Geomechanics and Engineering, *Vol. 31, No. 4 (2022) 353-364*
https://doi.org/10.12989/gae.2022.31.4.353

A hybrid algorithm for classifying rock joints based on improved artificial bee colony and fuzzy C-means clustering algorithm

Duofa Ji[1,2a], Weidong Lei*[3] and Wenqin Chen[3b]

[1]*Key Lab of Structures Dynamic Behavior and Control of the Ministry of Education, Harbin Institute of Technology, Harbin, 150090, China*
[2]*Key Lab of Smart Prevention and Mitigation of Civil Engineering Disasters of the Ministry of Industry and Information Technology, Harbin Institute of Technology, Harbin, 150090, China*
[3]*Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, 518055, China*

(*Received July 2, 2019, Revised October 5, 2022, Accepted November 8, 2022*)

Abstract. This study presents a hybrid algorithm for classifying the rock joints, where the improved artificial bee colony (IABC) and the fuzzy C-means (FCM) clustering algorithms are incorporated to take advantage of the artificial bee colony (ABC) algorithm by tuning the FCM clustering algorithm to obtain the more reasonable and stable result. A coefficient is proposed to reduce the amount of blind random searches and speed up convergence, thus achieving the goals of optimizing and improving the ABC algorithm. The results from the IABC algorithm are used as initial parameters in FCM to avoid falling to the local optimum in the local search, thus obtaining stable classifying results. Two validity indices are adopted to verify the rationality and practicability of the IABC–FCM algorithm in classifying the rock joints, and the optimal amount of joint sets is obtained based on the two validity indices. Two illustrative examples, i.e., the simulated rock joints data and the field-survey rock joints data, are used in the verification to check the feasibility and practicability in rock engineering for the proposed algorithm. The results show that the IABC–FCM algorithm could be applicable in classifying the rock joint sets.

Keywords: fuzzy C-means clustering algorithm; improved artificial bee colony algorithm; joint set; rock mass

1. Introduction

The Rock mass is often encountered in many engineering fields (Cui *et al*. 2019, Hudson and Harrison 2002, Mirsalimov 2021), such as civil engineering, water conservancy and hydropower engineering, as well as in mining, and nuclear-power engineering. The rock mass is consisted of intact rocks and rock discontinuities, with the latter predominating the mechanical properties of the rock mass (such as deformation, strength, and seepage characteristics). However, in engineering, the discontinuities are widely distributed in rock masses, and geometry data of their discontinuities could only be collected in outcrops, tunnels or boreholes; thus, it is impossible to measure them individually (Galindo *et al*. 2021, Zhao *et al*. 2021). Therefore, to outline the geometry and accurately express the mechanical and hydraulic properties of rock masses, it is necessary to analyze and try to clarify the distribution of discontinuities (Haghnejad *et al*. 2018, Lei 2011, Lei *et al*. 2007, Zhang and Einstein 2000, Zhang *et al*. 2016).

In 1941, for the first time, Arnord (1941) statistically studied the orientations of rock joints in terms of the distribution characteristics, which laid the solid foundation for subsequent works. In 1964, Bingham (1964) proposed the method by presenting the orientation data of the 3-D structural plane on the 2-D projection plane, which has been gradually developed into the well-known Hemispherical Projection Method and has been used to classify rock joints. Besides, rose diagrams or pole-density maps are usually used to analyze and classify rock joints as well. However, those methods are somewhat subjective by checking the density of the orientation data, and it is difficult to effectively deal with the attribution of data in groups and group boundaries. The partitioning results are often dependent on personal professional ability and engineering experience. Therefore, those methods seem not so objectively convincing. Moreover, noises caused by the random measurements or planes could not effectively be excluded. Therefore, those methods are suitable for roughly classifying rock joints.

With the rapid development of computer technology, clustering analysis based on statistical principles to describe and analyze rock joints has become one of the popular topics for classifying rock joints. In 1976, Shanley and Mahtab (1976) proposed the original clustering algorithm that described the orientation data and applied the objective function to classify the given orientations. This clustering algorithm directly described the relative-position relationship among orientations, avoiding errors caused by using different coordinate systems. After that, many efforts have been devoted to investigating different kinds of clustering analysis, such as fuzzy K-means (FKM) cluster (Hammah and Curran 2000, 1999, 1998), C-means cluster (Zhou and Maerz 2002), neuro-fuzzy inference system

*Corresponding author, Associate Professor
E-mail: leiwd@hit.edu.cn
[a]Ph.D.
[b]Ph.D. Student

ISSN: 2005-307X (Print), 2092-6219 (Online)

(Jalalifar *et al*. 2011, Xue and Zhou 2018), and spectrum cluster (Jimenez-Rodriguez and Sitar 2006, Jimenez 2008), to identify the amount of the discontinuity sets. For example, Hammah and Curran (1998) isolated and identified the number of joint sets using the FKM clustering analysis. The FKM clustering algorithm usually needs to specify in advance the initial guesses for the K cluster centroids, and different partitioning results are obtained from different initial guesses. Meanwhile, the most important features of the FKM clustering analysis method include the use of the predefined objective function, distance-measurement method among the rock joints, and iterative calculation using the predefined algorithm to obtain the minimum value of the objective function. It means that the FKM clustering algorithm needs to pre-define the relevant indicators to guarantee the optimal number of packets, and the FKM algorithm has several disadvantages. For instance, it is unable to determine the reasonable search range during the search for the optimal partitioning, which easily makes it fall into the local optimum. The drawbacks greatly restrict the use of the FKM algorithm, and thus several studies investigated discontinuities using other methods (Chen *et al*. 2018, Chiu *et al*. 2016, Gallant and Marshall 2016, Regassa *et al*. 2018, Yang *et al*. 2017).

To solve the difficulties and to improve the shortcomings of the FKM algorithm, some scholars have proposed the concept of a global optimization algorithm by introducing genetics, chaos (Xu *et al*. 2013), and other intelligent algorithms (Li *et al*. 2015) into the fuzzy-clustering analysis for classifying rock joints. However, the reasonable initial parameters for the above methods are not easily determined (Liu *et al*. 2017), which significantly affects the output results. In addition, the above-mentioned intelligent algorithms are not the most suitable algorithm for dealing with global optimization problems, as indicated by the literature (Karaboga and Akay 2009). For example, Karaboga and Akay (2009) compared the results obtained by the artificial bee colony (ABC) algorithm with those by the genetic, particle-swarm optimization (PSO), and differential-evolution algorithms, as well as different evolution strategies. The results showed that the performance of the ABC algorithm was better than that of the other four algorithms. Furthermore, the fuzzy C-means (FCM) algorithm was an efficient tool for solving fuzzy-clustering problems (Majdi and Beiki 2019, Wang *et al*. 2020), and the ABC–FCM algorithm was proven to be the evolutionary algorithm that could provide better accuracy (Doğan and Korürek 2012, Li *et al*. 2011, Menon and Ramakrishnan 2015, Ozturk *et al*. 2015a, b, Su *et al*. 2021).

Therefore, the objective of the present study is to propose a hybrid algorithm that consists of the improved ABC (IABC) and FCM clustering algorithms to classify rock joints. The coefficient, called the self-adaptive step size coefficient, is used to reduce the number of blind random searches and speed up convergence. The hybrid algorithm takes advantage of the IABC algorithm and overcomes the defects of FCM. Two case studies are conducted on two sets of rock joint data, namely, the simulated rock joint data and the field-survey rock joint data, to verify the proposed algorithm.

2. FCM clustering algorithm

For the given data set $\boldsymbol{X}=\{X_1,X_2,\cdots,X_n\}\in\boldsymbol{R}^n$, the expression can be considered as hard c-partitions ($2\le c\le n$), when c natural subsets can be identified in $\boldsymbol{X}$ as follows(Cannon *et al*. 1986, Chuang *et al*. 2006, Pal and Bezdek 1995)

$$\begin{cases} X_1\bigcup X_2\bigcup\cdots\bigcup X_c=\boldsymbol{X} \\ X_i\bigcap X_k=\varnothing, 1\le i\ne k\le c \\ X_i\ne\varnothing, X_i\ne\boldsymbol{X}, 1\le i\le c \end{cases} \tag{1}$$

where c is the number of clusters, and $\boldsymbol{R}^n$ represents the set of n tuples of real numbers.

The fuzzy partition of $\boldsymbol{X}$ can be arrayed as ($c\times n$) matrix $\boldsymbol{U}$ as follows

$$\boldsymbol{U}=\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{c1} & u_{c2} & \cdots & u_{cn} \end{bmatrix} \tag{2}$$

where u_{ij} is the degree of membership of i in the j_{th} cluster, and it needs to satisfy the following criteria

$$\begin{cases} 0\le u_{ij}\le 1, \quad i=1,2,\cdots,c, \quad j=1,2,\cdots,n \\ \sum_{i=1}^{c}u_{ij}=1, \quad j=1,2,\cdots,n \\ 0<\sum_{j=1}^{n}u_{ij}<n, \quad i=1,2,\cdots,c \end{cases} \tag{3}$$

The set of partition matrix can be expressed as follows

$$M_{fc}=\left\{\boldsymbol{U}\in\boldsymbol{R}^{cn}\mid u_{ik}\in[0,1],\forall i,k;\ \sum_{i=1}^{c}u_{ik}=1,\forall k;0<\sum_{k=1}^{n}u_{ik}<n,\forall i\right\} \tag{4}$$

The FCM objective function can be expressed as follows

$$J(\boldsymbol{U},\boldsymbol{v})=\sum_{j=1}^{n}\sum_{i=1}^{c}u_{ij}^{m}\left\|x_j-v_i\right\|^2 \tag{5}$$

where $\boldsymbol{v}=\{v_1,v_2,\cdots,v_c\}$ is the vector cluster center of the fuzzy partition of matrix $\boldsymbol{U}$, and m is the fuzzy weighting index.

In the FCM clustering algorithm, as the number of iterations increases, the formulations of u_{ij} and y_i need to be updated, which can be performed by calculating the minimum value of objective function $\boldsymbol{J}(\boldsymbol{U},\boldsymbol{v})$ as follows

$$\begin{aligned}\min\left\{J(\boldsymbol{U},\boldsymbol{v})\right\}&=\min\left\{\sum_{j=1}^{n}\sum_{i=1}^{c}u_{ij}^{m}\left\|x_j-v_i\right\|^2\right\} \\ &=\sum_{j=1}^{n}\min\left\{\sum_{i=1}^{c}u_{ij}^{m}\left\|x_j-v_i\right\|^2\right\}\end{aligned} \tag{6}$$

To solve Eq. (6), an additional equation ($\boldsymbol{F}$ function) is

given as follows

$$F=\sum_{i=1}^{c}u_{ij}^{m}\left\|x_j-v_i\right\|^2+\lambda\left(\sum_{i=1}^{c}u_{ij}-1\right) \tag{7}$$

In addition, a necessary condition needs to be introduced to obtain the minimum value of the objective function [see Eq. (5)] as follows

$$\frac{\partial}{\partial y_i}J(\boldsymbol{U},\boldsymbol{v})=0 \tag{8}$$

The partial derivatives of the $\boldsymbol{F}$ function with respect to u_{ij} and λ can be calculated as

$$\frac{\partial F}{\partial u_{ij}}=mu_{ij}^{m-1}\left\|x_j-v_i\right\|^2-\lambda=0 \tag{9}$$

$$\frac{\partial F}{\partial \lambda}=\sum_{i=1}^{c}u_{ij}-1=0 \tag{10}$$

Eq. (9) can be reformulated as

$$u_{ij}=\left[\frac{\lambda}{m\left\|x_j-v_i\right\|^2}\right]^{\frac{1}{m-1}} \tag{11}$$

Introducing Eq. (11) into Eq. (10) yields the following formula

$$\begin{aligned}\sum_{l=1}^{c}u_{lt}&=\sum_{l=1}^{c}\left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}\left[\frac{1}{\left\|x_t-v_l\right\|^2}\right]^{\frac{1}{m-1}}\\&=\left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}\sum_{l=1}^{c}\left[\frac{1}{\left\|x_t-v_l\right\|^2}\right]^{\frac{1}{m-1}}=1\end{aligned} \tag{12}$$

where

$$\left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}=\frac{1}{\sum_{l=1}^{c}\left[\frac{1}{\left\|x_t-v_l\right\|^2}\right]^{\frac{1}{m-1}}} \tag{13}$$

Introducing Eq. (13) into Eq. (11) yields the following formula

$$u_{ij}=\frac{1}{\sum_{l=1}^{c}\left[\frac{\left\|x_j-v_i\right\|}{\left\|x_j-v_l\right\|}\right]^{\frac{2}{m-1}}} \tag{14}$$

The values of v_{i} can also be obtained using the same procedures mentioned above, as follows

$$v_i=\frac{\sum_{j=1}^{n}u_{ij}^{m}x_j}{\sum_{j=1}^{n}u_{ij}^{m}} \tag{15}$$

After setting the initial parameters of the FCM algorithm, $\boldsymbol{U}$ and $\boldsymbol{v}$ can be updated according to Eqs. (14) and (15), and the optimal cluster results can be obtained when min $\boldsymbol{J}(\boldsymbol{U}, \boldsymbol{v})$ is reached.

To find the optimal number of the joint sets, two fuzzy cluster validity indices are adopted in the study, namely, Xie–Beni (1991) and Fukuyama–Sugeno (1989) validity indices. The Xie–Beni validity index can be expressed as

$$V_{\mathrm{XB}}=\frac{\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{2}\left\|x_j-v_i\right\|^2}{n\left(\min_{i\neq k}\left\{\left\|v_i-v_k\right\|^2\right\}\right)} \tag{16}$$

The Fukuyama–Sugeno validity index can be expressed as

$$V_{\mathrm{FS}}=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}\left(\left\|x_j-v_i\right\|_{\mathrm{A}}^2-\left\|v_i-\overline{v}\right\|_{\mathrm{A}}^2\right) \tag{17}$$

where $\overline{v}$ is the mean vector of $\boldsymbol{X}$ and A is a 1×1 positive definite symmetric matrix.

3. IABC algorithm

The partitioning results of the clusters can be rapidly obtained using the FCM algorithm. However, the algorithm can be regarded as a local searching optimization intrinsically, in which hill-climbing algorithm is adopted to gain the optimal solution in the optimizing process. Therefore, it often falls into the local optimum during the local search, and might not obtain relatively reasonable and effective clustering results, when the initial fuzzy partition matrix or the number of classification partitioning is unreasonable. Furthermore, the results are sensitive to the amount of operational processing when the FCM algorithm is used to deal with the rock joint plane data. Therefore, there is no guarantee that the final result is the optimal result.

Tereshko and his colleagues (Tereshko 2000, Tereshko and Lee 2002, Tereshko and Loengarov 2005) presented the prediction model that can be used to simulate the foraging behavior of a honeybee colony. In the model, they defined three essential components, namely, food sources, employed foragers, and unemployed foragers. The workflow of the foraging behavior of the honeybee colony (as shown in Fig. 1) can be divided into the following steps:

(1) Initially, potential foragers are unemployed foragers.

(2) The unemployed foragers scout and search around the nest for food sources (see S in Fig. 1).

(3) The foragers are recruited by watching the waggling dances of the honeybees, which are gathering the nectar, and after that, start to search for the food source according to the information from the dancing honeybees (see R in Fig. 1).

(4) The unemployed foragers become employed foragers. When they find the food source, they start to gather the nectar.

The gathering honeybees go back to the nest to unload the nectar. Then, they have three options as follows:

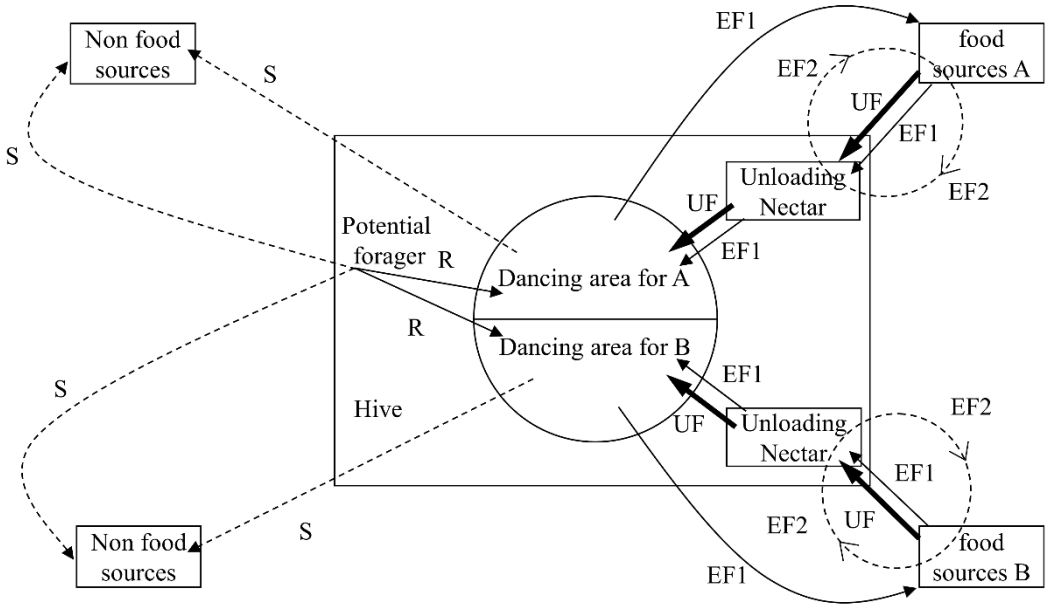


Fig. 1 Flowchart of honeybee foraging for nectar

(i) They abandon the food source and become recruits (see UF in Fig. 1).

(ii) They share the information of the food source through waggling dances in the dancing area to attract unemployed foragers and then return to the food source to gather the nectar (see EF1 in Fig. 1).

(iii) They directly go back to the food source to gather the nectar (see EF2 in Fig. 1).

Three procedures for each searching process can be seen in Fig. 1, namely, (i) the employed bees reach the food source and then assess their fitness, (ii) the employed bees share the information about the food source in the dancing area, and the onlooker bees determine and select the food source, or (iii) the latter become scout bees and search for the new food source. The above three bee classifications can interchange under given conditions, as shown in Fig. 2. For example, the employed bees become scout bees when they abandon the food source, whereas the opposite condition is established when the scout bees find the new source.

In the IABC algorithm, the number of employed or onlooker bees is equal to the number of food sources that represent the number of solutions in the population. The position of the food source corresponds to the potential solution of the optimization problem, and the fitness of the food source corresponds to the quality of the possible solution to the problem. The fitness of the food source, i.e., *fit*, is represented by its benefit from the nectar, as expressed in the following equation

$$fit = \frac{1}{J(\boldsymbol{U},\boldsymbol{v})+1} \tag{18}$$

Initially, the IABC algorithm specifies the problem with SN initial solutions x_i (i = 1,2,…,SN) in the D-dimensional vector. Here, D is the number of optimization parameters. Furthermore, each food source attaches one employed bee to gather the nectar; thus, the location of the employed bee represents the location of the food source. The number of food sources remains constant in the whole process. The employed bees share the information about the food source through waggling dances in the dancing area, and the onlooker bees decide whether they would gather the nectar or not, according to the fitness of the food source. The observers predetermine the limited area based on their memory after they arrive at the food source and then carry out a comprehensive search for the potential food source.

When they find a new food source, they abandon the older one if the new one is more suitable. The probability that the food source is selected by an onlooker bee can be expressed as

$$P_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \tag{19}$$

To create the candidate food position from the old one in the memory, IABC uses the following expression

$$y_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \tag{20}$$

where y_{ij} (i = 1,2,…,SN and j = 1,2,…,D), x_{ij}, x_{kj} and ϕ_{ij} respectively represent the location of the calculated food source, the location of the current food source, the location of the new food source randomly selected in the neighbourhood of the current food source, and the random numbers in the interval [−1,1].

The above evolutionary process continuously cycles, until the predefined conditions are satisfied. In the ABC algorithm, the employed and onlooker bees find the optimal solution for their position based on the food-source updating formula (see Eq. (15)). However, because the employed and onlooker bees are randomly distributed in a certain area, their search has no clear range and direction of

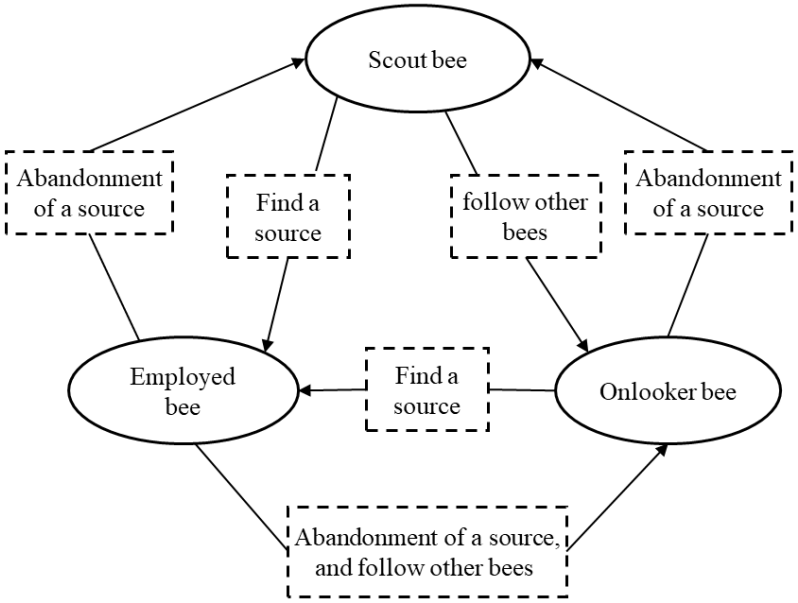


Fig. 2 Interchanging process in the three bee identities

motion. Therefore, the search is of a certain degree of blindness, which slows down the convergence speed of the algorithm, and leads to a long time to find the best food-source location. During each iteration of the ABC algorithm, the difference, $\phi_{ij}(x_{ij}- x_{kj})$, exists between the updated and current food sources. Because ϕ_{ij} and the choice of location of the food source (x_{ij}) are somewhat random, the difference during the iterations is also random.

To optimize the algorithm and improve the computation efficiency, the difference would gradually decrease as the number of iterations increases. Thus, this paper introduces the coefficient, called the self-adaptive step-size coefficient, which is related to the number of iterations. The coefficient is used to optimize and improve the ABC algorithm, thereby reducing the number of blind random searches and speeding up the convergence. The mathematical expression for the self-adaptive step-size coefficient used in this study is expressed as follows

$$step_i = \frac{\eta(N-t)}{N} \tag{21}$$

where N, t and η are respectively the total number of iterations, the current iteration ($t \le N$), and the random number generated in the process, $\eta \in [1.1,1.5]$.

The self-adaptive step-size coefficient is closely related to the number of searches (i.e., the number of iterations in the algorithm), and its value tends to decrease as the number of iterations increases. Therefore, the variation in the food-source position would gradually shrink during the search process of the algorithm until the convergence is achieved. The use of the coefficient to optimize the position does not change the basic principle of the food-source location search. In general, the step size tends to decrease, which not only accelerates the convergence of the algorithm but also ensures the accuracy and stability of the results.

The new food-position equation can be reformulated as

$$y_{ij} = x_{ij} + \phi_{ij} step_i (x_{ij} - x_{kj}) \tag{22}$$

Eq. (22) is called the self-adaptive step based food source updating formula, and the ABC algorithm is improved by Eq. (22) to form the IABC algorithm. Similar to the ABC algorithm, the IABC algorithm has the same three control parameters: number of food sources SN, limit time of food source mining, and the maximum number of cycles (MCN). The IABC algorithm is the self-organizing process, that uses basic elements to achieve the specific behavior, i.e., the overall behavior of the bee colony can be highlighted by the individual behavior of the single bee.

In the IABC algorithm, the quality of solutions remains constant when the employed bee continues to mine the food source, and the number of mining is larger than its predefined limitation. In this manner, the employed bee is transformed into the scout bee, and the food source is replaced by the new food source, which is randomly generated, as expressed in the following equation

$$x_{ij} = x_{\min}^{j} + rand(0,1) \cdot (x_{\max}^{j} - x_{\min}^{j}) \tag{23}$$

where rand(0,1) represents the random number generated between the range [0,1], with $x_{\max}$ and $x_{\min}$ representing the maximum and minimum values of the position variables of the food source, respectively, and j representing the randomly selected subscript, where $j \in \{1,2,\ldots,D\}$ and D =3.

The IABC algorithm is applied to classify the rock joints by combining the FCM clustering algorithm. The results from the IABC algorithm are treated as the initial parameters, which are input into the FCM clustering algorithm for further processing. The basic ideas of the method are as follows: first, the IABC algorithm is used to process the orientations of the rock joints, and several initial food sources (initial cluster centers) are randomly generated, based on which the IABC algorithm is

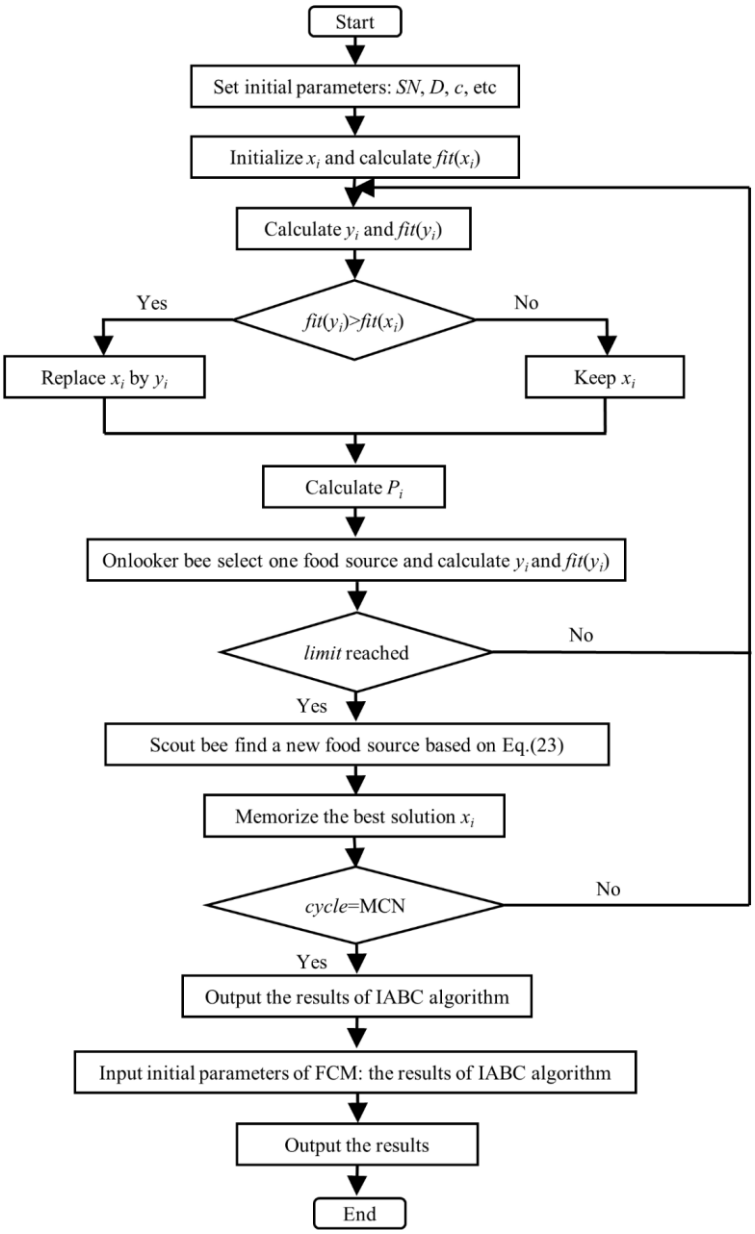


Fig. 3 Flowchart of the IABC–FCM algorithm

subsequently iteratively updated to find new locations of food sources. After that, the cluster center point is continuously updated until the algorithm converges to achieve the objective, which is considered the initial parameter of the FCM clustering algorithm for cluster analysis. Finally, the final partitioning result is obtained by implementing the FCM clustering analysis.

The calculation procedure of the hybrid algorithm can be described as follows (see Fig. 3):

1: Setting the initial parameters of the IABC algorithm: the total number of bees SN, the parameter dimension *D*, the limit times of the food source mining, number of clusters c, and other related parameters

2: Initializing solutions x_i (i =1,2,…,SN), and calculating the fitness function value of each food source

3: Cycling: cycle = 1

4: Repeating

5: Finding new food sources by the employed bee using Eq. (22), and calculating the fitness value of new food sources

6: Applying the greedy selection process for the employed bee

7: Calculating Pi by using Eq. (19)

8: Selecting one food source by the onlooker bees, and calculating y_i and $fit(y_i)$

9: Applying the greedy selection process for the onlooker bees

10: Determining the abandoned solution of the scout (if it exists), and replacing it with a new randomly generated solution x_i using Eq. (23)

11: Memorizing the best solution achieved so far

12: Cycling: cycle = cycle + 1

13: **until** cycle = MCN

14: Outputting the results of the IABC algorithm

Table 1 Parameters of the rock-mass joint surface parameters generated by artificial simulation

Joint sets	Dip direction (°)			Dip angle (°)			Number of rock joints
	Mean	Variance	Range	Mean	Variance	Range	
1	50	120	18–89	45	40	31–61	60
2	80	50	65–101	70	40	55–86	40
3	160	80	133–175	20	20	9–30	50
4	200	100	175–222	25	24	16–37	50
5	275	200	244–301	55	40	42–66	40
6	315	40	300–333	50	30	38–61	60

Table 2 Cluster analysis results of the joint surfaces generated by the simulation

Number of clusters	Dip direction (°)	Dip angle (°)	Number of rock joints
2	285.15	33.86	169
	66.96	45.02	131
3	301.74	50.33	100
	65.56	54.11	100
	182.05	20.67	100
4	79.24	69.22	42
	51.04	43.64	58
	182.15	20.79	100
	302.77	50.33	100
5	65.39	54.04	100
	157.74	19.25	52
	200.43	24.84	48
	314.13	50.51	66
	271.23	54.51	34
6	50.76	43.59	57
	79.04	69.42	43
	159.14	19.47	52
	200.83	25.00	48
	271.80	54.79	36
	314.30	50.60	64
7	59.92	45.87	36
	159.21	19.47	52
	200.84	25.01	48
	314.30	50.61	64
	271.85	54.79	36
	39.78	42.20	24
	79.38	70.17	40
8	39.70	42.28	24
	159.20	19.47	52
	200.78	24.99	48
	315.09	50.62	62
	281.17	54.80	25
	59.88	45.79	36
	79.36	70.13	40
	260.48	55.05	13
9	59.93	45.85	36
	159.20	19.49	52
	200.77	24.99	48
	316.92	50.73	54
	229.73	50.41	13
	39.84	42.21	54
	79.39	70.12	13
	276.83	55.18	24
10	257.13	55.42	9
	39.88	42.16	25
	159.58	22.51	36
	59.85	45.94	35
	316.87	50.74	56
	299.47	50.39	11
	161.91	13.44	17
	79.36	70.13	40
	276.86	55.92	24
	257.11	55.43	9
	201.42	25.26	47

15: Using the results of the IABC algorithm as initial parameters of the FCM clustering analysis

16: Outputting the results of the FCM clustering analysis

4. Case studies

4.1 Case study on artificially stimulated rock joints data

This sub-section presents the case study on the artificially simulated orientation data of rock joints to validate the proposed algorithm. Several probability models can be used to simulate the rock joint orientations, such as the normal, the lognormal–normal, the exponential, as well as the Fisher, and the Bingham distributions. The normal distribution model is chosen to produce the artificial rock joint data in this case study, because most of the real rock joint orientations statistically obey the normal distribution (Cai 2011, Pan *et al*. 2019). In total, 300 orientation data for six joint sets are generated by artificial computer simulation, as listed in Table 1.

It is noted that the artificial rock joint data could be artificially manipulated by the adopted probability models and statistical distribution centers, and by the added noises as well. Therefore, the comparison between the classification results by the proposed algorithm and the artificial rock joint data is of clear standard. It means that the artificially manipulated data might be more reasonable to verify the proposed algorithm than the comparison, where the practical rock joint data are used.

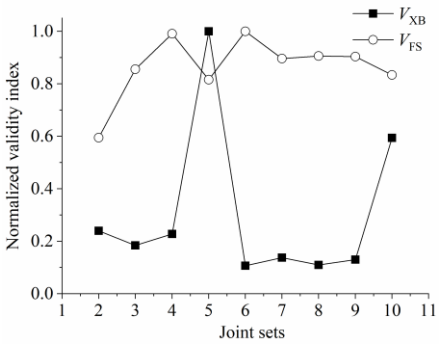
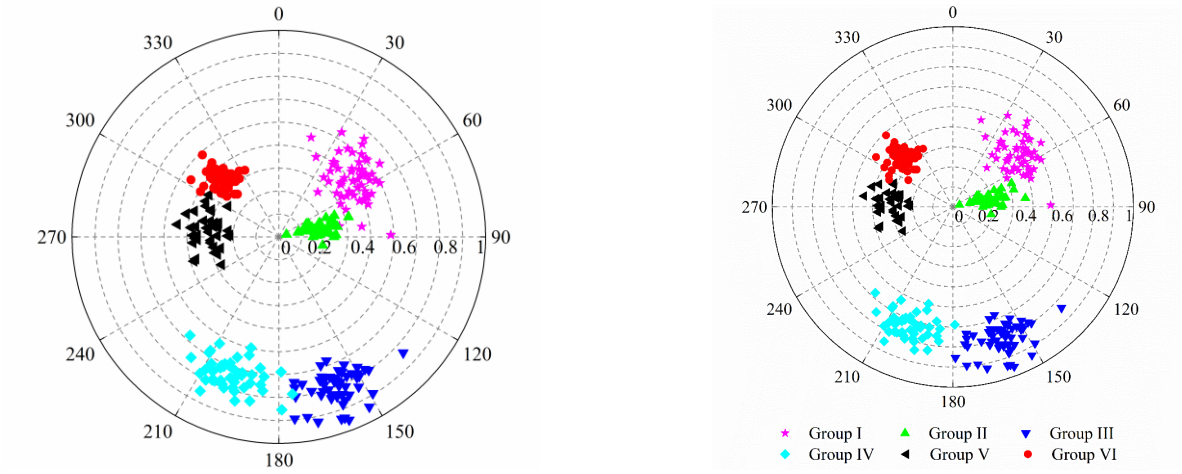


Fig. 4 Two normalized validity indices in each joint set



(a) Pole projection of the joint surface in the artificial data (b) Pole projection of the joint surface calculated by the IABC–FCM algorithm

Fig. 5 Comparisons between the artificial data and the calculated results by the IABC–FCM algorithm in the pole projection

The initial parameters are set as: SN = 15, D = 3, $c\in$[2–10], m = 2, limit = 45, and MCN = 50. Table 2 lists the results of the rock joint planes calculated by the IABC–FCM algorithm. The two validity indices are calculated by using Eqs. (16) and (17), and listed in Table 3. To facilitate the comparative analysis, V_{XB} and V_{FS} were normalized in the range between 0 and 1 by using their maximum values, as shown in Fig. 4. Table 3 and Fig. 4 show that the minimum values of the validity indices correspond to six, meaning that the optimal partitioning of the rock joints is six. The result is consistent with the simulation results, as listed in Table 1. Table 4 lists the comparison of the results obtained by the FCM and IABC–FCM algorithms with the original results for the average values of the whole sample.

Compared with the FCM algorithm, the IABC–FCM algorithm shows better calculation results. Besides, the errors computed by the IABC-FCM algorithm are less than 4% for the dip direction and dip angle and less than 10% for the number of rock joints. Figs. 5(a) and 5(b) show the pole projection view of the artificial rock joint data and those from the IABC–FCM algorithm, respectively.

From Fig. 5, it can be seen that the dip direction and dip angle calculated by the IABC–FCM algorithm are consistent with those of the artificial data, meaning that the IABC–FCM algorithm exhibits good prediction accuracy in every single sample. However, it should be noted that some artificial data may be mis-judged by another group in the IABC-FCM when the dip direction and dip angle of those data are close to those of the other group.

4.2 Case study on rock joints data from the rock slope along Beijing–Zhuhai Expressway

To further verify the feasibility and practicability of the algorithm in practical engineering, a case study is carried out on 337 orientation data collected from the rock slope along Beijing–Zhuhai Expressway. More details of the site could be found in the literature (Jia *et al*. 2008).

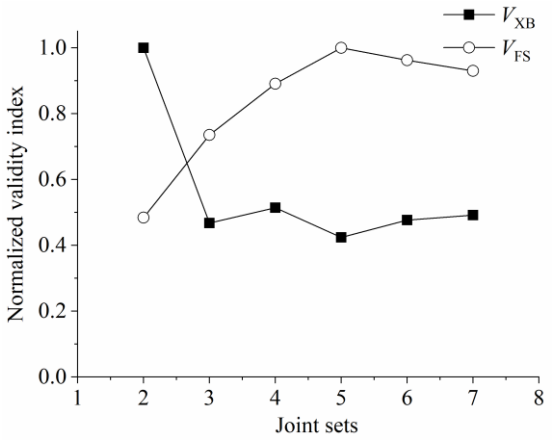


Fig. 6 Two normalized validity indices in each joint set

The initial parameters are set as: SN = 15, D = 3, $c\in$[2–7], m = 2, limit = 45, and MCN = 50. Table 5 lists the result of the rock joints calculated by the IABC–FCM algorithm. Two validity indices are calculated by using Eqs. (16) and (17), and listed in Table 6. To facilitate comparative analysis, V_{XB} and V_{FS} are converted in the range [0 1], as shown in Fig. 6. Table 6 and Fig. 6 show that the minimum values of the validity indices correspond to five, which means that the optimal partitioning of the rock joints is five. The result is consistent with the manual judgment based on the rose diagrams (Jia *et al*. 2008) and the judgment based on the PSO–FCM algorithm in the literature (Song *et al*. 2012). Table 7 compares the results obtained by the IABC–FCM algorithm with those calculated by the FKM (Jia *et al*. 2008) and PSO–FCM algorithms (Song *et al*. 2012) for the average values of the whole sample. Table 7 indicates that the results calculated by the IABC–FCM, and PSO–FCM algorithms show a consistent trend, whereas the results from the IABC–FCM algorithm are slightly different from those calculated by FKM, which might be caused by the improper setting of the initial parameters. The comparison results of rock joints are shown in Fig. 7, where Figs. 7(a)

Table 3 Two validity indices in each joint set

Number of clusters	2	3	4	5	6	7	8	9	10
V_{XB}	0.2894	0.2229	0.2755	1.2088	0.1290	0.1666	0.1324	0.1572	0.7179
V_{FS}	−47.3112	−68.0351	−78.8046	−64.8955	−79.5657	−71.2667	−72.0856	−71.8796	−66.3495

Table 4 Comparisons of the artificial data and calculated results from the IABC–FCM algorithm

Joint sets	Category	Dip direction (°)	Dip angle (°)	Number of rock joints
1	Artificial	50.00	45.00	60
	FCM	52.20	42.48	57
	IABC–FCM	50.76	43.59	57
2	Artificial	80.00	70.00	40
	FCM	79.01	68.70	43
	IABC–FCM	79.04	69.42	43
3	Artificial	160.00	20.00	50
	FCM	153.63	20.94	52
	IABC–FCM	159.14	19.47	52
4	Artificial	200.00	25.00	50
	FCM	204.45	25.59	48
	IABC–FCM	200.83	25.00	48
5	Artificial	275.00	55.00	40
	FCM	271.26	55.91	36
	IABC–FCM	271.80	54.79	36
6	Artificial	315.00	50.00	60
	FCM	313.54	50.99	64
	IABC–FCM	314.30	50.60	64

Table 5 Cluster analysis results of the joint surfaces generated by the simulation

Number of clusters	Dip direction (°)	Dip angle (°)	Number of rock joints
2	310.69	40.03	99
	180.81	42.08	238
3	303.97	50.64	73
	197.64	45.16	199
	99.42	55.22	65
4	300.18	55.28	59
	218.83	49.41	127
	146.37	52.74	109
	38.31	61.18	42
5	24.87	67.85	32
	159.57	51.31	92
	221.90	51.16	104
	95.69	69.32	56
	298.92	60.10	53
6	338.47	45.70	19
	221.82	51.29	117
	158.00	51.35	96
	37.16	66.89	20
	295.44	58.82	49
	96.62	61.37	36
7	340.30	32.76	2
	221.75	51.63	111
	159.27	51.44	101
	96.83	60.17	1
	298.34	57.90	60
	26.68	65.66	31
	96.83	60.17	31

and (b) show the pole projection view of the results calculated by the FKM and IABC–FCM algorithms, respectively. Fig. 7 shows that the dip direction and dip angle calculated by the IABC–FCM algorithm are generally consistent with those calculated by FKM.

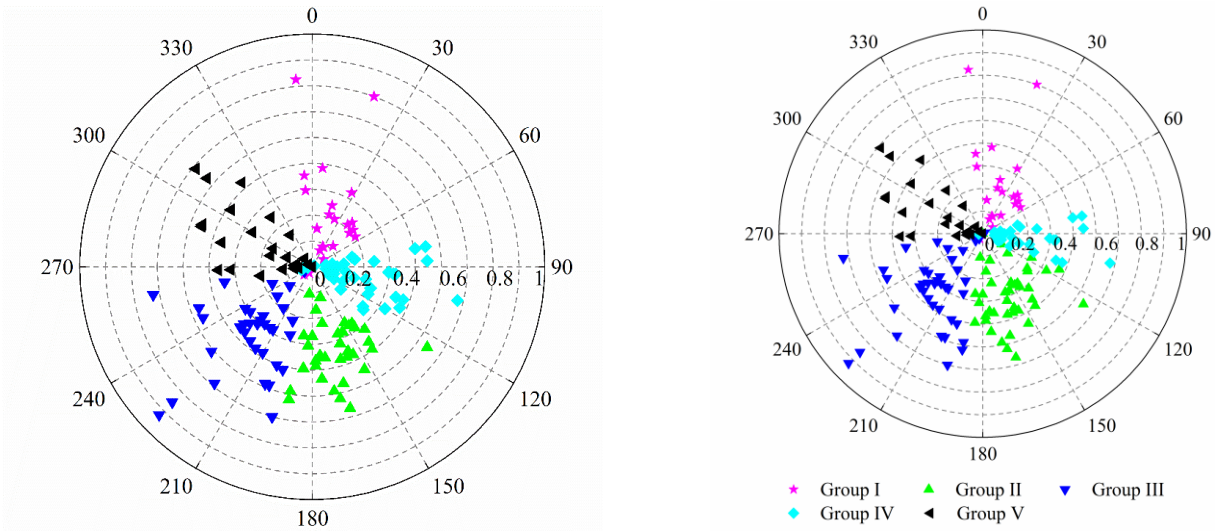
5. Conclusions

To classify the rock joints, this study presents the IABC–FCM algorithm, in which the self-adaptive step size coefficient is proposed to improve and optimize the ABC algorithm. The IABC algorithm reduces the randomness in the search process and accelerates the convergence speed. The main conclusions are summarized as follows:

1. The self-adaptive step size coefficient is related to the number of iterations and can reduce the number of blind random searches, thereby optimizing and improving the ABC algorithm.
2. The FCM clustering algorithm is sensitive to initial partitioning and often falls into the local optimal loop. Meanwhile, using the results of the IABC algorithm as initial parameters of the FCM clustering algorithm can avoid subjective bias and improve the local optimization of the FCM clustering algorithm. The IABC–FCM algorithm has better global optimization and yields stable classifying results.
3. The two validity indices, namely, V_{XB} and V_{FS}, show the proper evaluation of the partitioning results and could be used to quickly determine the optimal partitioning result, which avoids the drawbacks of specifying the number of packets in advance.
4. From the case studies on the artificial rock joint data and the field measurements, the feasibility and practicability of the proposed method are verified. By comparing and analyzing the calculation results, the proposed method obtains accurate partitioning results. The proposed method is suitable for analyzing and processing a large number of rock joint data and offers good practical engineering applications.

Table 6 Two validity indices in each joint set

Number of clusters	2	3	4	5	6	7
V_{XB}	0.2108	0.0985	0.1083	0.0893	0.1005	0.1036
V_{FS}	−50.7145	−76.9743	−93.3283	−104.7499	−100.7879	−97.4212



(a) Pole projection of the joint surface of the results calculated by the FKM algorithm (b) Pole projection of the joint surface of the results calculated by the IABC–FCM algorithm

Fig. 7 Comparisons of the calculated results by the FKM and IABC–FCM algorithms using the pole projection.

Table 7 Comparisons of the results of the joint sets and rock joints obtained by the different algorithms

Joint sets	Category	Dip direction (°)	Dip angle (°)	Number of rock joints
1	FKM	25.00	67.00	36
	PSO–FCM	24.50	68.20	34
	IABC–FCM	24.87	67.85	32
2	FKM	163.00	49.00	90
	PSO–FCM	161.20	51.00	93
	IABC–FCM	159.57	51.31	92
3	FKM	220.00	49.00	101
	PSO–FCM	221.40	51.20	101
	IABC–FCM	221.90	51.16	104
4	FKM	103.00	66.00	53
	PSO–FCM	97.00	72.10	60
	IABC–FCM	95.69	69.32	56
5	FKM	295.00	71.00	57
	PSO–FCM	298.90	58.10	49
	IABC–FCM	298.92	60.10	53

Acknowledgments

The authors would like to acknowledge the financial support from research grants No.s 2019YFC1511105, 2021YFC3001002 by National Key R&D Program of China grant No. JCYJ20210324121402008 by Shenzhen Science and Technology Innovation Commission, grant No.s 52008142 and 51778193 by the National Natural Science Foundation of China, grant No. LH2020E057 by the Heilongjiang Provincial Natural Science Foundation of China.

References

Arnold, K.J. (1941), On spherical probability distributions, Massachusetts Institute of Technology, USA.

Bingham, C. (1964), Distributions on the sphere and on the projective plane, PhD dissertation , Yale University.

Cai, M. (2011), "Rock mass characterization and rock property variability considerations for tunnel and cavern design", *Rock Mech. Rock Eng.*, **44**(4), 379-399.

Cannon, R.L., Dave, J.V. and Bezdek, J.C. (1986), "Efficient Implementation of the Fuzzy c-Means Clustering Algorithms", *IEEE T. Pattern Anal. Mach. Intell.*, **PAMI-8**(2), 248-255. https://doi.org/10.1109/tpami.1986.4767778.

Chen, S., Walske, M.L. and Davies, I.J. (2018), "Rapid mapping and analysing rock mass discontinuities with 3D terrestrial laser scanning in the underground excavation", *Int. J. Rock Mech. Min. Sci.*, **110**, 28-35. https://doi.org/10.1016/j.ijrmms.2018.07.012.

Chiu, C.C., Weng, M.C. and Huang, T.H. (2016), "Modeling rock joint behavior using a rough-joint model", *Int. J. Rock Mech. Min. Sci.*, **89**, 14-25. https://doi.org/10.1016/j.ijrmms.2016.08.001.

Chuang, K.S., Tzeng, H.L., Chen, S., Wu, J. and Chen, T.J. (2006), "Fuzzy c-means clustering with spatial information for image segmentation", *Comput. Med. Imaging Graph.*, **30**(1), 9-15. https://doi.org/10.1016/j.compmedimag.2005.10.001.

Cui, L., Zheng, J., Sheng, Q. and Pan, Y. (2019), "A simplified procedure for the interaction between fully-grouted bolts and rock mass for circular tunnels", *Comput. Geotech.*, **106**, 177-192. https://doi.org/10.1016/j.compgeo.2018.10.008.

Doğan, B. and Korürek M. (2012), "A new ECG beat clustering method based on kernelized fuzzy c-means and hybrid ant colony optimization for continuous domains", *Appl. Soft Comput.*, **12**(11), 3442–3451. https://doi.org/10.1016/j.asoc.2012.07.007.

Fukuyama Y. and Sugeno M. (1989), "A new method of choosing the number of clusters for the fuzzy c-mean method", in *proceedings of fifth fuzzy system symposium*, 247-250.

Galindo, R., Andres, J.L., Lara, A., Xu, B., Cao, Z. and Cai, Y. (2021), "Theoretical model for the shear strength of rock discontinuities with non-associated flow laws", *Geomech. Eng.*, **24**(4), 307-321. https://doi.org/10.12989/gae.2021.24.4.307.

Gallant, M.J. and Marshall, J.A. (2016), "Automated rapid mapping of joint orientations with mobile LiDAR", *Int. J. Rock Mech. Min. Sci.*, **90**, 1-14. https://doi.org/10.1016/j.ijrmms.2016.09.014.

Haghnejad, A., Ahangari, K., Moarefvand, P. and Goshtasbi, K. (2018), "Numerical investigation of the impact of geological

discontinuities on the propagation of ground vibrations", *Geomech. Eng.*, **14**(6), 545-552. http://dx.doi.org/10.12989/gae.2018.14.6.545.

Hammah, R.E. and Curran J.H. (1998), "Fuzzy cluster algorithm for the automatic identification of joint sets", *Int. J. Rock Mech. Min. Sci.*, **35**(7), 889-905. https://doi.org/10.1016/s0148-9062(98)00011-4.

Hammah, R.E. and Curran J.H. (1999), "On distance measures for the fuzzy K-means algorithm for joint data", *Rock Mech. Rock Eng.*, **32** (1), 1–27. https://doi.org/10.1007/s006030050041.

Hammah, R.E. and Curran J.H. (2000), "Validity measures for the fuzzy cluster analysis of orientations", *IEEE T. Pattern Anal. Mach. Intell.*, **22**(12), 1467–1472. https://doi.org/10.1109/34.895981.

Hudson, J.A. and Harrison, J.P. (2002), Engineering Rock Mechanics: An Introduction to the Principles, Pergamon, Elsevier Science, Oxford.

Jalalifar, H., Mojedifar, S., Sahebi, A.A. and Nezamabadi-pour, H. (2011), "Application of the adaptive neuro-fuzzy inference system for prediction of a rock engineering classification system", *Comput. Geotech.*, **38**(6), 783-7905. https://doi.org/10.1016/j.compgeo.2011.04.005.

Jia H.B., Tang, H., Liu, Y. and Ma, S. (2008), Theory and engineering application of 3-d network modeling of discontinuities in rockmass, Science Press, Beijing.

Jimenez-Rodriguez, R. and Sitar, N. (2006), "A spectral method for clustering of rock discontinuity sets", *Int. J. Rock Mech. Min. Sci.*, **43**(7), 1052-1061. https://doi.org/10.1016/j.ijrmms.2006.02.003.

Jimenez, R. (2008), "Fuzzy spectral clustering for identification of rock discontinuity sets", *Rock Mech. Rock Eng.*, **41**(6), 929-939. https://doi.org/10.1007/s00603-007-0155-6.

Karaboga, D. and Akay, B. (2009), "A comparative study of Artificial Bee Colony algorithm", *Appl. Math. Comput.*, **214**(1), 108-132. https://doi.org/10.1016/j.amc.2009.03.090.

Lei, W.D. (2011), "A new method for processing wave amplitude for numerical analysis in rock dynamics", *Int. J. Rock Mech. Min. Sci.*, **48**(4), 674-680. https://doi.org/10.1016/j.ijrmms.2011.01.006.

Lei, W.D., Hefny, A.M., Yan, S. and Teng, J. (2007), "A numerical study on 2-D compressive wave propagation in rock masses with a set of joints along the radial direction normal to the joints", *Comput. Geotech.*, **34**(6), 508-523. https://doi.org/10.1016/j.compgeo.2007.01.002.

Li, H., Li, J. and Kang, F. (2011), "Risk analysis of dam based on artificial bee colony algorithm with fuzzy c-means clustering", *Can. J. Civ. Eng.*, **38**(5), 483-492. https://doi.org/10.1139/l11-020.

Li, Y., Wang, Q., Chen, J., Xu, L. and Song, S. (2015), "K-means algorithm based on particle swarm optimization for the identification of rock discontinuity sets", *Rock Mech. Rock Eng.*, **48**(1), 375–385. https://doi.org/10.1007/s00603-014-0569-x.

Liu, J., Zhao, X.D. and Xu Z. (2017), "Identification of rock discontinuity sets based on a modified affinity propagation algorithm", *Int. J. Rock Mech. Min. Sci.*, **94**, 32–42. https://doi.org/10.1016/j.ijrmms.2017.02.012.

Majdi, A. and Beiki, M. (2019), "Applying evolutionary optimization algorithms for improving fuzzy C-mean clustering performance to predict the deformation modulus of rock mass", *Int. J. Rock Mech. Min. Sci.*, **113**, 172-182. https://doi.org/10.1016/j.ijrmms.2018.10.030.

Menon, N. and Ramakrishnan, R. (2015), "Brain tumor segmentation in MRI images using unsupervised artificial bee colony algorithm and FCM clustering", *Proceedings of the 2015 International Conference on Communications and Signal Processing (ICCSP)*, IEEE, 0006–0009. https://doi.org/10.1109/iccsp.2015.7322635.

Mirsalimov, V.M. (2021), "Optimal design of shape of a working in cracked rock mass", *Geomech. Eng.*, **24**(3), 227-235. https://doi.org/10.12989/gae.2021.24.3.227.

Ozturk, C., Hancer, E. and Karaboga, D. (2015a), "Dynamic clustering with improved binary artificial bee colony algorithm", *Appl. Soft Comput.*, **28**, 69-80. https://doi.org/10.1016/j.asoc.2014.11.040.

Ozturk, C., Hancer, E. and Karaboga, D. (2015b), "Improved clustering criterion for image clustering with artificial bee colony algorithm", *Pattern Anal. Appl.*, **18**(3), 587-599. https://doi.org/10.1007/s10044-014-0365-y.

Pal, N.R. and Bezdek, J.C. (1995), "On cluster validity for the fuzzy c-means model", *IEEE T. Fuzzy Syst.*, **3**(3), 370-379. https://doi.org/10.1109/91.413225.

Pan, D., Li, S., Xu, Z., Zhang, Y., Lin, P. and Li, H. (2019), "A deterministic-stochastic identification and modelling method of discrete fracture networks using laser scanning: Development and case study, *Eng. Geol.*, **262**, 105310.

Regassa, B., Xu, N. and Mei, G. (2018), "An equivalent discontinuous modeling method of jointed rock masses for DEM simulation of mining-induced rock movements", *Int. J. Rock Mech. Min. Sci.*, **108**, 1-14. https://doi.org/10.1016/j.ijrmms.2018.04.053.

Shanley, R.J. and Mahtab, M.A. (1976), "Delineation and analysis of clusters in orientation data", *J. Int. Assoc. Math. Geol.*, **8**(1) 9–23. https://doi.org/10.1016/0148-9062(76)91825-8.

Song, J., Huang R. and Pei, X. (2012), "Particle swarm optimization algorithm based fuzzy c-means cluster analysis for discontinuities occurrence in rock mass", *J. Eng. Geol.*, **20**(4), 591-598. (in Chinese).

Su, Z.G., Zhou, H.Y. and Hao, Y.S. (2021), "Evidential evolving C-means clustering method based on artificial bee colony algorithm with variable strings and interactive evaluation mode", *Fuzzy Optim. Decis. Mak.*, **20**(3), 293-313. https://doi.org/10.1007/s10700-020-09344-7.

Tereshko, V. (2000), Reaction-diffusion model of a honeybee colony's foraging behaviour, in *Parallel Problem Solving from Nature VI,Lecture Notes in Computer Science* M. Schoenauer(Editor), Springer–Verlag, Berlin, 807-816. https://doi.org/10.1007/3-540-45356-3_79.

Tereshko, V. and Lee, T. (2002), "How information-mapping patterns determine foraging behaviour of a honey bee colony", *Open Syst. Inf. Dyn,*. **09**(2), 181-193. https://doi.org/10.1023/a:1015652810815.

Tereshko, V. and Loengarov, A. (2005), "Collective decision making in honey-bee foraging dynamics", *Comput. Inf. Syst.*, **9**(3), 1-7.

Wang, M., Liu, Y., Qi, B., Li, D. and Wang, W. (2020), "BSE-based dispersion quantification of carbon fibres in cementitious composite through FCM-CA approach", *Constr. Build. Mater.*, **260**, 119789. https://doi.org/10.1016/j.conbuildmat.2020.119789.

Xie, X.L. and Beni, G. (1991), "A validity measure for fuzzy clustering", *IEEE T. Pattern Anal. Mach. Intell.*, **13**(8), 841-847. https://doi.org/10.1109/34.85677.

Xu, L.M., Chen, J.P., Wang, Q. and Zhou F.J. (2013), "Fuzzy C-means cluster analysis based on mutative scale chaos optimization algorithm for the grouping of discontinuity sets", *Rock Mech. Rock Eng.*, **46**(1), 189-198. https://doi.org/10.1007/s00603-012-0244-z.

Xue, X. and Zhou, H. (2018), "Neuro-fuzzy based approach for estimation of concrete compressive strength", *Comput. Concret*, **21**(6), 697-703. http://dx.doi.org/10.12989/cac.2018.21.6.697.

Yang, X.X., Jing, H.W., Tang, C.A. and Yang, S.Q. (2017), "Effect of parallel joint interaction on mechanical behavior of jointed rock mass models", *Int. J. Rock Mech. Min. Sci.*, **92**, 40-53. https://doi.org/10.1016/j.ijrmms.2016.12.010.

Zhang, W., Dai, B., Liu, Z. and Zhou, C. (2016), "Modeling

discontinuous rock mass based on smoothed finite element method", *Comput. Geotech.*, **79**, 22-30. https://doi.org/10.1016/j.compgeo.2016.05.020.

Zhang, L. and . Einstein, H. (2000), "Estimating the intensity of rock discontinuities", *Int. J. Rock Mech. Min. Sci.*, **37**(5), 819-837. https://doi.org/10.1016/s1365-1609(00)00022-8.

Zhao, Z., Jing, H., Shi, X., Yang, L., Yin, Q. and Gao, Y. (2021), "Study on bearing characteristic of rock mass with different structures: Physical modeling", *Geomech. Eng.*, **25**(3), 179-194. https://doi.org/10.12989/gae.2021.25.3.179.

Zhou, W. and Maerz, N.H. (2002), "Implementation of multivariate clustering methods for characterizing discontinuities data from scanlines and oriented boreholes", *Comput. Geosci.*, **28**(7), 827-839. https://doi.org/10.1016/s0098-3004(01)00111-x.

CC