



OPEN

# A New Representative Sampling Method for Series Size Rock Joint Surfaces

Man Huang<sup>1,2</sup>✉, Chenjie Hong<sup>1</sup>, Chengrong Ma<sup>1</sup>, Zhanyou Luo<sup>3</sup>, Shigui Du<sup>1</sup> & Fei Yang<sup>1</sup>

The greatest variability in both shear strength and roughness exists for joint samples with smaller size, which underscores the necessity of performing representative sampling. This study aims to provide a representative sampling method for series size joint surfaces. The progressive coverage statistical method is introduced to provide the sufficient sample capacity for series sampling sizes by setting different propulsion spaces. The statistical law of the joint surface morphology at different sampling sizes is measured by the 3D roughness parameter with  $\theta_{\max}^*/(C + 1)$ . Through an application in nine natural large-scale rock joints, nine consecutive sampling sizes from 100 mm × 100 mm to 900 mm × 900 mm are selected and 121 samples are successfully acquired from each sampling size. According to the frequency distribution of roughness statistics, a new sampling method combining the layering principle and K-medoids clustering algorithm is proposed to screen representative joint samples for each sampling size. The sampling results that meet the test accuracy requirements suggest the possibility of realizing an intelligent sampling method. In addition, the representative of the interlayer cluster center is validated. Finally, the comparison results with the traditional stratified sampling method prove that the proposed method has better stability.

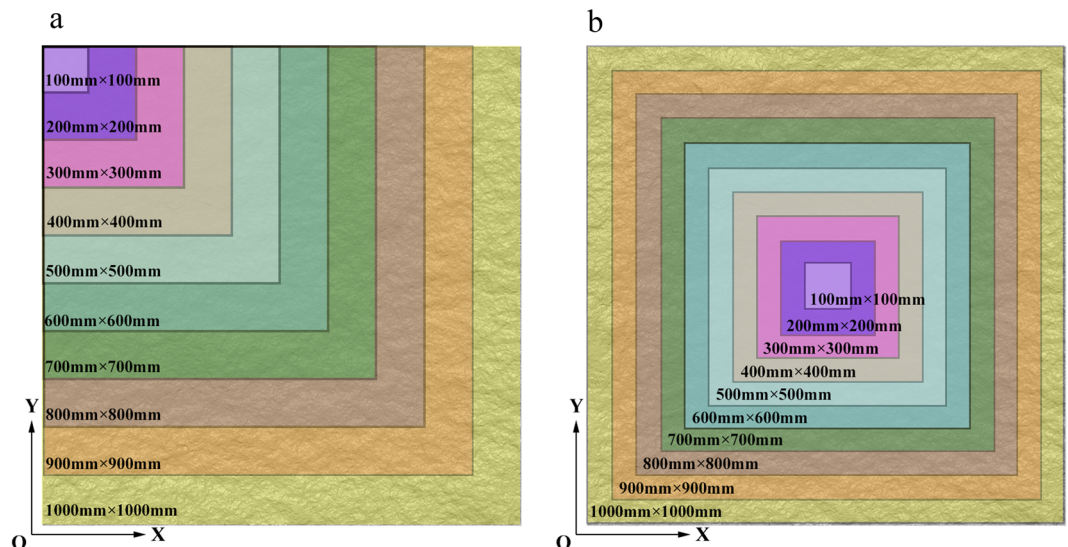
The shear behavior of rock joints is largely determined by size<sup>1,2</sup>. Hence, numerous shear test investigations on the scale effect in rock joints have been conducted<sup>3-7</sup>. Such investigations usually divide a large natural or artificially reproduced rock joint specimen into various joint models with small sizes. Direct shear tests under a given normal stress are then conducted on these joint surfaces of series sizes. Finally, the scale effect on the peak shear strength is obtained by comparing the average peak shear strength of the rock joints of each specimen size to that of the original specimen<sup>8</sup>. In this process, the material properties of different-sized joint specimens are the same in those of the original specimen. Therefore, the acquisition of each specimen size is equal to obtain the surface morphology (roughness) of the specimen at this size.

Roughness parameterization provides methods for quantifying the characterization of joint surface morphology, including empirical<sup>9</sup>, statistical<sup>10,11</sup>, and fractal methods<sup>12-14</sup>. Then, the scale effects of joint roughness and shear strength are established<sup>15-17</sup>. However, the representativeness of different-sized specimens is often neglected when studying the scale dependency of the two. The mechanical test of joint models with different sizes requires that each specimen represents the undulating characteristics of a specific size. However, the results of scale effect obtained by selecting a single sample to replace all specimens for experimental studies are questionable. Therefore, the representativeness of all specimens should be systematically analyzed.

Recently, some scholars have put forward methods for the statistics of joint samples. Yong *et al.* proposed a method for obtaining continuously sized joint profiles with overlapping length<sup>18</sup>. Considering that the statistical method of joint samples around two-dimensional profile is noncomprehensive, Huang *et al.* proposed a progressive coverage statistical method based on the idea of overlapping sampling, which can realize the statistics of three-dimensional (3D) joint morphology in series size<sup>19</sup>. However, the representative sampling of joint samples is not involved.

To explore the most representative roughness samples on a joint surface, Huang *et al.* first proposed a stratified sampling method for the representative sampling of joint samples<sup>20</sup>. However, a large number of probability calculations limit the use of this methods as the scale of joint research increases. In recent years, clustering, which

<sup>1</sup>Department of Civil Engineering, Shaoxing University, 508 Huancheng West Road, Shaoxing, 312000, Zhejiang, China. <sup>2</sup>Department of Civil and Environmental Engineering, Louisiana State University, 3255 Patrick F. Taylor hall, Baton Rouge, Louisiana, 70803, USA. <sup>3</sup>Geotechnical Engineering Institute, Zhejiang University of Science and Technology, 318 Liuhe Road, Hangzhou, 310023, Zhejiang, China. ✉e-mail: [hmcadx@126.com](mailto:hmcadx@126.com)



**Figure 1.** Processive magnifying sampling method. (a) Side amplification; (b) middle amplification.

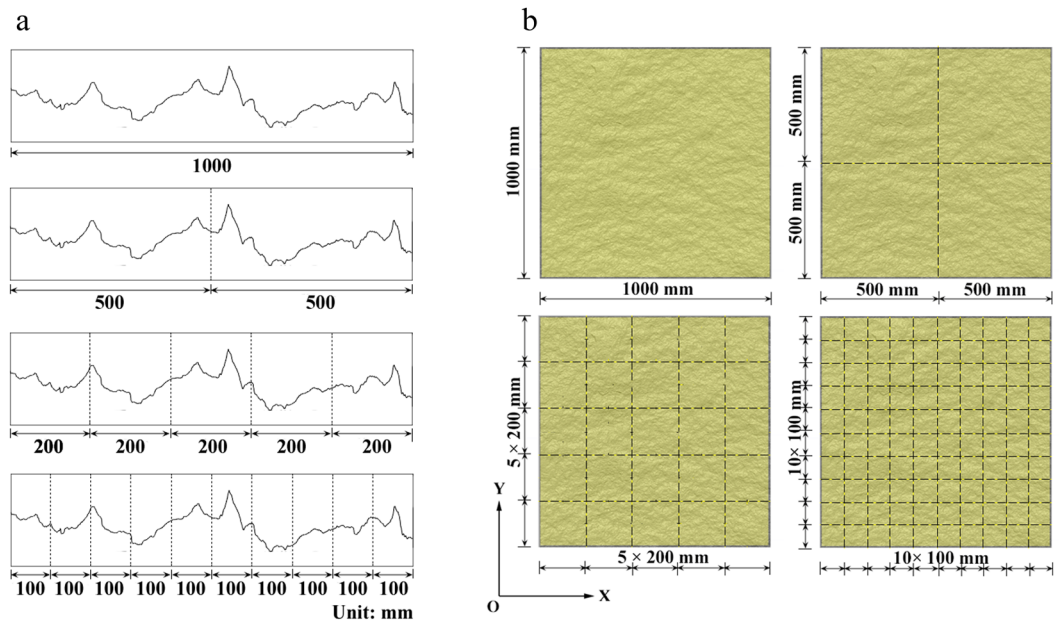
plays an important role in exploring data, has been used in the traditional joint roughness analysis<sup>21–23</sup>. Among them, the K-medoids clustering algorithm is widely recognized for its insensitivity to processing data outliers<sup>24</sup>. Therefore, the stratified sampling method may be optimized by the K-medoids clustering algorithm.

In this study, we investigate the traditional sampling methods in rock joints and propose an improved stratified sampling method, which can be combined with K-medoids clustering algorithm to perform intelligent sampling. The progressive coverage statistical method is introduced to obtain series size joint samples. Furthermore, the representative assessment of sampling results with the proposed method is tested with natural rock joints, and its feature and advantages are compared with those of the traditional stratified sampling method. In addition, the representative verification of a cluster center and sensitivity analysis of k-value in K-medoids sampling process is discussed in this paper. In doing so, the accuracy of the mechanical test is validated, and the work efficiency is improved.

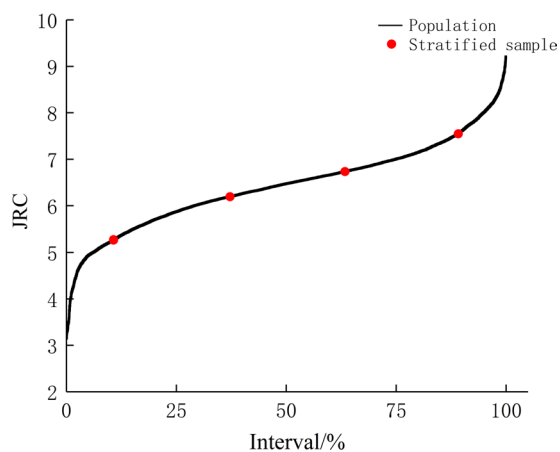
## Methodology

**Traditional sampling methods in rock joints.** The mechanical test of joint models of different sizes requires small-scale model specimens to be included in large-scale ones<sup>20</sup>. Four existing sampling methods follow the above requirements, namely, simple random sampling method, processive magnifying sampling method, equal-partition sampling method, and stratified sampling method.

- (1) Simple random sampling method. Considering the difficulty of performing roughness measurements in field rough joints, samples of different sizes are often arbitrarily taken from the original surface. The locations of different-sized samples primarily rely on the personal judgment and choice of researchers. Therefore, representing the corresponding size of the surface morphology with such random and irregular joint samples is not comprehensive.
- (2) Processive magnifying sampling method<sup>18</sup>. The processive magnifying sampling method refers to the sampling process in which the large-sized samples are obtained through a regular amplification of small-sized samples from a side or middle section (Fig. 1), which can overcome the irregularity of the simple random sampling method. This sampling method has been widely used in scale effect research on rock joints<sup>25–28</sup>. However, this sampling method is still a one-sample characterization method, the number of samples in each sampling size is small and their representativeness are unclear.
- (3) Equal-partition sampling method<sup>18</sup>. To obtain different positions of the original samples, the equal-partition sampling method is applied to joint sampling<sup>3,6</sup>, which provide profile samples through equivalence partitioning (Fig. 2a). On the basis of this definition, the 3D surface morphology may be evenly divided, as shown in Fig. 2b. Here, it is found that too many samples of small size and too few at the large size would be provided with the equal-partition sampling method. Moreover, this sampling method cannot guarantee full coverage samples for each sampling size because some sampling sizes cannot be divisible.
- (4) Stratified sampling method. The stratified sampling method introduced by Huang *et al.* is used to improve the case of having excessive samples in the small sampling size<sup>20</sup>. This method considers two key factors: stratified proportions are determined through the quartile method of JRC and sampling capacity is calculated within the permissible error range. The representative joint samples are then selected in the respective layer with the product of the sample quantity and stratified proportions, which will minimize the sample size and make them reasonably distributed (Fig. 3). However, due to the insufficient acquisition of large-sized sample sizes in current sample statistical methods, this method is limited to the application of small size representative samples. In addition, this method does not define the value of the interlayer samples.



**Figure 2.** Equal-partition sampling method. (a) schematic diagram of uniform profile; (b) equal partition of the surface.



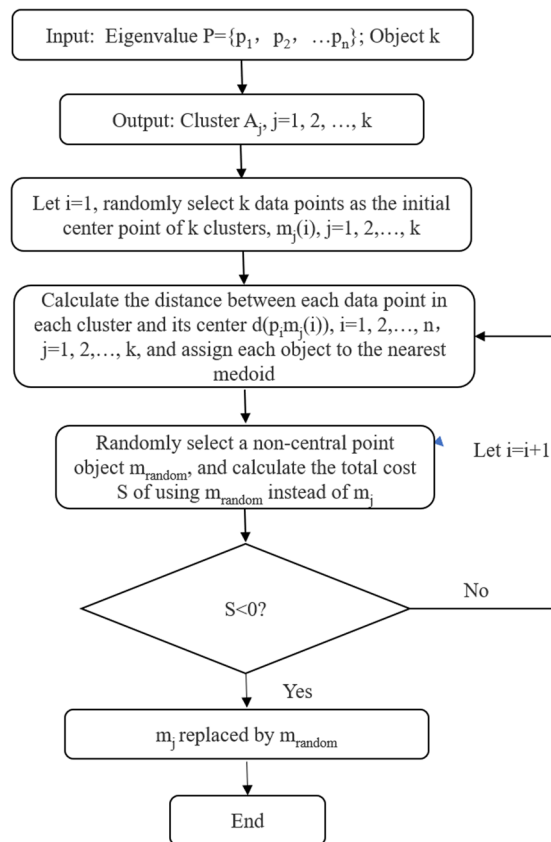
**Figure 3.** Stratified sampling method.

Hence, the arbitrarily selected joint samples may be less representative.

In summary, there are still two major shortcomings of the conventional sampling method. The first is uneven sampling (i.e., too many samples of small size and too few at the large size). The second is that the representativeness of the obtained samples is not clear. Therefore, establishing a new sampling method that can ensure the equal number of samples for different sampling sizes and therefrom select the representative samples is necessary.

**Progressive coverage statistical method.** To provide a comprehensive sampling basis for the sample statistics of series sample sizes, Huang *et al.* proposed a new progressive coverage statistical method<sup>19</sup>. Through this method, unit samples of different sizes are propelled in orthogonal directions along the entire rock joint with different propulsion spaces (see Supplementary Fig. S1). Here, with the propulsion of the unit sample, more complete unit samples are obtained, which can cover the partial missing area morphology, increase the sample capacity, and improve the accuracy of the overall samples. Therefore, this statistical method can effectively solve the problem of excessive or insufficient samples in different sampling sizes.

**K-medoids clustering algorithm.** K-medoids clustering algorithm is a data mining tool whose aim is to find K representative objects from the data set in such a way that the sum of the within-cluster dissimilarities is minimized<sup>29</sup>. It assigns every object to the nearest centroid by calculating the Euclidean distance, and the expression is given by Park *et al.* as



**Figure 4.** K-medoids clustering algorithm flowchart.

$$d_{ij} = \sqrt{\sum_{a=1}^p (X_{ia} - X_{ja})^2} \quad i = 1, \dots, m; j = 1, \dots, m, \quad (1)$$

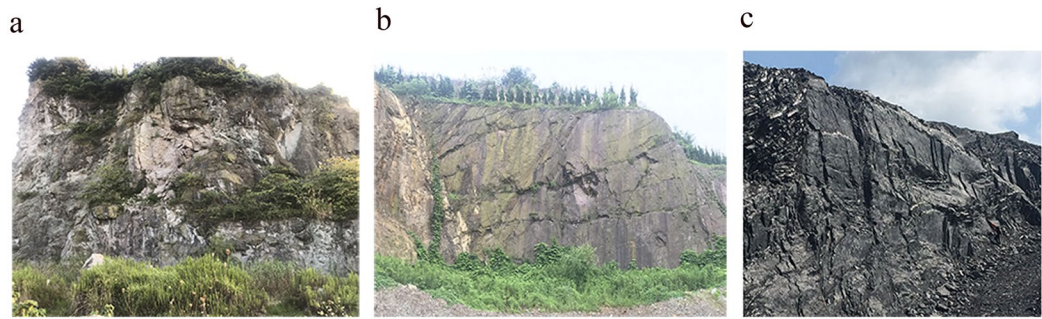
where  $m$  is the total objects,  $p$  is the number of variables, and  $X_{ia}$  and  $X_{ja}$  are the  $a$ th variable of objects  $i$  and  $j$ , respectively<sup>24</sup>. Figure 4 depicts the basic operation process of this clustering algorithm. The flowchart shows that the K-medoids clustering algorithm has a classification function similar to that of the stratified sampling method. Moreover, all the obtained cluster centers are real data and the most representative sample in each cluster, which means that the defect of random selection in interlayer samples is overcome. Therefore, the K-medoids clustering algorithm can be combined with the stratified sampling method for the representative sampling of joint surface morphology.

**New combined sampling method.** The progressive coverage statistical method can provide enough sample size for series sampling sizes and the stratified sampling method with K-medoids clustering algorithm has great advantage in representative sampling. Therefore, we combine the two methods for the representative sampling of rock joints. To characterize the surface morphology of the joint samples, the quantitative parameters need to be determined first. In the current study of roughness parameters, Grasselli *et al.* established a relationship between total potential contact area ratio  $A_{\theta^*}$  and apparent dip angle  $\theta^*$  as follows<sup>3</sup>:

$$A_{\theta^*} = A_0 \left( \frac{\theta_{\max}^* - \theta^*}{\theta_{\max}^*} \right)^C, \quad (2)$$

where  $A_0$  is the maximum potential contact area ratio,  $\theta_{\max}^*$  is the maximum apparent dip angle in the shear direction, and  $C$  is the roughness fitting coefficient. Then, Tatone and Grasselli<sup>11</sup> proposed a 3D roughness parameter  $\theta_{\max}^*/(C + 1)$  to evaluate the rock joint roughness, which is widely recognized<sup>4</sup>. Therefore, we take  $\theta_{\max}^*/(C + 1)$  as the quantitative parameters of the sample for the combined sampling method. The specific process is composed of the following steps:

**Step 1: Sample statistics.** The corresponding propulsion space  $\Delta d$  for different sampling sizes is determined, and then the square sampling unit is progressively advanced along the surface morphology of the joints to derive the appropriate sample capacity. The sample capacity  $N$  is calculated as



**Figure 5.** Large-scale natural rock joint. (a) tuff; (b) sandstone; (c) limestone.

$$N = \left[ \frac{L - l}{\Delta d} + 1 \right]^2 \quad (3)$$

where  $L$  is the side length of the original square joint and  $l$  is the side length of the target sample size, such that  $l \leq L$ . Then, the 3D roughness parameters  $\theta_{\max}^*/(C + 1)$  of each joint sample at a given shear direction are calculated.

Step 2: Sample stratification. In accordance with the definition of the stratified sampling method,  $\theta_{\max}^*/(C + 1)$  of each sampling size is arranged from small to large, and the sample layer division is defined by the relative range into two cases. When the relative range is greater than 10%, the quartile is used as the boundary to divided all the statistical values into three intervals, that is, 0–25%, 25–75%, and 75–100%, in which the distribution proportion  $W$  of samples in each layer is 1/4, 1/2, and 1/4, respectively; and when the relative range is less than 10%, an interval of 0%–100% is defined for the statistical values because of the small variation in the sample roughness, and its distribution proportion  $W$  is 1. The sampling quantity  $n$  is calculated through the stratified sampling equation as follows:

$$n = \frac{\sum W_h S_h^2}{V + \sum W_h S_h^2 / N}, \quad (4)$$

$$V = \left( \frac{\gamma \bar{Y}}{t} \right)^2, \quad (5)$$

where  $h$  is the layer number,  $S^2$  is the variance,  $V$  is the mean variance,  $t$  is the upper quantile of the standard normal distribution,  $\gamma$  is the permissible error,  $\bar{Y}$  is the population mean, and  $N$  is the total sample number.

Step 3: Representative sampling with the K-medoids clustering algorithm. For one sampling size, the eigenvalue  $P$  of the K-medoids clustering algorithm for the stratified samples is determined as follows:

$$P_h = \{(x_{h1}, l), (x_{h2}, l), \dots, (x_{ht}, l)\}, \quad (6)$$

where  $x_{ht}$  represents the  $\theta_{\max}^*/(C + 1)$  values of the  $t$ th joint sample in  $h$  layer. The  $k$ -value in the K-medoids clustering algorithm is determined based on the results of the sample distribution of each layer as

$$K_h = \lceil n \times W_h \rceil. \quad (7)$$

Then, the K-medoids clustering algorithm is run to obtain the  $K$  cluster centers from the data of the eigenvalue. The joint samples corresponding to these center points are the representative samples of the corresponding sampling size.

To verify the feasibility of the new sampling method, we will use it to perform the sampling in natural rock joints. In this method, the different propulsion spaces will be selected to obtain a sufficient sample size for series sampling sizes and the  $k$ -values in different layers for each sampling size will be determined to carry out representative sampling. Finally, the method evaluates the representativeness of the sampling results to show good applicability of the combined sampling method in the rock joints.

## Application

**Acquisition of a large-scale joint.** A large-scale and well-preserved natural rock joint should be prepared before conducting the statistics of the joint samples. After many field investigations, we found three sets of different lithological joints (tuff, sandstone, and limestone), whose width and height are all more than 1 m (Fig. 5). To collect the 3D surface information, a portable laser scanner (MetraSCAN 3D, Creafom, Canada) with a scanning accuracy of 0.5 mm is used to scan the field site. Three 1000 mm  $\times$  1000 mm digitized surfaces are selected for each lithology as the original joints, labeled as T1, T2, T3, S1, S2, S3, L1, L2, and L3 (see Supplementary Fig. S2).

**Progressive coverage statistical procedure.** To make a comprehensive division of rock joints, nine consecutive sampling sizes with a dimension of 100 mm × 100 mm to 900 mm × 900 mm are selected in this work. Then, the progressive coverage statistical method is used to set the propulsion spacing to 90, 80, 70, 60, 50, 40, 30, 20, and 10 mm, which can achieve 121 samples per sampling size. Hence, the progressive coverage statistical method can create conditions for the representative sampling of large-sized samples.

## Results

**Joint sample roughness distribution.** Roughness measurement is performed on the obtained joint samples by means of roughness parameter  $\theta_{\max}^*/(C + 1)$  along shear direction at 0°, in which the roughness statistics of T1, S1, and L1 joints are given in the form of histograms, as shown in Supplementary Fig. S3. It shows that the distribution in roughness is quite wide at small sample sizes, but quite narrow at the larger sample sizes. This means that the smaller the sampling size, the greater dispersion of joint sample roughness. Therefore, the representativeness of the joint samples is very important. In addition, from the perspective of frequency distribution of the joint sample roughness, although the normal, skewed distribution, and irregular distributions appeared, the roughness is still distributed throughout the different intervals. This finding indicates that such data sets are suitable for screening representative samples with the stratified sampling method.

**Determination of the sampling quantity.** According to the sample stratification (step 2) in Section 2, the permissible error  $\gamma$  is set as 0.15, and the upper standard quantile  $t$  in the corresponding standard normal distribution table is 1.96 when the confidence level is 95%. The interlayer variance ( $S_{0-25}^2$ ,  $S_{25-75}^2$ ,  $S_{75-100}^2$ , or  $S_{0-100}^2$ ) of each sample size is calculated and the calculated data is substituted into Eq. (4) to obtain the sampling quantity  $n$  and the  $k$ -values in accordance with Eq. (7). The calculation results of the T1, S1, and L1 joints can be found as Supplementary Table S1. The sample capacity of different sampling sizes gradually decreases due to the decrease in the difference between the joint sample surface morphology as the sampling size increases.

**Sampling with the K-medoids clustering algorithm.** The sampling work starts by determining the  $k$ -value for the interlayer sample in each sampling size. After acquiring the eigenvalue of the samples for each sampling size, we sequentially input the eigenvalue and corresponding  $k$ -value into the K-medoids clustering algorithm. The clustering results under different sampling sizes are shown in Supplementary Fig. S4, where the cluster centers are distributed in different sampling intervals.

**Representative assessment of the sampling results.** Sampling of the joint model requires that the selected sample be representative of the undulating surface morphology under a specimen size. Considering that the proposed sampling method is based on the roughness statistics, Fig. 6 shows the variation of mean  $\theta_{\max}^*/(C + 1)$  of the cluster centers ( $L_c$ ) and the mean  $\theta_{\max}^*/(C + 1)$  of population ( $L_p$ ). The relative error  $\delta$  is calculated as follows:

$$\delta = \frac{|L_p - L_c|}{L_p} \times 100\%. \quad (8)$$

The obtained results have good coherence, in which the maximum error is 6.78% (see Supplementary Table S2). Therefore, the sampling results obtained by the K-medoids clustering algorithm can meet the test accuracy requirements.

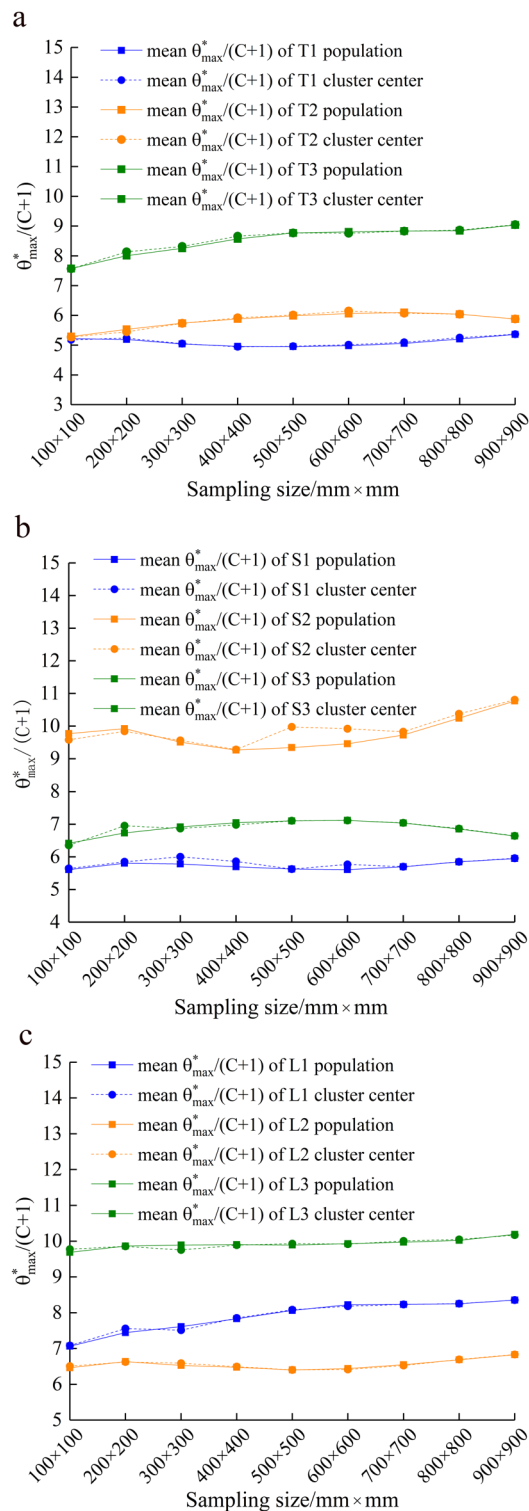
## Comparison and Discussion

**Representative verification of interlayer cluster centers.** From the process of K-medoids stratified sampling, the whole representativeness of the sample depends on the representativeness of the interlayer cluster centers. To demonstrate the representativeness of the cluster centers in each layer, we take the T1 joint as the research object and compare the mean  $\theta_{\max}^*/(C + 1)$  of the cluster samples of three layers and the mean  $\theta_{\max}^*/(C + 1)$  of the cluster centers (Fig. 7). The results show that each interlayer cluster center is almost identical to the average of the cluster samples. This finding indicates that the K-medoids clustering algorithm can ensure the consistency of the cluster results, which is the biggest feature and advantage of this method in representative sampling.

**Comparison with different sampling method.** To compare with the different sampling methods, we take the sampling size at 100 mm × 100 mm of the T1 joint as an example. First, the progressive coverage statistical method is used to get population samples, and then the roughness  $\theta_{\max}^*/(C + 1)$  of all samples are calculated to obtain population mean. Second, five repeated samplings are performed with Simple random sampling, Processive magnifying sampling, Equal partition sampling, Stratified sampling and K-medoids sampling methods, and then the roughness  $\theta_{\max}^*/(C + 1)$  of samples are calculated to obtain sample mean. Finally, the relative error between the sample mean of different sampling times and the population mean is calculated, as shown in Fig. 8. Although the results show that the relative error using the Equal partition sampling method is more constant and smaller, the large sample capacity will increase the burden for successive test. Therefore, the K-medoids sampling method with relative errors basically stabilized below 5% and appropriate sample capacity is the best choice for the representative sampling of joint surface morphology.

## Conclusions

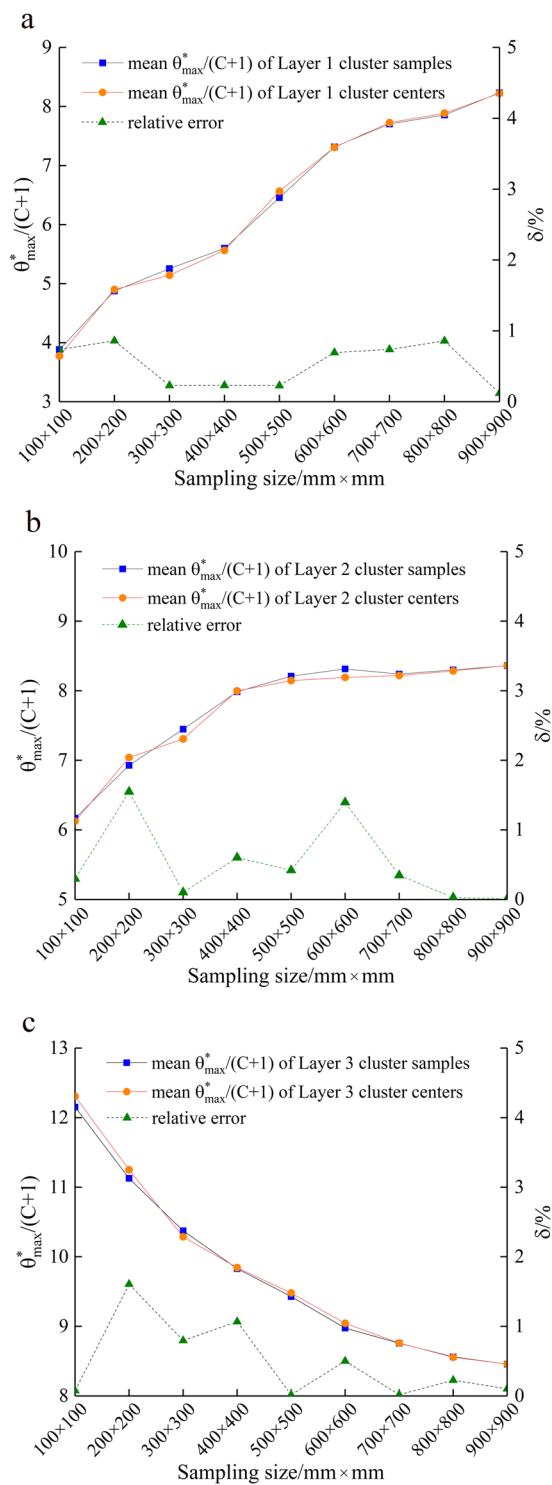
A new representative sampling method for rock joint surfaces presented in this paper, the major summaries and conclusions are as follows:



**Figure 6.** Evaluation of the cluster results. (a) tuff; (b) sandstone; (c) limestone.

The performance of four traditional sampling methods that are frequently used in the joint model test, namely, the simple random sampling method, processive magnifying sampling method, equal-partition sampling method, and stratified sampling method, are investigated in representative sampling. The results show that the conventional sampling methods have different drawbacks in the sampling of series size joint samples, and their performances do not involve the representative characterization of large-sized ones.

A new sampling method that combines the progressive coverage statistical and K-medoids stratified sampling methods is proposed to achieve an intelligent representative sampling for rock joints. The reasonably allocated samples of each sampling size are selected, and the roughness representativeness of the samples is quantitatively

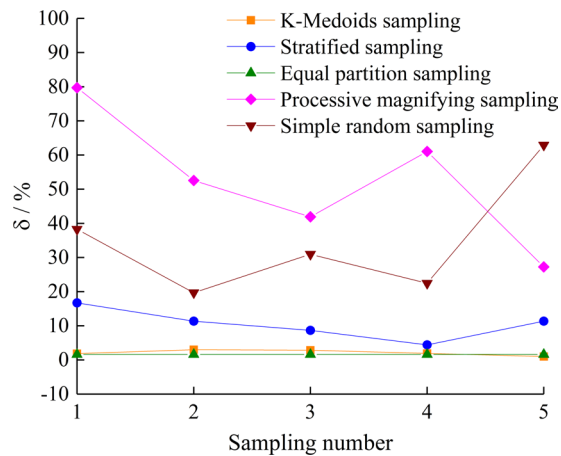


**Figure 7.** Evaluation of the interlayer cluster results.

evaluated on the basis of the statistical analysis of the  $\theta_{\max}^*/(C+1)$  values. The relative errors between the mean  $\theta_{\max}^*/(C+1)$  of the samples and the mean  $\theta_{\max}^*/(C+1)$  of the population are almost entirely below 5%. This finding indicates that the new sampling method can effectively provide representative joint samples for the joint model tests.

The representative verification of the interlayer cluster centers is carried out. The results show that the K-medoids clustering algorithm can effectively achieve a reasonable allocation of cluster centers and ensures the representativeness of the sampled samples. In addition, in accordance with the comparison with the traditional stratified sampling method, the K-medoids clustering algorithm enables a stable representative sampling of joint samples.





**Figure 8.** Sample distribution of different sampling methods.

To verify the mechanical reliability of the selected samples, the comparative analysis of the mechanics need be further studied. Furthermore, the scale effect of shear behavior based on the representative samples will be investigated via direct shear tests in the future.

Received: 31 December 2019; Accepted: 14 May 2020;

Published online: 04 June 2020

## References

- Hencher, S. R. & Richards, L. R. Assessing the shear strength of rock discontinuities at laboratory and field scales. *Rock Mechanics and Rock Engineering* **48**, 883–905 (2014).
- Muralha, J. *et al.* ISRM Suggested Method for Laboratory Determination of the Shear Strength of Rock Joints: Revised Version. *Rock Mechanics and Rock Engineering* **47**, 291–302 (2014).
- Bandis, S., Lumsden, A. & Barton, N. Experimental studies of scale effects on the shear behaviour of rock joints. *International Journal of Rock Mechanics and Mining Sciences* **18**, 1–21 (1981).
- Re, F., Scavia, C. & Zaninetti, A. Variation in contact areas of rock joint surfaces as a function of scale. *International Journal of Rock Mechanics and Mining Sciences* **34**, 254.e1–254.e12 (1997).
- Vallier, F., Mitani, Y., Boulon, M., Esaki, T. & Pellet, F. A Shear Model Accounting Scale Effect in Rock Joints Behavior. *Rock Mechanics and Rock Engineering* **43**, 581–595 (2010).
- Bahaaddini, M., Hagan, P. C., Mitra, R. & Hebblewhite, B. K. Scale effect on the shear behaviour of rock joints based on a numerical study. *Engineering Geology* **181**, 212–223 (2014).
- Johansson, F. Influence of scale and matedness on the peak shear strength of fresh, unweathered rock joints. *International Journal of Rock Mechanics and Mining Sciences* **82**, 36–47 (2016).
- Ueng, T. S., Jou, Y. J. & Peng, I. H. Scale effect on shear strength of computer-aided-manufactured joints. *Journal of GeoEngineering* **5**, 29–37 (2010).
- Barton, N. & Choubey, V. The shear strength of rock joints in theory and practice. *Rock Mechanics* **10**, 1–54 (1977).
- Maerz, N. H., Franklin, J. A. & Bennett, C. P. Joint roughness measurement using shadow profilometry. *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts* **27**, 329–343 (1990).
- Tatone, B. S. A. & Grasselli, G. A method to evaluate the three-dimensional roughness of fracture surfaces in brittle geomaterials. *The Review of scientific instruments* **80**, 125110 (2009).
- Lanaro, F. A random field model for surface roughness and aperture of rock fractures. *International Journal of Rock Mechanics and Mining Sciences* **37**, 1195–1210 (2009).
- Fardin, N., Stephansson, O. & Jing, L. The scale dependence of rock joint surface roughness. *International Journal of Rock Mechanics and Mining Sciences* **38**, 659–669 (2001).
- Kulatilake, P., Balasingam, P., Park, J. & Morgan, R. Natural rock joint roughness quantification through fractal techniques. *Geotechnical and Geological Engineering* **24**, 1181–1202 (2006).
- Barton, N., Bandis, S. & Bakhtar, K. Strength, deformation and conductivity coupling of rock joints. *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts* **22**, 121–140 (1985).
- Fardin, N. Influence of Structural Non-Stationarity of Surface Roughness on Morphological Characterization and Mechanical Deformation of Rock Joints. *Rock Mechanics and Rock Engineering* **41**, 267–297 (2008).
- Li, Y., Oh, J., Mitra, R. & Cambulat, I. A Fractal Model for the Shear Behaviour of Large-Scale Opened Rock Joints. *Rock Mechanics and Rock Engineering* **50**, 67–79 (2017).
- Yong, R. *et al.* An Innovative Sampling Method for Determining the Scale Effect of Rock Joints. *Rock Mechanics and Rock Engineering* **52**, 935–946 (2019).
- Huang, M. *et al.* A progressive coverage statistical method for series scales three-dimensional morphology of rock joints. *Chinese Journal of Rock Mechanics and Engineering* **38**, 1533–1541 (2019).
- Huang, M., Luo, Z. & Du, S. Experimental study of sampling representativeness of structural plane of rock model. *Chinese Journal of Rock Mechanics and Engineering* **32**, 2008–2014 (2013).
- Jimenez, R. & Sitar, N. A spectral method for clustering of rock discontinuity sets. *International Journal of Rock Mechanics and Mining Sciences* **43**, 1052–1061 (2006).
- Xu, L., Chen, J., Wang, Q. & Zhou, F. Fuzzy C-Means Cluster Analysis Based on Mutative Scale Chaos Optimization Algorithm for the Grouping of Discontinuity Sets. *Rock Mechanics and Rock Engineering* **46**, 189–198 (2013).
- Li, Y., Wang, Q., Chen, J., Xu, L. & Song, S. K-means Algorithm Based on Particle Swarm Optimization for the Identification of Rock Discontinuity Sets. *Rock Mechanics and Rock Engineering* **48**, 375–385 (2015).
- Park, H. & Jun, C. A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications* **36**, 3336–3341 (2009).

25. Fardin, N., Feng, Q. & Stephansson, O. Application of a new *in situ* 3D laser scanner to study the scale effect on the rock joint surface roughness. *International Journal of Rock Mechanics and Mining Sciences* **41**, 329–335 (2004).
26. Du, S., Hu, Y. & Hu, X. Measurement of joint roughness coefficient by using profilograph and roughness ruler. *Journal of Earth Science* **20**, 890–896 (2009).
27. Tatone, B. S. A. & Grasselli, G. An investigation of discontinuity roughness scale dependency using high-resolution surface measurements. *Rock Mechanics and Rock Engineering* **46**, 657–681 (2013).
28. Chen, S., Zhu, W., Yu, Q. & Liu, X. Characterization of Anisotropy of Joint Surface Roughness and Aperture by Variogram Approach Based on Digital Image Processing Technique. *Rock Mechanics and Rock Engineering* **49**, 855–876 (2015).
29. Assis, E. & Souza, R. A K-medoids clustering algorithm for mixed feature-type symbolic data. In: IEEE International Conference on Systems. Anchorage, Alaska, USA; 2011.

### Acknowledgements

The study was funded by the National Natural Science Foundation of China (Nos. 41572299, 41427802), Natural Science Foundation of Zhejiang Province (No. LY18D020003), and Key Research and Development Projects of Zhejiang Province (No. 2019C03104). Their support is gratefully acknowledged.

### Author contributions

M.H. and C.J.H. designed the study and wrote the main manuscript text, C.R.M. did most work of algorithm writing and analyzed the joint roughness data, Z.Y.L. and S.G.D. conducted the laser scanning in field, F.Y. prepared figures and tables. All authors reviewed the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-66047-0>.

**Correspondence** and requests for materials should be addressed to M.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020