Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multi-Scale Topographical Criteria for Systematic Automated Structural Mapping

Niloufarsadat Sadeghi

Niloufarsadat.sadeghi.1@ens.etsmtl.ca

École de Technologie Supérieure, ÉTS

Jonathan D. Aubertin

École de Technologie Supérieure, ÉTS

Mohammad Niknezhad Golafzani

École de Technologie Supérieure, ÉTS

Research Article

Keywords: Automated structural mapping, unsupervised machine learning, scale-dependent analysis of rock surfaces, fractals, structural geology, rock mass characterization, remote sensing, topographical data sets.

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10589396/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

MULTI-SCALE TOPOGRAPHICAL CRITERIA FOR SYSTEMATIC AUTOMATED STRUCTURAL MAPPING

Niloufarsadat Sadeghi, Jonathan D. Aubertin, Mohammad Niknezhad Golafzani
École de Technologie Supérieure, ÉTS, 1100 Notre-Dame St W, Montréal, Québec, H3C 1K3, Canada

ABSTRACT:

Remote sensing-assisted structural mapping has become a prevalent tool in rock mass characterization. Practitioners use topographical models and unsupervised machine learning (*ML*) algorithms to objectively group and extract geological structures from exposed rock outcrops. A major limitation of current *ML*-assisted solutions arises from the inherent scale-dependent nature of the input features. The scales at which topographical features (e.g. roughness, local orientation) are measured directly impact statistical variability and the amount of noise or unclassified data generated by highly fractured surfaces and edges. Current methods rely on arbitrary selection for the scale of measurement or extensive and redundant pre-processing, making the workflow iterative, computationally demanding, and subjective.

To address these challenges, this research proposes a scale-dependent analysis framework that derives systematic, site-specific topographical criteria for unsupervised *ML* clustering. Rooted in fractal theory principles, the methodology computes surface roughness across varying radii of measurement. These measurements are compiled using power-law trends and quantified via the Hurst exponent and an amplitude coefficient to identify transitional scales.

Once these characteristic site-specific scales are established, roughness and local orientation are computed for the topographical data set and processed using *K*-means clustering to differentiate structural elements. By integrating these features at their characteristic scale, the algorithm isolates edges and noisy data into a distinct, extra cluster. This cluster is filtered out for automated mapping reflective of the natural variability.

The presented framework was validated on two case studies: a surface quarry rock slope and an underground tunnel face. When benchmarked against digital manual mapping the proposed algorithm demonstrated remarkable accuracy, achieving combined orientation deviations of less than 3.5° for both environments. Computational time, reproducibility of the workflow and convenient implementation of the proposed workflow offers a meaningful contribution towards automated structural mapping.

KEYWORDS:

Automated structural mapping, unsupervised machine learning, scale-dependent analysis of rock surfaces, fractals, structural geology, rock mass characterization, remote sensing, topographical data sets.

1 INTRODUCTION

Rock mass discontinuities, defined as physical interruptions in the rock continuum, appear as faults, joints, bedding planes, or blast-induced fractures along rock outcrops (Priest and Hudson, 1976). These features define planes of weakness along which rock blocks can detach and fail. Proper identification and description of their orientation, frequency, spacing, persistence, and surface roughness are essential for evaluating rock mass stability, and corresponding failure models (Hammah and Curran, 1999). Conventional mapping of geological structures using a Brunton compass or similar apparatus tends to be tedious, time-consuming, and potentially hazardous. Technological analogues using remote sensing techniques have therefore become prevalent in many rock engineering applications.

Automated solutions have emerged to improve efficiency by detecting and extracting discontinuity orientations directly from topographical data sets (Chen, Zhu and Li, 2016; Gallant and Marshall, 2016; Wang *et al.*, 2017; Drews *et al.*, 2018; Ge *et al.*, 2018; Singh *et al.*, 2022; Chen *et al.*, 2023). Early attempts at normal vector-based structural mapping often failed to differentiate relevant structural elements from noise (e.g., vegetation, debris) or procedural artifacts arising from edges and ridges (Gigli and Casagli, 2011; Lato and Vöge, 2012; Zhang *et al.*, 2018; Battulwar *et al.*, 2021; Singh, Raval and Banerjee, 2021).

Recent developments in automated structural mapping have leveraged machine-learning (*ML*) algorithms for systematic clustering based on physical measurements. Slob et al. (2005) applied an adjusted fuzzy *k*-means clustering approach to facet normals extracted from triangulated meshes. The approach relied on iterative variance checks to delineate cluster boundaries and assuming Fisher distribution. The process required important parameter tuning and post-processing to refine discontinuity sets. Assali et al. (2014) applied *k*-means clustering directly to normal vectors. This approach provided a systematic mapping framework but featured important sensitivity to noise and arbitrary choice of neighborhood radius for normal estimation.

Cluster-based methods were proposed to filter irrelevant features. Riquelme et al. (2014) integrated local curvature, coplanarity tests, and stereonet-based density estimation with Density Based Scan Algorithm with Noise (*DBSCAN*) clustering. This method improved reliability of the results by filtering and merging local maxima but still relied on arbitrary parameter selection and was prone to underrepresenting low-density sets. Chen, Zhu and Li (2016) introduced an improved *k*-means framework combined with digital surface model triangulation, discontinuity segmentation, and Random Sample Consensus (*RANSAC*) plane fitting, further enhancing automation at the costs of heavy parametric and iterative optimization.

Recent efforts have focused on multi-parameter descriptors and hybrid methods. Singh, Raval and Banerjee (2021) incorporated multiple geometric descriptors (e.g., curvature, orientation, fast point feature histogram) with *k*-medoids clustering, supported by noise and vegetation filters. This framework improved the discrimination of discontinuity planes, but it required processing many input features, which increased computational cost and made the analysis

more complex. Singh et al. (2022) proposed an amplitude–phase decomposition (*APD*) technique to measure orientations at the intra-cluster level, achieving very low orientation deviations. Similar to many recently developed solutions, this approach imposes subjective quantification to orientation variability, directly impacting intended applications.

Current *ML*-assisted automated structural mapping solutions effectively integrate one or more topographical features computed on raw data-sets to support systematic clustering. These topographical features (e.g. local orientation, roughness, undulation, planarity, etc.) are inherently scale dependent. The radius of measurement at which these parameters are computed reflects a characteristic scale associated with the considered surface. For rock outcrops, these scales of measurement may represent crystalline texture (i.e. small scale), fracture intensity (medium scale) or excavation geometry (large scale) (Aubertin and Hutchinson, 2022). Consider for example the surface normal $\vec{N}$ computed at different radii of measurements r (m) along a rock outcrop (Figure 1). The corresponding textures (represented by lighting shadows) will reflect fracture intensity and structural edges at small scale, and excavation geometry at larger scale. Figure 1 conceptually illustrates the influence of radius of measurement for normal orientation along a rock slope with two different radii of measurements. By extension, most topographical features are greatly influenced by the scale at which they are measured. Consequently, *ML*-assisted structural mapping is inherently scale-dependent.

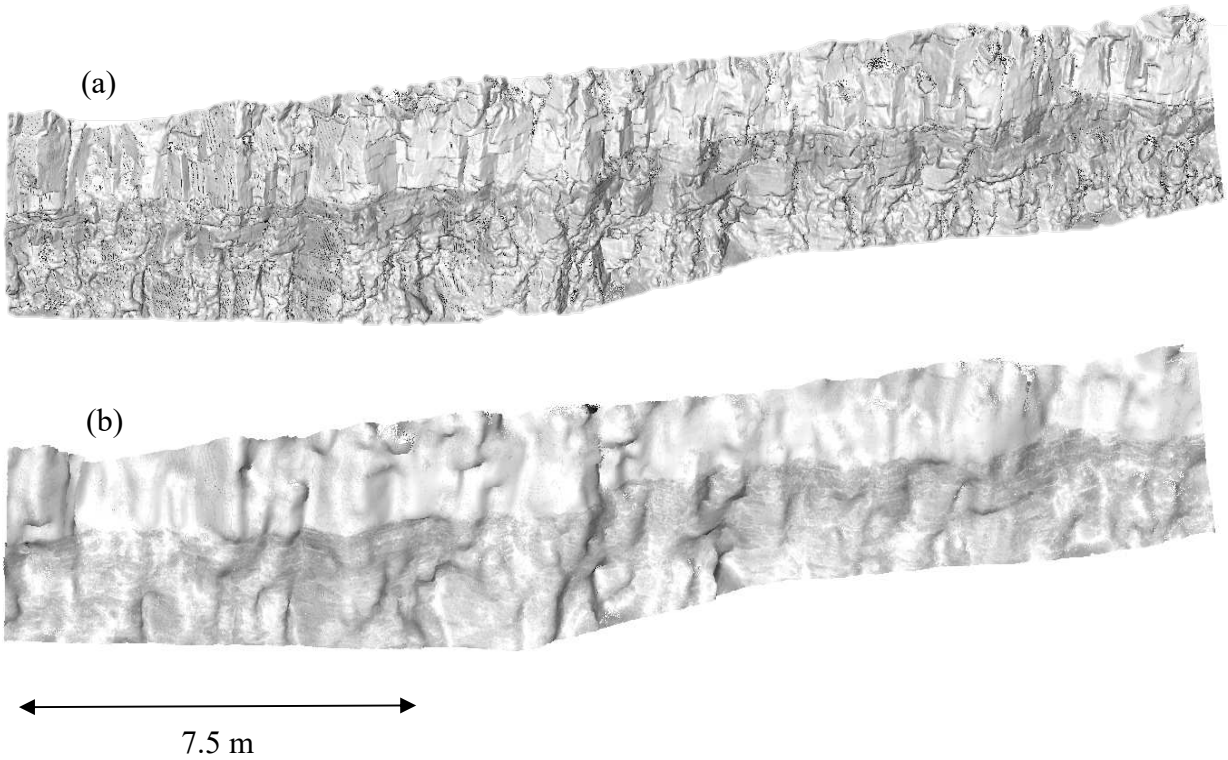


Figure 1: Influence of (a) small (r = 0.02m) and (b) large (r = 0.5m) neighborhood radii on point-normal orientation $\vec{N}$ computation. Visuals generated with CloudCompare V2.14 (Girardeau-Montaut D., 2024).

Despite the fundamentally scale-dependent nature of *ML*-assisted automated structural mapping, current solutions either arbitrarily impose scales of measurements based on iterative site-specific empirical evidence or incorporate extensive redundancy in the scale of measurements fed to the *ML* algorithms. Consequently, available *ML*-assisted automated structural mapping require extensive site-specific adjustment difficult to reproduce systematically, or involve computationally intensive pre-processing and large data sets difficult to integrate pragmatically in operational settings.

This work presents site-specific scale-dependent criteria for topographical characterization of rock surfaces. These criteria are implemented in a pre-processing workflow to generate a robust input for unsupervised *ML* extraction of structural geology. In this research, surface orientation and topographical roughness are computed at the derived scale-dependent criteria. These parameters were selected for their intuitive nature, high correlation with structural geology, and simple computation with commercially available or open-source software. The topographical data set including the computed multi-scale roughness measurements and surface orientations are integrated to k-means clustering. The characteristic scales incorporated in the clustering groups the structural families and outputs an additional cluster associated with local edges and high fracture-intensity noise. This later cluster is filtered out, leaving distinct structural clusters for systematic characterization at characteristic scales with representative corresponding variability (captured here with Fisher *K* parameter).

The proposed framework was tested on a surface quarry rock slope and a tunnel face. The method was benchmarked against digital manual mapping. Additional benchmarking with respect to other *ML*-assisted solutions is presented for the tunnel face case. The results demonstrate that explicitly accounting for the scale dependency of geometric feature extraction allows meaningful roughness and normal-estimation scales to be identified and integrated into an unsupervised clustering framework. This scale-informed approach reduces the influence of edge-related noise and density variations, improves the separation of structural domains, and produces orientation clusters that are consistent with manually mapped reference measurements. The workflow operates directly on point cloud datasets without requiring preprocessing, surface generation, or explicit plane fitting.

2 COMPUTATIONAL METHODS

2.1 Topographical representation of structural elements

In this work, point cloud visuals were generated using the open-source software CloudCompare V2.14 (Girardeau-Montaut D., 2024). Python scripts were developed to systematically compute relevant geometric features at different radii of measurements. Stereographic projections for structural orientations were generated using RocScience software *DIPS* (Rocscience Inc., 2006).

The present work implements vectorized normal vectors $\vec{N} = [n_x, n_y, n_z]$ and topographical roughness R (m) as input parameters for automated segmentation and differentiation. The mathematical definitions to evaluate a best-fitted plane, the resulting normal orientation, and the distance between a plane and a point (i.e. roughness) can be found in most linear algebra textbooks (e.g., Stewart (2003)). Relevant equations are summarized in the appendices for reference.

Roughness R (m) is computed as the distance from a best-fitted plane for points within distance r (see appendix). Consequently, roughness values are higher near edges. This zone of higher roughness is explicitly controlled by the measurement radius r (m). Figure 2 exemplifies this behavior for the point cloud of a simple cube with color scale for roughness values. Points located near edges exhibit higher roughness (red), as the computed distances (i.e., roughness R) are obtained from best-fitted planes that span two intersecting surfaces. As can be deducted from Figure 1, structural data extraction from normal vector orientations suffers the same computational artefact and influences greatly measured variability around structural characterization .

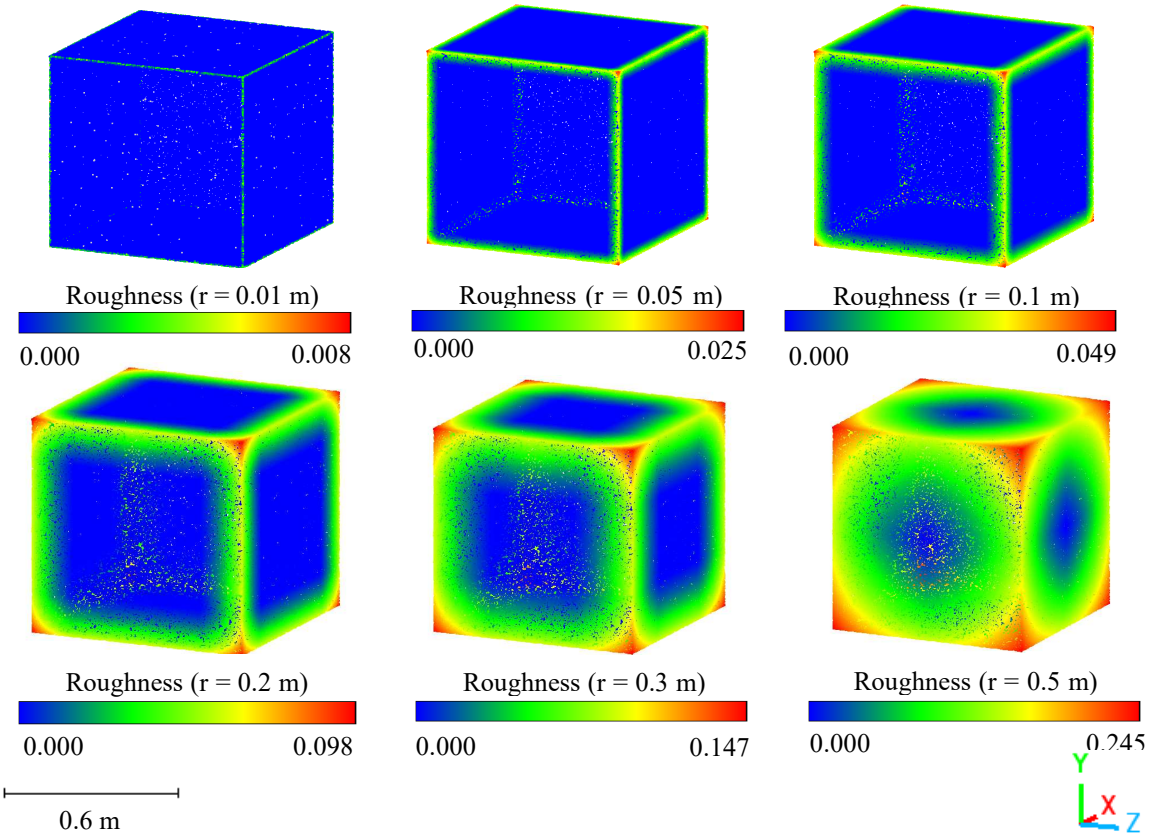


Figure 2: Scale-dependent roughness of a cube model for neighborhood radii between 0.01 m and 0.5 m, illustrating the transition from fine-scale edge irregularities to broader surface curvature with increasing radius.

Figure 3 expands upon the synthetic cube example. Surface normals were computed at radius of measurement $r = 0.05$ m and plotted in a stereographical projection. The corresponding plot presents concentrated data points at the cardinal edges due to plane orientation, and extensive noise from edge-induced artifacts. In the figure, these edge data points were filtered out, and the corresponding stereographic plot regenerated, featuring idealistic precision.

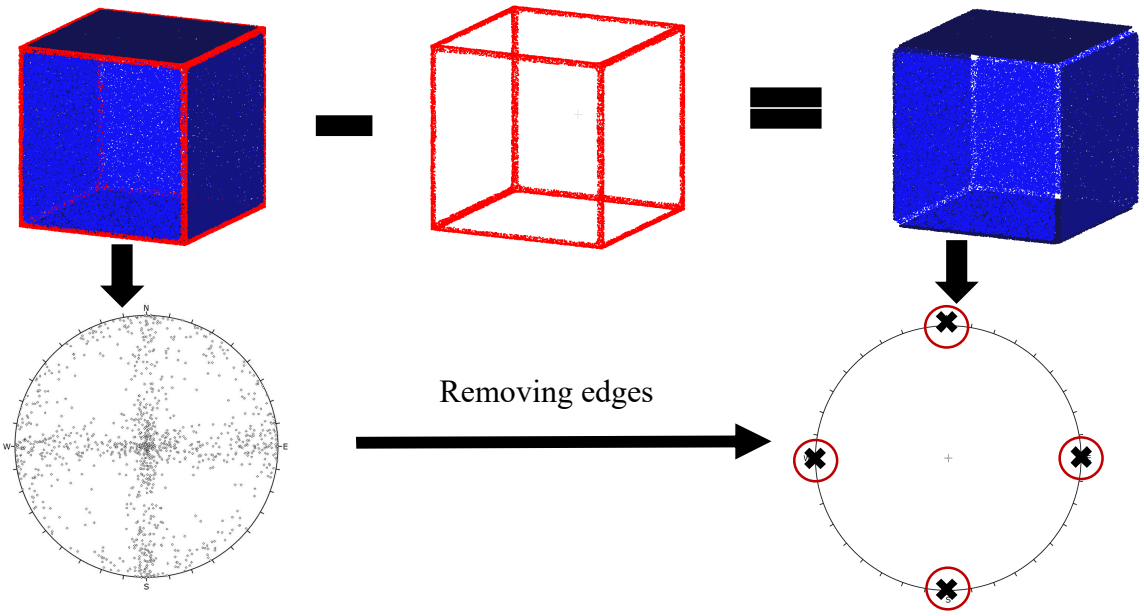


Figure 3: Illustration of edge-induced artefacts in normal vector orientation analysis and their removal through edge filtering.

The logic exemplified in Figure 2 and Figure 3 were implemented in the proposed framework. The next section develops the scale-dependent analysis proposed to evaluate systematically topographical criteria for objective and reproducible characterization.

2.2 Multi-scale criteria for topographical measurements

In this work, a multi-scale approach analogous to fractals (Mandelbrot, 1967) is implemented to evaluate the proper scale of measurement for reproducible automated structural mapping. The fractal-like rationale is developed for topographical roughness following work by Shepard et al. (1995, 2001), Aubertin and Huthinson (2022) and Sadeghi and Aubertin (2025). The overall roughness ξ_R (m) of a surface is evaluated as the Root Mean Square Deviation (*RMSD*) of measured point roughness values for radius of measurement *r:*

$$\xi_R = \sqrt{\frac{1}{n-1}\sum_1^n (R_r - \bar{R}_r)^2} \quad [1]$$

where *n* is number of points along the considered area, and $\bar{R}_r$ is the average of point roughness values for radius of measurement r.

The $\xi_R - r$ relationship for various geomaterials will typically present a power relationship of the form (Mandelbrot and Wheeler, 1983; Fardin, Stephansson and Jing, 2001; Aubertin and Hutchinson, 2022; Sadeghi and Aubertin, 2025) :

$$\xi_R = A(r)^H \quad [2]$$

where H has been referred to as the Hurst exponent (Hurst, 1951) and A is an analogue to roughness amplitude (Kulatilake and Um, 1997).

Plots for $\xi_R - r$ data points measured along an exposed rock surface will typically present ranges for which parameters of Eq. 4 remain constant. The corresponding values of parameters A and H represent characteristic signatures relating to the influence of scale on overall topography of the considered surface (Sadeghi and Aubertin, 2025)

Figure 4 plots the $\xi_R - r$ for the rock surface presented in Figure 1. For $0.06 \geq r\,(\text{m}) \geq 0.6$, roughness ξ_R exhibits a linear trend in the log–log space, indicating power-law behavior. This scale range corresponds to geometric features governed by similar dominant textures (Aubertin and Hutchinson, 2022). The plot also features two additional distinctive trends in the log-log scale. The small-scale trend ($r < 0.06$ m) is influenced by point spacing of the data set and crystalline texture. It will tend to further plateau horizontally as r approaches minimum point spacing. The large-scale trend ($r > 0.6$ m) represents excavation geometry. In the considered example, the large scale of measurement leads to a lower correlation between scale and roughness as the local best-fitted plane of each point converges to one best-fit plane representative of the overall surface.

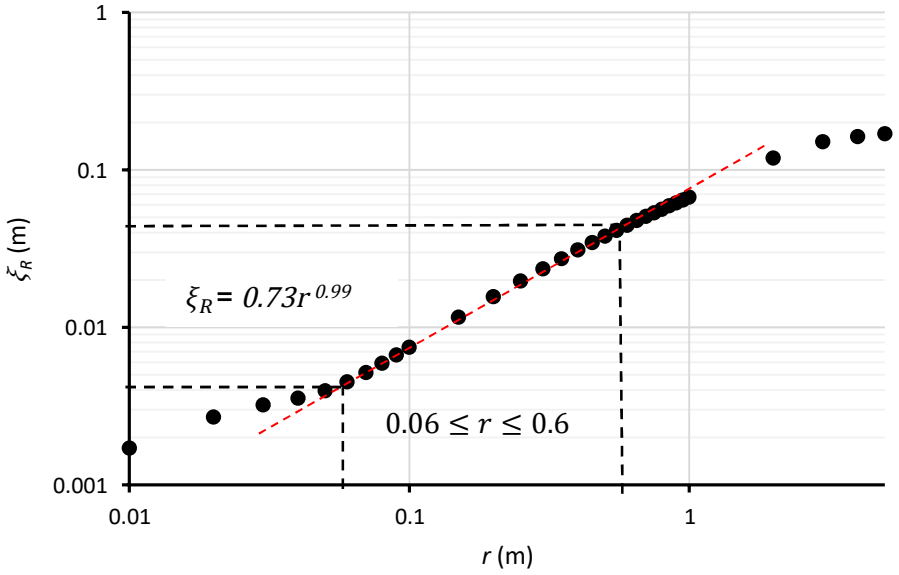


Figure 4: Log–log plot of ξ_R versus r, highlighting the linear scale range representing characteristic roughness behavior.

The characteristic trend for $0.06 \geq r\,(\text{m}) \geq 0.6$ presented in Figure 4 is therefore bounded by scale thresholds with physical significance. The distinctive trend, and corresponding upper and

lower threshold boundaries represent topographical scale criteria for reproducible and objective automated structural mapping. Topographical measurements within the lower and upper bounds r_1^* and r_2^* (m) represent the structural elements that characterize the rock mass at the considered scale.

2.3 Unsupervised clustering

In the proposed workflow, unsupervised *K*-means clustering (MacQueen, 1967) is performed using normal vector orientations and surface roughness measured at meaningful scales determined from multi-scale topographical criteria. Surface roughness supports the explicit segregation of edges and irregular subsets, while normal orientation partitions the dataset according to the number of discernible discontinuity sets, resulting in structurally meaningful clusters.

The similarity between points is quantified based on the angular difference θ (°) between their unit normal vectors, defined by the following equation:

$$\theta = cos^{-1}\left(\frac{\vec{N}_{r,i}.\vec{N}_{r,j}}{\left|\vec{N}_{r,i}\right|.\left|\vec{N}_{r,j}\right|}\right) \quad [3]$$

where $\vec{N}_{r,i}.\vec{N}_{r,j}$ is the dot product of two normal vectors $\vec{N}_{r,i}$ and $\vec{N}_{r,j}$ with respective magnitude $\left|\vec{N}_{r,i}\right|$ and $\left|\vec{N}_{r,j}\right|$

The silhouette method (Rousseeuw, 1987) is used to determine the optimal number of clusters K. This approach evaluates clustering performance by measuring how similar each point is to its own cluster compared to other clusters. The average silhouette score is calculated for different values of *K*, and the value that yields the highest score is considered optimal. The silhouette score ranges from –1 to +1, where values close to +1 indicate that the point is well-matched to its own cluster and poorly matched to neighboring clusters. A score near 0 suggests overlapping clusters, and negative values indicate misclassification.

The effectiveness of each feature included in the *K*-means clustering is evaluated through quantitative analysis using the Davies–Bouldin Index (*DBI*) in conjunction with the silhouette score. *DBI* evaluates clustering quality by comparing the compactness of individual clusters with their separation from the most similar neighboring cluster. A lower *DBI* values reflect better-defined, more distinct clusters (Davies and Bouldin, 1979). Roughness and local orientation measured for $r_1^* \leq r \leq r_2^*$ were tested as part of the detailed analysis presented in the case studies. In subsequent sections, it is demonstrated with the computed DBI ratings that targeted values derived from the topographical criteria r_1^* and r_2^* can be implemented systematically for efficient automated structural mapping and avoid excessive pre-processing redundancy.

2.4 Topographical filters and post-processing

Unsupervised clustering considering surface orientation and roughness yields M+1 number of clusters where M represents the number of distinct discontinuity sets. The extra cluster corresponds to structural edges and noise. This later cluster is explicitly filtered out. Normal vectors are recomputed within each remaining cluster separately. The results are converted into dip D (°) and dip direction DD (°) values to generate stereographic projections.

A density-based filter is applied to the pole projections in the stereonet to remove scattered or unreliable poles. For each pole, the number of neighbors within a defined angular radius is calculated using a *k-d* tree search. A *k-d* tree is a space-partitioning data structure that enables efficient nearest-neighbor queries in multi-dimensional space and provides a measure of local pole density (Bentley, 1975). The angular radius specifies the maximum angular separation on the unit sphere within which neighboring poles are counted and is converted to a chord distance d (m) such that:

$$d = 2 \cdot sin\,(\alpha/2) \quad [6]$$

where α (°) is the angular radius.

The distribution of neighbor density values is analyzed, and an automatic cut-off is defined as the 95th percentile of the density distribution. Poles with densities below this threshold are classified as isolated or noise-related and are excluded. This objective filtering step retains poles associated with meaningful joint sets while removing those generated by random orientation scatter.

3 APLICATION EXAMPLES AND VALIDATION

3.1 Scope and validation approach

The procedural framework presented in the previous section yields pole projections for structural measurements represented in a stereonet. The mean pole orientation for a given structure can be derived using built-in functions of the software *DIPS* (Rocscience Inc., 2006). Two case studies are presented next to exemplify the procedural method and evaluate its performance in surface and underground settings.

Clustering results were benchmarked against manual digital mapping performed on each data set. Deviations were calculated for each set and compared between the considered approach. Fisher's K parameter (Fisher, 1953) was computed to evaluate statistical variability of the resulting pole distributions (see appendices for an explicit mathematical definition of Fisher's K parameter).

3.2 Case study 1: Aggregate surface quarry

3.2.1 Settings

The proposed procedural workflow was applied to an aggregate quarry located south of Montreal, Canada. The site is characterized by distinct tabular geological units and sub-vertical structures resulting in various block arrangements. Figure 5 shows a picture of a 9 m vertical rock slope excavated at the quarry by conventional drilling and blasting. The figure highlights the structural prevalence of the rock mass considered for this case study.



Figure 5:Picture of a 9 m vertical bench excavated by drilling and blasting at the aggregate quarry considered for this case study.

A FARO X-130 laser scanner (FARO Technologies, 2015) was used to survey the rock surfaces and generate a topographical model. Figure 6 shows the point cloud model of the surveyed area, captured from seven scanning positions and registered using the open-source software CloudCompare (Girardeau-Montaut D., 2024). Scan placement and registration processes followed the methodology outlined by Aubertin (2025). The studied section of the rock surface is highlighted in blue in Figure 6. It covers an area of approximately 30 × 5 m and contains around 2,500,000 points with average point spacing of 1 cm.

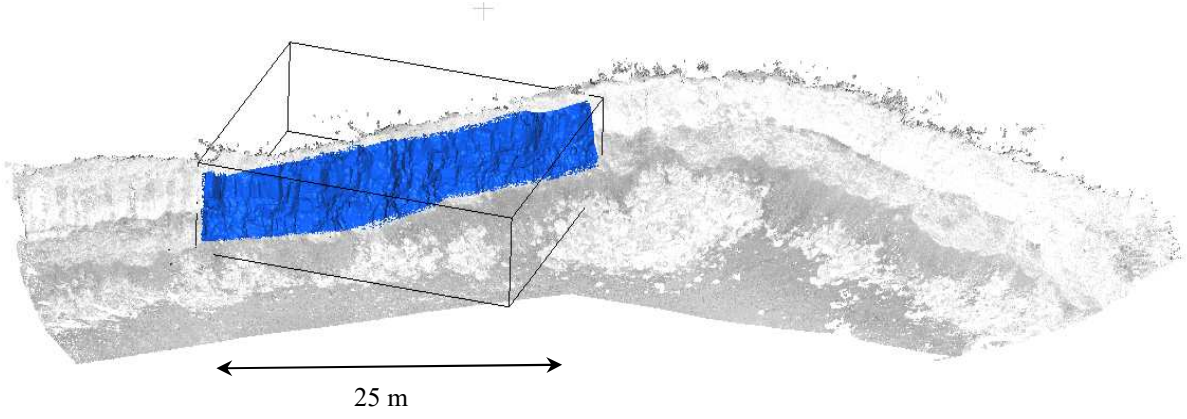


Figure 6: 3D point cloud model of the same area, with the studied span highlighted in blue.

3.2.2 *Manual digital mapping*

Manual digital mapping was conducted on the topographical model using CloudCompare V2.14 (Girardeau-Monteau, 2022) and the built-in virtual compass plugin. Figure 7 shows the resulting equal-angle lower hemisphere stereonet plot based on 45 manual measurements. Three joint sets are identified from manual mapping. Two sub-vertical structures dipping near 80 degrees are noted. A sub-horizontal lithology common around the greater Montreal area is identified as the third set. The orientation details of each set and respective Fisher's *K* parameters are outlined in *Table 1*.

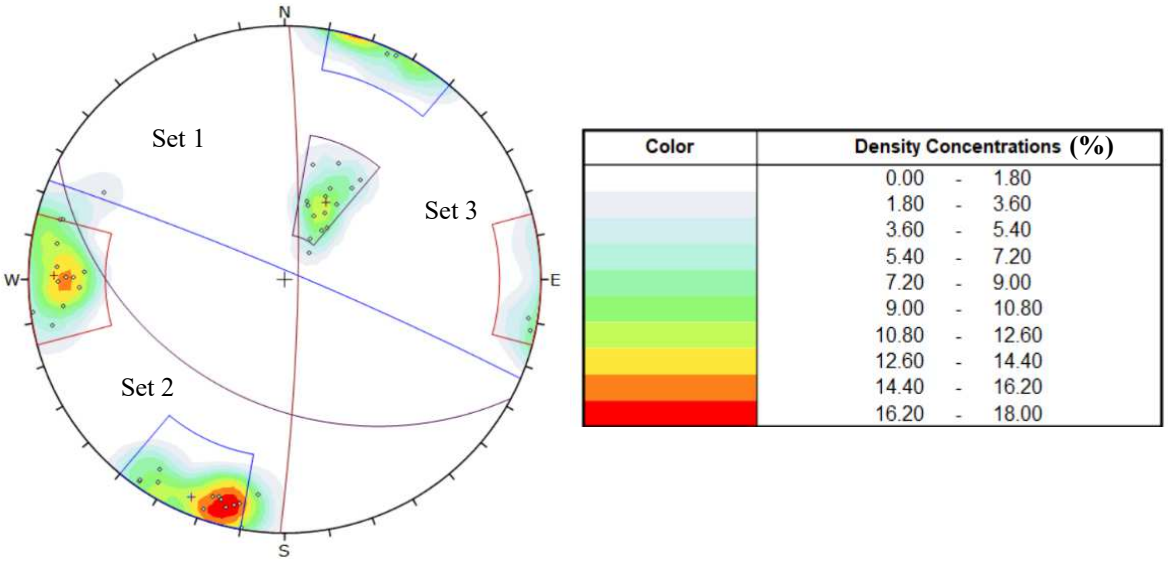


Figure 7: Plotted orientation of manually collected data from LiDAR scan on stereonet for the quarry case study.

Table 1: Orientation of three joint sets identified from manual mapping in the quarry case study.

ID	Dip (°)	Dip Direction (°)	Fisher's *k*
Set 1	83.59	90.70	86.13
Set 2	86.33	22.63	76.13
Set 3	38.21	207.84	52.35

3.2.3 *Scale-dependent criteria*

Scale-dependent analysis was performed across the range of $0.01 \leq r\ (\mathrm{m}) \leq 5.$ to derive the characteristic $\xi_R - r$ plot of the exposed rock outcrop (Figure 8). This range was established based on point spacing (lower threshold) and dimension of the bench for the upper bound (r equals half bench height for a circle of diameter equals to the bench height).

Figure 8 features a consistent linear trend in the plotted log-log plane along $0.035 \leq r \leq 0.5$ m. It is inferred that the lower bound range $r_1^* = 0.035$ m corresponds to the scale at which crystalline features from the exposed minerals start to dominate topographical texture. The

upper bound $r_2^* = 0.5$ m of the captured trend captures the broader topographical features associated with excavation geometry. The scale-dependent to characteristics for this rock surface are therefore represented within the range $0.035 \leq r\ (\text{m}) \leq 0.5$ defined by scale threshold criteria $r_1^* = 0.035$ m and $r_2^* = 0.5$ m.

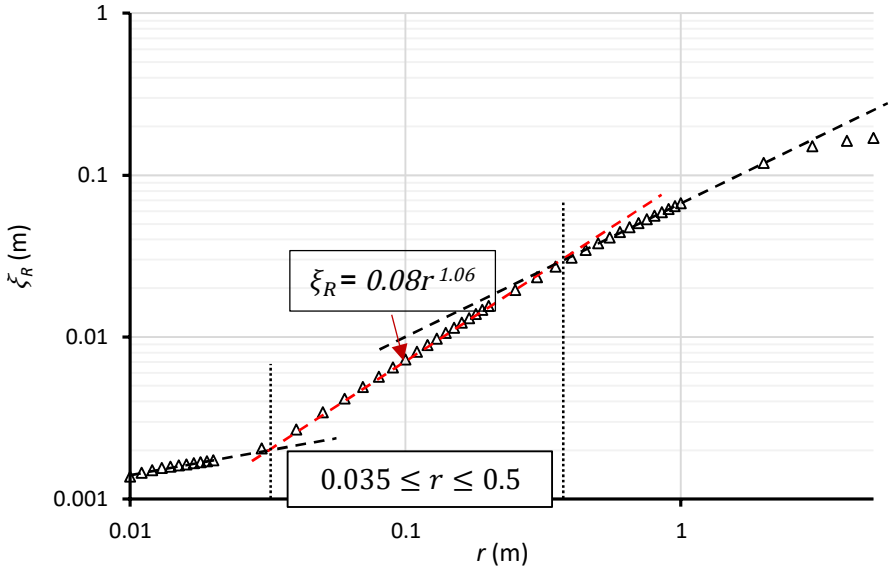


Figure 8: Log–log plot of ξ_R versus r for the quarry case study, illustrating scale-dependent variation in roughness.

3.2.4 Automated clustering results

Structural clustering was performed by integrating unit normal vectors with roughness values computed at three characteristic radii derived from the scale-dependent topographical criteria measured for the evaluated rock surface. Roughness and normals are evaluated at the lower and upper characteristic thresholds $r_1^* = 0.035$ and $r_2^* = 0.5$ m, and at the logarithmic mean between r_1^* and r_2^*, $\tilde{r} = 0.17$ m. Topographical measurements captured for the three characteristic scales r_1^*, r_2^*, and $\tilde{r}$ effectively represent the textural behavior of the features dominating the fractal behavior at that range. The characteristic criteria for scale of measurement provide an objective, reproducible, and computationally efficient framework for systematic clustering.

Clustering performance was evaluated for different combination inputs by considering *DBI* and average deviation from manual mapping. Relevant sample results are presented in Table 2 with computational radii for normals and roughness r_N (m) and r_R (m). It is noted that roughness input for the three characteristic radii of measurements consistently presented the best clustering performance. On the other hand, clustering performance drops substantially when surface orientation normals are computed for $r > \tilde{r}$. It is surmised that this result stems from the physical definition of the upper threshold r_2^* corresponding to the transition of dominant textural features. It is further noted that clustering performance with normals computed only for $r = r_1^*$ presented the best results for this case. For this optimal case, the silhouette method indicated an optimal number of clusters of $k = 4$.

Table 2: Clustering performance for samples of tested input combinations with varying radii of measurements.

Normal radii r_N (m)	Roughness radii r_R (m)	*DBI*	Average deviation (°)
[0.035, 0.17, 0.5]	[0.035, 0.17, 0.5]	1.1345	5.96
[0.035, 0.17]	[0.035, 0.17, 0.5]	1.1035	3.44
[0.035]	[0.035, 0.17, 0.5]	1.1038	3.62
[0.035, 0.17, 0.5]	[0.035, 0.17]	1.1102	6.64
[0.035, 0.17, 0.5]	[0.035]	1.1575	7.03

Figure 9 shows the clustering output obtained using r_N = 0.035, r_R = 0.035, 0.17, 0.5 m. Three main structural sets are identified, plus an extra cluster associated with edge and noise. The structural sets (Set 1, Set 2, and Set 3) appear as distinct, continuous regions that follow dominant orientations in the rock face. The edge cluster mainly captures noisy areas and irregular surfaces, especially around the boundaries of blocks.

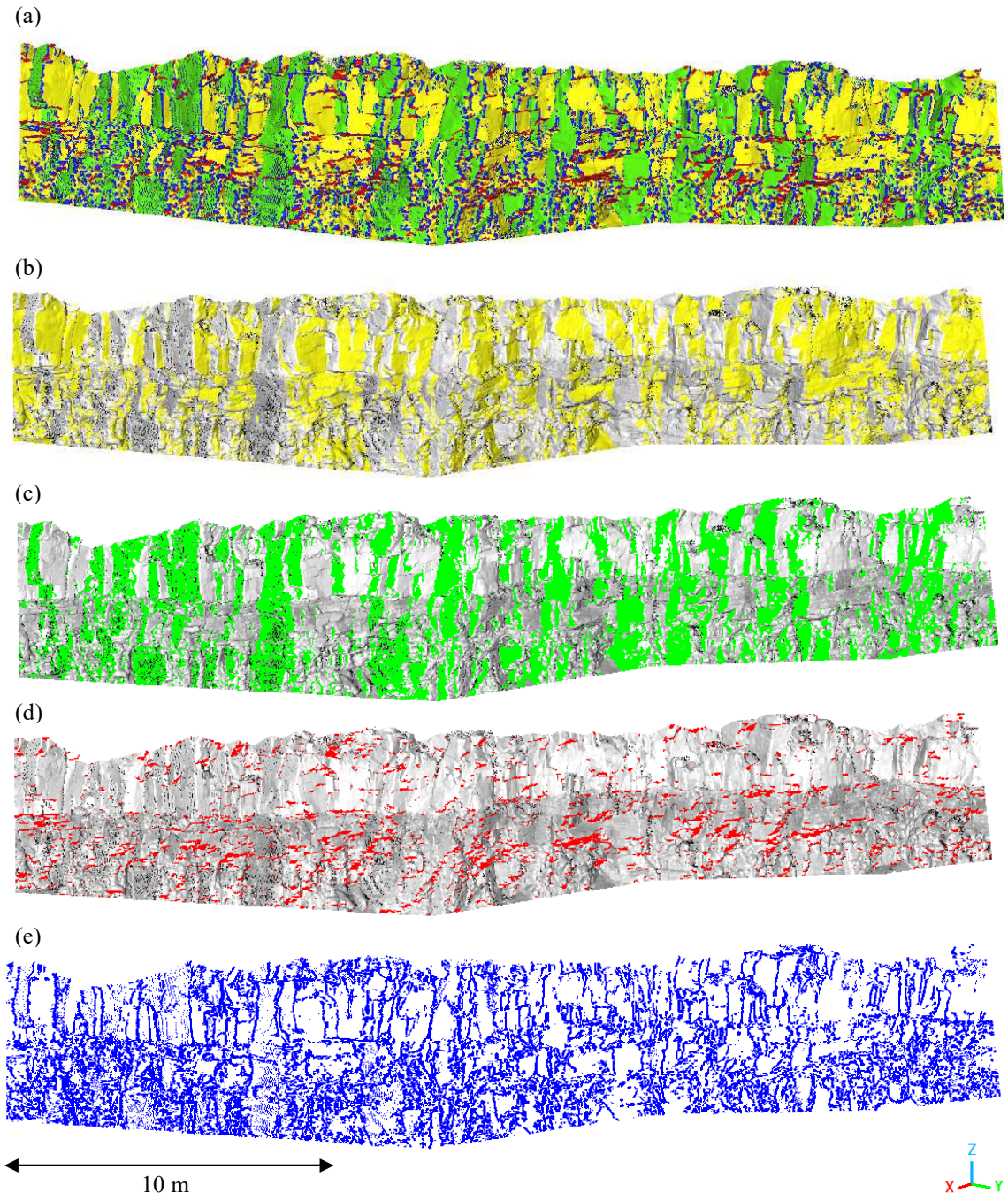


Figure 9: Clustering output at r_N = 0.035, r_R = 0.035, 0.17, 0.5 m for the quarry case study, showing (a) all clusters combined, (b) set 1, (c) set 2, (d) set 3, and (e) the edge cluster. Clusters (b–d) correspond to dominant orientations, while (e) captures noisy and irregular surfaces along block boundaries.

Normal vectors were recomputed independently for each cluster and converted to dip and dip-direction values. Density-based angular filtering was applied using a 95th-percentile threshold of local pole density. Figure 10 presents the stereographic projections for the individual structural families extracted. The figure also demonstrates the results from density-based angular filtering. The results presented and average deviation of 3.54° and a Fisher's *K* value of 24.22. Figure 11 compares the mean orientations obtained from manual mapping with that derived from the best-performing clustering configuration. The three dominant sets identified

through the automated workflow show close alignment with the manually mapped reference orientations, with only minor angular deviations. It is inferred that the statistical variability obtained with the proposed approach captures realistic behavior representative of the scale of measurement chosen, and therefore surpasses manual mapping achieved.

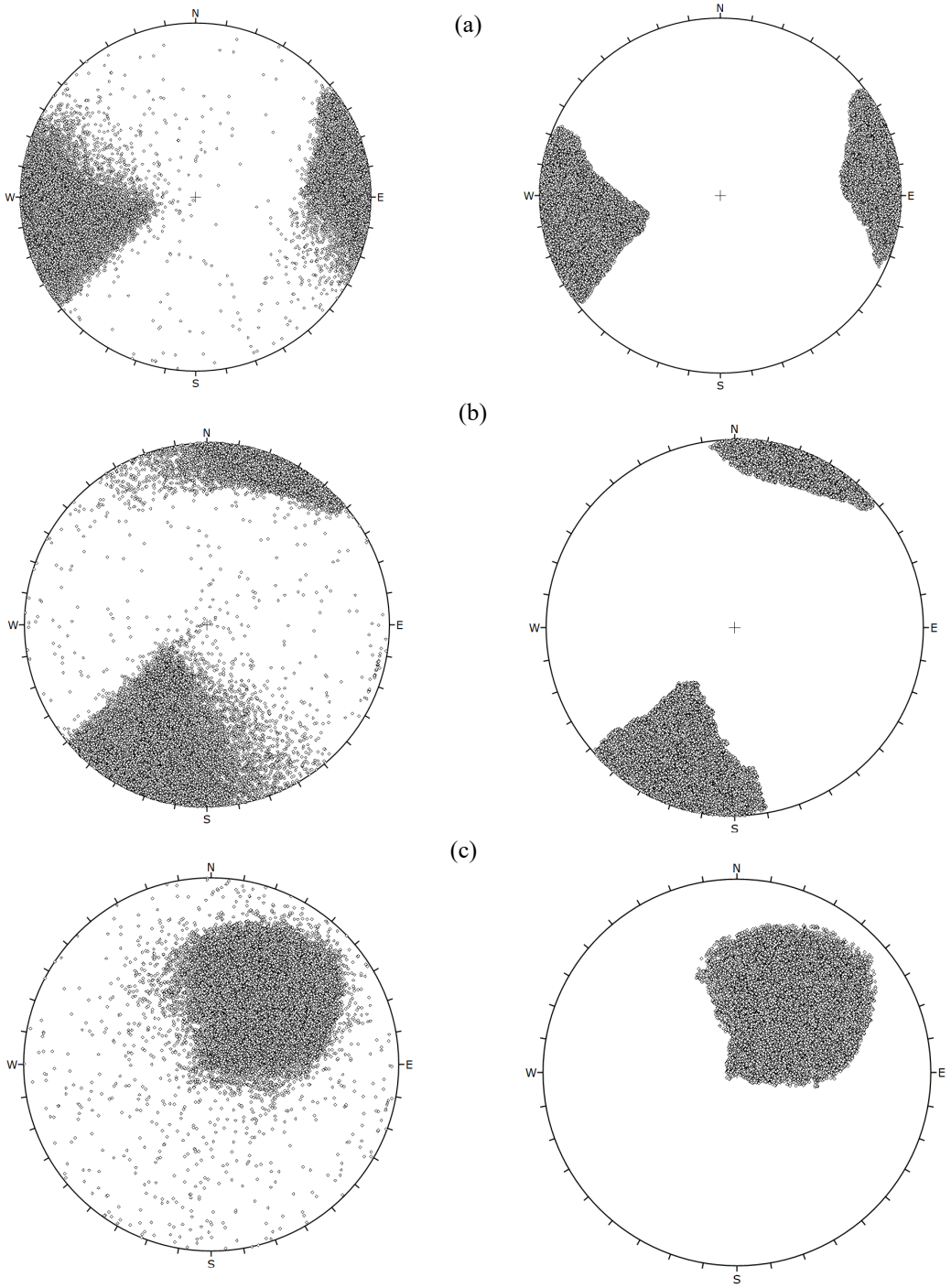


Figure 10: Effect of density-based filtering on orientation data for the quarry case study: (a) set 1, (b) set 2, and (c) set 3. Left: raw pole plots before filtering; right: results after applying a density-based filter.

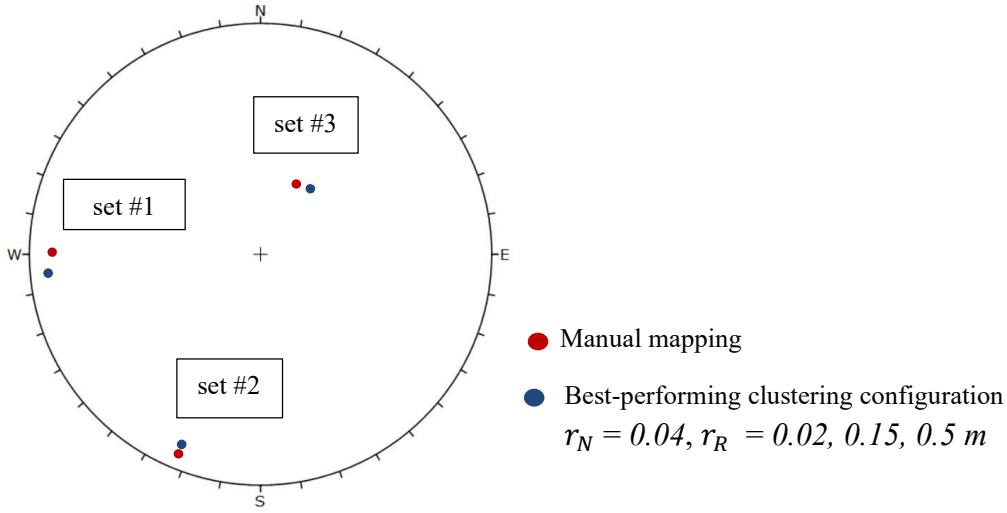


Figure 11: Comparison of mean orientations from manual mapping and the best-performing clustering configuration (r_N = 0.035, r_R = 0.035, 0.15, 0.5 m) in the quarry case study.

Density contour plots of the three joint sets obtained from clustering at this configuration are shown individually on stereonets (Figure 12 (a-c)), while a combined stereonet (Figure 12 (d)) is produced by randomly subsampling all clusters to equal point counts (200,000), ensuring uniform visual weighting and minimizing bias caused by density differences across clusters.

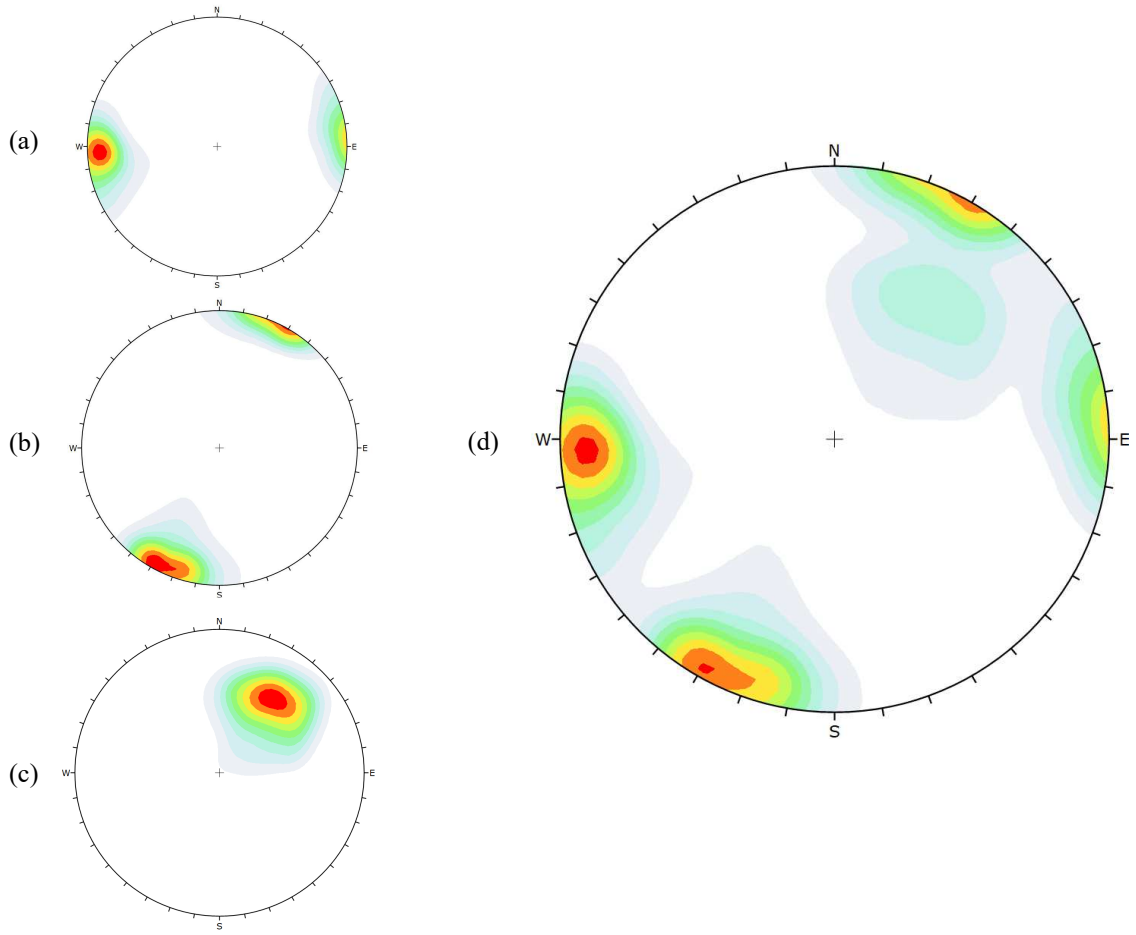


Figure 12: Stereonet contour plots of clustered discontinuity orientations at r_N = 0.04, r_R = 0.02, 0.15, 0.5 m *for the quarry case study: (a–c) individual joint sets and (d) combined stereonet with equalized point counts.*

3.3 Case study 2

3.3.1 Settings

The second case study is an underground tunnel face. The topographical model of the excavation was sourced from the open-access Rockbench repository (Lato *et al.*, 2013). The selected tunnel-face section is shown in Figure 13. The model contains approximately 4 million points with an average point spacing of 5 mm. The dataset was previously investigated by Singh et al. (2022), who identified four dominant discontinuity sets using an amplitude–phase decomposition–based approach.

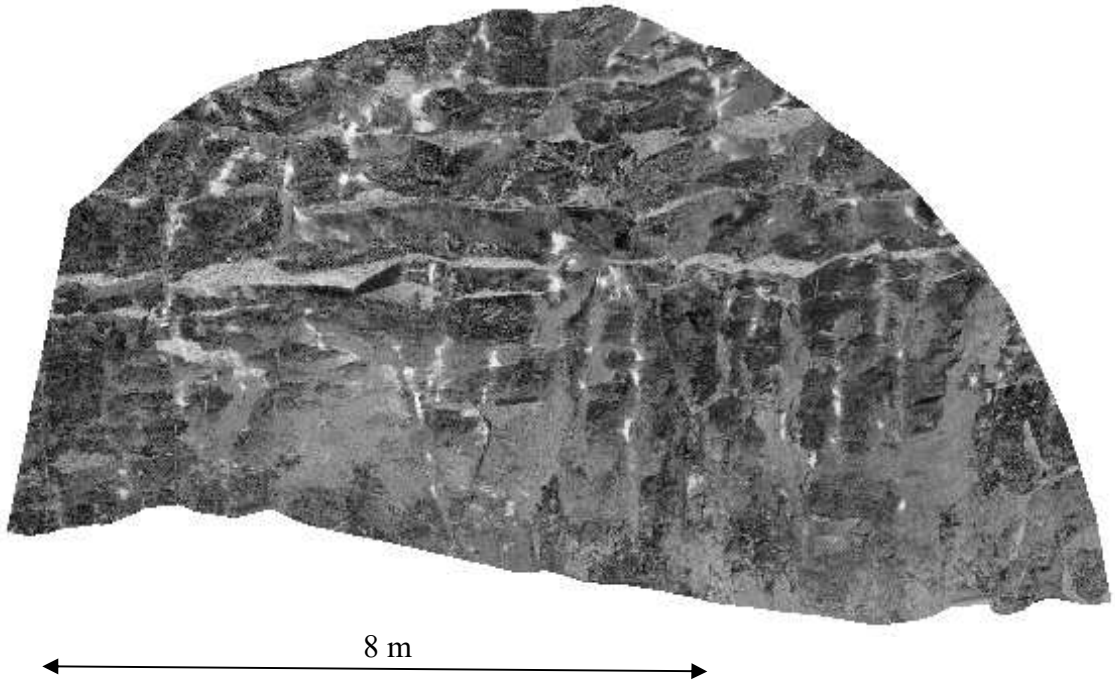


Figure 13: 3D point cloud model of the studied area along the underground tunnel face case study.

3.3.2 Manual digital mapping

Manual mapping was performed on the tunnel face using CloudCompare's virtual compass tool. Figure 14 presents the resulting equal-angle lower hemisphere stereonet plot, based on 40 manually mapped orientations. Four distinct joint sets were identified from the topographical model. The specific orientation characteristics for each set are summarized in Table 3.

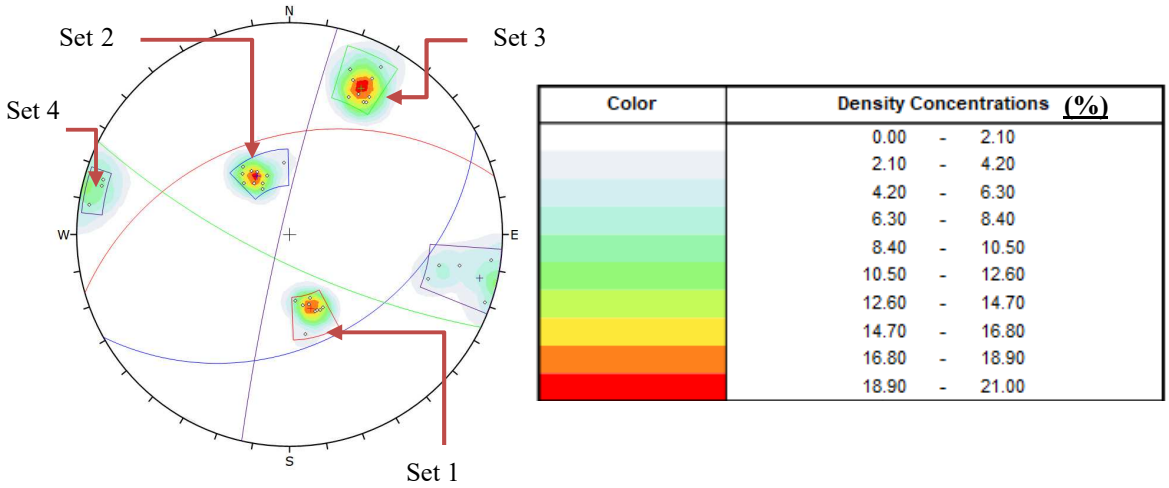


Figure 14: Plotted orientation of manually collected data from LiDAR scan on stereonet for the underground tunnel case study.

Table 3: Orientation of three joint sets identified from manual mapping in the underground tunnel case study.

	Dip (°)	Dip Direction (°)	Fisher's k
Set 1	40.13	344.09	158.1
Set 2	35.11	150.93	115.7
Set 3	75.43	206.30	192.0
Set 4	84.67	283.24	50.0

3.3.3 Scale-dependent criteria

Figure 15 plots the $\xi_R - r$ trend for the tunnel face. A consistent trend is measured along the range defined by lower and upper scale criteria $r_1^* = 0.02$ m and $r_2^* = 0.5$ m. It is surmised that this range, again, corresponds to the scale dominated by structural elements observed along the rock surface. These values were implemented in *ML* clustering for automated structural mapping.

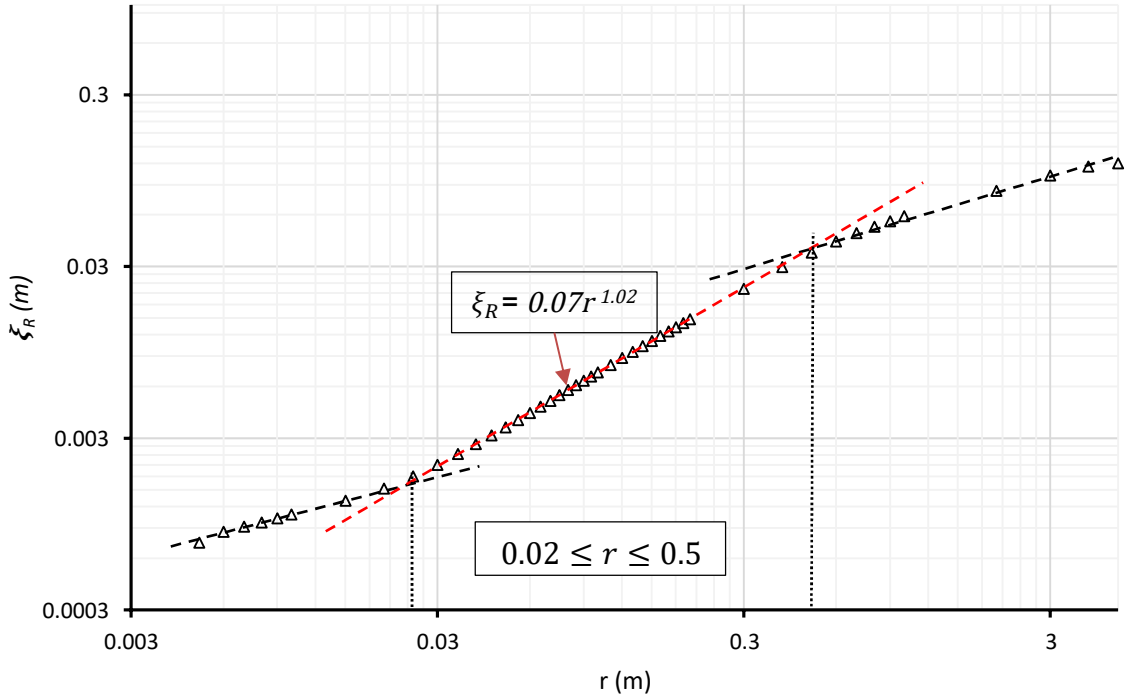


Figure 15: Log–log plot of $\xi_R - r$ for the underground tunnel case study, illustrating scale-dependent variation in roughness.

3.3.4 Automated clustering results

K-means clustering was performed on the tunnel face following the method presented above. Roughness and normal orientation were computed as input clustering parameters for radii of measurements r_R and r_N at $r_1^* = 0.02$ m, $r_2^* = 0.5$ m and $\tilde{r} = 0.15$ m (i.e. logarithmic mean of r_1^* and r_2^*). Combinations of input parameters for clustering were evaluated for optimal

clustering performance. Pertinent sample results are presented in Table 4 with computed *DBI* and average deviation with respect to manual mapping.

Table 4: Clustering performance for samples of tested input combinations with varying radii of measurements.

Normal radii r_N (m)	Roughness radii r_R (m)	*DBI*	Average deviation (°)
[0.02, 0.15, 0.5]	[0.02, 0.15, 0.5]	1.0736	5.17
[0.02, 0.15]	[0.02, 0.15, 0.5]	1.0386	3.48
[0.02]	[0.02, 0.15, 0.5]	1.0391	3.61
[0.02, 0.15, 0.5]	[0.02, 0.15]	1.0892	6.16
[0.02, 0.15, 0.5]	[0.02]	1.0911	7.29

Results presented in Table 4 highlight that clustering performance consistently performs better with all three r_R inputs considered. Similar to the previous example, clustering performance with $r_N > \tilde{r}$ presents low performance. The best configuration was achieved with $r_N = r_1^*$ and $r_R \in [r_1^*, r_2^*, \tilde{r}]$. The silhouette method indicated an optimal number of clusters of $k = 5$.

Figure 16 displays the clustering results for the optimal input combination. Four principal structural sets (Sets 1–4) are distinguished, appearing as continuous regions that align with the dominant orientations of the rock face. A fifth cluster is identified, associated with noisy areas and irregular surfaces along block boundaries.

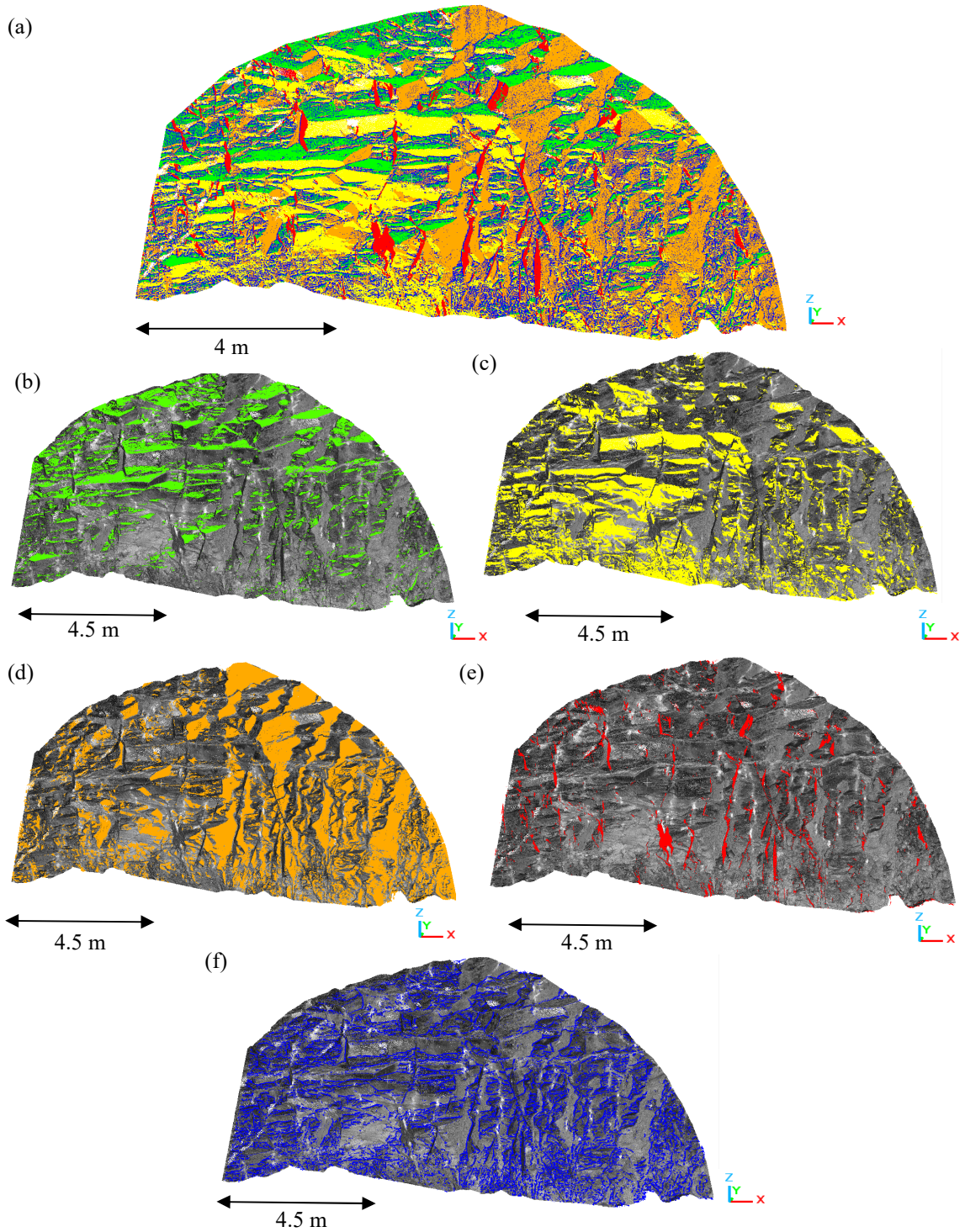


Figure 16: Clustering output at $r_N = 0.02$ and *$r_R = 0.01, 0.11, 0.4$ m for the underground tunnel case study, showing (a) all clusters combined, (b) set 1, (c) set 2, (d) set 3, (e) set 4, and (f) the edge cluster. Clusters (b–e) correspond to dominant orientations, while (f) captures noisy and irregular surfaces along block boundaries.*

Normal orientations were computed again after the fifth was filtered out of the data set. Density-based filter was subsequently applied to the pole plots to remove noise and scatter points. Figure 17 illustrates the resulting improvement in cluster clarity and structural interpretability. Fisher's K value and deviation from the manual baseline are computed for the four joint sets. The

selected configuration, combining a normal radius of r_N = 0.02 m with roughness radii of rR = 0.02, 0.15, and 0.5 m, yields a combined orientation deviation of 3.58° together with a Fisher's *K* value of 25.83, indicating strong clustering concentration and close agreement with manually mapped structural orientations.

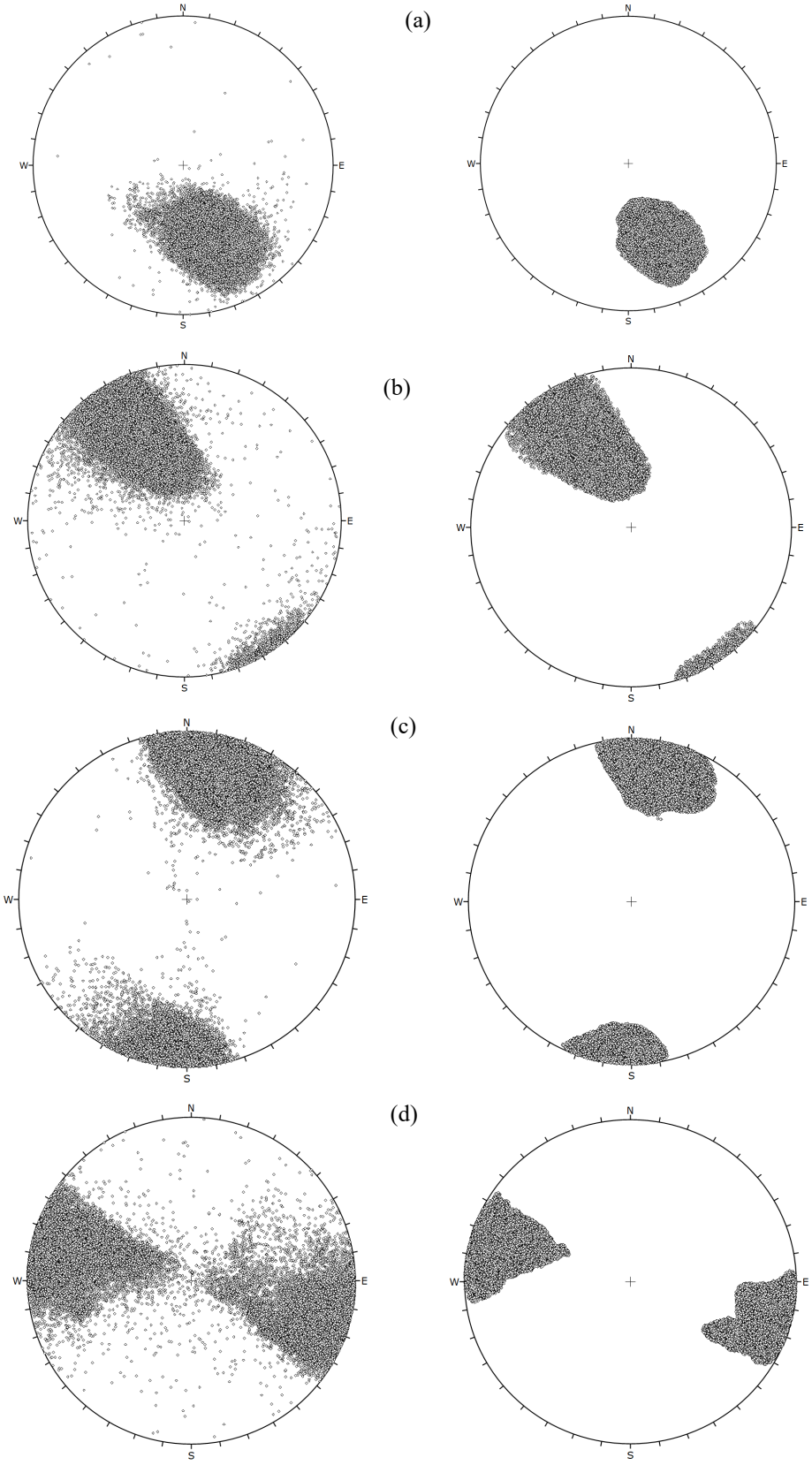


Figure 17: Effect of density-based filtering on orientation data for the underground tunnel case study: (a) set 1, (b) set 2, (c) set 3, and (d) set 4. Left: raw pole plots before filtering; right: results after applying a density-based filter.

Figure 18 presents a comparison between the mean orientations from manual mapping and those derived from the best-performing clustering configuration. The four dominant sets detected by the automated workflow closely match the manually mapped reference orientations, exhibiting only minor angular differences.

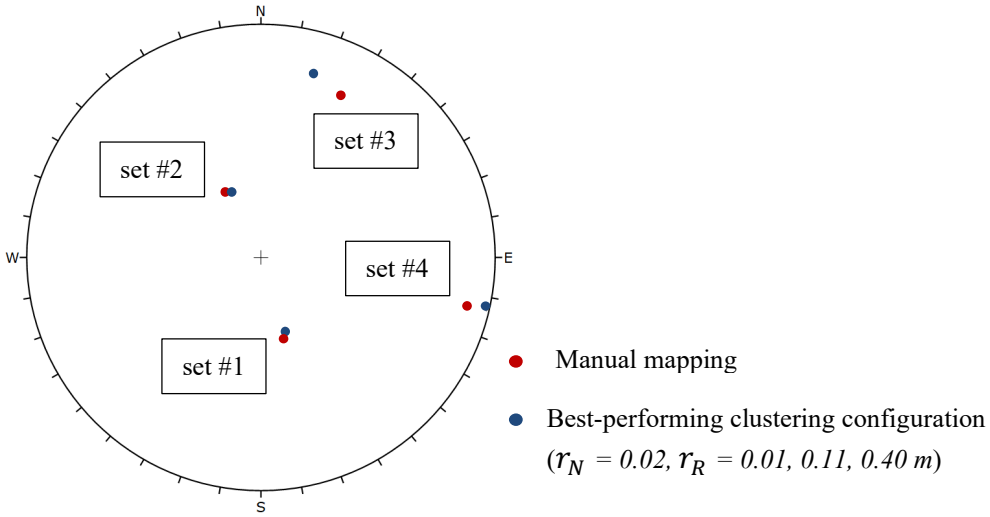


Figure 18: *Comparison of mean orientations from manual mapping and automated clustering in the underground tunnel case study.*

Individual stereonet density contours for the four joint sets identified as the best performing configuration are shown in Figure 19 (a–d), while Figure 19 (e) displays a combined stereonet constructed using random subsamples of 200,000 poles per cluster.

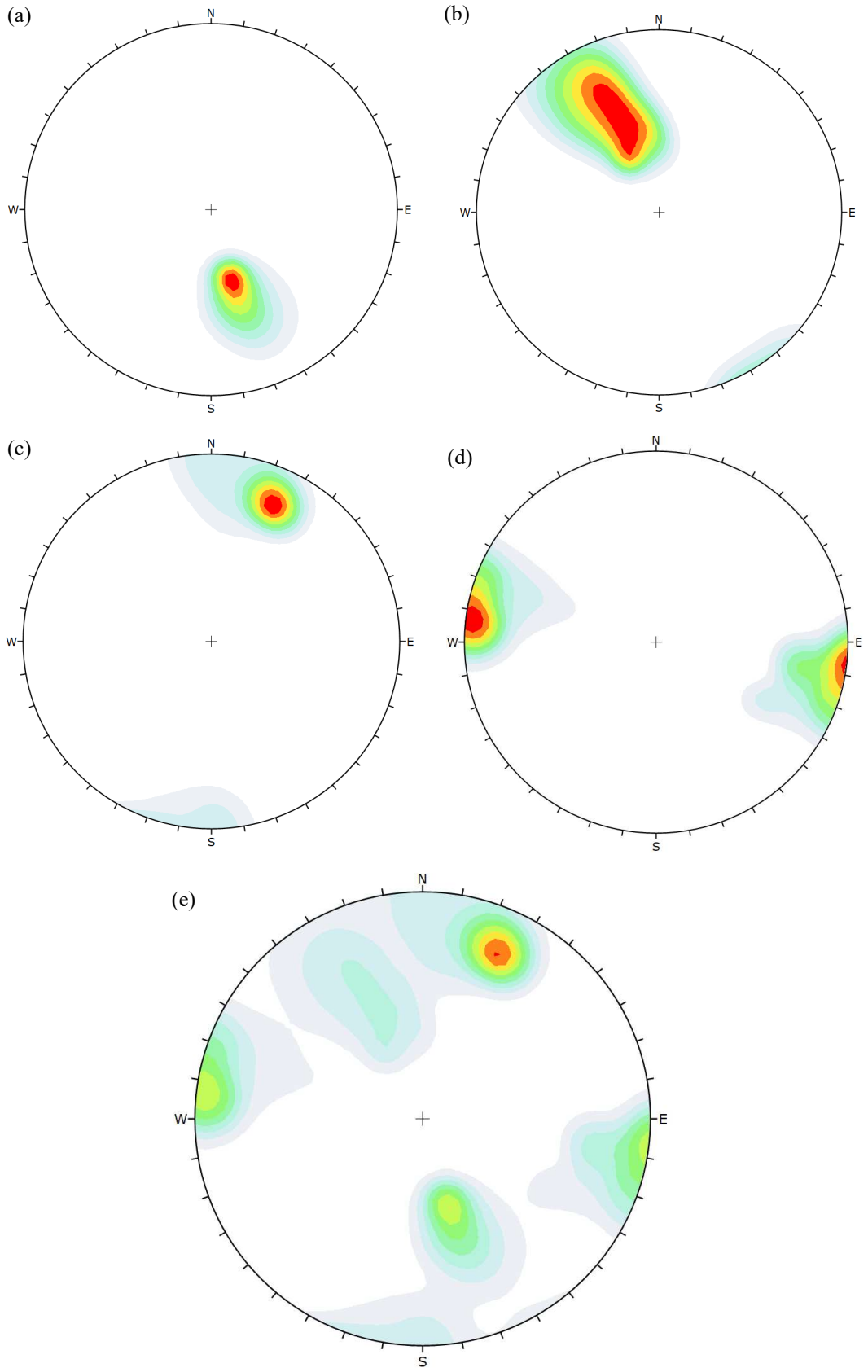


Figure 19: Stereonet contour plots of clustered discontinuity orientations for the underground tunnel case study: (a–d) individual joint sets and (e) combined stereonet with equalized point counts.

4 DISCUSSION

The workflow presented in this work introduced objective and reproducible topographical criteria for automated clustering inputs. The analyses carried out for these investigations were conducted on office laptops with 16 GB of RAM. A complete analysis could be performed under 5 minutes. It is acknowledged that clustering performance could maybe be improved by increasing drastically and iteratively the number of input parameters and radii of measurement for clustering. Such an approach would however be very computationally intensive due to the inherent redundancy involved with large number of clustering inputs without objective rationale for consideration. It would also considerably limit reproducibility of the analysis. It is further asserted that iterative generation of large clustering input layers may incorporate erroneously large scale (i.e $r > r_2^*$) and consequently apply clustering to textural features that do not correspond to structural elements along the exposed outcrop. The proposed criteria offer robust and intuitive input parameters with clear physical rationale.

The application of clustering results to subsequent analysis such as explicit kinematic analysis or transposed DFN modelling can be limited to some extent by the statistical variably inherently present in the results obtained. Additional post-processing may be considered to facilitate facet grouping (e.g. connected Component (*CC*) filtering, see also Rusu and Cousins, 2011) Figure 20 and Figure 21 illustrate *CC* filtering applied to case study 1 and 2. The filtering process effectively removes small, disconnected patches and yields prevalent surfaces. Practitioners are cautioned to apply engineering judgments to further determine the statistical validity of the results before and after such filtering. The resulting statistical properties of the stereographic projections are smoothed by post-facet clustering analysis. It is asserted that the analysis presented without CC filtering offers a more robust scheme to represent statistical variability for structural joint orientation. Facet clustering may be implemented subsequently for specific applications with proper engineering judgment.

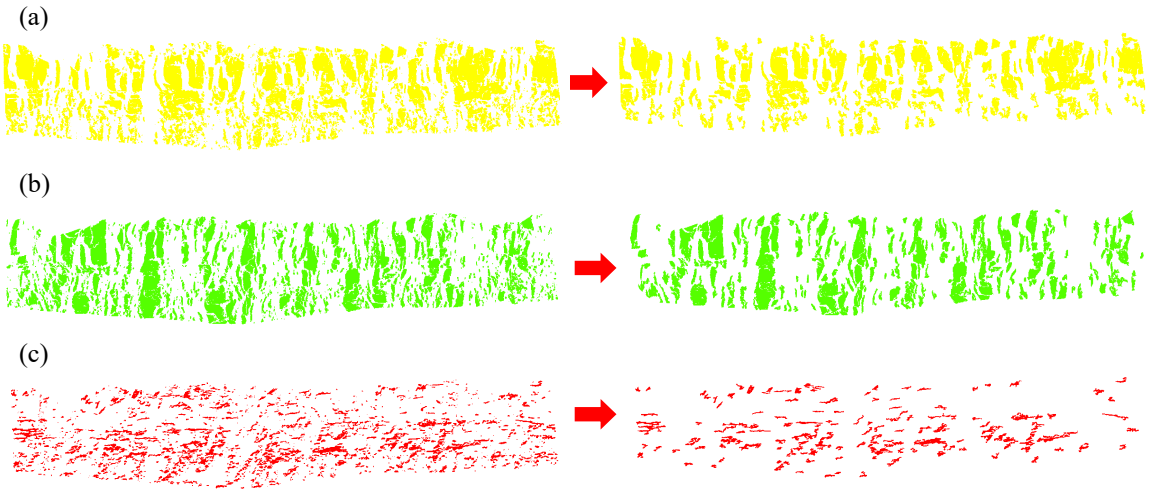


Figure 20: Clusters from the quarry case study before (left) and after (right) applying the CC filter: (a) set 1, (b) set 2, and (c) set 3.

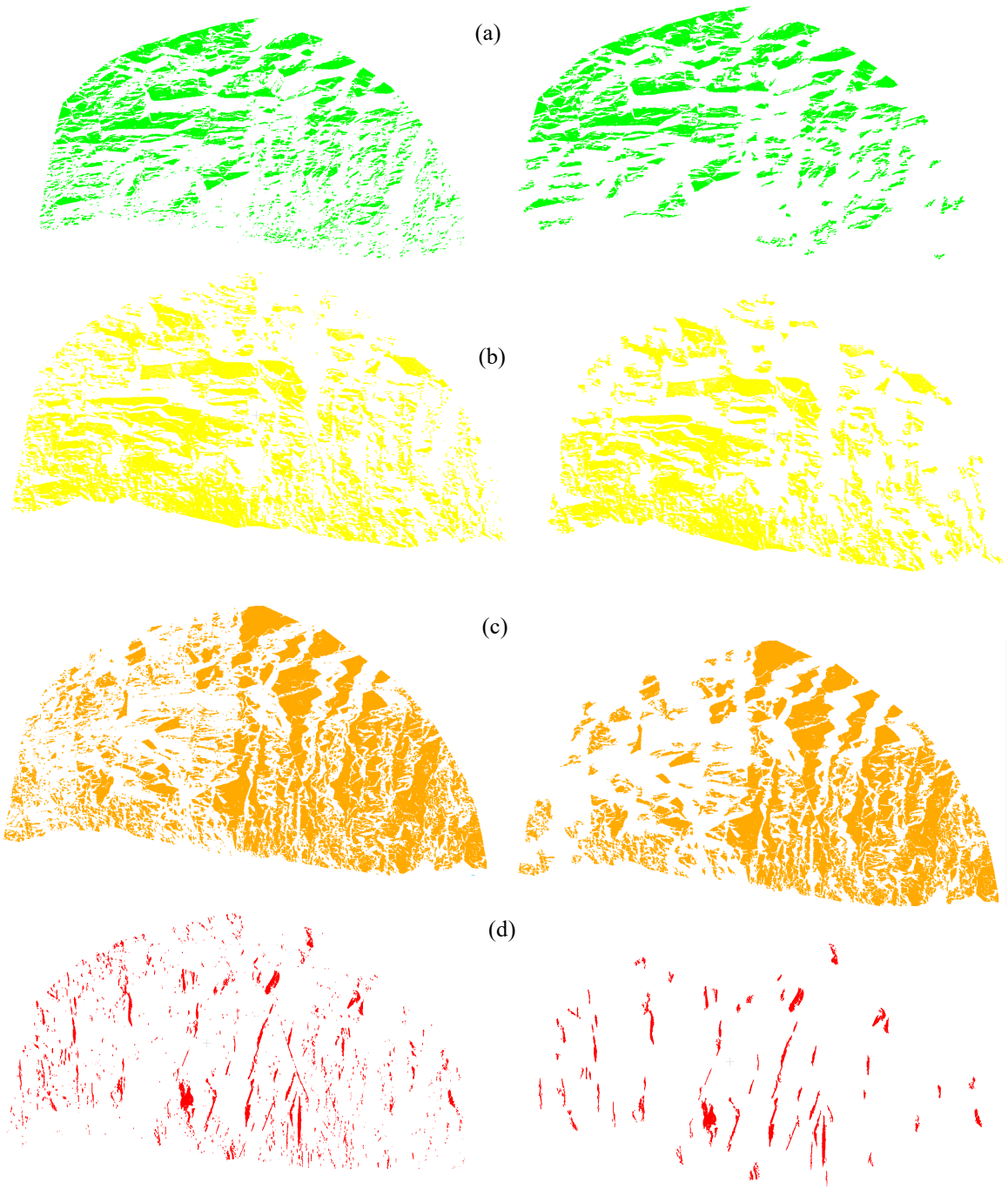


Figure 21: Clusters the underground tunnel case study shown before (left) and after (right) applying the CC filter: (a) set 1, (b) set 2, (c) set 3, and (d) set 4.

Validity of the presented clustering workflow is influenced by point cloud resolution. Point spacing were in the considered case examples ranged between 0.5 to 1 cm on average. Such resolution can yield sizeable data sets and may not be achievable with *UAV*-mounted photogrammetry technologies. It is demonstrated next that the proposed approach can be applied to a lower cloud resolution, at the cost of coarser statistical variability (e.g. lower Fisher's *K* parameter).

The influence of point cloud density is exemplified by revisiting the tunnel face case study. The original point cloud featured and average point spacing of 0.005 m. The model was artificially

subsampled to generate datasets with average spacings of 0.01 m and 0.02 m. Centimetric pixel size is certainly achievable with moder photogrammetry solutions.

Figure 22 plots the characteristic $\xi_R - r$ relationship for the tunnel face model. The plot presents the original and subsample data sets. It is noted from the figure that similar scale dependent trends are featured. The lower bound scale r_1^* appears to be inversely proportional to point spacing. Figure 23 plots the stereographic projections following the clustering framework presented in this research applied to the subsampled data sets. The results are consistent with previous sections (Figure 19 (e)), although the distributions appear coarser. This increased dispersion corresponds to lower Fisher's *K* values of 21.76 and 17.34 for 0.01 m and 0.02 m subsampled point spacing respectively. These results demonstrate that the proposed method remains applicable to lower-resolution point cloud datasets, albeit with increased spread in statistical precision of the structural orientations reported.

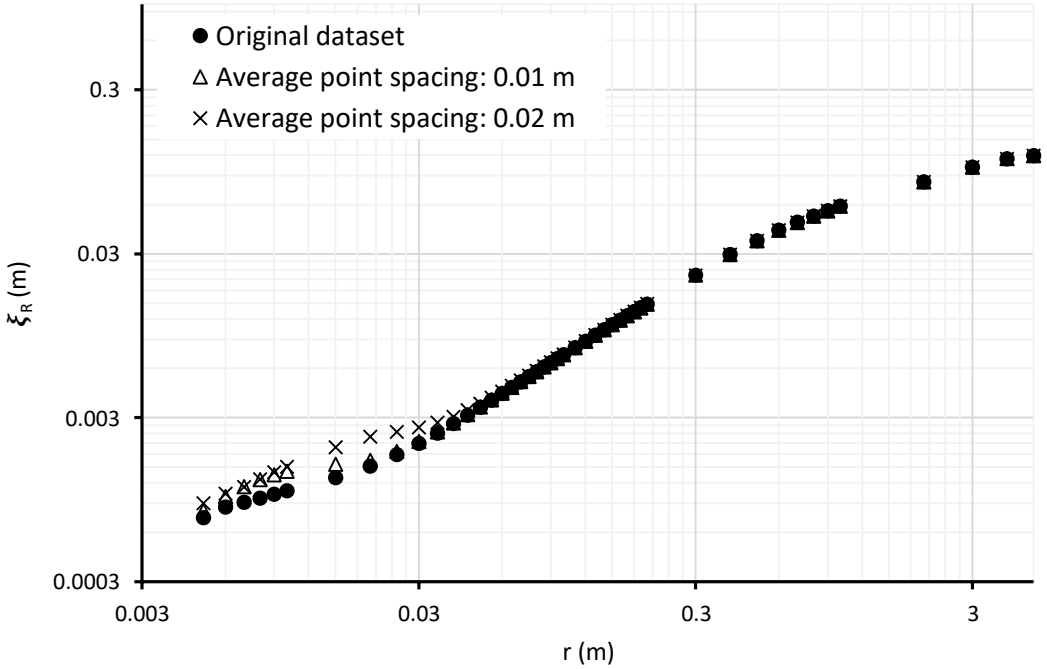
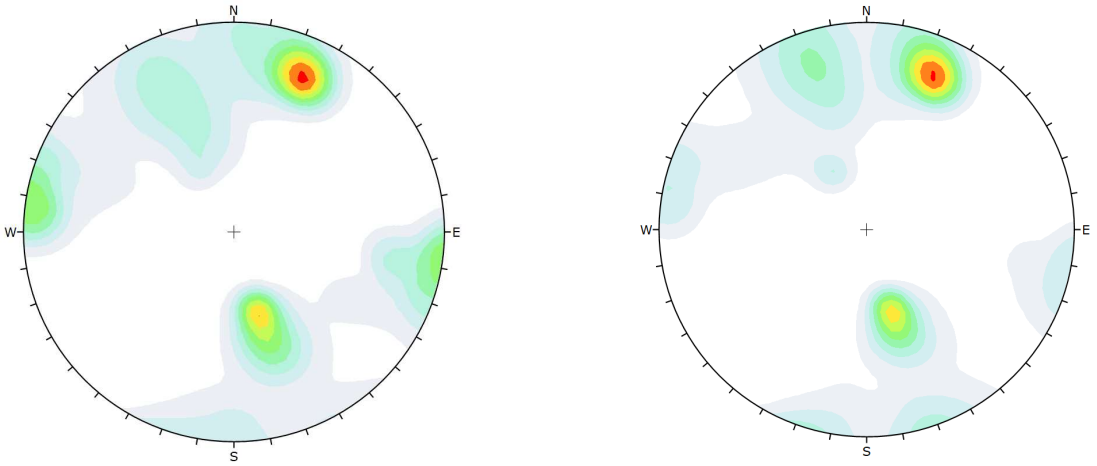


Figure 22: Scale dependent trend $\xi_R - r$ for case study 2 and subsampled cloud data sets.



(a) (b)

Figure 23: Resulting stereographic projections for the tunnel case study after subsampling the point cloud to average point spacing of (a) 0.01 m and (b) 0.02 m.

5 CONCLUSION

This research presented an *ML*-assisted workflow for automated structural mapping of rock faces. The developed framework establishes coherent and objective topographical scale criteria to support systematic and reproducible automated clustering of structural elements. The developed workflow can be summarized as follows:

1. Characteristic scale-dependent representation of topographical roughness for the exposed rock outcrop with $\xi_R - r$ relationship.
2. Identification of topographical scale criteria representative of structural element texture r_1^* and r_2^*.
3. Computation of surface normal $\vec{N}$ with radius of measurement $r_N = r_1^*$, and topographical roughness R with radius of measurement $r_R \in [r_1^*, r_2^*, \tilde{r}]$.
4. Determine optimal number of clusters with silhouette method.
5. Perform clustering to generate $M + 1$ clusters for M number of natural discontinuity sets.
6. Filter out cluster $M + 1$.
7. Recompute topographical normal for the M clusters, and represent in a sterographical projection.
8. Apply density-based filtering with angular error for each clusters in the stereographicl plots.

This work demonstrated that the presented methodology can be applied to models of varying resolutions and benchmarked advantageously to manual digital mapping. Computational time for this analysis was proven manageable while the process itself is fundamentally reproducible. The leveraged topographical scale criteria offer robust input parameters related to physical and measurable characteristics derived from fractals theory.

6 REFERENCES

Abellán, A., Vilaplana, J.M. and Martínez, J. (2006) "Application of a long-range Terrestrial Laser Scanner to a detailed rockfall study at Vall de Núria (Eastern Pyrenees, Spain)," *Engineering Geology*, 88(3–4), pp. 136–148. Available at: https://doi.org/10.1016/j.enggeo.2006.09.012.

Assali, P. *et al.* (2014) "Surveying and modeling of rock discontinuities by terrestrial laser scanning and photogrammetry: Semi-automatic approaches for linear outcrop inspection," *Journal of Structural Geology*, 66, pp. 102–114. Available at: https://doi.org/10.1016/j.jsg.2014.05.014.

Aubertin, J.D. (2025) "Enhanced Scan Registration Through Recursive ICP and Bias Filtering for Geoengineering Applications," *Remote Sensing in Earth Systems Sciences*, 8(2), pp. 718–732. Available at: https://doi.org/10.1007/s41976-025-00214-7.

Battulwar, R. *et al.* (2021) "A state-of-the-art review of automated extraction of rock mass discontinuity characteristics using three-dimensional surface models," *Journal of Rock Mechanics and Geotechnical Engineering*, 13(4), pp. 920–936. Available at: https://doi.org/10.1016/j.jrmge.2021.01.008.

Bentley, J.L. (1975) "Multidimensional binary search trees used for associative searching," *Commun. ACM*, 18(9), pp. 509–517. Available at: https://doi.org/10.1145/361002.361007.

Berti, M., Corsini, A. and Daehne, A. (2013) "Comparative analysis of surface roughness algorithms for the identification of active landslides," *Geomorphology*, 182, pp. 1–18. Available at: https://doi.org/10.1016/j.geomorph.2012.10.022.

Chen, J. *et al.* (2021) "Towards semi-automatic discontinuity characterization in rock tunnel faces using 3D point clouds," *Engineering Geology*, 291. Available at: https://doi.org/10.1016/j.enggeo.2021.106232.

Chen, J. *et al.* (2023) "A critical review of automated extraction of rock mass parameters using 3D point cloud data," *Intelligent Transportation Infrastructure*. Oxford University Press. Available at: https://doi.org/10.1093/iti/liad005.

Chen, J., Zhu, H. and Li, X. (2016) "Automatic extraction of discontinuity orientation from rock mass surface 3D point cloud," *Computers & Geosciences*, 95, pp. 18–31. Available at: https://doi.org/10.1016/j.cageo.2016.06.015.

Drews, T. *et al.* (2018) "Validation of fracture data recognition in rock masses by automated plane detection in 3D point clouds," *International Journal of Rock Mechanics and Mining Sciences*, 109, pp. 19–31. Available at: https://doi.org/10.1016/j.ijrmms.2018.06.023.

FARO Technologies, Inc. (2015) *FARO Focus3D X 130 laser scanner*. Lake Mary, FL, USA.

Fisher, R.A. (1953) "Dispersion on a sphere," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 217(1130), pp. 295–305. Available at: https://doi.org/10.1098/rspa.1953.0064.

Gallant, M.J. and Marshall, J.A. (2016) "Automated rapid mapping of joint orientations with mobile LiDAR," *International Journal of Rock Mechanics and Mining Sciences*, 90, pp. 1–14. Available at: https://doi.org/10.1016/j.ijrmms.2016.09.014.

Ge, Y. *et al.* (2018) "Automated measurements of discontinuity geometric properties from a 3D-point cloud based on a modified region growing algorithm," *Engineering Geology*, 242, pp. 44–54. Available at: https://doi.org/10.1016/j.enggeo.2018.05.007.

Gigli, G. and Casagli, N. (2011) "Semi-automatic extraction of rock mass structural data from high resolution LIDAR point clouds," *International Journal of Rock Mechanics and Mining Sciences*, 48(2), pp. 187–198. Available at: https://doi.org/10.1016/j.ijrmms.2010.11.009.

Girardeau-Montaut D. (2024) "CloudCompare."

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

Hurst, H.E. (1951) "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, 116(1), pp. 770–799.

Kreslavsky, M.A. and Head III, J.W. (1999) "Kilometer-scale slopes on Mars and their correlation with geologic units: Initial results from Mars Orbiter Laser Altimeter (MOLA) data," *Journal of Geophysical Research: Planets*, 104(E9), pp. 21911–21924. Available at: https://doi.org/10.1029/1999JE001051.

Kulatilake, P. and Um, J. (1997) "Requirements for accurate quantification of self-affine roughness using the roughness–length method," *International Journal of Rock Mechanics and Mining Sciences*, 34(3–4), pp. 166.e1-166.e15. Available at: https://doi.org/10.1016/S1365-1609(97)00202-5.

Lato, M. *et al.* (2013) "Rock bench: Establishing a common repository and standards for assessing rockmass characteristics using LiDAR and photogrammetry," *Computers & Geosciences*, 50, pp. 106–114. Available at: https://doi.org/10.1016/0148-9062(76)90818-4.

Lato, M.J. and Vöge, M. (2012) "Automated mapping of rock discontinuities in 3D lidar and photogrammetry models," *International Journal of Rock Mechanics and Mining Sciences*, 54, pp. 150–158. Available at: https://doi.org/10.1016/j.ijrmms.2012.06.003.

MacQueen, J. (1967) "Some methods for classification and analysis of multivariate observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. University of California press, pp. 281–298.

Priest, S.D. and Hudson, J.A. (1976) "Discontinuity spacings in rock," *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 13(5), pp. 135–148. Available at: https://doi.org/10.1016/0148-9062(76)90818-4.

Riquelme, A.J. *et al.* (2014) "A new approach for semi-automatic rock mass joints recognition from 3D point clouds," *Computers & Geosciences*, 68, pp. 38–52. Available at: https://doi.org/10.1016/j.cageo.2014.03.014.

Rocscience Inc. (2006) "Dips." Toronto, ON, Canada.

Rusu, R.B. and Cousins, S. (2011) "3D is here: Point Cloud Library (PCL)," *2011 IEEE International Conference on Robotics and Automation*, pp. 1–4. Available at: https://doi.org/10.1109/ICRA.2011.5980567.

Shepard, M.K. *et al.* (2001) "The roughness of natural terrain: A planetary and remote sensing perspective," *Journal of Geophysical Research E: Planets*, 106(E12), pp. 32777–32795. Available at: https://doi.org/10.1029/2000JE001429.

Singh, S.K. *et al.* (2022) "Automated rock mass discontinuity set characterisation using amplitude and phase decomposition of point cloud data," *International Journal of Rock Mechanics and Mining Sciences*, 152. Available at: https://doi.org/10.1016/j.ijrmms.2022.105072.

Singh, S.K., Raval, S. and Banerjee, B.P. (2021) "Automated structural discontinuity mapping in a rock face occluded by vegetation using mobile laser scanning," *Engineering Geology*, 285, p. 106040. Available at: https://doi.org/10.1016/j.enggeo.2021.106040.

Slob, S. *et al.* (2005) "Method for Automated Discontinuity Analysis of Rock Slopes with Three-Dimensional Laser Scanning," *Transportation Research Record: Journal of the Transportation Research Board*, 1913(1), pp. 187–194. Available at: https://doi.org/10.1177/0361198105191300118.

Stewart, J. (2003) *Calculus: Early Transcendentals*. 6th ed. Belmont, California: Brooks/Cole.

Sturzenegger, M. and Stead, D. (2009) "Quantifying discontinuity orientation and persistence on high mountain rock slopes and large landslides using terrestrial remote sensing techniques," *Natural Hazards and Earth System Sciences*, 9(2), pp. 267–287. Available at: https://doi.org/10.5194/nhess-9-267-2009.

Wang, X. *et al.* (2017) "A region-growing approach for automatic outcrop fracture extraction from a three-dimensional point cloud," *Computers & Geosciences*, 99, pp. 100–106. Available at: https://doi.org/10.1016/j.cageo.2016.11.002.

Zhang, P. *et al.* (2018) "Semi-automatic extraction of rock discontinuities from point clouds using the ISODATA clustering algorithm and deviation from mean elevation," *International Journal of Rock Mechanics and Mining Sciences*, 110, pp. 76–87. Available at: https://doi.org/10.1016/j.ijrmms.2018.07.009.

7 APPENDIX

7.1 Geometric features

For a point cloud P, the normal vector of each point p_i is evaluated with the best fitted plane considering the subset P_i of all points located within radius r of p_i. Roughness of a point p_i at radius of measurement r, $R_{r,i}$, is computed as the distance between p_i and the best fitted plane from all points located with radius r of p_i, i.e., $\vec{N}_{r,i}$. A plane is defined by its normal vector $\vec{N}$ and reference point *d* such that:

$$\vec{N} \cdot \vec{P}i + d = 0 \quad [1]$$

$$\vec{N} = [\, n_x, n_y, n_z] \quad [2]$$

Where n_x, n_y, and n_z are the normal vector components.

For Cartesian coordinates with *X*, *Y*, and *Z* axes, and for a point cloud with its shortest span along the *Z*-axis, the normal vector $\vec{N}$ of a best-fit plane and parameter *d* can be determined by solving the equations below:

$$D = \sum X^2 . \sum Y^2 - (\sum XY)^2 \quad [3]$$

$$D.n_x = \sum YZ . \sum XY - \sum XZ . \sum Y^2 \quad [4]$$

$$D.n_y = \sum XY . \sum XZ - \sum X^2 . \sum YZ \quad [4]$$

$$-d = n_x x_c + n_y y_c + n_z z_c \quad [6]$$

where (x_c, y_c, z_c) represents the centroid of the point cloud.

Point normal vectors are computed using Principal Component Analysis (*PCA*) applied to local neighborhoods defined around each point. For a given point, all neighboring points within a spherical radius *r* are selected, and the covariance matrix of the neighborhood is constructed as:

$$C = \frac{1}{n}\sum_{i=1}^{n}(p_i - \bar{p})(p_i - \bar{p})^T \quad [7]$$

where p_i is the position vector of the ith neighboring point, n is the number of points in the neighborhood, and $\bar{p}$ is the centroid of the neighborhood.

Eigenvalue decomposition of *C* yields three eigenvalues and associated eigenvectors. The eigenvector corresponding to the smallest eigenvalue defines the normal vector of the locally best-fit plane and is assigned as the point normal $\vec{N}$.

For a given point $p_i = [x_i, y_i, z_i]$, the roughness $R_{r,i}$ (m) is defined as the orthogonal distance from the point to a best-fit plane derived from its local neighborhood, defined by a user-specified radius *r* (Kreslavsky and Head III, 1999; Shepard *et al.*, 2001; Berti, Corsini and Daehne, 2013). Roughness is computed as:

$$R_{r,i} = \frac{n_{xi}x_i + n_{yi}y_i + n_{zi}z_i}{\sqrt{n_{xi}^2 + n_{yi}^2 + n_{zi}^2}} \quad [8]$$

where n_{xi}, n_{yi}, and n_{zi} are the normal vector components of the ith pole measured at the radius of *r*.

In this work, the orientation of structures is represented by surface dip *D* (°) and dip direction *DD* (°) (Priest, 1993):

$$DD = tan^{-1}\left(\frac{n_x}{n_y}\right) + Q \quad [5]$$

$$D = tan^{-1}\left(\frac{n_z}{\sqrt{n_x^2 + n_y^2}}\right) \quad [6]$$

where Q is a quadrant correction such that $0 \geq D \geq 360$, and the unit normal vector $\vec{N} = [n_x, n_y, n_z]$.

7.2 Mean pole orientation for a discontinuity set

To quantify the representative orientation of each cluster, the mean orientation is calculated from the set of normalized pole vectors. Let $\vec{N}_{r,i}$= (n_{xi}, n_{yi}, n_{zi}) denote the unit normal vector of the ith pole, where $i=1, 2, ..., n$. The orientation matrix M is then constructed as the average of the outer products of all normalized vectors:

$$M = \frac{1}{n}\sum_{i=1}^{n} \vec{N}_{r,i}\vec{N}_{r,i}^{T} \quad [9]$$

Eigenvalue decomposition of M yields three eigenvalues (λ_1, λ_2, λ_3) and corresponding orthogonal eigenvectors (e_1, e_2, e_3). The eigenvector associated with the smallest eigenvalue represents the mean pole orientation, since it corresponds to the axis of least variance in the distribution. This vector is then converted back into dip and dip direction values to express the mean orientation of the corresponding plane.

7.3 Fisher distribution

This parameter provides a statistical measure of how tightly clustered a set of orientation vectors is on the unit sphere, analogous to the variance in linear statistics. First, the resultant vector length (VL) is computed as the magnitude of the vector sum of all normalized poles:

$$VL = \sqrt{(\sum_{i=1}^{n} n_{xi})^2 + (\sum_{i=1}^{n} n_{yi})^2 + (\sum_{i=1}^{n} n_{zi})^2} \quad [10]$$

where n is the total number of vectors. Using this, Fisher's concentration parameter is estimated as:

$$K = \frac{n-1}{n-R} \quad [11]$$

Higher values of K indicate greater concentration of vectors around the mean orientation (i.e., a well-defined joint set), while smaller values indicate more dispersed orientations. This statistic therefore provides a robust means of evaluating the reliability of the mean orientation derived for each cluster.