Journal of Structural Geology 27 (2005) 929–936

JOURNAL OF STRUCTURAL GEOLOGY

www.elsevier.com/locate/jsg

A hierarchical cluster approach for forward separation of heterogeneous fault/slip data into subsets

Yehua Shan[a,*], Norman Fry[b]

[a]Department of Marine Geology, Ocean University of China, Qingdao City 266003, P.R. China
[b]School of Earth, Ocean and Planetary Sciences, Cardiff University, Cardiff CF10 3YE, UK

Available online 2 April 2005

Abstract

A new simple method of stress inversion uses hierarchical cluster analysis for forward separation of heterogeneous fault/slip data into subsets. Fault/slip data are classified into homogeneous fault classes, and a clustering routine classifies these into subsets. The method includes a way of discarding some residual data at the first stage that makes it fairly easy to recognize and eliminate some spurious fault data. However, this method is a type of hard division that overlooks the indeterminate nature of fault data. The more heterogeneous the data, the larger the calculation needed to find from a K-data set the homogeneous fault class that agglomerates a pair of 5-data subsets, sampled in a binomial distribution, with the maximum similarity in estimated stress vector between them. The K-data set is a working data group successively taken from the whole data. Given P phases of different stress state, the minimum value of K is $5P+1$. Results from applying the method to two examples, artificial and real, demonstrate the feasibility of the method.

Keywords: Heterogeneity; Fault/slip data; Stress inversion; Hierarchical clustering

1. Introduction

Inversion of stress from geological structures is becoming a significant tool for structural geologists (Ramsay and Lisle, 2000). It is usually based upon measured fault/slip data. These data are referred to as heterogeneous if they record multiphase deformation, or homogeneous if they record a single-phase deformation. The former is quite common by virtue of both spatial and temporal variation of the stress field. In nature, faults might be either newly produced in one individual phase or reactivated in a subsequent phase. Field observations in a study area, including fault orientations, properties, cross-cutting relationships, and so forth, are normally required to determine the relative age of individual faults, and to establish faulting phases, but these are not always available and complete. Homogeneous deformation is assumed by stress inversion (Carey and Brunier, 1974; Angelier, 1979; Etchecopar et al., 1981; Angelier, 1994). Conventional inversion methods are based on single-phase deformation, and are applicable to homogeneous fault/slip data. They lose validity in the case of heterogeneous fault/slip data (Will and Powell, 1991; Nemcok and Lisle, 1995; Nemcok et al., 1999).

Nearly all methods developed for separating a set of heterogeneous fault/slip data into homogeneous subsets are numerical. They can be categorized as grid-searching schemes (e.g. Hardcastle and Hills, 1991; Yamaji, 2000), hierarchical cluster analyses (Nemcok and Lisle, 1995; Nemcok et al., 1999), dynamic cluster analysis (e.g. Huang, 1988; Shan et al., 2003, 2004a), and graphic presentations (Simón-Gómez, 1986; Fry, 1992; Célérier, 1995; Célérier and Séranne, 2001; Shan et al., 2004b). The advantages and disadvantages of these methods in separating heterogeneous data have been discussed by Shan et al. (2003) and Nemcok et al. (1999). Interested readers are encouraged to refer to their papers.

To solve for stress requires a minimum number of fault/slip data. It is generally 5 or 6, or another figure that was no less than the division number of the parameter space. That is to say, homogeneous subsets identified in any way, although variable in number, all contain at least the minimum number of fault/slip data. Hierarchical cluster

* Corresponding author. Current address: Computational Geosciences Research Center, Central South University, Chengsha City 410083 P.R. China. Tel.: +86 731 8877077, fax: +86 731 8879602
E-mail address: shanyehua@yahoo.com.cn (Y. Shan).

0191-8141/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsg.2005.02.001

analysis cannot be directly adopted for this purpose because of its emphasis on the individual fault datum. It requires some kind of preprocessing of the fault data. After discretizing the parameter space, Nemcok and Lisle (1995) introduced a variable having its components assigned 0/1 according to the fit/misfit under a prescribed resolution between measured fault slip and calculated fault slip under the given discretized stress. So each fault/slip datum is characterized by an assigned value of the new variable, to which hierarchical cluster analysis is applied.

Their preprocess focuses on fault/slip data similarity in the parameter space; hence, it is somewhat reverse in a search for stress solutions. By contrast, the theme of this communication is to develop a forward preprocess, accomplished during the search for stress solutions. This paper tests this preprocess using sample data and discusses its advantages and disadvantages.

2. Methodology

2.1. Parameter space

After some transformation, stress inversion may be turned into a linear problem that can be simply manipulated (Fry, 1999; Shan et al., 2003). Further, under auxiliary constraints, the parameter space is reduced to a five-dimensional unit sphere centred at the origin. In the reduced space, homogeneous fault data tend to lie in a hyperplane through the origin, normal to which is the vector representing the optimal stress. Accordingly, heterogeneous fault data tend to lie in many such hyperplanes (Fry, 1999). There is an analytical solution of the optimal stress vector from a given fault data set (Shan et al., 2003).

2.2. Homogeneous fault class

For a heterogeneous fault/slip data set, we assume that each homogeneous subset has a number of data that are larger than the dimension of the parameter space (5). This is explicitly or implicitly necessary for all inversion methods. We take a set of K fault/slip data from the whole L data ($5 < K \leq L$). K must be large enough to ensure that there is usually a homogeneous subset having more than 5 data within the selected set. We call this subset a homogeneous fault class. We propose to arrive at it as follows:

1. Select all 5-datum subsets from the K-data set in a binomial distribution,
2. Calculate the similarity coefficient between the stresses estimated from any two different 5-datum subsets using the analytical scheme (Shan et al., 2003), and
3. Fuse the two 5-datum subsets having the largest similarity coefficient to produce a homogeneous fault class. If there is more than one pair of such a kind, only the first pair is taken into account. The arbitrary selection of any one among such pairs has no effect on the final result.

In this way, the homogeneous fault class has a data number >5 and ≤ 10. There are other possible ways to look for a homogeneous fault class, but they are beyond the scope of this paper.

When K equals L, the selected fault set becomes the whole data, from which exhaustive selection would give too many 5-datum subsets. This enormous calculation is not necessarily worthwhile as a way to look for a single homogeneous fault class, because our strategy is to find and cluster all the classes of the same kind. Therefore, it is practical to have a value of K as small as possible. The value of K, as well as its influence on defining such a class, will be discussed below.

Once a homogeneous fault class is obtained, its data are excluded from the whole L-data set, and the above-mentioned procedure is then repeated for another class. This process does not terminate until the residual data are less than K in number.

2.3. Residual data

At the final stage, the method given above is inapplicable to the residual data, due to their small number (<5), or to the existence of some spurious result, as will be discussed below. Either way, the residual data need to be processed differently. Each datum must have one of three fates: assigned to an old fault class, or to a new fault class, or discarded. The choice depends upon its angular cosines with the stress vectors of the obtained fault classes.

Take Wi to be the cosine of the angle between a specific residual fault datum and the stress vector that is normal to datum vectors in the ith homogeneous fault class calculated. Let us introduce W as a limit (maximum cosine, minimum angle). For example, W was assigned the value 0.1 in this paper. If $W > Wi$, we assign the datum to the ith fault class; if not, we make no decision until all data are checked. For the latter data, we will discard them if they are <5 in number, or open a new homogeneous fault class for them if they are $\geqq 5$. That is to say, a few spurious fault data, if they exist, would be recognized and eliminated in that way before classification.

2.4. Systematic classification

Using the above means, we eventually obtain a set of homogeneous fault classes from fault data. These classes are considered as entities, upon which our classification is based. There are a variety of hierarchical cluster schemes, such as single linkage, complete linkage, average linkage, median, centroid, increase in sum of squares (or Ward's (1963) method), and so forth (see Everitt et al., 2001). According to the study by Mojena (1977), Ward's (1963) method is the optimum choice for producing a monotonic series of solutions. However, these conventional schemes cannot be applied directly to the fault/slip data of the homogeneous fault classes, as our purpose is to classify entities on the sole basis of the most similar stress vectors. Unlike their stress vectors, the fault/slip data of entities that are classified together may not be spatially close to each other in the parameter space.

To achieve our purpose, we develop a new classification scheme that minimizes the sum of squares of cosines between the stress vector for a certain entity, and the individual fault datum vectors from which it is determined. This sum was also defined by Shan et al. (2003, 2004a) as the objective function, of which stress vectors were unknown variables, solved for by minimizing it under some auxiliary constraints. Let E_p and E_q be the sum for entities p and q, respectively, at the first stage. These two entities are classified into a new large entity with a sum of E_{pq}. It is proved in Appendix A that:

$$E_{pq} \geq E_p, \quad E_{pq} \geq E_q \tag{1}$$

This indicates that the new entity inevitably becomes more dispersed than its two daughter entities. In each stage of classification, only those two entities with the least sum of square cosines, or E_i in the ith stage, are agglomerated into a new entity. Finally we have a series of E_i that increase monotonically. Subscripts denote the number of entities resulting at each stage: $E_1 \geqq E_2 \geqq \ldots \geqq E_{n-2} \geqq E_{n-1}$. This is similar to Ward's (1963) method.

2.5. Number of subsets

For a given heterogeneous fault/slip data set, we wish to know through stress inversion how many phases are mixed in the set. The number of homogeneous subsets is generally unknown, and needs to be solved for. Can it be determined from hierarchical cluster analysis? This is a difficult problem since there are no satisfactory methods for determining it (Mojena, 1977; Everitt, 1979; Hartigan, 1985).

Given a monotonic increase in values of E_i ($i=1, 2,\ldots,n-1$), a common empirical procedure for this purpose is to utilize the E_i distribution, and look for a 'significant' change in E_i as a stopping point, above which classification seems unnecessary. In a statistical sense, a critical value E_{limit} is defined as one lying in the upper tail of the E_i distribution (Mojena, 1977):

$$E_{\text{limit}} = \mu + k\delta \tag{2}$$

where k is the standard deviate; the μ and δ are the mean and unbiased standard deviation of the E_i distribution, respectively. Mojena (1977) argued that values of k in the range 2.75–3.50 might give a good overall classification.

If no value for E_i is larger than the critical value, we must choose: (a) one cluster; (b) the ith stage for which the $i+1$th stage yields the largest standard deviate; or (c) some other appropriate heuristic rule.

However, as noted below, E_{limit} generally depends upon the value of E_n that is much larger than E_{n-1}. The stopping rule adopted is thus most likely to produce a smaller optimal number of clusters. Moreover, we can imagine that the knickpoint determined in this way would be of no value in the case of fault data randomised in the parameter space. For the sake of compensation, we introduce the square root of the mean of squares of cosines between individual fault datum vectors and the stress vector estimated from them (d_i):

$$d_i = \sqrt{\frac{E_i}{m}} \tag{3}$$

where m is the number of data in the cluster at the ith stage whose sum of squares of cosines is E_i. Further, a subtended angle is simply calculated from the root of the mean of square cosines (see the inset of Fig. 1). Both are good indices showing the degree of dispersion among the grouped data as a whole. The more dispersed the data, the larger are the indices. Our calculation shows (Fig. 1) that the parameter space is divided into two subspaces of equal area by the hyperlines intersected in that angle with the estimated stress vector, if half of the subtended angle reaches ca. 22°. This seems to represent the nature of fault data randomised in the parameter space, and may be considered as an upper limit of subtended angles calculated from fault data sets.

Both Mojena's (1977) stopping rule and the concept of subtended angle were used below in this paper to look for the optimal number of clusters.

2.6. Procedure

The proposed inversion method has two parts: to classify fault/slip data into homogeneous fault classes, and further to classify these classes into subsets. The entire procedure to realize it is as follows:

1. Input measured fault/slip data, the data number K for selecting K-datum sets, and angular cosine limit W,
2. Calculate direction cosines of each fault/slip datum, and then its datum vector according to Fry's (1999) equations,
3. Find homogeneous fault classes from K-datum sets until there are fewer than K data in the residual set,
4. Assign each of the residual data to an old or new homogeneous fault class, or discard it,
5. Apply the above cluster scheme to the homogeneous fault classes obtained, as follows:

(a) determine optimal stress vectors for each class, and for every permutation of pairs of classes, using the method developed by Fry (1999) and Shan et al. (2003),
(b) calculate the increase in sum of squares of cosines for every permutation of pairs of classes,
(c) fuse those two entities with the smallest increase in sum

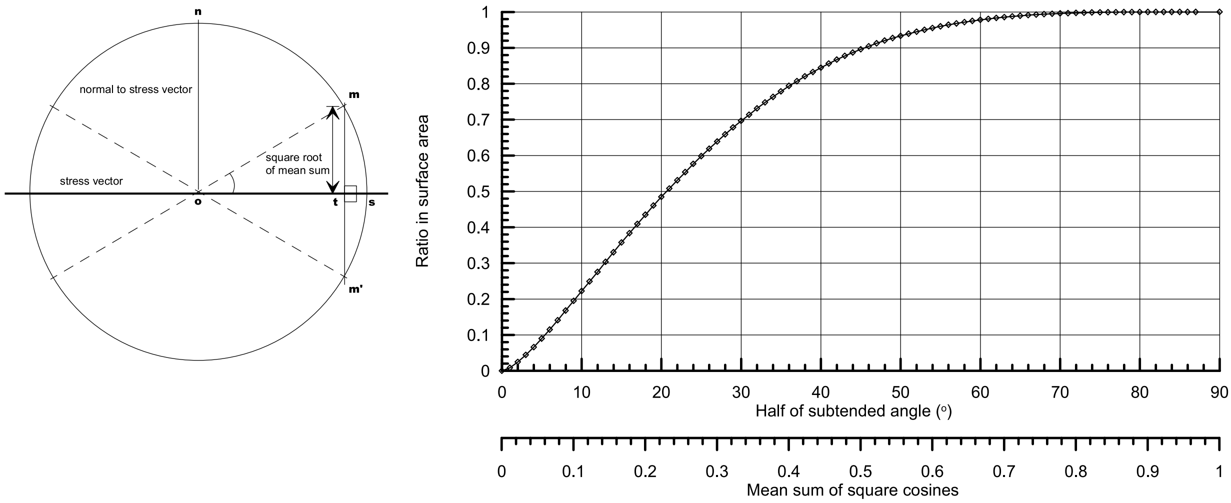

Fig. 1. Ratio in surface area of a belt around a stress vector to the unit hyper sphere in the 5D parameter space. The belt has a subtended angle ∠ mom′ between vectors from the upper and lower boundaries to the origin. The graph at the right side was obtained by Monte Carlo sampling—a number of 100,000 vectors evenly distributed in the hyper sphere were generated, and the ratio was estimated by counting the number of vectors having a smaller angle than half of the subtended angle, and dividing it by the whole number. Half of the subtended angle reaches ca. 22° while the ratio becomes 0.5. See the text for further explanation.

of square cosine (that is, least E_i in the ith stage) into a new class, and

(d) terminate when there one class remains; otherwise, return to step (a)
(Should two entities have equal smallest E_i at (c) above, one is taken arbitrarily. This will not affect the overall result.),

6. Determine the best number of subsets from the E_i distribution and from the square root of mean E_i, and separate the heterogeneous fault data into these homogeneous subsets, and
7. With the constraints of measured fault slips, restore the stress tensor from the calculated stress vector of each homogeneous subset, according to Fry's (1999) equations.

3. Test

In order to validate the proposed method, a set of artificial heterogeneous fault/slip data was taken from Shan et al. (2003, listed there as case one in appendix C). The data set is equally mixed from three prescribed phases. Each phase has 20 fault data, of which each fault slip is strictly parallel to the maximum resolved shear along the fault plane under the prescribed stress. Data 1–20, 21–40 and 41–60 belong to the three phases, respectively.

Let data number K and angular cosine limit W be 16 and 0.1, respectively. Results through applying the proposed method to the data are shown in Figs. 2 and 3. Nine homogeneous fault classes were recognized by the method. Classification of these classes into three groups corresponding to the three prescribed phases is clearly efficient because of the abrupt change in the E_i distribution between E_5, which is nearly zero, and the relatively much larger E_6 (Fig. 3a). The abrupt change corresponds to a lower value of standard deviate k, <2.5, which suggests an underestimated number of clusters using Mojena's (1977) stopping rule. Two groups remain when $k \geq 2.5$. This fairly wide range would lead to a biased acceptance of an optimal number of three subsets without knowledge of controlling stresses.

Similar to the E_i distribution is the distribution of half subtended angles (Fig. 3b). Half of the subtended angle is in the range 0.00–2.89° when all classes are fused into three clusters.

4. Case study

The real example, taken from Xie and Liu (1989), consists of 198 fault/slip data measured in the middle segment of the active NEE-trending transcurrent Altun fault. The fault defines the northern boundary of the Tibetan Plateau. It has a cumulative sinistral strike slip of about 350 km as a consequence of extrusion of thickened Tibetan Plateau created by the collision between India and Eurasia since 45 Ma (e.g. Molnar and Tapponnier, 1975).

A number of inversion methods have recently been applied to this fault/slip data set, including conventional inversion (Xie and Liu, 1989), a fuzzy C-lines clustering algorithm (Shan et al., 2004a), and graphic presentation (Shan et al., 2004b). All these results indicate the heterogeneity of the data set. Two distinct tectonic phases were recognized and the determined orientation and stress ratio varied slightly with inversion method. According to Xie and Liu (1989), the maximum principal stress has a bearing of 166.71° in subset 1, and of 72.15° in subset 2. Projection of these data by the stereonet method of Shan

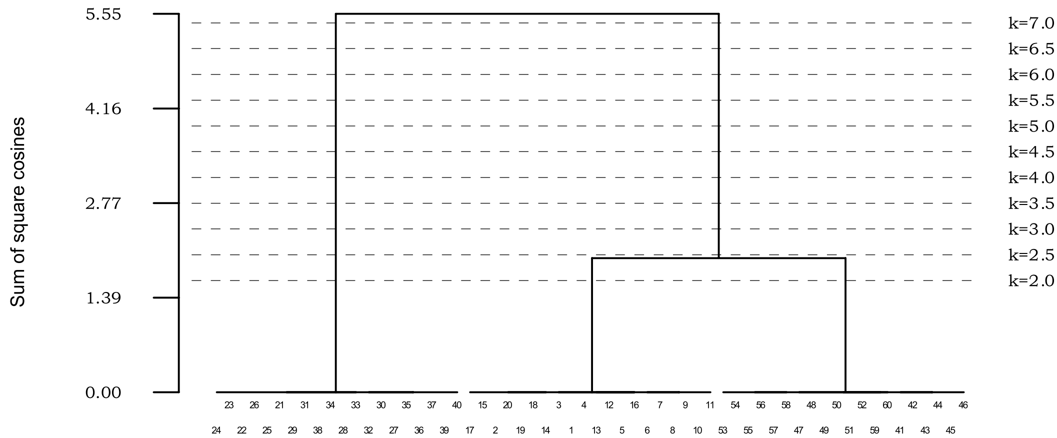

Fig. 2. A dendrogram of classification of the artificial fault data through applying the proposed method. The mean and the unbiased deviation of the least increases in sum of square cosines in varying stages are 0.094 and 0.528, respectively. Dashed lines represent critical values under varying standard deviate *k*. See the text for further explanation.

et al. (2004b) showed more heterogeneous features, one major and two or more minor girdles, but these remain in doubt because the validity of the method's assumption of Andersonian stress state had not been ascertained in this case.

In the light of these considerations, data number *K* was assigned a value of 16—large enough to process three-phase fault data. The angular cosine limit *W* was 0.1. By means of our proposed method, 31 homogeneous fault classes were recognized from the data set, and then were further classified (Figs. 4 and 5). In the dendrogram, all classes were eventually fused into a single group at $E_{30}=15.44$ and a half subtended angle of 4.64°, and two groups at $E_{29}=8.68$ and a half subtended angle of 4.98°. A significant change in the distribution takes place in the E_i distribution between E_{25} and E_{26} (Fig. 5a) where *k* is nearly in the range 2.0–2.5. This indicates an optimal number of 6 subsets, thus implying a more heterogeneous nature of data set than previously considered.

In terms of precise estimation, the previous model of a two-phase mixture (Xie and Liu, 1989; Shan et al., 2003, 2004a) seems to be a very rough approximation to the real complex data. It is certain that a simple mixture of two-phase data fails to account for many subtle linear structures in the parameter space, as illustrated by the stereonet projection assuming Andersonian stress state showing a loosely defined girdle with a large maximum (Shan et al., 2004b). It may be significant that Xie and Liu (1989) estimated stresses only from fault data measured in the subsegments of the middle Altun segment. Their two distinct phases accord with field observations. The presence of many potential homogeneous subsets cannot be explained by measurement errors at outcrop. Other possible reasons for fluctuation from stress inversion assumptions must be included to account for the complexity of the data, such as intermittent fault slip, slip not synchronous everywhere along the fault (e.g. Price, 1988; Gutscher et al.,

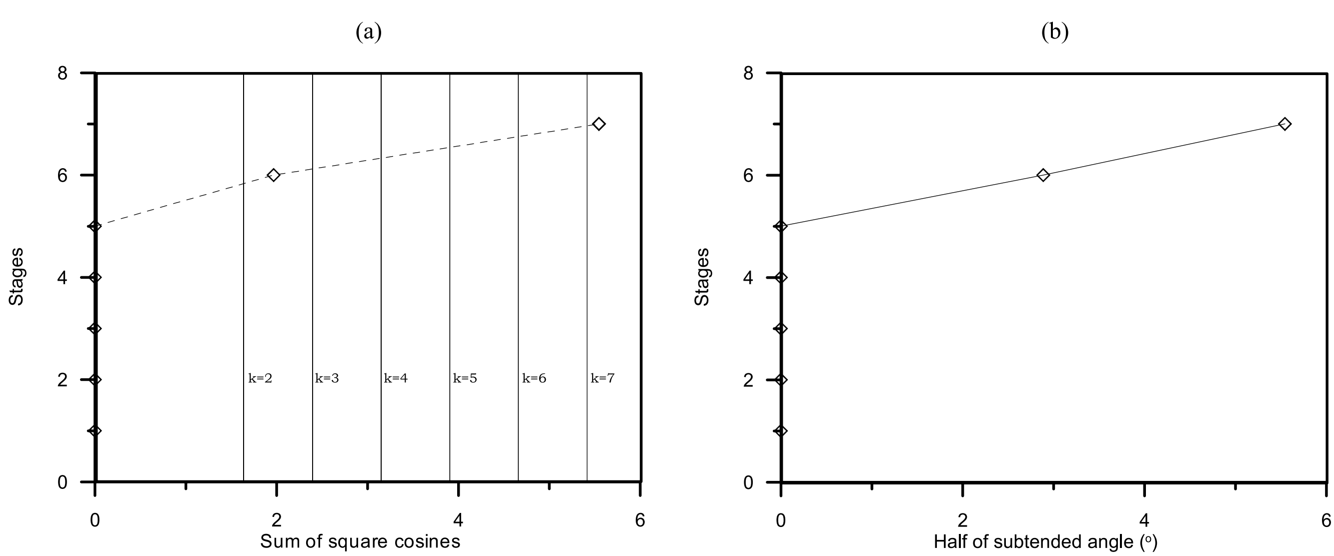

Fig. 3. Sum of square cosines (a) and half of the subtended angle (b) in stages. Thin lines represent critical values under varying standard deviate *k*.

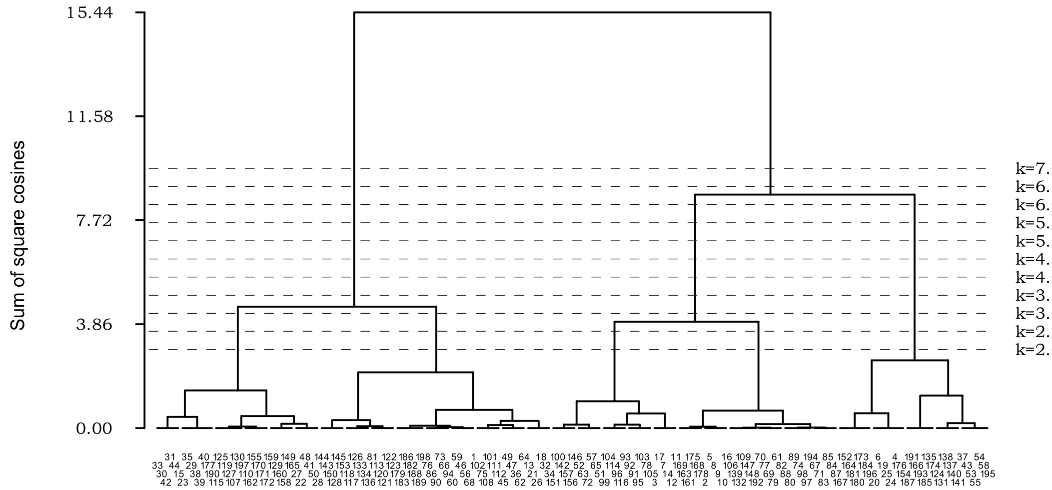

Fig. 4. A dendrogram of classification of the 198 fault/slip data taken from Xie and Liu (1989), using the method proposed. The mean and the unbiased deviation of the least increases in sum of square cosines in varying stages are 0.070 and 0.338, respectively. Dashed lines represent critical values under varying standard deviate *k*. Data number *K* and angular cosine limit *W* were assigned values of 16 and 0.1, respectively. There are 33 homogeneous fault classes, each having 6 or a little more fault data and being represented by a horizontal short line at the bottom.

1996), heterogeneous distributions of earthquake shear stress drop over much of the rupture surface (e.g. Day et al., 1998), and heterogeneous deformation (i.e. stress) distributions in structurally complex areas (e.g. Mitra, 1987; Koyi, 1995).

5. Discussion

5.1. Advantages and disadvantages

The proposed inversion method is different from many pre-existing methods in that it is simple and direct. Hierarchical classification of fault data in a forward way parallels our thinking. The way of processing the residual fault data allows one to recognize and leave out some spurious data, which is of practical use. Spurious fault data, although in a minority, seem rather common, presumably due to diverse factors, including mistakes at measurement or in recording, fluctuation from the assumptions of stress inversion, and so forth.

This method is in the category of hard division meaning that each datum only belongs to a single subset. However, fault/slip data are normally fuzzy by virtue of indeterminability of the individual datum as well as fluctuation from inversion assumptions. Hard division overlooks this common feature, rather like a wave filter. It tends to fail in recognizing subtle linear structures among fault data in the parameter space (Shan et al., 2003). Other disadvantages are discussed below.

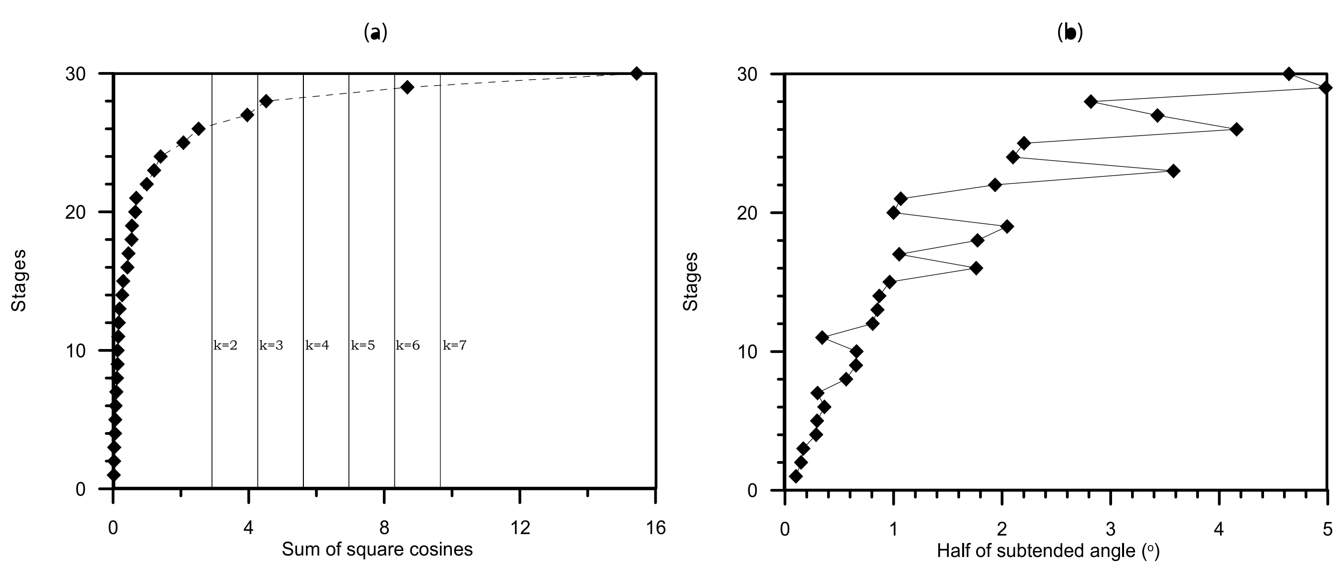

Fig. 5. Sum of square cosines (a) and half of the subtended angle (b) in stages. Thin lines represent critical values under varying standard deviate *k*.

5.2. Data number K

From the method for finding a homogeneous fault class from K data, it is fairly easy to appreciate that K has a minimum number of $5P+1$ if there are P phases in the fault data. Variation in value of K might have some influence on defining a homogeneous fault class and on run time. For values of K smaller than the minimum, we will have shorter run times but likely a heterogeneous fault class. For example, if a set of four-phase fault data has two data under each phase at $K=8$, the cluster obtained is absolutely heterogeneous. This possibility decreases to zero with increasing values of K. However, this would involve the cost of enormous calculations. In accord with the binomial theorem, the number of 5-datum subsets is $C(K, 5)=K!/5!$. It increases dramatically with K, for instance 2,118,760 at $K=50$.

For an unknown number of phases there is no general rule for assigning a definite number K that minimises run time. We have to find some remedy—K as large as possible, at the cost of satisfactory run time. This seems not to be a big problem for today's large-memory, fast-running personal computers. In this paper, K has a prescribed value of 16, so that fault/slip data of no more than three phases could be processed. We believe this may be large enough to process most heterogeneous fault data sets measured at outcrop or in core, which have homogeneous subsets.

5.3. Homogeneity of fault class

The first step of the method is to find a series of homogeneous fault classes from a fault/slip data set. Given a large enough value of K, the great majority of fault classes obtained are homogeneous, but there remains a very small, theoretical possibility that a class is heterogeneous. If a fault datum lies in or near the intersection of two optimal hyperplanes in the parameter space, we will have equal possibility of classifying it to either subset. Its final assignment is dependent only on the way that we take the K-data set from the whole data. A datum belonging to subset A may be classified to subset B. This problem is not obvious until there are quite a number of data in or near the intersections. Although such a disposition in the parameter space can be made visible by the methods of Fry (1992, 1999) and Shan et al. (2004b), it is almost impossible to determine the influence on stress estimation by most inversion methods when applied to a real data set in advance of separation into homogeneous subsets and estimation of stress.

6. Conclusion

An inversion method has been developed for forward separation of heterogeneous fault/slip data into subsets through applying hierarchical cluster analysis. It has two parts: fault/slip data are classified into many homogeneous fault classes, and these classes are then classified into subsets. In the first stage, we take a series of K-data sets from the whole data. With each, we proceed, among all pairs of 5-datum subsets selected in a binomial distribution from the selected K data set, to find the homogeneous fault class that consists of the two 5-datum subsets having maximum similarity in their optimal stress and then exclude this class from the whole data. This is repeated until there is insufficient data to make up a K set. The residual data are then assigned to an old fault class or a new fault class, or discarded.

In the second stage, the homogeneous fault classes obtained are hierarchically classified by using our novel cluster method. This defines an increase in the sum of squares of cosines between the optimal stress vector and individual fault datum in a fault class as the partition criterion. The sum is always greater than or equal to zero, so a monotonic increase in solutions is guaranteed. Both Mojena's (1977) stopping rule and the concept of subtended angle are adopted to determine the preferred number of subsets.

The proposed method is simple, direct and forward. The method of discarding some residual data at the first step makes it fairly easy to recognize and eliminate some spurious fault data. Spurious data, although in a minority, are common and present a big challenge for many inversion methods. On the other hand, this method is a type of hard division that overlooks the indeterminate nature of fault data. The more heterogeneous the data the more enormous the calculation needed to find the homogeneous fault class from the K data set. Given P as the number of phases, the minimum number of K is $5P+1$.

Two examples, one artificial and one real, have demonstrated the feasibility of the proposed method.

Acknowledgements

This paper was reviewed by Drs M. Nemcok and C.L. Liesa who made valuable suggestions for improving it.

Appendix A. Monotonic increase in sum of square cosines with grouping

Let us consider two entities p and q, each having sum of square cosines, E_p and E_q, respectively, between the optimal stress vector and fault datum. They are classified into a new entity with a sum of E_{pq}. The method developed by Fry (1999) and Shan et al. (2003) was used in this paper to solve for the optimal stress vector for a certain entity. As proven by Shan et al. (2003), the optimal stress vector is the least eigen vector of the data matrix. Let A_p, A_q and A_{pq} stand for data matrices of entities p, q and pq, respectively. From definition, we know $A_{pq}=A_p+A_q$. Let x_p, x_q and x_{pq} stand for the eigen vectors corresponding to the least eigen values for entities p, q and pq, respectively:

$$\begin{aligned} E_{pq} &= x_{pq}A_{pq}x_{pq} \\ &= x_{pq}(A_p + A_q)x_{pq} \\ &= x_{pq}A_px_{pq} + x_{pq}A_qx_{pq} \end{aligned}$$

$$\because \quad x_{pq}A_px_{pq} \geq x_pA_px_p = E_p, \quad x_{pq}A_qx_{pq} \geq x_qA_qx_q = E_q$$

$$\therefore \quad E_{pq} \geq E_p + E_q$$

Therefore, in this way, an agglomerated entity always has no less sum of square cosine than its two daughter entities.

References

Angelier, J., 1979. Determination of the mean principal directions of stresses for a given fault population. Tectonophysics 56, T17–T26.

Angelier, J., 1994. Fault slip analysis and palaeostress construction. In: Hancock, P.L. (Ed.), Continental Deformation. Pergamon Press, pp. 53–100.

Carey, M.E., Brunier, M.B., 1974. Analyse theorique et numerique d'un modele mecanique elementaire applique a l'etude d'une population de failles. Compte Rendus Hebdomadaires des Seances de I'Academie des Sciences 279, 891–894.

Célérier, B., 1995. Tectonic regime and slip orientation of reactivated faults. Geophysical Journal International 121, 143–161.

Célérier, B., Séranne, M., 2001. Breddin's graph for tectonic regimes. Journal of Structural Geology 23, 789–801.

Day, S.M., Yu, G., Wald, D.J., 1998. Dynamic stress changes during earthquake rupture. Bulletin of the Seismologic Society of America 88, 512–522.

Etchecopar, A., Vasseur, G., Daignieres, M., 1981. An inverse problem in microtectonics for the determination of stress tensors from fault striation analysis. Journal of Structural Geology 3, 51–65.

Everitt, B.S., 1979. Unresolved problems in cluster analysis. Biometrics 35, 169–181.

Everitt, B.S., Landau, S., Leese, M., 2001. Cluster Analysis, 4th edition Edward Arnold, London. 237pp.

Fry, N., 1992. Stress ratio determinations from striated faults: a spherical plot for cases of near-vertical principal stress. Journal of Structural Geology 14, 1121–1131.

Fry, N., 1999. Striated faults: visual appreciation of their constraint on possible palaeostress tensors. Journal of Structural Geology 21, 7–22.

Gutscher, M.A., Kukowski, N., Malavieille, J., Lallemand, S., 1996. Cyclical behaviour of thrust wedges: insights from high basal friction sandbox experiments. Geology 24, 135–138.

Hardcastle, K.C., Hills, L.S., 1991. Brute3 and Select: QuickBasic 4 programs for determination of stress tensor configurations and separation of heterogeneous. Computer & Geosciences 17, 23–43.

Hartigan, J.A., 1985. Statistical theory in clustering. Journal of Classification 2, 63–76.

Huang, Q., 1988. Computer-based method to separate heterogeneous sets of fault-slip data into subsets. Journal of Structural Geology 10, 297–299.

Koyi, H., 1995. Mode of internal deformation in sand wedges. Journal of Structural Geology 17, 293–300.

Mitra, S., 1987. Regional variations in deformation mechanisms and structural styles in the central Appalachian orogenic belt. Geological Society of America Bulletin 98, 569–590.

Mojena, R., 1977. Hierarchical grouping methods and stopping rules: an evaluation. The Computer Journal 29, 359–363.

Molnar, P., Tapponnier, P., 1975. Cenozoic tectonics of Asia: effects of a continental collision. Science 189, 419–426.

Nemcok, M., Lisle, R.J., 1995. A stress inversion procedure for polyphase fault/slip data sets. Journal of Structural Geology 17, 1445–1453.

Nemcok, M., Kovac, D., Lisle, R.J., 1999. Stress inversion procedure for polyphase calcite twin and fault/slip data sets. Journal of Structural Geology 21, 597–611.

Price, R.A., 1988. The mechanical paradox of large thrusts. Geological Society of America Bulletin 100, 1898–1908.

Ramsay, J.G., Lisle, R.J., 2000. The Techniques of Modern Structural Geology Volume 3: Applications of Continuum Mechanics in Structural Geology. Academic Press, London. 346pp.

Shan, Y., Suen, H., Lin, G., 2003. Separation of polyphase fault/slip data: an objective-function algorithm based on hard division. Journal of Structural Geology 25, 829–840.

Shan, Y., Li, Z., Lin, G., 2004a. A stress inversion procedure for automatic recognition of polyphase fault/slip data sets. Journal of Structural Geology 16, 919–925.

Shan, Y., Lin, G., Li, Z., 2004b. A simple stress inversion of from fault/slip data assuming Andersonian stress state. Journal of Geophysical Research 109, B04408.

Simón-Gómez, J.L., 1986. Analysis of a gradual change in stress regime: example from the eastern Iberian Chain. Tectonophysics 124, 37–53.

Ward, J.H., 1963. Hierarchical evaluating clustering methods. Journal of American Statistics Association 58, 236–244.

Will, T.M., Powell, R., 1991. A robust approach to the calculation of paleostress fields from fault plane data. Journal of Structural Geology 13, 813–821.

Xie, F.R., Liu, G.X., 1989. Analysis of neotectonic stress field in area of the central segment of Altun fault zone, China. Earthquake Research in China 5, 26–36 (in Chinese with an English abstract).

Yamaji, A., 2000. The multiple inverse method: a new technique to separate stresses from heterogeneous fault-slip data. Journal of Structural Geology 22, 441–452.