



Spatial prediction of loose aquifer water abundance mapping based on a hybrid statistical learning approach

Qi Zhang¹ · Zaiyong Wang^{2,3}

Received: 2 March 2021 / Accepted: 25 May 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

In order to study and prevent water hazards in coal mines under thick loose strata, we need to make correct assessments of their water abundance levels based on the limited borehole data. According to the multi-source information composite principle, five main influencing factors of water abundance are chosen, and they are: the aquifer thickness, the thickness ratio between sandy and clayey layers, the consumption of the drilling fluid, the core recovery rate, and the hydraulic conductivity. Their spatial variations in the whole study area could be inferred from Kriging interpolation. Next, we have developed a novel off-the-shelf two-step assessment approach. In the first step, we apply a dimensionality reduction technique called Fisher discriminant analysis (FDA) to project the original normalized data into a low-dimensional space, which is convenient for data visualization. In this projection process, we want to keep the original information as much as possible. In the second step, we train three classification algorithms on the same transformed low-dimensional data to predict the water abundance level, and leave-one-out cross-validation is used to validate our proposed method due to data sparsity. Finally, the comprehensive zoning map of the study area's water abundance level is established, which can provide scientific guidance for the mining operations and prevention of mine water hazards in this region. The whole process is further elaborated through a case study of the Baodian coal mine, from which we are able to know the location with the highest water abundance level.

Keywords Mining · Aquifer water abundance · Machine learning · Fisher discriminant analysis · Multiclass classification

Introduction

Coal resources are an essential component of the energy structure in China (Liu et al. 2020). However, hundreds of water inrush accidents have occurred in China coal mines since the 1950s, deeply affecting the coal mine workers' safety and coal production (Meng et al. 2012; Wang et al. 2020c; Zhang et al. 2019, 2020c). Moreover, many coal mines in North China are covered by vast thick loose strata,

and a portion of the water inrush accidents actually happen in these strata due to disturbances of aquifer properties. These water inrush accidents belong to the category of mine roof water inrush hazards (Li et al. 2017). A comprehensive study of the associated water inrush mechanisms and water abundance levels would be instrumental in controlling water hazards in coal mines. Although numerous theories on water inrush mechanisms have been proposed in the past decade (Guo et al. 2017; Zhang 2005), accurate modeling of the water flow in the strata and its interaction with mining activities are considerably challenging owing to heterogeneity, anisotropy, multiphysics coupling (Wang et al. 2020b; Zhang 2020a, Zhang et al. 2021), and dynamic flow evolution (Shao et al. 2020a, b; 2021). In this study, we focus on mathematical methods based on actual measurements rather than simulation of the actual water inrush process, which can significantly decrease the computational complexity (Ke et al. 2020).

Research on the water abundance of loose aquifer has only been recently conducted. The simplest method is by calculating the water inrush coefficient (Meng et al. 2012). Some geophysical prospecting techniques have

Communicated by: H. Babaie

✉ Zaiyong Wang
zaiyongwang@163.com

¹ Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305, USA

² State Key Laboratory of Mining Disaster Prevention and Control, Shandong University of Science and Technology, Qingdao, 266590, China

³ College of Mining Engineering, Taiyuan University of Technology, Taiyuan, 030024, China

been carried out to directly evaluate the water abundance level. For example, Yang et al. (2016) obtained the relationship between the water abundance parameter and apparent resistivity through a layered geo-electrical model using the resistivity surveying method and the transient electromagnetic method. Another direct evaluation method is known as the unit water inflow method, which infers the water abundance level from the borehole flow rate, radius of influence, and borehole radius of a typical pumping experiment. However, this method fails to consider the impact of sedimentary condition. Besides, the data obtained from these methods may be considered to be sparse, thus the water abundance interpretations between two nearest field observations could be limited. Furthermore, the single-variable model might be biased. In addition to direct prospecting techniques, mathematical models could be great indirect alternatives, which have been successfully applied to many real problems such as mine floor water inrush risk prediction (*e.g.*, vulnerability index method) (Liu et al. 2018; Shi et al. 2017; Wang et al. 2012; Wu et al.

2014), groundwater spring potential modeling (Naghbi et al. 2020), urban flooding (Ke et al. 2020), forecasting time series data (Phan and Nguyen 2020), groundwater contaminant transport (Wang and Lu 2020a), and landslide susceptibility (Chen et al. 2017), among many others. However, a sophisticated mathematical model on loose aquifer water abundance (mine roof water inrush risk) evaluation still remains rare.

In this study, we propose a new off-the-shelf combined evaluation method for water abundance level prediction, and further apply it to the Baodian coal mine. In this method, low-dimensional embedding is applied to samples through Fisher discriminant analysis (FDA) (Hou et al. 2016; Mika et al. 1999; Qian et al. 2016; Sugiyama 2007; Zhang et al. 2013, 2020c), followed by various classification algorithms (Hastie et al. 2009; Ng 2020) to arrive at a convincing conclusion. It is the first time, to the best of our knowledge, that these statistical learning algorithms have been integrated within the context of mine water and the environment, particularly for the spatial characteristics

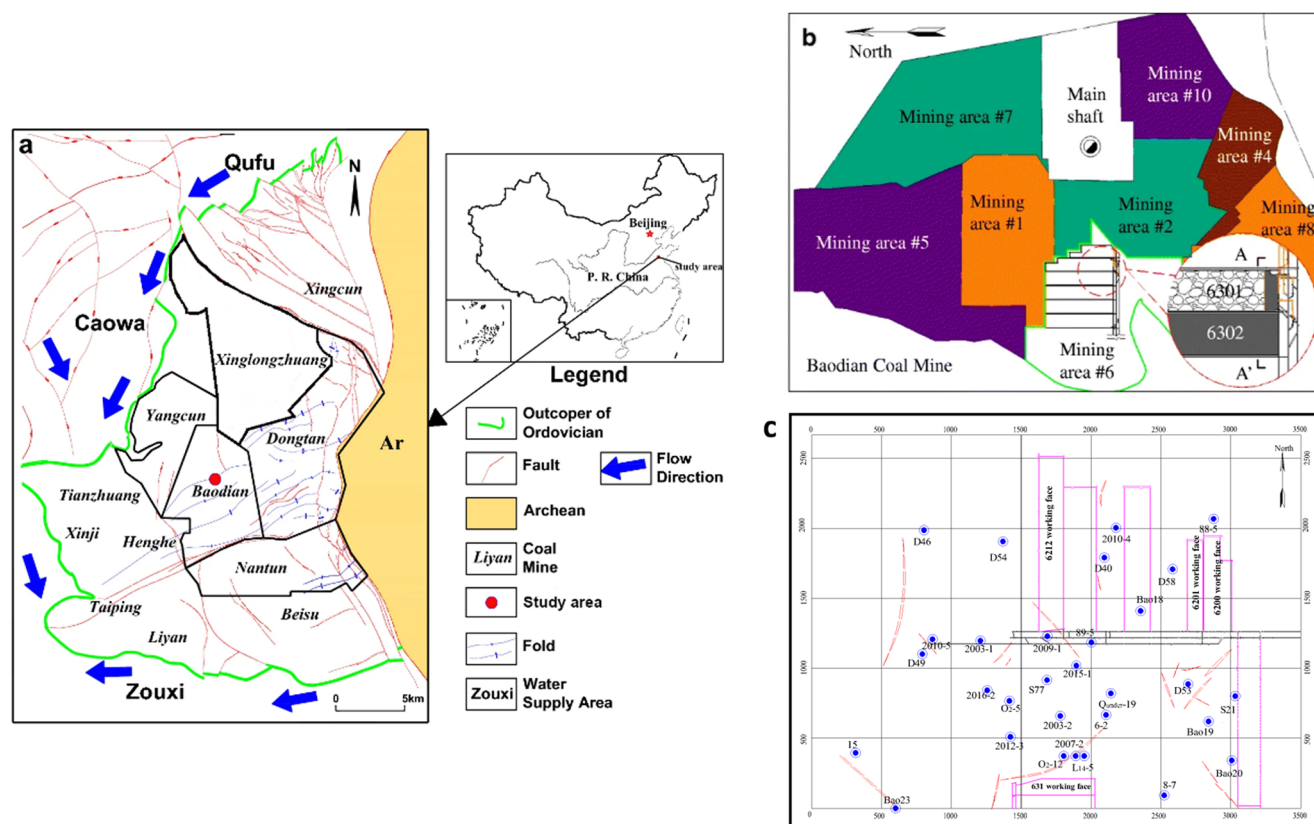


Fig. 1 a The geographical location of the Baodian coal mine (Qiao et al. 2020). b The mining areas in Baodian coal mine (Zhang et al. 2018a). c The sites of hydrogeological boreholes and locations of working faces in mining area #6

of loose aquifer water abundance. The novel contributions of this paper are five-fold: (1) The resulting zoning map is not a direct superposition of each thematic layer with some specified weights (Chen et al. 2018; Wu et al. 2014), but instead, the combination is much more complicated than simple analytic expressions; (2) Our results are more reliable than the Indicator Kriging (IK) interpolation of the response variable since the latter cannot consider any covariates; (3) Our proposed method requires much fewer hyperparameters than deep learning methods, and it still works in the case of limited data (deep learning methods are typically data-hungry (Zhang et al. 2020b)); (4) Our proposed method is more objective than the analytic hierarchy process (AHP) since the latter requires expert opinions (Wu and Wang 2007); (5) Comparing with the principal component analysis (PCA), PCA is an unsupervised procedure that only finds the direction with the highest sample variance, which can be irrespective of the sample class. In contrast, the FDA is a supervised procedure that uses the response variable, which would be a better choice for our data.

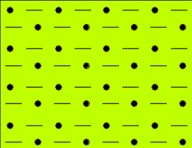
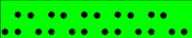



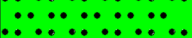


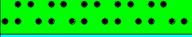





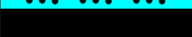

The rest of this paper is organized as follows. The sedimentary characteristics and hydrogeological information of our study area are first presented, and the influencing

factors of the water abundance level are then selected. Subsequently, the assessment procedures are described in detail and validated on available datasets. Finally, the application results are discussed.

Overview of the coal mine study area

The Baodian coal mine, which belongs to the Yanzhou Coal Mining Company Limited, is part of the Yanzhou coalfield and covers an area of 37.01 km². In this paper, we choose the sixth mining area (# 6) of the Baodian coal mine as our study area (Fig. 1). Its stratigraphic profiles are given as (from the bottom to the top): Ordovician (O2), Carboniferous (C2), Permian (P1, P2), Jurassic (J3), and Quaternary (Q). Among all the strata, the main coal-bearing regions are the Shanxi and Taiyuan groups of the Carboniferous stratum and the Permian stratum, which belong to the North China coal-bearing rock series. In these regions, the coal seams are thick, stable, and of good quality, with relatively flat terrains. The overall geological structure is a simple monocline. There are six mineable coal seams in this region, and the main mineable coal seam is named as No.3, which includes No.3^{upper} and No.3^{lower}.

Fig. 2 A representative borehole diagram which provides some lithological strata information and the location of the No.3 coal seam

System	Formation	Average thickness (m)	Column	Marker beds	Lithology
Quaternary	Upper group	76.7		Aquifuge group 1	Clay, sandy clay
	Middle group	59.96		Aquifer group 1	Medium sand
				Aquifuge group 2	Clay
					Sandy clay
	Upper segment of lower group	20.34			Sandy gravel
	Lower segment of lower group	32			Sandy clay
				Aquifer group 2	Fine sand
Permian	Shanxi group				Sandy clay
					Medium sand
		19		Aquifer group 3	Standstone
		5.64		Aquifuge group 3	
		17.07		Aquifer group 4	Standstone
		1.03		No.2 coal seam	
		18.38		Aquifer group 5	Standstone
		8.38		No.3 coal seam	
		19		Aquifer group 6	Standstone

The aquifers that affect the coal mining process are (from the top to the bottom): the lower group loose sandy layer of the Quaternary system, the sandstone of the Jurassic Santai group, the sandstone of the Shanxi group, the limestone of the Taiyuan group, and the limestone of the Benxi group. In this paper, we focus on the first one, namely the lower group loose aquifer of the Quaternary system. A sample schematic is shown in Fig. 2. Because there is an aquifuge (water-resisting layer) above this lower group loose aquifer, precipitation and other boundary conditions have negligible influences on water abundance in this region. This aquifer belongs to the category of the pore-fissure confined aquifer, and the borehole unit flow rates (during borehole construction) are between 0.001 to 0.794 L/m/s (Qiao et al. 2020).

Before the start of mining operations in our study area, 30 boreholes were drilled to explore the hydrogeological conditions of the strata above the working face. Based on these 30 boreholes, we determine that the range of the aquifer initial depth is from 133 m to 214.9 m. Taking borehole “2010-4” as an example, the construction began on January 4th, 2011 and ended on January 25th, 2011, with the completion of a 355.61 m drilling footage. From the borehole log, the buried depth of the lower group aquifer of the Quaternary system is from 214.9 m to 221.88 m; in other words, the aquifer thickness is 6.98 m. The loose stratum is mainly composed of quartz, followed by a certain amount of feldspar, and its permeability is good. The core recovery rate is 92.62%. The consumption of the drilling fluid is about 0.12 m³/h.

Water abundance level and its influencing factors

With regard to the actual water abundance level, according to the “China exploration specification of hydrogeology and engineering geology in mining areas (Appendix C)” and with reference to the actual hydrogeological conditions of the sixth mining area of the Baodian coal mine, we categorize the borehole unit flow rate q (L/m/s) (during borehole construction) into five levels by labeling the values in the range $q \leq 0.05$ as 1, $0.05 < q \leq 0.1$ as 2, $0.1 < q \leq 1$ as 3, $1 < q \leq 5$ as 4, and $q > 5$ as 5. We regard these numbers as the actual water abundance level.

As for the influencing factors, based on the “China provisions on prevention and control of water in coal mines”, and by summarizing the previous research on the water abundance of aquifers and water inrush risk evaluation, we think that the main influencing factors are the aquifer thickness (m), the thickness ratio between sandy and clayey layers (-), the consumption of the drilling fluid (m³/h), the core recovery rate (%), and the hydraulic

conductivity (m/d). The supporting evidences are provided below.

- Generally speaking, the aquifer water abundance is proportional to the thickness of the aquifer (Gao et al. 2018; Wei et al. 2020), which is as expected.
- The thickness ratio between sandy and clayey layers, which reflects the relative content of the sand (gravel) and the clay in the stratum, is used to represent the sedimentary characteristics of an aquifer. This factor would increase the reliability of our model compared with the unit water inflow method. In general, if this ratio gets larger, the aquifer’s water storage capacity would be higher (Wei et al. 2020).
- According to the characteristics and technical requirements of the borehole construction, the original mud is selected as the drilling fluid for the entire hole, and the consumption (circulation loss) of the drilling fluid can be used to indicate the water abundance for a certain rock stratum. While drilling the aquifer, the drilling fluid is consumed to a certain degree. If the consumption of the drilling fluid is high, fractures in the rock stratum might be well-connected and highly developed, indicating a large water flow capacity and a high water abundance level.
- In the borehole construction, the core recovery rate is used to indicate the fragmentation degree of the rock stratum and the intersection degree of the rock fractures. If the integrity of the rock stratum is poor and the fragmentation degree is relatively high (high water storage capacity), the measured value of the core recovery rate may be relatively low (Wei et al. 2020).
- Hydraulic conductivity is closely related to the particle size composition of the rock and soil mass, the distribution characteristics of the pores, the development of the fissures, and the physical properties of the fluid in an aquifer. In Gao et al. (2018), it was revealed that water inflow increases as a power function with respect to the permeability coefficient.

In the following paragraphs, we name above factors as the attributes, and name the water abundance level as the response. Having correctly identified the attributes and response, we can collect our original data which is attached in the Appendix A. Using the original data, we can perform an inversion analysis through Kriging spatial interpolation (Zhang and Zhu 2018b) to obtain the spatial variation contour map of each attribute. The interpolation results are depicted in Fig. 3. It can be seen from Fig. 3 that different attributes possess different distributions, and some distributions are multimodal. Thus, it is necessary to develop a comprehensive analysis algorithm to predict the water abundance, which is discussed in the following section.

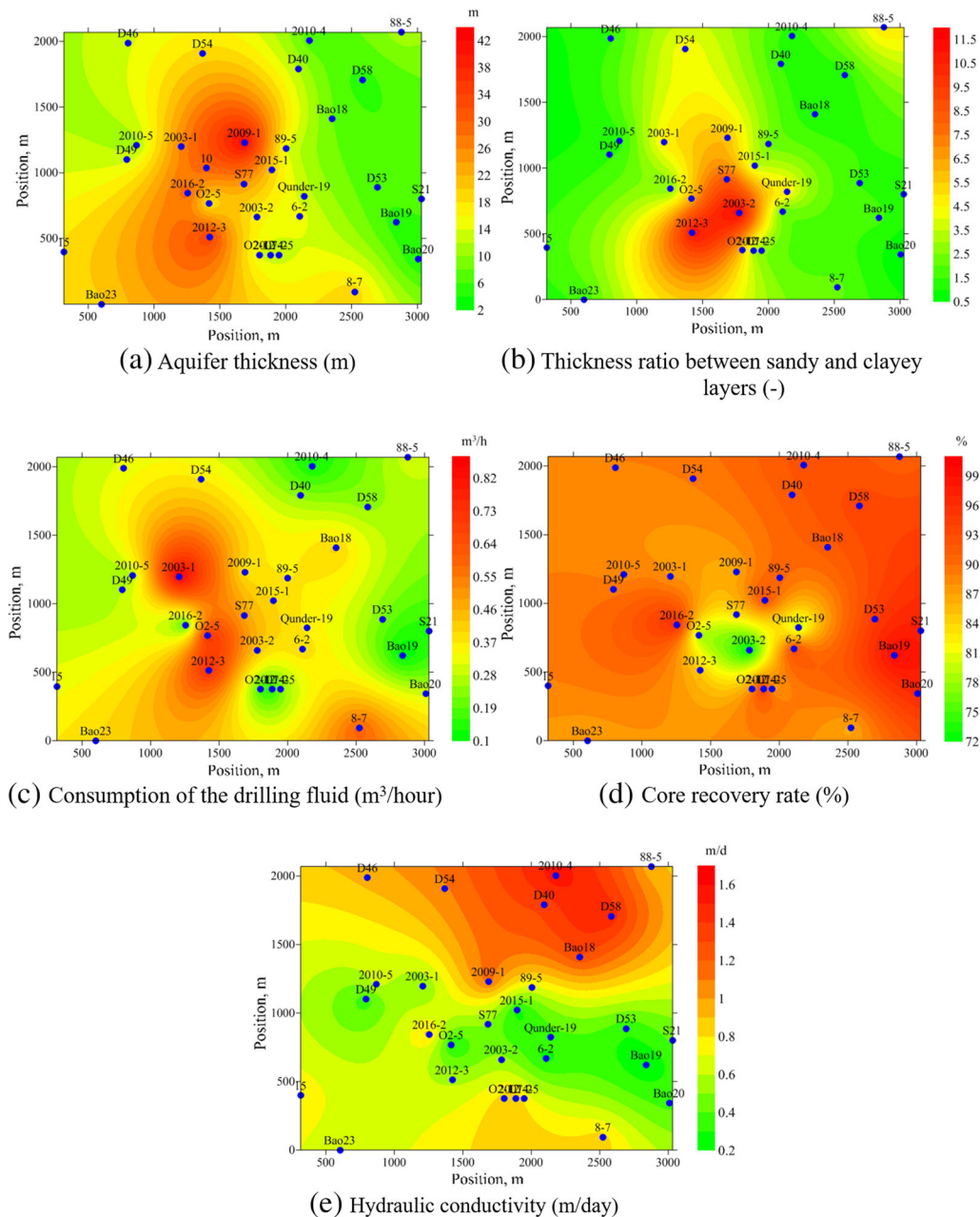


Fig. 3 Graphical overview of influencing factors for the lower group aquifer water abundance in our study area. The trends in (b) and (e) seem to be incompatible with each other; however, this could be true in real geological strata. On one hand, when the sandy layer is thick, the water saturation in the sandy layer could be low, and in addition, a low-permeability flow barrier could significantly affect the system drainage capacity (that is why $k_V < k_H$ for most strata). On the other

hand, when the clayey layer is thick, the mining-induced residual fracture aperture can increase the hydraulic conductivity. This intuitive incompatibility confirms the need for a sophisticated method to predict the water abundance in our study area. Note three sample labels overlap with each other, and they are (from left to right): O2-12, 2007-2, and L14-5

Water abundance assessment process

Fisher discriminant analysis for dimensionality reduction

The Fisher discriminant analysis (FDA) is a popular supervised dimensionality reduction technique which aims

to project a high-dimensional point onto an embedded low-dimensional space, while still keeping the information in the original dataset as much as possible (Hou et al. 2016; Mika et al. 1999; Qian et al. 2016; Sugiyama, 2007; Zhang et al. 2013, 2020c). In the low-dimensional space, the machine learning algorithms generally converge faster since each sample has smaller number of attributes, sometimes in 1D,

2D, or 3D, we can classify the samples visually. The FDA is briefly described as follows. By default, all the vectors are assumed to be column vectors.

Suppose we have n labeled samples $x_1, x_2, \dots, x_n \in \mathbb{R}^p$ where p is the number of attributes or covariates, they form a $n \times p$ sample matrix X :

$$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}. \quad (1)$$

For each sample $i = 1, 2, \dots, n$, we also have a label $y_i \in \{1, 2, \dots, K\}$ that tells which class does the sample belong to, and K is the number of classes ($K = 2$ is called binary classification and $K \geq 3$ is called multiclass classification). Now we want to find a transformation matrix $W \in \mathbb{R}^{p \times r}$ where $r \leq \min\{K - 1, p\}$ is the dimension of the reduced space such that

$$Z = \begin{pmatrix} z_1^T \\ z_2^T \\ \vdots \\ z_n^T \end{pmatrix} = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix} W \quad (2)$$

still remains the large between-class separability and small within-class scatter. In Eq. 2, $z_i \in \mathbb{R}^r = W^T x_i$ for $i = 1, 2, \dots, n$ stand for the transformed samples in the reduced space. The W for FDA is obtained by solving the following optimization problem with variable $W \in \mathbb{R}^{p \times r}$:

$$\begin{aligned} \max. \quad & \text{Tr}(W^T S^{(b)} W) \\ \text{s. t.} \quad & W^T S^{(w)} W = I_r, \end{aligned} \quad (3)$$

where $S^{(b)} \in \mathbb{S}_+^{p \times p}$ represents the between-class symmetric scatter matrix (positive semi-definite), $S^{(w)} \in \mathbb{S}_{++}^{p \times p}$ represents the within-class symmetric scatter matrix (positive definite), and I_r is the $r \times r$ identity matrix. In FDA, the $S^{(b)}$ and $S^{(w)}$ are constructed as follows (Mika et al. 1999; Sugiyama 2007):

$$S^{(b)} = \sum_{k=1}^K n_k (\mu_k - \mu) (\mu_k - \mu)^T, \quad (4)$$

$$S^{(w)} = \sum_{i=1}^n (x_i - \mu_{y_i}) (x_i - \mu_{y_i})^T, \quad (5)$$

where n_k is the number of samples that belong to class k which can be mathematically represented using indicator function $\mathbf{1}_{(\cdot)}$ ($k = 1, \dots, K$):

$$n_k = \sum_{i=1}^n \mathbf{1}_{(y_i=k)}, \quad (6)$$

and $\sum_{k=1}^K n_k = n$, $\mu_k \in \mathbb{R}^p$ is the mean of samples in the class k and $\mu \in \mathbb{R}^p$ is the mean of all samples ($k = 1, \dots, K$):

$$\mu_k = \frac{1}{n_k} \sum_{i=1}^n \mathbf{1}_{(y_i=k)} \times x_i, \quad (7)$$

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i. \quad (8)$$

The solution to Eq. 3 is called the symmetric generalized eigenvalue problem:

$$S^{(b)} \varphi = \lambda S^{(w)} \varphi. \quad (9)$$

This problem has p real non-negative eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, and the p corresponding right eigenvectors $\varphi_1, \varphi_2, \dots, \varphi_p \in \mathbb{R}^p$ can be chosen to be $S^{(w)}$ orthogonal to each other, i.e., $\varphi_i^T S^{(w)} \varphi_j = \mathbf{1}_{(i=j)}$. The optimal point W^* of Eq. 3 is:

$$W^* = (\varphi_1 \cdots \varphi_r), \quad (10)$$

and the corresponding optimal objective $\text{Tr}(W^{*T} S^{(b)} W^*)$ is $\lambda_1 + \dots + \lambda_r$. There is a subtle point of W^* which is the sign of the eigenvectors, thus in order to make the W^* unique, we look at the sign of

$$\varphi(\arg\max_{\text{index}=1, \dots, p} |\varphi(\text{index})|) \quad (11)$$

and multiply φ with -1 if this sign is negative, where we use the angle bracket $\langle \cdot \rangle$ to represent a component of the vector $\varphi \in \mathbb{R}^p$. Above operation is repeated for every column of W^* .

Classification in the reduced space

After performing FDA, we have a low-dimensional representation $Z \in \mathbb{R}^{n \times r}$, and in this reduced space, we can train our classification algorithm to learn the data and make predictions of the water abundance level. In this paper, three classification algorithms are considered.

Quadratic discriminant analysis

The quadratic discriminant analysis (QDA) (James et al. 2013) is one of the most widely used generative

classification algorithm in machine learning. In this algorithm, we assume a prior distribution of $\text{Prob}(Y = k) = \pi_k$ where $k = 1, \dots, K$, and a conditional distribution of $Z \in \mathbb{R}^r$ for a given Y which is a multivariate Gaussian, *i.e.*:

$$\text{Prob}(Z = z | Y = k) = \frac{1}{(2\pi)^{r/2} \sqrt{\det \Sigma_k}} \exp \left[-\frac{(z - v_k)^T \Sigma_k^{-1} (z - v_k)}{2} \right], \quad (12)$$

where $v_k \in \mathbb{R}^r$ and $\Sigma_k \in \mathbb{S}_{++}^{r \times r}$ are mean vector and covariance matrix (positive definite) for class $k = 1, \dots, K$, respectively. According to the Bayes rule, the posterior distribution is obtained as:

$$\text{Prob}(Y = k | Z = z) = \frac{\text{Prob}(Z = z | Y = k) \times \text{Prob}(Y = k)}{\text{Prob}(Z = z)}, \quad (13)$$

and we choose the class with the highest $\text{Prob}(Y = k | Z = z)$. We notice that the denominator of above equation does not depend on the class k , therefore we only need to compare the numerator. If we take the natural logarithm of the numerator and further ignore the constant that is the same for all classes, we end up with the score function (James et al. 2013)

$$\delta_k(z) = -\frac{1}{2} \log \det \Sigma_k - \frac{1}{2} (z - v_k)^T \Sigma_k^{-1} (z - v_k) + \log \pi_k \quad (14)$$

for $k = 1, 2, \dots, K$. Again, we choose the class with the highest $\delta_k(z)$. From Eq. 14, we can see that the score for each class depends on three factors: the generalized variance $\det \Sigma_k$, the prior probability π_k , and the squared Mahalanobis distance from z to the mean v_k , which seems to be a pretty comprehensive and convincing measure. The reason for the name “quadratic discriminant analysis” is because for two classes k and l , the decision boundary given by $\delta_k(z) = \delta_l(z)$ is a quadratic function of z due to $\Sigma_k \neq \Sigma_l$.

The remaining task is to estimate π_k , v_k , and Σ_k for all $k = 1, \dots, K$ from the training data. Here we shall derive the maximum likelihood estimation (MLE) results. After changing of variables such that $S_k = \Sigma_k^{-1}$ for all $k = 1, \dots, K$ and considering the affine constrain that $\sum_{k=1}^K \pi_k = 1$, the log-likelihood of the data is given as:

$$\begin{aligned} \ell(\pi_1, v_1, S_1, \dots, \pi_K, v_K, S_K, \xi) = & -\frac{nr}{2} \log(2\pi) + \sum_{i=1}^n \log \pi_{y_i} + \frac{1}{2} \sum_{i=1}^n \log \det S_{y_i} \\ & - \frac{1}{2} \sum_{i=1}^n (z_i - v_{y_i})^T S_{y_i} (z_i - v_{y_i}) - \xi \left(\sum_{k=1}^K \pi_k - 1 \right), \end{aligned} \quad (15)$$

where ξ is called the Lagrangian multiplier. Next we maximize ℓ with respect to all the parameters, *i.e.*, calculate

the partial derivatives and set them to zero, we get for $k = 1, \dots, K$ and ξ :

$$\frac{\partial \ell}{\partial \xi} = \sum_{k=1}^K \pi_k - 1 = 0, \quad (16)$$

$$\frac{\partial \ell}{\partial \pi_k} = \frac{\sum_{i=1}^n \mathbf{1}_{(y_i=k)}}{\pi_k} - \xi = \frac{n_k}{\pi_k} - \xi = 0, \quad (17)$$

$$\frac{\partial \ell}{\partial v_k} = \sum_{i=1}^n S_k (z_i - v_k) \times \mathbf{1}_{(y_i=k)} = 0, \quad (18)$$

$$\frac{\partial \ell}{\partial S_k} = \frac{n_k}{2} S_k^{-1} - \frac{1}{2} \sum_{i=1}^n (z_i - v_k)^T (z_i - v_k) \times \mathbf{1}_{(y_i=k)} = 0. \quad (19)$$

Here the definition of n_k is exactly the same as in Eq. 6. By solving above equations simultaneously, we get our MLE results that are given as follows (we add a hat symbol on all the parameters to emphasize that they come from the training data):

$$\hat{\xi} = n, \quad (20)$$

$$\hat{\pi}_k = \frac{n_k}{n}, \quad (21)$$

$$\hat{v}_k = \frac{1}{n_k} \sum_{i=1}^n z_i \times \mathbf{1}_{(y_i=k)}, \quad (22)$$

$$\hat{\Sigma}_k = \frac{1}{n_k} \sum_{i=1}^n (z_i - \hat{v}_k)^T (z_i - \hat{v}_k) \times \mathbf{1}_{(y_i=k)}. \quad (23)$$

We need to mention that above results are valid only when $n_k \geq 1$ and $\hat{\Sigma}_k > 0$ for all $k = 1, \dots, K$, the discussions of other cases are beyond the scope of this paper. In practice, for Eq. 23, instead of the MLE result, people also use the unbiased estimation of the covariance matrix which just replaces the denominator n_k with $n_k - 1$ given $n_k \geq 2$, *i.e.*:

$$\hat{\Sigma}_k = \frac{1}{n_k - 1} \sum_{i=1}^n (z_i - \hat{v}_k)^T (z_i - \hat{v}_k) \times \mathbf{1}_{(y_i=k)}. \quad (24)$$

Support vector machines

Support vector machines (SVM) were originally designed for the binary classification (Ke et al. 2020; Panahi et al. 2020; Yadav et al. 2020), and it is probably the best “off-the-shelf” supervised learning algorithm (Hastie et al. 2009; Ke et al. 2020). Here we apply the SVM (with linear kernel) to the multiclass classification by using the one-against-all method. Since we have totally K classes, we could construct K SVM models. The k^{th} SVM is trained with all of the examples in the k^{th} class with positive

labels, and all the other examples with negative labels. It solves the following convex optimization problem (Boyd and Vandenberghe 2004) with variables $w^k \in \mathbb{R}^r$, $b^k \in \mathbb{R}$, and $s^k \in \mathbb{R}^n$:

$$\begin{aligned} \min. \quad & \frac{1}{2} \|w^k\|_2^2 + C \sum_{i=1}^n s_i^k \\ \text{s. t.} \quad & (w^k)^T z_i + b^k \geq 1 - s_i^k, \text{ if } y_i = k \\ & (w^k)^T z_i + b^k \leq -1 + s_i^k, \text{ else} \\ & s_i^k \geq 0, \end{aligned} \quad (25)$$

where C is called the penalty parameter and \geq for a vector means element-wise operation (*i.e.*, every component of s^k should be non-negative). In the following application, C is always assumed to be 1 and we do not tune it. The basic concept behind SVM is to search for a balance between a large margin and a small number of misclassified points. After solving (25) for each $k = 1, \dots, K$ using the primal-dual interior-point method (Boyd and Vandenberghe 2004), we have K decision functions and now for a given $z \in \mathbb{R}^r$, we assign it to the class that has the largest value of the decision function, *i.e.*:

$$y_{\text{pred}} = \underset{k=1, \dots, K}{\operatorname{argmax}} \left[(\hat{w}^k)^T z + \hat{b}^k \right]. \quad (26)$$

Softmax regression (Generalized linear model)

Softmax regression is a generalized version of logistic regression for more than 2 classes' scenario (discriminative classification algorithm), and both of them are just special cases of so-called generalized linear model (GLM) which assumes the response variable Y follows the exponential family distribution given our sample $z \in \mathbb{R}^r$ and model parameter Θ . This can be written in the form (Ng 2020):

$$\operatorname{Prob}(Y = y | z; \Theta) = b(y) \exp \left[\hat{z}^T \Theta^T T(y) - a(\Theta \hat{z}) \right], \quad (27)$$

where $\hat{z} \in \mathbb{R}^{r+1}$ means we have added a single row with the value 1 to z , $\Theta \hat{z}$ is always denoted as the natural parameter η (a vector), $T(y)$ is called the sufficient statistic and it has the same shape as η , $a(\eta) \in \mathbb{R}$ is the log-partition function, and $b(y) \in \mathbb{R}_{++}$ is the base measure.

In the softmax regression, we have totally K classes to classify and we always choose a reference class, therefore our model parameter Θ has the size $\Theta \in \mathbb{R}^{(K-1) \times (r+1)}$. Without loss of generality, let us assume the K^{th} class is the reference class, and we have:

$$\eta = \Theta \hat{z} = \begin{pmatrix} \theta_1^T \hat{z} \\ \vdots \\ \theta_{K-1}^T \hat{z} \end{pmatrix}, \quad (28)$$

where each $\theta_k \in \mathbb{R}^{r+1}$ has length $r+1$ for $k = 1, \dots, K-1$. For $T(y)$, it is a length $K-1$ vector with component $(T(y))_k = \mathbf{1}_{(y=k)}$, and $b(y) = 1$. For $a(\eta)$, the expression is given as:

$$a(\Theta \hat{z}) = \log \left(1 + \sum_{k'=1}^{K-1} \exp(\theta_{k'}^T \hat{z}) \right). \quad (29)$$

Now we can substitute above expressions back into the exponential family distribution (27), and we get a multinomial distribution of Y given z and Θ which takes the form (Ng 2020):

$$\operatorname{Prob}(Y = k | z; \Theta) = \frac{\exp(\theta_k^T \hat{z})}{1 + \sum_{k'=1}^{K-1} \exp(\theta_{k'}^T \hat{z})}, \quad (30)$$

where $k = 1, \dots, K-1$ and

$$\operatorname{Prob}(Y = K | z; \Theta) = \frac{1}{1 + \sum_{k'=1}^{K-1} \exp(\theta_{k'}^T \hat{z})}. \quad (31)$$

The right-hand side of either (30) or (31) is sometimes called the softmax function because it guarantees the output is bounded between 0 and 1, and that's how this method got its name. Finally, we need to do parameter fitting. We begin by writing down the log-likelihood of n training samples (z_i, y_i) ($i = 1, 2, \dots, n$) as:

$$\ell(\theta_1, \dots, \theta_{K-1}) = \sum_{i=1}^n \log \left\{ \prod_{k=1}^K \left[\frac{\exp(\theta_k^T \hat{z}_i)}{\sum_{k'=1}^K \exp(\theta_{k'}^T \hat{z}_i)} \right]^{\mathbf{1}_{(y_i=k)}} \right\}, \quad (32)$$

where we have defined $\theta_K \equiv 0$ and this is not the parameter that we need to fit. This is an unconstrained convex optimization problem (Boyd and Vandenberghe 2004) and we can solve it using either gradient method or Newton's method (Boyd and Vandenberghe 2004) to get the estimated parameters $\hat{\theta}_1, \dots, \hat{\theta}_{K-1}$, in which the iteration update rules are:

$$\mathbf{Vec}[\Theta^{(t+1)}] = \mathbf{Vec}[\Theta^{(t)}] + \alpha \frac{\partial \ell}{\partial \{\mathbf{Vec}[\Theta^{(t)}]\}} \quad (33)$$

for gradient method, and

$$\begin{aligned} \mathbf{Vec}[\Theta^{(t+1)}] &= \mathbf{Vec}[\Theta^{(t)}] \\ &- \alpha \left[\frac{\partial^2 \ell}{\partial \{\mathbf{Vec}[\Theta^{(t)}]\} \partial \{\mathbf{Vec}[\Theta^{(t)}]\}} \right]^{-1} \frac{\partial \ell}{\partial \{\mathbf{Vec}[\Theta^{(t)}]\}} \end{aligned} \quad (34)$$

for Newton's method. The **Vec** symbol means that we have flattened Θ into a column vector of length $(K - 1) \times (r + 1)$. α is known as the step size or learning rate, which can be chosen from exact line search or backtracking line search for convex problems (Boyd and Vandenberghe 2004), while for non-convex problems such as training a neural network, we treat α as a pre-specified hyperparameter (always very small) before the training process. Finally, for a new input, we assign it to the class that has the largest probability using estimated parameters substituted into Eqs. 30 and 31.

Estimation of the test error through cross-validation

To evaluate the performance of a machine learning algorithm, we always rely on the test error due to the bias-variance tradeoff. Since if we rely on the training error, we end up selecting a high-variance model, which is often a bad choice. Furthermore, in our application the borehole data is sparse due to the expensive costs of drilling, so we still want to use as much data as possible to train our model. In this situation, we consider a method called leave-one-out cross-validation (LOOCV), which means that we repeatedly train on all but one of the training examples in our dataset, and test on that held-out example. For our application, the LOOCV algorithm goes as follows:

1. For $i = 1, 2, \dots, n$ ($n = 30$ in our application)
 - (a) Perform FDA and train each of our three classification algorithms (QDA, SVM, GLM), *all on* $x_1 \cup \dots \cup x_{i-1} \cup x_{i+1} \cup \dots \cup x_n$ (*i.e.*, use all data except x_i) to get our hypothesis/model $M_i^{\text{QDA, SVM, GLM}}$. In this process, it is not surprised to see different results of $S^{(b)}$, $S^{(w)}$, and W^* every time because each time we use different $n - 1$ samples.
 - (b) Test model $M_i^{\text{QDA, SVM, GLM}}$ on x_i to tell whether x_i has been misclassified by comparing the predicted class with the true class label y_i . It is possible that for some x_i , different algorithms predict different responses. In addition, in the prediction of x_i , it is required to normalize x_i using the same statistics (mean and standard deviation vectors) used to normalize the training set $x_1 \cup \dots \cup x_{i-1} \cup x_{i+1} \cup \dots \cup x_n$.
2. The estimated test error is then calculated as the total number of misclassified samples divided by n for QDA, SVM, and GLM, respectively.
3. Refit the whole training data to get our final hypothesis/model for QDA, SVM, and GLM, respectively. For the softmax regression, we also output p-values to check the statistical significance ($p \leq 0.05$).

Results and discussions

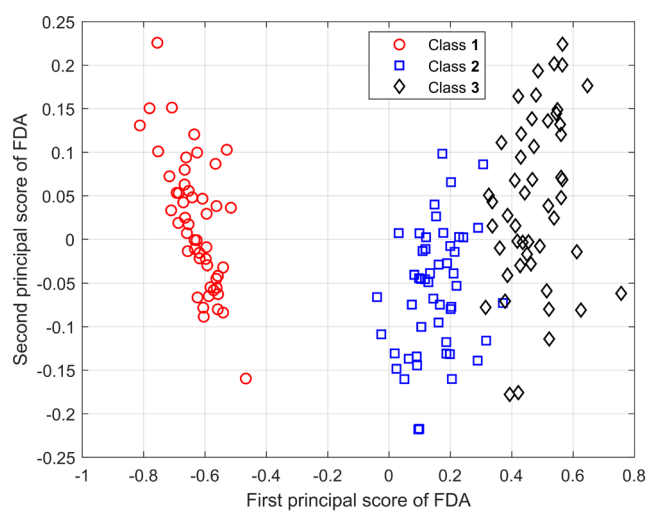
General qualitative comparison of different classification methods

In general, QDA makes stronger modeling assumptions about the data than does softmax regression. It turns out that when these modeling assumptions are correct or at least approximately correct, then QDA would find better fits to the data, and is a better model. Specifically, when the conditional distribution of Z given Y is indeed Gaussian, then QDA is more efficient (*i.e.*, requires less training data to learn “well”). In contrast, by making weaker assumptions, softmax regression is significantly more robust and less sensitive to incorrect modeling assumptions. Specifically, when the data is indeed non-Gaussian, then in the limit of large datasets, softmax regression often performs better than QDA.

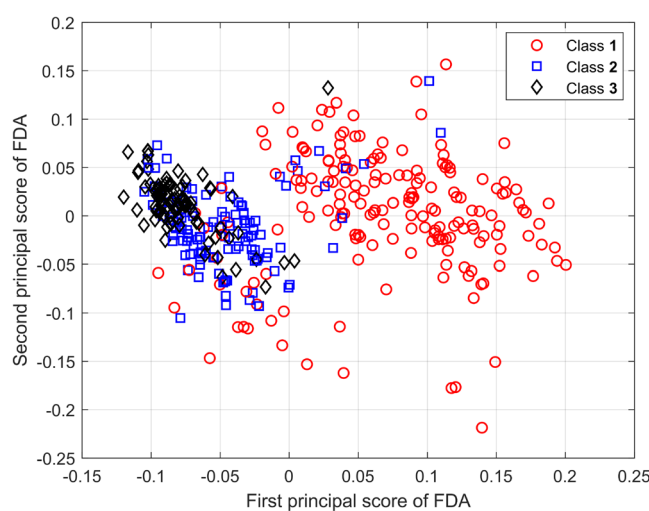
The SVM algorithm is much more geometrically motivated since it is an extension of the support vector classifier (SVC). A SVM is just a SVC applied on an expanded set of predictors (although in our application, we just use SVC). Instead of assuming a probabilistic model, we are trying to find a particular optimal separating hyperplane. The optimal separating plane aims to maximize the width of the margins, with some budget for “violations of the margins” (James et al. 2013). This is the geometric motivation for SVM. We do not have anything resembling the statistical model we use in the generalized linear model or QDA. However, the SVM is more robust to data with very non-linear patterns (for example, the data cannot be separated by linear or quadratic curves). This is achieved by using the kernel trick such as the radial basis function (RBF), and an example is shown in FIGURE 9.8 and FIGURE 9.9 of James et al. (2013). In practice, the SVM with linear kernel and softmax regression should perform comparably. For binary classification, we can even show that the SVM loss and logistic regression loss are quite close to each other (James et al. 2013).

Validation of the FDA

In order to verify the effectiveness of FDA, we apply it to the available sample datasets stored in Matlab Statistics and Machine Learning Toolbox (MathWorks 2021): `carbig.mat` and `fisheriris.mat`. The data pre-processing procedure is described in the next subsection, and here we only display the final results in Fig. 4. For `fisheriris.mat`, when you load the data, the predictor and response variables are already stored in `meas` matrix and `species` cell, respectively. For the `carbig.mat`, the predictor variables are the acceleration, engine displacement, horsepower, and weight of the cars. The response



(a) fisheriris.mat



(b) carbig.mat

Fig. 4 FDA validation results on two sample datasets. For the latter one, the performance of FDA is affected by the choice of categorization and deletion of NaN values (rows)

variable is miles per gallon (MPG) which is continuous, but in order to perform FDA, we need to make it discrete. In this work we categorize MPG into three levels from 9 to 48 mpg by labeling the response values in the range 9–22 as 1, 23–30 as 2, and 31–48 as 3. From Fig. 4, we can see that for the fisheriris.mat dataset, the FDA works pretty well, while for the carbig.mat dataset, its performance is still acceptable. Now we are confident and ready to try our method on the real geological data.

Dimensionality reduction result of the borehole data

For each covariate or attribute, we subtract the mean of the corresponding column, *i.e.*, we center the attributes. Most of the time, we do not care about the absolute numerical value, we care about the value relative to the spread observed in the sample. For many quantities, the choice of unit may lead to an arbitrary scaling of the data. In that case, in addition to centering each attribute or covariate, we also multiply it with a constant to make its sample variance equal to 1. This process is called data normalization. Next, we can calculate $S^{(b)}$ and $S^{(w)}$ from the normalized $X \in \mathbb{R}^{n \times p}$ using Eqs. 4 and 5, and then solve for W^* as described in Section Fisher discriminant analysis for dimensionality reduction. In our application, the total number of classes K is 3, so the size of W^* is $p = 5$ by $r = K - 1 = 2$. The results are obtained as follows:

$$S^{(b)} = \begin{pmatrix} 6.0546 & 5.4272 & 0.3545 & -2.3378 & 7.6803 \\ 5.4272 & 4.9748 & -0.3582 & -1.4832 & 6.3469 \\ 0.3545 & -0.3582 & 4.1716 & -3.8976 & 3.7500 \\ -2.3378 & -1.4832 & -3.8976 & 4.3099 & -5.9557 \\ 7.6803 & 6.3469 & 3.7500 & -5.9557 & 12.3666 \end{pmatrix}, \quad (35)$$

$$S^{(w)} = \begin{pmatrix} 22.9454 & 11.0843 & 15.5247 & -8.6526 & -8.2219 \\ 11.0843 & 24.0252 & 15.2566 & -19.5476 & -11.4532 \\ 15.5247 & 15.2566 & 24.8284 & -12.3432 & -11.1505 \\ -8.6526 & -19.5476 & -12.3432 & 24.6901 & 9.5369 \\ -8.2219 & -11.4532 & -11.1505 & 9.5369 & 16.6334 \end{pmatrix}, \quad (36)$$

$$W^* = \begin{pmatrix} 0.1100 & 0.0907 \\ 0.1688 & 0.2606 \\ -0.0239 & -0.2205 \\ 0.0180 & 0.2246 \\ 0.2780 & -0.1021 \end{pmatrix}. \quad (37)$$

Then we can project the normalized X onto the reduced (discriminant) space, *i.e.*, $Z = XW^*$. The visualization of

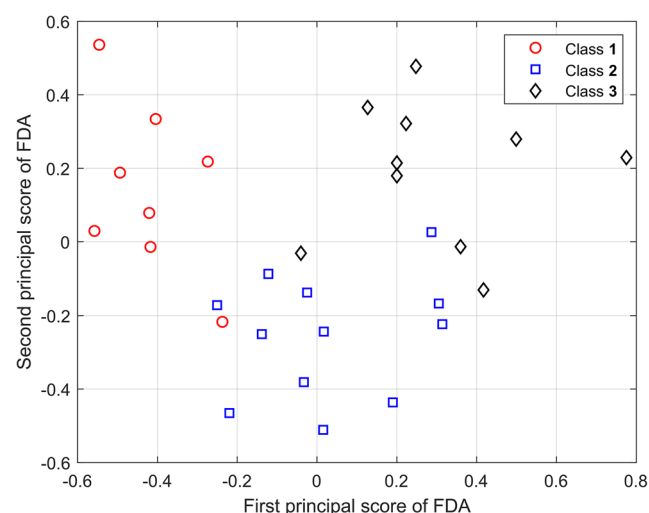


Fig. 5 Scatter diagram of the borehole data points (properties) in the reduced (discriminant) space

Table 1 Misclassified samples in LOOCV (a measure of test error)

Number i	ID	True y_i	QDA	SVM	GLM
1	Qunder-19	1	$y_{\text{pred}} = 2$	$y_{\text{pred}} = 2$	$y_{\text{pred}} = 2$
4	2015-1	1	✓	✓	$y_{\text{pred}} = 3$
6	2003-2	2	$y_{\text{pred}} = 3$	$y_{\text{pred}} = 3$	$y_{\text{pred}} = 1$
7	2010-5	2	$y_{\text{pred}} = 1$	✓	$y_{\text{pred}} = 1$
11	L14-5	3	$y_{\text{pred}} = 2$	$y_{\text{pred}} = 2$	$y_{\text{pred}} = 2$
18	2010-4	3	$y_{\text{pred}} = 2$	$y_{\text{pred}} = 2$	$y_{\text{pred}} = 2$
28	D58	2	✓	$y_{\text{pred}} = 3$	✓

$Z \in \mathbb{R}^{n \times r}$ with different class labels is given in Fig. 5. It can be seen from Fig. 5 that the overall performance of FDA is relatively good because it gives small intra-group dispersion and large inter-group separability.

Test error and training error

The misclassified samples in LOOCV are summarized in Table 1. It is interesting to notice that for $i = 1, 11, 18$,

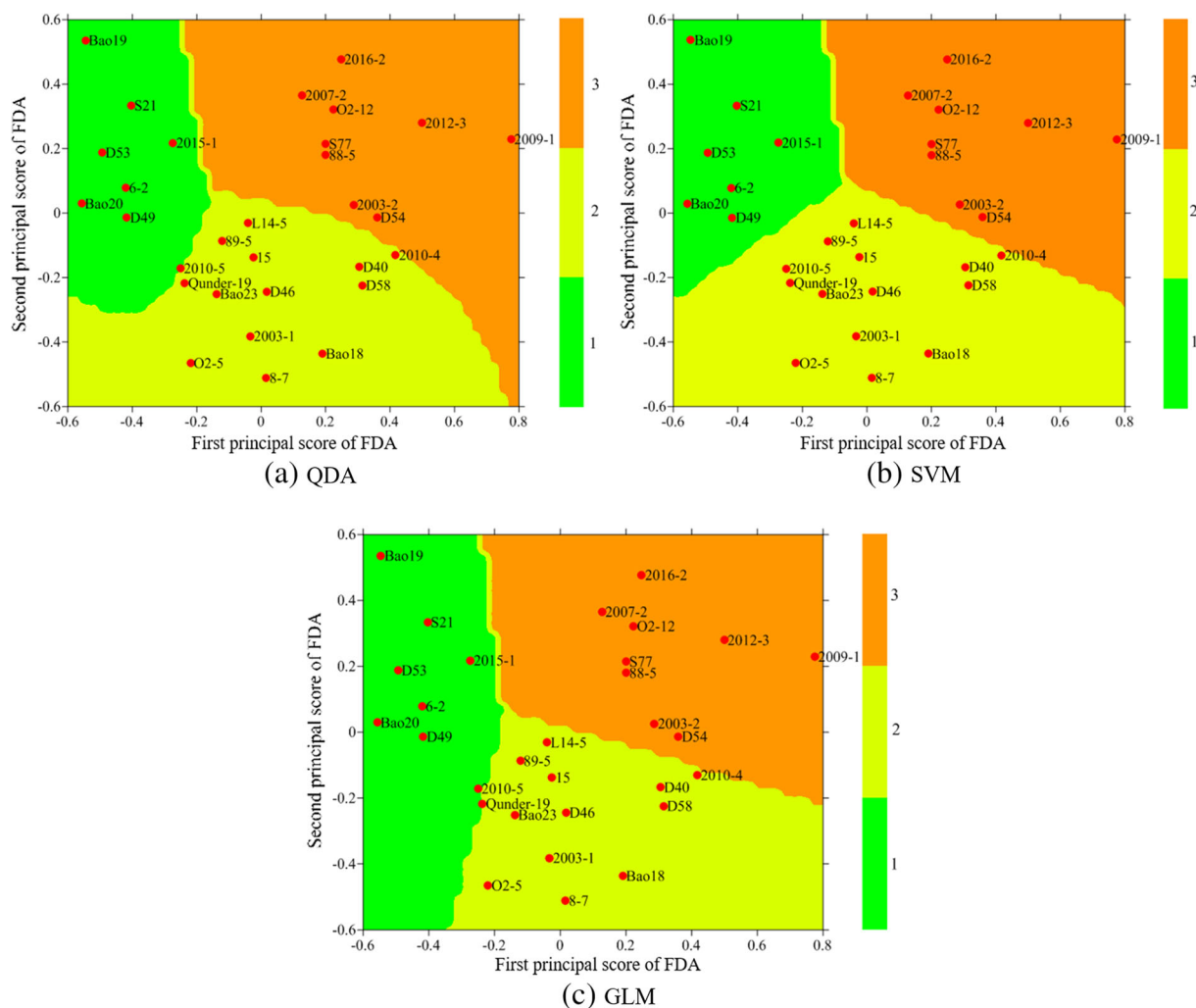


Fig. 6 A visualization of decision boundaries in the reduced (discriminant) space using different classification methods in step 3 of Section Estimation of the test error through cross-validation

all three methods give same \hat{y}_i . We believe this indicates some undiscovered factors that either inhibit the flow rate in Qunder-19 or promote the flow rates in L14-5 and 2010-4. Another possibility is that our proposed model is still biased for these three locations, and further model improvement could be performed. The p-values of the softmax regression are shown in Eq. 38:

$$p = \begin{pmatrix} 0.1437 & 0.9402 \\ 0.0392 & 0.3087 \\ 0.6398 & 0.0501 \end{pmatrix}, \quad (38)$$

which indicates that the first principal score (horizontal axis of Fig. 5) is statistically significant ($p = 0.0392$) on the

relative risk of being level 1 versus being level 3, and the second principal score (vertical axis of Fig. 5) is statistically significant ($p \approx 0.05$) on the relative risk of being level 2 versus being level 3. However, the other four model coefficients including the intercept terms are not statistically significant which might be the reason why GLM performs slightly worse on our dataset. Figure 6 depicts the decision boundaries for each of these three classification algorithms in step 3 of Section [Estimation of the test error through cross-validation](#), from which we can observe the non-linear decision boundaries for QDA (Fig. 6a), while for SVM and GLM, their decision boundaries are approximately linear. We can also find that the decision boundaries of SVM maximize the margin with data points on both sides (Fig. 6b),

Table 2 Training results on the borehole data

ID	True y_i	QDA	SVM	GLM
Qunder-19	1	<u>2</u>	<u>2</u>	<u>2</u>
6-2	1	1	1	1
Bao19	1	1	1	1
2015-1	1	1	1	1
2003-1	2	2	2	2
2003-2	2	<u>3</u>	<u>3</u>	<u>3</u>
2010-5	2	2	2	<u>1</u>
O2-5	2	2	2	2
2016-2	3	3	3	3
O2-12	3	3	3	3
L14-5	3	<u>2</u>	<u>2</u>	<u>2</u>
2007-2	3	3	3	3
D49	1	1	1	1
D53	1	1	1	1
2012-3	3	3	3	3
89-5	2	2	2	2
2009-1	3	3	3	3
2010-4	3	<u>2</u>	<u>2</u>	<u>2</u>
15	2	2	2	2
D40	2	2	2	2
D46	2	2	2	2
D54	3	3	3	3
Bao18	2	2	2	2
8-7	2	2	2	2
Bao20	1	1	1	1
Bao23	2	2	2	2
S77	3	3	3	3
D58	2	2	2	2
S21	1	1	1	1
88-5	3	3	3	3

We have underlined the inconsistent predictions (a measure of training error)

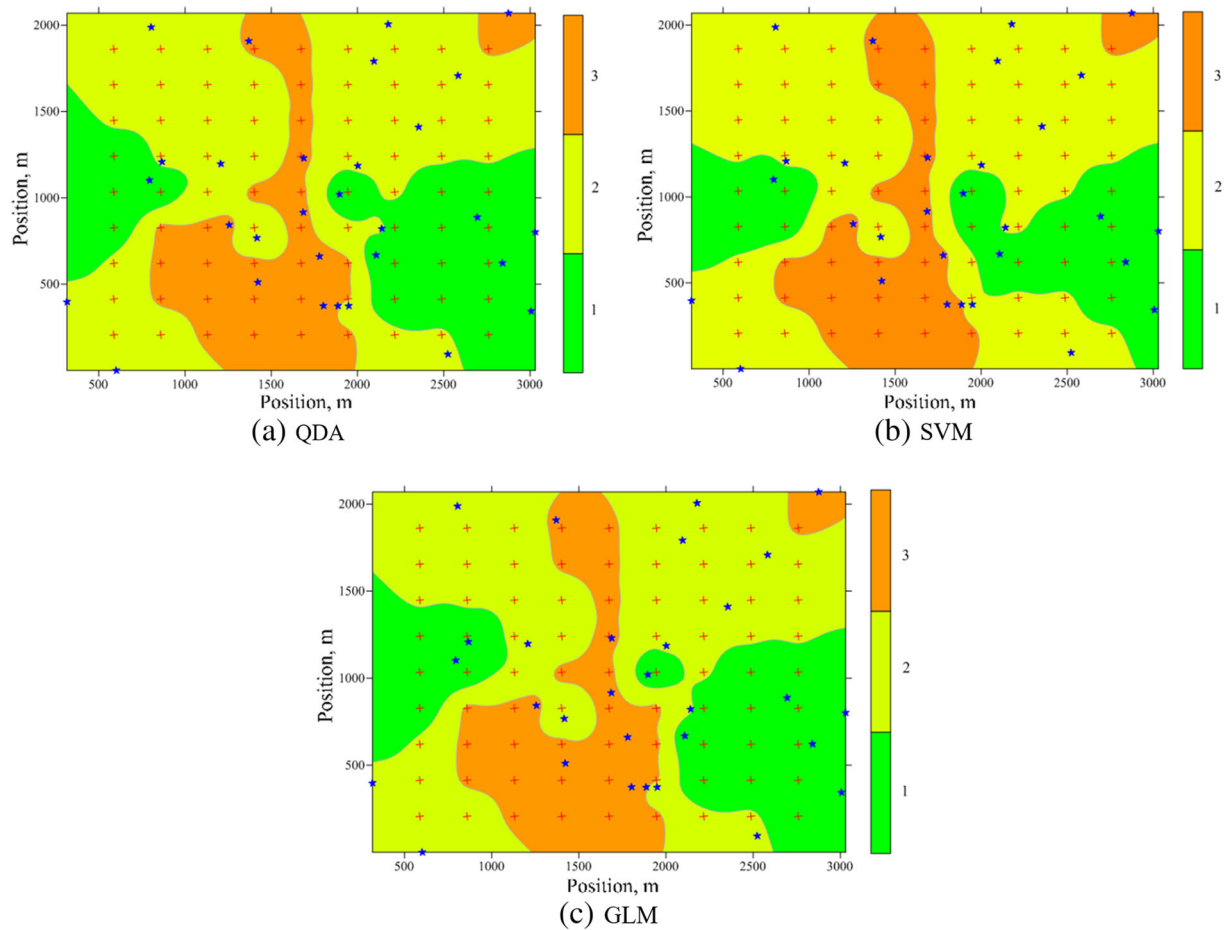


Fig. 7 Zoning map of lower group aquifer water abundance level in our study area using different classification methods. Note the red “plus” marker represents the 81 evenly distributed points and the blue “star” marker represents the 30 borehole points

which is not the case for the other two methods. These decision boundaries can be understood as the projections of hypersurfaces in the original high-dimensional spaces (Ke et al. 2020). From Fig. 6, we could obtain the model prediction values (*i.e.*, the water abundance level) for all borehole points. Table 2 visually compares these prediction values with the true values, while it is important to keep in mind that they are different from the predictions in Table 1: Table 1 is for the test error and Table 2 is for the training error.

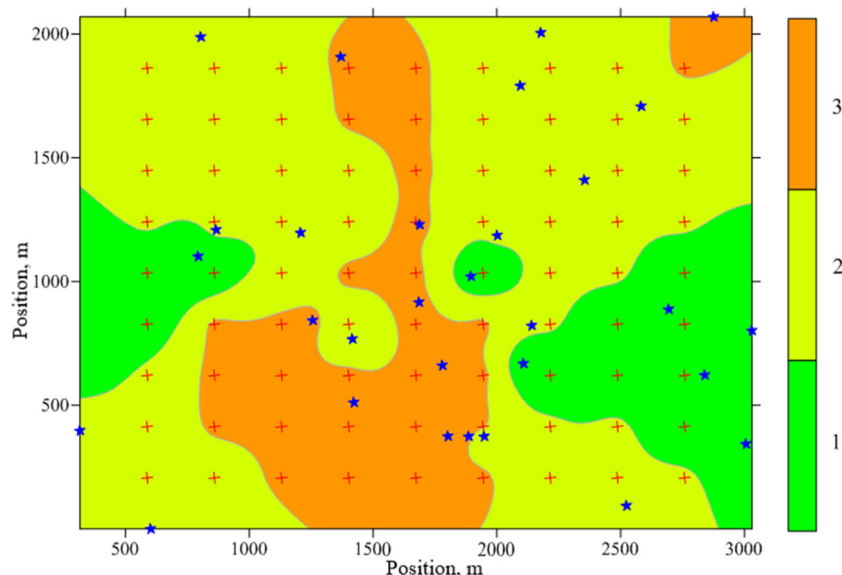
Prediction of the water abundance level

In this section, we make a comprehensive prediction of water abundance level in our study area. To achieve that, we have sampled 81 evenly distributed points in our study area, and their covariates/attributes' values or properties are inferred from Fig. 3. Then we predict the water abundance

level on these 81 points together with the existing 30 borehole points using the trained model of Section [Test error and training error](#). The prediction results are shown in Fig. 7.

It can be seen from Fig. 7 that three classification methods give nearly consistent results of water abundance in our study area. To be more specific, nearly all the western and eastern areas present an extremely weak to weak water abundance. However, the central region presents a medium water abundance, especially for the southern part. In addition, the northeast corner also shows a medium water abundance, which might be attributed to a thin clayey layer (aquifuge) and a high hydraulic conductivity, as shown in Fig. 3. For engineering practice, we should combine the three subfigures of Fig. 7 and obtain the final comprehensive zoning map. Here, we have two options. The first option is based on the conservative prevention principle: we always choose the largest predicted level from three subfigures, which can be mathematically written as:

Fig. 8 The final comprehensive zoning map for option 1. Note the red “plus” marker represents the 81 evenly distributed points and the blue “star” marker represents the 30 borehole points



$\hat{y}_{\text{final}} = \max\{\hat{y}_{\text{QDA}}, \hat{y}_{\text{SVM}}, \hat{y}_{\text{GLM}}\}$. The second option is less conservative and it emphasizes the high-risk area. The detailed explanation can be found in Algorithm 1. Figures 8 and 9 show the final results. Through Figs. 8 and 9, we may conclude that the regions around boreholes 2003-2, 2012-3, S77, O2-12, 2009-1, D54, and 88-5 are more likely to suffer from the mine roof water inrush, which indicates that when coal seams are being mined, in accordance with the corresponding water abundance characteristics of the upper thick loose strata, different mining methods and different coal pillar thickness protection methods should be taken. Furthermore, if we compare the final comprehensive zoning map with the selected influencing factors shown in Fig. 3, we may conclude that the aquifer thickness and consumption of the drilling fluid mainly control the pattern

of water abundance level, while the core recovery rate seems to have negligible effects. As for the hydraulic conductivity, it should have a medium effect on the water abundance level of the north region in our study area.

Algorithm 1 The algorithm of the second option.

```

if  $\hat{y}_{\text{QDA}} = \hat{y}_{\text{SVM}} = \hat{y}_{\text{GLM}} = 3$  then
  |  $\hat{y}_{\text{final}} = 3$ 
else if  $\max\{\hat{y}_{\text{QDA}}, \hat{y}_{\text{SVM}}, \hat{y}_{\text{GLM}}\} = 3$  or
   $\hat{y}_{\text{QDA}} = \hat{y}_{\text{SVM}} = \hat{y}_{\text{GLM}} = 2$  then
  |  $\hat{y}_{\text{final}} = 2$ 
else
  |  $\hat{y}_{\text{final}} = 1$ 
end

```

Fig. 9 The final comprehensive zoning map for option 2 by using Algorithm 1. Note the red “plus” marker represents the 81 evenly distributed points and the blue “star” marker represents the 30 borehole points

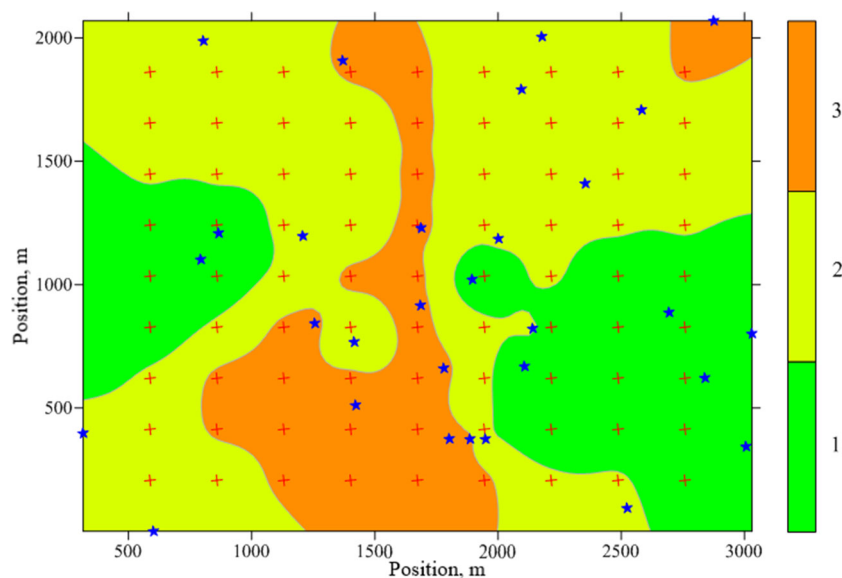


Table 3 The detailed attributes and water abundance level of our training samples

Number i	ID	Th (m)	SC ratio (-)	Q (m ³ /h)	Core (%)	k_h (m/d)	True y_i
1	Qunder-19	12.93	4.92	0.36	78.72	0.3734	1
2	6-2	15.05	2.1	0.40	93.15	0.2624	1
3	Bao19	4.96	1.65	0.10	99.33	0.2231	1
4	2015-1	20	3.43	0.40	92.35	0.2732	1
5	2003-1	26.65	4.76	0.87	87.8	0.4801	2
6	2003-2	16.8	11.42	0.43	71.79	0.5363	2
7	2010-5	10.99	1.19	0.27	88.43	0.6321	2
8	O2-5	17.37	6.22	0.72	78.49	0.2809	2
9	2016-2	29.09	2.5	0.18	96.09	0.8451	3
10	O2-12	13.8	5.75	0.17	87.98	0.8435	3
11	L14-5	13.2	2.73	0.23	87.42	0.7523	3
12	2007-2	13.35	3.13	0.14	95.05	0.8969	3
13	D49	13.5	1.43	0.28	89.55	0.3409	1
14	D53	6.5	1.48	0.22	94.63	0.3213	1
15	2012-3	34.07	10.88	0.74	84.84	0.5713	3
16	89-5	14.15	2.36	0.37	90.37	0.6676	2
17	2009-1	41.88	5.88	0.38	86.48	1.1452	3
18	2010-4	6.98	1.16	0.12	92.62	1.5637	3
19	15	18.55	1.73	0.33	88.86	0.7739	2
20	D40	11.87	1.5	0.24	91.57	1.3265	2
21	D46	13.68	1.44	0.32	88.93	0.9349	2
22	D54	18.78	4.38	0.37	89.15	1.0839	3
23	Bao18	8.05	1.14	0.41	91.51	1.2946	2
24	8-7	19.35	2.15	0.60	86.73	0.8371	2
25	Bao20	3.82	0.51	0.30	96.35	0.3652	1
26	Bao23	19.78	1.61	0.40	87	0.6285	2
27	S77	24.2	8.36	0.48	83.37	0.4988	3
28	D58	5.14	1.22	0.24	93.61	1.4675	2
29	S21	7.8	1.17	0.15	97.38	0.4217	1
30	88-5	11.34	5.12	0.30	91.08	0.9152	3

Closure

This paper presents a novel off-the-shelf combined evaluation method to predict the spatial characteristics of water abundance in the loose aquifer. Based on an analysis of the underlying data within the area, five main factors (attributes) that influence aquifer water abundance were selected: the aquifer thickness, the thickness ratio between sandy and clayey layers, the consumption of the drilling fluid, the core recovery rate, and the hydraulic conductivity. The spatial variation of each assessment attribute in our study area was investigated, and its relative importance to the water abundance level was incorporated into our proposed two-step advanced mathematical model. This study was encouraged by the advancement of machine learning in the past decade. It is the first time, to the authors' knowledge, that FDA and several classification methods have been integrated within

the context of mine water and the environment. Based on the case study of the Baodian coal mine, our model exhibits a relatively high prediction accuracy, and the zoning map of the water abundance in our study area is of great importance. This model provides a useful reference for the water abundance in the nearby mining areas of the Yanzhou coalfield.

Appendix A: Original data

The Baodian coal mine has a geological type of thick loose strata and thin bedrock (Li et al. 2017). The water abundance of the lower group aquifer of this thick loose stratum has a significant impact on the actual coal mining process. Therefore, it is scientifically feasible to choose the Baodian coal mine as our study area. As a result, 30 samples are collected from the hydrogeological borehole data. The

details are presented in the following Table 3, where “ID” denotes the corresponding borehole ID, “Th” denotes the aquifer thickness, “SC ratio” denotes the thickness ratio between sandy and clayey layers, Q represents the consumption of the drilling fluid, “Core” indicates the core recovery rate, and k_h represents the hydraulic conductivity.

Acknowledgments This paper was supported by grants from the National Natural Science Foundation of China (51774199) and the Natural Science Foundation of Shandong Province for the support of major basic research projects (ZR2018ZC0740). The authors would like to sincerely thank the Yanzhou Coal Mining Company Limited for providing the hydrogeological data.

Declarations

Competing Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press. <https://web.stanford.edu/~boyd/cvxbook/>
- Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *CATENA* 151:147–160
- Chen L, Feng X, Xu D, Zeng W, Zheng Z (2018) Prediction of water inrush areas under an unconsolidated, confined aquifer: the application of multi-information superposition based on GIS and AHP in the Qidong coal mine, China. *Mine Water Environ* 37:786–795
- Gao R, Yan H, Ju F, Mei X, Wang X (2018) Influential factors and control of water inrush in a coal seam as the main aquifer. *Int J Min Sci Technol* 28:187–193
- Guo W, Zhao J, Yin L, Kong D (2017) Simulating research on pressure distribution of floor pore water based on fluid-solid coupling. *Arab J Geosci* 10
- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. Springer. <https://web.stanford.edu/~hastie/ElemStatLearn/>
- Hou E, Tong R, Wang S, Feng J, Chen T (2016) Prediction method for the water enrichment of weathered bedrock based on fisher model in northern shaanxi jurassic coalfield. *J Chin Coal Soc (Chin)* 41:2312–2318
- James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning with applications in R. Springer. <https://www.statlearning.com/>
- Ke Q, Tian X, Bricker J, Tian Z, Guan G, Cai H, Huang X, Yang H, Liu J (2020) Urban pluvial flooding prediction by machine learning approaches a case study of Shenzhen city. *China Adv Water Resour* 145:103719
- Li B, Zhang W, Ma L (2017) Influencing factors and prediction of mine water inrush disaster under thick unconsolidated layers and thin bedrock. *J Shandong Univ Sci Technol (Nat Sci) (Chin)* 36
- Liu W, Li Q, Zhao J, 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
- Liu W, Pang L, Xu B, Sun X (2020) Study on overburden failure characteristics in deep thick loose seam and thick coal seam mining. *Geomatics Nat Hazards Risk* 11:632–653
- MathWorks (2021) Sample data sets. <https://ww2.mathworks.cn/help/stats/sample-data-sets.html?lang=en>
- Meng Z, Li G, Xie X (2012) A geological assessment method of floor water inrush risk and its application. *Eng Geol* 143:144: 51–60
- Mika S, Ratsch G, Weston J (1999) Fisher discriminant analysis with kernels. In: *Neural Networks Signal Process. IX Proc. 1999 IEEE Signal Process. Soc. Work.*, pp 41–48
- Naghibi SA, Hashemi H, Berndtsson R, Lee S (2020) Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors. *J Hydrol* 589:125197
- Ng A (2020) Stanford CS 229: Machine Learning lecture notes. <http://cs229.stanford.edu/syllabus-spring2019.html>
- Panahi M, Sadhasivam N, Pourghasemi HR, Rezaie F, Lee S (2020) Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). *J Hydrol* 588:125033
- Phan T-T-H, Nguyen XH (2020) Combining statistical machine learning models with ARIMA for water level forecasting: The case of the Red river. *Adv Water Resour* 142:103656
- Qian Z, Ren G, Chu F, Qin S (2016) Rock mass quality classification based on PCA and Fisher discrimination analysis. *Rock Soil Mech. (in Chinese)* 37
- Qiao W, Howard KWF, Li W, Zhang S, Zhang X, Niu Y (2020) Coordinated exploitation of both coal and deep groundwater resources. *Environ Earth Sci* 79
- Shao J, Zhang Q, Sun W, Wang Z, Zhu X (2020a) Numerical Simulation on Non-Darcy Flow in a Single Rock Fracture Domain Inverted by Digital Images. *Geofluids* 2020:1–13
- Shao J, Zhang Q, Wu X, Lei Y, Wu X, Wang Z (2020b) Investigation on the Water Flow Evolution in a Filled Fracture under Seepage-Induced Erosion. *Water* 12(11):3188
- Shao J, Zhang Q, Zhang W, Wang Z, Wu X (2021) Effects of the borehole drainage for roof aquifer on local stress in underground mining. *Geomech Eng* 24(5):479–490
- Shi L, Gao W, Han J, Tan X (2017) A nonlinear risk evaluation method for water inrush through the seam floor. *Mine Water Environ* 36:597–605
- Sugiyama M (2007) Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *J Mach Learn Res* 8:1027–1061
- Wang H, Lu W (2020a) Recognizing groundwater DNAPL contaminant source and aquifer parameters using parallel heuristic search strategy based on Bayesian approach. *Stoch Env Res Risk A*
- Wang Q, Dong S, Wang H, Yang J, Huang H, Dong X, Yu B (2020b) Hydrogeochemical processes and groundwater quality assessment for different aquifers in the Caojiatan coal mine of Ordos Basin, northwestern China. *Environ Earth Sci* 79
- Wang Z, Zhang Q, Shao J, Zhang W, Wu X, Zhu X (2020c) New Type of Similar Material for Simulating the Processes of Water Inrush from Roof Bed Separation. *ACS Omega* 5(47):30405–30415
- Wang Y, Yang W, Li M, Liu X (2012) Risk assessment of floor water inrush in coal mines based on secondary fuzzy comprehensive evaluation. *Int J Rock Mech Min Sci* 52:50–55
- Wei J, Zhao Z, Xie D, Yu G, Wu X (2020) Water abundance evaluation of sandstone aquifer based on lithologic and structural characteristics. *J Shandong Univ Sci Technol (Natural Sci) (in Chinese)* 39
- Wu Q, Wang M (2007) A framework for risk assessment on soil erosion by water using an integrated and systematic approach. *J Hydrol* 337:11–21

- Wu Q, Fan Z, Zhang Z, Zhou W (2014) Evaluation and zoning of groundwater hazards in Pingshuo No. 1 underground coal mine, Shanxi Province, China. *Hydrogeol J* 22:1693–1705
- Yadav B, Gupta PK, Patidar N, Himanshu SK (2020) Ensemble modelling framework for groundwater level prediction in urban areas of India. *Sci Total Environ* 712:135539
- Yang C, Liu S, Liu L (2016) Water abundance of mine floor limestone by simulation experiment. *Int J Min Sci Technol* 26:495–500
- Zhang J (2005) Investigations of water intrushes from aquifers under coal seams. *Int J Rock Mech Min Sci* 42:350–360
- Zhang H, Wan Z, Ma Z, Zhang Y (2018a) Stability control of narrow coal pillars in gob-side entry driving for the LTCC with unstable overlying strata: a case study. *Arab J Geosci* 11
- Zhang Q, Zhu H (2018b) Collaborative 3D geological modeling analysis based on multi-source data standard. *Eng Geol* 246:233–244
- Zhang Q (2020a) Hydromechanical modeling of solid deformation and fluid flow in the transversely isotropic fissured rocks. *Comput Geotech* 128:103812
- Zhang W, Zhang G, Li W, Hua X (2013) A model of Fisher's discriminant analysis for evaluating water intrush risk from coal seam floor. *J China Coal Soc (Chin)* 38
- Zhang W, Wang Z, Shao J, Zhu X, Li W, Wu X (2019) Evaluation on the stability of vertical mine shafts below thick loose strata based on the comprehensive weight method and a fuzzy matter-element analysis model. *Geofluids*
- Zhang Q, Chen Y, Yang Z, Darve E (2020b) Multi-Constitutive Neural Network for Large Deformation Poromechanics Problem. [arXiv:2010.15549](https://arxiv.org/abs/2010.15549)
- Zhang Q, Yan X, Shao J (2021) Fluid flow through anisotropic and deformable double porosity media with ultra-low matrix permeability: A continuum framework. *J Pet Sci Eng* 200:108349
- Zhang W, Wang Z, Zhu X, Li W, Gao B, Yu H (2020c) A risk assessment of a water-sand intrush during coal mining under a loose aquifer based on a factor analysis and the Fisher model. *J Hydrol Eng* 25
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.