Multivariate Clustering Analysis of Discontinuity Data: Implementation and Applications

W. Zhou
Department of Geological Engineering, University of Missouri-Rolla

N.H. Maerz
Rock Mechanics and Explosives Research Center, University of Missouri-Rolla

Rock Mechanics in the National Interest; Proceedings of the 38th U.S. Rock Mechanics Symposium, Washington, D.C., July 7-10, 2001, pp 861-868.

ABSTRACT This paper presents the results of ongoing research on the characterization of rock mass structure from discontinuity data. Multivariate clustering analysis represents a relatively recent development in characterizing the structure of rock masses. Multivariate clustering allows characterization of discontinuities into subsets according to multiple parameters, such as orientation, spacing, and roughness, where, rather than considering one variable at a time, a number of parameters can be treated simultaneously, so that the interactions between parameters are taken into account. The comprehensive algorithm has been developed into a software package called CYL. It enables fully automated multivariate clustering analysis and offers various visualization tools, such as a three-dimensional stereonet, a stereoscopic view, and a statistical table. In this paper we focus on the implementation of the algorithms and the application of this method to field data, both from oriented core and mapped road cut discontinuity data.

1 INRODUCTION

The characterization of rock discontinuities is an essential requirement for most of the rock engineering projects. Discontinuities within rock dominate the mechanical and hydraulic behaviors of rock masses (Figure 1). Because of lack of effective and comprehensive tools for the interpretation of the data, the discontinuity analyses suffer from the inability to incorporate more than one parameter simultaneously.

Multivariate clustering analysis represents a relatively recent development in characterizing the structure of rock masses. It characterizes discontinuities into subsets according to multiple parameters, such as orientation, spacing, and roughness, where rather than considering one variable at a time, a number of parameters can be treated simultaneously, so that the interactions between parameters are taken into account (Dershowitz & Einstein 1988).

Rock characterization using oriented bore-hole data or linear mapping data is often more useful because it is cost effective, and can target the exact location of a proposed underground structure. But because of the lack of effective tools for the interpretation of this data, it is more often than not underutilized.

In the previous papers (Maerz & Zhou 1999; Maerz & Zhou 2000) a new approach to the analysis of oriented borehole discontinuity data was introduced. Rather than considering parameters such as orientation, spacing, infilling, wall rock strength, roughness and mineralization individually, a multivariate approach was used. The new approach is designed around utilizing multivariate-clustering algorithm, considering both spherical orientation data and spatial data. The data can be visualized in a "three-dimensional" stereonet where joint poles are plotted on individual "stacked" stereonets, where each pole is plotted with respect to its own stereonet, and each stereonet is plotted in a linear position that corresponds to the position where the joint corresponding to that joint normal intersects the bore hole (Figure 2). This idea was first used by Wenk (Wenk et al. 1987) to represent the pattern of lattice preferred orientation in deformed rocks.

Figure 1. Outcrop of weathered gneiss close to I-70 Exit 259 in Colorado.

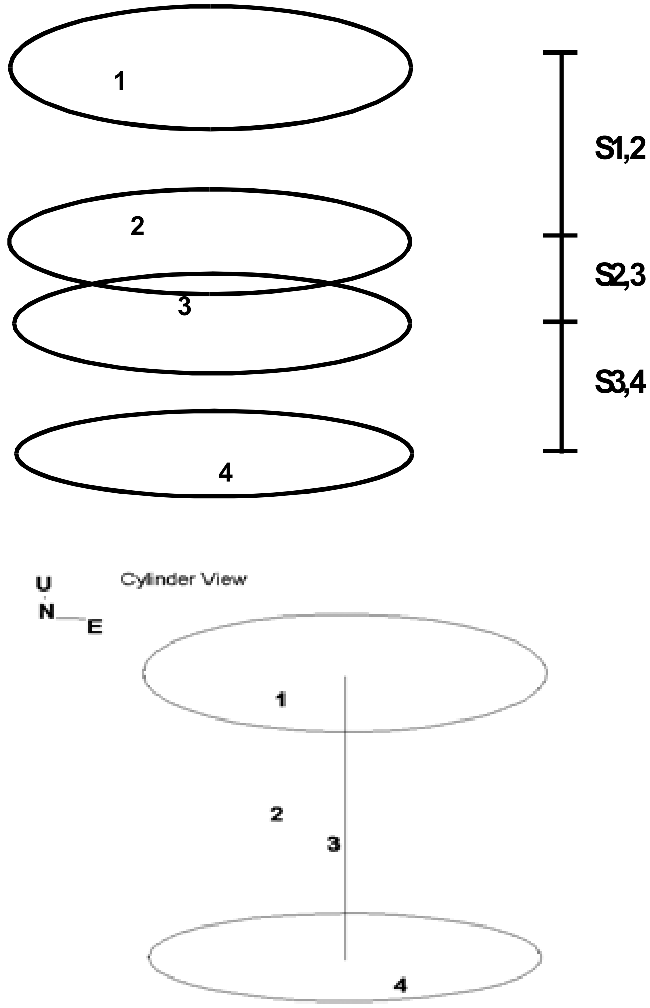


Figure 2. Top: The individual stereonets are stacked, with each spacing in proportion to the spacing between discontinuities in the borehole. S1,2 is the spacing distance between discontinuity 1 and 2. Bottom: "Three-dimensional stereonet". (Maerz & Zhou 2000).

In this paper we focus on the implementations of the multivariate clustering methods, usages and the visual tools of the CYL software, and the applications of CYL to real data. In addition to spherical orientation data and spatial data, roughness has been added to the analysis as a fourth parameter in the multivariate analysis. Weighting factors, which have values between 0.0 and 1.0, is assigned to the parameters according to the degree of importance of each parameter. Discussions about how weighting factors affects the clustering results are also presented in this paper.

2 MULTIVARIATE CLUSTERING METHODOLOGY

In previous papers (Maerz and Zhou, 1999; Maerz and Zhou 2000), four clustering methods were described. Discontinuity orientation data is a combination of vector and scale data types, and the following is a discussion of how this is incorporated into a multivariate clustering analysis.

2.1 *Nearest neighbor clustering method*

For the nearest neighbor method (also called single linkage method), the similarity between joint attributes is based on Euclidean distance measurements. A thorough description about this method can be found in Dillion and Goldstein (1984).

To implement this method for discontinuity data analysis, data transformations have to be carried out. There are two different types of discontinuity parameters. Orientation parameters (dip direction and dip) are vector variables, while the rest of the parameters are scalars. For orientation vectors we are interested in the differences in direction, while for scalar variables we are interested in the differences between their magnitudes.

When using the nearest neighbor method, spherical coordinates are used to convert orientation variables (dip/dip direction) into vectors. The arc length between any two vectors on the unit sphere is calculated and used as the Euclidean distance. During the calculation of the arc length, both original vector and its corresponding reverse vector are used. The smaller arc length between the two is used in further analysis. A vector flag is used to track whether the original vector or the reverse vector is chosen. This is very important for calculating the mean within each clustering group later on. In addition to the arc length, other variables, for example distance between each pair of discontinuities can be considered by the absolute difference of the positions of the two discontinuities.

2.2 *K-means clustering method*

A detailed discussion about K-means method can also be found in Dillion and Goldstein (1984).

For this 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, and is defined in Figure 3. The smaller Euclidean distance 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 very important for calculating the mean orientation within each clustering group.

2.3 *Fuzzy c-means clustering method*

Fuzzy sets (Zadeh, 1965) are an approach to represent vagueness in everyday life. Often, there are situations that some data points are located in between several clustering centers. This situation can be described mathematically by a fuzzy membership factor. If a data point located exactly on a clustering center, then the fuzzy membership factor of this data

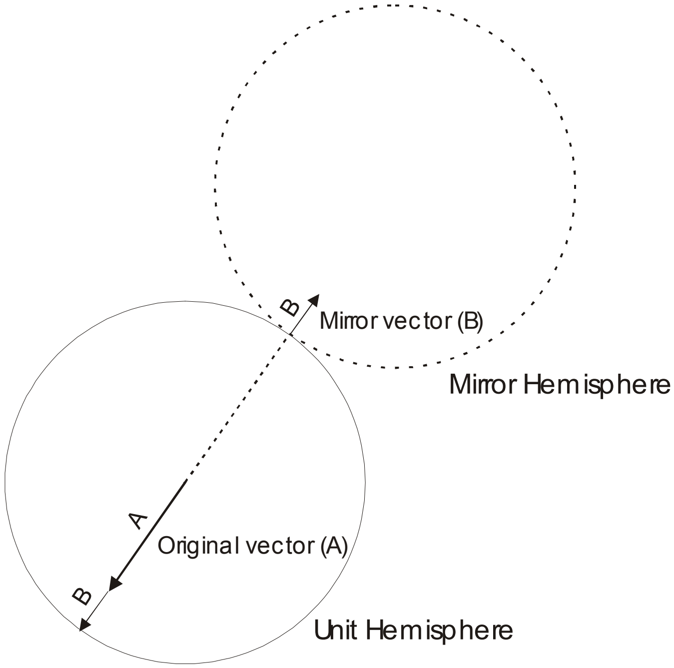


Figure 3. Schematic drawing shows the concept of mirror vector.

point relative to that clustering center is 1.0. If the distance from a data point to a clustering center is far enough then the fuzzy membership factor of this data point relative to that clustering center is almost 0.0. In most cases the fuzzy membership factor is between 0.0 and 1.0. By using the fuzzy c-means method (FCM), the vagueness of real data is taken into account. Detailed description about this method can be found in many publications (Bezdek, 1981; Gath and Geva, 1989; Xie and Beni, 1991; Harrison, 1992; Hammah and Curran, 1998; etc.).

To apply fuzzy c-means method to discontinuity data, data transformation had to be carried out. Since spherical coordinates were adopted the data transformation procedures of fuzzy c-means method are the same as that of nearest-neighbor method.

2.4 *Vector quantization clustering method*

Most of the above algorithms require the number of clusters to be known in advance. However, in real situations a-priori knowledge of the number of the meaningful clusters to describe the discontinuity data from an oriented bore hole may not be available. Therefore, unsupervised learning methods become highly desirable. The vector quantization method (VQM) is an unsupervised learning method. For detailed treatment about this method, please refer to Pandya and Macy (1996).

Since cylinder coordinates were used in the vector quantization method, the data transformation procedures of the vector quantization method are the same as the k-means method.

3 APPLICATIONS

Examples of applications using program CYL are given in the following sections. In two cases discontinuities along road cuts were mapped along near-horizontal scan lines; in a third, borehole data was used.

Although CYL was designed for analysis of borehole data, scan line mapping data can also be used. Often, because of the ease of data collection scan line mapping date is easier to get, and there is more flexibility in which parameters can be logged.

The sampling methodology follows Piteau and Martin (1979). The parameters measured and recorded for each discontinuity include orientation, position, roughness, waviness, persistence, type of discontinuity, lithology, filling and color. The data was then analyzed using CYL.

3.1 *Granitic Rocks, Golden Colorado*

The first application involves a road cut along interstate 70, near Golden Colorado. A 190 m section of rock cut through granite, gneiss, and weathered granite was measured (Figure. 4).

The clustering method used in this example is the supervised nearest neighbor method (NNM). Clustering analysis using orientation only (traditional method) is shown in Figure 5. The numbers in the stereonet represent the group number of each discontinuity set, and the location of each number represents the pole of each individual joint in a lower hemisphere equal angle projection. It reveals there are two sub-vertical discontinuity sets (sets 1 and 5).

Three discontinuity sets have a dip angle of about 50º (sets 2, 3, and 6), and one relatively gently dipping discontinuity set with a dip angle of 35º (set 4).

Figure 4. Outcrop of weathered granite close to I-70 Exit 259 in Colorado.

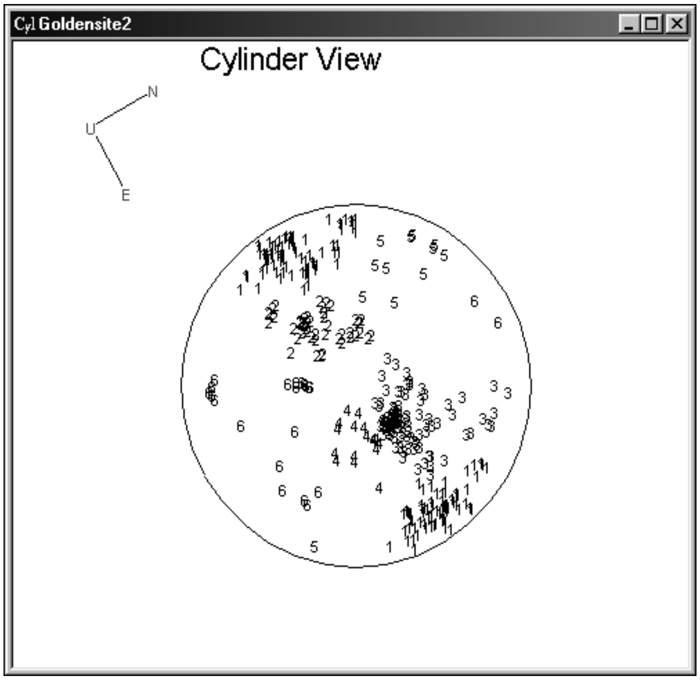


Figure 5. Clustering analysis by nearest neighbor method based on orientation only. Six joint subsets were identified by their dip and dip angle.

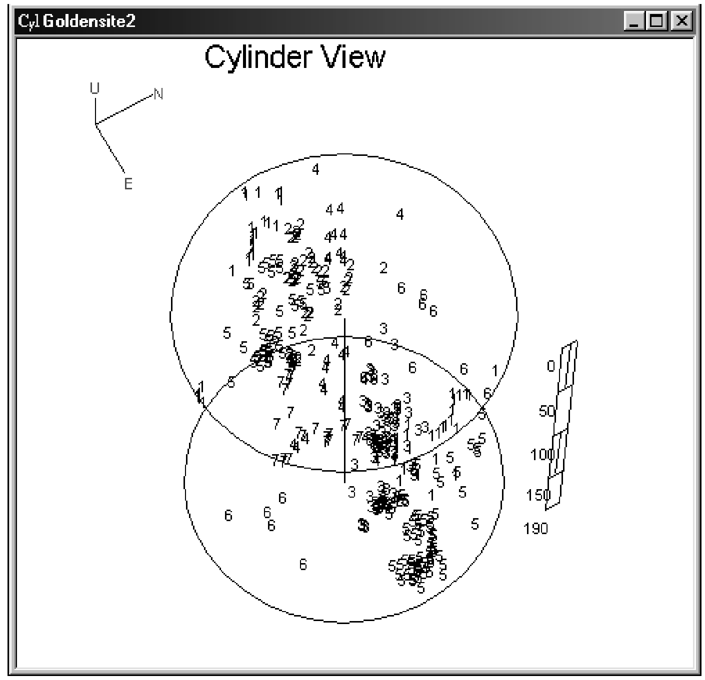
Figure 6 shows the results using multivariate clustering analysis considering orientation as well as spacing. The weighting factor for position is 0.5.

Seven discontinuity sets are identified. There are a series of three sub-vertical discontinuity sets (sets 1, 5, and 6), three joint sets with a moderate dip angle (sets 2, 4, and 7), and a set with a relatively gentle dip angle (set 3). Among the three sub-vertical discontinuity sets, set 1 is distinguished by its dip direction and position along the borehole, set 5 and set 6 have similar position along the borehole, but are distinguished from each other by their nearly perpendicular dip direction. Set 2 and set 7 have similar dip direction and dip angles, but are distinguished from each other by their position. Since the roughness is not considered in the clustering analysis, the difference of roughness between each set are relatively uniform comparing the results shown in Figure 6.

Figure 7 shows the results of considering orientation and both spacing and roughness. For roughness, Bartons joint roughness coefficient (JRC) is used (Barton, 1977). The weighting factors for position and roughness are both 0.5. In this scenario, discontinuity sets are partitioned by their one or more distinguished parameters.

Adding this 4th dimension results in a redistribution of clustering centers. The Cylinder View of Figure 7, which is a 3-dimentional stereonet, cannot adequately display the results of the 4-dimensional analysis.

From the tabular view of Figure 7, we can determine that sets 3 and 5 are both of high roughness, distinguishable from each other by dramatically differing values for the spacing parameter, and slightly



Goldensite2

Clustering Analysis

set #	joints	dip dir	dip ang	position	spacing	roughness
1	63	45	82	29.2	1.3	5.6
2	56	3	52	33.2	1.5	6.5
3	89	206	35	112.6	2.1	9.1
4	29	356	65	38.1	3.2	6.5
5	122	212	79	142.1	0.8	6.5
6	15	311	79	148.5	5.0	4.7
7	23	9	43	162.8	2.8	6.2

Figure 6. Top: NNM method; 3D; considering orientation and position; number of clusters = 7; weighting factor for position = 0.5. Bottom: Tabular output summary of parameter averages by set.

differing values for average dip angles (dip directions are almost identical).

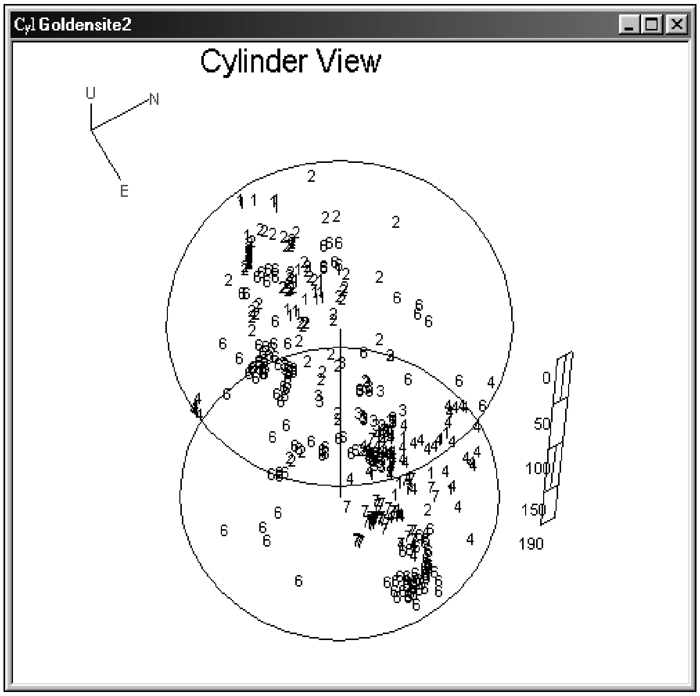
Three sets (2, 4, 6) have low and consistent roughness, but are located along different positions along the scan line as well as dramatically different dip directions (dip angles are similar).

Sets 1 and 7 are of moderate roughness but are of completely different position and attitude.

3.2 *Gabbroic Rocks, Huntsville, Ontario, Canada*

The next application involves a road cut along Highway 11, north of Hunstville, Ontario, Canada. A 118 m section of rock cut through fresh gabbro and gneiss, was measured (Figure. 8).

The clustering method used in this example is the supervised nearest neighbor method (NNM). Clustering analysis using orientation only (traditional



Cyl Goldensite2

Clustering Analysis

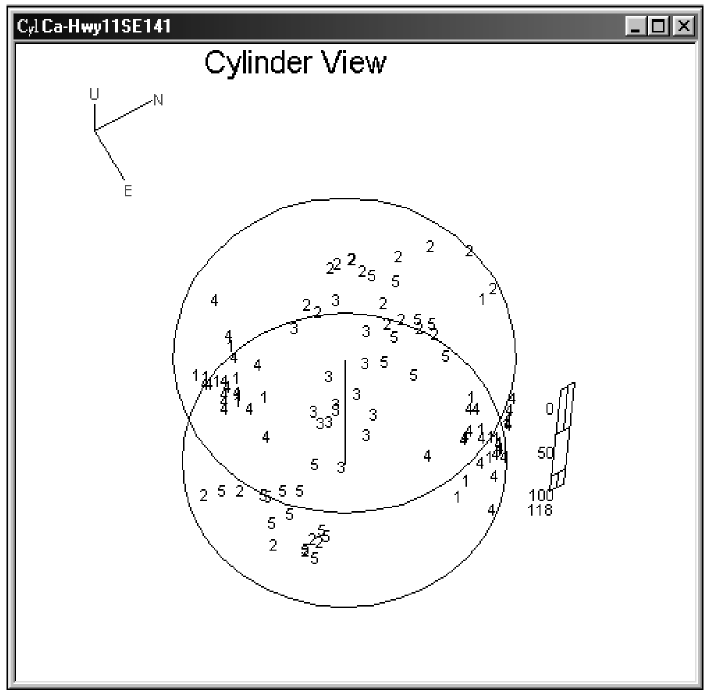
set #	joints	dip dir	dip ang	position	spacing	roughness
1	31	343	81	45.4	3.3	8.0
2	81	2	61	31.4	1.1	5.6
3	51	189	33	71.3	2.4	10.2
4	70	198	64	84.7	2.4	5.6
5	2	186	65	102.1	41.5	14.0
6	122	334	76	145.8	1.0	5.9
7	40	201	43	165.7	1.5	9.2

Figure 7. Top: Colorado site, NNM method; 4D; considering orientation, position, and roughness; number of clusters = 7; weighting factor for position = 0.5; weighting factor for ropughness = 0.5. Bottom: Tabular output summary of parameter averages by set.

method) shows that there are predominantly three discontinuity sets (one sub-horizontal set and sub-vertical sets). The sub-horizontal set has an average dip direction of 135º and dip angle of 15º. The two sub-vertical discontinuity sets are almost orthogonal to each other, one set has an average dip direction of 319º, and the other set has an average dip direction of 234º.

Figure 9 shows the 3-dimensional stereonet and tabular output indicating discontinuity grouped by considering orientation, position along the scan line, and roughness simultaneously. The weighting factor for position is 0.4, and weighting factor for roughness is 0.6. There are 4 sub-vertical joint sets (sets 1, 2, 4, and 5), and one sub-horizontal set (set 3). Joint sets 1 and 4 are partially overlapped in orientation, but are distinguished by the different average position and roughness, as are sets 2 and 5.

Figure 8. Outcrop of gabbro along highway 11, Huntsville, Ontario, Canada.



Cyl Ca-Hwy11SE141

Clustering Analysis

set #	joints	dip dir	dip ang	position	spacing	roughness
1	20	243	80	53.1	2.7	9.2
2	23	324	79	54.8	4.9	16.2
3	14	347	26	48.0	7.3	13.8
4	32	232	81	52.8	3.6	16.2
5	20	131	75	65.9	4.8	8.8

Figure 9. Top: Ontario site, NNM method; 4D; considering orientation, position, and roughness; number of clusters is 5; weighting factor for position is 0.4; weighting factor for JRC is 0.6. Bottom: Tabular output of summary of parameter averages by set.

3.3 *Sulphide Ore, Northern Labrador, Canada*

For this example the discontinuity data was collected from orientated boreholes in a nickel deposit in northern Labrador, Canada. The geological formation mainly consists of sulphide, gneiss, and pegmatite. The borehole is 184 meters long. Joint roughness was measured by index of joint roughness (J_r), which ranges from 0.0 to 3.0. The clustering method used in this example is the fuzzy c-means method (FCM).

Figure 10 is the tabular summary of clustering results by considering orientation only. The five discontinuity sets have relatively uniform average attributes including dip, but reveal different dip directions, distributed almost evenly in the low hemisphere projection.

Figure 11 shows the 3-dimensional stereonet and tabular output indicating discontinuity grouped by considering orientation and position along the orientated borehole. The weighting factor for position is 0.5. There are 6 discontinuity sets (sets 1, 2, 3, 4, 5, and 6) that have moderate dip angle, and one sub-horizontal set (set 7). The different average dip direction, roughness, or both distinguish the 6 moderate sets from one and another. Notice that the average roughness by set are relatively uniform in the table of Figure 11 since we have not included the roughness in this analysis yet.

Figure 12 shows the tabular results of considering orientation, spacing and roughness. The weighting factors for position and roughness are both 0.5. In this scenario, discontinuity sets are partitioned by orientation, position, roughness, or the combination of these distinguished parameters.

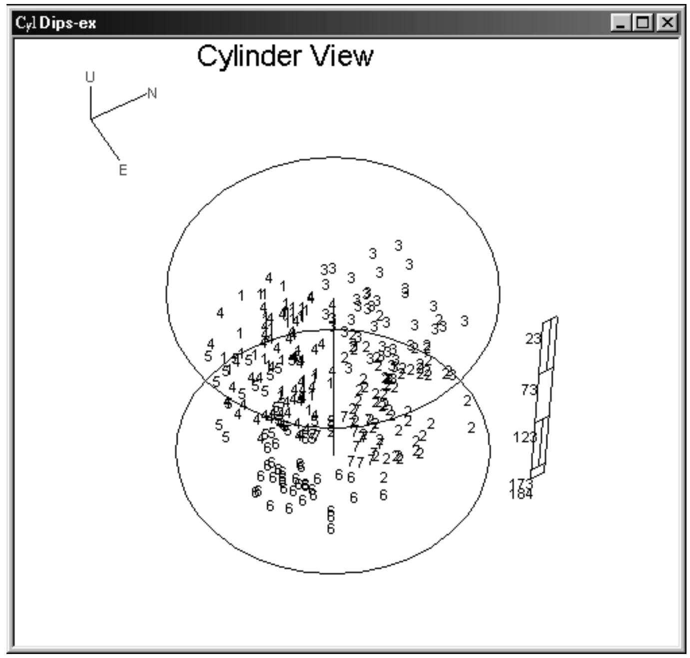
From the tabular view of Figure 12, we can determine that set 5 is the roughest set (recall that J_r ranges from 0.0 to 3.0), while set 7 is the smoothest set.

Cyl Dips-ex

Clustering Analysis

set #	joints	dip dir	dip ang	position	spacing	roughness
1	45	77	56	103.3	3.6	2.3
2	55	142	58	99.4	2.8	1.8
3	76	220	51	92.6	2.1	2.0
4	74	16	42	114.5	2.0	1.5
5	30	304	50	93.9	4.6	2.2

Figure 10 The Tabular output of summary of parameter averages by set; FCM method; 2D; considering orientation only; number of cluster is 5.



Cyl Dips-ex

Clustering Analysis

set #	joints	dip dir	dip ang	position	spacing	roughness
1	36	84	46	42.8	1.3	2.2
2	65	200	57	76.8	1.6	2.2
3	45	287	50	83.6	2.3	2.2
4	59	16	48	133.5	1.9	1.4
5	26	93	56	90.5	1.6	1.6
6	34	121	54	157.3	2.5	1.9
7	15	264	28	171.7	3.9	1.7

Figure 11 Top: Northern Labrador site, FCMC method; 3D; considering orientation, and position; number of clusters is 7; weighting factor for position is 0.5. Bottom: Tabular output of summary of parameter averages by set.

Cyl Dips-ex

Clustering Analysis

set #	joints	dip dir	dip ang	position	spacing	roughness
1	51	76	49	61.9	1.8	2.0
2	40	167	59	68.2	2.6	1.9
3	73	242	53	85.2	2.0	2.2
4	35	3	49	121.4	4.0	1.5
5	25	13	47	145.9	3.8	3.0
6	33	125	51	162.2	1.8	1.8
7	23	1	38	134.4	4.3	0.0

Figure 12 The Tabular output of summary of parameter averages by set; FCM method; 4D; considering orientation, position, and roughness; number of cluster is 7; weighting factor for position is 0.5; weighting factor for roughness is 0.5.

3.4 *Analysis of Weighting Factor Influence*

During the multivariate clustering analysis, it is very important to identify the most dominant parameters in cluster partition. In our analysis, orientation (dip and dip angle) is usually considered as the primary parameter. Position, roughness and others are treated as secondary parameters. A weighting factor (between zero and one) is assigned to the secondary parameters. The selection of a proper weighting factor with respect to a specific parameter is vital to the recovery of the true cluster structure (Hammah and Curran, 2000). In the Ontario example, we show the impacts of the weighting factor toward secondary parameters by calculating the standard deviation of the JRC with various weighting factors.

To study the influence of the weighting factor, analyses were conducted on the Ontario data. The weighting factor for position was varied from 0.3 to 0.7 in an increment of 0.1. At the same time the weighting factor for roughness was varied from 0.7 to 0.3 with the same increment. During this, the sum of the two weighting factors was always maintained at 1.0. Then the standard deviation of grouped data of average position and average roughness was calculated for each scenario.

Figure 13 shows the relationship between the standard deviation of roughness (JRC) in grouped data and the weighing factor assigned. The dispersion of the average roughness of each discontinuity set increases with the increase of the weighting factor. Further study will determine the optimal weighting factor for individual parameters.

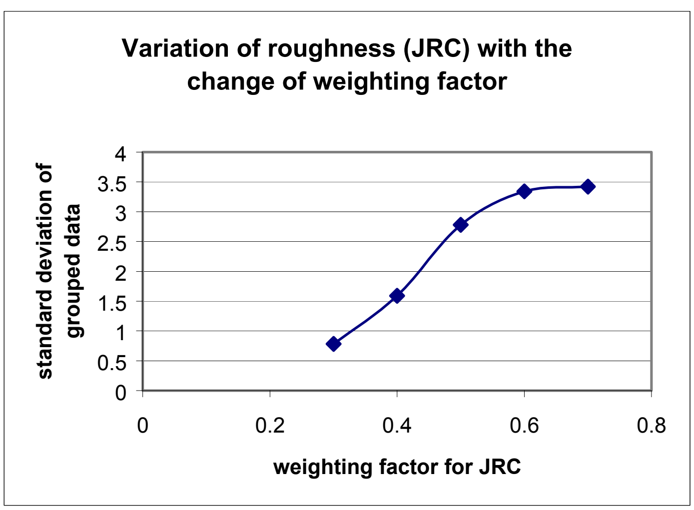


Figure 13. The standard deviation of roughness (JRC) in grouped data increase with the increase of the weighting factor.

4 DISCUSSION

This paper demonstrates an approach that clusters discontinuity into subsets based on multiple attributes, so that discontinuities within the same subset will have similar geometric properties. The multivariate clustering approach was adapted to analysis discontinuity data, which is a combination of vector and scalar data.

A new analytical tool (CYL software package) was developed in PC platform in Visual C++®. It has the ability to incorporate additional discontinuity parameters, such as roughness and infilling materials, into the analysis. Multivariate clustering implemented into CYL with user-friendly interface is a cost effective and efficient method for characterizing discontinuity rock. CYL software package was applied to analysis real discontinuity data collected from linear mapping. CYL offers an automatic, fast and cost efficient way to characterize discontinuity data.

5 ACKNOWLEGMENTS

The authors would like to express our gratitude to the University of Missouri System Research Board and National Science Foundation for funding this research. We also wish to thank Golder Associates for providing discontinuity data used in this study.

6 REFERENCES

Barton, N., and 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: 256 pp.

Dershowitz, W. S., and Einstein H. H., 1988, *Characterizing rock joint geometry with joint system models,* Rock Mechanics and Rock Engineering, 21, 21-51

Dillon, W. R. and Goldstein, M., 1984, *Multivariate Analysis Methods and Applications*, John Wiley & Sons, New York: 487 pp.

Gath, I., and Geva, A. B., 1989, *Unsupervised optimal fuzzy clustering*, IEEE Trans. Pattern Anal. Machine Intell 11(7): 773-781

Hammah, R. E., and Curran, J. H., 1998, *Optimal delineation of joint sets using a fuzzy clustering algorithm*, Int. J. of Rock Mech. & Min. Sci. 35 (7): 889-905.

Hammah, R. E., and Curran, J. H., 2000, *Standardization and weighting of variables for the fuzzy K-means clustering of discontinuity data*, Pacific Rocks 2000, Girard, Liebman, Breeds and Doe (eds) © 2000 Balkema, Rotterdam, ISBN 90 5809 155 (4): 659-666

Harrison, J. P., 1992, *Fuzzy objective functions applied to the analysis of discontinuity orientation data,* Eurock'92: 25-30.

Maerz, N. H., and Zhou, W., 1999. *Multivariate analysis of bore hole discontinuity data***.** Rock Mechanics for Industry, Proceedings of the 37th US Rock Mechanics Symposium, Vail Colorado, June 6-9, 1999, v. 1, pp. 431-438.

Maerz, N. H., and Zhou, W., 2000. *Discontinuity data analysis from oriented boreholes***.** Pacific Rocks; Proceedings of the Fourth North American Rock Mechanics Symposium, Seattle, Washington, July 31- Aug.1, 2000, pp. 667-674.

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

Piteau and Martin, *Rock Slope Engineering reference Manual Part G (Field Manual): Description of Detail Line Engineering Geology Mapping Method*, Piteau (D. R.) And Associates Ltd., Vancouver (British Columbia), January1979: 30 pp.

Wenk, H. R., Takeshita, T., Bechler, E., Erskine, B. G., and Matthies S., 1987, *Pure shear and simple shear calcite textures, Comparison of experimental, theoretical and natural data*, Journal of Structure Geology, Vol. 9, No. 5/6 pp. 731-745

Xie, X. L. and Beni, G., 1991, *A validity measure for fuzzy clustering,* IEEE Trans. Anal. Machine Intell., vol. PAMI-13 (8): 841-847.

Zadeh, L. A., 1965, *Fuzzy sets*, Information Control, vol. 8, p. 338-353.

®Visual C++ is a registered trademark of Microsoft Corporation.