Implementation of multivariate clustering methods for characterizing discontinuities data from scanlines and oriented boreholes [*]

Wei Zhou[a], Norbert H. Maerz[b,*]

[a] Department of Mining and Geological Engineering, University of Alaska Fairbanks, Fairbanks, AK 99775, USA
[b] Rock Mechanics and Explosives Research Center, University of Missouri-Rolla, 1006 Kingshighway, Rolla, MO 65409-0660, USA

Abstract

In geological engineering, discontinuities are typically analyzed by grouping (clustering) them into subsets based on similar orientations, and then characterizing each set in terms of position, spacing, persistence, roughness and other parameters. Multivariate analysis can be used to incorporate some of these other parameters directly into the cluster analysis. The implementation of four methods of cluster analysis that consider orientation, spacing and roughness are described here: nearest neighbor, k-means, fuzzy c-means, and vector quantization. The net result is a better grouping of discontinuities, so that members of a subset might be more uniform in terms of mechanical or hydrological properties. This paper presents the implementation of this analysis in a Windows[TM] based program CYL that also serves as a graphical visualization tool. © 2002 Elsevier Science Ltd. All rights reserved.

1. Introduction

The behavior of a rock mass is largely governed by the discontinuities within that rock mass. Understanding the nature of the discontinuities (joints, fractures, bedding planes, faults, and other breaks in the continuity of the rock) is a fundamental requirement of discontinuous rock mass characterization. With a few exceptions, the engineering properties of most rock masses are influenced to some extent or another by discontinuities. Whereas we understand much about the mechanical properties of intact solid rock, our understanding of discontinuous rock is significantly less well developed. Among the most important parameters describing the discontinuities are orientation, spacing, persistence (length), roughness,

aperture and infilling materials (Hudson and Priest, 1979; Priest, 1993; Hudson and Harrison, 1997; Mauldon et al., 2001). Most of these can be measured from oriented borehole core logs, or borehole video logs.

There are many papers on methods of discontinuity analysis (Mahtab et al., 1972; Priest and Hudson, 1981; Baecher, 1983; Mahtab and Yegulap, 1984; Kulatilake and Wu, 1984; Maerz et al., 1990; Tsoutrelis et al., 1990; Willis-Richards and Jupe, 1995; Hudson and Priest, 1979; Mauldon et al., 2001). Hammah and Curran (1998, 2000) have used joint roughness in clustering joint sets. Dershowitz et al. (1998) have used discontinuity orientation and spacing to define structural domains. However, most existing analytical methods tend to underutilize the available data. Usually, discontinuity parameters are analyzed individually, such as grouping the discontinuities into subsets based on orientation only. Consequently, the analyses suffer from the inability to consider more than one parameter at a time. As a result, discontinuities within the same subset may have various mechanical or hydrological properties.

This paper presents an approach and an analytical tool to characterize discontinuities from oriented borehole data. It characterizes discontinuities into subsets according to multiple parameters, such as orientation, spacing, and roughness. During the discontinuity clustering analysis, a number of variables can be treated simultaneously, so that not only the variance but also the covariance is considered. In this way the interactions between variables are taken into account.

Four clustering methods, namely nearest neighbor, k-means, fuzzy c-means and vector quantization methods are adopted to cluster discontinuity data. Among them, the first method uses hierarchical techniques, and the rest use partitioning techniques. Hierarchical techniques perform successive merging or splitting of the data. One of the primary features distinguishing hierarchical techniques from other clustering algorithms is that the allocation of an object to a cluster is irrevocable, that is, once an object joins a cluster it is never removed. Unlike hierarchical clustering techniques, methods that effect a partition of the data do not require that the allocation of an object to a cluster is irrevocable. That is, objects may be reallocated if their initial assignments were indeed inaccurate.

2. Implementation of multivariate clustering analysis

2.1. Basic concepts

The attributes for discontinuity classification are described in International Society of Rock Mechanics

Commission on Standardization of Laboratory and Field Tests (1978), and shown in Fig. 1. These include:

(1) The discontinuity attitude (orientation).
(2) The distance between adjacent discontinuities (spacing).
(3) The physical extent of discontinuities (persistence).
(4) The surface characteristics of the discontinuities (roughness, strength, mineralization and alteration).
(5) The filling material (filling or infilling).

For the purpose of this paper, Barton's Joint Roughness Coefficient (JRC) (Barton and Choubey, 1977) is used. The JRC values range from 0 to 20. A JRC value of 0 represents the most smooth surface, whereas 20 represents the most rough surface.

The conventional analysis is done by grouping the discontinuities into families or sets based on orientation only, and then summarising the other attributes by these groupings. Although this approach works well in situations where the clustering of joints into orientation families is obvious, it breaks down under typical conditions where the geologic structures change, especially in long boreholes. Under these conditions, the joints often do not cluster well into families, and the geological engineer must make arbitrary decisions on how to interpret the data. Consequently, the available data are typically underutilized.

This research is designed around building a new software package for characterizing discontinuities
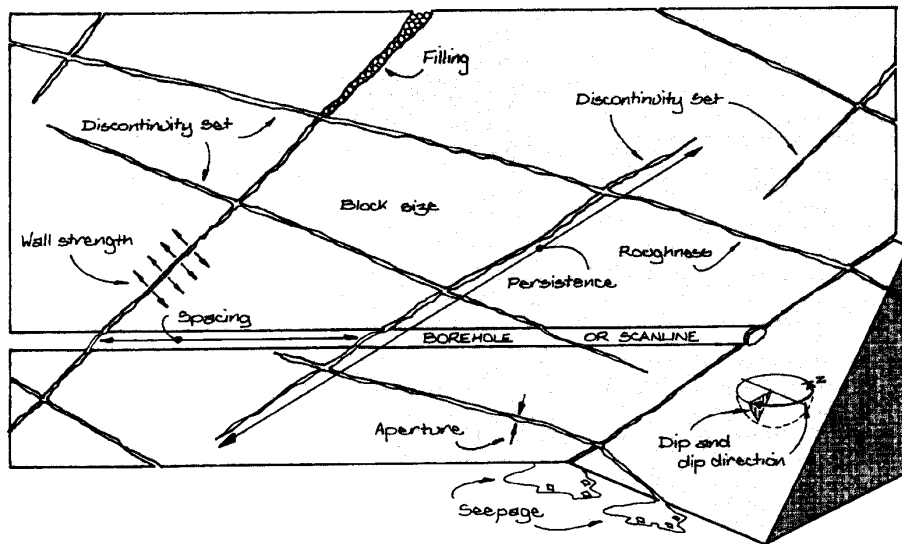

Fig. 1. Schematic drawing of borehole intersecting rock mass shows that discontinuous rock is perceived as set of blocks separated by series of discontinuities (adapted from Hudson, 1989).

based on multivariate-clustering algorithms. The new analysis tool being developed, incorporates the various attributes of the discontinuities in a multivariate analysis, and uses multi-dimensional clustering to identify joint sets and geological/geomechanical domains (statistically homogeneous regions with respect to geological structure). The method makes use of both statistical analysis and three-dimensional visualization tools, in an integrated and automated package of computer algorithms.

In order to offer an intuitive view of discontinuity distribution in a three-dimensional space, different visualization tools are incorporated in the software package. Among them, the three-dimensional stereonet is the most useful one. The concept of the three-dimensional stereo net involves plotting each joint normal on a separate stereonet and stacking them together on top of each other according to their position along the borehole. This idea was first used by Wenk et al. (1987) to represent the pattern of lattice-preferred orientation in deformed rocks. A detailed introduction about this idea can be found in Maerz and Zhou (1999, 2000), and Zhou and Maerz (2001).

2.2. Nearest neighbor clustering method

For the nearest neighbor method (also called single linkage method), the similarity of the discontinuities is evaluated by distance-type measure. Some measures of distance are special cases of the Minkowsi metric, which is defined by (Dillon and Goldstein, 1984)

$$d_{ij} = \left\{ \sum_{k=1}^{p} |X_{ik} - X_{jk}|^r \right\}^{1/r}, \tag{1}$$

where d_{ij} denotes the Minkowsi distance between two objects i and j, X is the value of a parameter, and p is the number of parameters considered in the analysis.

If $r = 2$, then the familiar Euclidean distance between object i and j is obtained (Dillon and Goldstein, 1984; Mardia et al., 1979).

$$Ed_{ij} = \left\{ \sum_{k=1}^{p} |X_{ik} - X_{jk}|^2 \right\}^{1/2}, \tag{2}$$

where Ed_{ij} denotes the Euclidean distance between two joints (i and j) in a p-dimensional space.

If there are n discontinuities, an n-by-n matrix of Euclidean norm is created, where the element in the ith row and jth column represents the distance between individuals i and j induced by the Eq. (2).

At the first stage of the merging process, the two individual objects (for example ith and jth discontinuities) with the smallest entry are joined to form a cluster since they are the closest. Once the cluster is formed, the distance between the cluster and the remaining individuals is computed, and the smaller

value of Ed_{ik} and Ed_{jk} remains in the matrix, $k = 1, 2, ..., n$ $(k \neq i, k \neq j)$.

Next a matrix of Euclidean norms can be constructed; whose elements are inter-individual and inter-group distances. Performing the above merging procedures repeatedly forms new or larger clusters. The number of rows and columns of the matrix is reduced by one every time a merge is made. At the last stage, the two clusters are merged to form a single cluster containing all the individuals.

The following example (Fig. 2) summarizes the various stages at which merging are made. The individuals joined together first have the most similarity and individuals joined together last have least similarity. Fig. 2 is a tree diagram (dendrogram). The numbers in horizontal axis represent the individual objects and the vertical axis represents the element values in a Euclidean distance matrix.

The number of clusters is specified before processing, and the merging is concluded when the appropriate number of clusters is reached. Instead of specifying the number of clusters required, a threshold value of the Euclidean norm may also be specified which indirectly determines the number of clusters. Fig. 3 illustrates the overall process of the nearest neighbor method.

For discontinuity data analysis, variable transformations are first required. There are two different types of discontinuity parameters: vector and scalar. To calculate the mean of the orientation within each discontinuity set, first the normal of every discontinuity is represented by three components (Vx, Vy, and Vz). Then the sums of each of the components of all the discontinuities with the same set are calculated, and finally the three
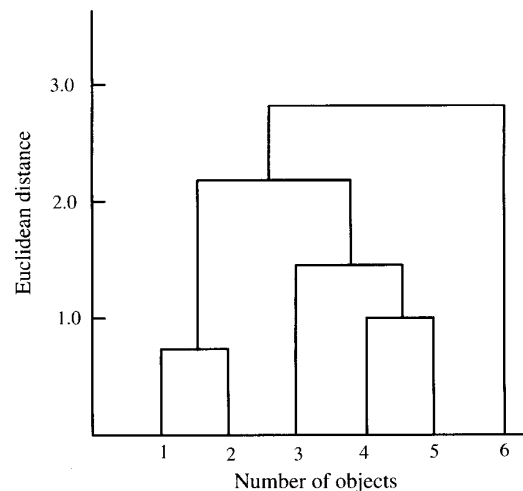

Fig. 2. Example of tree diagram (dendrogram) of nearest neighbor method shows various pairwise merging stages.
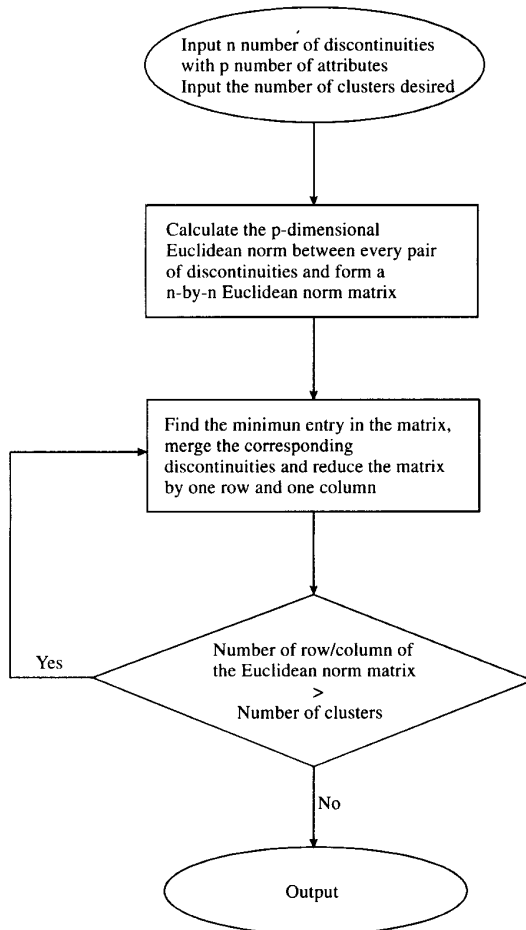
Fig. 3. Flow chart shows overall process of nearest neighbor method.

components vectors are combined to form the mean normal (orientation) for each discontinuity set. For the scalar parameters, arithmetic means are used.

2.3. K-means clustering method

A detailed discussion about K-means method is given in Dillon and Goldstein (1984). The K-means method assumes that the number of the final clusters is known and specified in advance. The basic steps for K-means clustering include: (1) form the initial clusters, (2) allocate the discontinuities to clusters, and (3) reallocate some or all of the discontinuities already clustered if they are inaccurate.

Assume n discontinuities have been measured and consider three attributes (parameters) for each disconti-

nuity. Denote by $x(i, j)$ the value of the ith discontinuity on the jth parameter (variable); $i = 1, 2, ..., n$ and $j = 1, 2, 3$. Let $P(n, K)$ be the partition that results in each of the n discontinuities being allocated to one of the K clusters, and let l denote the lth cluster; $l = 1, 2, ..., K$. K is the number of clusters. The mean of the jth variable in the lth cluster will be denoted by $\bar{x}(l,j)$, and the number of discontinuities belong to the lth cluster by $n(l)$. Based upon the above notation the Euclidean norm, $D[i, l(i)]$, between the ith joint and the center of the mean of lth cluster can be expressed as (Dillon and Goldstein, 1984)

$$D[i, l(i)] = \left(\sum_{j=1}^{p} [x(i,j) - \bar{x}(l,j)]^2 \right)^{1/2}. \tag{3}$$

The error component, $E[P(n, K)]$, of the partition is defined by (Dillon and Goldstein, 1984):

$$E[P(n, K)] = \sum_{i=1}^{n} D[i, l(i)]^2, \tag{4}$$

where $l(i)$ is the cluster that contains the ith discontinuity, and $D[i, l(i)]$ is the Euclidean norm between discontinuity i and the mean of the cluster containing the discontinuity which is defined by Eq. (3).

This procedure is repeated until the minimum error component of partition is determined. The specific procedures are summarized in the following steps:

(1) Form the initial clusters arbitrarily.
(2) Calculate $\bar{x}(l,j)$ the mean of the jth variable over all discontinuities in the lth cluster.
(3) Calculate the Euclidean distance between the ith discontinuity and lth cluster as given by Eq. (2), in which, p is the number of parameters considered.
(4) In order to minimize the error component given by Eq. (5), The following value has to be computed for every discontinuity. It will check to see if any allocation of any discontinuity from one cluster to another results in reduction in the initial cluster error.

$$R_{l(i),l} = \frac{n(l)D(i, l)^2}{n(l) + 1} - \frac{n(l(i))D(i, l(i))^2}{n(l(i)) - 1}. \tag{5}$$

(5) Form the final clusters.

Fig. 4 illustrates the overall process of the K-means method. Data transformations are required as before. For the nearest neighbor method, cylindrical coordinates are used for orientation and spacing variables.

When calculating the Euclidean distance, both original vector and its corresponding "mirror" vector are used. A mirror vector is used to cluster "around the outside" of the steronet, as defined in Fig. 5. The mirror vector is constructed by taking the complementary
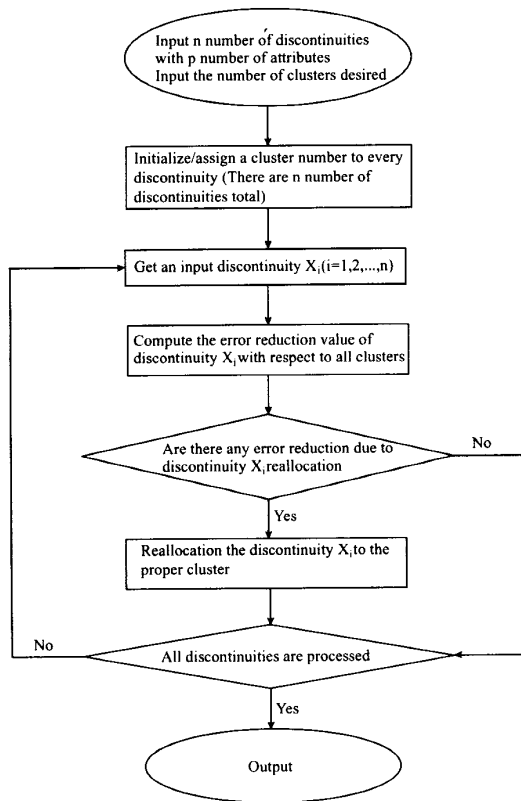
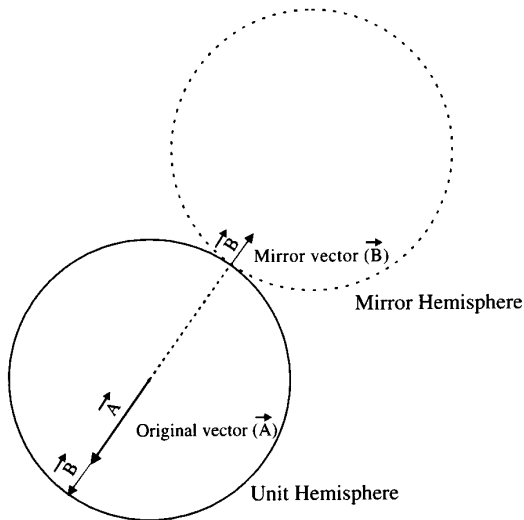Fig. 4. Flow chart shows overall process of k-means method.

Fig. 5. Schematic drawing shows concept of mirror vector.

vector to vector A (magnitude $= 1 - |A|$) and adding this to the unit vector in a direction opposite to vector A. The smaller arc length between the two is used in further analysis. A vector flag is used to trace whether the original vector or the "mirror" vector is chosen. This is important for calculating the mean orientation within each clustering group.

2.4. Fuzzy c-means clustering method

Zadeh (1965) first introduced fuzzy sets as a new way to represent vagueness in everyday life. The "fuzzy" concept has been widely adopted in the fields such as pattern recognition, neural networks, image processing, and expert systems. However, not much attention has paid in the geoscience and rock engineering literatures. Examples of such works include Tao and Peng (1983), Harrison (1992), Hudson and Hudson (1993), Feng et al. (1997), and Hammah and Curran (1998, 2000). Descriptions of this method can be found in several publications (eg. Bezdek, 1981; Gath and Geva, 1989; Xie and Beni, 1991).

The fuzzy c-means method (FCM) is based on minimization of the following objective function (Bezdek, 1981):

$$J_m = \sum_{j=1}^{n} \sum_{i=1}^{c} (u_{ij})^m d^2(X_j, V_i) \quad (c \leqslant N), \tag{6}$$

where u_{ij} is the fuzzy membership, V_i is the cluster centroid, c is the number of clusters, n is the number of discontinuities, $d^2(X_j, V_i)$ is any inner product metric (distance between X_j and V_i), the variable m is the degree of fuzzification that controls the fuzziness of the memberships and is a real number greater than 1. As the values of m become progressively higher, the resulting memberships become fuzzier. No theoretical optimal value for m has been determined, however, $m = 2$ is frequently used by researchers.

The FCM algorithm involves the following steps:

(1) Select c numbers of initial cluster centers. There are different ways to do this. One method is to select the first c data points as an initial guess. Another method would be to select the initial cluster centers randomly. In order to get k well separated initial centers, one data point is picked as the first initial center, then each subsequent initial guess is picked such that its distance from each of those already be chosen is not less than a specified minimum.

(2) Calculate the degree of membership of all discontinuities in all clusters by the following equation (Gath and Geva, 1989):

$$u_{ij} = \frac{(1/d^2(X_j, V_i))^{1/(m-1)}}{\sum_{j=1}^{c} (1/d^2(X_j, V_i))^{1/(m-1)}}. \tag{7}$$

(3) Re-calculate the centroids of the clusters and update the degree of membership from u_{ij} to $\hat{u}_{ij}$. For orientation variables, eigenanalysis is used for finding means, for scalar variables the weighted mean of the variable in a cluster is calculated by (Gath and Geva, 1989)

$$V_i = \frac{\sum_{j=1}^{n}(u_{ij})^m X_j}{\sum_{j=1}^{n}(u_{ij})^m}. \tag{8}$$

(4) Repeat steps (2) and (3) till the following termination criterion is met:

$$\max_{ij}|u_{ij} - \hat{u}_{ij}| < \varepsilon, \tag{9}$$

where ε is a termination criterion between 0 and 1.

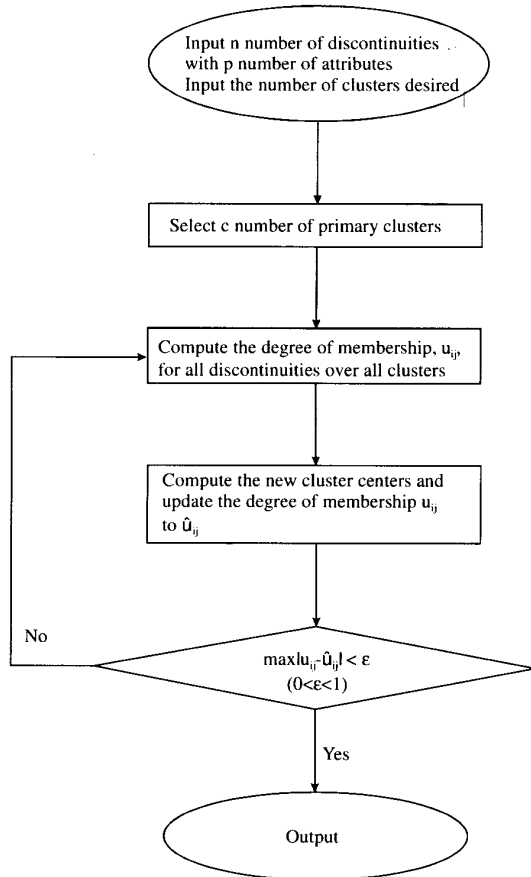Fig. 6 illustrates the overall process of the fuzzy c-means method. To apply fuzzy c-means method to

discontinuity data, the same variable transformations as the nearest neighbor method were applied.

2.5. Vector quantization clustering method

The vector quantization method (VQM) begins with no clusters allocated. The first discontinuity is forced to create a cluster to hold it. After that, with each new discontinuity input, the Euclidean norm between it and any previous cluster is calculated. Once the distance between the current discontinuity and all previous clusters is known, the cluster closest to the input discontinuity may be chosen so that (Pandya and Macy, 1996)

$$|X^{(i)} - C_k| < |X^{(i)} - C_j| \quad (j = 1, \ldots, M\ j \neq k), \tag{10}$$

where $X^{(i)}$ is the ith input discontinuity, C_j is the jth cluster center, C_k is the cluster closest to the input object, and M is the number of allocated clusters.

After determining the closest cluster, the Euclidean norm must be tested against a distance threshold that is selected by the user. If the Euclidean norm is less than the threshold, then the current discontinuity joins the
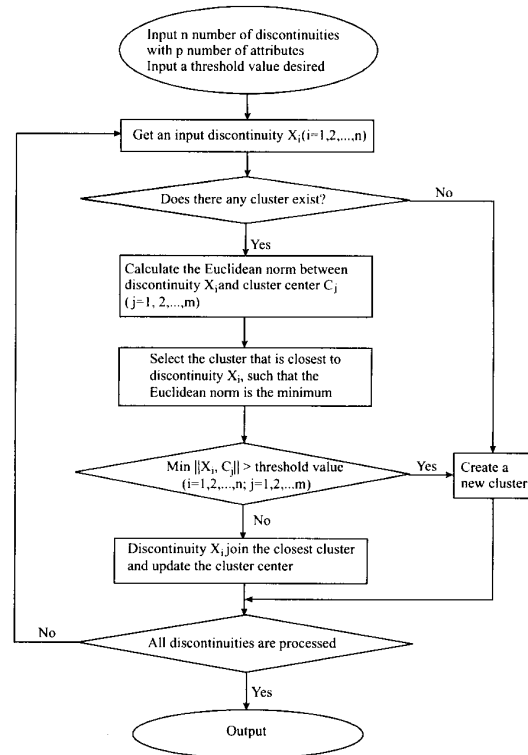

Fig. 6. Flow chart shows overall process of fuzzy c-means method.

Fig. 7. Flow chart shows overall process of vector quantization method (modified from Pandya and Macy, 1996).

Data5 - Notepad

r 1

1	232	89	0.58	5.	sandstone
2	76	88	0.95	4.	sandstone
3	300	80	1.32	6.	sandstone
4	117	78	1.87	7.	sandstone
5	183	2	3.1	2.	sandstone
6	309	70	4.28	4.	sandstone
7	94	75	5.01	6.	sandstone
8	278	84	5.85	4.	sandstone
9	185	11	6.2	3.	sandstone
10	268	70	7.21	7.	sandstone
11	259	85	7.9	12.	sandstone
12	187	3	7.93	5.	sandstone
13	285	73	8.08	5.	sandstone
14	121	79	9.14	5.	sandstone
15	345	89	9.38	5.	sandstone
16	192	4	9.42	4.	sandstone
17	68	70	10.56	7.	sandstone
18	256	88	10.78	6.	sandstone
19	181	5	11.01	3.	sandstone
20	75	85	12.52	5.	sandstone
21	201	5	12.96	3.	sandstone
22	281	70	13.18	12.	sandstone
23	105	69	13.73	6.	sandstone
24	84	63	13.94	7.	sandstone
25	67	84	14.19	14.	sandstone
26	88	88	14.55	12.	sandstone
27	92	74	14.71	9.	sandstone
28	87	79	15.18	7.	limestone
29	299	89	15.58	7.	limestone
30	188	2	15.92	3.	limestone
31	261	57	16.21	6.	limestone
32	182	7	17.18	4.	limestone
33	218	90	17.5	5.	limestone
34	94	74	17.74	5.	limestone

Fig. 8. Standard format of input data file contains discontinuity number, dip direction, dip, position along the borehole or scanline, and roughness and lithology.

Open — Save — Print — Rotate Up — Zoom In — Zoom Out — Move Up — Split Cylinder — Cylinder — Stereoscope — Chart

Rotate Left — Rotate Right — Rotate Down — Move Left — Move Right — Move Down — Plot Toggle — Color Toggle — Table — Help
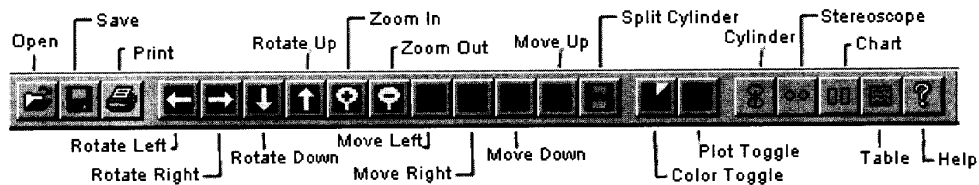
Fig. 9. Toolbar functions.

closest cluster. Otherwise, a new cluster is allocated for the current discontinuity. For detailed treatment about this method, see Pandya and Macy (1996). Fig. 7 illustrates the overall process of the vector quantization method.

The variable transformation procedures of vector quantization method are same as the k-means method.

3. The CYL program

CYL is a new analysis tool, developed in Visual C + +[®] ([®]Windows, Visual C + + are registered trademarks of Microsoft Corporation), which incorporates the various attributes of the discontinuities in a multivariate analysis, and uses multi-dimensional clus-

tering to identify joint sets and geological/geomechanical domains (statistically homogeneous regions with respect to geological structure). The method makes use of both clustering analysis and various visualization tools, in an integrated and automated package of computer algorithms.

3.1. Data input

An ASCII input file is used for data input. An example input file format is shown in Fig. 8. This can be generated using a text editor or a custom report generator from other software.

The first line of the input data file contains two variables. They are the roughness and the non-enumerable variable flags. These indicate the presence or

absence of roughness data or non-enumerable data (color, lithology, infilling, persistence, etc.), respectively. The remaining lines of the input data file have uniform line length, and contain information about the different attributes of every discontinuity. These are discontinuity number, dip direction, dip magnitude, position along the borehole or scan line, and roughness and descriptive data if present.

3.2. CYL user interface

The interaction between users and the program takes place by using menus, toolbars, shortcut keys and dialog boxes. Fig. 9 shows the icons and functions of various toolbars. Toolbars are enabled or disabled depending on the current display modes chosen.

Each clustering method has its own dialog box. Fig. 10 shows the dialog box of fuzzy *c*-means method (FCM). It allows users to change parameters, such as number of clusters, factors for determining the initial cluster centers, and weighting factors interactively.
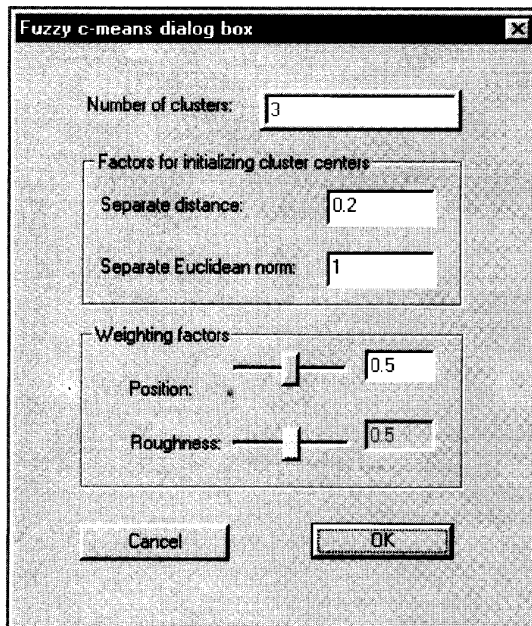

Fig. 10. The FCM dialog box. Number of clusters denotes user specified discontinuity sets. Separate distance and separate Euclidean norm are parameters to control distribution of initial cluster centers. Weighting factors are parameters that control influence of secondary variables. They range from zero to one, with larger values exerting more influence on value of variable.

3.3. Output display

In order to display the results, CYL offers many different views to output the analysis results. These are cylinder view, stereoscopic view, chart view and table view. Each view has different interactions, accessed through toolbars, menus or mouse.

The cylinder view (Fig. 11) offers a three-dimensional perspective of a lower hemisphere stereonet, where each pole of discontinuity is plotted on a "virtual" stereonet. Each discontinuity is plotted both in relation to its own steroenet and vertically in relation to its position along the borehole. Discontinuities are grouped within a set as identified by the multivariate cluster analysis, and each set number is designated by color and or and identifying number or symbol.

The "cylinder" in this view can be rotated about two axes for enhanced viewing. It gives information about the location of the poles of discontinuity in the stereonet and the position of the discontinuity along the scanline. More specifically, if the "cylinder" is rotated to the N–E plane, it gives a two-dimensional view of low-hemisphere stereonet projection of discontinuity poles. If the "cylinder" is rotated to the U–N or U–E plane, the scale bar is in its original position and it helps to read the distance along the scanline for any given data point.

The movements of the "cylinder" can be done via menu items, tools on the toolbar, or moving the mouse
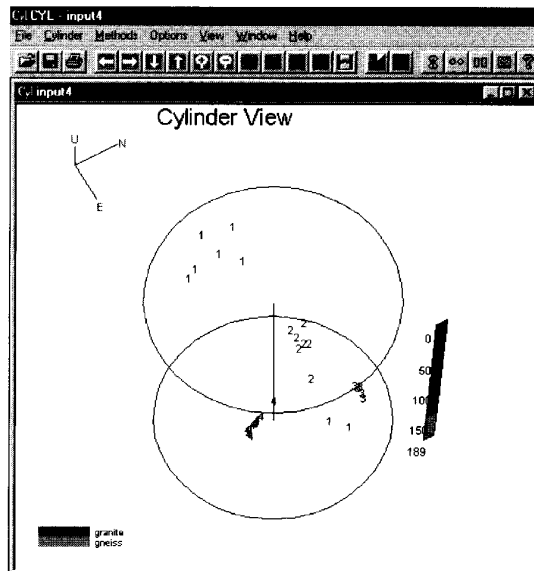

Fig. 11. Cylinder view (series of stacked lower hemisphere plots) displays distribution of discontinuities along borehole, discontinuity sets to which each discontinuity belongs, distance scale bar, and rock types that boreholes and scanlines intersect.
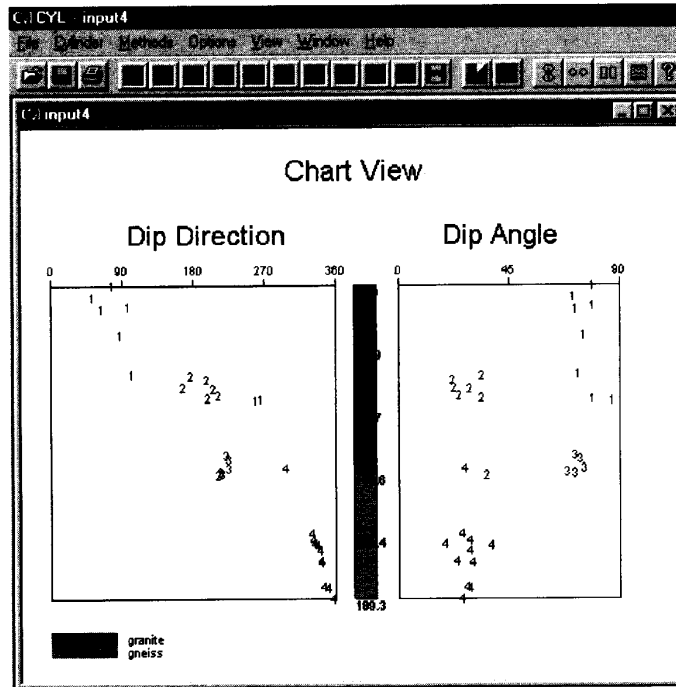
Fig. 12. Chart view offers another perspective of discontinuity distribution along borehole. Selection of split location is chosen from chart view. Notice that some tool bar functions are disabled in this view, since no rotations are needed.

Clustering Analysis

set #	joints	dip dir	dip ang	position	spacing	roughness
1	8	126	77	32.3	7.8	10.4
2	7	196	29	69.4	8.5	4.7
3	5	221	73	109.0	2.2	8.8
4	11	337	28	161.7	7.1	3.4

Fig. 13. Table summarizes average parameters within each discontinuity set.

on the over the view. In addition, the cylinder can be scaled or moved around the page. Other features include options to toggle the color numbers to black and white (for printing on laser printers), and a toggle to display numbers as symbols.

The chart view (Fig. 12) offers distribution of dip direction of discontinuities along the borehole and the distribution of dip angle of discontinuities along the borehole. In this figure, the data points with the same number belong to the same discontinuity set. The sets are distinguished from each other predominantly by their different orientations.

Using the chart view, the data set can be split into two separate data sets at the position of the cursor.
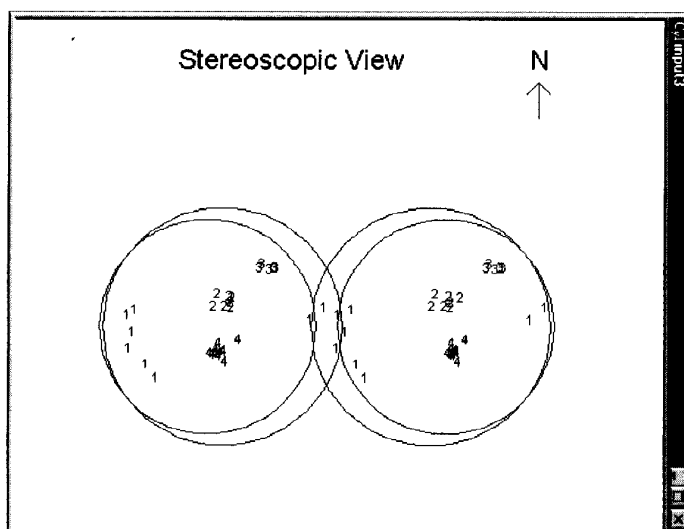
Fig. 14. Stereoscopic view offers three-dimensional picture as if seen through stereoscope.

This is used once GMUs (Geotechnical Mapping Units) are identified, that would best be analyzed separately.

The table view (Fig. 13) displays the summary statistics for each cluster. The stereoscopic view (Fig. 14) offers a thee-dimensional view of the borehole, designed to be printed and viewed with a pocket stereoscope.

4. Applications

The first example of using CYL program is given in Figs. 15A–D. For this example the discontinuities along a section of outcrop of gneiss and gabbro were mapped along a nearly horizontal scan line. The parameters measured of each discontinuity include orientation, distance, JRC, nature of the discontinuity and lithology.

Cluster analysis using orientation only by the nearest neighbor method is shown in Fig. 15(B). It reveals there is one sub-horizontal discontinuity set (set 2, which is the bedding planes), and two sub-vertical joint sets (set 1 and 3).

Clustering results using multivariate analysis by the nearest neighbor method considering orientation and position show a near horizontal discontinuity set and a series of three sub-vertical discontinuity sets (Figs. 15C and D). The sub-horizontal discontinuity set has a JRC of 14.0 and average spacing of 9.3 feet (set 2). Sets 3 and 4 have similar mean JRC values (13.6, 13.5), but are distinguishable by their different orientation, position

along the scan line, and average spacing (1.8, 2.4). Sets 1 and 5 are both sub-vertical and have similar mean JRC values (12.4, 12.8), but are distinguished from each other by their substantially different average spacing (4.3, 2.7 m) and position along the scan line (31.1, 81.5 m).

A long borehole might traverse more than one mapunit (GMU). There are different distribution patterns of discontinuities at different GMUs. Consequently, during the analysis, it often proves useful to split the data set into different GMUs.

Figs. 16A–C show an example of a data set that is split into two portions by the user. This results in two new windows opening for each data set. Each of the new windows can be saved as a separate data file, and each of the new data files can further be split, as required.

Fig. 16A is the original single data file. After splitting it results in two separate data files by manually selecting the split position. Further clustering analysis shows that three joint sets are identified in the upper portion of the borehole (Fig. 16B), and five joint sets are identified in the lower portion of the borehole (Fig. 16C). This result indicates that in some situations the clustering analysis can give valuable insights by partitioning into subsets. The problem is how to chose the best split locations? In order to provide a systematic guideline further studies are needed. The suggested rule of thumb includes looking at the lithologic changes and the transitions of different distribution patterns.
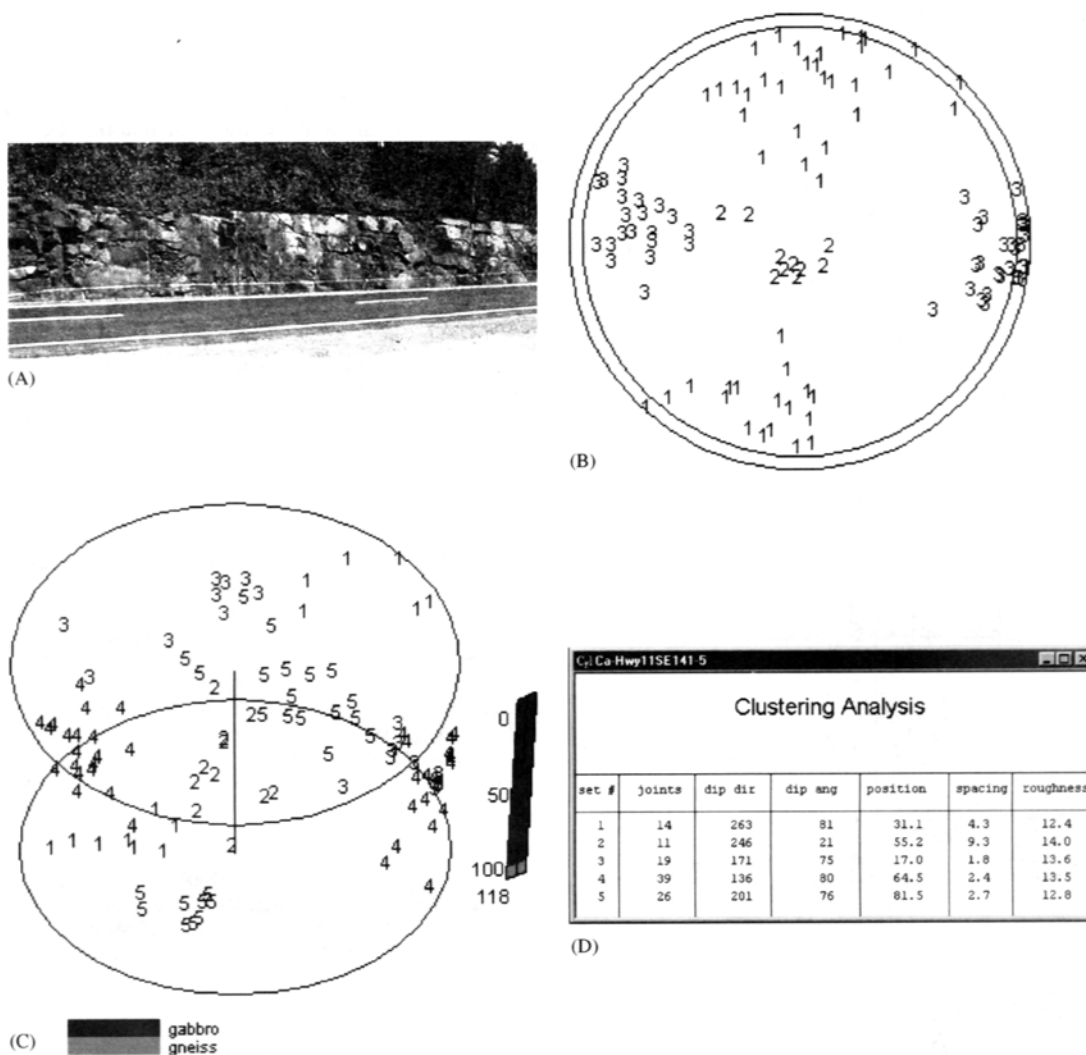
Fig. 15. (A) Discontinuities along section of highway outcrop of gneiss and gabbro were sampled along nearly horizontal scan line. (B) Cluster analysis by nearest neighbor method based on orientation only. Three joint subsets were identified by their dips and dip angles. (C) Multivariate cluster analysis by nearest neighbor method based on orientation and spacing. Scale bar on right indicates depth of scan line as well as rock types. (D) Tabular output from CYLINDER program shows discontinuity sets number, number of discontinuities in each set, mean orientation (dip and dip angle), mean position along the scan line (meter), mean apparent spacing (meter) and mean roughness (JRC).

5. Conclusions

This paper demonstrates a methodology that clusters discontinuity into subsets based on multiple attributes, so that discontinuities within the same subset will have similar geometric properties.

The CYL software package has the ability to incorporate discontinuity parameters, such as roughness and infilling materials, into the analysis. Multivariate clustering implemented in CYL is an effective and efficient method for characterizing rock discontinues.
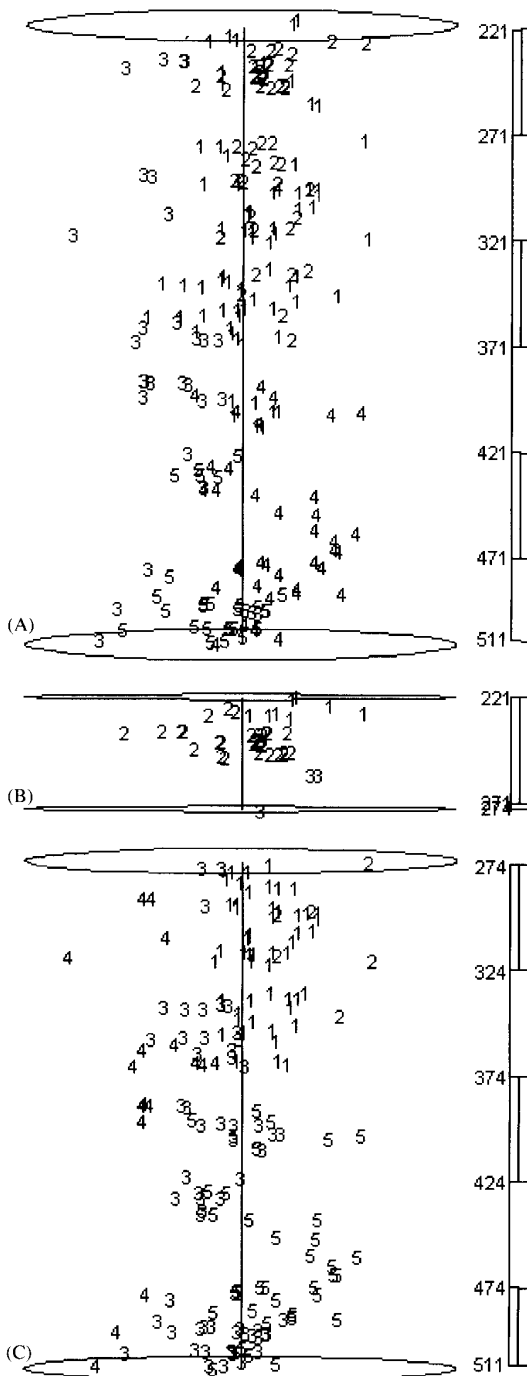
References

Baecher, G.B., 1983. Statistical analysis of rock mass fracturing. Mathematical Geology 15 (2), 329–347.

Barton, N., Choubey, V., 1977. The shear strength of rock joints in theory and in practice. Rock Mechanics 10, 1–54.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 256pp.

Dershowitz, W., LaPoint, P., Cladouhos, T., 1998. Derivation of fracture spatial pattern parameters from bore hole data. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts 35 (4/5), Paper No. 134.

Dillon, W.R., Goldstein, M., 1984. Multivariate Analysis Methods and Applications, Wiley, New York, 587pp.

Feng, X., Katsuyama, K., Wang, Y., Lin, J., 1997. A new direction: intelligent rock mechanics and rock engineering. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts 34 (1), 135–141.

Gath, I., Geva, A.B., 1989. Unsupervised optimal fuzzy clustering. IEEE Transitions on Pattern Analysis and Machine Intelligence 11 (7), 773–781.

Hammah, R.E., Curran, J.H., 1998. Optimal delineation of joint sets using a fuzzy clustering algorithm. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts 35 (4/5), Paper No. 136.

Hammah, R.E., Curran, J.H., 2000. Standardization of variables for the fuzzy K-means clustering of discontinuity data. In: Proceedings of the Fourth North American Rock Mechanics Symposium, Seattle, Washington, July 31–August 1, 2000, pp. 659–666.

Harrison, J.P., 1992. Fuzzy objective functions applied to the analysis of discontinuity orientation data. In: International Society of Rock Mechanics Symposium: Eurock'92, September 14–17, 1992, Chester, England, pp. 25–30.

Hudson, J.A., 1989. Rock Mechanics Principles in Engineering Practice, Butterworths, London, 72pp.

Hudson, J.A., Harrison, J.P., 1997. Engineering Rock Mechanics: An Introduction to the Principles. Elsevier Science, Ltd., Oxford, 444pp.

Hudson, J.A., Hudson, J.L., 1993. Establishing potential behavioural modes of rock engineering systems by computer simulation of interaction matrix energy flux. International

Fig. 16. (A) Five discontinuity sets are identified along 300 feet borehole. Both orientation and position are considered in analysis. (B) Upper portion of split borehole (from elevation of 221 feet to about 274 feet) shows three discontinuity sets. (C) Lower portion of split borehole (from elevation of 274 feet to about 511 feet) shows five discontinuity sets.

Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts 30 (4), 457–468.

Hudson, J.A., Priest, S.D., 1979. Discontinuity and rock mass geometry. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts 16, 339–362.

International Society of Rock Mechanics Commission on Standardization of Laboratory and Field Tests 1978, 1978. Suggested methods for the quantitative description of discontinuities in rock masses. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts 15, 319–368.

Kulatilake, P.S.H.W., Wu, T.H., 1984. Estimation of mean trace length of discontinuities. Rock Mechanics and Rock Engineering 17, 215–232.

Maerz, N.H., Franklin, J.A., Bennett, C.P., 1990. Joint roughness measurement using shadow profilometry. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts 27 (5), 329–432.

Maerz, N.H., Zhou, W., 1999. Multivariate analysis of bore hole discontinuity data. In: Proceedings of the 37th US Rock Mechanics Symposium, Vail, Colorado, June 6–9, 1999, pp. 431–438.

Maerz, N.H., Zhou, W., 2000. Discontinuity data analysis from oriented boreholes. In: Pacific Rocks, Proceedings of t he Fourth North American Rock Mechanics Symposium, Seattle, Washington, July 31–August 1, 2000, pp. 667–674.

Mahtab, M.A., Bolstat, D.D., Alldredge, J.R., Shanley, R.J., 1972. Analysis of fracture orientations for input to structural models of discontinuous rock. US Bureau of Mines Report of Investigations 7669, 76pp.

Mahtab, M.A., Yegulap, T.M., 1984. A similarity test for grouping orientation data in rock mechanics. In: Proceedings of 25th US Symposium on Rock Mechanics, Evanston, Illinois, June 25–27, 1984, pp. 495–502.

Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. Multivariate Analysis. Academic Press, London, 521pp.

Mauldon, M., Dunne, W.M., Rohrbaugh Jr., M.B., 2001. Circular scanlines and circular windows: new tools for characterizing the geometry of fracture traces. Journal of Structural Geology 23 (2/3), 247–258.

Pandya, A.S., Macy, R.B., 1996. Pattern Recognition with Neural Networks in C++. CRC Press and IEEE Press, Boca Raton, FL, 410pp.

Priest, S.D., 1993. Discontinuity Analysis for Rock Engineering, 1st Edn. Chapman & Hall, London, 473p.

Priest, S.D., Hudson, J.A., 1981. Estimation of discontinuity spacing and tracing length using scanline survey. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts 18, 183–197.

Tao, Z.Y., Peng, Z.Z., 1983, Classification system of fuzzy sets for rock engineering. International symposium on engineering geology and underground construction. Lisbon, Portugal, September 12, 1983, pp. II.1–II.9.

Tsoutrelis, C.E., Exadactylos, G.E., Kapenis, A.P., 1990. Study of rock mass discontinuity system using photoanalysis. In: Rossmanith, H.P. (Ed.), Mechanics of Jointed and Faulted Rock. Balkema, Rotterdam, pp. 103–112.

Wenk, H.R., Takeshita, T., Bechler, E., Erskine, B.G., Matthies, S., 1987. Pure shear and simple shear calcite textures: comparison of experimental, theoretical and natural data. Journal of Structural Geology 9 (5/6), 731–745.

Willis-Richards, J., Jupe, A., 1995. Evaluation of fracture orientation statistics from borehole imaging. Scientific Drilling 5, 55–60.

Xie, X.L., Beni, G., 1991. A validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-13 (8), 841–847.

Zadeh, L.A., 1965. Fuzzy sets. Information Control 8, 338–353.

Zhou, W., Maerz, N.H., 2001. Multivariate clustering analysis of discontinuity data: applications. In: Proceedings of the 38th US Rock Mechanics Symposium, Washington, DC, July 7–10, accepted for publication.