Journal of Rock Mechanics and Geotechnical Engineering 14 (2022) 366–376

Contents lists available at ScienceDirect

Journal of Rock Mechanics and Geotechnical Engineering

journal homepage: www.jrmge.cn

Full Length Article

A new multiple-factor clustering method considering both box fractal dimension and orientation of joints

Tiexin Liu [a], Annan Jiang [a,*], Jianhui Deng [b], Jun Zheng [c], Zhenghu Zhang [d]

[a] Department of Civil Engineering, Dalian Maritime University, Dalian, 116026, China
[b] State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource & Hydropower, Sichuan University, Chengdu, 610065, China
[c] Department of Civil Engineering, Zhejiang University, Hangzhou, 310058, China
[d] State Key Laboratory of the Coastal and Offshore Engineering, School of Civil Engineering, Dalian University of Technology, Dalian, 116024, China

ARTICLE INFO

Article history:
Received 27 March 2021
Received in revised form
29 May 2021
Accepted 17 July 2021
Available online 1 November 2021

Keywords:
Joint clustering
Box fractal dimension (BFD)
Orientation
Hard boundary

ABSTRACT

The paper proposes a new multiple-factor clustering method (NMFCM) with consideration of both box fractal dimension (BFD) and orientation of joints. This method assumes that the BFDs of different clusters were uneven, and clustering was performed by redistributing the joints near the boundaries of clusters on a polar map to maximize an index for estimating the difference of the BFD (DBFD). Three main aspects were studied to develop the NMFCM. First, procedures of the NMFCM and reasonableness of assumptions were illustrated. Second, main factors affecting the NMFCM were investigated by numerical simulations with disk joint models. Finally, two different sections of a rock slope were studied to verify the practicability of the NMFCM. The results demonstrated that: (1) The NMFCM was practical and could effectively alleviate the problem of hard boundary during clustering; (2) The DBFD tended to increase after the improvement of clustering accuracy; (3) The improvement degree of the NMFCM clustering accuracy was mainly influenced by three parameters, namely, the number of clusters, number of redistributed joints, and total number of joints; and (4) The accuracy rate of clustering could be effectively improved by the NMFCM.

1. Introduction

The mechanical and hydraulic behaviors of rock masses are influenced by the properties of joints (Hammah and Curran, 2000). Usually, the number of joints in the rock mass is too large to analyze them individually (Liu et al., 2016). With similar geometric properties, different joints are grouped into the same cluster. Thus, the joints could be analyzed by mathematical statistical methods simultaneously, and lots of workload can be saved. The detection of joint clusters is a main part of the statistical analysis of field data (Esmaeilzadeh and Shahriar, 2019). The clustering of joints is an important premise for civil engineering projects and mining applications (Hammah and Curran, 1998).

According to the incorporated amount of joint properties, clustering methods can be divided into two main categories: single-factor clustering methods (SFCM) and multiple-factor clustering methods (MFCM) (Liu et al., 2021). Orientation is the only property taken into account in the SFCM since it is the prime geological factor influencing the stability of rock mass (Duncan and Christopher, 2004), whereas at least one another property is simultaneously considered in the MFCM.

The most common SFCM is a contouring method (Klose et al., 2005; Xu et al., 2013), in which joint clusters are judged according to the density of joint poles. The method has been widely used due to its convenience. However, the clustering results strongly depend on the experience of handlers (e.g. the size of reference circle). Therefore, many researchers have devoted themselves to developing new clustering methods with less manual intervention and higher clustering accuracy. Two main aspects have been investigated, including the automatic judgments of the amount and center positions of initial clusters (Xie and Beni, 1991; Yamaji and Sato, 2011; Joopudi et al., 2013; Ma et al., 2015; Liu et al., 2017), and the developments of further excellent rules for clustering (Shanley and Mahtab, 1976; Hammah and Curran, 1998, 1999; Klose et al., 2005; Jimenez-Rodriguez and Sitar, 2006; Xu et al., 2013; Esmaeilzadeh and Shahriar, 2019). After a series of studies, the

* Corresponding author.
E-mail address: jiangannan@163.com (A. Jiang).
Peer review under responsibility of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences.

https://doi.org/10.1016/j.jrmge.2021.07.014

problems of these two aspects have been solved well. However, there is still a drawback that cannot easily be avoided for the SFCM. Because only the orientation is taken into account, the actual physical relationship among different properties of a joint cannot be considered in the analysis of the distance between joint poles on the polar map, which is often taken as an important index for clustering. Thus, joints are assigned to the closest cluster center in most of the SFCM, and the orientation distributions are truncated in the overlap areas of the polar map contributing to the boundaries between joint clusters being rigid rather than natural (Dershowitz et al., 1996; Joopudi et al., 2013). To solve this problem of hard boundary and make the boundaries between joint clusters more natural, the orientation distribution of the entire or a part of a cluster could be assumed to conform to Fisher distribution (Zhan et al., 2017), bivariate normal distribution (Kulatilake, 1986; Marcotte and Henry, 2002), or some other distributions (Yamajia and Sato, 2011).

Tokhmechi et al. (2011) found that the classification accuracy was about 49% when only the orientation was taken into account. The finding illustrated the importance of MFCM in identifying the joint clusters. Dershowizt et al. (1996) developed a MFCM, and suggested some weighting factors for the concerned properties of joints (e.g. rock type, joint coating and infilling, termination mode, weathering, and planarity). The weighting factors quantify the importance of each property in clustering. Considering the orientation, spacing, and roughness of joints, Zhou and Maerz (2002) developed an analytical platform that use both clustering analysis and various visualization tools to identify joint clusters. Tokhmechi et al. (2011) utilized different properties (orientation, continuity, roughness, aperture, and hardness) to group joints, and they found that the results are more accurate when more properties are considered during clustering. Taking the orientation, trace length, opening degree, and undulation of joints into account, Ding et al. (2018) developed a multi-parameter joint clustering method based on an improved iterative self-organizing data analysis (ISODATA) algorithm. With the principle of the minimum cluster validity index, the optimal weight configuration was determined for every property considered in the method. Assuming that the trace lengths of joints obey a specific distribution, Liu et al. (2021) developed a MFCM by redistributing the joints to make their trace lengths more closely obey a specific distribution.

Theoretically, the problem of hard boundary can be solved with these methods. However, the methods still have limitations because many properties cannot be easily selected and abundantly obtained in engineering practice, such as roughness, strength, and aperture (Pusch 1995; Liu et al., 2021). Additionally, some properties cannot be quantified objectively and readily, such as spacing and weathering, because these values are variable for well-developed joints. Therefore, it is necessary to develop a new multivariate clustering method in which the considered joint properties should be conveniently obtained and the joints are clustered with enough physical meaning.

The geometric features of joints are widely considered to have fractal characteristics (Kulatilake et al., 1997; Zhou and Xie, 2003; Li et al., 2020). The characteristics can be quantified by the fractal dimension. As a type of fractal dimensions, the box fractal dimension (BFD) of joint clusters could capture the combined changes of density and trace length (Kulatilake et al., 1997). The density is defined as the number of joints in unit volume of rock mass. When joints intersect with rock surfaces, traces of these joints arise, and the trace length can directly reflect the scale of joint (Hammah and Curran, 1998). Both of these properties are important for the identification of different clusters in addition to the orientation, which is also easy to be obtained. A fracture cluster is defined as a group of fractures with statistically similar properties, which means that the properties (joint density and trace length) of different clusters are different. Therefore, different clusters should have different BFDs.

This paper proposes a new MFCM (NMFCM) with the assumption that the BFDs of different clusters are uneven, with the corresponding algorithm code developed in MATLAB. The NMFCM performs clustering by redistributing the joints near the boundaries of clusters on a polar map to maximize an index for estimating the difference of BFDs (DBFD). The paper is organized as follows. First, procedures of the NMFCM and reasonableness of the assumption are introduced. Second, the main influencing factors of the NMFCM are investigated by numerical simulations in which the joint data are synthetic and the clustering results can be clearly captured. Finally, an engineering case of a rock slope is used to illustrate the practicability of the NMFCM.

2. The new multi-factor clustering method

The NMFCM in this research includes three main modules: calculation of membership degree value, judgment of optimum number of clusters, and fractal redistribution of joints. The first two modules constitute the fuzzy c-means method. This research mainly focuses on the third module, and the reasonableness of the assumption is also demonstrated in this module.

2.1. Calculation of membership degree value

The relevant upper hemispherical polar map could be plotted after obtaining the orientation of all the joints in the studied rock mass. Based on the polar map, the cluster number and the initial cluster centers could be assumed. According to the assumption, the membership degree values of the joints could be calculated. Taking Fig. 1 as an example, there are two clusters of joints on the polar map. The red points in the figure are the cluster centers.

The distance between a joint pole and a cluster center can be acquired as follows (Hammah and Curran, 1998):

$$d^2\left(\boldsymbol{X}_i, \boldsymbol{V}_j^n\right) = 1 - \left(\boldsymbol{X}_i \cdot \boldsymbol{V}_j^n\right)^2 \tag{1}$$

where $d\left(\boldsymbol{X}_i, \boldsymbol{V}_j^n\right)$ is the distance between the joint pole i and the cluster center j in the nth iteration, $\boldsymbol{X}_i$ is the orientation of observation, and $\boldsymbol{V}_j^n$ represents the orientation of the cluster center j in the nth iteration. In Fig. 1, d_1 and d_2 represent the distances

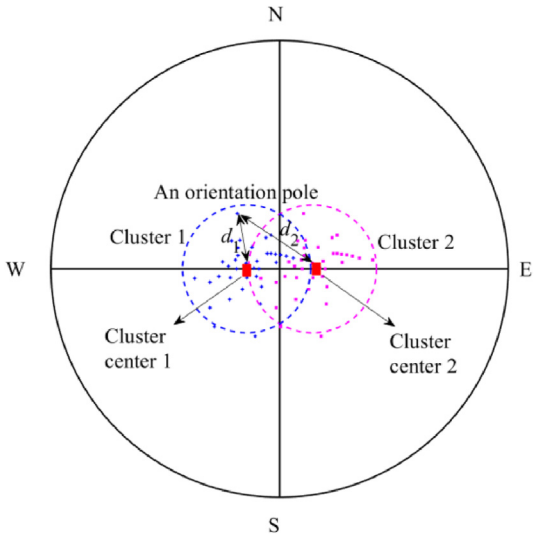

Fig. 1. Diagram of calculation of membership degree value (Liu et al., 2021).

between a joint pole and the two cluster centers, respectively. After obtaining the distances, the membership degree values can be computed as (Hammah and Curran, 1998):

$$u_{ij} = \left[d^2(\boldsymbol{X}_i, \boldsymbol{V}_j)^{1/(1-m)}\right]\left[\sum_{k=1}^{K} d^2(\boldsymbol{X}_i, \boldsymbol{V}_k)^{1/(1-m)}\right]^{-1} \quad (2)$$

where u_{ij} is the membership degree value (measure of belonging) of $\boldsymbol{X}_i$ in the cluster j; m represents the degree of fuzzification and $m = 2$ is believed to be the best value for most applications; and K is the total number of clusters (generally ranging from 2 to 4), with each value corresponding to a different clustering result. From Fig. 1, the joint represented by the orientation pole can be intuitively perceived to belong to cluster 1 because d_1 is smaller than d_2. This kind of relationship can be quantified as the membership degree value.

After determining the membership degree values, new cluster centers can be recalculated as (Hammah and Curran, 1998):

$$\boldsymbol{V}_j = \sum_{i=1}^{N} u_{ij}^m \boldsymbol{X}_i \Big/ \sum_{i=1}^{N} u_{ij}^m \quad (3)$$

$$\Delta\boldsymbol{V} = \sum_{i=1}^{K}\left[\left(x_{v_i^n} - x_{v_i^{n-1}}\right)^2 + \left(y_{v_i^n} - y_{v_i^{n-1}}\right)^2\right]^{0.5} \quad (4)$$

where N is the total number of joints required to be clustered, and $N \geq K$; x and y signify the coordinates of $\boldsymbol{V}$; and $\Delta\boldsymbol{V}$ is the difference of the cluster center vectors in two adjacent iterations. If $\Delta\boldsymbol{V}$ is greater than the critical value (0.01 is recommended), then the old cluster centers are replaced by the new ones, and a new iteration begins; otherwise, the iterations are terminated and the membership degree values and the cluster centers are output. It should be noted that different values of K have to be tested in this module to obtain the optimal result.

2.2. Judgment of optimum number of clusters

The Xie–Beni index (Xie and Beni, 1991) could estimate the ratio of compactness to separation of the resulting fuzzy division after a set of data has been divided into K clusters. Ma et al. (2015) proposed a simplified Xie–Beni index, in which acute angle is utilized to represent the distance between fractures. The simplified index has the advantages of simple programming, high precision and practicability. Therefore, the simplified Xie–Beni cluster validity index was adopted in this research to judge the optimum number of clusters, which can be determined as

$$SXB_K = \frac{\sum_{i=1}^{N}\sum_{j=1}^{K} I_{ij} s^2(\boldsymbol{X}_i, \boldsymbol{V}_j)}{\min_{j \neq l}\left[s^2(\boldsymbol{V}_j, \boldsymbol{V}_l)\right]} \quad (5)$$

where SXB_K is the simplified Xie–Beni cluster validity index when N joints are divided into K clusters; and I_{ij} is a Boolean value, which is equal to 1 if $\boldsymbol{X}_i$ belongs to the cluster j, otherwise it is equal to 0. The simplified distance between two unit normal vectors $\boldsymbol{A} = (x_1, y_1, z_1)$ and $\boldsymbol{B} = (x_2, y_2, z_2)$ in Cartesian coordinates could be computed as

$$s(\boldsymbol{A}, \boldsymbol{B}) = \arccos\left|\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}\right| = \arccos|x_1x_2 + y_1y_2 + z_1z_2| \quad (6)$$

The value of K that minimizes SXB_K was determined as the optimum number of clusters.

2.3. Fractal relocation of joints

To make the numerical simulations closer to the situation of real engineering projects, the trace lines of joints used in this research were generated by the intersection of a sampling window with a disk joint model (Fig. 2). This method was employed due to its good coincidence with real rock masses in the aspects of mechanical and hydraulic behaviors (Dershowitz and Einstein, 1988; Kulatilake et al., 1993; Liu et al., 2015; Lei et al., 2017). The disk joint model is mainly controlled by three parameters, namely, the orientation, radius, and density. When the model is divided by a sampling window, the joint disks are transformed into trace lines on the window (Fig. 3). The sizes of the joint models and the sampling windows were kept the same in this research, as shown in Fig. 2.

To illustrate the reasonableness of using the DBFD as an additional parameter to cluster joints, two kinds of tests were executed using the numerical simulation method introduced above. Two joint clusters were generated in all of the models described in this section, and correct cluster code of each joint that indicates the cluster affiliation was clearly known.

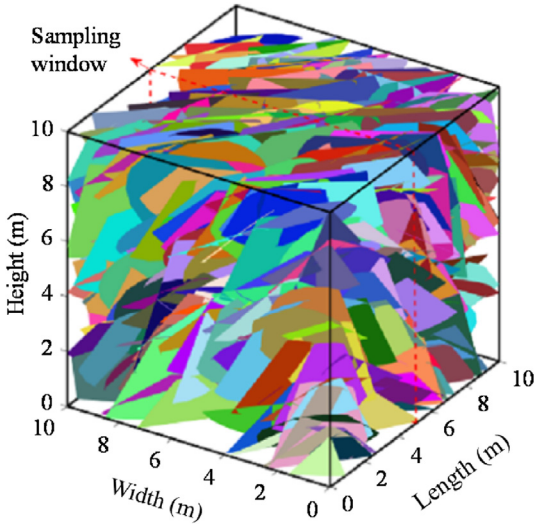

Fig. 2. Sketch map of the disk joint model.

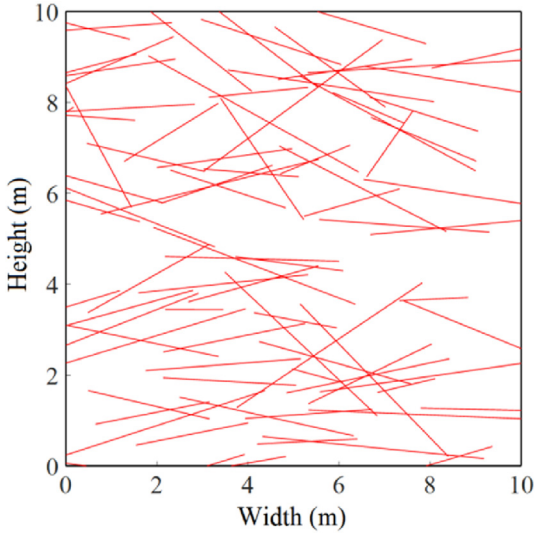

Fig. 3. Trace lines on the sampling window.

Table 1
Parameters used in different disk joint models.

Cluster No.	Volume density range (m^{-3})	Mid-value of the volume density range (m^{-3})	Orientation distribution (Fisher distribution)			Radius distribution (gamma distribution)		
			Dip direction (°)	Dip angle (°)	K_F	Mean value range (m)	Mid-value of the mean value range (m)	Variance (m^2)
1	0.05–0.15	0.1	90	20	10	2.3–2.7	2.5	0.01
2	0.1–0.2	0.15	270	20	10	1.3–1.7	1.5	0.01

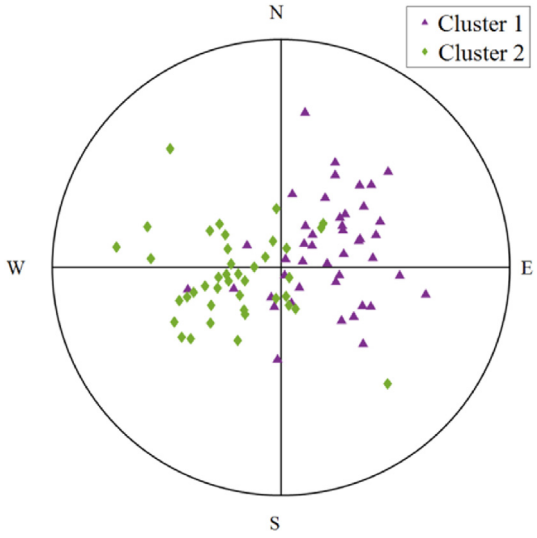

Fig. 4. Polar map of joints in each cluster of the generated model.

The first kind of tests was executed using the parameters listed in Table 1. It should be noted that the mid-values of the density and the radius were adopted in the disk joint model. After the model was generated, the polar map of the joints in the sampling window can be easily obtained, as presented in Fig. 4. It is worth noting that not all the joints generated in the model intersect with the sampling window. Fig. 4 only shows the poles of joints that intersect with the sampling window, and this was taken as the standard for correct clustering. Figs. 5 and 6 show the procedures of the BFD calculation for each joint cluster corresponding to Fig. 4.

Three main steps were required to finish the BFD calculation in the NMFCM. First, the sampling window was discretized into small square boxes with different sizes (denoted as r in Figs. 5 and 6). The box sizes should be a geometric progression with a fixed ratio. In this study, box sizes of 2 m, 1 m, 0.5 m, 0.25 m, and 0.125 m were selected to make the number of small square boxes in the sampling window integers in the NMFCM. Figs. 5a and 6a show the box networks when the box size is equal to 1 m, and Figs. 5b and 6b show the box networks when the box size takes all the values. Second, the amount of square boxes needed to completely cover the traces in the sampling window were counted when the box size took different values. Finally, the amount of the square boxes (N) was plotted versus the inverse of the box sizes ($1/r$) in a log-log plot, and the slope of the fitting line obtained by the least-square regression was determined as the BFD of the traces, as shown in Figs. 5c and 6c.

There were 44 and 41 traces in the two clusters, and the corresponding BFDs were 1.46 and 1.27, respectively. Consequently, the DBFD of the two clusters was equal to 0.19, and the accuracy rate of clustering (defined as the number of joints grouped into the right clusters divided by the total number of joints) was equal to 100%.

Then, some joints were randomly exchanged to another cluster to reduce the accuracy rate. The number of exchanged joints ranged from 1 to 20. The upper limit of the range was the maximum integer that is smaller than half of the number of joints in the smaller cluster, and it was taken as 20 in this study. If the number of exchanged joints exceeds the upper limit, a whole cluster exchange would take place and seriously affected the accuracy rate. The box dimension and the accuracy rate were recalculated when the number of exchanged joints varied.

Fig. 7 shows the relationship between the box dimensions and the accuracy rates of different clusters. Based on Fig. 7, the following conclusions could be drawn:

(1) The BFD of clusters, B_F, could capture the combined change of density and trace length.
(2) The DBFD (denoted as D_B) reached the maximum when the accuracy rate of clustering was equal to 100%.
(3) With the decrease of the accuracy rate, the geometric features of joints in different clusters became closer and closer, and the DBFD decreased accordingly.

The second kind of tests was performed to eliminate the influence of discreteness. One hundred groups of parameters were generated evenly in the ranges of Table 1 and 100 models were established based on these parameters. The statistical relationship between the DBFD and the accuracy rate is drawn in Fig. 8. The greater the difference, the higher the accuracy rate of clustering. The trace lengths and densities of different clusters were different in most cases. As a consequence, the BFDs were also different. The worse the clustering result, the more random the joint distribution in different clusters, and the smaller the difference of the fractal dimensions. The results of the two kinds of tests meant that using the DBFD as an auxiliary clustering index could improve the accuracy rate.

Based on the above conclusions, a new method was proposed in this paper and used to perform clustering by redistributing the joints near the boundaries of clusters on the polar map to maximize the index for estimating the DBFD. The joints required to be redistributed could be determined by

$$u_i^{\max} - u_i^{\min} \le \Delta u \tag{7}$$

where $u_i^{\max}$ and $u_i^{\min}$ represent the maximum and minimum membership degree values of joint i, respectively (it should be noted that one joint has K membership degree values); and Δu is a threshold required to be specified, and its advised values are given in Section 3. The smaller the value of Δu, the closer the location to the boundary of joint i.

The full permutations can be obtained by redistributing all of the joints selected by Eq. (7). The BFD of the clusters in the full permutations is then calculated, and the permutation that maximizes Eq. (8) is selected, where F_{obj} is the index for estimating the DBFD. Fig. 9 shows the flowchart of main steps of NMFCM.

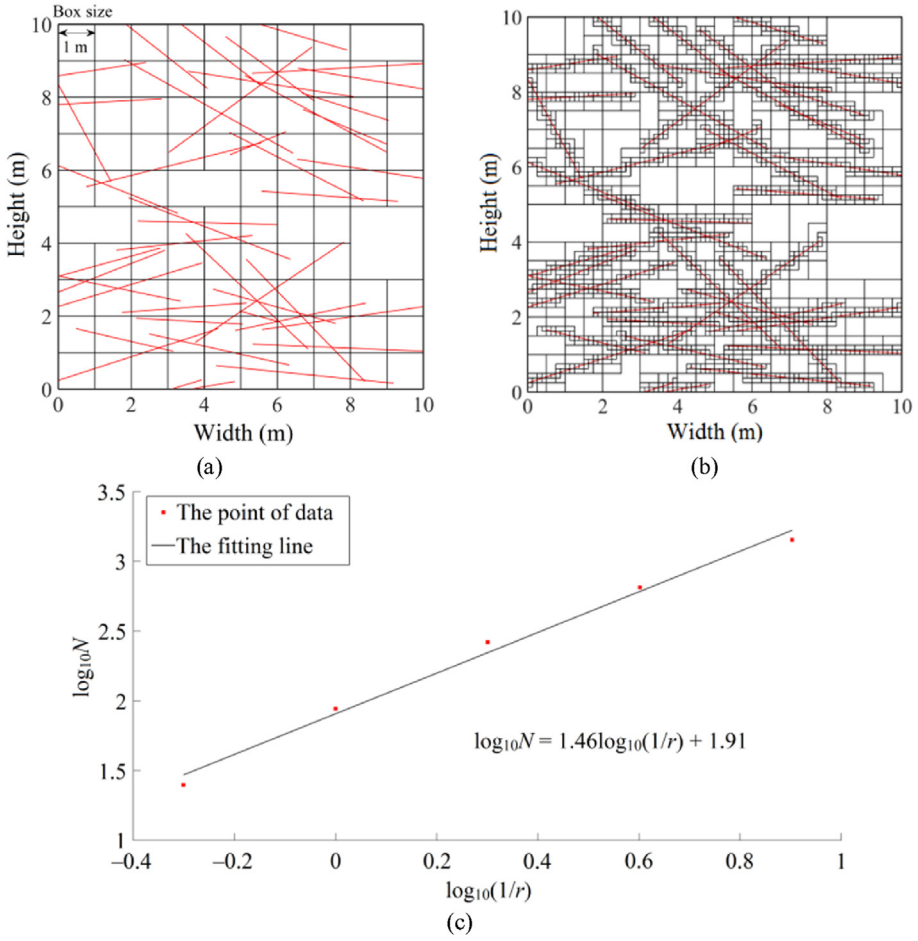

Fig. 5. Procedures of BFD calculation for cluster 1: (a) The box network when $r = 1$ m; (b) The box networks when r takes different values; and (c) The relationship between the number of boxes needed to completely cover the traces in cluster 1 and the corresponding box sizes ($BFD_1 = 1.46$).

$$F_{\text{obj}} = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} |\text{BF}(i) - \text{BF}(j)| \tag{8}$$

It is worth noting that if the number of joints selected by Eq. (7) was quite large, too many permutations (K^n, where n is the number of the selected joints) would be required to be tested and cause a very long execution time. This problem could be solved by determining the number of joints that could be redistributed simultaneously. This number denoted as N_o could be selected from the range of 7–15 according to the computation capacity. Based on the number, joints were selected randomly within the range determined by Eq. (7). Then, the process of selection was repeated to ensure that each joint had the opportunity to be selected. Taking 0.01 as the interval length, the calculation results were classified according to Eq. (8). The clustering results were averaged at the maximum of the intervals to obtain the final result.

3. Analysis of the new multiple-factor clustering method with synthetic data

Because the clustering results in real rock masses are difficult to be verified, a series of numerical simulations using disk joint models was designed to characterize the NMFCM. This section analyzes the main influencing factors of NMFCM and demonstrates the executing processes of NMFCM.

3.1. Analysis of main influencing factors

As described in Section 2, three parameters directly affected the running efficiency and the precision of the NMFCM, namely, Δu, N_o, and K. Δu controlled the redistribution range of joints. The greater the value, the wider the range. N_o governed the computation time when the full permutations determined by Δu could not be completely tested due to the limitations of the computation capacity. The greater the value is, the more time the computation costs; and the greater the number of clusters is, the more complicated the clustering results are.

Effects of Δu and N_o on the accuracy rate were studied firstly. According to the parameter range listed in Table 1, 100 disk joint models were generated. The number of clusters in all the models remained the same of 2. The values of Δu ranged from 0.1 to 0.9 at regular intervals of 0.1, and N_o ranged from 7 to 12 with an interval length of 1. Fig. 10 shows the influence of the two parameters on the promotion of the accuracy rate. The promotion rate was equal to the accuracy rate of the NMFCM minus that of the fuzzy c-means method, and this was the contribution made by the module of the fractal relocation of the joints.

From Fig. 10, the following findings could be obtained:

(1) The promotion of the accuracy rate increases with the increase of N_o.
(2) The promotion presents a trend of increasing firstly and then decreasing as Δu increases.

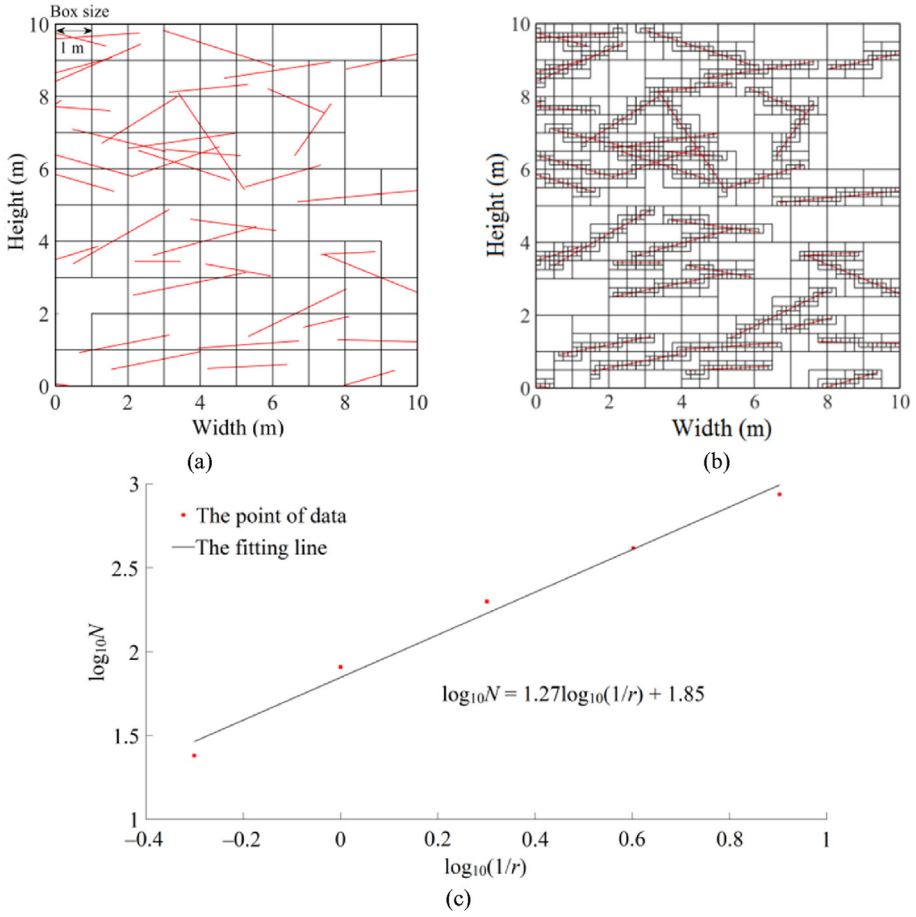

Fig. 6. Procedures of BFD calculation for cluster 2: (a) The box network when $r = 1$ m; (b) The box networks when r takes different values; and (c) The relationship between the number of boxes needed to completely cover the traces in cluster 2 and the corresponding box sizes ($BFD_2 = 1.27$).

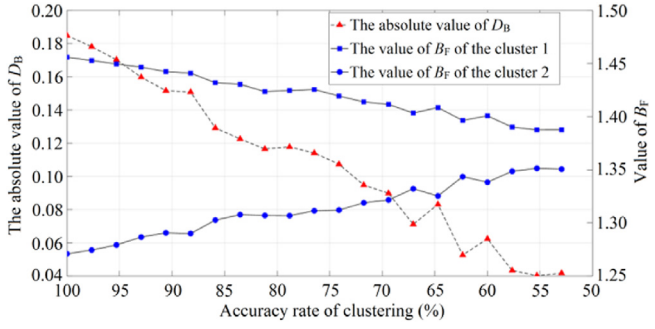

Fig. 7. The relationship between the difference of box fractal dimension (D_B) and the accuracy rate. BF denotes the box fractal dimension of clusters.

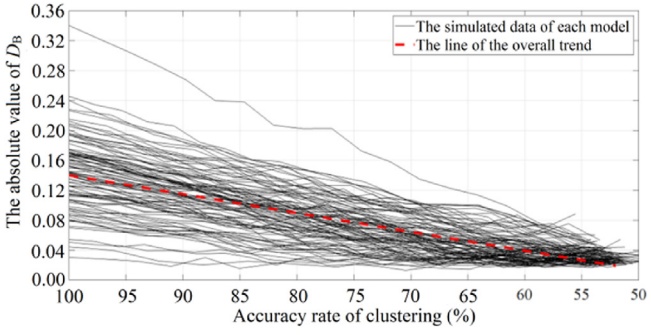

Fig. 8. The statistical relationship between D_B and accuracy rate (100 samples).

(3) The influence degree of N_o on the promotion rises with the increase of Δu when Δu is smaller than 0.4.

When Δu is small, the number of redistributed joints determined by Δu is less than N_o, thus the changes of N_o do not affect the clustering results. With the increase of Δu, the number of redistributed joints rises quickly, and the promotion of accuracy rate increases with the increase of N_o. With the further increase of Δu, more and more joints far from the boundary are selected, and the promotion of accuracy rate begins to decrease. Consequently, a larger value of Δu may not generate a better result. Considering the calculation efficiency and accuracy, the optimal value of Δu was considered as 0.4 when $K = 2$.

Subsequently, the effect of K on the promotion of accuracy rate was studied. Two kinds of disk joint models were generated (50 models were built for each model), and the amount of clusters in the models was $K = 3$ and $K = 4$, respectively. The parameters of clusters 1–3 were used, as listed in Table 2. Fig. 11 shows the effect of K on the promotion of accuracy rate at different values of N_o and Δu. Fig. 12 presents the running time required to finish the numerical simulations. The CPU of the computer was an Intel(R) Core(TM) i7-9700 with 3.00 GHz.

From Figs. 11 and 12, the following conclusions could be drawn:

(1) The promotion of accuracy rate linearly rises with the increase of N_o.

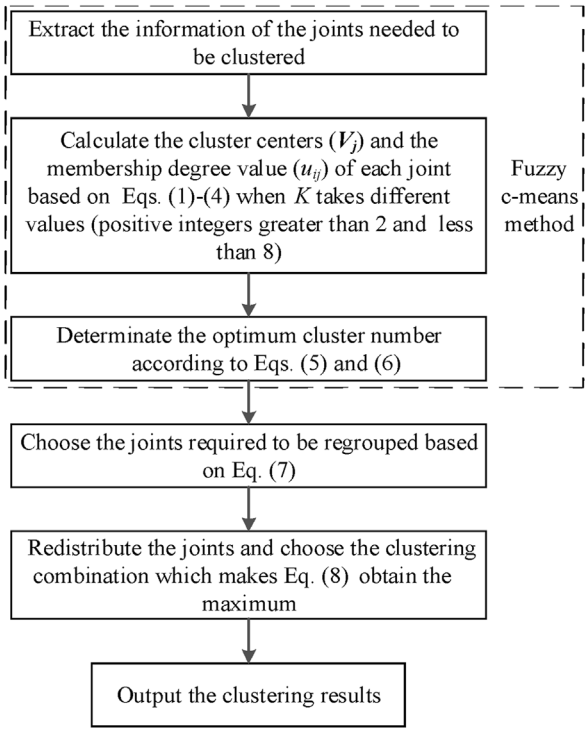

Fig. 9. Flowchart of the main steps for the NMCM.

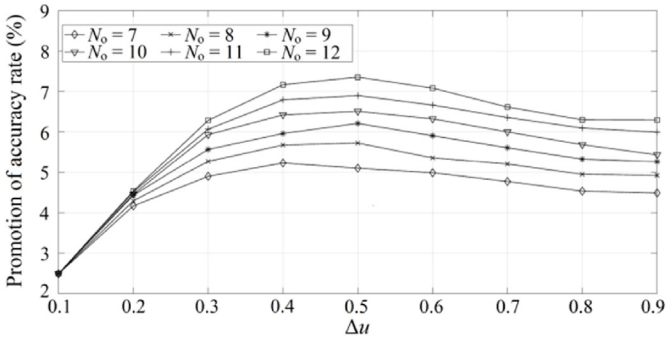

Fig. 10. Effect of Δu on the promotion of accuracy rate when N_o takes different values.

Table 2
Parameters used in different models.

Cluster No.	Volume density range (m^{-3})	Orientation distribution (Fisher distribution)			Radius distribution (gamma distribution)	
		Dip direction (°)	Dip angle (°)	K_F	Mean value range (m)	Variance (m^2)
1	0.1–0.14	0	20	10	0.3–0.7	0.01
2	0.08–0.12	90 (120)	20	10	0.8–1.2	0.01
3	0.06–0.1	180 (240)	20	10	1.3–1.7	0.01
4	0.04–0.08	270	20	10	2.3–2.7	0.01

Note: Data in and out of the brackets are used for cases of $K = 3$ and $K = 4$, respectively.

(2) For $K = 3$, the promotion reaches the peak at $\Delta u = 0.3$, whereas this occurs at $\Delta u = 0.2$ for the case of $K = 4$.
(3) The optimum value of Δu has a descending trend with the increase of K. The optimum values of Δu were suggested as 0.4, 0.3, and 0.2 when K was equal to 2, 3 and 4, respectively.

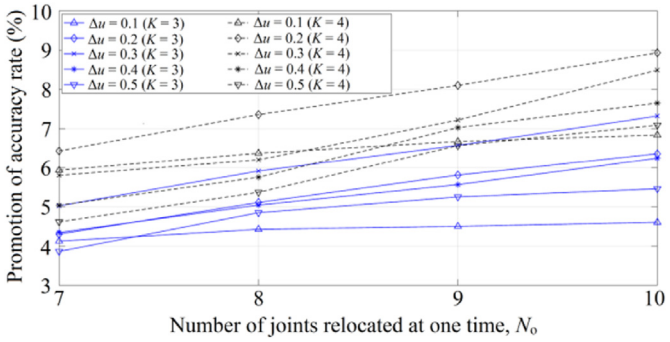

Fig. 11. The effect of K on the promotion of accuracy rate when N_o takes different values.

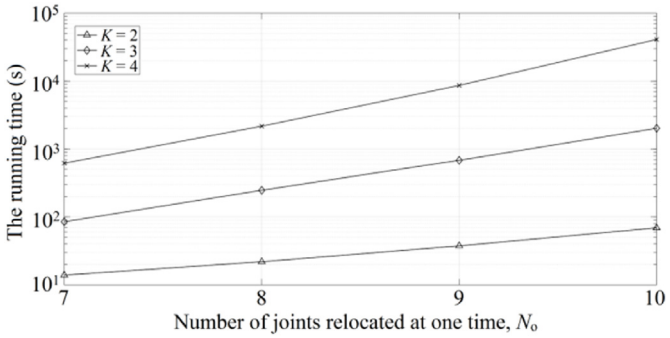

Fig. 12. The running time of the new multiple-factor clustering method when N_o and K take different values.

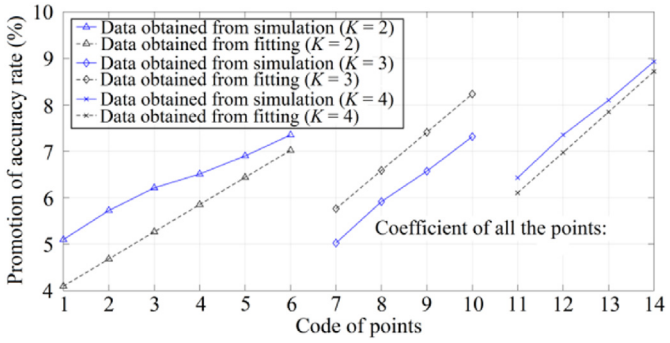

Fig. 13. The estimation goodness of Eq. (9) for the promotion of the accuracy rate.

(4) The running time exponentially rises with the increase of N_o and K, and the effect of K is more significant.

Considering the effect of K and N_o on the promotion of accuracy rate, Eq. (9) was used to estimate the value of the promotion when Δu was equal to the optimum values. The data points satisfying the conditions of Δu in Figs. 10 and 11 are plotted and displayed as the blue marks in Fig. 13. The corresponding fitting data points obtained by Eq. (9) are displayed by the dark marks. From Fig. 13, it is easy to find that the precision of the equation is good. The correlation coefficient of two sets of variables is a measure of their linear dependence, as defined in Eq. (10). The correlation coefficient of the fitting equation and numerical simulation for the promotion of accuracy rate reaches 0.85, revealing that there is a strong correlation between them. The relatively large error is attributed to two reasons when $K = 2$. First, two more data points are required to be

fitted when $K = 2$. Second, the number of joints when $K = 2$ is less than that when $K = 4$, which increases the volatility of the data.

$$E(p) = \frac{(K-1)N_o}{KN} \tag{9}$$

$$\rho(A, B) = \frac{1}{M-1}\sum_{i=1}^{M}\frac{A_i - \mu_A}{\sigma_A}\frac{B_i - \mu_B}{\sigma_B} \tag{10}$$

where $E(p)$ represents the expectation of the promotion of accuracy rate; M is the number of variables in each set; μ_A and σ_A are the mean and the standard deviation of A, respectively; and μ_B and σ_B are the mean and the standard deviation of B, respectively.

3.2. Demonstration of executing processes

The generated joint data described in Section 2.3 were analyzed to demonstrate the executing processes of NMFCM. The information of orientations and traces of the joints required to be clustered is shown in Figs. 4–6, but the correct cluster codes of the joints were assumed to be unknown in this part.

According to Fig. 9, different numbers of clusters were tested to find the optimal number. The simplified Xie–Beni cluster validity index was calculated based on Eqs. (5) and (6). The values of the index when the number of clusters ranged from 2 to 8 are given in Fig. 14. It was easy to determine that the optimal number is 2. Then

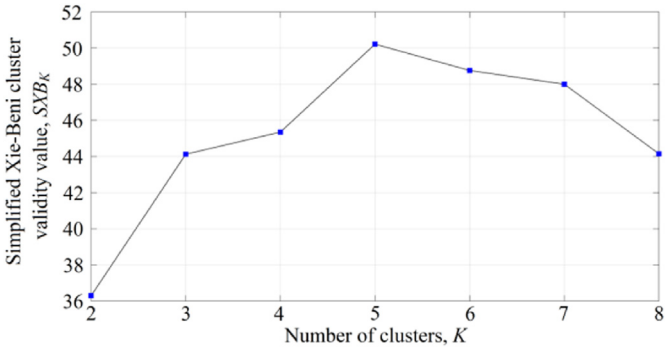

Fig. 14. The simplified Xie–Beni cluster validity index corresponding to different numbers of clusters.

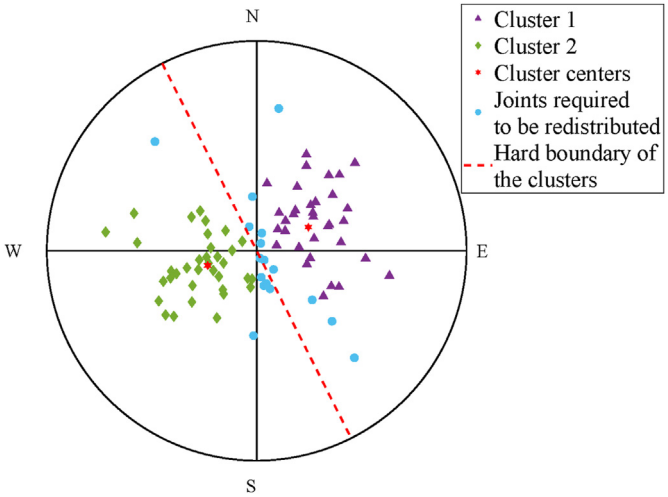

Fig. 15. The polar map of the clustering result according to the membership degree values of the fuzzy c-means method ($\Delta u = 0.4$).

the corresponding cluster centers and membership degree values could be determined with Eqs. (1)–(4).

Based on the suggestion provided in Section 3.1, Δu takes the value of 0.4. Fig. 15 shows the polar map of the clustering result according to only the membership degree values. There was a hard boundary dividing the joints into two clusters that was almost in the middle of the poles. The poles located at the upper half of the boundary fell into cluster 1, and those located at the lower half belonged to cluster 2. Joints near the boundary required to be redistributed were selected based on Eq. (7). Seventeen joints satisfied the conditions. These joints are displayed as the blue circle marks in Fig. 15.

When there were two clusters, the number of calculations ($2^{17} = 1.3 \times 10^5$) was acceptable. However, as the number of clusters became larger, the computation became very large (taking 4 for example, it was $4^{17} = 1.7 \times 10^{10}$). To demonstrate that the executing processes of the NMFCM is more effective, the reductive method introduced in Section 2.3 was adopted. N_o was selected as 10, therefore, 10 of the 17 joints were randomly selected and assigned to the two clusters. The processes were repeated to ensure that each joint had the opportunity to be selected. In this study, the processes were repeated ten times, and the DBFD of each combination of joints was calculated according to Eq. (8).

Taking 0.01 as the interval length, the calculation results were classified with the DBFD (see Fig. 16). The clustering results were averaged at the maximum of the intervals to get the final result. Fig. 17 shows the polar map of the final clustering result.

The following data were obtained during the execution. The DBFD was equal to 0.18 when the accuracy rate was 100%. The difference and the accuracy rate obtained with the fuzzy c-means method were equal to 0.11 and 83.53%, respectively. They increased to 0.17 and 88.24% after execution of the fractal redistribution of joints, respectively. The real promotion of accuracy rate was 4.71%, and the expectation of the promotion based on Eq. (9) was 5.88%. Based on the above data and Figs. 16 and 17, the following results could be concluded:

(1) With the increase of the DBFD, the accuracy rate rises. The correlation coefficient is 0.97, which suggests a strong correlation between the DBFD and the accuracy rate.
(2) The problem of the hard boundary that appears in the fuzzy c-means method can be effectively solved by the NMFCM.
(3) The reductive method introduced in Section 2.3 is practical.
(4) The accuracy rate of clustering can be improved by the NMFCM.

4. Case study

The Changhe dam hydropower station is a rockfill dam located in Sichuan Province, southwestern China (Fig. 18). Two different

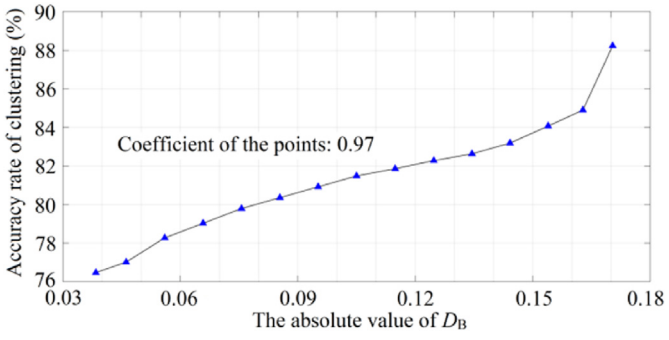

Fig. 16. The relationship between the DBFD and the accuracy rate.

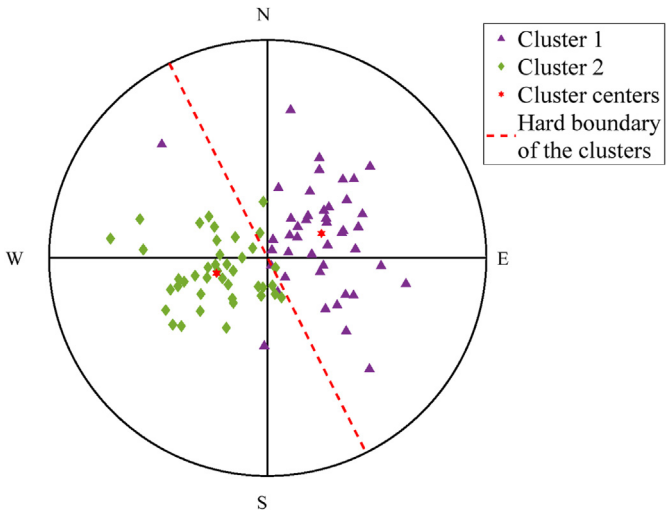

Fig. 17. The polar map of the final clustering result.

slope sections of its #3 spillway tunnel were studied (namely, cases 1 and 2, respectively, as shown in Fig. 19). According to field investigation, the studied sections of the slope mainly consist of two types of lithological rocks: medium–coarse grained granite and quartz diorite. Both types of rock masses are complete, hard, and developed notably well (Liu et al., 2016).

The indispensable input information for the NMFCM includes the data of orientations and traces of joints. Close-range digital photogrammetry was used as the mapping tool to obtain these data of the slope. The method of joint extraction and its accuracy have been extensively introduced by Liu et al. (2016). The accuracy could meet the requirements of engineering applications. All the joint traces are displayed in Fig. 20, and 64 and 51 joints with trace lengths greater than 2 m were analyzed for the two cases, respectively.

The validity indices of simplified Xie–Beni cluster were calculated for the two cases, and the optimal numbers of the clusters were determined as 2 and 3 for the two cases, respectively. Then the corresponding cluster centers and the membership degree

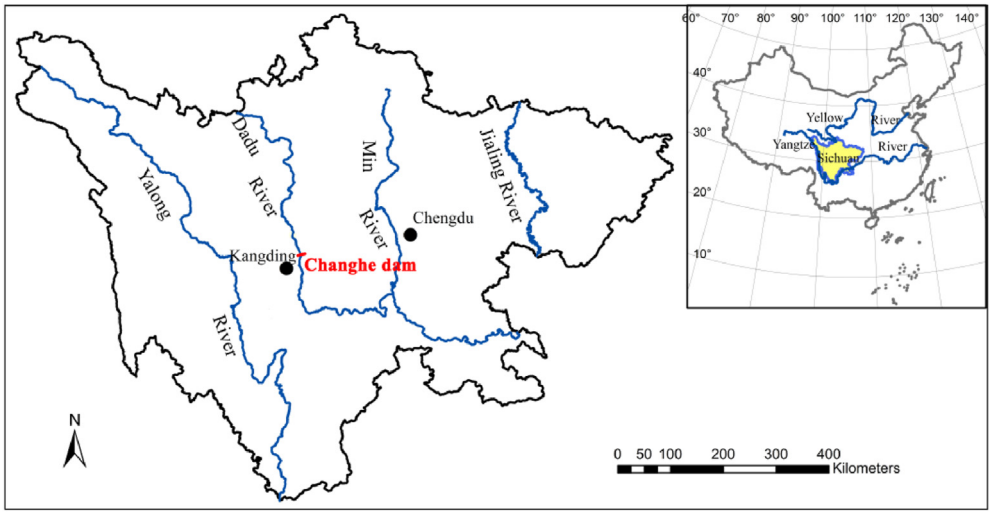

Fig. 18. Regional map of Changhe dam (Liu et al., 2016).

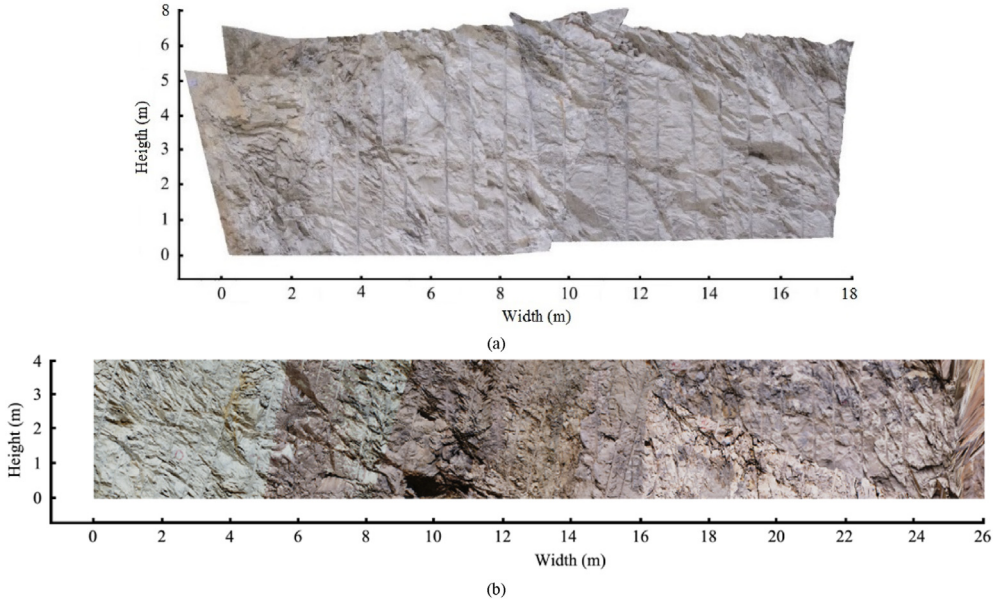

Fig. 19. A view of the two different sections of the slope surface: (a) Case 1 and (b) case 2.

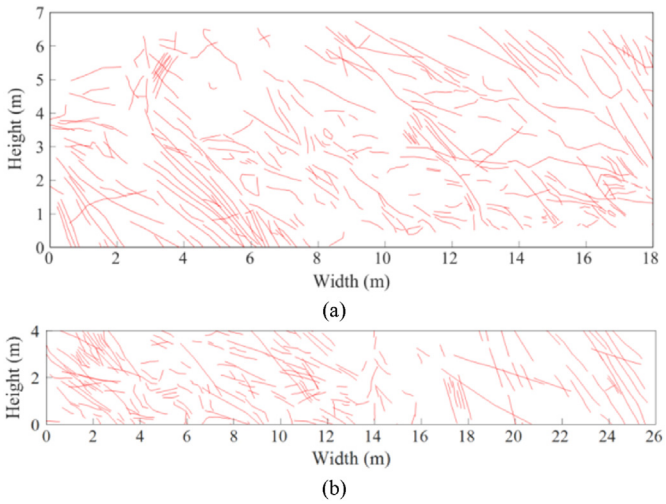

Fig. 20. The joint traces of the study area: (a) Case 1 and (b) case 2.

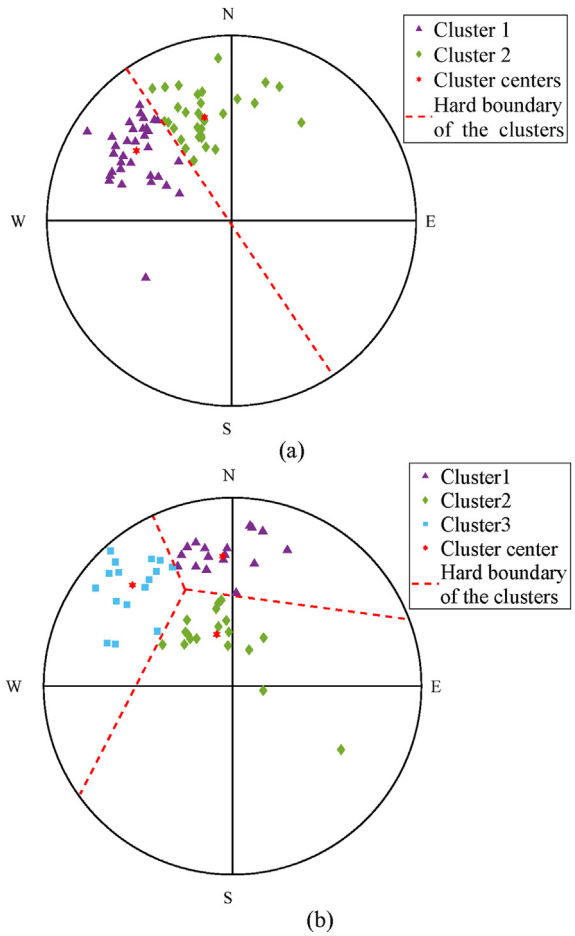

Fig. 21. The polar map of the clustering result according to the membership degree values of the fuzzy c-means method: (a) Case 1 and (b) case 2.

values were determined. Fig. 21 shows the polar map of the clustering result according to only the membership degree values. The joints were divided into 2 and 3 clusters. The red dotted lines in the figure represent the hard boundaries of different clusters. The boundaries were basically perpendicular bisectors of the lines between the cluster centers.

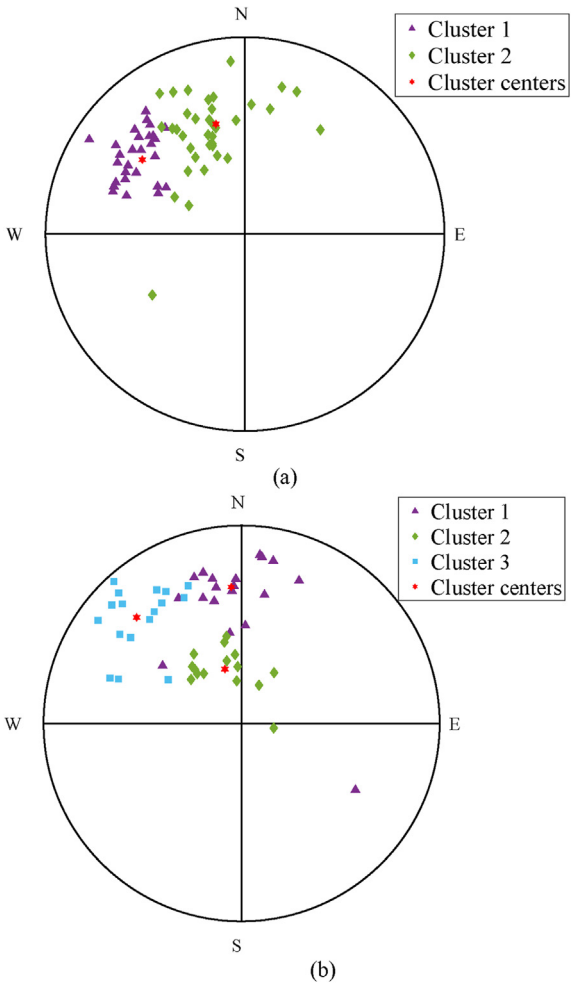

Fig. 22. The polar map of the clustering result obtained by the new multiple-factor clustering method: (a) Case 1 and (b) case 2.

For case 1, Δu and N_0 took the values of 0.4 and 12, respectively, and the DBFD was equal to 0.08 before execution of the fractal relocation. The expectation of the promotion based on Eq. (9) was 9.38%, and the DBFD of the two clusters increased to 0.16 after execution of the fractal relocation. For case 2, Δu and N_0 took values of 0.3 and 8, respectively, and the DBFD was equal to 0.09 before execution of the fractal relocation. The expectation of the promotion based on Eq. (9) was 10.46%, and the DBFD of the two clusters increased to 0.11 after execution of the fractal relocation. The running times for the two cases were 7.8 s and 14.7 s, respectively. Fig. 22 presents the polar map of the clustering result obtained with the NMFCM.

According to the case study, it can be found that: (1) The problem of hard boundary that appears in the fuzzy c-means method can be relieved effectively by the NMFCM; (2) The NMFCM could make the DBFD of the clusters increase; and (c) The NMFCM is practicable for engineering projects.

5. Discussion and conclusions

The number of clusters (K) had a significant effect on the promotion of accuracy rate and computation time. The effect was studied in detail when K had the values of 2, 3, and 4. However, situations where K is greater than 4 might be occasionally encountered in some engineering projects. The NMFCM was also applicable, but the data of joints had to be preprocessed before

execution of joint redistribution. The joints were selected in batches by cluster by judging whether the clusters have the same boundary. Then the selected joints could be clustered using the NMFCM. The process was repeated until all the boundaries were analyzed.

The development of NMFCM was described in this paper. The method accounted for the BFDs and orientations of joints simultaneously. This meant that the density, trace length, and orientation were considered together during clustering in the NMFCM. In this method, it was assumed that the BFDs of different clusters were uneven, and clustering was performed by redistributing the joints near the boundaries of clusters to maximize the DBFD. The reasonableness of the assumption and the main influencing factors were demonstrated with synthetic data. An equation was proposed to estimate the promotion of accuracy rate of clustering. An engineering case was studied to verify the practicability of the NMFCM. The results clearly demonstrated that:

(1) The assumption was reasonable, and with the improvement of the clustering accuracy, the DBFD tended to increase.
(2) The improvement degree of the clustering accuracy of the NMFCM was mainly controlled by three parameters, namely, the number of clusters, the number of redistributed joints, and the total number of joints.
(3) The promotion of accuracy rate was positively correlated with the number of redistributed joints, but negatively correlated with the number of clusters and the total number of joints.
(4) The method was practical and it could effectively alleviate the problem of hard boundary.
(5) The accuracy rate of clustering could be improved by using the NMFCM.

Declaration of 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.

Acknowledgments

The work was funded by the National Natural Science Foundation of China (Grant Nos. 41972264 and 52078093), Liaoning Revitalization Talents Program, China (Grant No. XLYC1905015).

References

Dershowitz, W., Busse, R., Geier, J., Uchida, M., 1996. A stochastic approach for joint set definition. In: Aubertin, M., Hassani, F., Mitri, H. (Eds.), Rock Mechanics Tools and Techniques: Proceedings of the 2nd North American Rock Mechanics Symposium (NARMS'96). A.A. Balkema, Rotterdam/Brookfield, pp. 1809–1813.
Dershowitz, W., Einstein, H.H., 1988. Characterizing rock joint geometry with joint system models. Rock Mech. Rock Eng. 21 (1), 21–51.
Ding, Q., Huang, R., Wang, F., Chen, J., Wang, M., Zhang, X., 2018. Multi- parameter dominant grouping of discontinuities in rock mass using improved ISODATA algorithm. Math. Probl Eng., 5619404, 2018.
Duncan, C.W., Christopher, W.M., 2004. Rock Slope Engineering, fourth ed. CRC Press, London, UK, pp. 7–10.
Esmaeilzadeh, A., Shahriar, K., 2019. Optimized fuzzy c-means–fuzzy covariance–fuzzy maximum likelihood estimation clustering method based on deferential evolutionary optimization algorithm for identification of rock mass discontinuities sets. Period. Polytech. Civ. Eng. 63 (2), 674–686.
Hammah, R.E., Curran, J.H., 1998. Fuzzy cluster algorithm for the automatic identification of joint sets. Int. J. Rock Mech. Min. Sci. 35 (7), 889–905.
Hammah, R.E., Curran, J.H., 1999. On distance measures for the fuzzy K-means algorithm for joint data. Rock Mech. Rock Eng. 32 (1), 1–27.
Hammah, R.E., Curran, J.H., 2000. Validity measures for the fuzzy cluster analysis of orientations. IEEE Trans. Pattern Anal. Mach. Intell. 22 (12), 1467–1472.
Jimenez-Rodriguez, R., Sitar, N., 2006. A spectral method for clustering of rock discontinuity sets. Int. J. Rock Mech. Min. Sci. 43 (7), 1052–1061.
Joopudi, S., Rathi, S.S., Narasimhan, S., Rengaswamy, R., 2013. A new cluster validity index for fuzzy clustering. IFAC Proc. 46 (32), 325–330.
Klose, C.D., Seo, S., Obermayer, K., 2005. A new clustering approach for partitioning directional data. Int. J. Rock Mech. Min. Sci. 42 (2), 315–321.
Kulatilake, P.H.S.W., 1986. Bivariate normal distribution fitting on discontinuity orientation clusters. Math. Geol. 18 (2), 181–195.
Kulatilake, P.H.S.W., Fiedler, R., Panda, B.B., 1997. Box fractal dimension as a measure of statistical homogeneity of jointed rock masses. Eng. Geol. 48 (3–4), 217–229.
Kulatilake, P.H.S.W., Wathugala, D.N., Stephansson, O., 1993. Joint network modelling with a validation exercise in Stripa mine, Sweden. Int. J. Rock Mech. Min. Sci. Geomech. Abstr. 30 (5), 503–526.
Lei, Q., Latham, J.P., Tsang, C.F., 2017. The use of discrete fracture networks for modelling coupled geomechanical and hydrological behaviour of fractured rocks. Comput. Geotech. 85, 151–176.
Li, Y., Sun, S., Yang, H., 2020. Scale dependence of waviness and unevenness of natural rock joints through fractal analysis. Geofluids 2020, 8818815.
Liu, J., Zhao, X., Xu, Z., 2017. Identification of rock discontinuity sets based on a modified affinity propagation algorithm. Int. J. Rock Mech. Min. Sci. 94, 32–42.
Liu, T., Deng, J., Zheng, J., Zheng, L., Zhang, Z., Zheng, H., 2016. A new semi-deterministic block theory method with digital photogrammetry for stability analysis of a high rock slope in China. Eng. Geol. 216, 76–89.
Liu, T., Zheng, J., Deng, J., 2021. A new iteration clustering method for rock discontinuity sets considering discontinuity trace lengths and orientations. Bull. Eng. Geol. Environ. 80, 413–428.
Liu, R., Jiang, Y., Li, B., Wang, X., 2015. A fractal model for characterizing fluid flow in fractured rock masses based on randomly distributed rock fracture networks. Comput. Geotech. 65, 45–55.
Ma, G., Xu, Z., Zhang, W., Li, S., 2015. An enriched K-means clustering method for grouping joints with meliorated initial centers. Arab. J. Geosci. 8, 1881–1893.
Marcotte, D., Henry, E., 2002. Automatic joint set clustering using a mixture of bivariate normal distributions. Int. J. Rock Mech. Min. Sci. 39 (3), 323–334.
Pusch, R., 1995. Rock Mechanics on a Geological Base. Elsevier, Amsterdam, Netherlands.
Shanley, R.J., Mahtab, M.A., 1976. Delineation and analysis of clusters in orientation data. Math. Geol. 8 (1), 9–23.
Tokhmechi, B., Memarian, H., Moshiri, B., Rasouli, V., Noubari, H.A., 2011. Investigating the validity of conventional joint set clustering methods. Eng. Geol. 118 (3–4), 75–81.
Xie, X., Beni, G., 1991. A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 13 (8), 841–847.
Xu, L., Chen, J., Wang, Q., Zhou, F., 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.
Yamaji, A., Sato, K., 2011. Clustering of joint orientations using a mixed Bingham distribution and its application to paleostress analysis from dike or vein orientations. J. Struct. Geol. 33 (7), 1148–1157.
Zhan, J., Chen, J., Xu, P., Zhang, W., Han, X., Zhou, X., 2017. Automatic identification of rock joint sets using finite mixture models. Math. Geosci. 49 (8), 1021–1056.
Zhou, H., Xie, H., 2003. Direct estimation of the fractal dimensions of a joint surface of rock. Surf. Rev. Lett. 10 (5), 751–762.
Zhou, W., Maerz, N.H., 2002. Implementation of multivariate clustering methods for characterizing discontinuities data from scanlines and oriented boreholes. Comput. Geotech. 28 (7), 827–839.

Tiexin Liu obtained his BSc degree in Water Resources and Hydropower Engineering from Sichuan University, China (2012), and PhD degree in Civil Engineering from the same university (2017). He has worked in the Department of Civil Engineering of Dalian Maritime University since 2017. His work focuses on collection and analysis of geometric parameters of joints.