Journal of Structural Geology 27 (2005) 855–858

JOURNAL OF STRUCTURAL GEOLOGY

www.elsevier.com/locate/jsg

Obtaining a best fitting plane through 3D georeferenced data

O. Fernández*

Grup de Geodinàmica i Anàlisi de Conques, Dept. de Geodinàmica i Geofísica, Universitat de Barcelona, 08028 Barcelona, Spain

Received 16 December 2003; received in revised form 1 October 2004; accepted 18 December 2004

Abstract

The orientation of a surface can be derived from the X, Y, Z position of three or more points lying on it. Two different methods are presented to obtain average surface orientations from points belonging to the surface. One consists of calculating a best-fit plane through a planar regression of data, which yields an average orientation for sets of more than three points. The second approach consists of analyzing the moment of inertia of the set of points to obtain the orientation of the best-fit plane and a measure of the spatial distribution of points. The quality of the orientation measurement depends strongly on the spatial distribution of points and can be evaluated with the use of eigenvalues.

Keywords: Best-fit plane; Outcrop trace; Bedding orientation; Moment of inertia analysis; Planar regression

1. Introduction

The three-point method to obtain the strike and dip of geological surfaces from 3D georeferenced data has been used by many authors (e.g. Berger et al., 1992; Bilotti et al., 2000) and is described in many handbooks (e.g. Groshong, 1999). This method has been applied classically on outcrop traces and can also be implemented on any set of 3D georeferenced data including depth-migrated seismic horizons or even isolated points.

The three-point approach consists of picking three points with known X, Y, Z coordinates and calculating the orientation of a plane fitting the three points. In doing so one assumes that the surface under study is planar and that the three points selected to define the surface attitude are representative of the surface.

The main drawback behind this approach is that the quality and reliability of the result depend heavily on the set of points selected to perform the calculations (e.g. Banerjee and Mitra, 2004). Frequently, more than only three points with known X, Y, Z coordinates are available for any one surface. For instance, digitized outcrop traces and seismic horizons have as many points as digitized nodes. Therefore, to obtain representative measurements it is preferable to use the largest amount of points in calculations, as long as they define a single planar surface.

Two methods can be used to obtain average orientation values that are statistically reliable by incorporating more than three points in calculations. One method consists of defining the surface orientation and position by calculating the best-fit plane for a set of points through a planar regression (briefly described by Banerjee and Mitra (2004)). The other approach consists of performing a moment of inertia analysis on the set of points. The moment of inertia analysis also provides parameters to evaluate the reliability of the orientation measurement.

2. Planar regression

The first approach to obtain an orientation from a set of points consists of finding a best-fit plane through a planar regression. A best-fit plane can be defined with the equation:

$$(x_i - \bar{x}) = B(y_i - \bar{y}) + C(z_i - \bar{z}) \tag{1}$$

where $\bar{x}$, $\bar{y}$, and $\bar{z}$ are the respective mean values of X, Y, and Z coordinates of all points. To find the equation of the best-fit plane for a given set of points, Press et al. (1986) present the following equations that have to be solved for B and C:

* Current address: Midland Valley Expl., 14 Park Circus, Glasgow G3 6AX, UK

E-mail address: oscar@mve.com.

0191-8141/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.jsg.2004.12.004

$$\sum((x_i - \bar{x})(y_i - \bar{y})) = B\sum(y_i - \bar{y})^2 + C\sum((y_i - \bar{y})(z_i - \bar{z})) \quad (2)$$

$$\sum((x_i - \bar{x})(z_i - \bar{z})) = B\sum((y_i - \bar{y})(z_i - \bar{z})) + C\sum(z_i - \bar{z})^2 \quad (3)$$

The result of the regression is a plane that passes through a point with coordinates ($\bar{x}$, $\bar{y}$, $\bar{z}$) and is returned in the form of a vector normal to the best-fit plane that can be converted to dip direction/dip notation.

The equations by Press et al. (1986) need to be corrected to deal with traces striking perfectly E–W by replacing Eq. (1) with the following definition:

$$(y_i - \bar{y}) = A(x_i - \bar{x}) + C(z_i - \bar{z}) \quad (4)$$

and modifying Eqs. (2) and (3) accordingly.

3. Moment of inertia analysis

The second approach consists of estimating the moment of inertia of a set of nodes and using the axis of maximum moment of inertia as the pole to the best-fit plane. This approach is equivalent to the process used in standard structural analysis to define girdle distributions (Woodcock, 1977; Davis, 2002). To estimate the moment of inertia of a set of points, it is assumed that the best-fit plane passes through a point whose coordinates are the average of the X, Y, and Z coordinates of the nodes (the nodes' center of mass ($\bar{x}$, $\bar{y}$, $\bar{z}$)). The vectors linking the center of mass with each node are calculated, and an orientation matrix **T** is constructed (as proposed by Woodcock, 1977):

$$\mathbf{T} = \begin{pmatrix} \sum l_i^2 & \sum l_i m_i & \sum l_i n_i \\ \sum m_i l_i & \sum m_i^2 & \sum m_i n_i \\ \sum n_i l_i & \sum n_i m_i & \sum n_i^2 \end{pmatrix} \quad (5)$$

where l_i, m_i, and n_i are the X, Y, and Z components of the individual vectors, without normalizing. Matrix **T** is a symmetrical matrix and can therefore be solved to obtain its eigenvalues (λ_1, λ_2, and λ_3) and eigenvectors (v_1, v_2, v_3). Eigenvalue λ_1 and eigenvector v_1 correspond to the orientation with maximum density of vectors. Eigenvalue λ_3 and eigenvector v_3 correspond to the orientation with the minimum density of vectors and maximum moment of inertia, and therefore the pole to the best-fit plane.

4. Evaluating quality of the orientation measurement

There are two concepts that define the quality of an orientation measurement obtained from a set of points: the degree of fit of the plane to the points, and the reliability of the measurement. The degree of fit of the calculated plane is inversely proportional to the distance of the nodes to the plane, whereas the reliability is a concept associated with the stability of the solution.

4.1. Degree of fit of a planar regression

For planes obtained through a planar regression, the degree of fit of the plane calculated is defined with the value of the correlation coefficient R_Z (the correlation coefficient of Z with respect to X and Y) (Press et al., 1986):

$$R_Z = \sqrt{1 - \frac{E}{D}} \quad (6)$$

where:

$$E = \frac{1}{n}\sum\left[z_i - \bar{z} - ((x_i - \bar{x}) - B(y_i - \bar{y}))/C\right]^2 \quad (7)$$

$$D = \frac{1}{n}\sum(z_i - \bar{z})^2 \quad (8)$$

and B and C are as in Eq. (1).

The values of R_X and R_Y can also be calculated to complement the information provided by R_Z.

An alternative measure of the fit of the plane to the source data can be provided by the mean distance ($\bar{dist}$) of individual points to the plane calculated as:

$$\bar{dist} = \frac{1}{n}\sum|P + (x_i + By_i + Cz_i)|/\sqrt{1 + B^2 + C^2} \quad (9)$$

where:

$$P = \bar{x} - B\bar{y} - C\bar{z} \quad (10)$$

This measure is particularly useful for surfaces with elevated dips, which yield misleading values of R_Z.

4.2. Degree of fit and reliability for moment of inertia analysis

For best-fit planes obtained with a moment of inertia analysis, the degree of fit and reliability of the best-fit plane can be defined by the ratio between the three eigenvalues of the orientation matrix, following the criteria proposed by Woodcock (1977).

The degree of fit can be defined by the ratio between λ_1 and λ_3:

$$M = \ln(\lambda_1/\lambda_3) \quad (11)$$

The larger this ratio, the more co-planar the nodes (Fig. 1), and thus the smaller the distance between the nodes and the best-fit plane.

However, a high degree of fit is not necessarily synonymous with good orientation quality, as the measurement is not necessarily reliable. For a set of points to be optimal to obtain a reliable orientation measurement, the nodes must be as far from co-linear as possible, and as

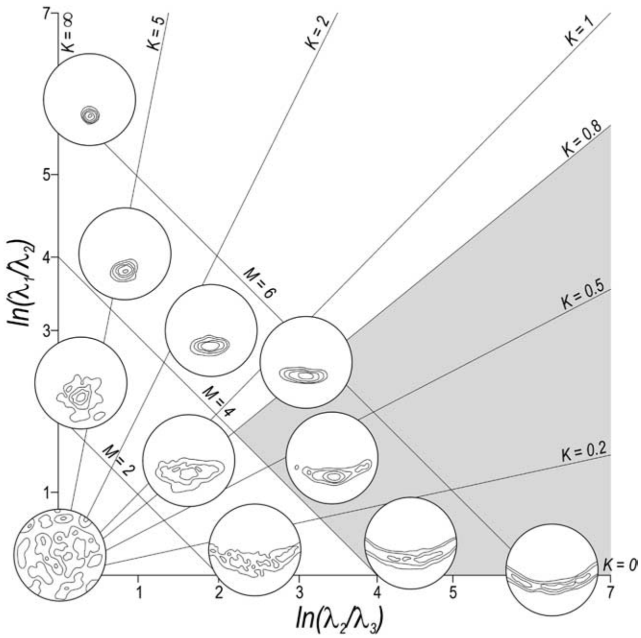


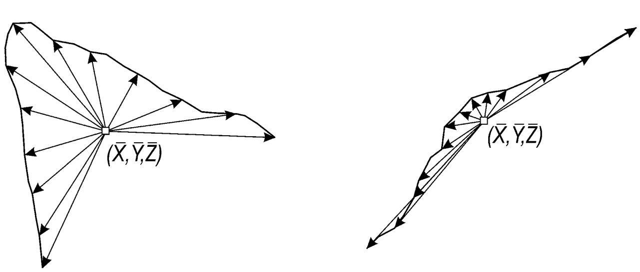
Fig. 1. Spatial distribution of point data according to the different ratios between eigenvalues. $M = \ln(\lambda_1/\lambda_3)$ and $K = \ln(\lambda_1/\lambda_2)/\ln(\lambda_2/\lambda_3)$. Sets of points with values falling in the gray shaded area yield best-fit planes with good degrees of fit and reliability. Modified from Woodcock (1977).

distant from the nodes' center of mass as possible (Fig. 2). For a set of highly co-linear nodes many best-fit planes can be defined with a similar degree of fit. An extreme case would be a perfectly linear set of points, through which infinite planes with different orientations can be defined, all with an equal degree of fit. On the other hand, for non-co-linear nodes, slight variations in the best-fit plane's orientation will cause significant differences in the degree of the fit. A best-fit plane's orientation will be more reliable if a slight change in the plane's orientation causes a significant decrease in the degree of the fit. Thus, the less co-linear the nodes, the more reliable the plane's orientation, irrespective of the degree of fit.

The advantage of the moment of inertia analysis over the planar regression is that it provides a quantitative measure of the shape of the trace, and therefore its reliability. As the vectors that are introduced in the orientation matrix are not normalized, those with a larger modulus weigh more in the orientation analysis. Therefore, even if vectors have a wide range of orientations, if there is a particular preferred orientation with a marked maximum of vector lengths, the distribution will approach a cluster (Fig. 2). The less co-linear the nodes, the more the distribution of vectors approaches a girdle distribution, and the better the reliability of the measure of the orientation of the best-fit plane (Fig. 2). The degree of co-linearity of a point (and thus the reliability of the best-fit plane) can be expressed as a ratio of the orientation matrix eigenvalues (Woodcock, 1977). For any orientation matrix, a value K can be defined such that (Woodcock, 1977):

$$K = \ln(\lambda_1/\lambda_2)/\ln(\lambda_2/\lambda_3) \tag{12}$$

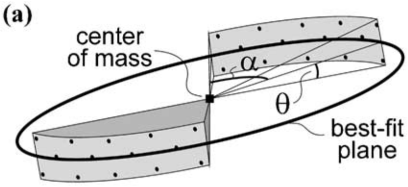
To quantify the relationship between the values of M and K and the quality of the orientation measurement a test was carried out on sets of points with different distributions (Fig. 3). Points were distributed around a center of mass, at a fixed distance (Fig. 3a). Their distribution was measured with angles θ and α (Fig. 3a). A moment of inertia analysis was performed on sets with varying values of θ and α to



(a) (b)

Fig. 2. The reliability of the orientation measured from an outcrop trace depends on the distribution of its nodes in space. A trace whose nodes are distributed evenly around its center of mass (a) will provide a more reliable measure of orientation than a trace with more co-linear nodes (b). Even if the orientations of vectors in (b) are widely distributed (over 180° in range), the longer vectors weigh more than the shorter ones, yielding a cluster distribution.

(a)



(b)

θ	*M*
1°	7.81
2°	6.42
3°	5.61
4°	5.03
5°	4.59
6°	4.22
7°	3.91
8°	3.64
9°	3.40
10°	3.18
15°	2.35

(c)

α	*K*
180°	3.1×10^{-4}
150°	2.9×10^{-2}
120°	0.10
90°	0.21
75°	0.28
60°	0.38
45°	0.52
30°	0.76
15°	1.33

Fig. 3. Test to define equivalences between eigenvalues and the spatial distribution of points. (a) A set of points or nodes distributed around a center of mass defines a best-fit plane. Points are contained in two symmetrical prisms, which can be defined by angles θ and α. (b) Table indicating the correlation between the value of angle θ and M if all points are equidistant from the center of mass and distributed homogeneously above and below the mean plane. The value of $\alpha = 180°$. (c) Table showing the correlation between the value of angle α and K if all points are equidistant from the center of mass. Points are distributed homogeneously above and below the mean plane with an angle of $\theta = 2°$.

derive two tables (Fig. 3b and c). Two threshold values are proposed for M and K according to their correlation with θ and α. For the fit of the best-fit plane to be good, values for M should be greater than four, corresponding to points that are distributed homogeneously within less than 6° of the best-fit plane ($\theta < 6°$) (Figs. 1 and 3b). Reliable orientation values are found to be provided by sets of points distributed along the best-fit plane over a range of more than 30° ($\alpha > 30°$) that yield values of K below 0.8 (Figs. 1 and 3c).

5. Conclusions

The use of planar regression and moment of inertia analysis algorithms are an effective manner of calculating representative values of surface orientation from sets of points. These tools can be applied recursively to different sets of points belonging to a same surface to analyze the spatial variation in surface attitude.

The eigenvalues derived from the moment of inertia analysis prove to be a powerful tool to estimate the quality of the orientations calculated by providing a quantitative measure of the spatial distribution of points.

Acknowledgements

Research by Oscar Fernández is funded by a pre-doctoral grant from the Direcció General de Recerca (Generalitat de Catalunya), by the Generalitat de Catalunya (Grup de Recerca de Geodinàmica i Anàlisi de Conques, 2001SGR-000074), and by the MCyT (Proyecto BTE2001-3650). The author also wishes to thank R. Groshong and T. Blenkinsop for helpful reviews and comments.

References

Banerjee, S., Mitra, S., 2004. Remote surface mapping using ortophotos and geologic maps draped over digital elevation models: application to the Sheep Mountain anticline, Wyoming. American Association of Petroleum Geologists Bulletin 88, 1227–1237.

Berger, Z., Williams, T.H.L., Anderson, D.W., 1992. Geologic stereo mapping of geologic structures with SPOT satellite data. American Association of Petroleum Geologists Bulletin 76, 101–120.

Bilotti, F., Shaw, J.H., Brennan, P.A., 2000. Quantitative structural analysis with spectroscopic remote sensing imagery. American Association of Petroleum Geologists Bulletin 84, 727–740.

Davis, J.C., 2002. Statistics and Data Analysis in Geology. John Wiley & Sons, New York.

Groshong Jr., R.H., 1999. 3-D Structural Geology: a Practical Guide to Surface and Subsurface Map Interpretation. Springer-Verlag, Berlin.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1986. Numerical Recipes. Cambridge University Press, Cambridge.

Woodcock, N.H., 1977. Specification of fabric shapes using an eigenvalue method. Geological Society of America Bulletin 88, 1231–1236.