Proceedings of the 8th WSEAS International Conference on APPLIED MATHEMATICS, Tenerife, Spain, December 16-18, 2005 (pp99-103)

Reliability Assessment for Vector Data Statistics

WANWU GUO
School of Computer and Information Science
Edith Cowan University
2 Bradford Street, Mount Lawley, WA 6050
AUSTRALIA

Abstract –In traditional vector data statistics, the reliability of the mean vector can not be quantified. In this paper, a quantified method is proposed to assess the reliability of the vector mean of a set of vectors. This method is based on the combination of the conventional vector statistics and the error analysis theory. The paper outlines the processes of the vector statistics and reliability assessment using this new method. The case studies show that this reliability assessment is coincident with the natural judgment, and can be used as a sufficient criterion to judge if the vector mean of a set of vectors is reliable.

Key-Words: - Vector data, Statistics, Coefficient of variation, Angular deviation, Vector mean reliability

1 Introduction

A vector is defined by both its magnitude and its orientation. In vector data statistics, traditionally, the vector sets are decomposed to three scalar components. Each component can then be analysed using the conventional statistical methods. In the end, the statistical means of the three components are combined together to determine the mean vector using its average magnitude and associated directional angles [1][2][3]. The reliability of this mean vector depends on two factors: the deviation of the magnitude and the dispersion in orientation. The former can be measured using the combination of standard deviations in the three axes, but the latter can not be quantified uniformly because the angular dispersion may vary with different axes.

In this paper, a quantified method is proposed to assess the reliability of the vector mean of a set of vectors. This method is based on the combination of the conventional vector statistics and the error analysis theory. The paper firstly introduces the fundamentals of vectors in different coordinating systems, and then outlines the processes for the vector statistics and reliability assessment using this new method. This new method is finally applied to statistics of magnetization data to demonstrate its usefulness.

2 Fundamentals of Vector Analysis

A vector can be expressed in different forms depending on its usage for different purposes. In rectangular coordinates, a vector $\boldsymbol{V}$ is defined by the three components (X, Y, and Z) in the x-axis, y-axis, and z-axis as

$$\boldsymbol{V} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}, \tag{1}$$

or

$$\begin{aligned} X &= V\cos(\mathbf{i},\boldsymbol{V}), \\ Y &= V\cos(\mathbf{j},\boldsymbol{V}), \\ Z &= V\cos(\mathbf{k},\boldsymbol{V}), \end{aligned} \tag{2}$$

where

$$V = \sqrt{(X^2 + Y^2 + Z^2)}\,. \tag{3}$$

In other words, a vector $\boldsymbol{V}$ can be defined by its magnitude V and three direction cosines between $\boldsymbol{V}$ and the three axes, respectively.

In spherical coordinates, a vector is defined by its magnitude ρ, polar angle θ and azimuthal angle φ. Relations between these two coordinates are summarised below:

$$\begin{aligned} \rho^2 &= V^2 = X^2 + Y^2 + Z^2 \\ X &= \rho\sin\varphi\cos\theta \\ Y &= \rho\sin\varphi\sin\theta, \\ Z &= \rho\cos\varphi. \end{aligned} \tag{4}$$

In Earth science, the coordinating system is set as x towards the north, y towards the east and z downwards for the rectangular coordinates, and the azimuthal angle starting from horizontal downwards as inclination (I), the polar angle starting horizontally from the north clockwise as declination (D) shown in Figure 1. In this system, the three components X, Y and Z of a vector is expressed as

Proceedings of the 8th WSEAS International Conference on APPLIED MATHEMATICS, Tenerife, Spain, December 16-18, 2005 (pp99-103)

$$X = V\cos I\cos D,\quad Y = V\cos I\sin D,\quad Z = V\sin I. \tag{5}$$

These three coordinating systems are mutually convertible so any system can be used to carry out data analysis. Since geomagnetism data will be used for testing the usefulness of the newly proposed method in this paper, the last coordinating system is chosen to outline the mathematical processes of this new approach.

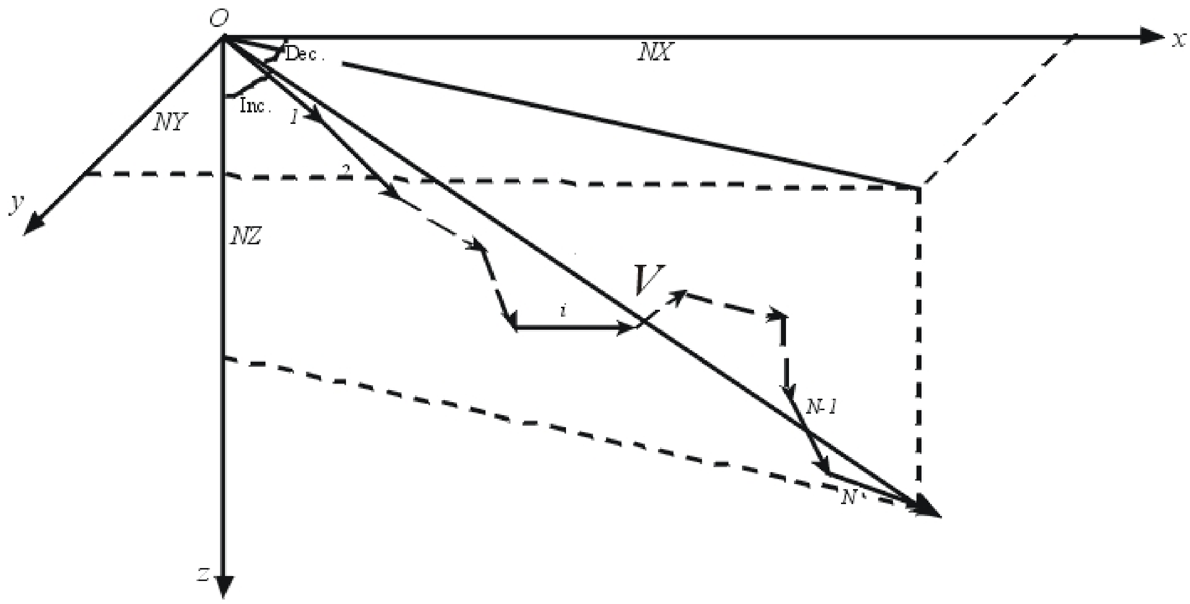

Figure 1. Illustration of coordinates used in Earth Science

3 Methodology

To calculate the vector mean of a set of N vectors, the three components X_i, Y_i and Z_i of any vector can be expressed by its declination, inclination, and magnitude as

$$X_i = V_i\cos I_i\cos D_i,\quad Y_i = V_i\cos I_i\sin D_i,\quad Z_i = V_i\sin I_i. \quad (i = 1, 2, \dots N) \tag{6}$$

The three components (X, Y, and Z) of the vector mean are determined by

$$X = \frac{\sum_{i=1}^{N} X_i}{N},\quad Y = \frac{\sum_{i=1}^{N} Y_i}{N},\quad Z = \frac{\sum_{i=1}^{N} Z_i}{N}. \tag{7}$$

The magnitude of this mean is given by

$$V^2 = X^2 + Y^2 + Z^2. \tag{8}$$

The declination and inclination of this mean can be derived from Relations (7) and (8) as

$$\tan D = \frac{Y}{X} \text{ and } \sin I = \frac{Z}{V}. \tag{9}$$

The standard deviations of the three components (σ_x, σ_y, σ_z) are given by

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{N}(j_i - \bar{j})^2}{N-1}},\ (j = X, Y, Z) \tag{10}$$

where $\bar{j}$ is the mean of one of the three components and given by

$$\bar{j} = \frac{1}{N}\sum_{i=1}^{N} j_i\ . \tag{11}$$

If x, y and z with standard deviations σ_x, σ_y and σ_z are used to compute the function $F(x,y,z)$, then the deviation or error caused in $F(x,y,z)$ is determined by [5]

$$\sigma = \sqrt{\left(\frac{\partial F}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial F}{\partial y}\sigma_y\right)^2 + \left(\frac{\partial F}{\partial z}\sigma_z\right)^2} . \qquad (12)$$

From Relations (8) and (12), the deviation (σ) of V can be determined by

$$\sigma = \frac{1}{V}\sqrt{(X\sigma_x)^2 + (Y\sigma_y)^2 + (Z\sigma_z)^2} . \qquad (13)$$

The ratio of the deviation to the mean of a data set is called the *relative uncertainty* in the error analysis theory [5]. It is a measure to estimate the precision of the mean and the extent of scatter of a set of data/measurements. The smaller the ratio, the more precise the mean and the more centred the data set.

This parameter is adopted in this study for estimating the precision of the vector mean and the extent of scatter of a set of vectors. It is named the *coefficient of variation* (C_v), i.e.,

$$C_v = \frac{\sigma}{V} = \tan\lambda , \qquad (14)$$

where λ is called the equivalent *angular deviation* of the mean shown in Figure 2. The mean vector is determined by its declination (D) and inclination (I), both possibly varying from $D - \lambda$ (or $I - \lambda$) to $D + \lambda$ (or $I + \lambda$), and its intensity (V) varying from $V(1 - C_v)$ to $V(1 + C_v)$. Therefore, C_v can be used as a measure to indicate the precision of the mean and the extent of scatter of a set of vectors.

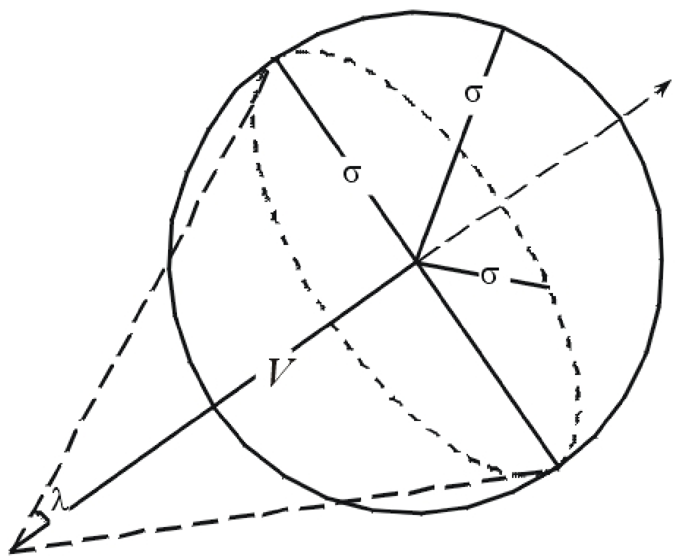

Figure 2. Schematic diagram illustrating the parameters of vector statistics

Simply using the coefficient of variation (C_v) and the angular deviation (λ) does not give a clear indication whether the mean vector is reliable in use. Hence, a new parameter, *reliability* (R_v), is defined as

$$R_v = (1 - C_v)\,\% = (1 - \tan\lambda)\%. \qquad (15)$$

Normally in data analysis, if the standard deviation is greater than its statistical mean, the statistical result becomes unreliable. This case refers to the situation where, for example, the average height of human beings is 1.70 m with a standard deviation of 2.00 m. This result does not make sense at all because it indicates that some people may be as high as 3.70 m (which is highly unlikely if impossible!) whereas others may be as tall as -0.3 m that is surely impossible. Therefore, the threshold of reliability is to be $C_v < 1$ or $\lambda < 45°$. On the other hand, the absolute reliability is indicated by $C_v = 0$ or $\lambda = 0°$ where all vectors in the data set are the same. By using these two natural benchmarks, a reliability list is given in Table 1. It shows that, for example, to achieve a reliability of 85% or higher, the statistical mean must satisfy $C_v \leq 0.15$ or $\lambda \leq 8.5°$.

Table 1. List of reliabilities for vector statistics

Reliability (R_v)	Coefficient of Variation (C_v)	Angular Deviation (λ)
95%	0.05	2.9°
90%	0.10	5.7°
85%	0.15	8.5°
80%	0.20	11.3°
75%	0.25	14.0°
70%	0.30	16.7°
65%	0.35	19.3°
60%	0.40	21.8°
55%	0.45	24.2°
50%	0.50	26.6°
0	1.00	45.0°

4 Case Studies

Naturally rocks on the Earth all carry magnetism more or less. Knowledge of magnetism carried in rocks is fundamental to palaeomagnetism [5], rock magnetism [6], petrophysics and exploration geophysics [7]. There are many magnetic properties in magnetism, but different subjects deal with some specific parameters of them. For example, palaeomagnetism focuses mainly on the orientations of remanent magnetisations carried in rocks, whereas both remanent and induced magnetisations carried in rocks are required in exploration geophysics. The two cases provided below deal with statistics of remanent magnetisations carried in rocks (Tables 2 and 3).

Table 2. Magnetisation data as vectors of case 1 [8]

Declination (D°)	Inclination (I°)	Magnetisation (M)
325.9	-12.0	888.1
319.2	-14.0	842.0
321.4	-19.1	872.9
330.6	-19.0	877.0
320.8	-17.1	875.6
326.3	-18.0	847.3
328.9	-17.2	882.2
293.5	-17.1	869.1
319.7	-19.7	803.5
326.9	-23.6	874.5
328.4	-23.1	903.9
332.8	-23.6	855.5
326.4	-12.7	905.0
318.8	-19.7	861.5
330.4	-11.3	880.6

There are 15 vectors in the first case that are listed in Table 2. Magnetisations are measured using a relative unit scale that is specific to the instrument used. The 15 vectors coherent closely in the direction of (320° and -20°) around 850 units, which should result in a mean vector with a higher reliability. This proves to be true as its Reliability is 88.8% shown in Table 4.

Table 3. Magnetisation data as vectors of case 2 [8]

Declination (D°)	Inclination (I°)	Magnetisation (M)
219.3	77.6	3823
296.8	77.9	1542
263.8	68.4	463
296.5	78.7	2257
275.4	79.1	2543
193.0	77.9	5388
184.5	73.7	5128
181.0	74.7	6051
201.1	73.1	4326
294.2	17.7	1098
258.4	61.7	1439
191.6	69.3	4524
191.1	73.4	5061
289.7	40.5	1429
278.6	32.7	2679
287.1	29.4	2273
280.6	48.2	1616
284.2	52.4	1124
261.1	57.9	1156
306.1	40.1	830

On the other hand, the 20 vectors listed in Table 3 show great variations in magnetisation (463 – 6051), declination (181° – 306.1°), and inclination (17.7° – 79.1°). A low reliability should be expected for the mean vector of this data set. This turns to be true as its Reliability is only 29.5% (Table 4).

Table 4. Vector means and their reliabilities of case studies

Case	N	D_{mean}	I_{mean}	M_{mean}	R_v	C_v	λ
Case 1	15	323.4°	-18.0°	857.5	88.8%	0.112	6.4°
Case 2	20	243.9°	73.4°	2521	29.5%	0.705	35.2°

5 Conclusion

By introducing the error analysis theory to the vector statistics, the reliability of a vector mean can be assessed quantitatively. This can not be measured in conventional vector statistics. The case studies proven that this reliability assessment is coincident with the natural judgment.

However, it should be noticed that this reliability assessment can only be used as the sufficient condition to judge if the vector mean of a set of vectors is reliable, i.e., a high reliability ensures that the vector mean is reliable, but a low reliability does not exclusively indicate that the mean is unreliable. This is because this reliability assessment applies the maximum constraint (σ) equally to all three variables of V, D, and I (or ρ, φ, and θ). In reality, the main contributor to σ may be only one of the three variables, which should affect that variable mostly.

References

[1] N.A.C. Cressie, *Statistics for spatial data*, Wiley, 1993.

[2] W.R. Dillon and M. Goldstein, *Multivariate analysis methods and applications*, Wiley, 1984.

[3] G.J.G. Upton and B. Fingleton, *Spatial data analysis by examples (Vol. 2): categorical and directional data*, Wiley, 1989.

[4] J.R. Taylor, *An introduction to error analysis*, University Science Books, 1997.

[5] R.F. Butler, *Paleomagnetism*, Blackwell Scientific Publications, 1992.

[6] D.J. Dunlop and O. Ozdemir, *Rock magnetism*, Cambridge University Press, 1997.

[7] J.H. Schön, *Physical properties of rocks: fundamentals and principles of petrophysics, Handbook of geophysical exploration – seismic exploration*, Pergamon, 1996.

[8] W. Guo, Magnetic petrophysics and density investigations of the Hamersley Province, Western Australia: implications for magnetic and gravity interpretation, PhD thesis, The University of Western Australia, 1999.