Rock Mech Rock Engng (2008) 41: 929–939
DOI 10.1007/s00603-007-0155-6
Printed in The Netherlands

Rock Mechanics
and Rock Engineering

Technical Note

Fuzzy spectral clustering for identification of rock discontinuity sets

By

R. Jimenez

Department of Civil and Environmental Engineering,
Imperial College London, London, UK
ETS Caminos, C. y P. Universidad Politecnica de Madrid, Spain

Received July 5 2006; Accepted July 5 2007; Published online November 13 2007

Keywords: Joint sets, K-means, orientation analysis, structural analysis

1. Introduction

The characterization of rock masses for engineering applications generally includes analyses for the identification of discontinuity sets and the characterization of their orientation which are commonly performed using hemispherical projections of discontinuity unit normal vectors (see e.g., Priest, 1985, 1993b; Harrison, 1992). Visual inspection of contour plots of orientation density computed by counting the number of unit normal vectors that fall inside a reference circle have been traditionally employed for such task. However, they have been found to present problems due to sampling bias (Terzaghi, 1965; Priest, 1993a); to clustering results heavily depending on the size of the reference circle (Harrison, 1992); and to subjectivity in the interpretation of the results (Mahtab and Yegulalp, 1982; Priest, 1993b; Hammah and Curran, 1998). Such problems have lead to a recent interest in the development of alternative techniques for automatic identification of rock discontinuity sets based on their orientation. For instance, spectral clustering methods that use eigenvectors of matrices constructed using measures of similarity between the data points have been recently employed for clustering of rock discontinuities (Jimenez-Rodriguez and Sitar, 2006).

Correspondence: R. Jimenez, Department of Civil and Environmental Engineering, Imperial College London, Skempton Building, Room 335, SW7 2AZ London, UK
e-mail: r.jimenez@imperial.ac.uk

In this work we demonstrate the integration of spectral methods and fuzzy K-means for clustering of discontinuity orientations based on their orientation. In general, clustering techniques based on fuzzy algorithms are believed to provide improved partitions with respect to non-fuzzy clustering methods (see e.g., Hammah and Curran, 1998). Fuzzy clustering algorithms assign, for each observation in the data set, degrees of membership to the different clusters that provide information about the uncertainty of the clustering assignments. That is, when an observation belongs to a cluster, it tends to have a high degree of membership to that cluster and low degrees of membership to the remaining clusters (Hammah and Curran, 1998). A further advantage of fuzzy clustering techniques is that degrees of membership can be regarded as probabilities that can be exploited in simulations using Monte Carlo sampling. This could be of significant practical interest if there are other properties, such as size, transmissivity, etc., associated to fractures of each discontinuity set (Munier, 2006).

We also present a novel method for graphical representation of discontinuity sets, in which the degree of membership of each discontinuity to its assigned set (i.e., a measure of the certainty of the assignment) is explicitly indicated by means of a color scale. Finally, we further illustrate the performance of the algorithm in a number of test cases, and we compare the clustering results computed using the proposed fuzzy spectral clustering algorithm with results computed using other clustering algorithms employed in rock engineering applications.

2. Algorithm for fuzzy spectral clustering

In this work we integrate fuzzy K-means with the spectral clustering algorithm. As an initial step to perform clustering, however, we need to define a notion of similarity (or distance) between orientations of discontinuities in the data set. The goal of the fuzzy clustering algorithm is then to identify discontinuities that are ‘‘similar’’ and to assign them (considering the uncertainty of the assignment) to the same discontinuity set.

We use the sine of the acute angle between discontinuity unit normal vectors as a measure of their similarity. The sine-based similarity measure has the advantage that it is simple to compute, at the same time that (as we will show) it performs well in most cases, successfully identifying anisotropic clusters in a number of test cases. Note that, however, the election of an adequate distance metric is a crucial aspect of discontinuity clustering methods. For instance, in the context of regular fuzzy K-means clustering, the topology induced by the sine-based distance measure favors isotropic cluster shapes, which leads to difficulties with discontinuity sets of elliptical shape (see e.g., Harrison, 1992; Hammah and Curran, 1998, 1999).

Given a data set with N measurements of rock discontinuity orientations to be clustered into K discontinuity sets, spectral clustering is performed by a transformation of the N discontinuity orientations to a K-dimensional space. The coordinates of the observations in the transformed space are computed using the main eigenvectors of a matrix that includes information of the (normalized) distance between observations. The advantage of spectral clustering is that transformed points cluster around K

mutually orthogonal points on the surface of the K-dimensional unit hypersphere,[1] where it is easy to perform clustering; in addition, partitions in the transformed K-dimensional space correspond to partitions in the original space (see Ng et al., 2002; Jimenez-Rodriquez and Sitar, 2006 for details).

In this work, we extend previous crisp applications of spectral clustering, with the objective of obtaining information about the uncertainties of the assignments of observations to each discontinuity set. To that end, we perform the transformation into the K dimensional space mentioned above; however, instead of performing hard clustering in the transformed space via K-means, we perform fuzzy clustering, as follows (see also Ng et al., 2002):

1. Cluster the data points in the transformed space into K subsets using the fuzzy K-means algorithm. Observations in the transformed space are assigned to the cluster to which they have the highest degree of membership; observations in the original space are assigned to the cluster to which their corresponding transformed point is assigned.
2. Assign computed degrees of membership of points to each cluster in the transformed space to their corresponding points in the original space of discontinuity orientations.

In this way, we compute estimates of the degrees of membership of discontinuity observations to each discontinuity set that can also be used to perform Monte Carlo sampling. One further advantage is that, as we show in Sect. 3, degrees of membership can be incorporated into stereographical projection plots of discontinuity orientations to indicate the uncertainty of the assignments performed.

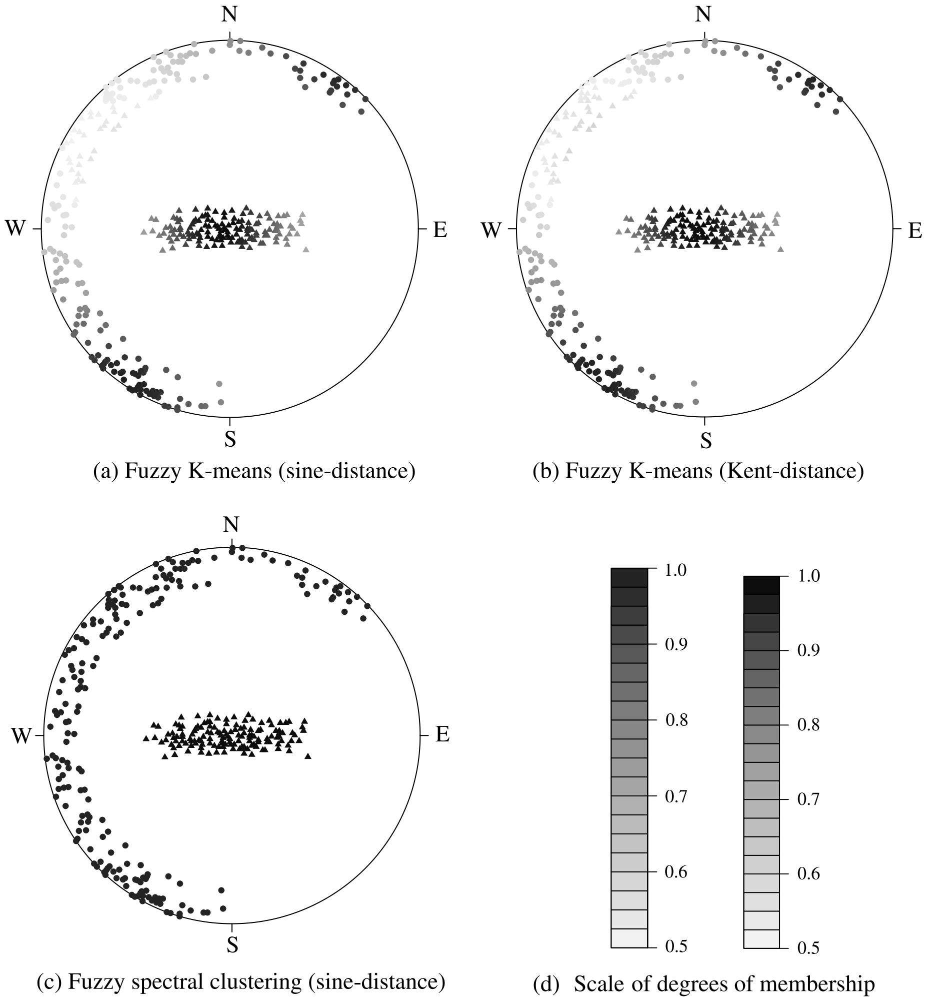
3. Example analyses

3.1 Herda et al. (1991) and Hammah and Curran (1999) data sets

In this section we work with discontinuities corresponding to Site c1904.1 from the MIT fracture-attitude data collection (Herda et al., 1991), and also with data of discontinuity orientations taken from Hammah and Curran (1999). The near-vertical fractures (i.e., near-horizontal poles) in Fig. 1, which are aligned along a girdle, are taken from Herda et al. (1991), whereas the near-horizontal fractures (i.e., near-vertical poles) are taken from Hammah and Curran (1999), and they correspond to a discontinuity set with elliptical shape.

Figure 1 shows a comparison between clustering results (considering two discontinuity sets) computed with the fuzzy K-means algorithm and with the fuzzy spectral clustering algorithm proposed in this paper. Uncertainties in the clustering assignments are represented by means of the color scales shown in Fig. 1d. That is, discontinuities are plotted using a blue point if they are assigned to set number 1 (i.e., if their degree of membership to set number 1 is higher than their degree of membership

[1] For two discontinuity sets ($K = 2$), the 2-dimensional unit hypersphere is reduced to a circle; for three discontinuity sets ($K = 3$) we have a regular unit 3-dimensional sphere; whereas for four and more discontinuity sets ($K \geqslant 4$) we have a K-dimensional unit hypersphere.



(a) Fuzzy K-means (sine-distance)

(b) Fuzzy K-means (Kent-distance)

(c) Fuzzy spectral clustering (sine-distance)

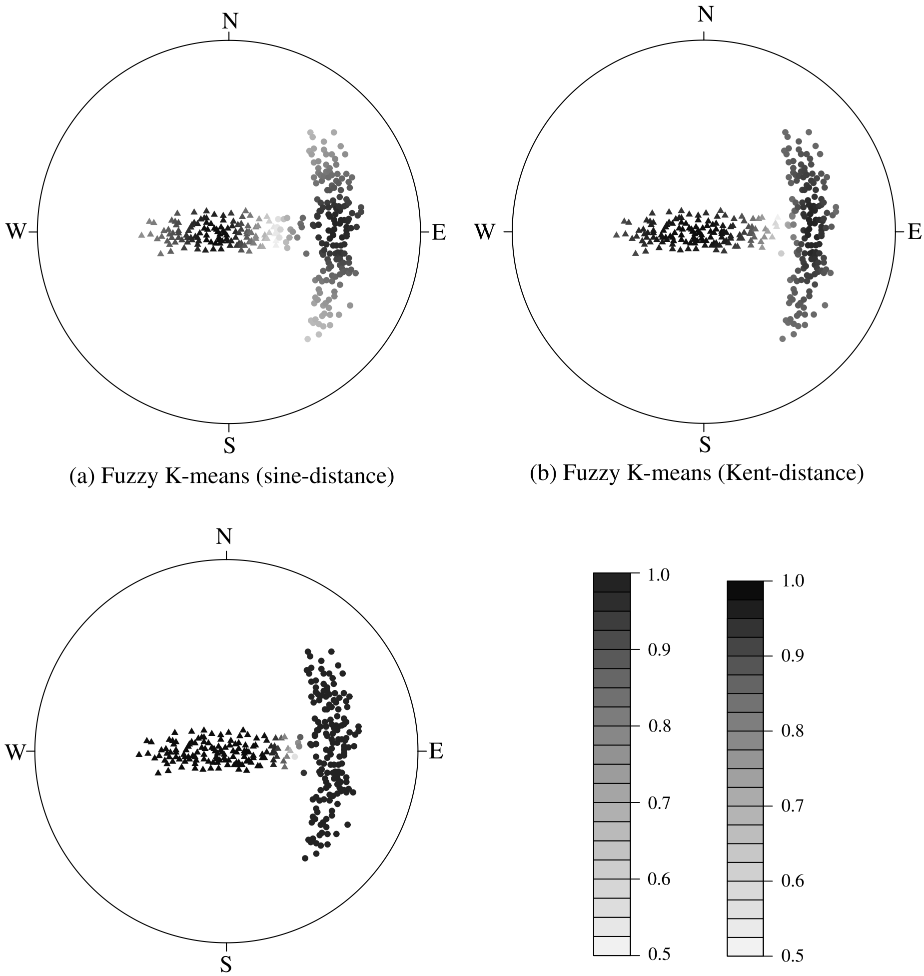
(d) Scale of degrees of membership

Fig. 1. Comparison of fuzzy K-means clustering and fuzzy spectral clustering with an orientation data set composed of an elliptical discontinuity set and discontinuities aligned along a girdle. (● set number 1; ▲ set number 2)

to any other set), whereas they are plotted using a red point if they are assigned to set number 2. Similarly, note that shading of colors indicates the degree of membership of the observations to their assigned set. In other words, points plotted with solid colors indicate certain assignments, whereas points plotted with shaded colors indicate uncertain assignments.

Figure 1c shows that the fuzzy spectral clustering method (using a sine-based similarity measure) performs well, successfully identifying discontinuities aligned along the girdle and discontinuities belonging to the horizontal elliptical discontinuity set. However, Fig. 1a and b, show that, in this case, the fuzzy K-means algorithm (both with sine-distance and Kent-distance metrics) fails to identify the natural partitions in the data set. Similarly, the solid colors in Fig. 1c (when compared with the shaded colors in Fig. 1a and b) indicate that the clustering partitions provided by traditional fuzzy K-means are significantly more uncertain (i.e., less crisp) than those computed with the fuzzy spectral clustering algorithm. (Indeed, Jimenez-Rodriguez and Sitar (2006) show that these data form tight clusters in the transformed space; accordingly, it is easy to perform fuzzy K-means and observations have high degrees of membership to their corresponding clusters.)

Figure 2 shows another comparison between clustering results computed with the fuzzy K-means algorithm and the spectral fuzzy algorithm. Discontinuity orientation data in Fig. 2 correspond to two (simulated) anisotropic discontinuity sets taken from



(a) Fuzzy K-means (sine-distance)

(b) Fuzzy K-means (Kent-distance)

(c) Fuzzy spectral clustering (sine-distance)

(d) Scale of degrees of membership

Fig. 2. Comparison of fuzzy K-means clustering and fuzzy spectral clustering with an orientation data set composed of two anisotropic discontinuity sets. (● set number 1; ▲ set number 2)

Hammah and Curran (1999). (The near-vertical pole set is the same as in Fig. 1.) As expected in this case (see Hammah and Curran (1999) for details), the fuzzy K-means algorithm with the Kent-based distance metric (Fig. 2b) performs better than with the sine-based distance metric (Fig. 2a), and it provides an improved partition of discontinuity sets. (Uncertainties in the clustering assignments are represented by means of the color scale shown in Fig. 2d.) Results also show that discontinuity sets identified with the fuzzy spectral clustering algorithm (Fig. 2c) are very similar to those computed with the fuzzy K-means algorithm and the Kent distance metric, although in this case the fuzzy K-means (Kent distance) algorithm slightly outperforms the spectral clustering algorithm. (Five discontinuities are assigned to the wrong discontinuity set in Fig. 2c, whereas only one is assigned to the wrong discontinuity set in Fig. 2b.) Finally, the solid colors in Fig. 2b and 2c indicate that, in this case, both the spectral fuzzy clustering algorithm (with the sine-based distance metric) and the fuzzy K-means algorithm provide crisp partitions into discontinuity sets.

3.2 ASM000205 outcrop data set

In this section we further test the fuzzy spectral clustering algorithm with a data set of $N = 1,173$ discontinuity orientation measurements performed by the Swedish

Fig. 3. Map of discontinuity traces recorded at ASM000205 outcrop. (Courtesy of Raymond Munier)

Nuclear Fuel and Waste Management Company (SKB) at outcrop ASM000205 in Simpevarp Peninsula, Sweden (Darcel et al., 2004). Measurements of discontinuity orientations at this outcrop are made in the context of a wider effort for site characterization at this area, with the objective of assessing the adequacy of alternative locations of a proposed geological repository for spent nuclear fuel (SKB, 2005).

The map of discontinuity traces at the outcrop is shown in Fig. 3. The rock type at the outcrop is diorite with very small grain size (Darcel et al., 2004). A high density of very small fractures can be observed at the outcrop, and fracture traces were mapped over the whole surface of the outcrop with a lower cut-off length of approximately 50 cm (Darcel et al., 2004).

Figure 4 shows an equal-area (Schmidt) lower hemisphere projection of the normal unit vectors of each discontinuity mapped. Contours that represent the estimated frequency of discontinuity normals in each direction in space are presented as well. In this work we do not discuss the identification of the optimal number of discontinuity sets (for a discussion, see e.g., Harrison, 1992; Hammah and Curran, 1998, 2000). However, contour maps of unit normal vectors such as that in Fig. 4 could help a trained analyst to decide the number of discontinuity sets that should be used in each particular application.

In this case, for instance, it seems reasonable to cluster the data into three discontinuity sets. As shown in Fig. 4, there are two obvious sets corresponding to areas of high density in the contour plot of discontinuity orientations (one with poles trending

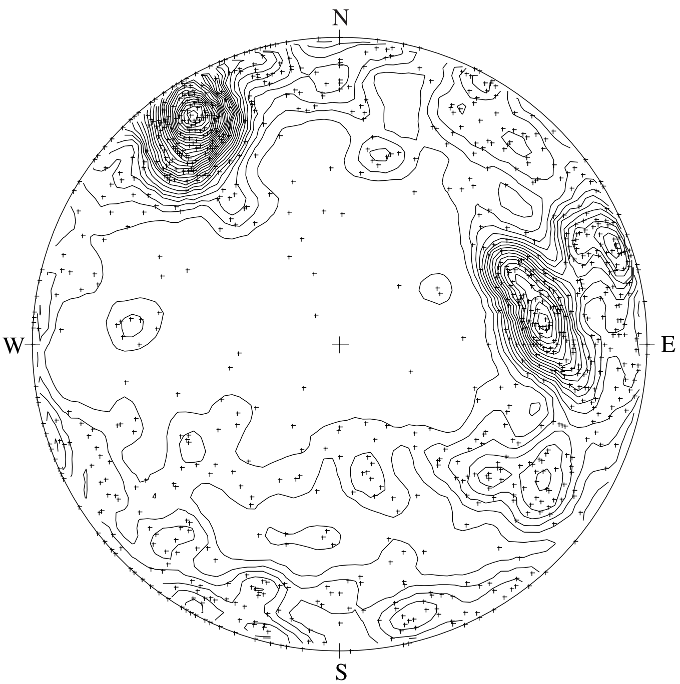
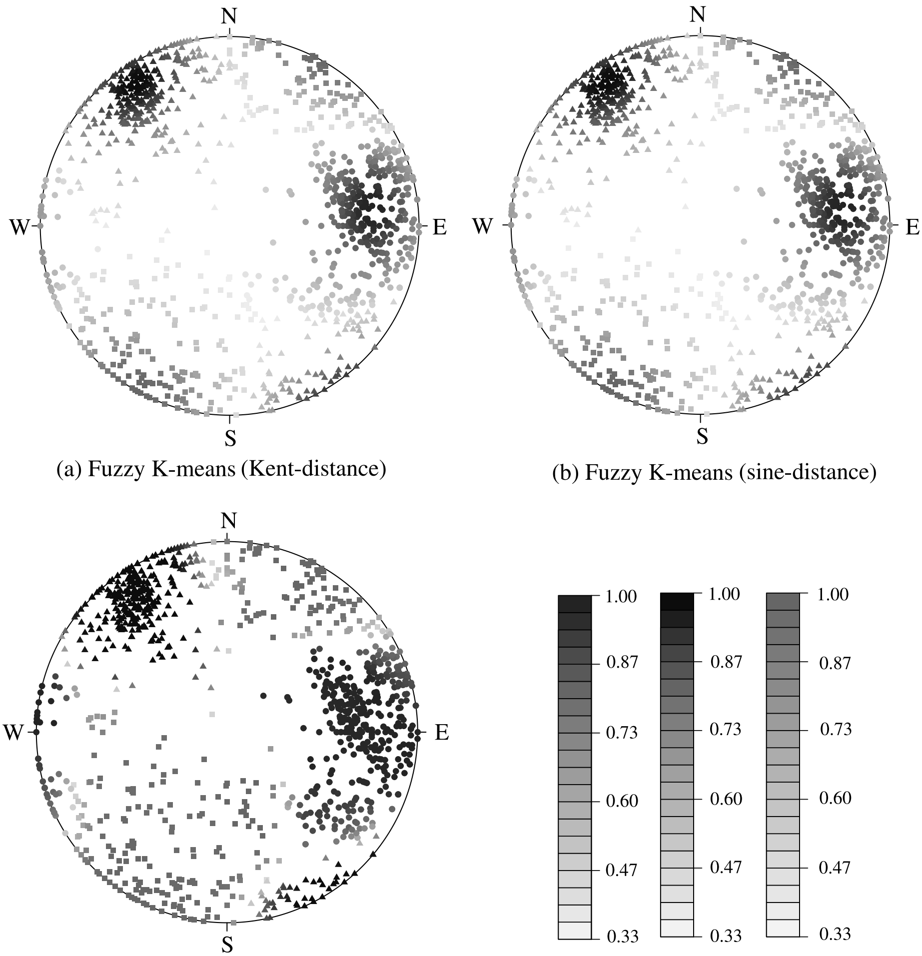


Fig. 4. Representation of Kamb's contours of discontinuity orientations in the ASM000205 data set (contour interval 2σ; significance level 3σ)

approximately in direction E–W, and the other with poles trending approximately in direction NW–SE); whereas the third set is expected to cluster discontinuity poles trending approximately in direction NE–SW. This working assumption is further supported by the observation that Darcel et al. (2004) use to partition the data set of discontinuity orientations at outcrop ASM000205 into three discontinuity sets based on their orientation.

Therefore, we also consider three discontinuity sets and, in Fig. 5, we compare the computed clustering partitions when different algorithms are employed to identify three discontinuity sets. The results show that the partitions provided by the fuzzy spectral clustering algorithm agree well with those provided by the other algorithms, and it is difficult to know which partition should be preferred in general (Klose et al.,



(a) Fuzzy K-means (Kent-distance)

(b) Fuzzy K-means (sine-distance)

(c) Fuzzy spectral clustering (sine-distance)

(d) Scale of degrees of membership

Fig. 5. Comparison of clustering results for different clustering algorithms considering $K = 3$ discontinuity sets (ASM000205 data set)

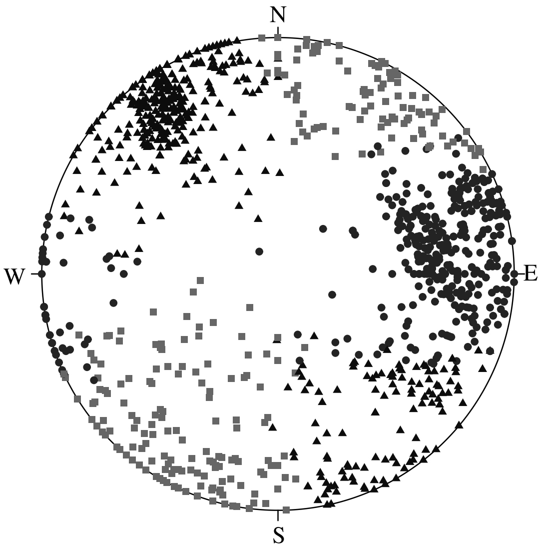


Fig. 6. Clustering results for the vector quantization clustering algorithm (Klose et al., 2005), with crisp assignments, considering $K = 3$ discontinuity sets (ASM000205 data set)

2005). However, based on visual observation, we consider that partitions computed with the fuzzy spectral clustering algorithm are at least as natural as partitions computed with the other algorithms. Similarly, our results show that, in this case, higher degrees of membership to the different partitions (i.e., lower uncertainties in the assignments) are obtained with the fuzzy spectral clustering algorithm than with the fuzzy K-means algorithm. (Differences in the uncertainties of the assignments can be observed when the more solid colors in Fig. 5c are compared with the more shaded colors in Fig. 5a and b; the scale of degrees of membership is shown in Fig. 5d.)

For the sake of completeness, partitions computed using the vector quantization algorithm proposed by Klose et al. (2005) for $K = 3$ discontinuity sets are presented in Fig. 6. As shown, such partitions agree well with the partitions of the fuzzy spectral clustering algorithm and of the regular fuzzy K-means algorithm. However, note that no information of the uncertainty of the assignments is available in this case, since the vector quantization algorithm is a crisp algorithm. Accordingly, degrees of membership are not plotted in this case.

4. Conclusions

We present a new approach for the identification of rock discontinuity sets based on their orientations by means of fuzzy spectral clustering. We also present a novel approach for graphical representation of discontinuity sets, in which the degree of membership of each discontinuity to its assigned discontinuity set (i.e., a measure of the uncertainty of the assignment) is explicitly indicated by means of a scale of shaded colors.

For clustering into K discontinuity sets, the fuzzy spectral clustering algorithm is based on performing a transformation of the original data of discontinuity orientations into a transformed space where clustering is performed. The advantage of such a transformation is that points in the transformed space are expected to cluster around K orthogonal points on the surface of the K-dimensional unit sphere. Clustering partitions in the transformed space are in general easier to identify, hence providing reduced uncertainties (i.e., higher degrees of membership) in the assignments of data points to discontinuity sets using the fuzzy K-means algorithm.

The performance of the fuzzy spectral clustering algorithm is studied using several real data sets of rock discontinuity orientations. The results show that the algorithm exhibits good clustering capabilities, even when a simple sine-based similarity measure is considered. In particular, the discontinuity set partitions computed with the spectral clustering algorithm are shown to be very similar to those computed with other clustering algorithms that are used in rock engineering applications. We also show an example case in which the fuzzy spectral clustering algorithm provides more natural partitions than the other algorithms considered, and one example in which the fuzzy K-means algorithm (with the Kent-based distance metric) performs slightly better than the fuzzy spectral clustering algorithm. Furthermore, our results indicate that, given the tight clusters that observations are expected to form in the transformed space, the computed degrees of membership of discontinuity observations to discontinuity sets are higher in general for the fuzzy spectral clustering algorithm than for the fuzzy K-means algorithm.

Acknowledgments

Christian Klose kindly provided a computer implementation of their vector quantization algorithm that was used during development of our implementation. The Swedish Nuclear Fuel and Waste Management Company (SKB) provided access to their databases of discontinuity data. Raymond Munier (SKB) provided explanations of the on-going site investigation program and reviewed an early version of the manuscript.

References

Darcel C, Davy P, Bour O, de Dreuzy J-R (2004) Alternative DFN model based on initial site investigacions at Simpevarp. Tech Rep SKB Rapport R-04-76, The Swedish Nuclear Fuel and Waste Management Co., Stockholm, Sweden. Also available at http://www.skb.se/upload/publications/pdf/R-04-76webb.pdf

Hammah RE, Curran JH (1998) Fuzzy cluster algorithm for the automatic identification of joint sets. Int J Rock Mech Min Sci 35: 889–905

Hammah RE, Curran JH (1999) On distance measures for the fuzzy K-means algorithm for joint data. Rock Mech Rock Engng 32: 1–27

Hammah RE, Curran JH (2000) Validity measures for the fuzzy cluster analysis of orientations. IEEE T Pattern Anal 22: 1467–1472

Harrison JP (1992) Fuzzy objective functions applied to the analysis of discontinuity orientation data. In: Hudson JA (ed) Rock characterization. Proc. ISRM symposium, Eurock '92. British Geotechnical Society, Thomas Telford, London

Herda HHW, Einstein HH, Dershowitz WS (1991) Problems with representation of rock fracture clusters. J Geotech Engng 117: 1754–1771

Jimenez-Rodriguez R, Sitar N (2006) A spectral method for clustering of rock discontinuity sets. Int J Rock Mech Min Sci 43: 1052–1061

Klose C, Seo S, Obermayer K (2005) A new clustering approach for partitioning directional data. Int J Rock Mech Min Sci Geomech Abs 42: 315–321

Mahtab MA, Yegulalp TM (1982) A rejection criterion for definition of clusters in orientation data. In: Goodman RE, Heuze FE (eds) Proc. 22nd US Symposium Rock Mechanics, Soc Min Eng Am Inst Min Metall Petrol Eng, pp 116–123

Munier R (2006) Personal communication

Ng AY, Jordan M, Weiss Y (2002) On spectral clustering: analysis and an algorithm. In: Dietterich TG, Becker S, Ghahramani Z (eds) Advances in neural information processing systems 14. MIT Press, Cambridge, MA, pp 849–856

Priest SD (1985) Hemispherical projection methods in rock mechanics. Allen & Unwin, London, Boston

Priest SD (1993a) The collection and analysis of discontinuity orientation data for engineering design, with examples. In: Hudson JA (ed) Comprehensive rock engineering: principles, practice & projects: rock testing and site characterization. Pergamon Press, Oxford, pp 167–192

Priest SD (1993b) Discontinuity analysis for rock engineering, 1st edn. Chapman & Hall, London, New York

SKB (2005) Preliminary site description: simpevarp subarea – version 1.2. Tech Rep SKB Rapport R-05-08; ISSN 1402-3091, The Swedish Nuclear Fuel and Waste Management Company (SKB), Stockholm, Sweden. Also available at http://www.skb.se/upload/publications/pdf/R-05-08webb-light.pdf

Terzaghi RD (1965) Sources of errors in joint surveys. Geotechnique 15: 287–304