Environ Geol (2006) 49: 821–827
DOI 10.1007/s00254-006-0177-5

ORIGINAL ARTICLE

J. Hofrichter
G. Winkler

Statistical analysis for the hydrogeological evaluation of the fracture networks in hard rocks

Received: 30 August 2004
Accepted: 12 December 2005
Published online: 25 February 2006

J. Hofrichter (✉)
Institute of Applied Statistics and System Analysis, Joanneum Research GmbH, Steyrergasse 25a, 8010 Graz, Austria
E-mail: Johannes.hofrichter@joanneum.at
Tel.: +43-316-8761556
Fax: +43-316-87691556

G. Winkler
Institute of Water Resources Management – Hydrogeology and Geophysics, Joanneum Research GmbH, Elisabethstraße 16/II, 8010 Graz, Austria
E-mail: gerfried.winkler@joanneum.at
Tel.: +43-316-8761492
Fax: +43-316-87691492

Abstract The hydrogeological effectiveness of fracture sets is determined and evaluated by the fuzzy c-mean and hierarchical clustering. These cluster analyses combine the geological spatial attributes and the hydraulic relevant attributes of fractures. Based on the results of the clustering the fracture set volumes are estimated.

Keywords Fracture network · Fracture set volume · Fuzzy c-mean clustering · Hierarchical clustering · Austria

Introduction

Large areas of Austria, like the Central Alps or the Bohemian Massif, are built up by crystalline hard rocks. Tectonic deformations of geological units cause the development of fracture networks forming fissured/fractured aquifers in hard rocks. The water flow in fractured hard rocks is predominant within its fracture network, which mostly consists of several fracture sets. For the determination of favoured flow directions in fractured hard rocks, it is necessary to determine the fracture network and the hydrogeological effectiveness of the fracture sets. The hydrogeological effectiveness of fracture sets result from two groups of fracture attributes: the first group contains the geological spatial attributes like trend, plunge and the frequency of fractures. The second group includes the hydraulic relevant attributes like aperture, trace length and linear degree of separation of the fractures. The hydrogeological effectiveness of individual fracture sets is answered with a new approach. Statistical clustering integrating both attribute types enhances the determination of the hydrogeological effectiveness of the fracture sets. Further on, the volume of each fracture set can be estimated.

Methodology

The hydrogeological effectiveness of fracture sets can be estimated by the statistical analysis of two groups of fracture attributes. The first group contains the geological spatial parameters like trend, plunge and the frequency of fractures. The second group includes the hydraulic relevant attributes like averaged aperture, length and linear degree of separation of the fractures. The terms fracture, discontinuity, joint and their networks are described by many authors in an inconsistent way. For example Bridges (1975) defines a fracture as "... a discrete break in a rock which is not parallel with a visible fabric". Excluding discontinuities that have been the result of the exploitation of cleavage Bridges concurs with Whitten and Brooks (1972). In these studies the term fracture will be used as a synonym of the term "discontinuity" in Priest (1993), who defines discontinuity "... as any significant mechanical break or fracture of negligible tensile strength in a rock." This term makes no distinction concerning the age, geometry or mode of origin of the feature. The only differentiation is between natural, that are of geomorphologic or geological origin, and artificial discontinuities, that are caused by drilling or blasting in excavation. For the three-dimensional structure of fractures, the authors use the terms fracture set and in addition fracture network. The intact, unfractured rock is referred to as the rock material, which together with the fracture network form the rock mass.

Data recording

To record and describe the fracture network of an exposure, several methods have been developed. The ISRM (1978) describes two basic levels at which a rock mass survey may be carried out upon the amount of detail that is required.

- In a *subjective* survey only those fractures, which appear to be important, are described. This level is biased by the experiences of the person (geologist) recording the data.
- In an *objective* level the fractures intersecting a fixed line (scanline) or area (grid) of rock outcrops are recorded. So, this level is standardized and independent of individual persons experiences recording the data.

At exposures, the attributes of the fractures are recorded with the scanline method based on objective and standardized criterions. As Priest (1993) mentioned, there is no universally accepted scanline sampling, so the specific data and the sampling method have to be modified and adjusted to the concrete demands. Generally, it is very important to obtain a fixed level of quality of the recorded data, independent of the person and the circumstances at the recording time. So, in the run-up to the data recording, standards for the exposure-logging and the scanline-logging have to be developed.

For this work the objective level with fixed lines (scanlines) was adopted. The scanline method is a commonly used sampling technique, that has been described and discussed by a number of authors including ISRM (1978), Long (1993) and Priest (1993).

Each exposure is marked with a point (white point on the left exposure end near the scale bar in Fig. 1), that is the coordinative origin for the localisation of the scanlines. Depending on the orientation of the main fracture sets, the scanlines are fixed by the recording geologist. The scanlines themselves are simply measuring tapes, pinned with masonry nails and wires to the rock face. Two scanlines intersecting more or less orthogonal are the minimum observing the fracture network at an exposure. The data of more than 100–150 fractures should be recorded to enhance statistical analysis. Additional scanlines, pinned on further different orientated rock faces of the same exposure enable a better spatial imagination of the fracture network and minimise orientation sampling biases. The length of the scanline depends on the spacing of the fractures regarding the quantity of fractures for satisfying statistical analysis.

In advance of scanlines logging some assumptions have to be defined. The aperture should be measured on several points along one fracture and then averaged. Fractures of the exposure face of which the aperture is immeasurable are marked separately with "n.m." (not measurable). For the statistical analysis the attributes of

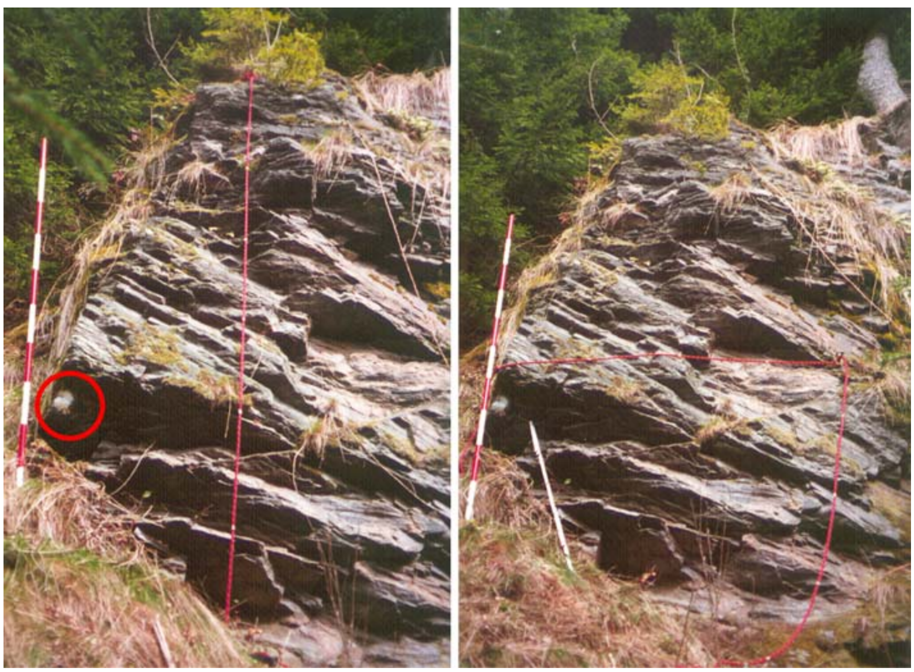

Fig. 1 Outcrop with two more or less orthographic scanlines for the data recording

these fractures are estimated with data of similarly orientated measurable fractures of the same exposure or fracture network. The total length is the complete measurable stretch of a fracture. The linear degree of separation is the sum of the trace length sections where an aperture can be observed.

Statistical data analysis: orientation related fracture volumes

Frequency analysis and structural–geological clustering are generally used for geological and tectonic interpretation of fracture networks. These statistic methods are discussed in detail by a couple of authors including Barton et al. (1993), Priest (1993) and Wallbrecher (1986). So, these methods will not be discussed closer. The new approach was to combine the spatial geological and the hydraulic relevant attributes to determine the fracture network and additionally to estimate the fracture set volumes.

Pre-processing

In advance to the statistical analysis some biases have to be corrected by weightings and assumptions/estimations. Biases are caused by: (a) sampling technique (Kulitulake et al. 1993; Bäumle et al. 1998), (b) the immeasurable attributes (aperture, length) of fractures forming the exposure face and (c) the varying orientation of scanlines to the orientation of the intersecting fractures.

Weightings The fractures are weighted by their spatial position to the scanline orientation. In the further formula the fractures are represented by their pole points (Wallbrecher 1986); so, the weight is related to the angle between the pole of the fracture and the scanline. There must be consideration that the higher the angle the higher the fracture must be weighted. The weighting function can be written as:

$$\tilde{g}_i = \frac{1}{|\cos(\varphi_i)|} = \frac{1}{|\langle x_s, x_i \rangle|},$$

where ϕ_i is the angle between scanline x_s and pole x_i and where $x_i = (x_{1i}, x_{2i}, x_{3i})$ are the x,y,z-coordinates on the surface of the unit sphere. For $\phi_i \rightarrow 0$ this weighting function goes to ∞, therefore the authors set a limit at 5, and the weighting function becomes

$$g_i^* = \min\left(\frac{1}{|\langle x_s, x_i \rangle|}, 5\right).$$

Because of the dimension of the finite-sized exposures the length of the scanlines differs. This influences the number of intersecting fractures along one scanline depending on the total length of the scanline. For comparing different scanlines, the fractures must be weighted by the length of the scanline. The final weighting function can be defined as

$$g_i = \frac{g_i^*}{l_s} = \frac{1}{l_s}\min\left(\frac{1}{|\langle x_s, x_i \rangle|}, 5\right),$$

where l_s is the length of the scanline, on which the fracture is detected. So this weighting considers for each fracture the angle to the scanline and its length.

Estimations and assumptions Estimation of immeasurable fracture attributes: The aperture and sometimes the lengths of fractures forming the exposure face are not measurable. The rock face is subjected to the fracture network. As fractures are characterized by a negligible tensile strength the loosening of hard rocks follows the fracture network. So, fractures forming rock faces are assumed to be hydrogeologically relevant and are considered as open. In most cases exposures have several rock face directions that scanlines can be stretched. So, it is possible to measure the aperture and length of similar orientated fractures. With that data the immeasurable attributes of fractures forming the rock face can be estimated. The attributes a_0 of a fracture forming the rock face are estimated by the attributes a_i of other fractures using the weighted mean

$$\hat{a}_0 = \bar{a}_g = \frac{\sum_i w_i g_i a_i}{\sum_i w_i g_i},$$

where w_i is the weight of the individual fractures and defined as

$$w_i = \begin{cases} \exp(-4(1 - \langle x_0.x_i \rangle^2)) & \varphi_i \leq 30 \\ 0 & \text{else} \end{cases},$$

where ϕ_i is the angle between x_0 and x_i. The weights are defined in a way that fractures parallel to the main direction x_0 influence the equation most strongly. At an angle greater than 30° the weights are set to zero. That means fractures with an direction deviation more than 30° are not considered for the estimation. If there are no other fractures within that angle deviation, then the estimated fracture attributes are set to zero. One problem has to be mentioned in that context; a very low number of fractures having lower angles than 30° do not give satisfying results. So, there is still some developing necessary; for the first approach, there are no further limitations used.

Assumptions for the weightings and the fracture volume: Another goal of the fracture analysis is the determination of the fracture set volume. Therefore the individual fractures are weighted by the attributes aperture and linear degree of separation. The authors assume that the attributes recorded on the surface continue through the whole rock mass. So the area defined by the aperture and linear degree of separation can be regarded as a representative weighting value. The maximum aperture is fixed with 1 cm. The maximum length of fractures extending out of the exposure are standardized to 2 m.

Considering these assumptions the weights gv_i are given by:

$$gv_i = \mathrm{o}_i \times \mathrm{d}_i, \tag{1}$$

where o_i is the aperture and d_i is the standardized linear degree of separation.

Combining the two weights g_i and gv_i the total weight v_i for one fracture is:

$$v_i = g_i \times gv_i. \tag{2}$$

This total weight describes the hydrogeological importance of a fracture based on the fracture volume.

Cluster analysis

The cluster analysis classifies objects to groups (clusters). The classification is based on the similarity of the objects attributes. The clusters are characterized by high similarity within the cluster and high dissimilarity between the clusters.

So the similarity respectively dissimilarity (=distance) between the objects have to defined.

Distances and similarities The similarity between objects or object groups can be defined in many ways. The cluster analysis describes the similarity of two objects with the distance between this two objects. That means the stronger the similarity the smaller the distance. The definitions of different statistical distances is well discussed by many authors including Hammah et al. (1998, 1999) and Steinhausen and Langer (1977). Let d_{ij} be a function of two objects x_i and x_j. The authors write

$$d_{ij} = d(x_i, x_j) \quad \text{with} \quad x_i = (x_{i1}, x_{i2}, \ldots x_{ip})\prime \quad \text{and} \quad i = 1, 2, \ldots, n,$$

where n is the number of observed objects and p the number of the attributes. If d_{ij} complies with the first three conditions:

1. $d(x,y) \geq 0$
2. $d(x,y) = 0 \Leftrightarrow x = y$
3. $d(x,y) = d(y,x)$
4. $d(x,y) \leq d(x,z) + d(z,y)$

then d_{ij} is a distance. A metric distance also complies with the fourth condition. As mentioned above, the attributes of an observed fracture can be separated into two groups. The measured values of these two different groups are defined in two different spaces. The measured values of the spatial attributes are spherical data and can be represented by the surface of the unit sphere in $\mathbf{R}^3$ space. The measured values of the hydrogeological attributes are numeric data in $\mathbf{R}^2$ space. So, two different kinds of distances are needed and should be combined in a suitable way.

For the spherical data the authors consider the sine-square distance and the Mahanalobis distance of spherical data (Hammah et al. 1999). Let x_i and x_j be two points on the surface of the unit sphere, then the sine-square distance is defined as

$$d^2(x_i, x_j) = 1 - \langle x_i, x_j \rangle^2.$$

For the definition of the Mahanalobis distance for spherical data see Hammah et al. (1999). The common distance of numeric data is the Euclidean distance

$$d(x_i, x_j) = \left\| x_i - x_j \right\|$$

where $||\cdot||$ is the Euclidean norm. The authors used the Euclidean distance for analysing the hydraulic attributes.

To combine the different kinds of distances, Steinhausen et al. 1977 propose a mixed distance d_{M}. Let d_{r} be the distance of the direction (sine-square or Mahalanobis of spherical data) and d_{e} the Euclidean distance, then d_{M} is defined as

$$d_{\mathrm{M}} = \frac{m_{\mathrm{r}} d_{\mathrm{r}}}{m} + \frac{m_{\mathrm{e}} d_{\mathrm{e}}}{m},$$

where m_{r} is the number of the variables representing the direction (in the discrete demand $m_{\mathrm{r}} = 3$), m_{e} is the number of variables included in the Euclidian distance (in the discrete demand $m_{\mathrm{e}} = 2$) and out of that $m = m_{\mathrm{r}} + m_{\mathrm{e}}$. This means that the distance of the direction is weighted more strongly. For the hydrogeological interpretation this fact makes sense because the first goal is to summarize fractures as clusters with similar orientation. But also includes their hydrogeological relevant attributes for the discrete classification to the different clusters.

Clustering Agglomerative hierarchical clustering: The agglomerative hierarchical clustering leads to an exact partition of objects, that means that one object is classified to exactly one group. At the beginning every cluster is defined by exactly one object. The two clusters with the lowest distance are combined to a new cluster. So, the total number of clusters is reduced by one. This step is repeated till the demanded number of clusters is reached. For the distance between two clusters the authors use the "mixed distance" as described above. After the first step one cluster can include one or more than one object. So the authors derive the distance between two clusters with the centroid method calculating the "mixed distance" between the mean of the clusters.

Fuzzy c-mean clustering: The fuzzy c-mean clustering differs from the hierarchical clustering in two points. The number c of cluster has to be determined before starting the cluster analysis and each object is classified to each cluster with a certain degree of membership. So, the fuzzy c-mean clustering is based on the definition of diffuse sets by generalised characteristic functions (Bezdek 1981; Bäck 1994).

Let M be a subset of the real numbers, the characteristic function m_M of the set M is defined by:

$$m_M : \mathrm{R} \to \{0,\ 1\} \quad \text{with} \quad m_M(x) = \begin{cases} 1 & x \in M \\ 0 & \text{else} \end{cases}$$

This gives an exact classification of each element to one set. The generalised characteristic function m_M of a set M is defined by:

$$m_M : \mathrm{R} \to [0,\ 1] \quad \text{and} \quad m_M(x) = \begin{cases} f(x) & x \in M \\ 0 & \text{else} \end{cases} \quad \text{with} \quad f : \mathrm{R} \to [0,\ 1].$$

That means that the elements can be classified to each group with a certain degree of membership. For the better understanding it will be explained by an example:

M is the set of elements which are "approximately 7". A characteristic function m_1 of the exact set M_1 can be:

$$m_1(x) = \begin{cases} 1 & 6.95 \leq x \leq 7.05 \\ 0 & \text{else} \end{cases}$$

and a generalised characteristic function m_2 of the set M_2 can be:

$$m_2(x) = \exp\left(-(x-7)^2\right) \quad x \in (-\infty, \infty).$$

With the characteristic function m_1 all the values between 6.95 and 7.05 are classified to the exact set M_1 (approximately 7). Under the use of the generalised characteristic function m_2 only the value 7 is exactly classified to the diffuse set M_2 and has a degree of membership equal one. The remained elements have a degree of membership less than one.

That means for the cluster analysis, that the objects are not classified to exactly one cluster. The classification to the different clusters is defined by the degree of membership w_{ij} with

$$0 \leq w_{ij} \leq 1 \quad \text{and} \quad \sum_j w_{ij} = 1,$$

where $i = 1,2,...n$ is the number of objects and $j = 1,2,...c$ is the number of clusters. The diffuse partition is uniquely fixed by the degrees of membership.

The fuzzy c-mean clustering is an iterative method. The first step is to determine the number c of cluster and c arbitrary cluster centeroids (prototypes) V_j with $j = 1,2,..c$. The next step is to calculate the distances between the objects x_i and V_j and the degrees of membership

$$w_{ij} = \frac{\left[1/\left(d^2(X_i, V_j)\right)\right]^{1/(m-1)}}{\sum_k \left[1/\left(d^2(X_i, V_k)\right)\right]^{1/(m-1)}}.$$

These new degrees of membership determine new cluster prototypes by

$$V_j = \frac{\sum_{i=1}^n (w_{ij})^m x_i}{\sum_{i=1}^n (w_{ij})^m}.$$

This process has to be repeated as long as the degrees of membership change less than an established tolerance. The weighting exponent $m > 1$ defines the "fuzzyness" of the partition. The closer the value is to 1 the exacter is the partition. Otherwise when the value m becomes bigger the weighting function gets smoother that leads to a diffuser partition. Every diffuse partition can be transformed to an exact one. The modified degrees of membership $w*_i$ of the next exact partition are defined by

$$w_i^* = \begin{cases} 1 & w_{ij} = \max_j(w_{ij}) \\ 0 & \text{else} \end{cases}$$

This definition classifies the objects to exact one cluster, so the results of both methods can be compared.

The result of the clustering is the definition of groups containing fractures with similar spatial and hydrological attributes which can be regarded as homogenous groups (clusters).

Cluster attributes: The centre of gravity describes the mean orientation of a fracture set (Wallbrecher 1986). The uncertainty of the centre of gravity can be described by cones of confidence (95 and 99%). In the structural geology the cone of confidence is parametrically estimated. Therefore, the data are assumed having a certain spatial distribution. Before calculating the angle of the cone of confidence, the parameters of the distribution have to be estimated. One approach which is not bound on a certain spatial distribution is based on the non-parametric bootstrap method (Davison and Hinkley 1997). The advantage of that method is that the shape of the cone of confidence is a result of the empirical distribution of the data. So, the shape of the cone of confidence is not bound on the shape of the assumed spatial distribution.

The orientation related fracture volumes define the hydrogeological effectiveness of fracture sets. For the calculation of the fracture volume, it was assumed that the aperture and the linear degree of separation continue into the rock mass as they are recorded on the surface. Multiplying the aperture and the standardized linear degree of separation (1) leads to the standardized area (equal to the weight gv_i). Correcting the standardized area like the weights in (2) leads to a weighted area that is proportional to the fracture volume. So, the sum of the weighted areas can be regarded as a first estimation of the fracture set volume.

Results and discussion

The practical application of this research work has been carried out in the area of Sonnwendstein/Semmering—Austria (S6 Semmering tunnel) (Harum et al. unpublished data). The fracture networks are observed at 17 exposures. The 1998 fractures and their attributes are recorded along 48 scanlines with a total length of 126.5 m. The data of each exposure are analysed with the two clustering methods considering the different distances. For each exposure a structural plot is generated by clustering combining the spatial geological and the hydraulic relevant fracture attributes like Fig. 2b. Further on, the volume of each cluster (fracture sets) is calculated.

The clustering combining the geological spatial attributes and the hydraulic relevant attributes lead sometimes to a different cluster distribution than the geological analysis (Fig. 2). The geological clustering is only bound on the orientation of the fractures. In the example SA65 the geological analysis determines four different clusters (Fig. 2a). The fuzzy c-mean clustering also leads to four different clusters but with a different distribution (Fig. 2b). Both methods show for Cl1 and Cl2 the same results. But the strongly scattering cluster Cl4 of Fig. 2a can be divided into two clusters having completely different fracture set volumes (Fig. 2b). The estimated cluster volume of Cl3 in Fig. 2b is about 1.68% and the volume of Cl4 in Fig. 2b is about 0.02%. The fractures of Cl3 in Fig. 2a have nearly the same hydraulic relevant attributes as the fractures dipping S with an angle about 80°. So these data are summarized to one cluster Cl4 in Fig. 2b. The hydraulic relevant fracture sets characterized by a high fracture set volume strike SW–NE (Cl3) and N–S (Cl1) both dipping nearly vertical.

The new clustering methods enhance (a) the calculation of the cluster volumes and (b) a more precise definition of the orientation of hydraulic relevant clusters, especially by strongly dispersing fracture data.

The analysis of all the exposures determine that the fuzzy c-mean clustering using the mixed distance (Euclidean and Mahalanobis), lead to the most satisfying results considering the geological, tectonic circumstances.

Conclusions

The clustering method enhances the description of the fracture sets in more detail, combining the geological spatial and the hydraulic relevant attributes. It is an

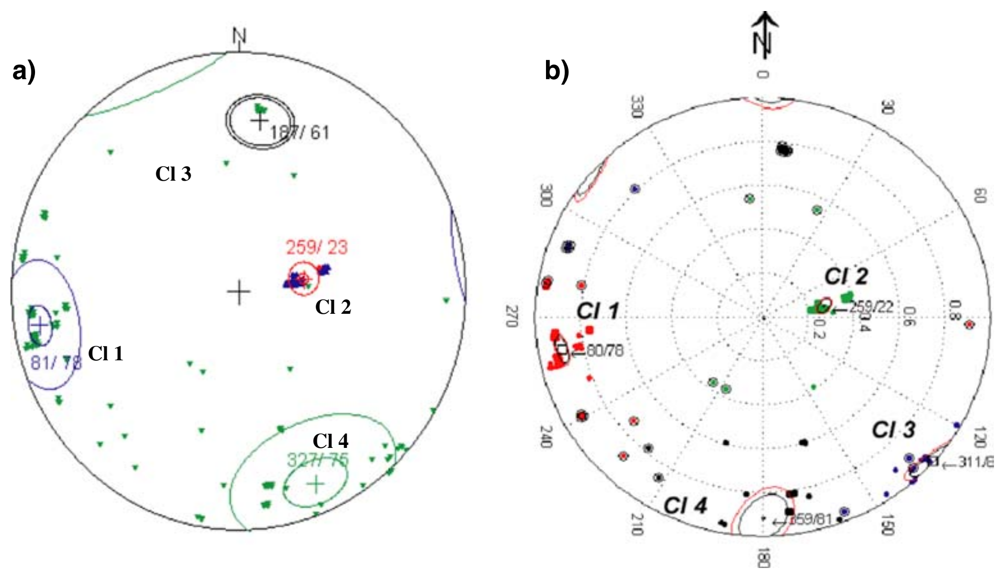

Fig. 2 The different spatial positions of the clusters Cl3 and Cl4 are figured out comparing the geological clustering (**a**) and the hydrogeological clustering (**b**) including the hydraulic relevant attributes

objective and comprehensible method which is based on data observed at surface exposures and boreholes. It enhances the determination of the hydrogeological relevant (dominant) fracture sets and separates clusters with equal geological but with different hydraulic attributes (Fig. 2).

Because of the linear sampling technique (scanline), boreholes can be included as an important complementary information for the interpretation into depth. So, this method enhances a better three-dimensional image of the fracture network and a better characterisation of hydrogeological units.

The combination of the results of the cluster analysis and results of hydraulic tests could quantify the hydraulic capacity of individual fracture sets. It can help determine and quantify the hydraulic attributes of a fracture network and their spatial distribution for a numerical realisation of fractured aquifers. On one side, the detailed fracture recording enhances the generation of a discrete fracture network with its attributes. On the other side, the hydraulic behaviour of the fracture network can be described by the hydraulic capacity of the fracture network resulted from the combination of cluster analysis and hydraulic tests. So, the results of this method can also be used for a continuum approach in numerical modelling.

The statistical analyses are based on several assumptions, which should be defined in more detail in the future. The estimation of the fracture set volumes has to be regarded as a first approximation and should be seen as an upper limit of the potential fracture volumes. It is planned to integrate the termination and a proper geometry of the individual fractures in the next investigation step and to combine the results of the clustering and volume estimation/calculation with results of hydraulic tests along boreholes. This combination could enable a quantification of the hydraulic effectiveness of the fracture network expressed in permeability and storativity.

Acknowledgements This methodology was developed within a research project supported by the Federal Ministry of Transport, Innovation and Technology of Austria and Joanneum Research Forschungsgesellschaft mbH.

References

Bäck C (1994) Fuzzy Clusterverfahren und Validitätsmaße für unscharfe Partitionen. Unveröffentl. Diplomarbeit. Institut für Statistik, TU- Graz, Graz

Barton CC, Larsen E, Page WR, Howard TM (1993) Characterizing fractured rock for fluid-flow, geomechanical and paleostress modelling: methods and preliminary results from Yucca Mountain. Open-File Report 92-269, US Geological survey, Denver, pp 64

Bäumle CC, Hötzl H, Thüringer C, Witthüser K (1998) Untersuchungen zur Einrichtung eines Festgestein-Naturmeßfeldes. AGK 50, Karlsruhe, pp 29–50

Bezdek J (1981) Pattern recognition with fuzzy objective function algorithms. Plenum, New York

Bridges MC (1975) Presentation of fracture data for rock mechanics. In: Proceedings of the 2nd Australia–New Zealand conference on geomechanics, Brisbane. Institution of Engineers, Australia

Davison AC, Hinkley DV (1997) Bootstrap methods and their application. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge

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 Eng 32(1):1–27

ISRM (1978) Suggested methods for the quantitative description of discontinuities in rock mass. Int J Rock Mech Min Sci 15:319–368

Kulitulake PHW, Wathgala DN, Stephansson O (1993) Joint network modelling with a valiation exercise in Stripa Mine Sweden. Int J Rock Mech Min Sci 30:503–526

Long JCS (1993) Construction of equivalent discontinuum for fracture hydrology. In: Hudson JA (ed) Comprehensive rock engineering, vol 3. Pergamon press, Oxford, pp 241–295

Priest SD (1993) Discontinuity analysis for rock engineering, 1st edn. Chapman and Hall, London, 473 pp

Steinhausen D, Langer K (1977) Clusteranalyse Einführung in Methoden und Verfahren der automatischen Klassifikation. Walter de Gruyter, Berlin New York

Wallbrecher E (1986) Tektonische und gefügeanalytische Arbeitsweisen. Ferdinand Enke Verlag Stuttgart, Stuttgart, 244 pp

Whitten DGA, Brooks JRV (1972) A dictionary of geology. Penguin, Harmondsworth