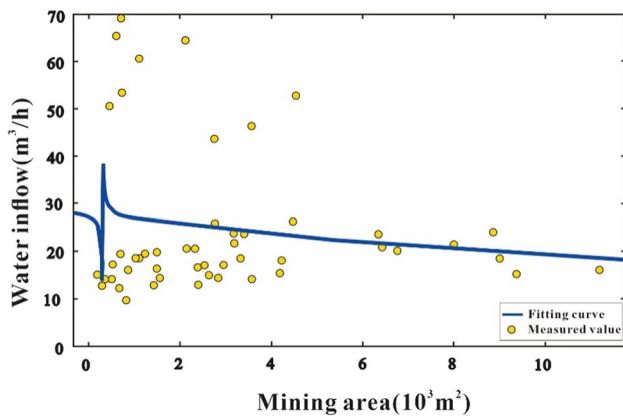

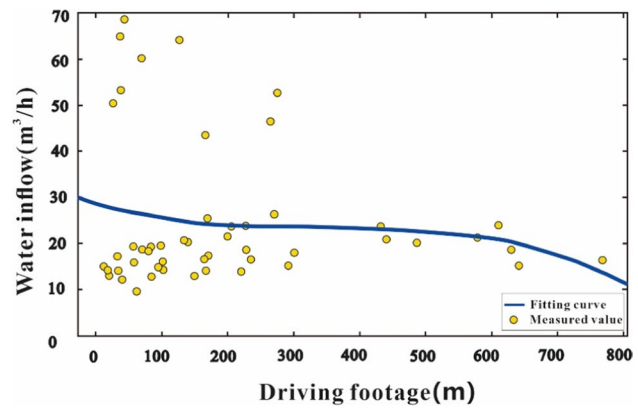**Fig. 7** Unary nonlinear fitting curve between mining area and water inflow

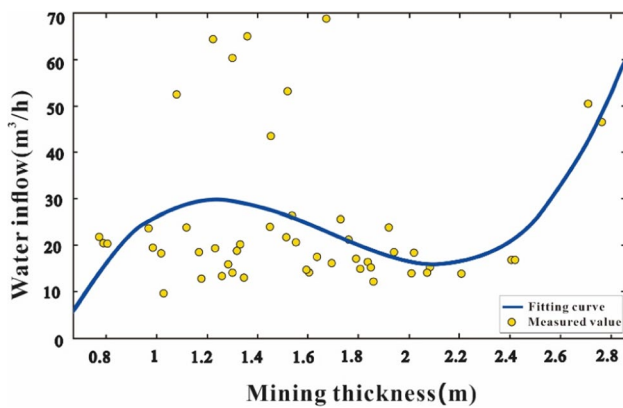**Fig. 10** Unary nonlinear fitting curve between driving footage and water inflow

Fig. 8 Unary nonlinear fitting curve between mining thickness and water inflow

Table 4 Function fitting indicators

Factors	SSE	R^2	Adjusted R^2	RMSE
P	11,546.51	0.04	-0.02	15.84
M	11,810.66	0.02	-0.02	15.85
A	11,607.93	0.04	-0.02	15.89
D	10,001.45	0.17	0.12	14.75
T	9988.57	0.17	0.12	14.74
L	11,790.57	0.02	-0.04	16.01

$$Q = b_1 \sin(b_2 M + b_3) \tag{8}$$

$$Q = \frac{c_1 A + c_2}{A^2 + c_3 A + c_4} \tag{9}$$

$$Q = d_1 + d_2 \cos(\lambda D) + d_3 \sin(\lambda D) \tag{10}$$

$$Q = e_1 T^3 + e_2 T^2 + e_3 T + e_4 \tag{11}$$

$$Q = f_1 L^3 + f_2 L^2 + f_3 L + f_4 \tag{12}$$

Where Q is water inflow(m^3/h); P is precipitation(mm); M is aquifer thickness (m); A is mining area (10^3m^2); D is mining depth (m); T is mining thickness (m); L is driving footage(m); a, b, c, d, e, f and λ are parameters to be solved.

After obtaining the unary nonlinear regression function relationships, to fully consider the important difference of the influencing factors to water inflow, the unary nonlinear regression function relationships of the influencing factors are weighted and summed. As shown in Eq. (13).

$$Q = 0.36(a_1 P^3 + a_2 P^2 + a_3 P + a_4) + 0.14(b_1 sin(b_2 M + b_3))$$
$$+ 0.27(\frac{c_1 A + c_2}{A^2 + c_3 A + c_4}) + 0.06(d_1 + d_2 cos(\lambda D) + d_3 sin(\lambda D)) \tag{13}$$
$$+ 0.13(e_1 T^3 + e_2 T^2 + e_3 T + e_4) + 0.04(f_1 L^3 + f_2 L^2 + f_3 L + f_4)$$

Where Q is water inflow(m^3/h); P is precipitation(mm); M is aquifer thickness;(m) A is mining area(10^3m^2);D is mining depth(m); T is mining thickness(m); L is driving footage(m); a, b, c, d, e, f and λ are parameters to be solved.

Using statistical analysis, a weighted multiple nonlinear regression prediction model is established when the residual sum of squares of the fitted value and the measured value is minimum. The parameter estimations for the influencing factors are shown in Table 5.

The fitting parameter estimations of influencing factors are substituted into Eq. (13) to get the weighted multiple nonlinear regression prediction model equation, as shown in Eq. (14).

$$Q = 0.16P + 0.99sin(0.73M + 3.6) + \frac{1.99A - 0.04}{(A - 0.68)^2 - 0.29}$$
$$+ 0.08cos(0.95D) + 2.69sin(0.95D) \tag{14}$$
$$+ 1.29T((^T + 0.49) + 0.07L - 12.92$$

Where Q is water inflow(m^3/h); P is precipitation(mm); M is aquifer thickness (m); A is mining area(10^3m^2); D is mining depth(m); T is mining thickness(m); L is driving footage (m).

Result

In order to evaluate the accuracy of the prediction results of the prediction model proposed herein, Mean Absolute Percentage Error (MAPE) and Root Mean Square Error

Table 5 Parameter estimations of the multiple nonlinear fitting function

Factors	Parameters	Estimate	Std.Error	95% Confidence Interval	
				Lower	Upper
P	a_1	0.00	0.00	0.00	0.00
	a_2	0.00	0.01	-0.07	0.05
	a_3	0.43	1.60	-4.59	6.07
	a_4	7.11	59.84	-237.28	250.61
M	b_1	6.18	20.93	-94.06	123.13
	b_2	0.73	0.03	0.69	0.82
	b_3	3.60	10.26	-2.86	7.44
A	c_1	8.03	7.80	-10.20	27.10
	c_2	-5.16	2.46	-10.96	8.47
	c_3	-1.35	0.32	-1.83	-0.54
	c_4	0.15	0.18	-0.03	0.73
D	d_1	-64.04	51.52	-238.25	816.28
	d_2	0.35	7.48	-418.71	432.73
	d_3	50.61	7.89	-70.55	718.09
	λ	0.95	0.10	0.60	1.02
T	e_1	9.56	240.27	-484.62	501.37
	e_2	1.85	10.79	-85.39	93.65
	e_3	3.19	45.04	-602.56	610.35
	e_4	-41.91	202.68	-6937.99	6944.26
L	f_1	0.00	0.00	0.00	0.00
	f_2	0.00	0.00	-0.01	0.00
	f_3	1.70	0.78	-1.30	1.91
	f_4	-785.50	17.02	-1223.86	34.83

(RMSE) are selected to compare the prediction results of the multiple linear regression prediction model, and the weighted multiple nonlinear regression prediction model with the measured values of mine water inflow. The error calculation formulas are Eqs. (15) and (16).

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} |\frac{\widehat{y}_i - y_i}{y_i}| \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\widehat{y}_i - y_i)^2} \tag{16}$$

Where $\widehat{y}_i$ is the predicted value; y_i is the measured value; n is the number of samples.

The test sample data in Table S1 (supplementary material) are respectively substituted into the multiple linear regression prediction model Eq. (6), and the weighted multiple nonlinear regression prediction model Eq. (14) to obtain the water inflow prediction values of the two prediction models, as shown in Table 6.

Table 6 The water inflow prediction values of the two prediction models

Test sample number	Measured value of water inflow	Multiple linear regression prediction model	Weighted multiple nonlinear regression prediction model
1	9.59	21.55	3.81
2	13.40	24.05	13.68
3	14.08	20.13	17.69
4	14.83	19.31	15.05
5	15.67	23.29	16.84
6	16.57	26.24	15.56
7	17.35	25.33	16.02
8	18.59	27.85	14.62
9	18.80	16.14	21.37
10	19.80	29.86	24.96
11	20.70	25.29	22.76
12	23.30	26.90	24.44
13	23.97	29.56	31.98
14	44.20	31.86	41.45
15	60.21	27.61	47.93

According to the measured values and predicted values in Table 6, comparative curves for the two prediction models are made, as shown in Fig. 11.

The measured and predicted values in Table 6 are substituted into Eqs. (15) and (16) to produce the error analysis tables for the two prediction models, as shown in Table 7.

According to Fig. 11, the predicted values of the multiple linear regression prediction model are basically larger than the measured values, and the overall trend of the prediction results is not consistent with the measured trend. However, the predicted values of the weighted multiple nonlinear regression prediction model agree well with the measured values, and the overall trend of the predicted results is almost in line with the measured trend. On the other hand, we found that the model fit accuracy is poor when predicting high water inflow.
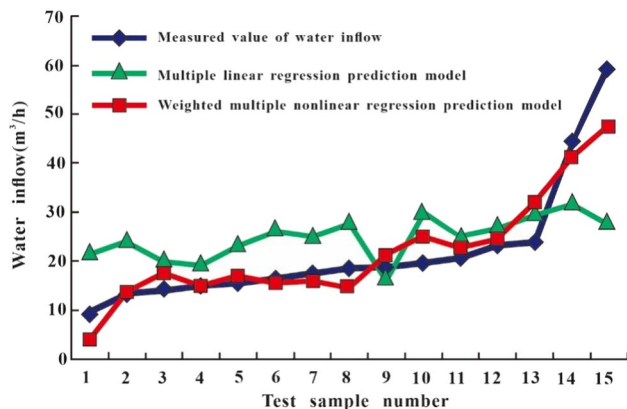
This may be due to the lack of extreme water inflow data under extreme conditions in our training samples, leading to the neglect of some extreme situations during training. The error analysis results in Table 7 show that the MAPE of the weighted multiple nonlinear prediction model is 16.44%, a great improvement compared to 45.88% of the multiple linear regression prediction model. The RMSE of the weighted multiple nonlinear prediction model is only 4.67, significantly lower than that of the multiple linear regression model.

Conclusion

This article analyzes the typical hydrogeological conditions in coal mines and, based on previous research, determines the main factors affecting the mine water flow to be rainfall, thickness of the aquifer, mining area, mining depth, mining thickness, and excavation length. Through correlation analysis, it is further concluded that the excavation length has a negative correlation with mine water flow, while the other factors have a positive correlation. Based on entropy value calculation, the weight of the impact factors on mine water flow is sorted as follows: rainfall > mining area > aquifer

Fig. 11 Comparative curves for the two prediction models

Table 7 Error analysis of predicted values

Evaluation indicators	Multiple linear regression prediction model	Weighted multiple nonlinear regression prediction model
MAPE	45.88%	16.44%
RMSE	11.56	4.67

thickness > mining thickness > mining depth > excavation length. Using multivariate regression theory, scatter analysis and MATLAB function programming, the article builds a weighted multivariate non-linear regression model for predicting mine water flow based on the calculated factor weights. This model considers both the impact of multiple factors on mine water flow and the differences in factor importance. The comparison analysis of the weighted multivariate non-linear regression model, the multivariate linear regression model, and the measured values of water flow shows that the newly established water flow prediction model can overcome the shortcomings of existing methods, minimize the prediction error caused by low hydrogeological survey degree, and improve the prediction accuracy.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12145-023-00985-x.

Data Availability The datasets used or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors declare no competing interests.

Ethics Approval Not applicable.

Conflict of Interest The authors declare that they have no competing fnancial interests or personal relationships that could have appeared to infuence the work reported in this manuscript.

References

Bai JW, Duan SL, Liu RT, Xin L, Tian JW, Zhang QS, Ma H (2021) Evolution of delayed water inrush in fault fracture zone considering time effect. Arab J Geosci 14(11):1001. https://doi.org/10.1007/s12517-021-06469-w

Bouw PC, Morton KL (1987) Calculation of mine water inflow using interactively a groundwater model and an inflow model. Int J Mine Water 6(4):31–50. https://doi.org/10.1007/BF02498188

Chen SL, Xu K, Yu SB, Li B, Cao B, Shen JJ (2015) Comparsion Study on Two Methods of Water Inflow Prediction of Coal Mining. Adv Mater Res 1092–1093:1379–1382. https://doi.org/10.1007/s12517-021-06469-w

Cohen J (1968) Multiple regression as a general data-analytic system. Psychol Bull 70(6):426–443. https://doi.org/10.1037/h0026714

Dong SN, Zhang WZ, Zhou WF, Chai R, Wang H, Zhao CH, Dong XL, Wang QM (2021) Discussion on Some Topical Issues of Water Prevention and Control in Coal Mines. Mine Water Environ 40(2):547–552. https://doi.org/10.1007/s10230-021-00773-3

Fiorillo F, Doglioni A (2010) The relation between karst spring discharge and rainfall by cross-correlation analysis (Campania, southern Italy). Hydrogeol J 18(8):1881–1895. https://doi.org/10.1007/s10040-010-0666-1

Gross AL (1975) Analyzing Multivariate Data. Contemp Psychol J Rev 20(9):743–744. https://doi.org/10.1037/014555

Guo H, Adhikary DP, Craig MS (2009) Simulation of mine water inflow and gas emission during longwall mining. Rock Mech Rock Eng 42(1):25–51. https://doi.org/10.1007/s00603-008-0168-9

Hou YJ (2012) Calculation of Mine Water Inflow and Cite Influencing Radius of Hetaoyu Coal with "Virtual Large Diameter Well" Method. Advanced Materials Research 610–613:2709–2712. https://doi.org/10.4028/www.scientific.net/AMR.610-613.2709

Hu WY, Zhao CH (2021) Evolution of Water Hazard Control Technology in China's Coal Mines. Mine Water Environ 40(2):334–344. https://doi.org/10.1007/s10230-020-00744-0

Katsaounis TI (2004) Analyzing Multivariate Data. Technometrics 46(2):254–255. https://doi.org/10.1198/tech.2004.s798

Krukovska V, Vynohradov Y (2019) Water stability influence of host rocks on the process of water filtration into mine working with frame and roof-bolting support, E3S Web of Conferences. 109(41). https://doi.org/10.1051/e3sconf/201910900041

Li GY, Zhou WF (2006) Impact of karst water on coal mining in North China. Environ Geol 49(3):449–457. https://doi.org/10.1007/s00254-005-0102-3

Li CX, Shi YJ, Zhou CC, Li X, Liu BY, Tang LM, Li SJ (2010) Evaluation of low amplitude and low resistivity pay zones under the fresh drilling mud invasion condition. Petrol Explor Dev+ 37(6):696–702. https://doi.org/10.1016/S1876-3804(11)60004-9

Li T, Song H, Huang G, Bi Y, Li X (2014) Assessment of groundwater changing trends through the generalized large well method with confined–unconfined flow model in open-pit mine area. Environ Earth Sci 72(11):4599–4606. https://doi.org/10.1007/s12665-014-3357-8

Li CC, Xu SG, Ba JJ (2015) The Prediction of the Water Inflow by Numerical Simulation Methods of Huijiabao Mine, Guizhou Province. Appl Mech Mater 744–746:1161–1164. https://doi.org/10.4028/www.scientific.net/AMM.744-746.1161

Li JB, Wang YH, Tan ZS, Du W, Liu ZY (2021) Study on Water Inflow Variation Law of No.1 Shaft Auxiliary Shaft in HighLiGongshan Based on Dual Medium Model. Symmetry 13(6):930. https://doi.org/10.3390/sym13060930

Li B, Zhang WP, Long J, Fan J, Chen MY, Li T, Liu P (2022a) Multi-source information fusion technology for risk assessment of water inrush from coal floor karst aquifer. Geomat Nat Haz Risk 13(1):2086–2106. https://doi.org/10.1080/19475705.2022.2108728

Li B, Zhang HL, Luo YL, Liu L, Li T (2022b) Mine inflow prediction model based on unbiased Grey-Markov theory and its application. Earth Sci Inf 15(2):855–862. https://doi.org/10.1007/s12145-022-00770-2

Liu WT, Li Q, Zhao JY, Fu B (2018) Assessment of water inrush risk using the principal component logistic regression model in the Pandao coal mine, China. Arab J Geosci 11(16):463. https://doi.org/10.1007/s12517-018-3815-9

Liu P, Yang M, Sun YJ (2019a) Hydro-geochemical processes of the deep Ordovician groundwater in a coal mining area, Xuzhou, China. Hydrogeol J 27(6):2231–2244. https://doi.org/10.1007/s10040-019-01991-4

Liu Q, Sun YJ, Xu ZM, Jiang S, Zhang P, Yang BB (2019b) Assessment of Abandoned Coal Mines as Urban Reservoirs. Mine Water Environ 38(2):215–225. https://doi.org/10.1007/s10230-019-00588-3

Ma D, Bai HB (2015) Groundwater inflow prediction model of karst collapse pillar: a case study for mining-induced groundwater inrush risk. Nat Hazards 76(2):1319–1334. https://doi.org/10.1007/s11069-014-1551-3

Ma D, Duan HY, Li WX, Zhang JX, Liu WT, Zhou ZL (2020) Prediction of water inflow from fault by particle swarm optimization-based modified grey models. Environ Sci Pollut R 27(33):42051–42063. https://doi.org/10.1007/s11356-020-10172-w

Miladinović B, Vakanjac VR, Bukumirović D, Dragišić V, Vakanjac B (2015) Simulation Of Mine Water Inflow: Case Study Of The Štavalj Coal Mine (Southwestern Serbia). Arch Min Sci 60(4):955–969. https://doi.org/10.1515/amsc-2015-0063

Ouedraogo I, Defourny P, Vanclooster M (2019) Application of random forest regression and comparison of its performance to multiple linear regression in modeling groundwater nitrate concentration at the African continent scale. Hydrogeol J 27(3):1081–1098. https://doi.org/10.1007/s10040-018-1900-5

Polak K, Różkowski K, Czaja P (2016) Causes and Effects of Uncontrolled Water Inrush into a Decommissioned Mine Shaft. Mine Water Environ 35(2):128–135. https://doi.org/10.1007/s10230-015-0360-6

Qiu M, Shi LQ, Teng C, Zhou Y (2017) Assessment of Water Inrush Risk Using the Fuzzy Delphi Analytic Hierarchy Process and Grey Relational Analysis in the Liangzhuang Coal Mine, China. Mine Water Environ 36(1):39–50. https://doi.org/10.1007/s10230-016-0391-7

Qiu M, Huang FJ, Wang Y, Guan T, Shi LQ, Han J (2020) Prediction model of water yield property based on GRA, FAHP and TOPSIS Methods for Ordovician Top Aquifer in the Xinwen Coalfield of China. Environ Earth Sci 79(10):214. https://doi.org/10.1007/s12665-020-08954-z

Shao AJ, Meng QX, Wang SW, Liu Y (2014) Prediction of Mine Inrush Water Based on BP Neural Network Method. Adv Mater Res 989–994:1814–1820. https://doi.org/10.4028/www.scientific.net/AMR.989-994.1814

Shi LQ, Gao WF, Han J, Tan XP (2017) A Nonlinear Risk Evaluation Method for Water Inrush Through the Seam Floor. Mine Water Environ 36(4):597–605. https://doi.org/10.1007/s10230-017-0449-1

Singh RN, Atkins AS (1985) Application of idealised analytical techniques for prediction of mine water inflow. Min Sci Technol 2(2):131–138. https://doi.org/10.1016/S0167-9031(85)90346-9

Singh RN, Atkins AS, Doulati Ardejani F (2012) Hydrogeological Issues Concerning the Thar Lignite Prospect. Int J Min Geo-Eng. 46(2):141–156. https://doi.org/10.22059/ijmge.2012.51324

Wang WD, Li SJ, Han JX (2015b) Analysis of the main global coal resource countries' supply-demand structural trend and coal industry outlook. China Min Mag 24(2):5–9. https://doi.org/10.3969/j.issn.1004-4051.2015.02.004

Wang H, Luo AK, Chai R, Liu QS (2015a) Application of GM Model in Coal Mine Water Inflow Prediction. 2015a Seventh International Conference on Measuring Technology and Mechatronics Automation. p192–195. https://doi.org/10.1109/ICMTMA.2015.54

Wei WX, Shi LQ, Lu XM, Zhang F (2011) Prediction of Mine Water Inflow Based on Support Vector Machine. 2011 Workshop on Digital Media and Digital Content Management. p326–329. https://doi.org/10.1109/DMDCM.2011.17

Wu Q, Fan SK, Zhou WF, Liu SQ (2013) Application of the Analytic Hierarchy Process to Assessment of Water Inrush: A Case Study for the No.17 Coal Seam in the Sanhejian Coal Mine, China. Mine Water Environ 32(3):229–238. https://doi.org/10.1007/s10230-013-0228-6

Wu Q, Guo XM, Shen JJ, Xu S, Liu SQ, Zeng YF (2017) Risk Assessment of Water Inrush from Aquifers Underlying the Gushuyuan Coal Mine, China. Mine Water Environ 36(1):96–103. https://doi.org/10.1007/s10230-016-0410-8

Wu C, Wu X, Zhu G, Qian C (2019) Predicting mine water inflow and groundwater levels for coal mining operations in the Pangpangta coalfield, China. Environ Earth Sci 78(5):130. https://doi.org/10.1007/s12665-019-8098-2

Xu C, Gong PP (2011) Water Disaster Types and Water Control Measures of Hanxing Coal Mine Area. Procedia Earth Planet Sci 3:343–348. https://doi.org/10.1016/j.proeps.2011.09.104

Xu J, Jing GX, Xu YY (2012) Prediction of the maximum water inflow in Pingdingshan No.8 mine based on grey system theory. J Coal Sci Eng (China) 18(1):55–59. https://doi.org/10.1007/s12404-012-0110-3

Xu Q, Wang XW, Chen JJ (2020) Quantitative evaluation of pore-scale heterogeneity based on statistical analysis of a pore network model of unconsolidated porous media. Hydrogeol J 28(5):1841–1852. https://doi.org/10.1007/s10040-020-02162-6

Xue JK, Shi L, Wang H, Ji ZK, Shang HB, Xu F, Zhao CH, Huang H, Luo AK (2021) Water abundance evaluation of a burnt rock aquifer using the AHP and entropy weight method: a case study in the Yongxin coal mine, China. Environ Earth Sci 80(11):417. https://doi.org/10.1007/s12665-021-09703-6

Zhang K, Cao B, Lin G, Zhao MD (2017) Using Multiple Methods to Predict Mine Water Inflow in the Pingdingshan No.10 Coal Mine, China. Mine Water Environ 36(1):154–160. https://doi.org/10.1007/s10230-015-0381-1

Zuo WZ, Wang GH, Li J, Gao LF (2011) Water Source Determination of Mine Inflow Based on Non-Linear Method and Hydrochemical Characteristics. Adv Mater Res 354–355:92–97. https://doi.org/10.4028/www.scientific.net/AMR.354-355.92