



# Principal component analysis and Fisher discriminant analysis of environmental and ecological quality, and the impacts of coal mining in an environmentally sensitive area

Hujun He  · Chong Tian · Gang Jin · Ke Han

Received: 27 August 2019 / Accepted: 18 February 2020 / Published online: 3 March 2020  
© Springer Nature Switzerland AG 2020

**Abstract** Most discriminant methods do not consider the problem of misjudgment related to the superposition of information from different discriminant indexes. Therefore, we used principal component and Fisher discriminant analysis to model, assess, and classify environmental and ecological quality, and the impacts of coal mining. The analysis uses the following input parameters as discrimination indexes: geomorphology, phreatic water depth, thickness of the phreatic water layer, bedrock thickness above the uppermost coal seam, and thickness of the uppermost coal seam. Twenty-three datasets from the Yushenfu coal mine

area, Shaanxi Province, China, were used to train the model. The validity of the model was tested by the backward substitution method, and the misjudgment rate was zero. Seven datasets were then used as test samples in a support vector machine model. Our results show that it is feasible to predict the environmental and ecological impacts of coal mining with principal component analysis and Fisher discriminant analysis, which can effectively eliminate the interaction between the sample variables. This results in a more accurate assessment of mine environmental quality and represents a new method for predicting the impacts of coal mining in environmentally sensitive areas.

H. He · C. Tian · G. Jin · K. Han  
School of Earth Science and Resources, Chang'an University,  
Xi'an 710054, China

C. Tian  
e-mail: tc@chd.edu.cn

G. Jin  
e-mail: gangjin@chd.edu.cn

K. Han  
e-mail: 877415765@qq.com

H. He  
Key Laboratory of Western Mineral Resources and Geological  
Engineering, Ministry of Education, Xi'an 710054, China

H. He (✉)  
State Key Laboratory of Water Resource Protection and  
Utilization in Coal Mining, Shenhua Group Co., Ltd,  
Beijing 100011, China  
e-mail: hsj2010@chd.edu.cn

**Keywords** Environmentally sensitive · Geological environment · Principal component analysis · Fisher discriminant analysis · Prediction

## Introduction

Coal is an important primary energy source and underpins the development of China's economy and society. However, coal mining and associated activities can have serious ecological and environmental impacts that are particularly problematic in environmentally sensitive areas. The Yulin, Shenmu, and Fugu mine areas in northern Shaanxi Province (i.e., the Yushenfu mine area) (Song 2009; Dou 2011, 2012) in China are located on the border between the Loess Plateau and the Maowusu Desert. This is one of the most important large-scale

coal mining areas in China, which produces 100 Mt/a of coal. The Yushenfu mine area is an environmentally sensitive region characterized by arid and semi-arid conditions, desertification, soil erosion, and scarce water resources. An important research question concerns how large-scale coal mining impacts the local environment and ecology, and how these impacts might be better classified and quantified.

The prediction and evaluation of mine-associated environmental impacts are important areas of research worldwide, and a number of approaches have been developed to undertake such research, including fuzzy mathematics (Liao and Wu 2004), analytical hierarchy processes (Chen et al. 2006), matter-element analysis (Feng et al. 2009), the back production (BP) neural network method (Li et al. 2015), the comprehensive index method (Lei et al. 2005), support vector machine (SVM) modeling (Dou 2011, 2012), the geographic information system (GIS) method (Shan 2013), and Fisher discriminant analysis (FDA) (Zhang et al. 2014). These various approaches each have unique methodologies and limitations. For example, the fuzzy mathematics method does not meet “non-negative bounded, additive, and polarity” measurement standards; the max–min principle has deficiencies related to the subjectivity of the weightings used in the expert scoring method (Cao et al. 2007); neural networks are prone to local optimality and have limited ability to solve small sample problems; and kernel functions and parameter selection for SVM models are complex (Shao and Xu 2015). Notably, most of these methods do not consider the problem of misjudgment related to the superposition of information from different discriminant indexes.

Given these limitations, we used principal component analysis (PCA) to refine the discriminant index data that affect the environment quality in mines, and converted multiple interrelated index variables into new, independent sample indexes by linear combinations. The variations in the original discriminant indexes are effectively extracted and eliminate the effects caused by the superposition of information among the indexes. This results in a more robust characterization of different mine environments. The processed data are then subjected to Fisher discriminant analysis (FDA), which enables the environmental and ecological impacts of coal mining to be predicted. We applied this method to the Yushenfu mine area and show that it produces

reliable results and thus is applicable to other coal mining areas in northwest China.

## Methodology

### PCA method

The PCA method is a data compression and feature information extraction technique. This method can effectively eliminate the correlation between high-dimensional datasets, reduce the data dimensions, and simplify the data structure. In the simplification process, multiple related variables are combined, based on the principle of minimizing information loss.

*PCA mathematical model* The  $p$  variables  $X_1, X_2, \dots, X_p$  of the original data matrix  $X$  are used for a linear combination, which is expressed as

$$\begin{cases} Y_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ \vdots \\ Y_p = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{cases} \quad (1)$$

In Eq. (1),  $a_{i1} + a_{i2} + \dots + a_{ip} = 1$ ,  $Y_i$ , and  $Y_j (i \neq j; i, j = 1, 2, \dots, p)$  are not related;  $Y_1$  is the largest variance combination in all linear combinations of  $(X_1, X_2, \dots, X_p)$ ; and  $Y_2$  is the largest variance combination in all unrelated linear combinations of  $(X_1, X_2, \dots, X_p)$ . The variance sum of  $Y_1, Y_2, \dots, Y_p$  is equal to the variance sum of  $(X_1, X_2, \dots, X_p)$ .

*Solving the principal component* The general steps for solving the principal component are as follows. (a) The original variable data are standardized, and then the covariance matrix  $\sum(S_{ij})_{p \times p}$  between the variables is calculated. (b) The eigenvalue  $\lambda_i$  of the covariance matrix  $\sum(S_{ij})_{p \times p}$  and the corresponding orthogonal unit eigenvector are calculated. The former  $m$  larger eigenvalues of  $\sum(S_{ij})_{p \times p}$  are  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ , which are the variances corresponding to the  $m$  principal components, and the unit eigenvector  $a_i$  corresponding to  $\lambda_i$  is the original variable coefficient of the principal component  $F_i$ , which is expressed as follows:

$$a_i = \lambda_i / \sum_{i=1}^p \lambda_i \tag{2}$$

The determination of the number of principal components generally depends on the cumulative variance contribution rate. When this rate exceeds 80%, the former  $m$  principal components in the data matrix are taken as the discriminant indexes (Lu et al. 2012; Qian et al. 2016).

FDA theory

The basic principle of FDA is that high-dimensional data points are projected onto low-dimensional space, such that the data points become more densely clustered. The projection is optimized to best discriminate classes of data, and the discriminant analysis function classifies these classes according to the largest distance between them and the smallest distance within a class. The model can then classify and discriminate data for new samples.

*Solution of the FDA model*  $G_1, G_2, \dots, G_k$  are a set of  $k$  populations, and the number of samples taken from it is  $n_1, n_2, \dots, n_k$ , respectively. In this case,  $n = n_1 + n_2 + \dots + n_k$  and  $x_a^{(i)} = (x_{a1}^{(i)}, \dots, x_{ap}^{(i)})$  are the observation vector for sample  $a$  of population  $i$ . The discriminant function is set to  $y(x) = c_1x_1 + \dots + c_px_p \triangleq c^T x$ , where  $c = (c_1, \dots, c_p)^T, x = (x_1, \dots, x_p)^T$ .  $\bar{x}^{(i)}$  and  $\bar{s}^{(i)}$  are the sample mean and covariance of  $x$  in population  $G_i$ , respectively, and the sample mean and variance of  $R(x)$  in  $G_i$  are  $\bar{y}^{(i)} = c^T \bar{x}^{(i)}, \sigma_i^2 = c^T \bar{s}^{(i)} c$ .  $\bar{x}$  is the total mean vector, and it then follows that  $\bar{y} = c^T \bar{x}$ .

The Fisher judgment criterion uses a variance analysis to select the coefficient vector  $c$ , such that  $\lambda = \sum_{i=1}^k n_i$

$$(\bar{y}^{(i)} - \bar{y})^2 / \sum_{i=1}^k (n_i - 1) \sigma_i^2 = \frac{c^T B c}{c^T E c}$$

is maximized, where  $E$  is the sum of the squares of deviations within the group and  $B$  is the sum of the squares of deviations between groups, leading to  $E = \sum_{i=1}^k n_i s^{(i)}, B = \sum_{i=1}^k n_i (\bar{x}^{(i)} - \bar{x})(\bar{x}^{(i)} - \bar{x})^T$ .

For the maximum value  $\lambda$ , given the necessary conditions for the existence of extreme values,  $\partial \lambda / \partial c = 0$  and  $Bc = \lambda Ec$  can be defined. Algebraically, the number  $m$  of non-zero eigenvalues for this formula cannot exceed

$\min(k - 1, p)$  and, because  $B$  is non-negative and definite, the non-zero eigenvalue must be positive. As such,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m > 0$  and  $m$  discriminant functions can be constructed, which are of the form  $y_l(x) = (c^{(l)})^T x (l = 1, \dots, m)$ . The first discriminant function is generally solved to obtain most of the sample information, but if it is difficult to discriminate, this can be considered in conjunction with subsequent discriminant functions.

*Discriminating criterion* For each discriminant function, there is an index contribution rate  $p_l = \lambda_l / \sum_{i=1}^k \lambda_i$

that measures its discriminating ability. The cumulative discriminating ability of the previous  $r (r \leq k)$  discriminant function is then defined as  $p_r = \sum_{i=1}^r \lambda_i / \sum_{i=1}^k \lambda_i$  and, if  $p > 0.85$ , then the  $r$  discriminant function can be used to discriminate. Two weighting methods can be used in the discrimination analysis, which are described as follows.

*Unweighted method:* If  $\bar{y}_l^{(i)} = (c^{(l)})^T \bar{x}^{(i)}$  ( $l = 1, \dots, r; i = 1, \dots, k$ ) and the unknown sample is  $x = (x_1, x_2, \dots, x_p)^T$ , then  $y_l(x) = (c^{(l)})^T x, D_i^2 = \sum_{l=1}^r [y_l(x) - \bar{y}_l^{(i)}]^2 (l = 1, \dots, r; i = 1, \dots, k)$ . If  $D_r^2 = \min_{1 \leq i \leq k} D_i^2$ , then  $x \in G_r$ .

*Weighting method:* Considering that each discriminant function has a different discriminability, let  $D_i^2 = \sum_{l=1}^r [y_l(x) - \bar{y}_l^{(i)}]^2 \lambda_l (l = 1, \dots, r; i = 1, \dots, k)$ , where  $\lambda_l$  is the eigenvalue obtained by solving  $Bc = \lambda Ec$ . If  $D_r^2 = \min_{1 \leq i \leq k} D_i^2$ , then  $x \in G_r$  (Zhong and Guo 2008; Hu et al. 2010; Wang 2010; Zhou and Shi 2010; Fu and Wang 2012; Lu et al. 2012; Hou et al. 2016).

Previous studies have shown that FDA does not limit the populations and distribution (He 2004; Yuan and Song 2009; Hu et al. 2010; Wang 2010), provided the population mean and covariance matrix have been calculated and the population covariance matrix is reversible. In such cases, the statistical test can be omitted. In general, if the backward substitution discriminant success is  $> 80\%$ , the FDA model is reliable. In this test, all training samples are taken as  $n$  new samples, which are then subjected to application of the discriminant function and criterion. The ratio of the number  $N$  of misjudgments to all samples is the misjudgment rate  $\gamma$ ,

$$\text{which is calculated as } \gamma = N / \sum_{i=1}^k n_i.$$

### Discriminant model and its application to the prediction of environmental and ecological impacts of coal mining

#### Geological background

The Yushenfu mine area is located in the central Ordos Basin, between the Loess Plateau and the Maowusu Desert. It covers the three administrative districts of Yuyang district, Shenmu city, and Fugu county. It consists of two mine areas: Yushen and Shenfu (Fig. 1). The area is dominated by hilly sandy and loess landforms. The main strata exposed in the study area are the Upper Triassic Yongping Formation ( $T_{3y}$ ), Lower Jurassic Fuxian Formation

( $J_{1f}$ ), Middle Jurassic Yan'an ( $J_{2y}$ ), Zhiluo ( $J_{2z}$ ), and Anding formations ( $J_{2a}$ ), Lower Cretaceous Tuoluo Formation ( $K_{1l}$ ), Pliocene Sanzhima Formation ( $N_{2s}$ ), lower Pleistocene Sanmen Formation ( $Q_{1s}$ ), middle Pleistocene Lishi Formation ( $Q_{2l}$ ), upper Pleistocene Salawusu ( $Q_{3s}$ ) and Malan formations ( $Q_{3m}$ ), and Holocene aeolian sand ( $Q_4^{eol}$ ) and alluvium ( $Q_4^{al}$ ). The study area is located in the northern Shaanxi slope along the northeastern edge of the Ordos Basin. The structure is simple, and faults and folds are not developed at the surface in this region. The surface geology is dominated by Quaternary strata. The main coal-bearing strata in the mine area are in the Yan'an Formation ( $J_{2y}$ ), which is divided into five coal sections (from top

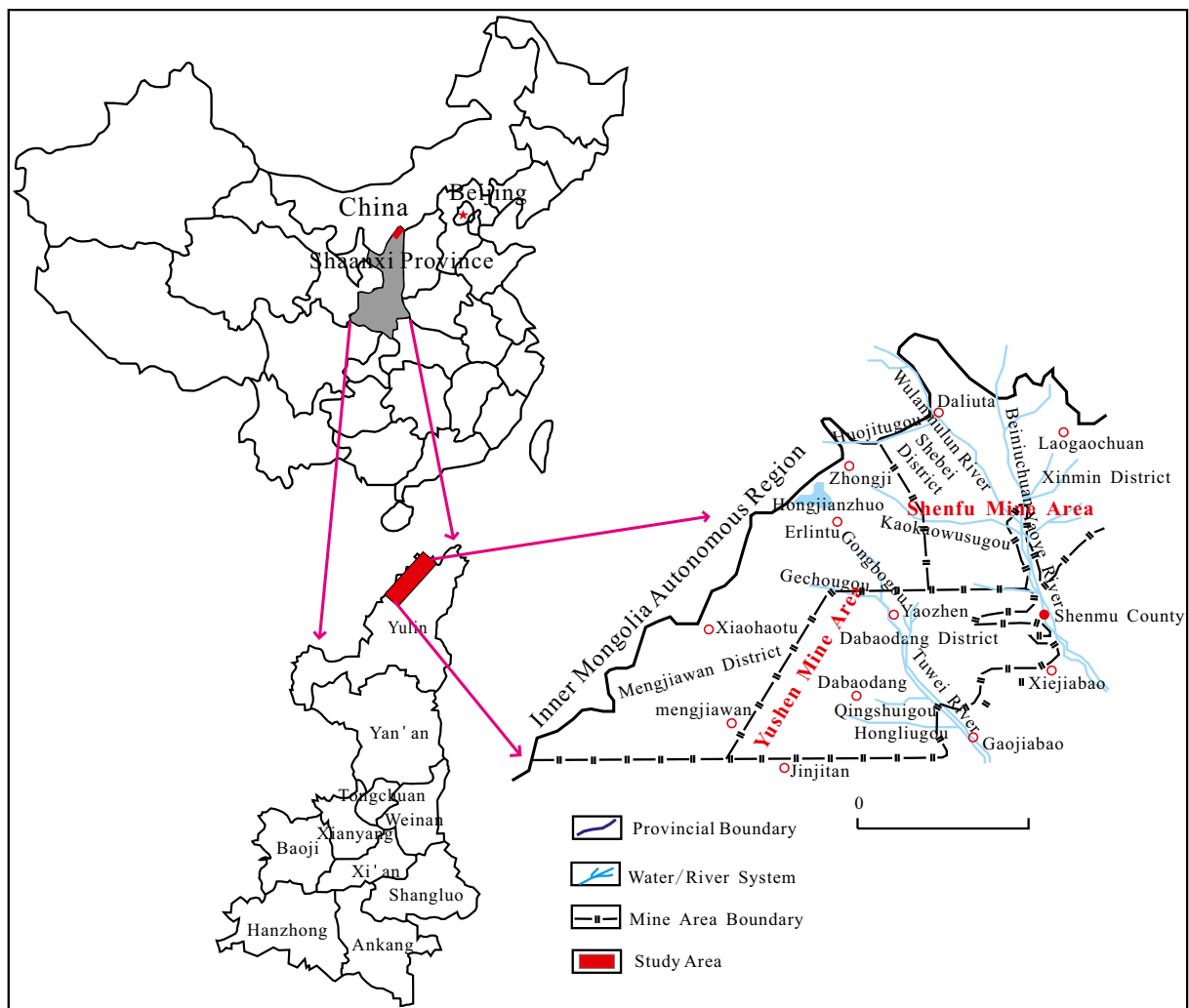


Fig. 1 Location of the Yushenfu mine area in Shaanxi Province, China

**Table 1** Classification and evaluation of the main factors affecting the environmental impacts of mining

Geomorphological type ( $X_1$ )	The depth of the previous water layer/m ( $X_2$ )	The sand layer thickness of the phreatic water/m ( $X_3$ )	Thickness of the bedrock/m ( $X_4$ )	The thickness of the upper coal seam/m ( $X_5$ )	Classification
River valley land (1)	0	> 100	> 400	< 2	I
Beach land (2)	0–3	20–100	200–400	2–4	II
Duce land (3)	3–10	20–5	120–200	4–6	III
Loess gully area (4)	10–20	0–5	30–120	6–8	IV
Bare bedrock area (5)	> 20	0	< 30	> 8	V

to bottom:  $J_2y^5$ ,  $J_2y^4$ ,  $J_2y^3$ ,  $J_2y^2$ , and  $J_2y^1$ ). There are five layers of coal seams in each of the Shenfu and Yushen mine areas ( $1^{-2}$ ,  $2^{-2}$ ,  $3^{-1}$ ,  $4^{-2}$ , and  $5^{-2}$ , and

$2^{-2}$ ,  $3^{-1}$ ,  $4^{-3}$ ,  $5^{-2}$ , and  $5^{-3}$ , respectively). The coal strata are better exposed and more shallowly buried in the eastern Shenfu mine area. The overburden

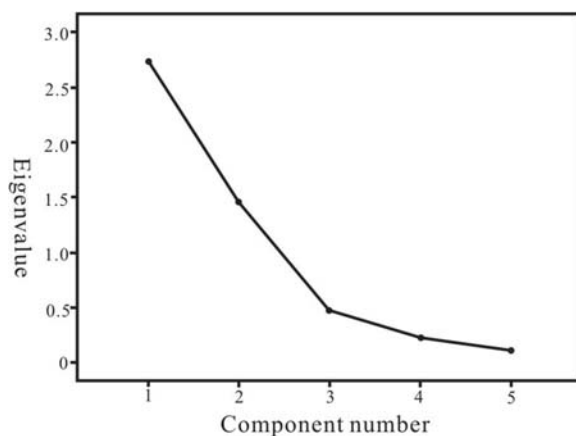
**Table 2** Sample learning data for the discriminant analysis model

Samples number	Discrimination factors (the main factors affecting the geological environment of mine area)					Discrimination result	
	Geomorphological type	The depth of the previous water layer/m	The sand layer thickness of the phreatic water/m	Thickness of the bedrock/m	The thickness of the upper coal seam/m	Actual result	PCA and FDA
1	1	0	30	500	5	I	I
2	2	0.8	80	220	6.2	I	I
3	2	2	80	240	6	I	I
4	2	1.8	70	130	6.2	II	II
5	2	2	80	150	6.5	II	II
6	2	2.2	80	140	6	II	II
7	2	16	40	110	6.5	III	III
8	2	10	20	120	6	III	III
9	2	13	23	115	6.5	III	III
10	2	6	12.2	60	6	III	III
11	4	10	0	20	2	IV	IV
12	4	12	0.5	25	3	IV	IV
13	4	15	0.8	26	3.5	IV	IV
14	4	18	1	30	4.5	IV	IV
15	4	20	1.5	35	5.6	IV	IV
16	5	22	0	25	6.5	V	V
17	5	28	0	35	7	V	V
18	5	30.5	0	40	6.5	V	V
19	5	40	0.5	50	7	V	V
20	5	50	0.3	80	8.5	V	V
21	3	60	35	80	9.5	V	V
22	3	45	30	45	8	V	V
23	3	50	20	50	8	V	V

**Table 3** Pearson correlation coefficient matrix for each index

Index	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
$X_1$	1.000				
$X_2$	0.546	1.000			
$X_3$	-0.734	-0.450	1.000		
$X_4$	-0.685	-0.457	0.534	1.000	
$X_5$	-0.015	0.642	0.194	0.003	1.000

thickness of the first coal seam  $2^{-2}$  is  $< 100$  m. Given the general westward dip of the strata, the burial depth in the Yushen mine area is greater, and the overburden thickness of the main coal seam  $2^{-2}$  is 100–400 m. In most of the Shenfu mine area and east of the Yushen mine area, the mining conditions of the  $2^{-2}$  coal seam are unusual, as the mined seam is shallowly buried beneath a thick and loose aquifer. The overburden lithology is generally a hard rock type. Especially when the structure type of overburden is a burnt rock type and a sand base type, the water diversion fissure zone is very developed and severely damages a highly water-rich aquifer. Therefore, destruction of the aquifer due to coal mining will have a serious impact on the water quality of the entire mine area, which will lead to serious ecological and environmental impacts (Song 2009; Dou 2011; Hou et al. 2016). As such, the Yushen and Shenfu mine areas are located in an environmentally sensitive region. Recent intensification of mining activity in this region is likely to lead to environmental impacts, such as destruction of surface vegetation, ground subsidence, degradation

**Fig. 2** PCA scree plot

of water resources, atmospheric pollution, desertification, and intensification of soil erosion.

### Model input parameters

Numerous natural and engineering factors govern the environmental impacts of coal mining (Dou 2011, 2012). These can be divided into natural factors (phreatic water depth, water quality, rock and soil type, topography, water resources, thickness and depth ratio of coal seams, geomorphology, and climatic conditions) and social factors (transport, mining, mineral processing, and ore washing). Based on the regional characteristics of the Yushenfu mine area, consulting experts' opinions and combining with the local actual situation, five factors affecting the environmental impacts in this area were selected for use in this study: (1) geomorphology ( $X_1$ ); (2) depth to the phreatic water layer ( $X_2$ ); (3) sand layer thickness of the phreatic water region ( $X_3$ ); (4) thickness of the overlying bedrock on the uppermost coal seam (referred to as the thickness of bedrock) ( $X_4$ ); and (5) thickness of the uppermost coal seam ( $X_5$ ). Among these inputs, geomorphology is a state parameter that must be described and quantified. The depth of the phreatic water layer, sand layer thickness of the phreatic water region, thickness of the bedrock, and thickness of the uppermost coal seam are measured values. The specific classification criteria and evaluation are listed in Table 1. The classification of the environmental impacts of coal mining is divided into groups I–V, which represent good, better, medium, poor, and extremely poor quality (Dou 2011, 2012).

### Learning samples and establishment of the discriminant analysis model

Thirty sets of representative environmental impact/quality data were selected from the Yushenfu mine area, and 23 datasets were used as training samples for the FDA model (Table 2); 7 other datasets were used for discrimination testing (Dou 2011).

### PCA data processing

The data in Table 2 were first standardized, and then the standardized data were processed by PCA. The correlation coefficient matrix for the indexes is provided in Table 3. There is a clear correlation between the five discriminant indexes that affect the mine environment,

**Table 4** Eigenvalues and group center values of the Fisher discriminant functions

Function	Eigenvalue	Variance contribution rate/%	Cumulative variance contribution rate/%	Canonical correlation	Group center values				
					G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	G <sub>5</sub>
1	27.081	3.4	93.4	0.982	-7.833	-5.432	-1.899	1.801	4.799
2	1.918	6.6	100.0	0.811	0.441	0.531	0.014	-2.226	1.020

such as between the geomorphological type and sand layer thickness of the phreatic water ( $r = 0.734$ ), and the information from the sample indexes overlaps. If the quality of the mine environment is discriminated using these five indexes, the prediction model will be affected and could lead to misjudgment. Therefore, it is necessary to perform PCA before input of the sample data. After PCA processing of the data, two principal components are extracted according to the distribution rules of the information content of each principal component in the PCA scree plot (Fig. 2). The selected principal components contain 83.93% of the information in the original data.

According to the PCA matrix, the relationship between the new principal components  $Y_1$  and  $Y_2$  and the original variables is as follows:

$$Y_1 = 0.54X_1 + 0.47X_2 - 0.49X_3 - 0.49X_4 + 0.11X_5 \quad (3)$$

$$Y_2 = -0.17X_1 + 0.47X_2 + 0.31X_3 + 0.14X_4 + 0.8X_5 \quad (4)$$

Fisher discriminant analysis

The environmental quality standards classified into five categories (I–V) were used as five different populations, and it was assumed that the covariance matrix of the environmental quality population of the five categories is equal. The two principal components  $Y_1$  and  $Y_2$  obtained by PCA were taken as the two discriminant indexes for the FDA model, and the Fisher discriminant function was obtained as follows:

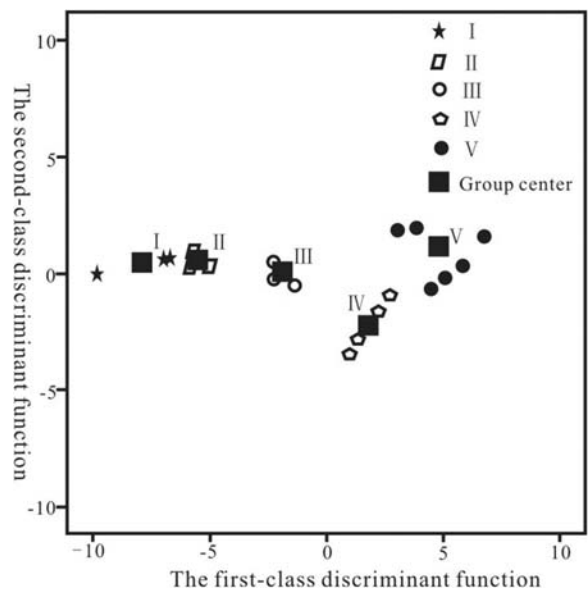
$$Z_1 = 1.023Y_1 - 0.025Y_2 \quad (5)$$

$$Z_2 = 0.244Y_1 + 0.994Y_2 \quad (6)$$

The discriminability of the two discriminant functions is significant (Table 4). The first discriminant function corresponds to an eigenvalue of 27.081, and the variance

contribution rate (discrimination efficiency) is 93.4%, which thus explains 93.4% of the sample information. The positive correlation coefficient is high (0.982), and this function can be used to discriminate most samples. When two discriminant functions are used, 100% of the sample information can be explained. Therefore, when Eq. (5) does not yield a clear judgment on sample classification, Eqs. (3) or (6) can be combined to obtain a clear judgment.

To initially test the validity and accuracy of the PCA and FDA model, 23 sets of training sample data were subjected to the backward substitution method using the established model, and compared with the actual results (Table 2). Figure 3 is a schematic diagram of the groupings divided by the first- and second-class discriminant functions, which illustrates the powerful predictive function of the model. The overall performance of the two discriminant functions is clearly evident in the figure. The misjudgment rate in this test was zero (Table 2). As such, the model appears stable and reliable, and it accurately predicted and classified the 23



**Fig. 3** Groupings of sample data using the first- and second-class discriminant functions

**Table 5** Test results of the PCA and FDA predictive model

Samples number	Discrimination factors (the main factors affecting the geological environment of mine area)					Discrimination results		
	Geomorphological type	The depth of the previous water layer/m	The sand layer thickness of the phreatic water/m	Thickness of the bedrock/m	The thickness of the upper coal seam/m	Actual result	SVM	PCA and FDA
1	2	1	120	300	6	I	I	I
2	2	2.6	100	180	6.5	II	II	II
3	3	12	30	280	6	III	III	III
4	2	15	50	110	6.5	III	III	III
5	5	30	0	30	7.5	V	V	V
6	5	19	0	20	8	V	V	V
7	3	25	20	30	8	V	V	V

datasets into the quality groupings (good, better, medium, poor, and extremely poor). The classification resulted in small distances within groups and large distances between groups. Therefore, PCA and FDA are able to robustly model and efficiently predict and classify the environmental impacts of mining in this area.

#### Model test and application

Another seven sets of sample data that were not used in the PCA and FDA model training were selected to further test the model. The calculated results are listed in Table 5. By comparison with a nonlinear model based on SVM (Dou 2011), it is evident that the discriminant results are exactly the same as the actual results (i.e., the classification accuracy rate is 100% and the misjudgment rate is zero). Therefore, this model is valid and reliable for making comprehensive predictions and classifications of environmental impact/quality in the mine areas. In addition, the method is relatively simple, stable, does not change with the number of sets of sample data, and has a high learning efficiency. Moreover, the method can not only effectively characterize the impacts of coal mining in an environmentally sensitive area with a small number of index variables, but also better eliminates the influence of superimposed information between the variables.

#### Conclusions

1. The impacts on environmental quality of coal mining in an environmentally sensitive area were evaluated by PCA and FDA. PCA was mainly used to

reduce the dimensionality of the variables and extract the most useful information to facilitate the subsequent processing of the algorithm. FDA enables different samples to be distinguished. The combination of the two methods makes it possible to discriminate and analyze the samples.

2. This model was applied to the Yushenfu mine area and yielded excellent discrimination results. This indicates that the model is robust, simple, and feasible, given that it has a high prediction accuracy and misjudgment rate of zero. As such, this new method can robustly assess and classify environmental and ecological quality and the impacts of coal mining.
3. The PCA and FDA model is based on measured engineering and natural data, and the representativeness and accuracy of these data can affect the model outcome. The environment in the mine areas is complex and influenced by many factors. Therefore, for practical application of our model, extensive and appropriate sample datasets need to be collected and the model needs to be trained. The accuracy of the model will improve as more sample datasets are processed.

**Funding information** This research was financially supported by the Open Fund of the State Key Laboratory of Water Resources Protection and Utilization in Coal Mining (SHJT-16-30.19) and the Special Fund for Basic Scientific Research of Central Colleges (310827173702).

#### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.



## References

- Cao, Q. K., Yang, Y. L., & Yu, R. L. (2007). The coal mines safety appraisal based on unascertained set. *Journal of China Coal Society*, 32(2), 181–185.
- Chen, Q., Hu, K., Luo, K. L., Li, F. L., & Zhao, W. (2006). Study on the synthetical assessment model of mine eco-environments based on AHP. *Journal of China University of Mining and Technology*, 35(3), 377–383.
- Dou, P. P. (2011). *Research on geologic environment effects induced by coal mining in ecological weakness area and assessment techniques—a case study taking Yushenfu mining district in northern Shanxi province*. China University of Mining and Technology, Xuzhou, China.
- Dou, P. P. (2012). The research of integrated geological environment evaluation based on support vector machine. *Journal of Mining & Safety Engineering*, 29(4), 555–558.
- Feng, J. C., Wu, Y., Zhao, S. P., Zhao, W. J., & Xing, Y. Q. (2009). Geological environment evaluation in Xiaolinling gold area. *Soil and Water Conversation in China*, 3, 29–31.
- Fu, Y. H., & Wang, X. M. (2012). Study and application of fisher discriminant analysis model to classification of rock mass quality. *Metal Mine*, 41(12), 107–110.
- He, X. Q. (2004). *Multivariate statistical analysis*. Beijing, China: China Renmin University Press.
- Hu, H. H., Liu, Z., Li, Z. J., & Cui, T. T. (2010). Fisher discriminant analysis to the classification of spontaneous combustion tendency grade of sulphide ores. *Journal of China Coal Society*, 35(10), 1674–1679.
- Hou, E. K., Tong, R. J., Wang, S. J., Feng, J., & Chen, T. (2016). Prediction method for the water enrichment of weathered bedrock based on Fisher model in Northern Shaaxi Jurassic coalfield. *Journal of China Coal Society*, 41(9), 2312–2318.
- Lei, W. R., Tang, C. M., Yu, G. W., & Huang, Y. Q. (2005). Study on the integrative assessment measures of mine geological environment in Jiangxi Province. *The Chinese Journal of Geological Hazard and Control*, 16(2), 63–66.
- Li, D., Zhou, K. F., Sun, W. D., Wang, J. L., Yu, H., & Liu, H. (2015). Application of BP neural network and SVM in mine environmental assessment. *Arid Land Geography*, 38(1), 128–134.
- Liao, G. L., & Wu, C. (2004). The application of fuzzy mathematical method on mine environmental comprehensive evaluation. *Environmental Science Trends*, 3, 15–17.
- Lu, J. T., Li, X. B., Gong, F. Q., Wang, X. R., & Liu, J. (2012). Recognizing of mine water inrush sources based on principal components analysis and fisher discrimination analysis method. *China Safety Science Journal*, 22(7), 109–115.
- Qian, Z. M., Ren, G. F., Chu, F. J., Qin, S. B. (2016). Rock mass quality classification based on PCA and Fisher discrimination analysis. *Rock and Soil Mechanics*, 37(S2), 427–432+441.
- Shan, J. H. (2013). *The evaluation of mining geo-environment in Northern Chaohu based on GIS and WebGIS design*. Hefei, China: Hefei University of Technology.
- Shao, L. S., & Xu, B. (2015). Prediction on classification of rock burst hazard based on Fisher discriminant analysis. *Metal Mine*, 44(1), 138–144.
- Song, S. J. (2009). The quantitative evaluation of eco-environment damage induced by coal exploitation in Yushefu mining area. Xi'an University of Science and Technology, Xi'an, China.
- Wang, D. H. (2010). *Multivariate statistical analysis and SPSS application*. East China University of Science and Technology Press, Shanghai, China.
- Yuan, Z. F., & Song, S. D. (2009). *Multivariate statistical analysis*. Beijing, China: Science Press.
- Zhang, Z. Z., Chu, C. M., Guan, W. M., Tian, Z. F., & Lin, W. C. (2014). Study on geological environment evaluation of Xinjiang rock gold mine and prevention measures. *Gold*, 35(6), 67–72.
- Zhong, C., & Guo, Q. (2008). Fisher discrimination method and its application. *Journal of Southwest Jiaotong University*, 43(1), 136–141.
- Zhou, J., & Shi, X. Z. (2010). Fisher discriminant analysis method for prediction of classification of rock burst risk. *Journal of China Coal Society*, 35(S0), 22–27.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.