

Genetic algorithm for fitting a mixed Bingham distribution to 3D orientations: A tool for the statistical and paleostress analyses of fracture orientations

ATSUSHI YAMAJI

Division of Earth and Planetary Sciences, Kyoto University, Kyoto 606-8502, Japan

E-mail: yamaji@kueps.kyoto-u.ac.jp

Abstract The clustering of fracture orientations is important for tectonic studies and for geotechnical engineering. In this study, real-coded genetic algorithm was adapted to fitting a mixed Bingham distribution to orientation data by maximizing the log-likelihood function of the distribution. The maximization is a difficult problem, because the function has multimodality and singularity. It was found that the algorithm was effective for this problem. Given the orientations of dilational fractures, the present method determines not only the stress axes and stress ratio of each of fracture groups but also the maximum non-dimensionalized fluid pressure at the time of their formation. In addition, the software calculates the 95% error ellipses of the concentration axes. The present method found that the orientations of ore veins of the Akenobe Mine, SW Japan, should be partitioned into three clusters. It is shown that two of the groups had distinctive Zn and Sn contents, and that the ore fluids had overpressures only slightly greater than the minimum principal stress at the time of the deposition of Zn and Sn rich veins.

Key words: tectonic stress, fluid pressure, confidence ellipse, Bingham distribution, statistical mixture model, vein-type ore, dike, dilational fracture

INTRODUCTION

Fracture orientations are important to predict the leakage of geofluid from reservoirs (e.g., Bigi *et al.* 2013) and to estimate the stability of open pit and tunnel walls (Priest 1993; Peel *et al.* 2001). The orientations of dilational fractures such as dikes and mineral veins are clues to paleostresses (Stevens 1911; Anderson 1942; McHone 1978; Delaney *et al.* 1986; Baer *et al.* 1994; Jolly & Sanderson 1997; Yamaji & Sato 2010;

Sato *et al.* 2013; Kanai *et al.* 2014). Since the poles to natural fractures often show multimodal distribution, the clustering of the fracture orientations are essential for the description and paleostress analysis of the fractures. Yamaji & Sato (2011) used mixed Bingham distribution for this purpose, and succeeded in detecting two paleostresses from the orientations of Miocene igneous dikes in SW Japan, the classification of which had not been recognized. The Bingham distribution is so flexible that it can describe circular, elliptical and girdle distributions (Fig. 1), and has been used for delineating the confidence regions of paleomagnetic directions (e.g. Onstott 1980; Evans *et al.* 2000; Love 2007) and structural orientations (e.g. Kelker & Langenberg 1982; Siemes *et al.* 2000; Johansen *et al.* 2004; Kunze & Schaeben, 2004; Ueda *et al.* 2012; Sáez *et al.* 2013; Riffert *et al.* 2014).

To fit a mixed Bingham distribution to a set of orientation data, a few researchers (Yamaji & Sato 2011; Niezgodá & Glover 2013) searched for the maximum point of a multimodal function, $L(\mathbf{X})$, by means of the expectation-maximization (EM) algorithm (e.g. Bishop 2006). That is, the maximum point of the function represents the best partition of the data. However, it is a difficult problem to discover the global maximum of such a multimodal function (Fig. 2) especially for the EM algorithm, which starts from a randomly chosen point and simply climbs the slope of the function from the starting point. The algorithm often arrives and stops not at the global but at a local maximum. Consequently, the algorithm should be started hundreds of times with random initial conditions to find the global maximum. The animation in the supplementary files, EM1.mov and EM2.mov, of this paper show the cases where the EM algorithm succeeded and failed, respectively, in detecting a good partition of the natural fracture data of Shanley & Mahtab (1976) into three clusters. In the latter case, the singularity of the function (Bishop 2006, p. 434) made red and blue clusters to shrink and most of the data to be classified into a green cluster. In addition, if a data set is partitioned into K clusters, the number of local maxima of the function increases with the increase of K . To make matters worse, the parameter space in which the global maximum is searched for is a $5K$ -dimensional space, because a Bingham distribution has five degrees of freedom (Bingham 1974; Borradaile 2003). For example, if a data set is partitioned into three groups, the global maximum must be discovered in

15-dimensional space. The high dimensionality makes it difficult to solve the search problem.

In this study, the genetic algorithm (GA) was found to be effective to search for the best partition of orientation data. The algorithm is one of the numerical optimization techniques that can escape from local maxima and detect the global maximum (e.g. Goldberg 1989). In the following sections, we introduce, first, the mathematical formulations of mixed Bingham distributions. Second, we document the algorithm. The method was applied to two natural data sets: one from the San Manuel copper mine, Arizona, and the other from the vein-type ores of the Akenobe mine, SW Japan.

The program was written in MATLAB version 7, and is available freely from the author.

METHOD

MIXED BINGHAM DISTRIBUTION

A mixed Bingham distribution is a linear combination of Bingham distributions with different concentration axes and concentration parameters. In this section, we briefly explain Bingham and mixed Bingham distributions, the parameter space of which is detailed in Appendix 1.

The Bingham distribution is so flexible a probability distribution that it can represent the uniform, circular, elliptical and girdle distributions of 3D orientations (Fig. 1). The distribution has the probability density function,

$$P_B(\mathbf{v}|\boldsymbol{\kappa}, \mathbf{E}) = \frac{1}{A} \exp(\mathbf{v}^T \mathbf{E}^T \boldsymbol{\kappa} \mathbf{E} \mathbf{v}),$$

where \mathbf{v} is the three-dimensional unit column vector indicating an orientation, A the normalizing factor, which is a function of the concentration parameters, κ_1 and κ_2 , \mathbf{E} the 3×3 orthogonal matrix denoting the orientations of maximum, intermediate and minimum concentration axes of the distribution, and $\boldsymbol{\kappa}$ the diagonal matrix, $\text{diag}(\kappa_1, \kappa_2, 0)$. When we deal with fractures, \mathbf{v} is the unit vector normal to a fracture plane. By definition the concentration parameters satisfy

$$\kappa_1 \leq \kappa_2 \leq 0. \quad (1)$$

Uniform and circular distributions are denoted by $\kappa_1 = \kappa_2 = 0$ and $\kappa_1 = \kappa_2 < 0$,

respectively; and smaller parameter values indicate a concentrated distribution. Elliptical and girdle distributions are denoted by $\kappa_1 \approx \kappa_2$ and $\kappa_1 \ll \kappa_2 \approx 0$, respectively (Fig. 1). The parameters characterizing a Bingham distribution are represented by a five-dimensional position vector, \mathbf{x} , the length of which is defined as

$$\rho \equiv |\mathbf{x}| = \sqrt{\kappa_1^2 + \kappa_1\kappa_2 + \kappa_2^2}. \quad (2)$$

For the paleostress analysis, the ratio, κ_2/κ_1 , is interpreted to be equal with the stress ratio, $\Phi = (\sigma_2 - \sigma_3)/(\sigma_1 - \sigma_3)$ (Yamaji *et al.* 2010), where σ_i is the i th principal stress of the stress condition that corresponds to the Bingham distribution.

A mixed Bingham distribution with K Bingham components has the probability density function (Yamaji & Sato 2011),

$$P_{\text{mB}}(\mathbf{v}|\mathbf{X}) = \sum_{k=1}^K \varpi^k P_{\text{B}}(\mathbf{v}|\mathbf{x}^k),$$

where \mathbf{x}^k is the five-dimensional position vector representing the k th Bingham distribution, ϖ^k the mixing coefficient (Bishop 2006) of the distribution, and the $5K$ -dimensional vector,

$$\mathbf{X} = (\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^K), \quad (3)$$

characterizes the mixed Bingham distribution. Each of the Bingham components represents a girdle, elliptical or circular cluster. The mixing coefficients, $\varpi^1, \dots, \varpi^K$, satisfy $0 \leq \varpi^k \leq 1$ and $\varpi^1 + \dots + \varpi^K = 1$. That is, ϖ^k stands for the significance of the k th Bingham distribution in the mixture. Figure 3 shows an example of mixed Bingham distribution and its Bingham components.

If a data set is composed of the N orientations, $\mathbf{v}_1, \dots, \mathbf{v}_N$, the quantity,

$$m_n^k = \frac{P_{\text{B}}(\mathbf{v}_n|\mathbf{x}^k)}{\sum_{k=1}^K P_{\text{B}}(\mathbf{v}_n|\mathbf{x}^k)}, \quad (4)$$

has a value between 0 and 1 denoting the membership of the n th datum to the k th group ($k = 1, \dots, K$). For example, the n th datum belongs to the k th cluster with a probability of 20% in case of $m_n^k = 0.2$.

FITTING

A mixed Bingham distribution is fitted to the orientations, $\mathbf{v}_1, \dots, \mathbf{v}_N$, by maximizing

the logarithmic likelihood function,

$$L(\mathbf{X}) = \sum_{n=1}^N \log_e P_{\text{mB}}(\mathbf{v}_n | \mathbf{X}). \quad (5)$$

Given a set of fracture orientations, the unit vector, \mathbf{v}_n , denotes the pole to the n th fracture surface. The distribution that best fits the N data is obtained by searching for the point, \mathbf{X}^{opt} , where this function has the maximum value. This function has numerous local maxima (Fig. 2), making the optimization of \mathbf{X} very difficult. In this work, the genetic algorithm, which is detailed in the next section, is used for this purpose.

Taking into account the measurement errors of orientations at outcrops, the optimal \mathbf{X} was searched for the global maximum of $L(\mathbf{X})$ in the region,

$$|\mathbf{x}^k| \leq 200 \quad (k = 1, \dots, K), \quad (6)$$

where \mathbf{x}^k is the k th component of \mathbf{X} (Eq. 3). That is, the measurement error is a few degrees, so a circular cluster with the radius of 4° is denoted by a Bingham distribution with $\kappa_1 = \kappa_2 \approx -115$ (Fig. 1); and we ignore smaller clusters. It follows from Eq. (2) and $\kappa_1 = \kappa_2 = -115$ that $\rho \approx 200$. Therefore, we deal with \mathbf{x}^k that satisfies the inequality (6).

Once $L(\mathbf{X})$ is maximized, the number of clusters, K , is evaluated using Bayesian information criterion, BIC (e.g., Bishop 2006). That is, the optimal K value is given by minimizing

$$\text{BIC} = -2L(\mathbf{X}^{\text{opt}}) + (6K - 1)\log_e N, \quad (7)$$

where \mathbf{X}^{opt} is a function of K , and represents the mixed Bingham distribution that best fits the data. To determine the appropriate K value for a given data set, the log-likelihood function, $L(\mathbf{X}^{\text{opt}})$, must be calculated for the cases, $K = 1, 2, 3, \dots$

ERROR ANALYSIS

Once the optimal partition is determined, we calculate the confidence ellipses of the concentration axes from the orientation matrix of the k th Bingham component,

$$\mathbf{T}^k = \sum_{n=1}^N m_n^k \mathbf{v}_n^T \mathbf{v}_n. \quad (8)$$

Since this is a symmetric matrix, \mathbf{T}^k has real eigenvalues, for which we assume the magnitude relation, $\tau_1^k \leq \tau_2^k \leq \tau_3^k$. Let κ_i^k be the i th concentration parameter of the

k th cluster, and let the third one be zero, i.e., $\kappa_1^k \leq \kappa_2^k \leq \kappa_3^k = 0$. Then, the 95% confidence ellipse of the k th Bingham component are given by

$$\alpha_{ij}^k = \frac{2.45}{2|(\kappa_i^k - \kappa_j^k)(\tau_i^k - \tau_j^k)|},$$

where α_{ij}^k is the semi-axis of the i th error ellipse towards the j th concentration axis of the k th cluster (Tanaka 1999).

REAL CODED GENETIC ALGORITHM

A real-coded genetic algorithm (e.g. Davis 1991), GA, was employed for the optimization of Eq. (5). The genetic algorithm is a search method based on the mechanics of biological evolution (Goldberg 1989) as follows. Each individual of a population has a fitness value: The individuals with low fitness die out, and others with high fitness have high probability to be mated with to have offspring(s) that may have better fitness. However, these processes have exceptions, because the survival of the fittest without other principles leads to the blind alley of the evolution. Those with low fitness can survive at low probabilities, and others with high fitness occasionally die out. In addition, mutation introduces random modification in the population that allows the population to escape from blind alleys. As a result of these processes, the best individuals of the population gradually improve their fitness values.

Now, an individual and its fitness are compared to a mixed Bingham distribution, \mathbf{X} , and its logarithmic likelihood, $L(\mathbf{X})$, for a given set of 3D orientations (Fig. 2). The flow chart of the program is shown in Fig. 4. The important procedures of the algorithm are as follows. We regard $5K$ -dimensional vector, \mathbf{X} , as an individual and its gene simultaneously; and the set of N_p individuals, $\{\mathbf{X}^1, \dots, \mathbf{X}^{N_p}\}$, as the population that evolves to determine the global maximum of $L(\mathbf{X})$. $L(\mathbf{X}^i)$ is termed the fitness of the i th individual, and determines the chance for the individual to survive and to have children. The population is improved iteratively with keeping the size of the population, N_p . A small population is favorable to speed-up the evolution, but GA with a small N_p value is liable to lead to a local maximum of $L(\mathbf{X})$, which is compared to a blind alley of the evolution. A few N_p values are used to determine the global maximum of $L(\mathbf{X})$ in our program. The GA with $N_p = 20$ in our software usually worked well.

INITIALIZATION

The population is initialized as follows. An individual represents K Bingham components, each of which is denoted by a point in the five-dimensional parameter space (Appendix 1). Each Bingham component is initialized by drawing a point, \mathbf{x} , randomly from a five-dimensional Gaussian distribution with the mean at the origin and the covariance matrix, $\text{diag}(1,1,1,1,1)$. Then, $|\mathbf{x}|$ which is identical with ρ (Fig. 1) obeys five-dimensional Rayleigh distribution (Ali & Woo 2005) with the mean at ~ 2.1 . That is, clusters with the radii as large as a few tens of degrees are chosen as the initial configuration of the GA.

The initial population is created by gathering the Bingham components generated in this way. We regard them as the 0th generation ($g = 0$). A negative value with a very large absolute value is assigned to L^{\max} as a tentative value.

In case of $K = 1$, the concentration axes and the concentration parameters were inferred through the method of Mardia & Zemroch (1977), and were used to form an individual. That is, the eigenvectors of the orientation matrix (Eq. 8) were identified with the concentration orientations and the concentration parameters were determined from its eigenvalues, where the value 1 was assigned to the memberships, m_1^1, \dots, m_N^1 .

EVALUATION

The fitness of an individual \mathbf{X} is evaluated as follows. First, \mathbf{X} is decomposed into K Bingham components, $\mathbf{x}^1, \dots, \mathbf{x}^K$ (Eq. 3). Second, the memberships of all the data are calculated by Eq. (4). Third, the mixing coefficients are obtained as $\varpi^k = m_1^k + \dots + m_N^k$ ($k = 1, \dots, K$). Finally, the log-likelihood function in Eq. (5), i.e., $L(\mathbf{X}^k)$, is assigned to the fitness of the individual.

CROSSOVER

An important step of the genetic algorithm is that individuals with high fitness values give birth to a child that resembles its parents. To choose a pair, we employ the rank selection (Baker 1985), meaning that the parents are randomly chosen from individuals. Those with high fitness values are chosen more often than those with low fitness values.

Then, two children are obtained by exchanging the ‘genes’ of the pairs through the unimodal normal distribution crossover technique (Kita *et al.* 1998; Ono *et al.* 2003). Figure 2 illustrates the basic idea of the technique, whereby individuals that have high fitness values tend to have children. Since a local or the global maximum can be expected to exist near the point with the high fitness values, this process tends to increase the best fitness of the population.

SELECTION

Rank selection (Baker 1985) is used to remove individuals from the population. That is, individuals with low fitness values survive at low probabilities. However, those with high fitness values are removed with non-zero but small probabilities. The denial of the survival of the fittest is very important to allow the population to get out of local maxima (Fig. 5a). When the highest fitness individual is removed, the population shows a sudden decline in the best fitness of the population.

MUTATION

An individual is replaced with randomly generated one. That is, the point representing the new individual, \mathbf{X} , is drawn from the multivariate normal distribution with the covariance matrix, $30^2 \text{diag}(1,1,1,1,1)$ or $70^2 \text{diag}(1,1,1,1,1)$, either of which is randomly chosen. The replaced individual is chosen by the rank selection. In case the length of \mathbf{x} , the component of \mathbf{X} , is larger than 200 (Eq. 7), \mathbf{x} is replaced with $200\mathbf{x}/|\mathbf{x}|$.

RANDOM WORK

In order to escape local maxima and for individuals to visit various points in the parameter space, a perturbation is applied every generation. That is, \mathbf{X} is replaced with $\mathbf{X} + 0.05\Delta\mathbf{X}$, where $\Delta\mathbf{X}$ is a random vector obeying the $5K$ -dimensional Gaussian distribution, the covariance of which is the identity matrix.

RECORDING

At the end of the main loop, the maximum fitness of the individual in the population is recorded along with the identification number and the point, \mathbf{X}^{opt} , of the best individual.

In this article, we refer to the maximum fitness at the generation g as ${}^g L^{\max}$. This quantity has a tendency to increase mainly by the crossover, but sometimes shows a sudden decline by the selection.

TERMINATION CONDITION OF GA

Let L^{\max} be the maximum among ${}^1 L^{\max}, \dots, {}^g L^{\max}$. The main loop is terminated by reaching the convergence of L^{\max} or by reaching a limit in g . Let g' be the generation when L^{\max} was last updated. The genetic algorithm is terminated when the condition,

$$g = g' + K\Delta g, \quad (9)$$

is met (Fig. 5b), where Δg is 100 or 1000 depending on the difficulty of the data set to be partitioned into K clusters. Time of computation required for the convergence generally increases with increasing K . The second term of the right-hand side of Eq. (9) deals with this effect. On the other hand, the computation is programmed to stop eventually at the 100,000th generation.

FINAL OPTIMIZATION BY THE EM ALGORITHM

The optimization by the genetic algorithm slowed down as X approached the global maximum of $L(\mathbf{X})$. Accordingly, the EM algorithm was employed to refine the partition (Fig. 4b). The EM algorithm is detailed by Yamaji & Sato (2011). The algorithm was terminated when the L values of succeeding iterations, ${}^g L$ and ${}^{g+1} L$, satisfied the condition, $|({}^{g+1} L - {}^g L) / {}^g L| < 10^{-6}$.

ITERATION OF THE ENTIRE ROUTINE

The optimal K value for a data set is evaluated from the BIC versus K plot. For this purpose, values of the log-likelihood function must be accurately determined for $K = 1, 2, 3, \dots$ (Eq. 7). This is a challenging task owing to the multimodality of the function. Fuzzy clustering, including the present case, is not easy due to the multimodality of its object function (e.g., Bishop 2006, p. 434–35). In order to escape local maxima and to search for the global maximum of the function, such clustering techniques run a computer program many times from different initial conditions. Elaborate clustering schemes such as Bayesian inference do so as well (Bishop 2006, p. 484). Our program is run for a data set several times with each K value starting from different initial

populations (Eq. 5).

How many times do we have to run the program? A heuristic approaches solves this problem, the key to which is the monotonous increase of $L(\mathbf{X}^{\text{opt}})$ as a function of K . That is, the graph, L^{max} versus K , should have a convex-upward, smooth curve (Fig. 6a). If the graph drawn with calculated values has a depression at a K value from such a smooth curve, L^{max} at the depression is inaccurate. Empirically, the present method requires the computation several times for each K value.

NUMERICAL EXAMPLES

FRACTURES

The present technique was tested with fracture orientations in the San Manuel Copper Mine, Arizona (Shanley & Mahtab 1976). The data set has been used for the benchmark tests of the clustering techniques of orientation data. An optimization process of $L(\mathbf{X})$ for the case of $K = 3$ is shown in Fig. 5. The supplementary file, GA.mov, shows the animation of this process, the snapshots of which are shown in Fig. 5c. The optimization of $L(\mathbf{X})$ was accomplished by repeating the entire routine in Fig. 4 several times for each of the K values, 1, 2, 3, etc.

The present technique was much more effective than the EM algorithm (Yamaji & Sato 2011) for the optimization, as the EM algorithm so often reached local maxima of $L(\mathbf{X})$ that the algorithm should have been run hundreds of times to detect the global maximum. The supplementary files, EM1.mov and EM2.mov, show the animation of the successful and unsuccessful optimization processes of the EM algorithm for partitioning 3D orientations into three clusters. The animation of the latter file demonstrates how the singularity of the function, $L(\mathbf{X})$, spoils the optimization: The singularity leads a cluster to collapse to a single data point (Bishop 2006, p. 434). Such failure was not rare for the EM algorithm, but was suppressed in the present genetic algorithm by the condition in Eq. (6).

ORE VEINS

The present technique was applied, next, to the orientation data from ore veins in the

Akenobe Mine, SW Japan (Sato *et al.* 1977; Sato & Akiyama 1980). The veins were formed in the Cretaceous–Paleocene (Ishihara & Shibata 1972; Watanabe *et al.* 1984). It is beyond the scope of this paper to discuss the tectonic and metallogenic implications of the orientations. Figures 5a and 5b show the final L values and the corresponding BIC values. BIC showed the minimum at $K = 3$, meaning that the data should be partitioned into three groups. The lower-hemisphere, equal-area projection in Fig. 6c and Table 1 show the data and their best partition. The minimum, intermediate and maximum concentration axes of each Bingham component correspond to the σ_1 -, σ_2 - and σ_3 -axes, respectively (Baer *et al.* 1994; Yamaji *et al.* 2010). These correspondences are based on the assumption that pre-existing planes of weakness to be dilated by overpressured fluid were randomly oriented. As a result, the three clusters, A, B and C, were identified from the data set. There was no data point with a neutral tint except for the orange one in the SW quadrant in Fig. 6c, meaning that the data were clearly separated into the three clusters.

The Bingham distribution of the poles to dilational fractures corresponds to the state of stress at the time of their formation, provided that the rock mass including the fractures were subjected to homogeneous stress field and to irrotational deformation by the formation. In case of conjugate vein arrays, the rocks between the veins are rotated (Beach 1975) to disturb the stress field around the veins. The spatial distribution of the NW-SE ore veins of the clusters B and C (Sato *et al.* 1977; Sato & Akiyama 1980) evidences that they were not formed as conjugate arrays. That is, the ore veins in the Akenobe area were deposited probably under three different stress conditions. Since a Paleozoic–Mesozoic accretion complex hosts the ore veins, it is difficult to correct the tilting that the host rock has experiences after the ore mineralization.

The cluster A made a girdle, so the error ellipses of its maximum and intermediate concentration axes were elongated along the girdle and were larger than that of the minimum concentration axis (Fig. 6c). Fifty-five percent of the data belonged to this cluster, i.e., $\varpi^A = 0.55$. The clusters B and C were composed of the green and blue data points, respectively. The veins belonging to the clusters had NW–SE trends, but those of the clusters B and C had northeastward and southwestward dips, respectively. The error ellipses of their maximum concentration axes were small compared to the intermediate

and minimum concentration axes. NW-SE trending ore veins in the Akenobe mine were recognized to form a group, but the present method separated them into two groups.

Since the concentration axes corresponds to the principal stress axes (Baer *et al.* 1994) and the stress ratio can be obtained from the estimated concentration parameters as $\Phi = \kappa_2/\kappa_1$ (Yamaji *et al.* 2010). The reduced stress tensors, \mathbf{S}^A , \mathbf{S}^B and \mathbf{S}^C , were calculated from the obtained Bingham components through the equation,

$$\mathbf{S}^k = (\mathbf{E}^k)^T \text{diag}(1, \Phi^k, 0) \mathbf{E}^k,$$

where k is the label to distinguish the three stresses, \mathbf{E}^k is the orthogonal matrix representing the principal axes of the k th cluster, and Φ^k is the stress ratio of the k th stress. Let \mathbf{v}_n be the unit vector normal to the n th vein wall, and let k stand for A, B or C. Figures 5d–f show the Mohr diagrams of the stresses. Let σ_n^k and τ_n^k be the normal and shear stresses on the n th vein due to the k th stress. They are calculated as $\sigma_n^k = \mathbf{v}_n^T \mathbf{S}^k \mathbf{v}_n$ and $\tau_n^k = |\mathbf{S}^k \mathbf{v}_n - \sigma_n^k \mathbf{v}_n|$, and are plotted with the membership, m_n^k , on the Mohr diagrams. The distribution of the normal stresses, $\sigma_1^k, \dots, \sigma_N^k$, indicates the lower bound of the fluid pressures that dilated the fractures when the ore veins were formed under the k th stress. That is, the diagrams just under the Mohr diagrams show the driving pressure $p = (p_f - \sigma_3)/(\sigma_1 - \sigma_3)$, at the time of the vein formation, where p_f is the pressure of ore fluid (Baer *et al.* 1994). For example, the memberships of the cluster A, z^A , are high only for the veins with driving pressures as low as ~ 0.2 (Fig. 5d), so the veins consisting of the cluster A were formed by the fluid with p at or smaller than ~ 0.2 . In contrast, the veins belonging to the clusters B and C exhibited very low driving pressures ($p < 0.1$) (Figs. 5e, f). The Zn- and Sn-rich veins were formed under low fluid pressure relative to the σ_1 and σ_3 values at the time of their formation.

Mining geologists thought that Zn-rich veins were deposited before Sn-rich ones in the area (Sato *et al.* 1977; Sato & Akiyama 1980). Figure 7 shows the Zn, Sn and Cu contents of the veins versus the memberships of the veins to the clusters A, B and C. Though there were a few exceptions, the veins belonging to the clusters B and C had low and high Zn contents, respectively. The veins of cluster C had relatively low Sn contents. Therefore, the stress indicated by the cluster C is probably older than the stress corresponding to the cluster B. Since the veins belonging to the cluster A had low Zn and Sn contents (though there are a few exceptions), it is difficult to infer the timing of

their formation based on this study. The Cu contents showed no correlation with the clusters.

DISCUSSION AND SUMMARY

The clustering of fracture orientations is important in geotechnical engineering and mining for safety and economic reasons (e.g., Priest 1993), where the discrimination of two clusters has little importance if they have common maximum concentration axes but different minimum concentration axes. However, it is essential for understanding paleostresses, as they indicate different σ_1 - and σ_3 -orientations. The data set in Fig. 6c is an example: The clusters B and C are more or less included in the girdle cluster A. Even if they have the maximum concentration axes in common, they are identified by the technique, provided that the minimum concentration axes make a large enough angle (Yamaji *et al.* 2010).

Peel *et al.* (2001) fitted mixed Kent distribution for the clustering the orientations of rock joints, and used BIC to estimate the number of clusters. The distribution was applied to remote sensing data by Lunga & Ersoy, (2011). The Kent distribution represents a circular cluster, but the paleostress analysis of dilational fractures requires not only circular but also elliptical and girdle clusters to determine the principal axes and Φ values. A Bingham or more complicated statistical distribution is necessary.

The maximization of the log-likelihood function has generally such singularity that a cluster tends to collapse onto a data point. That is, the probability density of the Bingham component goes to infinity at the data point, whereas the κ_1 and κ_2 values of the component go to $-\infty$. Bishop (2006, p. 434) details such singularity for the case of a mixed normal distribution. The singularity can spoil the result of the EM algorithm (see the animation in the supplementary file, EM2.mov), because there is no way for the algorithm to escape from the local maximum caused by the singularity. In contrast, the genetic algorithm is robust to the singularity, because it can escape from such a local maximum. The ore fluids had overpressures only slightly greater than the minimum principal stress (Figs. 6d–f) at the time of the deposition of Zn and Sn rich veins. The present method may be useful for the characterization of crystallographic preferred

orientations (Niezgoda & Glover 2013) and of multi-component paleomagnetic data.

ACKNOWLEDGMENTS

The author thanks Prof. T. Blenkinsop, Dr. F. Mazzarini and an anonymous reviewer for their constrictive comments on an early draft of this manuscript. This work was financially supported by JSPS (15H02141).

REFERENCES

- ALI, M. M. & WOO, J. 2005. Inference on reliability $P(Y < X)$ in a p -dimensional Rayleigh distribution. *Mathematical and Computer Modelling* **42**, 367–73.
- Anderson, E.M. 1942. *The Dynamics of Faulting and Dike Formation with Applications to Britain*. Oliver & Boyd, Edinburgh.
- Baer, G., Beyth, M. & Reches, Z. 1994. Dikes emplaced into fractured basement, Timna Igneous complex, Israel. *Journal of Geophysical Research* **99**, 24039–51.
- Baker, J. E. 1985. Adaptive selection methods for genetic algorithms. *Proceedings of the First International Conference on Genetic Algorithms*, 101–11.
- Beach, A. 1975. The geometry of en-echelon vein arrays. *Tectonophysics* **28**, 245–63.
- Bigi, S., Battaglia, M., Alemanni, A., Lombardi, S. & Campana, A. 2013. CO₂ flow through a fractured rock volume: Insights from field data, 3D fractures representation and fluid flow modeling. *International Journal of Greenhouse Gas Control* **18**, 183–99.
- Bingham, C. 1974. An antipodally symmetric distribution on the sphere. *Annals of Statistics* **2**, 1201–25.
- Bishop, C., 2006. *Pattern Recognition and Machine Learning*. Springer, N.Y.
- Borradaile, G. J., 2003. *Statistics of Earth Science Data: Their Distribution in Time, Space and Orientation*. Springer, Berlin.
- Davis, L. (Ed.), 1991. *Handbook of Genetic Algorithms*. Van Nostrand Reinbold, N.Y.
- Delaney, P. T., Pollard, D. D., Zioney, J. I. & McKee, E. H. 1986. Field relations between dikes and joints: Emplacement processes and paleostress analysis. *Journal of Geophysical Research* **91**, 4920–83.
- Evans, D. A. D., Li, Z. X., Kirschvink, J. L. & Wingate, M. T. D. 2000. A high-quality mid-Neoproterozoic paleomagnetic pole from South China, with implications for ice ages and the breakup configuration of Rodinia. *Precambrian Research* **100**, 313–34.
- Goldberg, D. E., 1989. *Genetic Algorithms in Search, Optimization, and Machine*

- Learning*. Addison-Wesley, Reading.
- Ishihara, S. & Shibata, K., 1972. Re-examination of the metallogenic epoch of the Ikuno-Akenobe province in Japan. *Mining Geology* **22**, 67–73.
- Johansen, A., Ruud, B. O. & Jakobsen, M. 2004. Effect of grain scale alignment on seismic anisotropy and reflectivity of shales. *Geophysical Prospecting* **52**, 133–49.
- Jolly, R. J. H. & Sanderson, D. J. 1997. A Mohr circle construction for the opening of a pre-existing fracture. *Journal of Structural Geology* **19**, 887–92.
- Kanai, T., Yamaji, A. & Takagi, H. 2014. Paleostress analyses by means of mixed Bingham distributions of healed microcracks in the Ryoke granites, Central Japan. *Journal of Geological Society of Japan* **120**, 23–35.
- Kelker, D. & Langenberg, C. W. 1982. A mathematical model for orientation data from macroscopic conical folds. *Mathematical Geology* **14**, 289–307.
- Kita, H., Ono, I. & Kobayashi, S. 1998. Theoretical analysis of the unimodal normal distribution crossover for real-coded genetic algorithms. *Proceedings of the 1998 IEEE International Conference on Evolutionary Computation, IEEE Congress on Evolutionary Computation*, 529–34.
- Kunze, K. & Schaeben, H. 2004. The Bingham distribution of quaternions and its spherical radon transform in texture analysis. *Mathematical Geology* **36**, 917–43.
- Lunga, D. & Ersoy, O. 2011. Kent mixture model for classification of remote sensing data on spherical manifolds. *Proceedings of Applied Imagery Pattern Recognition Workshop*, 6176337.
- Mardia, K. V. & Zemroch, P. J. 1977. Table of maximum likelihood estimates for the Bingham distribution. *Journal of Statistical Computation and Simulation* **6**, 29–34.
- McHone, J. G. 1978. Distribution, orientations, and ages of mafic dikes in central New England. *Geological Society of America Bulletin* **89**, 1645–55.
- Niezgoda, S. R. & Glover, J., 2013. Unsupervised learning for efficient texture estimation from limited discrete orientation data. *Metallurgical and Materials Transactions* **44A**, 4891–4905.
- Ono, I., Kita, H. & Kobayashi, S. 2003. A real-coded genetic algorithm using the Unimodal Normal Distribution Crossover, In: Ghosh, A., Tsutsumi, S. (Eds.) *Advances in Evolutionary Computing*, Springer, Berlin, pp. 213–37.
- Oncott, T. C., 1980. Application of the Bingham distribution function in paleomagnetic studies. *Journal of Geophysical Research* **85**, 1500–10.
- Peel, D., Whiten, W. J. & McLachlan, G. J., 2001. Fitting mixtures of Kent distribution to aid in joint set identification. *Journal of American Statistical Association* **96**, 56–63.

- Priest, S. D. 1993. *Discontinuity Analysis for Rock Engineering*. Chapman & Hall, London.
- Riffert, T. W., Schreiber, J., Anwander, A. & Krösche, T. R. 2014. Beyond fractional anisotropy: Extraction of bundle-specific structural metrics from crossing fiber models. *NeuroImage* **100**, 176–91.
- Sáez, P., Peña, E., Doblaré and Martínez, M.A. 2013. Hierarchical micro-adaptation of biological structures by mechanical stimuli. *International Journal of Solids and Structures* **50**, 2353–70.
- Sato, K., Yamaji, A. & Tonai, S. 2013. Parametric and non-parametric statistical approaches to the determination of paleostress from dilatant fractures: Application to an Early Miocene dike swarm in central Japan. *Tectonophysics* **588**, 69–81.
- Sato, N. & Akiyama, Y., 1980. Structural control of the Akenobe tin-polymetallic deposits, Southwest Japan. *Mining Geology, Special Issue* **8**, 175–88.
- Sato, N., Asada, I., Kasamatsu, S., 1977. Geological structure and ore deposits of the Akenobe mine, Hyogo Prefecture, Japan. *Mining Geology* **27**, 245–62.
- Siemes, H., Schaeben, H., Rosière, C. A. & Quade, H. 2000. Crystallographic and magnetic preferred orientation of hematite in banded iron ores. *Journal of Structural Geology* **22**, 1747–59.
- Shanley, R. J. & Mahtab, M. A. 1976. Delineation and analysis of clusters in orientation data. *Mathematical Geology* **8**, 9–23.
- Stevens, B. 1911. The laws of intrusion. *Transactions of American Institute of Mining Engineers* **49**, 1–23.
- Tanaka, H. 1999. Circular asymmetry of the paleomagnetic directions observed at low latitude volcanic sites. *Earth, Planets and Space* **51**, 1279–86.
- Ueda, K., Jacobs, J., Thomas, R. J., Kosler, J., Jourdan, F. & Motola, R. 2012. Delamination-induced late-tectonic deformation and high-grade metamorphism of the Proterozoic Nampula Complex, northern Mozambique. *Precambrian Research* **196–197**, 275–94.
- Watanabe, M., Shibata, K. & Soeda, A., 1984. K-Ar ages of base and precious metal mineralization in the Tungsten Province, southwest Japan. *Geochemical Journal* **18**, 189–93.
- Yamaji, A. & Sato, K. 2011. Clustering of fracture orientations using a mixed Bingham distribution and its application to paleostress analysis from dike or vein orientations. *Journal of Structural Geology* **33**, 1148–57.
- Yamaji, A., Sato, K. & Tonai, S. 2010. Stochastic modeling for the stress inversion of vein orientations: Paleostress analysis of Pliocene epithermal veins in southwestern

APPENDIX 1

FIVE-DIMENSIONAL PARAMETER SPACE

The five-dimensional parameter space for the Bingham distribution was detailed by Yamaji *et al.* (2010), so it is briefly introduced here. A set of the parameters of a Bingham distribution, $\boldsymbol{\kappa}$ and \mathbf{E} , can be identified with a point in a five-dimensional parameter space as follows. First, we introduce a ray denoted by the unit vector, $\boldsymbol{\xi}$, to indicate the position vector, \mathbf{x} , in the space such that

$$\mathbf{x} = \rho \boldsymbol{\xi}, \quad (\text{A1})$$

where ρ is the length of the position vector. The length is a function of the concentration parameters,

$$\rho = \sqrt{\kappa_1^2 + \kappa_1 \kappa_2 + \kappa_2^2}. \quad (\text{A2})$$

We have $\rho = 0$ only for the case of uniform distribution, $\kappa_1 = \kappa_2 = 0$. Non-uniform orientation distributions have positive ρ values. Small and large ρ values indicate broad and concentrated distributions, respectively.

It follows from the inequality (1) with Eqs. (A1) and (A2) that $\mathbf{x} = \mathbf{0}$ in case of $\kappa_1 = 0$. Given κ_1 , κ_2 and \mathbf{E} , we use the ratio, $r = \kappa_2/\kappa_1$, which satisfies $0 \leq r \leq 1$, to determine the corresponding position vector. To this end, we define the deviatoric tensor,

$$\boldsymbol{\varsigma} = \mathbf{E} \left[\frac{\text{diag}(2-r, 2r-1, -r-1)}{\sqrt{3r^2 - 3r + 3}} \right] \mathbf{E}^T. \quad (\text{A3})$$

Let ς_{ij} be the ij th component of this tensor, and ξ_i be the i th component of $\boldsymbol{\xi}$. Then, we have

$$\varsigma_{11} + \varsigma_{22} + \varsigma_{33} = 0, \quad (\text{A4})$$

and

$$\begin{cases} \xi_1 = -\left(\frac{\sqrt{2}}{4} + \frac{\sqrt{6}}{12}\right)\varsigma_{11} + \left(\frac{\sqrt{2}}{4} - \frac{\sqrt{6}}{12}\right)\varsigma_{22} + \frac{1}{\sqrt{6}}\varsigma_{33} \\ \xi_2 = \left(\frac{\sqrt{2}}{4} - \frac{\sqrt{6}}{12}\right)\varsigma_{11} - \left(\frac{\sqrt{2}}{4} + \frac{\sqrt{6}}{12}\right)\varsigma_{22} + \frac{1}{\sqrt{6}}\varsigma_{33} \\ \xi_3 = \varsigma_{23}, \quad \xi_4 = \varsigma_{31}, \quad \xi_5 = \varsigma_{12}. \end{cases} \quad (\text{A5})$$

These linear equations were originally introduced by Sato & Yamaji (2006) for the stress inversion of faults. The Bingham distribution with κ_1 , κ_2 and \mathbf{E} is accordingly identified with the point, \mathbf{x} , in the five-dimensional space through Eqs. (A1)–(A3) and (A5).

On the other hand, if \mathbf{x} is given, the corresponding parameters, κ_1 , κ_2 and \mathbf{E} , are obtained as follows. In case of $\mathbf{x} = \mathbf{0}$, we have $\kappa_1 = \kappa_2 = 0$, and \mathbf{E} is an arbitrary orthogonal matrix. Otherwise, we have $\rho = |\mathbf{x}|$ and $\boldsymbol{\xi} = \mathbf{x}/\rho$. It follows from Eqs. (A4) and (A5) that

$$\begin{cases} \varsigma_{11} = -\left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{6}}\right)\xi_1 + \left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{6}}\right)\xi_2 \\ \varsigma_{22} = \left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{6}}\right)\xi_1 - \left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{6}}\right)\xi_2 \\ \varsigma_{33} = \sqrt{2/3}(\xi_1 + \xi_2) \\ \varsigma_{23} = \varsigma_{32} = \xi_3, \quad \varsigma_{31} = \varsigma_{13} = \xi_4, \quad \varsigma_{12} = \varsigma_{21} = \xi_5. \end{cases} \quad (\text{A6})$$

Then, \mathbf{E} and r are obtained by solving the eigenproblem of $\boldsymbol{\varsigma}$ (Eq. A3). The three eigenvalues satisfying $\varsigma_1 \leq \varsigma_2 \leq \varsigma_3$ give $r = (\varsigma_2 - \varsigma_3)/(\varsigma_1 - \varsigma_3)$. Substituting $\kappa_2 = r\kappa_1$ into Eq. (A2), we have $\rho^2 = (r^2 + r + 1)\kappa_1^2$. Since κ_1 is negative in sign, we obtain $\kappa_1 = -\rho/\sqrt{r^2 + r + 1}$ and $\kappa_2 = r\kappa_1$.

FIGURE AND TABLE CAPTIONS

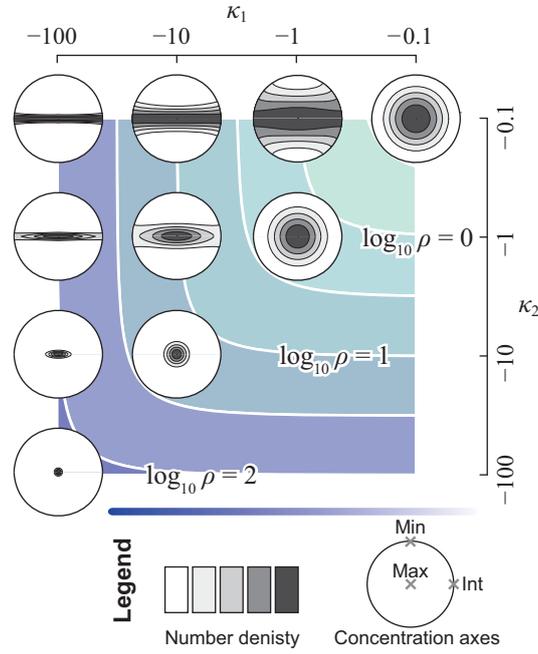


Fig. 1 Equal-area projections showing the probability densities of Bingham distributions with various κ_1 , κ_2 and $\rho = (\kappa_1^2 + \kappa_1\kappa_2 + \kappa_2^2)^{1/2}$. The values of $\log_{10}\rho$ are indicated by contours in the background. The maximum, intermediate and minimum concentration axes of a Bingham distribution make right angles with each other. The concentration parameters, κ_1 and κ_2 , are defined to be negative in sign or equal to zero: Uniform distribution is denoted by $\kappa_1 = \kappa_2 = 0$. Concentration along the great circle defined by the maximum and minimum concentration axes is denoted by $|\kappa_1|$, and that by the maximum and intermediate concentration axes is denoted by $|\kappa_2|$. A circular cluster is denoted by the condition, $\kappa_1 = \kappa_2$, and a girdle or elongated cluster by $\kappa_1 < \kappa_2$.

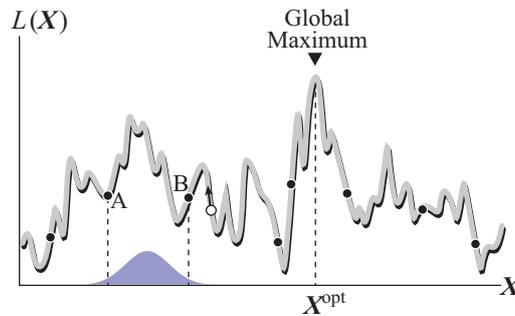


Fig. 2 Schematic illustration for the multimodal function, $L(X)$, the global maximum of which

represents the best partition of 3D orientation data. The local maxima of $L(\mathbf{X})$ are depicted by the peaks of the graph. The EM algorithm starts from a randomly chosen point (open circle), and climbs the slope of the function (arrow). The global maximum is detected by the algorithm only when the starting point is chosen by chance on the slope leading to the maximum. The genetic algorithm scatters dozens of ‘individuals’ (closed circles) at a time, and improves their position, \mathbf{X} . For example, if the two individuals A and B are chosen, the Gaussian distribution with the maximum at their midpoint (blue curve) is used to produce their ‘child,’ and an individual in the population is discarded instead. The global maximum is expected to be reached by an individual with the largest L value by chance, but the chance is increased by giving higher probabilities to the individuals with high L values to have children, and assigning lower probabilities to them to be excluded from the population.

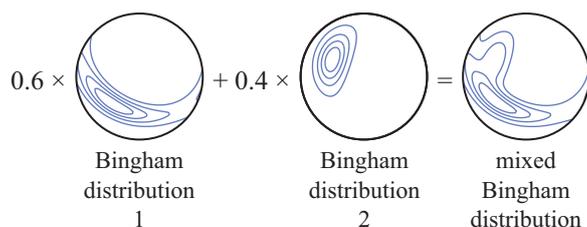


Fig. 3 Equal-area projections showing the density contours of a mixed Bingham distribution and its Bingham components. In this case, the mixing coefficients are $\varpi^1 = 0.6$ and $\varpi^2 = 0.4$; and $K = 2$.

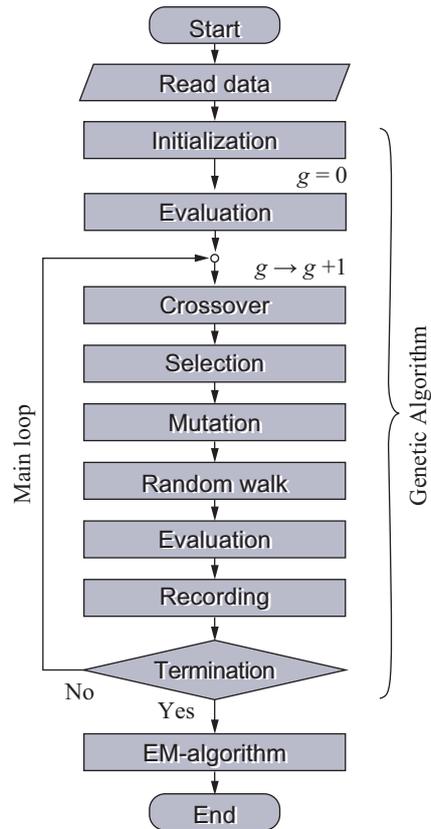


Fig. 4 Flow chart of the real-coded genetic algorithm to search for the vector, \mathbf{X} , that maximizes $L(\mathbf{X})$ for a given K value. The population created in the initialization stage is assumed to be of the 0th generation, and g denotes the number of generation.

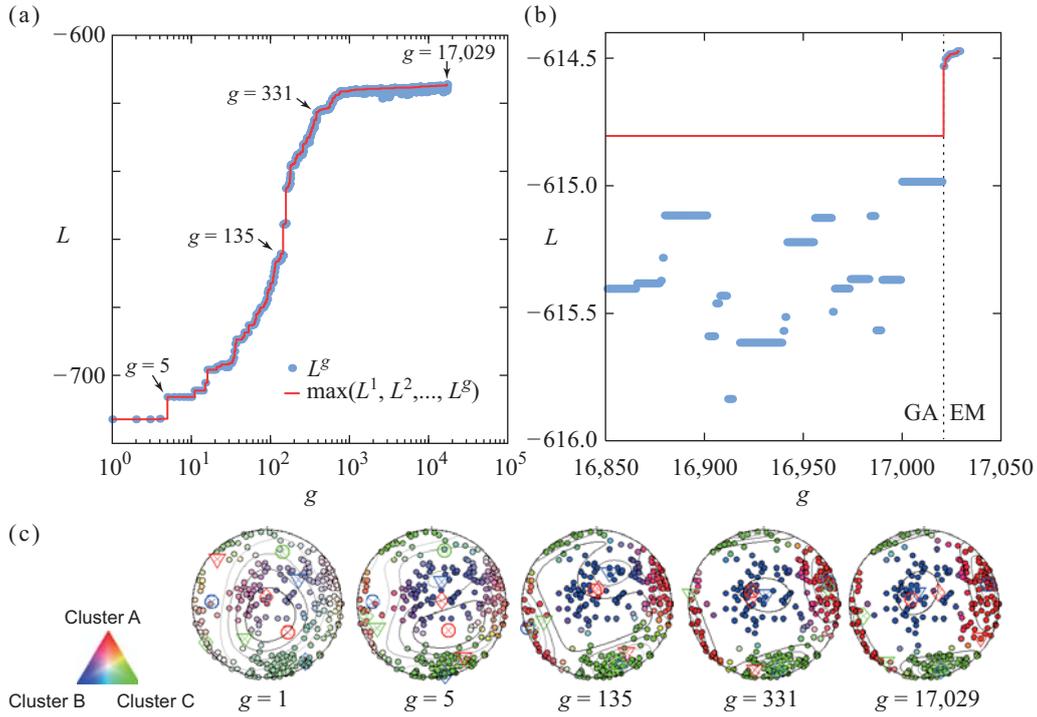


Fig. 5 Diagrams illustrating the partitioning process of the fracture orientations from the San Manuel Copper Mine, Arizona (Shanley & Mahtab 1976), into three groups ($K = 3$) by the present method ($\Delta g = 1000$). See the animation in the supplementary file, GA.mov, which shows this process. (a) The entire process. L indicates $L(X)$. (b) The close-up of the final part of the process. Data points are aligned in right-stepping echelon manner, because the crossover resulted in increasing L and the selection resulted in the sudden drops. The genetic algorithm (GA) was terminated at $g = 17,020$, and was switched to the EM algorithm, which was iterated 9 times, to improve the partition. (c) The partitions at several generations. The poles of fractures are shown by lower-hemisphere, equal-area projections; and their memberships, m_n^k , are indicated by colors. The ternary plot shows the correspondence of the memberships and colors.

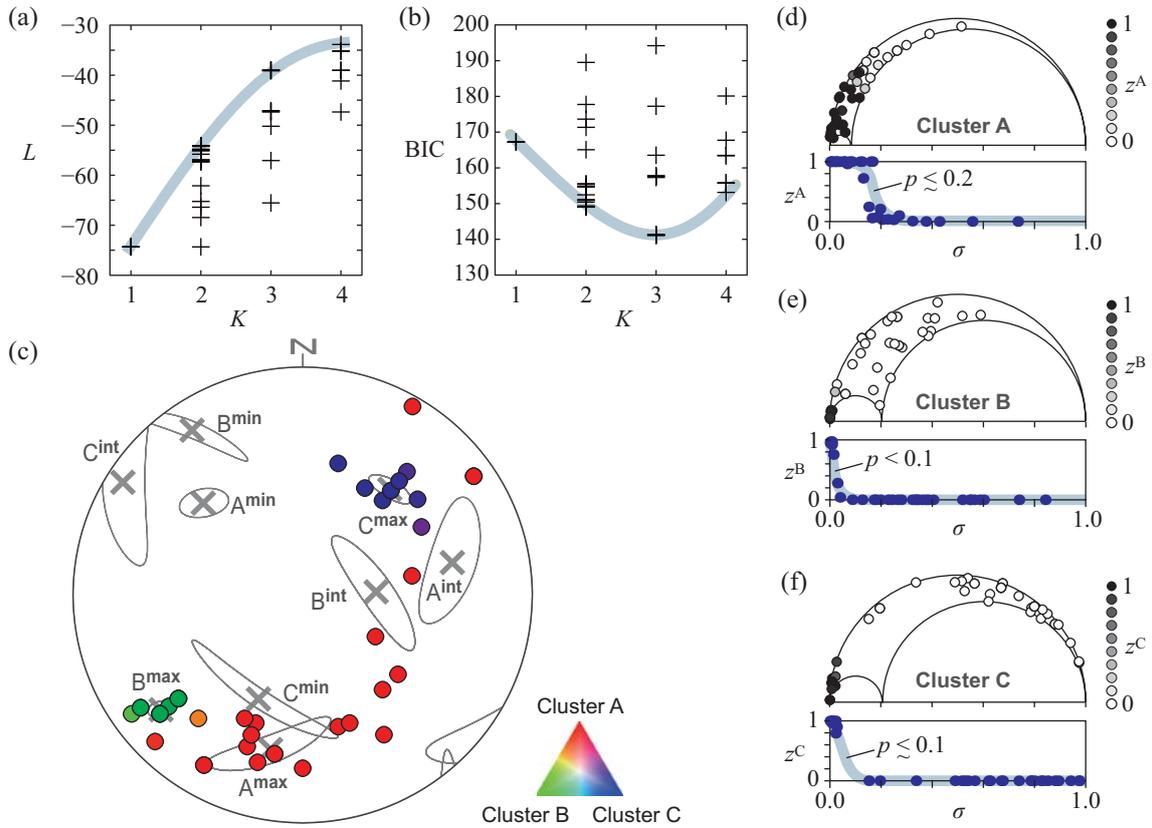


Fig. 6 The fitting of mixed Bingham distribution to the orientations of ore veins in the Akenobe Mine, SW Japan by the present technique. The fitting was done with $\Delta g = 1000$. (a) L values versus K . (b) BIC versus K indicating the minimum at $K = 3$. It means that the orientations should be partitioned into three groups. (c) Lower-hemisphere, equal-area projection for the orientations listed by Sato *et al.* (1977). Crosses and gray lines indicate the concentration axes and their 95% confidence ellipse of the three clusters, A, B and C. The memberships of the data are indicated by the colors in the ternary diagram. (d–f) Mohr diagrams and driving pressures of the veins belonging to the clusters, A, B and C. The memberships of the data points to the clusters are depicted by gray levels in the Mohr diagrams. The diagrams under the Mohr diagrams show the memberships versus normal stress of veins. The normal stress at which the membership declines indicates the driving pressure, p (Yamaji *et al.* 2010).

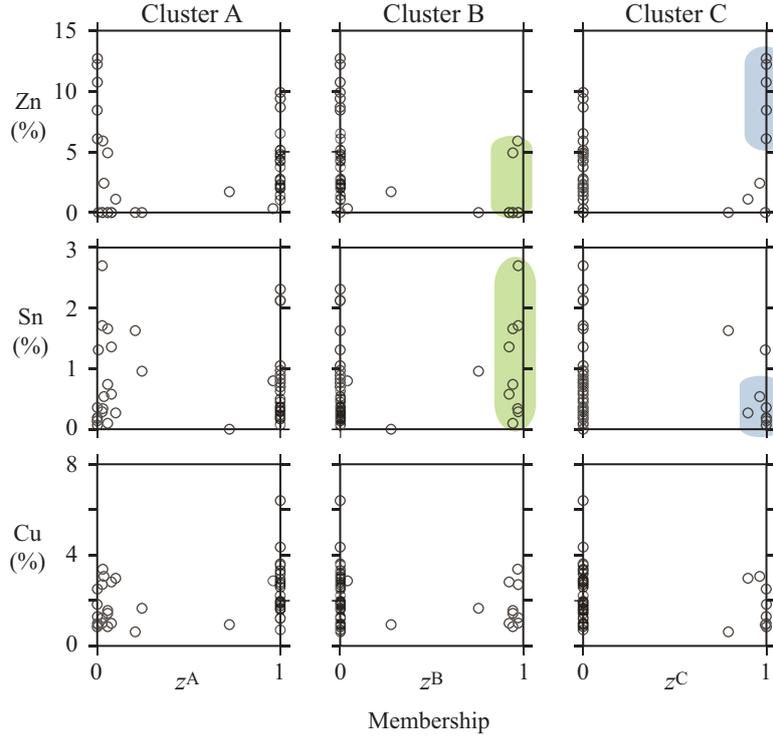


Fig. 7 The Zn, Sn and Cu contents (Sato *et al.* 1977; Sato & Akiyama 1980) versus the memberships of the Akenobe veins in Fig. 6c . The veins with high memberships to the cluster C, z^C , had high Zn and low Sn contents, though there were a few exceptions. The veins with high z^B values had low Zn contents. Their Sn contents showed a wide variety, but were high compared to the veins of the cluster C (there were exceptions).

TABLE 1 Clustering result for the orientations of ore veins in the Akenobe Mine (Fig. 6).

	concentration axes			κ_1	κ_2	Φ	ϖ
	maximum	intermediate	minimum				
Cluster A	192°/32°	078°/33°	313°/41°	-25.1	-2.00	0.08	0.55
Cluster B	230°/22°	087°/64°	327°/14°	-160	-32.2	0.20	0.24
Cluster C	039°/40°	302°/08°	203°/49°	-96.2	-20.1	0.21	0.21