Fitting Mixtures of Kent Distributions to Aid in Joint Set Identification

David PEEL, William J. WHITEN, and Geoffrey J. MCLACHLAN

When examining a rock mass, joint sets and their orientations can play a significant role with regard to how the rock mass will behave. To identify joint sets present in the rock mass, the orientation of individual fracture planes can be measured on exposed rock faces and the resulting data can be examined for heterogeneity. In this article, the expectation–maximization algorithm is used to fit mixtures of Kent component distributions to the fracture data to aid in the identification of joint sets. An additional uniform component is also included in the model to accommodate the noise present in the data.

KEY WORDS: Directional data; Expectation–maximization algorithm; Kent distribution; Mixture models.

1. INTRODUCTION

In nature, a rock mass is almost never a single block of uniform solid rock. Instead, it usually contains many fracture or discontinuity planes. Because these fracture planes arise from a force acting on the rock mass, they generally form in a nearly parallel fashion. These families of parallel fracture planes are called joint sets. Figure 1 shows a hypothetical cross section of a rock mass with a single joint set.

Because, over time, forces from more than one direction often act on a rock mass, a number of joint sets may be present (see Fig. 2). How the rock mass reacts to new external forces is greatly affected by the orientation of these joint sets, because the rock mass will most likely separate at its weakest point—the fracture planes.

In the mining industry it is advantageous to be able to predict how a rock mass will react. Obviously, when modeling a mine tunnel, the rock structure of the roof of a tunnel is extremely important. Other applications are in the design of ore extraction systems and the stability of open pit walls.

When more than one joint set is present, blocks are formed (see Fig. 3). The shape and location of these blocks is very significant. With certain configurations, the roof will be very unstable, and the blocks will fall and cause a cave in. This is the case in the block caving method of mining, where the roof is purposely caved in and the collapsed rock is progressively removed. In other cases, this situation can be very dangerous; at worst, costing lives; at best, costing extra time and resources to clear the debris. In both situations, it is vital to be able to predict the rock mass structure with some confidence.

To determine the joint sets that are present, measurements are taken on site of fracture planes evident on an exposed rock face, by taking one or more lines across an exposed rock surface. Any fracture planes that intersect this line are described by two angles, dip direction and dip angle (as shown in Fig. 4), which correspond to the direction of the normal to the fracture plane. Other measurements are also taken, such as distance along the measurement line and the nature of the fracture.

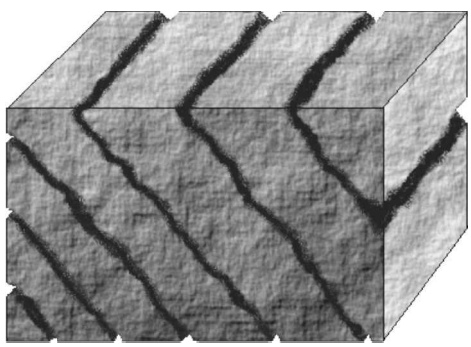


Figure 1. Hypothetical Cross Section of Rock Mass With a Single Joint Set.

In essence, a sample is being taken of the fracture planes present internally in the rock structure, which can have two major biases. First, only the fracture planes that are visible on the rock face are sampled and, second, only the fracture planes that intersect the horizontal line are sampled. In the first case, the probabilities of fracture planes being visible on the rock face are not equal; for example, fracture planes nearly parallel to the rock face will not be exposed. The second bias occurs

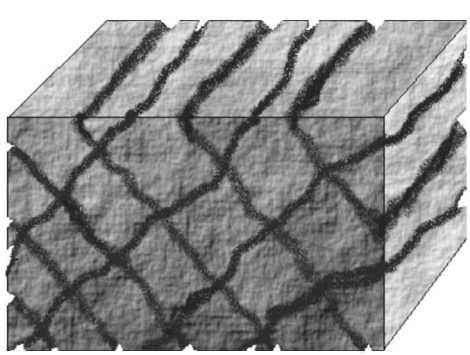


Figure 2. Hypothetical Cross Section of Rock Mass With Two Joint Sets.

David Peel is Postdoctoral Research Fellow and Geoffrey J. McLachlan is Professor, Department of Mathematics, University of Queensland, St. Lucia, Queensland 4072, Australia (E-mails: *dp@maths.uq.edu.au; gjm@maths.uq.edu.au*). William J. Whiten is a Professorial Research Fellow, Julius Kruttschnitt Mineral Research Centre, University of Queensland, St. Lucia, Queensland 4072, Australia (E-mail: *w.whiten@mailbox.uq.edu.au*). The authors thank Professor E. T. Brown, Dr. Gideon Chitombo, and other members of the mining research group at the JKMRC for providing the data used in this article and information regarding the application.

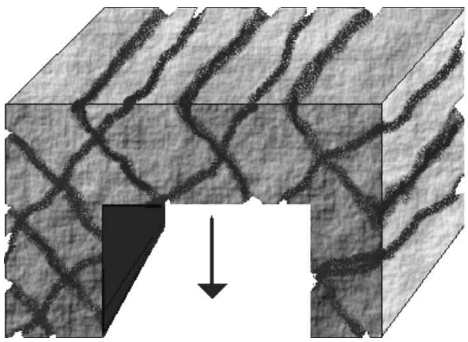


Figure 3. Hypothetical Cross Section of Rock Mass With Two Joint Sets That Form Blocks.

because a line is used for sampling and not all fracture planes are equally likely to be measured; that is, fracture planes that are nearly parallel to the horizontal will be unlikely to intersect with the horizontal measurement line. The bias of the sampling may need to be considered when interpreting the results, because it can result in gaps or holes in the data.

The first problem cannot easily be addressed, although if several rock faces are available for examination, this will not be as great a problem. The second problem is often reduced in practice by taking additional measurement lines perpendicular to the original line used to measure the data.

The fracture measurements are used to discern clusters of fractures that correspond to joint sets. Interpretation of joint set data is normally done with the aid of a projection of the joint normals from a hemisphere onto a plane. These are typically contoured (e.g., Palmstrom 1985; Priest 1993; Sullivan, Duran, and Eggers 1992; Webber and Gowans 1996) manually or by an automatic density contouring program (see Fig. 5). Interpretation of the density contours is then made manually, and thus subject to the biases of the person making the interpretation. Priest (1993) suggested a clustering method based on a search cone of a user-specified size, and then using a Fisher distribution if statistical properties of the distribution are required. He noted that that the Fisher distribution is not suitable if the cluster is not symmetric. Other methods for the cluster analysis of joint set data are discussed in Pal Roy (1995), Aler, Du Mouza, and Arnould (1996), and Hammah and Curran (1998).

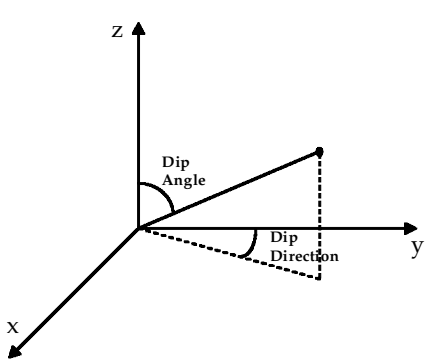


Figure 4. Representation of Dip Angle and Dip Direction.

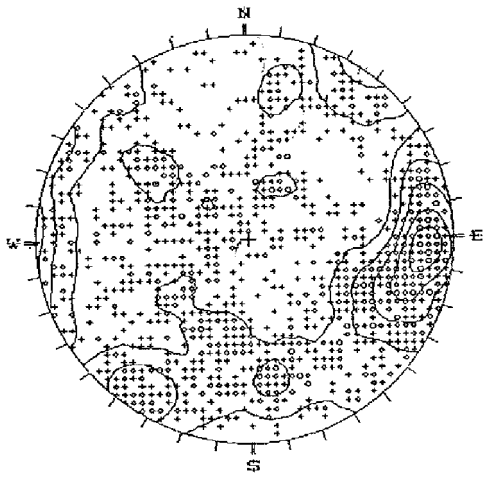


Figure 5. Example of a Polar Contour Plot of the Dip Direction and Dip Angle.

The aim of the method described in this article is to use clustering techniques to provide an automated tool to identify joint sets.

2. DIRECTIONAL DATA

The dip direction and dip angle measured are directional vectors, so the measured samples are two-dimensional directional data. A two-dimensional directional sample can simply be thought of as a sample of points that lie on the surface of a unit sphere. Hence the sample points can be represented in either polar or Cartesian coordinates. Denoting the polar coordinates by (γ_1, γ_2) $(0 \le \gamma_1 \le 2\pi,\ 0 \le \gamma_2 \le \pi/2)$, where γ_1 indicates the dip angle and γ_2 is the dip direction, as described previously, the Cartesian coordinates of a point are given by the directional cosines

$$x_1 = \cos(\gamma_2),$$
$$x_2 = \sin(\gamma_2)\cos(\gamma_1),$$
$$x_3 = \sin(\gamma_2)\sin(\gamma_1),$$

where we have followed the notation of Kent (1982) in using x_1, x_2, and x_3 to refer to the z, y, and x coordinates, respectively.

The polar coordinates (γ_1, γ_2) can be obtained from the directional cosines by

$$\gamma_1 = \tan^{-1}(x_3/x_2), \qquad \gamma_2 = \cos^{-1}(x_1).$$

It is also useful to note that the angle, θ, between two vectors $\mathbf{x}$ and $\mathbf{y}$ with directional cosines (x_1, x_2, x_3) and (y_1, y_2, y_3), respectively, is given by the inverse cosine of the dot product of the two vectors; that is,

$$\theta = \cos^{-1}(x_1y_1 + x_2y_2 + x_3y_3).$$

In the sequel, both polar and Cartesian representations will be used.

3. MIXTURE MODEL APPROACH USING KENT AND UNIFORM COMPONENT DISTRIBUTIONS

Finite mixtures provide a sound statistical based approach to cluster analysis as advocated in McLachlan and Basford (1988); see also Banfield and Raftery (1993) and Fraley and Raftery (1998). Let $\mathbf{y}_1, \ldots, \mathbf{y}_n$ denote an observed p-dimensional random sample of size n drawn from a heterogeneous population of g groups. The mixture model approach assumes each data point is a realization of the random p-dimensional vector $\mathbf{Y}$, which has probability density function

$$f(\mathbf{y}; \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i c_i(\mathbf{y}; \boldsymbol{\theta}_i), \tag{1}$$

where the mixing proportions π_i are nonnegative and sum to 1, c_i denotes the probability density function of the ith component, and $\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_{g-1}, \boldsymbol{\theta}^T)^T$, where $\boldsymbol{\theta}$ consists of the elements of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g$ known a priori to be distinct.

Under the assumption that $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are independent realizations of the feature vector $\mathbf{Y}$, the log-likelihood function for $\boldsymbol{\Psi}$ is given by

$$\log L(\boldsymbol{\Psi}) = \sum_{j=1}^{n} \log \sum_{i=1}^{g} \pi_i c_i(\mathbf{y}; \boldsymbol{\theta}_i). \tag{2}$$

With the maximum likelihood approach to the estimation of $\boldsymbol{\Psi}$, an estimate is provided by an appropriate root of the likelihood equation

$$\partial \log L(\boldsymbol{\Psi}) / \partial \boldsymbol{\Psi} = \mathbf{0}. \tag{3}$$

A common choice for the component probability density functions c_i in (1) is the multivariate normal due to its computational tractability. In the case of directional data, an obvious choice for the component distributions c_i is the Fisher distribution (due to R. A. Fisher), which can be thought of as an extension of the von Mises distribution from a circle to a sphere.

The Fisher distribution is analogous to a circular bivariate normal, which corresponds to a normal distribution with a diagonal covariance matrix with equal eigenvalues. This allows for distributions of circular shape of varying size on the surface of the sphere.

However, in this article, Kent distributions (Kent 1982) are used rather than Fisher distributions, so as to provide greater flexibility. The Kent distribution can be thought of as a generalization of the Fisher distribution or as a special case of the more general eight-parameter family of distributions known as Fisher–Bingham distributions; see Kent (1982). In the same way that the Fisher distribution is comparable to a bivariate normal with a constrained covariance matrix, the Kent distribution is comparable to a bivariate normal where the covariance matrices are unconstrained. This allows for distributions of any elliptical shape, size, and orientation on the surface of the sphere. Hence the use of Kent distributions provides a more flexible alternative to model the data.

The Kent density is given in terms of the directional cosines $\mathbf{y} = (y_1, y_2, y_3)^T$ and is given by

$$f_K(\mathbf{y}; \boldsymbol{\theta}) = C_K \exp\{\kappa(\mathbf{y}^T \boldsymbol{\xi}_1) + \beta(\mathbf{y}^T \boldsymbol{\xi}_2)^2 - \beta(\mathbf{y}^T \boldsymbol{\xi}_3)^2\}, \tag{4}$$

where for large κ

$$C_K \approx \frac{\exp(-\kappa)\sqrt{\kappa^2 - 4\beta^2}}{2\pi}$$

and $\boldsymbol{\theta} = (\kappa, \beta, \boldsymbol{\xi}_1^T, \boldsymbol{\xi}_2^T, \boldsymbol{\xi}_3^T)^T$ is the parameter vector. The parameter $\boldsymbol{\xi}_1$ is the vector of the directional cosines that define the mean or center of the distribution. The parameters $\boldsymbol{\xi}_2$ and $\boldsymbol{\xi}_3$ relate to the orientation of the distribution.

For a mixture of g Kent distributions, the probability density function is given by

$$f(\mathbf{y}; \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i f_K(\mathbf{y}; \boldsymbol{\theta}_i), \tag{5}$$

where $f_K(\mathbf{y}, \boldsymbol{\theta}_i)$ is the Kent density with parameter vector $\boldsymbol{\theta}_i = (\kappa_i, \beta_i, \boldsymbol{\xi}_{i1}^T, \boldsymbol{\xi}_{i2}^T, \boldsymbol{\xi}_{i3}^T)^T$ $(i = 1, \ldots, g)$.

The data collected in the actual application contain a significant amount of noise. To model this, an extra component is included in the mixture model, corresponding to Poisson noise, as proposed in Banfield and Raftery (1993); see also Campbell, Fraley, Murtagh, and Raftery (1997). In the present context, a uniform distribution on the unit sphere was used, which is defined by

$$f_0(\mathbf{y}) = \frac{1}{4\pi}, \qquad \mathbf{y} \in \mathcal{Y}, \tag{6}$$

where $\mathcal{Y}$ denotes the surface on the unit sphere. We let π_0 be the mixing proportion associated with the noise component so that now $\sum_{i=0}^{g} \pi_i = 1$. The mixture model can now be written as

$$f(\mathbf{y}; \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i f_K(\mathbf{y}; \boldsymbol{\theta}_i) + \pi_0 f_0(\mathbf{y}). \tag{7}$$

This formulation allows the data to determine the amount of noise present via the mixing proportion π_0, rather than the user setting the level in some ad hoc manner.

4. APPLICATION OF THE EXPECTATION–MAXIMIZATION ALGORITHM

Solutions of (3) can be found via the expectation–maximization (EM) algorithm of Dempster, Laird, and Rubin (1977); see also McLachlan and Krishnan (1997). On the $(k+1)$th iteration of the EM algorithm in the present context, the E step is equivalent to replacing the unobservable component-label indicators by their current conditional expectations, which are the current posterior probabilities of component membership of the data, $\tau_{ij}^{(k)}$, given by

$$\tau_{ij}^{(k)} = \frac{\pi_i^{(k)} f_K(\mathbf{y}_j; \boldsymbol{\theta}_i^{(k)})}{\sum_{h=1}^{g} \pi_h^{(k)} f_K(\mathbf{y}_j; \boldsymbol{\theta}_h^{(k)}) + \pi_0 f_0(\mathbf{y}_j)}$$

for $i = 1, \ldots, g$; $j = 1, \ldots, n$.

The M step requires finding the value of $\boldsymbol{\Psi}$, $\boldsymbol{\Psi}^{(k+1)}$ that globally maximizes the objective function $Q(\boldsymbol{\Psi};\ \boldsymbol{\Psi}^{(k)})$, which is defined to be the conditional expectation of the complete-data log-likelihood given the observed data, computed using $\boldsymbol{\Psi}^{(k)}$ for $\boldsymbol{\Psi}$. It effectively requires the calculation of the maximum likelihood estimates of the parameters of the component Kent distributions considered separately. On the $(k+1)$th iteration of the EM algorithm, it follows that the updated estimate $\boldsymbol{\theta}_i^{(k+1)}$ for $\boldsymbol{\theta}_i$ is obtained by solving

$$\sum_{j=1}^{n} \tau_{ij}^{(k)} \partial \log f_K(\mathbf{y}_j;\ \boldsymbol{\theta}_i)/\partial \boldsymbol{\theta}_i = \mathbf{0} \tag{8}$$

for each i $(i = 1, \ldots, g)$. However, this solution does not exist in closed form, and so has to be computed iteratively. A more convenient approach, as outlined by Kent (1982), is to use the moment estimates. Hence, the moment estimates were used here in place of the maximum likelihood estimates (see the Appendix). If $\boldsymbol{\Psi}^{(k+1)}$ denotes the updated estimate of $\boldsymbol{\Psi}$ so obtained on the $(k+1)$th iteration, it does not follow now that the objective function $Q(\boldsymbol{\Psi};\ \boldsymbol{\Psi}^{(k)})$ is globally maximized at $\boldsymbol{\Psi} = \boldsymbol{\Psi}^{(k+1)}$, because the moment rather than the maximum likelihood estimates of the component parameters $\boldsymbol{\theta}_i$ are being used. However, the use of the moment estimates here should make little difference, because Kent (1982) noted that if the eccentricity $2\beta/\kappa$ is small or if κ is large, the moment estimates are close to the maximum likelihood estimates. Indeed, in our work, it was found that the (incomplete-data) likelihood function was not decreased after each such EM iteration.

Of course, we could check whether the inequality

$$Q(\boldsymbol{\Psi}^{(k+1)};\ \boldsymbol{\Psi}^{(k)}) \geq Q(\boldsymbol{\Psi}^{(k)};\ \boldsymbol{\Psi}^{(k)}) \tag{9}$$

holds for each $\boldsymbol{\Psi}^{(k+1)}$. This inequality is sufficient to ensure that the likelihood is not decreased. Choosing $\boldsymbol{\Psi}^{(k+1)}$ so that (9) holds corresponds to using a generalized EM algorithm to compute the maximum likelihood (ML) estimate of $\boldsymbol{\Psi}$.

The algorithm was initialized by fitting a mixture without the noise component and allocating all points beyond a specified threshold from all group means to the noise component and refitting the mixture with a noise component.

To provide starting values for the EM algorithm, a small number of random starts and modified versions of various hierarchical methods and k means using the angle between vectors (see Section 2) as a distance measure were used. This provides an automated approach where the user simply provides the sample and specifies the methods to be utilized to provide starting values.

5. DETERMINING THE NUMBER OF JOINT SETS

An important question that needs to be addressed is how many joint sets are present or, in the mixture model framework, the value of g. This question was previously examined for directional data by Hsu, Walker, and Ogren (1986), who looked at a stepwise method for determining the number of components in a mixture with example applications to joint set data. Their stepwise procedure used a bootstrap method to determine the null distribution of Watson's U^2 statistic. An alternative is to use the likelihood ratio test statistic as described in McLachlan (1987).

With regard to the work presented in this article, it was felt a quick crude approximation would suffice due to the need for reasonably fast analysis. For this reason, the Akaike information criterion (AIC; Akaike 1973) and Bayesian information criterion (BIC; Schwarz 1978) were used; see Kass and Raftery (1995) on the BIC and other approximations to the Bayes factor. The AIC criterion tends to overfit, in general, and so for mixture models, it leads to the fitting of too many components. On the other hand, the BIC criterion tends to underestimate the number of components; see, for example, Cutler and Windham (1993) and Celeux and Soromenho (1996).

6. EXAMPLE

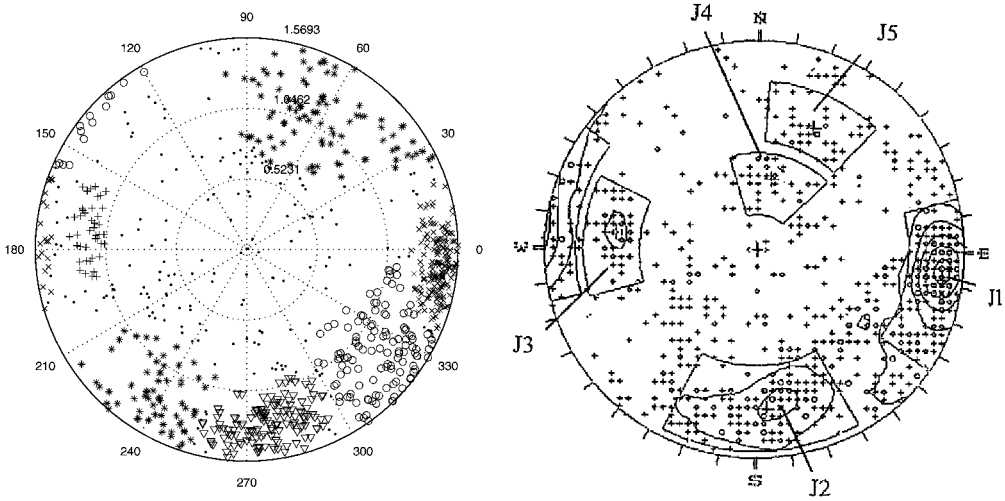
6.1 Introduction

To demonstrate the methods outlined in this paper, two samples, which shall be referred to as Site 1 and Site 2 [supplied by the Julius Kruttschnitt Mineral Research Centre (JKMRC) at the University of Queensland], were analyzed. Site 1 consists of 860 measurements of dip angle and direction and Site 2 consists of 531 measurements. Also available are the user interpretations based on the use of contour plots, as described in Section 1. These user-defined groupings are given in Figures 6 and 7 for Site 1 and Site 2, respectively, with the groupings denoted by rectangular regions bounded by the solid lines and labeled $J1, J2, \ldots, Jg$. The sample points are indicated by + and ∘ symbols (corresponding to the number of repeated values found at the indicated point), and contour graphs are superimposed on the plots.

A mixture of Kent distributions and a uniform distribution was fitted to the data using a modification of the program EMMIX (McLachlan, Peel, Basford, and Adams 1999). In this application, the data are axial (antipodally symmetric). Hence, as pointed out by a referee, it would have sufficed to fit mixtures of Bingham distributions, because the latter apply only for antipodally symmetric data; see Mardia (1972, sec. 8.5). In the fitting of our model with Kent component distributions, the problem of antipodal symmetry is handled by representing each datum point by the pole closest to the center of the cluster in question (using angular distance, as defined in Section 2). In effect, the data space is now restricted to a hemisphere and the uniform component distribution must be suitably changed to $f_0(\mathbf{y}) = 1/(2\pi)$. The use of Kent component distributions modified as in the preceding text should give similar results to Bingham components. The advantage of Kent components is that their use is not restricted to antipodally symmetric data.

As reported in Table 1, the number of groups specified by the user for the two examples is within the range defined by AIC and BIC. To initialize the EM algorithm, 10 random starts, 10 k means, and 7 hierarchical methods were used. The number of components fitted was taken to be same as the user interpretation. The results produced are shown in Figures 6 and 7 for Site 1 and Site 2, respectively. In these and subsequent figures, the points identified as noise are denoted by dots.

First, we compare the result obtained by fitting a mixture model to the user's interpretation for Site 1 (Fig. 6). There is



Mixture model *Human operator*

Figure 6. Polar Plot of the Results for Site 1.

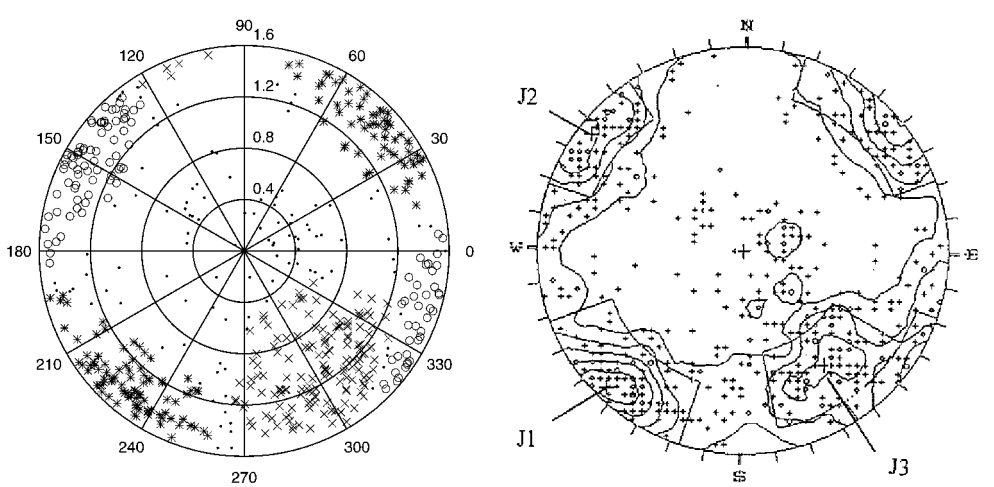
some agreement between the solutions, with the groups designated as J1, J2, and J3 by the user roughly corresponding to the groups denoted by the $\times$, $+$, and ∇ symbols in the mixture model solution. However, the user-defined groups J4 and J5 are modeled by a single group in the mixture model solution (denoted by the symbol $*$) and the extra component (denoted by the symbol $\circ$) used to fit the points to the left of J2.

The results produced when fitting a mixture model for Site 2 were much closer to the user-defined groups (see Fig. 8). The user-defined groups J1 and J2 matched very well to the mixture model solution (denoted by $*$ and $\circ$ symbols, respectively), whereas group J3 was shifted slightly in the mixture model solution (denoted by the symbol $\times$).

Overall, the computed results correspond very nicely to what the human user determined. It should be noted that, as stated in Section 1, the user's solutions are not necessarily the only solutions, but possible solutions. It is, however, encouraging that the two results are similar.

6.2 Simulations

To examine the use of the AIC and BIC criteria to provide a guide to the number of joint sets, a variety of simulated samples were analyzed. Whereas the true number of groups is known, some indication of the performance of AIC and BIC can be gained. Eight simulated samples, with various parameter configurations, were examined. A modification of the program EMMIX (McLachlan et al. 1999) was then used to examine the data fitting models from $g = 0$ to $g = 10$ with the criteria evaluated for each g. Here $g = 0$ corresponds to fitting just the uniform distribution. Each model was fitted using 10 random starts, 10 k-means starts, and 7 hierarchical methods to provide partitions of the sample to initialize the EM algorithm. The results are given in Table 1, which reports the number of groups estimated by AIC and BIC for the various samples. The number of sample and noise points is also given in Table 1. A comparison between the true grouping and the



Mixture model *Human operator*

Figure 7. Polar Plot of the Results for Site 2.

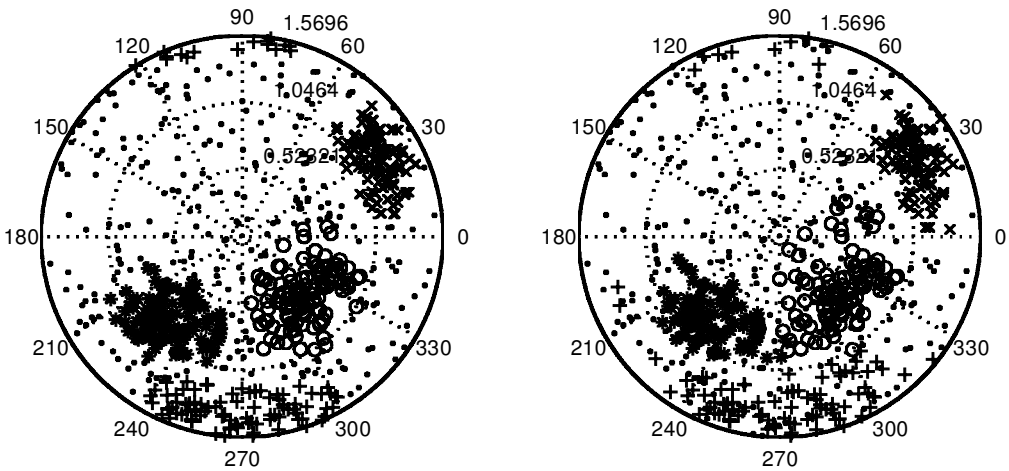


Figure 8. Comparison of True Grouping to That Obtained Fitting a Mixture Model for Simulated Sample 1.

grouping obtained by fitting the mixture model for sample 1 is given in Figure 6.2.

In all cases, the true number of components lies between the estimates provided by AIC and BIC. In this way, AIC and BIC provide a useful interval or range for the number of components. The estimate given by BIC was correct in six simulations out of eight, whereas AIC was correct in three of the simulations.

In the literature, AIC and BIC have been found to be very sensitive to the number of sample points and parameters (see Cutler and Windham 1993), with the methods failing for small sample size relative to the number of parameters. To determine how sensitive the criteria are within this application, a simple experiment was done to investigate the effect of sample size on the accuracy of the criteria. Samples of various sizes n were generated from four well separated groups, each consisting of $n/7$ points. The remaining $3n/7$ points were noise points generated from a uniform distribution. To each of these samples, AIC and BIC were calculated and the results are reported in Table 2. The experiment was repeated for three less separated groups and the results also are reported in Table 2. As expected, the criteria fail when the sample size is small. However, for reasonable sample sizes, the results were excellent, with both criteria repeatedly determining values for g that correspond to the true number of groups.

Table 1. Number of Groups Estimated by AIC and BIC for Various Samples

Sample	$\sum_{i=1}^{g} n_i$	n_0	True g	AIC	BIC
1	400	300	4	4	4
2	200	500	3	3	2
3	0	700	0	1	0
4	300	400	3	4	2
5	600	100	1	2	1
6	200	500	2	2	2
7	500	200	5	6	5
8	600	100	6	7	6
Site 1	860	—	5*	6	2
Site 2	531	—	3*	4	2

*The number of groups estimated by the user from contour plots.

7. DISCUSSION

The work presented in this article gives a useful application of nonnormal mixture models and demonstrates that the mixture model approach offers a practical method of analyzing joint sets in two ways. First, to identify joint sets in noisy data and, second, to determine a range for the number of joint sets suggested by the data. In the case of identifying joint sets, the mixture approach using Kent component distributions (plus a uniformly distributed component) performed very well on actual data. The results approximately matched those independently obtained by an operator using contour plots. An exact match was not expected, because even two experienced operators can produce differing results. However, the results produced by fitting a mixture of Kent and uniform distributions provided a reasonable interpretation of the data. An important feature of the model is that it is robust with respect to noise due to the inclusion of the extra uniform component in the model. With regard to determining the number of joint sets, in the simulation experiment conducted in this article, the AIC and BIC criteria were found to give useful and accurate bounds to the true number of simulated joint sets.

Table 2. Number of Groups Estimated by AIC and BIC for a Range of Sample Sizes n

	True $g = 4$		True $g = 3$	
n	AIC	BIC	AIC	BIC
100	0	0	1	1
200	4	0	3	1
300	4	0	3	1
400	4	0	3	2
500	4	4	3	2
600	4	4	3	3
700	4	4	3	3
800	4	4	4	3
900	4	4	3	3
1,000	4	4	3	3
1,500	4	4	3	3
2,000	4	4	4	3

In summary, our mixture model-based method provides a fully automated method for determining the number of joint sets present and their location and shape, in the form of model parameters, a partition of the sample, and/or posterior probabilities. The manual interpretation of joint sets using the progressive contouring has several disadvantages. First, the resulting interpretation is subjective and often varies according to who made the interpretation. Second, the presentation of the contours on a circle distorts the original hemisphere and may result in a biased interpretation. Third, it is difficult to remove accurately a particular joint set so that the remaining data can be be examined, particularly in the case of overlapping joint sets. Fourth, it is time-consuming and requires an experienced user.

Our mixture model approach is based on a sound statistical model of the joint set distribution that overcomes these problems because it (1) provides a quantitative method that is not dependent on the user's interpretation and (2) places no bias on the interpretation of joints at different angles. Moreover, it is capable of describing elongated clusters that often occur in practice, as well as circular clusters, and of dealing correctly with overlapping clusters in a satisfactory manner. Furthermore it requires only a limited amount of the user's time, this being mainly to read the computer output, and does not require any special experience.

Some work already exists with regard to analyzing heterogeneous directional data. Fisher, Lewis, and Embleton (1987) stated that there are three possible approaches to "multimodal" directional data: partition the data into groups visually using contour plots, use a clustering method, or use a probability model (such as a mixture of Fisher distributions). Fisher et al. (1987) went on to point out with regard to clustering methods for directional data that there is "little currently available in the literature which seems to be of practical use." One example is given of a clustering method based on a nonparametric density approach in Schaeben (1984).

With regard to fitting mixtures of Fisher distributions, Fisher et al. (1987) referred to Stephens (1969), who investigated fitting mixtures of Fisher distributions and found the estimation of the parameters to be tedious. However, it must be stated that Stephen's (1969) comment was made when the available computer power was considerably less than that available today.

Hsu et al. (1986) fitted a mixture of bivariate von Mises distributions; that is, the dip angle and direction were assumed to be distributed independently, each with a von Mises distribution. This assumption does not allow for any correlation between dip direction and dip angle, which can certainly occur. If we visualize the cluster's elliptical contours on the surface of the sphere, then the use of bivariate von Mises distributions constrains the axis of the cluster ellipses to be parallel to longitude and latitude lines of the sphere.

More recently, Hammah and Curran (1998) described the use of a fuzzy k-means algorithm for joint set identification, which utilizes extra fracture information, such as roughness or spacing in the clustering process.

Future work will investigate the use of a bootstrap based procedure as described in McLachlan (1987) to provide a more accurate estimation of the number of joint sets if required. Also, the joint set data as collected also include other information on the joint such as the actual position of the joint and qualitative information on the type of joint. This information could well be included as additional information to assist in the identification of joint sets by clustering.

APPENDIX: MOMENT ESTIMATES OF KENT DISTRIBUTION PARAMETERS

The following steps were proposed in Kent (1982) to estimate the parameters of a single Kent distribution from a sample $(\gamma_{11}, \gamma_{21})^T, \ldots, (\gamma_{1n}, \gamma_{2n})^T$. Let $(y_{11}, y_{21}, y_{31})^T, \ldots, (y_{1n}, y_{2n}, y_{3n})^T$ denote the respective directional cosines. Then the moment estimates are calculated as follows.

Step 1. Calculate the sample mean direction

$$\bar{\gamma}_1 = \sum_{j=1}^{n} \gamma_{1j}/n,\ \bar{\gamma}_2 = \sum_{j=1}^{n} \gamma_{2j}/n,$$

and

$$R^2 = S_{y_1}^2 + S_{y_2}^2 + S_{y_3}^2,$$

where $S_{y_1} = \sum_{i=1}^{n} y_{1i}$, $S_{y_2} = \sum_{i=1}^{n} y_{2i}$, and $S_{y_3} = \sum_{i=1}^{n} y_{3i}$. Next calculate the mean resultant length $\bar{R} = R/n$ and the matrix $\mathbf{S}$, given by

$$\mathbf{S} = \begin{bmatrix} \sum y_{1i}^2 & \sum y_{1i}y_{2i} & \sum y_{1i}y_{3i} \\ \sum y_{1i}y_{2i} & \sum y_{2i}^2 & \sum y_{2i}y_{3i} \\ \sum y_{1i}y_{3i} & \sum y_{2i}y_{3i} & \sum y_{3i}^2 \end{bmatrix}.$$

Step 2. Compute the matrix

$$\mathbf{H} = \begin{bmatrix} \cos\bar{\gamma}_2 & -\sin\bar{\gamma}_2 & 0 \\ \sin\bar{\gamma}_2\cos\bar{\gamma}_1 & \cos\overline{\gamma}_2\cos\bar{\gamma}_1 & -\sin\bar{\gamma}_1 \\ \sin\bar{\gamma}_2\sin\bar{\gamma}_1 & \cos\overline{\gamma}_2\sin\bar{\gamma}_1 & \cos\bar{\gamma}_1, \end{bmatrix},$$

and then compute the matrix $\mathbf{B}$ given by

$$\mathbf{B} = \mathbf{H}^T\mathbf{S}\mathbf{H}.$$

Then $\hat{\alpha}$ is defined by

$$\hat{\alpha} = \tfrac{1}{2}\tan^{-1}\{2b_{23}/(b_{22} - b_{33})\}.$$

Step 3. The matrix $\mathbf{K}$ is computed, where

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\hat{\alpha} & -\sin\hat{\alpha} \\ 0 & \sin\hat{\alpha} & \cos\hat{\alpha} \end{bmatrix}.$$

Put

$$\hat{\mathbf{G}} = \mathbf{H}\mathbf{K} = (\hat{\boldsymbol{\xi}}_1, \hat{\boldsymbol{\xi}}_2, \hat{\boldsymbol{\xi}}_3),$$

where $\hat{\boldsymbol{\xi}}_1, \hat{\boldsymbol{\xi}}_2$, and $\hat{\boldsymbol{\xi}}_3$ are 3×1 column vectors. Then calculate

$$\mathbf{V} = \hat{\mathbf{G}}^T\mathbf{S}\hat{\mathbf{G}}$$

and

$$W = v_{22} - v_{33},$$

where v_{ij} denotes the element of matrix $\mathbf{V}$ in the ith row and jth column.

Step 4. When κ is large, the parameter estimates of $\hat{\kappa}$ and $\hat{\beta}$ are given approximately by

$$\hat{\kappa} = (2 - 2\overline{R} - W)^{-1} + (2 - 2\overline{R} + W)^{-1} \tag{A.1}$$

and

$$\hat{\beta} = \tfrac{1}{2}[(2 - 2\overline{R} - W)^{-1} - (2 - 2\overline{R} + W)^{-1}], \tag{A.2}$$

and the mean direction $(\bar{y}_1, \bar{y}_2, \bar{y}_3)^T$ is given by $\hat{\boldsymbol{\xi}}_1$.

[Received March 1999. Revised October 2000.]

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," *Second International Symposium on Information Theory (Tsahkadsor 1971)*, Budapest: Akadémiai Kiadó, pp. 267–281.

Aler, J., Du Mouza, J., and Arnould, M. (1996), "Measurement of the Fragmentation Efficiency of Rock Mass Blasting and Its Mining Applications," *International Journal of Rock Mechanics and Mining Sciences*, 33, 125–139.

Banfield, J. D., and Raftery, A. E. (1993), "Model-Based Gaussian and Non-Gaussian Clustering," *Biometrics*, 49, 803–821.

Campbell, J. G., Fraley, C., Murtagh, F., and Raftery, A. E. (1997), "Linear Flaw Detection in Woven Textiles Using Model-Based Clustering," *Pattern Recognition Letters*, 18, 1539–1548.

Celeux, G., and Soromenho, G. (1996), "An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model," *Journal of Classification*, 13, 195–212.

Cutler, A., and Windham, M. P. (1993), "Information-Based Validity Functionals for Mixture Analysis," *Proceedings of the First US–Japan Conference on the Frontiers of Statistical Modeling*, ed. H. Bozdogan, Amsterdam: Kluwer, pp. 149–170.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 39, 249–282.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1987), *Statistical Analysis of Spherical Data*, Cambridge, U.K.: Cambridge University Press.

Fraley, C., and Raftery, A. E. (1998), "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis," *Computer Journal*, 41, 578–588.

Hammah, R. E., and Curran, J. H. (1998), "Fuzzy Cluster Algorithm for the Automatic Identification of Joint Sets," *International Journal of Rock Mechanics and Mining Sciences*, 35, 889–905.

Hsu, Y. S., Walker, J. J., and Ogren, D. E. (1986), "A Stepwise Method for Determining the Number of Component Distributions in a Mixture," *Mathematical Geology*, 18, 153–160.

Kass, R. and Raftery, A. E. (1995), "Bayes factors," *Journal of the American Statistical Association*, 90, 773–795.

Kent, J. T. (1982), "The Fisher–Bingham Distribution on the Sphere," *Journal of the Royal Statistical Society*, Ser. B, 44, 71–80.

Mardia, K. V. (1972), *Statistics of Directional Data*, London: Academic Press.

McLachlan, G. J. (1987), "On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture," *Journal of the Royal Statistical Society*, Ser. C, 36, 318–324.

McLachlan, G. J., and Basford, K. E. (1988), *Mixture Models: Inference and Applications to Clustering*, New York: Dekker.

McLachlan, G. J. and Krishnan, T. (1997), *The EM Algorithm and Extensions*, New York: Wiley.

McLachlan, G. J., Peel, D., Basford, K. E., and Adams, P. (1998), "The EMMIX Algorithm for the Fitting of Mixtures of Normal and *t*-Components," *Journal of Statistical Software* [on-line], 4 Available at (http://www.stat.ucla.edu/journals/jss/).

Palmstrom, A. (1985), "Application of the Volumetric Joint Count as a Measure of Rock Mass Jointing," *Proceedings of International Symposium on Fundamentals of Rock Joints*, Bjorkliden: Centek Publishers, pp. 103–110.

Pal Roy, P. (1995), "Breakage Assessment Through Cluster Analysis of Joint Set Orientations of Exposed Benches of Opencast Mines," *Geotechnical and Geological Engineering*, 13, 79–92.

Priest, S. D. (1993), *Discontinuity Analysis for Rock Engineering*, London: Chapman & Hall.

Schaeben, H. (1984), "A New Cluster Algorithm for Orientation Data," *Mathematical Geology*, 16, 139–154.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 78, 461–464.

Stephens, M. A. (1969), "Techniques for Directional Data," Technical Report 150, Stanford University, Dept. of Statistics.

Sullivan, T., Duran, A., and Eggers, J. (1992), "The Use and Abuse of Oriented Core in Open Pit Mine Design," *Proceedings of the Third Large Open Pit Conference*, Mackay, Australia, Melbourne: MM, pp. 387–396.

Webber, I. P., and Gowans, D. (1996), "A New Core Orientation Device," *Advances in Site Investigation Practice*, London: Thomas Telford, pp. 301–311.