Rock Mechanics and Rock Engineering (2024) 57:639–653
https://doi.org/10.1007/s00603-023-03572-0

ORIGINAL PAPER

A Comprehensive Method for Similarity Evaluation in Discrete Fracture Network Modeling of Jointed Rock Masses

Jiayao Chen[1] · Hyungjoon Seo[2] · Chengzhan Gao[3] · Qian Fang[1] · Dingli Zhang[1] · Hongwei Huang[4]

Received: 13 May 2023 / Accepted: 10 September 2023 / Published online: 29 September 2023

Abstract

To verify the accuracy and rationality of discrete fracture network modeling, a comprehensive method for evaluating the similarity between two-dimensional fracture maps is proposed in this study. Five evaluation indicators for two-dimensional discrete fracture networks are proposed, including spacing, length, angle, position, and density of rock fractures. Based on the quantitative characteristics of each indicator, methods for evaluating data distribution similarity based on the Wasserstein distance function and matrix similarity based on vector cosine values are proposed. Considering the grouping characteristics of trace lines that are neglected in existing similarity systems, the similarity between each indicator is calculated based on the grouping matching of trace line maps. This achieves the comprehensive calculation of grouping similarity and overall similarity indicators. The importance and size effects of indicators are discussed through fracture model examples. The results show that the indicators of grouping similarity are more rigorous and suitable for engineering modeling with strict requirements compared to those for overall similarity. The proposed method effectively alleviates the difficulty of comprehensive distribution similarity evaluation and achieves a more objective similarity evaluation of position and density indicators. Finally, the established similarity evaluation system is applied to real tunnel construction in Hebei Province, China, providing application value for the accuracy of fracture modeling in practical engineering.

Highlights

- A comprehensive method for evaluating similarity between 2D fracture maps is proposed
- Five indicators are suggested, including spacing, length, angle, position, and density.
- Grouping characteristics of trace lines are considered in similarity calculation.
- Proposed method achieves a more objective evaluation of position and density indicators.

Keywords Rock tunnel · Fracture trace · Similarity analysis · Discrete fracture network

✉ Qian Fang
qfang@bjtu.edu.cn

Jiayao Chen
jychen1@bjtu.edu.cn

1 Key Laboratory for Urban Underground Engineering of Ministry of Education and School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China

2 Department of Civil Engineering and Industrial Design, University of Liverpool, Liverpool, UK

3 College of Civil Engineering, Tongji University, Shanghai, China

4 Key Laboratory of Geotechnical and Underground Engineering of Ministry of Education and Department of Geotechnical Engineering, Tongji University, Shanghai, China

1 Introduction

The rock is divided into blocks of varying sizes by fractures, and the characteristics of the fractures largely determine the deformation and mechanical reaction mechanism of the rock, which has an important influence on the safety and stability of the rock mass (Bieniawski 1967; Hajiabdolmajid and Kaiser 2003). However, due to the frequent alternation of geological and engineering movements, the spatial morphology and distribution of rock fractures are extremely complex (Chen et al. 2022b; Kong et al. 2019; Yuan et al. 2021). Accurate characterization of fracture characteristic parameters, such as the group number, spacing, size, and position, has become a key focus of engineering attention in rationalizing fracture modeling (Rajwa et al. 2019). These parameters affect the discontinuity, heterogeneity, and anisotropy of the rock to varying degrees and also affect the rock mass classification, construction decisions, and optimization of parameter in tunnelling (Fraldi et al. 2019; Moomivand and Vandyousefi 2020). Therefore, the correct and reasonable evaluation of the accuracy and similarity of fracture modeling is of great significance for the safety and stability of rock construction.

With the rapid development of computer vision, the manual sketching method using geological compasses has been gradually replaced by machine vision recognition technology (Azizi and Moomivand 2021; Chen et al. 2021a; Riquelme et al. 2015; Shen et al. 2023; Weidner et al. 2019). The algorithm for feature extraction of rock fractures continues to merge with numerical modeling analysis, providing new ideas for the representation and analysis of real fractures in the field (Man et al. 2023; Zhao et al. 2020). After the International Society for Rock Mechanics proposed a quantitative description of structural surfaces in rock masses, scholars have emerged to reveal the statistical distribution of geometric parameters of rock mass discontinuities (Kong et al. 2021). Therefore, a method based on the stochastic simulation of fracture fractures began to be used to reconstruct rock mass models, namely discrete fracture networks (Fernández et al. 2023; Gottron and Henk 2021; Ma et al. 2019). Obviously, this is a statistical representation of rock fractures. However, the model established by this method does not strictly follow the geometric properties of the real rock fractures but is randomly generated based on the statistical parameters of these geometric properties (Wu et al. 2022). It is obvious that the generated fracture sets have certain randomness and uncertainty. Some scholars believe that discrete fracture networks with randomness and uncertainty can demonstrate the mechanical properties of real rock (Li et al. 2019). However, many scholars also believe that only the selection of suitable fractures can better reflect the mechanical performance of real rock masses (Berrone et al. 2019; Ma et al. 2022; Xiao and He 2022). Therefore, the problem that follows is how to comprehensively evaluate the differences between the model and reality.

The traditional graphical method usually performs only a qualitative analysis of real and simulated traces based on visual observation. However, due to the lack of reliable standards to define similarity, the validation results are often highly subjective (Reinhardt et al. 2022). With the development of digital image processing technology and artificial intelligence, researchers have gradually begun to use machine vision technologies to reveal the similarity of trace maps (Chen et al. 2021b; Xu et al. 2021). Although existing research has conducted extensive studies on the similarity of three-dimensional modeling, it is still difficult to meticulously evaluate the morphological similarity between actual fractures and modeling traces due to the complexity of statistical analysis of actual three-dimensional engineering traces (Cai et al. 2022; Han et al. 2020). On the other hand, this study mainly focuses on the research on the rationality of the fracture network modeling of rock masses, so it summarizes more about the research status of two-dimensional modeling similarity. Currently, common two-dimensional fracture characteristics used by researchers for modeling similarity evaluation include length, angle, position, density, and other common indicators (Battulwar et al. 2021; Han et al. 2018; Masciopinto and Alghalandis 2022; Wang et al. 2020; Wei et al. 2023). The proposal of these indicators provides important guarantees for the refinement of rock mass modeling and enhances the systematic evaluation of rock mass fractures. However, the current research is mainly confronted with the following three problems:

(1) The grouping and spacing of fractures are important indicators that reflect the characteristics of rock mass fractures (Sonmez et al. 2022), but they are rarely listed as indicators for similarity evaluation;
(2) Similarity analysis of individual parameter distribution is often based on comparison of the main control parameters of the distribution curve, while ignoring the distribution itself, which inevitably leads to statistical errors (Alghalandis et al. 2017);
(3) In terms of position and density statistics, relying on common window division areas for statistical analysis may affect similarity due to the subjective selection of window scope and size (Wang et al. 2020). In current research, the corresponding similarity is rarely analyzed from the perspective of global position and density distribution.

To improve the fracture similarity evaluation system, specific improvements were made in this study. To solve the first problem, this study introduces indicators for trace grouping and corresponding spacing to more comprehensively reflect the distribution law of fracture rock masses. Currently, K-means + +, DBSCAN, and other methods are commonly used for automatic grouping of trace angles, while the Silhouette validity index has also been proven to be useful for selecting the optimal group (Battulwar et al. 2021; Chen et al. 2022a). To address the second problem, Wasserstein distance function is introduced to evaluate the similarity of data distributions, replacing the research ideas that use core distribution control parameters to characterize distribution similarity evaluation (Panaretos and Zemel 2019). To address third problem, the study replaces the subjective statistical windowing method with a grid matrix cosine similarity evaluation method by dividing the trace map into grids (Rawat and Bajracharya 2015). This effectively converts local into global evaluations.

Therefore, this study establishes a complete discrete fracture network similarity evaluation system based on machine vision methods and statistical principles. Five two-dimensional indicators are proposed for evaluating discrete fracture networks modelling, including fracture spacing, length, angle, position, and density. Targeted similarity calculation methods are adopted for each indicator, and the importance and size effects of the indicators are discussed in conjunction with fracture model cases. Finally, the established similarity evaluation system is applied to real engineering projects in Hebei Province, China, to verify the practical technical value of this research method.

2 Proposed Method

The evaluation of similarity in 2D discrete fracture networks involves consideration of engineering and geometric indicators that reflect the characteristics of fracture traces. The most commonly used indicators in current research include direction, length, and position. However, several widely used surrounding rock classification systems, such as the Code for Design of Tunnels 2004, the Engineering Rock Mass Classification Standard, and the Rock Mass Rating (RMR) classification standard (Bieniawski 1993; Cai et al. 2023; China 2014; Liu et al. 2019), also emphasize the importance of describing the integrity of rock mass in terms of the spacing and density of structural planes. Thus, five types of similarity indicators were determined in this study, namely direction, length, spacing, position, and density, based on which a comprehensive similarity evaluation model was established. The main process of comprehensive similarity calculation is illustrated in Fig. 1, which involves two sets of methods: the pre-judgment of the existence of the same group between traces, followed by a more precise calculation of the similarity within the same trace group. The final similarity is calculated based on the global distribution characteristics for the trace maps that do not have the same trace group. The grouping of traces is critical for accurately representing the internal structure of discontinuities. Therefore, precise grouping data ensures minimal deviation in fracture modeling.

The proposed method determines the whole or group similarity index by judging whether the two comparative trace maps have the same number of trace groups. Matching occurs when the number of groups in the trace maps is the same, whereupon each group is compared individually. If the number of groups is different, the trace map

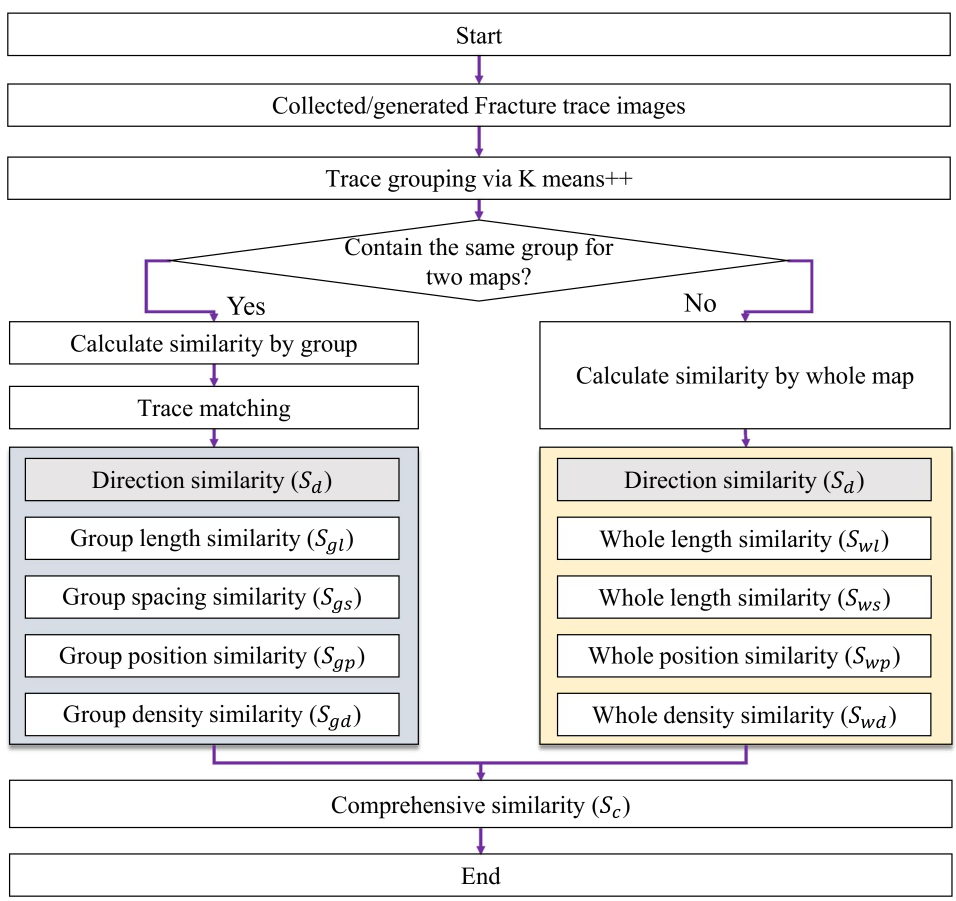


Fig. 1 Calculation flowchart of comprehensive similarity

is compared as a whole. The calculations of the comprehensive group and whole similarity are presented in Eqs. (1–2).

$$S_{gc} = \frac{aS_d + bS_{gl} + cS_{gs} + dS_{gp} + eS_{gd}}{5} \tag{1}$$

$$S_{wc} = \frac{aS_d + bS_{wl} + cS_{ws} + \mathrm{d}S_{wp} + eS_{wd}}{5}, \tag{2}$$

where S_{gc} and S_{wc} are the comprehensive group and whole similarities; S_d is the direction similarity; S_{wl} and S_{gl} represent the whole and group length similarities; S_{ws} and S_{gs} denote the whole and group spacing similarities; S_{wp} and S_{gp} are the whole and group position similarities; S_{wd} and S_{gd} refer to the whole and group density similarities. The weight of each similarity is indicated by $\{a, b, c, d, e\}$, which can be adjusted based on project requirements. Specifically, the definition of rock mass engineering can be determined by incorporating the emphasis of specific geological and hydrological factors, which vary across different engineering projects. It is to be noted that the self-adjustment of parameter weights is not aimed at maximizing or optimizing the overall similarity value, but rather at assessing the significance of each parameter on the engineering process. It is important to ensure that $a + b + c + d + e = 1$. For this study, to simplify the weight setting, all weights were assumed to be equal, i.e., $a = b = c = d = e = 0.2$. Regardless of trace grouping, direction similarity is a universal indicator for similarity evaluation, as evident from Eqs. (1–2). Calculation of the other four indicators can be automatically adjusted based on the trace grouping relationship.

3 Algorithm Implementation

3.1 Trace Grouping and Matching

To accurately depict the trace properties of rock masses in the field, this study builds on previous research by utilizing deep learning to extract trace contours (see Fig. 2). The trace contour skeleton is subsequently extracted, and key nodes of the trace are determined using the chain-code-based algorithm that approximates nodes to multiple segments. This method enables precise and effective extraction of key node coordinates required for trace grouping. The trace interruption algorithm, which is based on an angle threshold, is then employed to separate traces belonging to different groups. Disjoint traces are logically clustered using the improved K-means + + clustering algorithm. The Silhouette validity index (SVI) is utilized to determine the optimal group. These steps comprise the primary process of trace grouping. Further details can be found in Chen et al. (2022a).

To clearly and rigorously describe the process of automatically identifying group relationships and matching subordinate groups, a simple algorithm for group matching is proposed in this study, as shown in Eq. (3):

$$G_m = \min\left\{ \frac{\sum_{i=1}^{i=n} \left| D_{1,i} - D_{2,Ci} \right|}{n} \right\}, \tag{3}$$

where G_m is defined as an indicator to measure the rationality of group matching; $D_{1,i}$ refers to the mean direction of group i in map 1; $D_{1,Ci}$ represents the mean angle of Group C_i in map 2. Specifically, the number of groups in control map 1 is assumed to be $\{1, 2, \ldots, n\}$. Go through the number of map 2 in the way of arrangement and combination, and record it as $\{C_1, C_2, \ldots, C_n\}$. To obtain an optimal matching, the mean direction values of each group in the two maps are calculated based on the arrangement and combination. The basis for selecting the optimal group is the sum of the absolute values of the differences, as the matching combination with the minimum sum of differences is considered to be the best match. It is worth noting that the exhaustive method is used when reprogramming map 2, since the number of trace groups in real engineering projects is often between 3 and 5.

3.2 Direction Similarity

Dip and dip direction are essential factors that affect the stability of the rock mass with 3D fractures. Despite the feasibility of obtaining dip direction parameter through

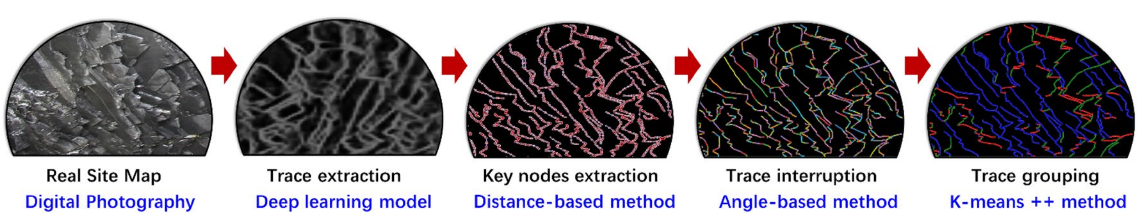


Fig. 2 Diagram from the original map to trace grouping of tunnel face

multi-angle photography and 3D reconstruction of rock mass, their primary application is found in the modeling of 3D DFN. Simultaneously, planar angle-related issues are often manifested due to the directional characteristics of trace lines in 2D fracture networks. Hence, the direction becomes a substitute for these two factors. Currently, research on direction similarity mainly focuses on expressing the distribution with key control parameters (e.g., mean and standard deviation) rather than the distribution form. However, this simplification may weaken the distribution information and even ignore key information. The difficulty of evaluating and quantifying the distribution it may be the potential reason.

To address this issue, a statistical distribution comparison method, Wasserstein distance method (WDM), is proposed to directly compare two distributions and determine their similarity (Panaretos and Zemel 2019). The first-order Wasserstein distance is defined as follows for two given distributions u and v in WDM.

$$l_1(u,v) = \inf_{\pi\in\Gamma(\mathrm{u,v})} \int_{R\times R} |x-y| d\pi(x,y), \tag{4}$$

where $\Gamma(u,v)$ is a set of probability distributions with a size of $R \times R$; $|x–y|$ denotes the change cost from x to y; inf represents the lower limit. Assuming the original distribution as u and the target distribution as v, the values at position x in the original and target distributions are denoted by $u(x)$ and $v(x)$, respectively. Specifically, if $u(x) > v(x)$, the value at x should be moved to another position, whereas if $u(x) \le v(x)$, the value should be moved to x from another position. The formula involves $\mathrm{d}\pi(x,y) = \pi(x,y)\mathrm{d}x\mathrm{d}y$, where $\pi(x,y)$ in the left formula represents the amount of probability density that is moved from x to y.

Furthermore, it should be noted that Eq. (4) entails determining the minimum cost among all possible methods of converting the probability distribution $u(x)$ to $v(x)$. It is important to highlight that the cost of transformation can be evaluated not only using the 1-norm $|x–y|$, but also by the p-norm $x - y_P$. Thus, the p-dimensional distribution of fractures can be formulated as shown in Eq. (5).

$$l_p(u,v) = \left(\inf_{\pi\in\Gamma(\mathrm{u,v})} \int_{R\times R} |x-y|^p d\pi(x,y) \right)^{1/p}. \tag{5}$$

In general, the WDM has the following advantages (Mémoli 2011; Panaretos and Zemel 2019): (1) it can naturally measure the distance between discrete distributions, (2) it can provide a scheme for converting one distribution to another rather than just measuring the distance between them, and (3) it can maintain the geometric properties of the distribution by continuously converting one distribution to another. Using WDM, the similarity of the direction distribution S_d can be defined as in Eq. (6).

$$S_d = \frac{\left|180 - \mathrm{WD}\left(A_1, A_2\right)\right|}{180}, \tag{6}$$

where A_1 and A_2 refer to the direction distributions of the two compared maps; WD denotes the Wasserstein distance calculation of the two distributions. In particular, larger WD values correspond to lower similarity, while smaller values indicate higher similarity. To standardize the comparison, the difference between WD and 180 is divided by 180 to control for the similarity between [0, 1]. Thus, if the two distributions are completely inconsistent, S_d equals 0, while if they are identical, S_d equals 1. The similarity is evaluated using the median value.

To demonstrate the effectiveness of the proposed method, two arbitrary comparison example maps and corresponding distribution statistics results are presented in Fig. 3. From a qualitative perspective, dissimilarities exist between the distribution rules of the blue and red histograms within several ranges, such as [37.5°, 50°], [87.5°, 100°], [125°, 175°], among others. From a quantitative perspective, the WD value and S_d were employed to calculate a final similarity score of 0.423. It is concluded that the distribution of the two maps is not consistent.

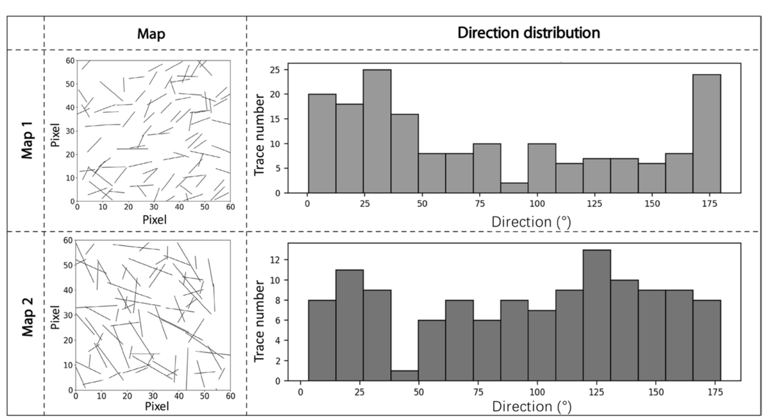


Fig. 3 Two arbitrary trace maps and corresponding direction distributions

3.3 Length Similarity

Similar to the statistics of angle parameters, current research pays more attention to the core variable controlling length, while neglecting the evaluation of the whole length distribution. To counteract this, the WDM method continues to be employed in this section to perform distribution statistics. Furthermore, the description of length similarity is refined by dividing it into group and whole length similarity based on group matching. In other words, if the two maps have the same groups, the group length similarity is calculated directly; otherwise, the whole length similarity is calculated.

(1) Group length similarity

After ensuring that the trace groups are consistent and matching is completed, Eq. (7) is proposed to calculate the grouping length similarity of comparative maps.

$$S_{gl} = \frac{\sum_{i=1}^{n} WD_l\left(L_{1i}, L_{2C_i}\right)}{n}, \tag{7}$$

where S_{gl} refers to the group length similarity; L_{1i}, L_{2C_i} are the length of the matched trace groups in maps 1 and 2 (i.e., group i in map 1 and group C_i in map 2); n denotes the number of trace groups; WD_l is the Wasserstein distance value of length distribution from the remodeling of the value range to [0, 1].

The main purpose of value conversion is to make a reasonable and uniform comprehensive similarity evaluation. The primary purpose of this value conversion is to enable a reasonable and uniform comprehensive similarity evaluation. The specific implementation process is defined in Eq. (8). The difference between the obtained distance and the larger value of the mean of the two distributions is taken and then divided by the mean. The max() function is used in the outermost layer of the formula molecule to ensure that the minimum value is greater than or equal to 0.

$$\mathrm{WD}_l\left(L_{1i}, L_{2Ci}\right) = \frac{\max\left(0, \left(\max\left(\mathrm{mean}\left(L_1\right),\ \mathrm{mean}\left(L_2\right)\right) - \mathrm{WD}\left(L_{1i}, L_{2Ci}\right)\right)\right)}{\max\left(\mathrm{mean}\left(L_1\right),\ \mathrm{mean}\left(L_2\right)\right)}, \tag{8}$$

where WD refers to the Wasserstein distance, more detailed calculation process can be seen in Sect. 3.1. Thus, the definition of interval value of S_{gl} is completed.

(2) Whole length similarity

The whole length similarity S_{wl} is different from S_{gl} calculation, which only needs to combine the global length distribution of the trace map. There are therefore slight differences in S_{wl} calculation, the specific definition is shown in Eq. (9).

$$S_{wl} = \frac{\max\left(0, \left(\max\left(mean\left(L_1\right), mean\left(L_2\right)\right) - DW\left(\left(L_{1i}, L_{2i}\right)\right)\right)\right)}{\max\left(mean\left(L_1\right), mean\left(L_2\right)\right)}, \tag{9}$$

where L_{1i}, L_{2i} is the length distribution of the fracture traces in the two contrast maps. Similar to Eq. (8), value conversion to [0, 1] is also applied in Eq. (9). In addition, since the method of calculating the similarity based on the global trace length is considered, there is no need to carry out the averaging operation.

(3) Comparison

To further analyze the performance of S_{gl} and S_{wl}, two arbitrary maps with the same trace groups are used for comparative analysis. Figure 4 shows the group and whole length distributions counted from two comparative maps. Then, the WDM method is used to calculate the values

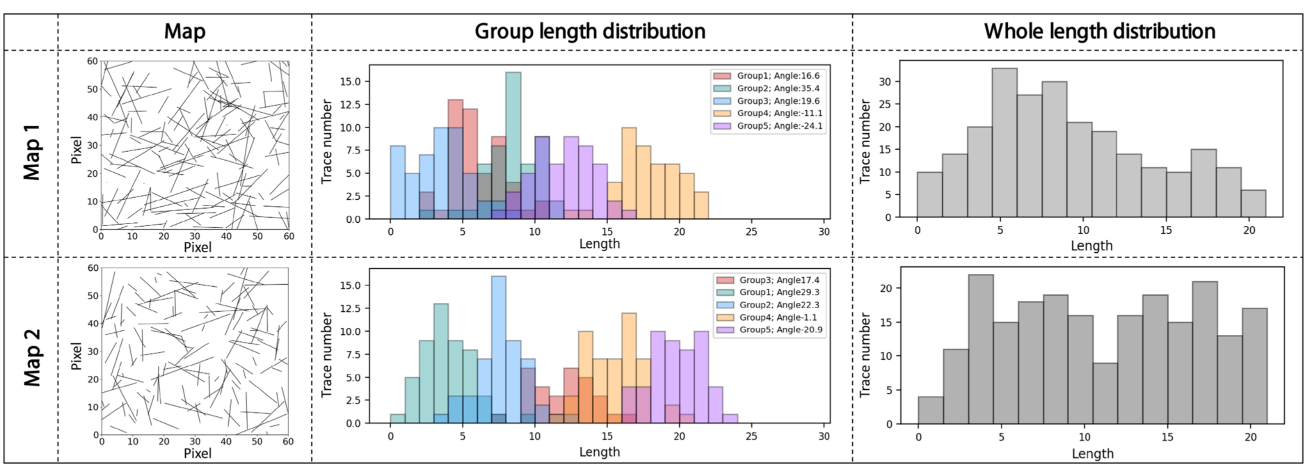


Fig. 4 Two arbitrary trace maps and corresponding length distributions

of S_{gl} and S_{wl}. The results show that the group and whole length similarity are 41.3% and 72.0%, respectively. It is seen that the statistical method of S_{gl} is strict than that of S_{wl}. Meanwhile, although the whole length distribution of the two maps is similar qualitatively, the distribution differences of each group are obvious after considering grouping. This also caused significant fluctuations in the length distribution within the same group.

3.4 Spacing Similarity

(1) Group spacing similarity

In the previous study, the spacing similarity has not yet emerged as a main indicator in the similarity evaluation system. Due to the importance of rock mass structure, spacing will be included in the similarity evaluation system for the first time. In this context, spacing is defined as the distance between adjacent individual trace lines. Furthermore, depending on the distinct adjacent attributes (whether they pertain to the same group or differ), spacing is subsequently classified into inter-group spacing and overall spacing. Herein, the division principles of group and whole similarity continue to be used. However, in a small range of rock mass area such as tunnel face, the spacing of each trace group does not differ greatly in quantity. Thus, the average value is directly applied instead of the distribution as the main feature to characterize the set spacing. On this basis, the mean value of each group spacing is executed according to the definition of Eq. (10), and the group spacing similarity S_{gs} is obtained.

$$S_{gs} = \frac{\sum_{i=1}^{n} \min\left(\frac{S_{1i}}{S_{2C_i}}, \frac{S_{2C_i}}{S_{1i}}\right)}{n}, \tag{10}$$

where S_{1i}, S_{2C_i} denote the mean spacings of the matched trace group in maps 1 and 2 (i.e., group i in map 1 and group C_i in map 2); n is the number of trace groups.

In Eq. (10), the smaller value of the ratio between S_{1i} and S_{2C_i} is directly used as the similarity of a single matched group to ensure that a [0, 1] value range. Then, sum and average the similarity of each group to achieve the calculation of group similarity within a reasonable range.

(2) Whole spacing similarity

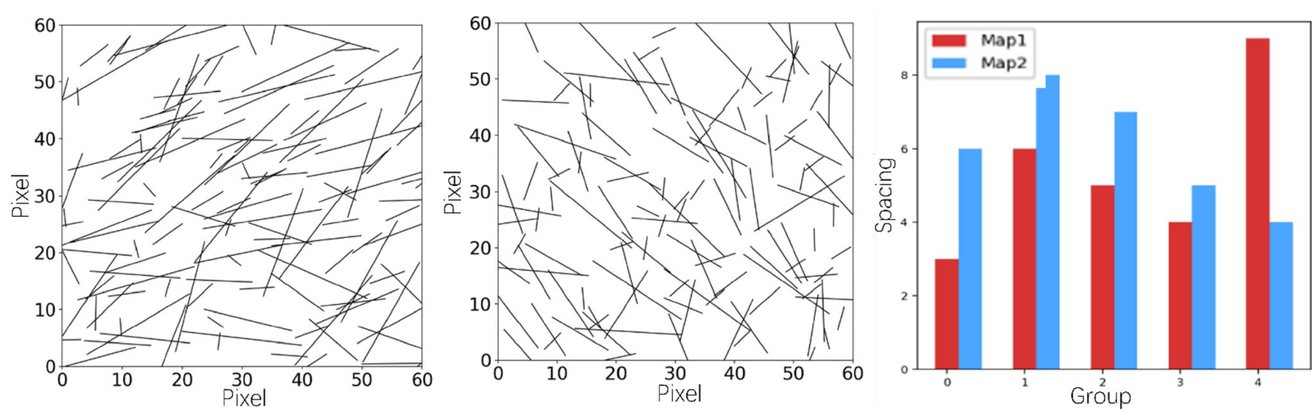
Similar to group spacing similarity, the average value of spacing is also used to characterize the spacing characteristics when calculating the whole spacing similarity. Since the number of groups in the comparison chart is different when calculating the overall spacing, the method of summing and averaging the average spacing of each group is used to calculate the spacing of the entire trace map. Finally, the smaller ratio of the two is used to ensure the value is less than 1. The specific definition is shown in Eq. (11).

$$\mathrm{S}_{ws} = \min\left\{ \frac{(\sum_{i=1}^{i=n} S_{1i})/n}{(\sum_{i=1}^{i=m} S_{2i})/m}, \frac{(\sum_{i=1}^{i=m} S_{2i})/m}{(\sum_{i=1}^{i=n} S_{1i})/n} \right\}, \tag{11}$$

where S_{ws} is whole spacing similarity; S_{1i}, S_{2i} denote the mean spacing of ith trace group in maps 1 and 2; n, m is the group number of two maps, respectively.

(3) Comparison

To compare the two similarity evaluation schemes, two maps with the same group number are selected for analysis (see in Fig. 5). Unlike the previous indicators, the number of groups here must be consistent to ensure complete correspondence between groups in the evaluation of S_{gs}. On the basis, the trace spacing distribution of the five groups is statistically analyzed, and Fig. 5c is obtained. According to



(a) Generated map 1 (b) Generated map 2 (c) Mean spacing distributions

Fig. 5 Two arbitrary trace maps and corresponding mean spacing distributions

Eqs. , the values of S_{gs}. and S_{ws} for the total average spacing in Fig. 5c are 0.642 and 0.9, respectively. It is seen that although the group is consistent, S_{gs} 's evaluation is much stricter than that of S_{ws}. The main reason is that the fluctuation of each group is ignored in the S_{ws} calculation, and only the change of the total average spacing is emphasized, largely ignoring the change in the spacing of each group is greatly. Therefore, the solution of each group's spacing has a great impact on the final spacing similarity.

3.5 Position Similarity

To comprehensively characterize the position similarity, a novel method is proposed for evaluating the position information. As shown in Fig. 6, the proposed method includes the following steps: (1) generating a discrete fracture mesh; (2) puncturing each continuous fracture column (line segment), i.e., replacing each line segment with a unit length by a point with a certain distance interval; (3) dividing the original complete fracture column area into grids; (4) calculating the centroid for the points in each area and recording the information in area (i, j) as centroid coordinates (x_{ij}, y_{ij}).

Solving the similarity of the centroids is undoubtedly cumbersome in this study, so all the centroids of the grids are combined and then determined in a matrix. The barycentric coordinates of each grid are listed in Eq. (12). They are defined by calculating the average value of the product of the barycenter coordinates and trace length in each grid.

$$\begin{cases} x_{ij} = \frac{\sum_{k=1}^{t} x_{ij,k} \times l_{ij,k}}{t} \\ y_{ij} = \frac{\sum_{k=1}^{t} y_{ij,k} \times l_{ij,k}}{t} \end{cases} \tag{12}$$

where $x_{ij,k}$, $y_{ij,k}$, $l_{ij,k}$ refer to the abscissa center, the ordinate, and the length of the kth lines selected in the ith row and jth column regions, respectively; t is the number of lines selected in corresponding regions.

According to the above calculation definition, the whole trace region is assumed to be divided into $n \times m$. Then, the barycentric coordinates $\left(x_{ij}, y_{ij}\right)$ are calculated for all points in each grid. Two matrices of the contrast curve are thus obtained from the compared maps. As shown in Eq. (13), the abscissa of the centroid of the area represented by each element in the matrix x_{ij} and ordinate y_{ij} two information.

$$\begin{bmatrix} \left(x_{11}, y_{11}\right) & \cdots & \left(x_{1k}, y_{1k}\right) \\ \vdots & \ddots & \vdots \\ \left(x_{h1}, y_{h1}\right) & \cdots & \left(x_{hk}, y_{hk}\right) \end{bmatrix}. \tag{13}$$

Through the above way, an $m \times n$ matrix composed of each grid centroid is formed. However, due to the complexity and disorder of matrices, it is not easy to evaluate their similarity. It is well known that the cosine value (defined in Eq. (14)) is commonly used to characterize the similarity between vectors (Dong et al. 2006; Ye 2011).

$$\cos\theta = \frac{\sum_{i=1}^{k} x_i y_i}{\sqrt{\sum_{i=1}^{k} \left(x_i\right)^2} \sqrt{\sum_{i=1}^{k} \left(y_i\right)^2}}, \tag{14}$$

where the cosine similarity $\cos\theta$ is a value with a range of $[-1,1]$; x_i, y_i refer to the ith values of two arbitrarily vectors of length k, i.e., $[x_1, x_2, x_3, \ldots x_k]$, $\left[y_1, y_2, y_3, \ldots y_k\right]$. Undoubtedly. If vectors of x_i and y_i have the same or opposite direction, the cosine similarity is 1 or -1. If they are perpendicular to each other, the result is 0.

Moreover, a matrix can be considered as a set of vectors. Inspired by vector similarity calculation, a matrix cosine calculation method is therefore proposed to characterize the similarity (Fouss et al. 2007). The definition is shown in Eq. (15).

$$\cos(A, B) = \frac{\overrightarrow{A_{h\times k}} \cdot \overrightarrow{B_{h\times k}}}{\left|\overrightarrow{A_{h\times k}}\right| \cdot \left|\overrightarrow{B_{h\times k}}\right|}, \tag{15}$$

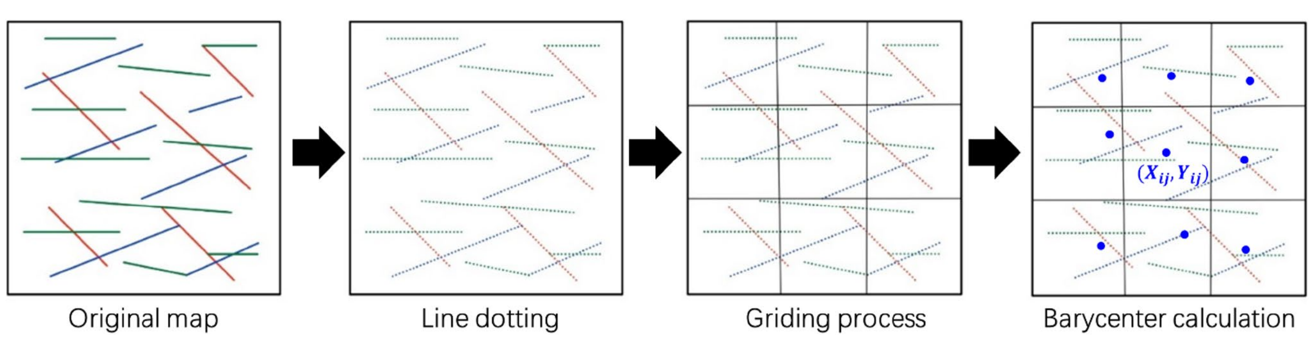


Fig. 6 Flowchart of position similarity evaluation

where cos(A, B) is the cosine similarity of matrices; $\overrightarrow{A_{h\times k}}$, $\overrightarrow{B_{h\times k}}$ refer to the centroid matrices of $h \times k$ generated from the trace maps. To keep consistent with the previous indicators, Eq. (16) was used to convert the values on the original [− 1, 1] interval to the interval [0, 1].

$$\cos(A,B)' = 0.5 + 0.5 \times \cos(A,B). \tag{16}$$

Based on the above cosine similarity method, the definitions of group and whole position similarity can be deduced as follows:

$$S_{\mathrm{gp}} = \frac{\sum_{i=1}^{i=n} \cos_i(A,B)'}{n} \tag{17}$$

$$S_{\mathrm{wp}} = \cos(A,B)', \tag{18}$$

where S_{gp} and S_{ws} represent the group and whole position similarity. It is found from Eqs. (17–18) that S_{gp} is the average value of the centroid matrix similarity of each group of traces, while S_{ws} donates directly the value of the centroid matrix similarity.

To compare S_{gp} and S_{ws}, two traces with the same number of groups are randomly selected to count the corresponding centroid coordinates and matrices. The visualization results are shown in Fig. 7, from which the values of S_{gp} and S_{ws} are calculated to be 0.584 and 0.827, respectively. The same conclusion can be thus drawn that the S_{gp} index is more rigorous and complex than S_{ws}. Also, the position similarity characteristics of each group have a great impact on the ultimate S_{gp} value. It is concluded that S_{gp} are more applicable to projects with strict modeling requirements.

3.6 Density Similarity

In the process of density similarity calculation, the method of "dotting and gridding" (see in Fig. 6) is also used here, but the information recorded in each grid has been changed from centroid coordinates to point numbers. That is, the operation of step 4 in Fig. 6 is replaced by counting the points in each grid. Similar to the barycentric matrix in the previous section, the trace map is also divided into $m \times n$ by gridding. Then, the points in each grid (i.e., ith row and jth column) are counted to obtain num_{ij}. Traversing the entire map in this way yields the quantity matrix, as shown in Eq. (19).

$$\begin{bmatrix} \mathrm{num}_{11} & \cdots & \mathrm{num}_{1k} \\ \vdots & \ddots & \vdots \\ \mathrm{num}_{h1} & \cdots & \mathrm{num}_{hk} \end{bmatrix}. \tag{19}$$

It is worth noting that the original definition of trace density is the length of traces per unit area. Since the trace width is 1 pixel in this study, the number of trace points is used to reflect the length in an equal proportion. Therefore, the density similarity can be obtained by comparing and averaging the point number of all grids in the two maps. The specific definition is shown in Eq. (20).

$$S_d = \frac{\sum_{j=1}^{j=k} \sum_{i=1}^{i=h} \min\left\{ \frac{\mathrm{num}_{2,hk}}{\mathrm{num}_{1,hk}}, \frac{\mathrm{num}_{1,hk}}{\mathrm{num}_{2,hk}} \right\}}{h \times k}, \tag{20}$$

where S_d is the density similarity, $\mathrm{num}_{1,hk}$, $\mathrm{num}_{2,hk}$ are the num_{hk} values of two compared trace maps. h, k refer to gridded rows and columns in maps.

The group and whole density similarity indexes have different calculation manners due to the different definitions. That is, the group density similarity S_{gd} is the summation and averaging of S_d values of each trace group, while the whole density

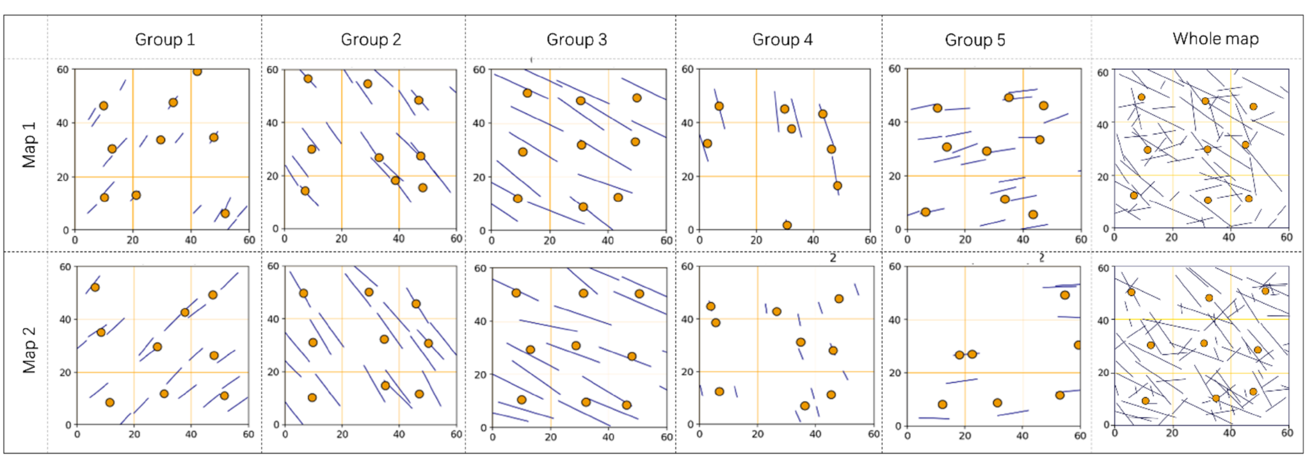


Fig. 7 Two arbitrary trace maps and corresponding group and whole position distributions (unit: pixel)

similarity S_{wd} is calculated by all traces, and the number is equal to S_d. The definitions are listed in Eqs. (21–22).

$$S_{gd} = \sum_{i=1}^{n} S_{di} \tag{21}$$

$$S_{wd} = S_d, \tag{22}$$

where S_{di} is the density similarity S_d of ith trace group, n represents the total group number in maps.

To compare S_{gd} and S_{wd}, two traces with the same number of groups are randomly selected to count the corresponding centroid coordinates and matrices. The visualization results are shown in Fig. 8, from which the values of S_{gd} and S_{wd} are calculated to be 0.173 and 0.428, respectively. The same conclusion can be thus drawn that the S_{gd} index is more rigorous than S_{wd}. Also, the density similarity characteristics of each group have a great impact on the ultimate S_{gd} value. Specifically, in the process of grouping similarity calculation, although the directions of each group are similar, significant differences in density similarity statistics arise due to differences in the position and number of traces in each group. This is also the direct reason why S_{gd} is more difficult to achieve satisfactory results than S_{wd} indicator.

4 Applications and Discussion

4.1 Engineering Application

In this study, an engineering case of a tunnel project under construction will be utilized to verify the effectiveness of the proposed similarity evaluation method. The Yangjiawopu Tunnel is located at the Lijiaying junction of the Changjiang-Shenzhen Highway (G25) in the Chengping section of the Ring Road in the Capital Region (see in Fig. 9a). It spans a total length of 5054 m, passing through Yingshouyingzi Mining Area and Xinglong County in the west, and reaching Pinggu district, Beijing in the south. The starting and ending pile numbers are K52 + 226 and K57 + 280 (see in Fig. 9b).

To obtain the discontinuity occurrences, a photography-based method is used to collect tunnel face images at K53 + 510-K54 + 140 on the right line. Feature extractions are performed via a FraSegNet-based deep learning method (Chen et al. 2021b). To verify the effectiveness of similarity evaluation, a tunnel section located at K53 + 810 is selected (see in Fig. 9c). The occurrence of the applied tunnel face is shown in Table 1, using the rock mass occurrence solution method developed by our research group. On this basis, the Discrete Fracture Network (DFN) method is employed to construct a new network to simulate the real structure in the field. The main purpose of this step is to verify the effectiveness of the proposed similarity model. As it is not possible to define the functional information of trace position and density in the original map, a uniform distribution is used as the main basis for modeling.

To assess the modeling effectiveness of the DFN and the similarity assessment method, five fracture networks depicted in Fig. 10 were generated using the provided occurrence distributions. Qualitative analysis of the overall occurrence distribution indicates a high level of modeling consistency across the compared maps. Subsequently, engineers may base their analysis on these results, without considering the occurrence similarity of the DFN models.

The proposed similarity evaluation method was employed to calculate the similarities between the generated trace maps and the real map, which are presented in Table 2. The mean similarity values for direction and length are notably higher, reaching 68.1 and 73.1, respectively, compared to the other three indicators. This could be attributed to the

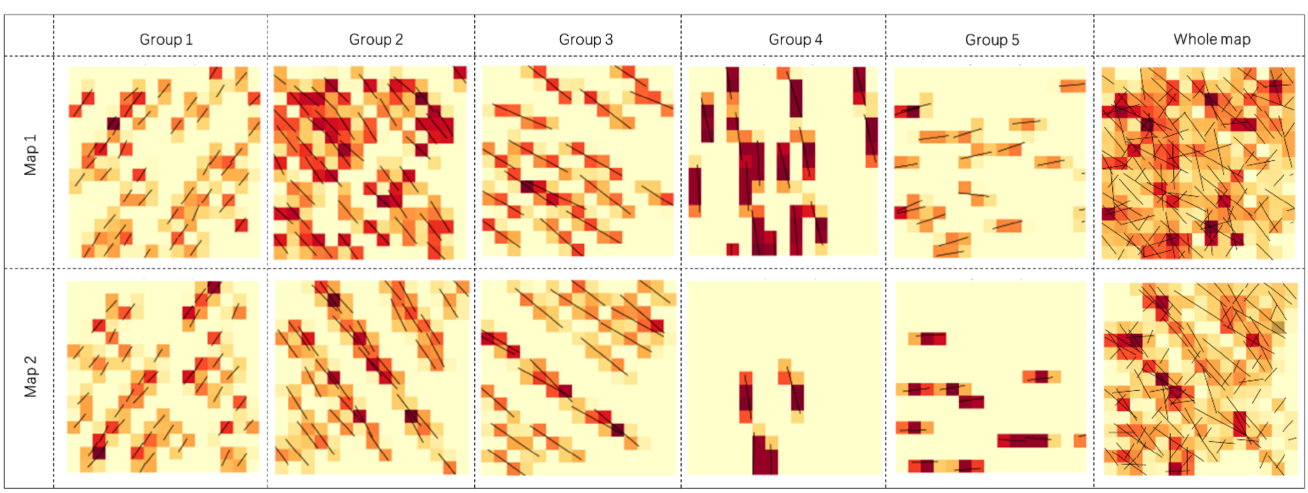


Fig. 8 Two arbitrary trace maps and corresponding mean spacing distributions

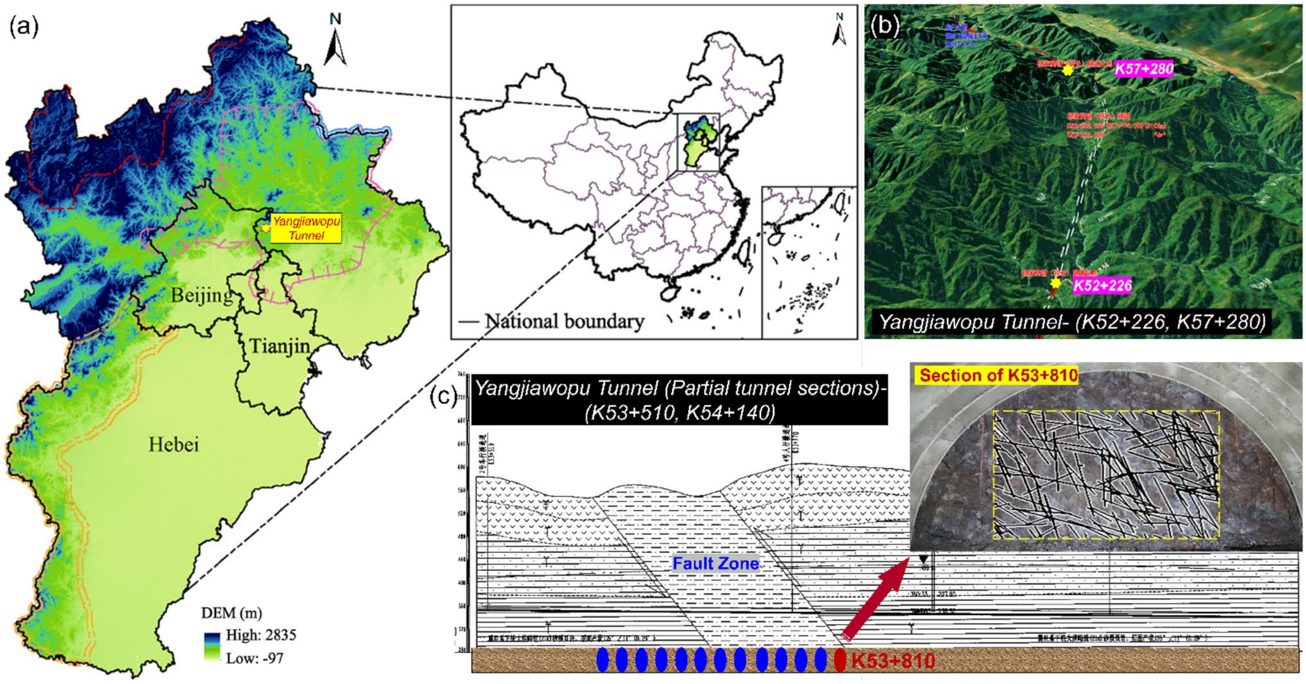


Fig. 9 Basic information of tunnel site, including **a** geographic location, **b** satellite map, **c** selected tunnel face and trace map

Table 1 Trace distribution features based on field acquisition and feature extraction

Group NO	Direction		Length		Spacing		Trace number
	Mean (°)	Fisher constant	Gamma coefficient K	Gamma coefficient θ	Mean (m)	Standard deviation	
1	162	3.79	3.7	0.8	0.45	0.04	47
2	36	2.94	2.8	1.5	0.61	0.08	18
3	118	5.11	4.3	1.1	0.33	0.10	74

DFN technology's ability to more accurately depict trace distributions, with length and direction directly influencing map generation. However, the assumption of uniform distribution for both position and density during DFN modeling leads to maps that deviate significantly from the real map, resulting in higher standard deviations for these two indicators. Notably, despite both density and position being based on grid statistics, they do not exhibit a positive or negative correlation in each graph, indicating that they are independent indicators that do not affect the final similarity statistical results. Generated map 5 exhibits a slightly superior comprehensive similarity value of 60, although this difference is difficult to discern quantitatively. Furthermore, the systematic similarity indexes provide valuable optimization directions for subsequent modeling.

4.2 Discussion

(1) Indicator contribution

Five similarity evaluation indicators were utilized in this study to determine the comprehensive similarity, with each index having varying effects on the final result. To highlight the importance of each indicator on the comprehensive indices, the absence of each indicator was investigated, and the detailed changes in the overall similarity are presented in Fig. 11. The results demonstrate that the length similarity has the greatest impact on the overall similarity, followed by direction and position similarities. Eliminating both direction and length indicators results in decreased overall similarity, while density and position similarity are stringent

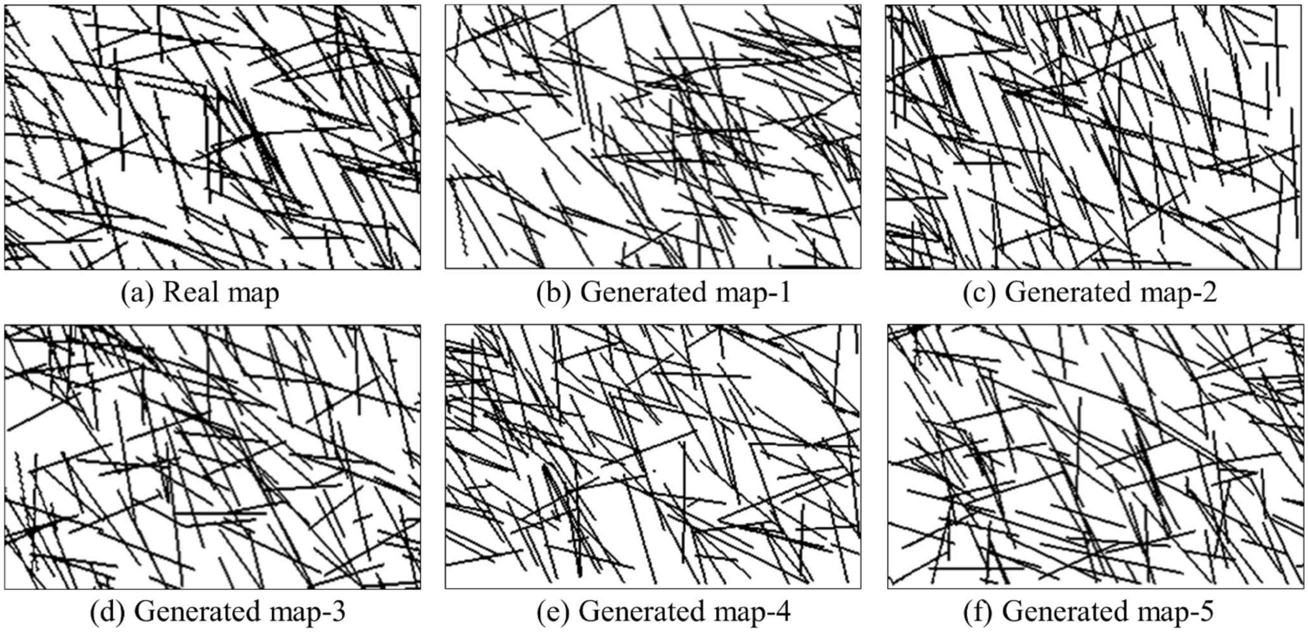


Fig. 10 On site statistical and correspondingly generated fracture trace maps

Table 2 Similarity statistics between original and generated maps

Generated maps	Similarity value (%)					
	Direction	Length	Spacing	Position	Density	Comprehensive
Map 1	67.5	72.6	49.1	34.4	50.6	54.8
Map 2	64.8	78.4	56.6	43.5	32.1	55.1
Map 3	69.6	65.4	47.8	48.9	38.3	54.0
Map 4	66.1	77.9	43.8	31.1	35.5	50.9
Map 5	72.3	73.1	55.6	44.1	55.9	60.2
Mean	68.1	73.5	50.6	40.4	42.5	
Std	2.7	4.7	4.8	6.6	9.2	–

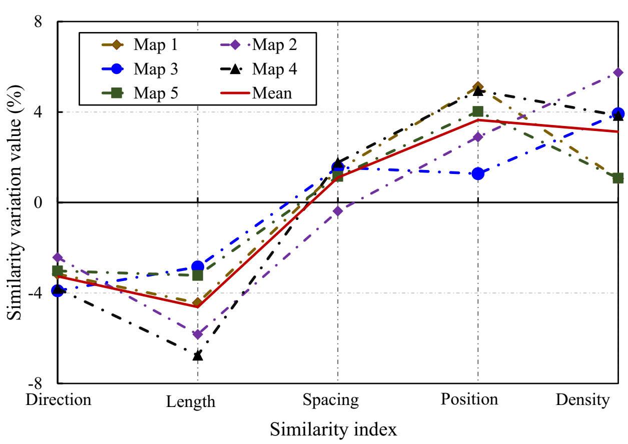


Fig. 11 Change of whole similarity index after removing one index

indicators that may improve the comprehensive similarity when removed. Therefore, it is concluded that increasing the similarity indicators of density and position can significantly enhance refined modeling. The spacing similarity is considered a compromise index that has the least influence on the final evaluation index. It is worth noting that this analysis is naturally random and only applicable to one typical case; however, it provides an improvement direction for refining the current method for modeling discrete fracture networks.

(2) Grid size effect

In the previous section, it was established that optimizing the comprehensive performance of position and density indicators can lead to further improvement in overall similarity.

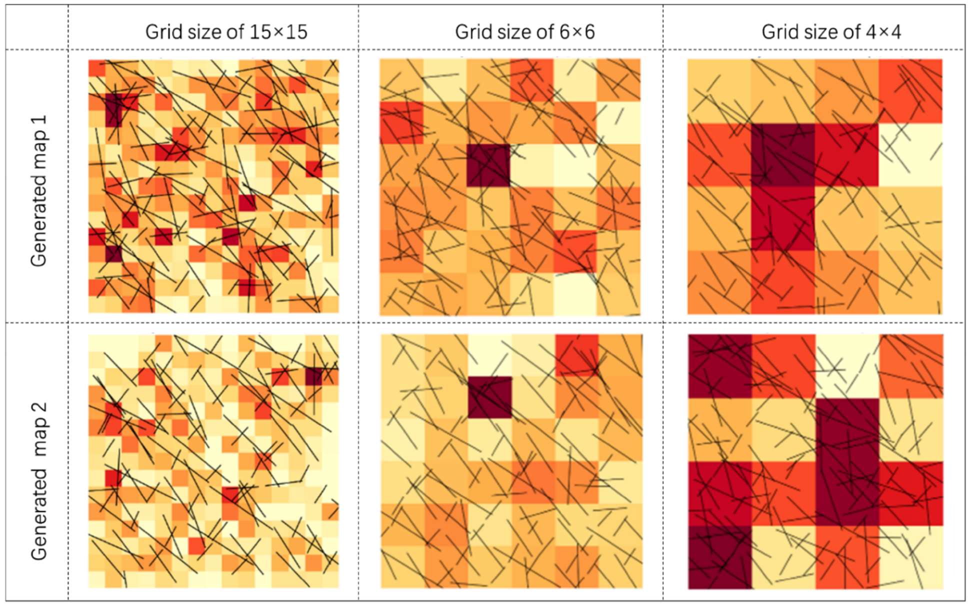


Fig. 12 Change of whole similarity index after removing one index

Table 3 Statistical density similarities of group and whole under different grid sizes

Grid size	Group density similarity						Whole density similarity
	Group 1	Group 2	Group 3	Group 4	Group 5	Mean	
4×4	0.669	0.743	0.68	0.089	0.435	0.5232	0.778
6×6	0.436	0.662	0.389	0.086	0.123	0.3392	0.669
15×15	0.085	0.191	0.167	0.014	0.011	0.0936	0.42

The calculation process for both indicators involves grid-based statistical evaluation methods. However, the size effect of the mesh may inevitably increase the similarity. Therefore, this study discusses the changes in corresponding density similarity caused by grid changes using examples. A random trace map with a size of 60×60 was generated by combining the trace distribution in Fig. 8 and was divided into grid sizes of 15×15, 6×6, and 4×4. The density grid diagrams for different levels are presented in Fig. 12, utilizing the density grid visualization method outlined in Sect. 3.6.

Based on this, the corresponding similarity values under each grid size are calculated and presented in Table 3. It was observed that group density similarity was generally lower than the overall density similarity, indicating that the group density similarity is more stringent. Additionally, finer subdivision of smaller grids provides more detailed statistics, making it more suitable for common trace maps with high similarity. Therefore, the finer the mesh subdivision, the lower the comprehensive density similarity value, while coarser subdivision leads to higher similarity values. In conclusion, the statistical value of similarity can adjust the severity of the similarity score by adjusting the fineness of the mesh. It is worth noting that the primary focus of this study lies on the grid effect of tunnel face site dimensions. For other site dimensions, appropriate actions should be taken based on the engineering control threshold or criteria for modeling similarity. In cases where a high similarity threshold is mandated, an increase in the grid cell size is warranted. Conversely, when a low similarity threshold is established, a reduction in the grid cell size is recommended to minimize similarity calculations.

5 Conclusions

This paper proposes a comprehensive similarity evaluation method for assessing the effectiveness of rock fracture trace map generation. Both group and whole similarity calculations are considered by raising five indicators: direction, length, spacing, position, and density similarities. To demonstrate the rationality and effectiveness of the proposed method, a typical engineering application is performed. The importance of indicators and grid size effect are also comprehensively analyzed to systematically discuss the future similarity improvement direction. The main findings are presented as follows:

(1) The proposed method considers fracture grouping that is often ignored during the diagenetic process of structural planes. Based on the grouping information, the comprehensive similarity is divided into group and whole similarity indicators. It is found that the total similarity calculated by grouping indicators is more rigorous than the overall calculation. Group and whole similarity are more suitable for projects with high requirements for fracture modelling and conventional engineering, respectively.
(2) The method employs dotting and gridding operations on the trace to calculate position and density similarity, converting the two abstract indicators into concrete parameters within the grid. Cosine similarity calculation and grid correspondence comparative research are conducted on the statistically obtained parameter matrix, position and density similarities are obtained, respectively. The research indicates that the fracture point segmentation combined with grid similarity statistics can provide an important idea for optimizing the modelling position.
(3) Combining the importance of indicators and grid scale effects, it is found that the contribution of inclination and length to the comprehensive similarity index is significantly greater than that of density and position due to their high relative values. Moreover, coarsening the mesh can help improve the indicator values of density and position. Engineers can use different statistical dimensions for sites of different engineering levels based on their own needs.

Although the proposed method adds the important impact of trace grouping and spacing to the existing trace similarity evaluation, a new position and density evaluation method based on grid matrix is also proposed. However, the practice of averaging five indicators to obtain comprehensive indicators remains to be discussed. In the future, it is possible to deeply explore the impact of various indicators on the properties of rock masses and redefine the importance based on this rule.

Acknowledgements The research presented in this paper is supported by Fundamental Research Funds for the Central Universities (No. 2023JBRC005) and Natural Science Foundation Committee Program of China (Grant No. 52308288).

Data availability The data that support the findings of this study are available from Dr. Jiayao Chen upon reasonable request.

References

Alghalandis YF, Elmo D, Eberhardt E (2017) Similarity analysis of discrete fracture networks. arXiv preprint. arXiv:1711.05257. Accessed 28 Sep 2023

Azizi A, Moomivand H (2021) A new approach to represent impact of discontinuity spacing and rock mass description on the median fragment size of blasted rocks using image analysis of rock mass. Rock Mech Rock Eng 54:2013–2038

Battulwar R, Zare-Naghadehi M, Emami E, Sattarvand J (2021) A state-of-the-art review of automated extraction of rock mass discontinuity characteristics using three-dimensional surface models. J Rock Mech Geotech Eng 13:920–936

Berrone S, Borio A, Vicini F (2019) Reliable a posteriori mesh adaptivity in discrete fracture network flow simulations. Comput Methods Appl Mech Eng 354:904–931

Bieniawski Z (1967) Stability concept of brittle fracture propagation in rock. Eng Geol 2:149–162

Bieniawski Z (1993) Classification of rock masses for engineering: the RMR system and future trends, rock testing and site characterization. Elsevier, New York, pp 553–573

Cai W, Zhu H, Liang W (2022) Three-dimensional tunnel face extrusion and reinforcement effects of underground excavations in deep rock masses. Int J Rock Mech Min Sci 150:104999

Cai W, Zhu H, Liang W, Wang X, Su C, Wei X (2023) A post-peak dilatancy model for soft rock and its application in deep tunnel excavation. J Rock Mech Geotech Eng 15:683–701

Chen J, Chen Y, Cohn AG, Huang H, Man J, Wei L (2022a) A novel image-based approach for interactive characterization of rock fracture spacing in a tunnel face. J Rock Mech Geotech Eng 14:1077–1088

Chen J, Huang H, Cohn AG, Zhang D, Zhou M (2022b) Machine learning-based classification of rock discontinuity trace: SMOTE oversampling integrated with GBT ensemble learning. Int J Min Sci Technol 32:309–322

Chen J, Yang T, Zhang D, Huang H, Tian Y (2021a) Deep learning based classification of rock structure of tunnel face. Geosci Front 12:395–404

Chen J, Zhou M, Huang H, Zhang D, Peng Z (2021b) Automated extraction and evaluation of fracture trace maps from rock tunnel face images via deep learning. Int J Rock Mech Min Sci 142:104745

China, N.S.C.G.o.P.s.R.o. (2014) TB 10121–2007 technical code for monitoring measurement of railway tunnel. China Planning Press, Beijing

Dong Y, Sun Z, Jia H (2006) A cosine similarity-based negative selection algorithm for time series novelty detection. Mech Syst Signal Process 20:1461–1472

Fernández A, Sanchidrián JA, Segarra P, Gómez S, Li E, Navarro R (2023) Rock mass structural recognition from drill monitoring technology in underground mining using discontinuity index and machine learning techniques. Int J Min Sci Technol 33(5):555–571

Fouss F, Pirotte A, Renders J-M, Saerens M (2007) Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. IEEE Trans Knowl Data Eng 19:355–369

Fraldi M, Cavuoto R, Cutolo A, Guarracino F (2019) Stability of tunnels according to depth and variability of rock mass parameters. Int J Rock Mech Min Sci 119:222–229

Gottron D, Henk A (2021) Upscaling of fractured rock mass properties–An example comparing discrete fracture network (DFN) modeling and empirical relations based on engineering rock mass classifications. Eng Geol 294:106382

Hajiabdolmajid V, Kaiser P (2003) Brittleness of rock and stability assessment in hard rock tunneling. Tunn Undergr Space Technol 18:35–48

Han S, Wang G, Li M (2018) A trace map comparison algorithm for the discrete fracture network models of rock masses. Comput Geosci 115:31–41

Han S, Li M, Wang G (2020) Copula-based simulating and analyzing methods of rock mass fractures. Comput Geotech 127:103779
Kong P, Jiang L, Shu J, Sainoki A, Wang Q (2019) Effect of fracture heterogeneity on rock mass stability in a highly heterogeneous underground roadway. Rock Mech Rock Eng 52:4547–4564
Kong D, Saroglou C, Wu F, Sha P, Li B (2021) Development and application of UAV-SfM photogrammetry for quantitative characterization of rock mass discontinuities. Int J Rock Mech Min Sci 141:104729
Li L, Jiang H, Wu K, Li J, Chen Z (2019) An analysis of tracer flowback profiles to reduce uncertainty in fracture-network geometries. J Petrol Sci Eng 173:246–257
Liu K, Liu B, Fang Y (2019) An intelligent model based on statistical learning theory for engineering rock mass classification. Bull Eng Geol Env 78:4533–4548
Ma G, Li T, Wang Y, Chen Y (2019) The equivalent discrete fracture networks based on the correlation index in highly fractured rock masses. Eng Geol 260:105228
Ma T, Liu J, Fu J, Wu B (2022) Drilling and completion technologies of coalbed methane exploitation: an overview. Int J Coal Sci Technol 9:68
Man J, Huang H, Ai Z, Chen J, Wang F (2023) Stability of complex rock tunnel face under seepage flow conditions using a novel equivalent analytical model. Int J Rock Mech Min Sci 170:105427
Masciopinto C, Alghalandis YF (2022) A new modeling approach for advective and dispersive pollutant transport in 3D discrete fracture network backbones of heterogeneous aquifers. Authorea Preprints.
Mémoli F (2011) A spectral notion of Gromov-Wasserstein distance and related methods. Appl Comput Harmon Anal 30:363–401
Moomivand H, Vandyousefi H (2020) Development of a new empirical fragmentation model using rock mass properties, blasthole parameters, and powder factor. Arab J Geosci 13:1–17
Panaretos VM, Zemel Y (2019) Statistical aspects of Wasserstein distances. Annu Rev Stat Appl 6:405–431
Rajwa S, Janoszek T, Prusek S (2019) Influence of canopy ratio of powered roof support on longwall working stability–a case study. Int J Min Sci Technol 29:591–598
Rawat DB, Bajracharya C (2015) Detection of false data injection attacks in smart grid communication systems. IEEE Signal Process Lett 22:1652–1656
Reinhardt M, Jacob A, Sadeghnejad S, Cappuccio F, Arnold P, Frank S, Enzmann F, Kersten M (2022) Benchmarking conventional and machine learning segmentation techniques for digital rock physics analysis of fractured rocks. Environ Earth Sci 81:71
Riquelme AJ, Abellán A, Tomás R (2015) Discontinuity spacing analysis in rock masses using 3D point clouds. Eng Geol 195:185–195
Shen Y, Zhang D, Wang R, Li J, Huang Z (2023) SBD-K-medoids-based long-term settlement analysis of shield tunnel. Transp Geotech. 42:101053
Sonmez H, Ercanoglu M, Dagdelenler G (2022) A novel approach to structural anisotropy classification for jointed rock masses using theoretical rock quality designation formulation adjusted to joint spacing. J Rock Mech Geotech Eng 14:329–345
Wang J, Zheng J, Liu T, Guo J, Lü Q (2020) A comprehensive dissimilarity method of modeling accuracy evaluation for discontinuity disc models based on the sampling window. Comput Geotech 119:103381
Wei X, Zhang L, Gardoni P, Chen Y, Tan L, Liu D, Du C, Li H (2023) Comparison of hybrid data-driven and physical models for landslide susceptibility mapping at regional scales. Acta Geotech 18(8):4453–4476
Weidner L, Walton G, Kromer R (2019) Classification methods for point clouds in rock slope monitoring: a novel machine learning approach and comparative analysis. Eng Geol 263:105326
Wu N, Liang Z, Zhang Z, Li S, Lang Y (2022) Development and verification of three-dimensional equivalent discrete fracture network modelling based on the finite element method. Eng Geol 306:106759
Xiao H, He L (2022) Implementation of manifold coverage for 3D rock fracture network modeling and its application in rock permeability prediction. Comput Geotech 145:104702
Xu W, Zhang Y, Li X, Wang X, Liu R, Zhao P, Zhang Y, Dai J (2021) Comprehensive identification of statistical homogeneity of fractured rock masses for a candidate HLW repository site. China Eng Geol 293:106279
Ye J (2011) Cosine similarity measures for intuitionistic fuzzy sets and their applications. Math Comput Model 53:91–97
Yuan Y, Xu T, Heap MJ, Meredith PG, Yang T, Zhou G (2021) A three-dimensional mesoscale model for progressive time-dependent deformation and fracturing of brittle rock with application to slope stability. Comput Geotech 135:104160
Zhao L, Zhang S, Huang D, Wang X, Zhang Y (2020) 3D shape quantification and random packing simulation of rock aggregates using photogrammetry-based reconstruction and discrete element method. Constr Build Mater 262:119986