Geophys. J. Int. (1990) **103**, 725–729

Classification of the reversal test in palaeomagnetism

P. L. McFadden[1] and M. W. McElhinny[2]

[1]*Geophysical Observatories & Mapping Branch, Bureau of Mineral Resources, GPO Box 378, Canberra, ACT 2601, Australia*
[2]*Gondwana Consultants, 112 Sealand Road, Fishing Point, NSW 2283, Australia*

Accepted 1990 July 13. Received 1990 July 13; in original form 1990 May 4

SUMMARY

It is standard practice that a positive reversal test is claimed on the basis of inability to reject the hypothesis that two distributions share a common mean direction, and thus the claim of a positive reversal test is in fact often based on a lack of information. This is unsatisfactory. Therefore it is suggested that positive reversal tests should be classified according to the amount of information that was available for the test. This amount of information is readily indicated by the critical angle (e.g., at the 95 per cent confidence level) between the two sample mean directions at which the hypothesis of common mean direction for the distributions would be rejected. It is recommended that 5°, 10° and 20° be used as the breakpoints in the classification.

Key words: database, palaeomagnetism, reversal test, stability tests, statistics.

1 INTRODUCTION

In attempting to reconstruct the ancient history of the geomagnetic field it is important to establish, as far as is possible, that the palaeomagnetic directions deduced from rocks relate to a single component of magnetization, and are not contaminated by other components. One stability test used as an aid to establish this is the reversal test, which (in effect) tests whether two sets of observations (each 'observation' usually being a site-mean direction), one with normal polarity and the other with reverse polarity, could have been drawn from distributions with mean directions 180° apart. The basis of this test is that if the magnetizations are contaminated with another component then both the normal and the reverse directions would be shifted towards the direction of the contaminating component, and they would no longer be 180° apart (Cox & Doell 1960).

In practice, the test is of course performed by inverting the observations from one polarity and testing whether the two sets of observations could have been drawn from distributions sharing a common mean direction (throughout the rest of this paper it is therefore understood that one of the two samples has had its directions inverted). Naturally this involves forming the null hypothesis that the two sets of observations are in fact drawn from distributions that share a common mean direction and then determining whether the observations are inconsistent with this hypothesis at some level of confidence (usually 95 per cent). If they are not inconsistent at that level of confidence the null hypothesis is accepted, otherwise it is rejected in favour of the alternative hypothesis that the distributions do not share a common mean direction (i.e., the original distributions do not have mean directions 180° apart).

If only a small amount of information is available (e.g. only a few observations or observations drawn from distributions with low precision), then almost inevitably the null hypothesis cannot be rejected, irrespective of the validity of this hypothesis. The unfortunate aspect of this with regard to the reversal test is that this is then claimed as being a 'positive' reversal test. Compilation of a global palaeomagnetic database (see Van der Voo & McElhinny 1989; McElhinny & Lock 1990a,b) has highlighted the fact that in many instances the claim of a positive reversal test is in fact based on just such a lack of information. In contrast, a negative reversal test indicates the availability of a reasonable amount of information. Clearly this is an unsatisfactory situation and there needs to be a simple classification indicating the amount of information actually available in the claim of a positive reversal test.

2 PROPOSED CLASSIFICATION

A simple, yet very effective, measure of the information available is the angle γ_c between the mean directions of the two sets of observations at which the null hypothesis of a common mean direction would be rejected with 95 per cent confidence, given the observed dispersion in the two samples. We propose that a 'positive' reversal test be classified as 'A' if $\gamma_c \leq 5°$, as 'B' if $5° < \gamma_c \leq 10°$, as 'C' if $10° \leq \gamma_c \leq 20°$, and as 'INDETERMINATE' if $\gamma_c > 20°$.

The attribute 'R' is used in the database to indicate that a reversal test has been performed. Thus it is proposed that the attributes 'Ra', 'Rb', 'Rc' and 'Ro' be used to indicate 'positive' reversal tests with classifications 'A', 'B', 'C', and 'INDETERMINATE' respectively, and that the attribute 'R-' be used to indicate a negative reversal test.

3 PERFORMING THE TEST AND CLASSIFICATION

The actual test used to determine whether the two samples could have been drawn from distributions sharing a common mean direction depends on the number of observations in each sample, and on whether the two distributions share a common precision. Throughout, it is assumed that the distributions are Fisherian (Fisher 1953). If there are several observations in one sample but only a single observation in the other sample, then a test should be performed on the isolated (single-sample) observation (see e.g., McFadden 1990) to see if it is discordant with the observations from the other set. If there are five or more observations in both sets and the hypothesis of a common precision cannot be rejected then the test may be performed as set out by McFadden & Lowes (1981). If there are fewer than five (but more than one) observations in either sample then the likelihood is that there is insufficient information available to reject the hypothesis of a common precision even when the precisions are substantially different. Thus it is prudent to revert to simulation without the assumption of a common precision (see e.g. McFadden 1990). If the precisions are different, it suggests that some problems may still exist in isolating a single component of magnetization in at least one of the polarity samples. However, this need not necessarily invalidate a reversal test, and (as in the previous situation) it is possible to use simulation without the assumption of a common precision.

3.1 Notation

Let

$$X_i = x_{i1} + \cdots + x_{iN_i},$$
$$Y_i = y_{i1} + \cdots + y_{iN_i}, \qquad (1)$$
$$Z_i = z_{i1} + \cdots + z_{iN_i},$$

where $i = 1$ or 2 (i.e., the two polarities) and (x_{ij}, y_{ij}, z_{ij}) are the direction cosines of the jth observation from the ith polarity. The length of the vector sum of the N_i unit vectors for each polarity is then

$$R_i = (X_i^2 + Y_i^2 + Z_i^2)^{1/2}. \qquad (2)$$

The length of the vector sum of all $N = N_1 + N_2$ unit vectors is

$$R = [(X_1 + X_2)^2 + (Y_1 + Y_2)^2 + (Z_1 + Z_2)^2]^{1/2}. \qquad (3)$$

or, equivalently, if γ is the angle between the two mean directions then

$$R^2 = R_1^2 + R_2^2 + 2R_1R_2 \cos\gamma. \qquad (4)$$

Thus if R_c is the critical value of R associated with the critical angle γ_c, then

$$R_c^2 = R_1^2 + R_2^2 + 2R_1R_2 \cos\gamma_c. \qquad (5)$$

Three other statistics that will be needed are the estimates

$$k_i = \frac{N_i - 1}{N_i - R_i} \qquad (6)$$

of the distribution precisions κ_i, the weighted sum

$$S_r = k_1R_1 + k_2R_2 \qquad (7)$$

of resultant lengths, and a weighted overall resultant length

$$R_w = (\hat{X}^2 + \hat{Y}^2 + \hat{Z}^2)^{1/2}, \qquad (8)$$

where

$$\hat{X} = k_1X_1 + k_2X_2,$$
$$\hat{Y} = k_1Y_1 + k_2Y_2, \qquad (9)$$
$$\hat{Z} = k_1Z_1 + k_2Z_2.$$

An alternative representation for R_w is

$$R_w^2 = (k_1R_1)^2 + (k_2R_2)^2 + 2k_1R_1k_2R_2 \cos\gamma. \qquad (10)$$

3.2 Test on an isolated observation

If only a single observation (i.e. an isolated observation) is available for the one polarity, choose the indices so that $N_2 = R_2 = 1$. Naturally it is not possible to estimate κ_2, but if it is assumed that the distributions share a common precision (and it is not possible to test this assumption) then, under the null hypothesis of a common mean direction, the isolated observation should be just another random observation from the same distribution as the N_1 observations. Thus, as with a fold test (McFadden 1990), it is simply a matter of testing whether the isolated observation is discordant with the other observations. It follows from McFadden (1982, equation 14) that

$$\cos\gamma_c = 1 - \frac{(R_1 + 1)(N_1 - R_1)}{R_1}\left[\left(\frac{1}{p}\right)^{1/N_1 - 1} - 1\right] \qquad (11)$$

with $p = 0.05$. If the observed angle γ_o exceeds γ_c then the hypothesis of a common mean direction may be rejected at the 95 per cent confidence level. Otherwise the reversal test is considered to be positive, but is classified according to the angle γ_c.

It must of course be recognized that this test is weak, and every attempt should be made to obtain more observations. The major problem is that all of the information available for classification comes from the one polarity; no information comes from the polarity with only one observation. Thus, to indicate this lack of information from the one polarity, it is proposed that an 'I' (for isolated) be appended to the classifications as 'AI', 'BI', and 'CI', and that the attribute be written in the database as 'Rai', 'Rbi', or 'Rci'.

3.3 Test with several observations per polarity and a common precision

Watson (1956) has shown that

$$2\kappa(N_i - R_i) \ni \chi^2_{2(N_i - 1)}, \qquad (12)$$

where the symbol '$\ni$' is to be read as 'is distributed as' and χ^2_ν is the chi-square distribution on ν degrees of freedom.

Thus

$$\frac{k_1}{k_2} \ni F[2(N_2-1), 2(N_1-1)], \tag{13}$$

where $F[\nu, \mu]$ is the F distribution on ν and μ degrees of freedom. Traditionally the test has been performed by choosing the subscripts so that $k_1 > k_2$ and accepting the null hypothesis of a common κ if k_1/k_2 does not exceed $F_{0.05}$ (the value that the relevant F-distributed variable will exceed with probability 0.05). This is in fact a two-tailed test at the 90 per cent confidence level, but in this application it is probably sensible since the reversal test itself can easily be performed by simulation if there is any question about a common κ.

With the assumption of a common κ, it follows from McFadden & Lowes (1981, equation 29) that

$$\frac{R_1+R_2-(R_c^2/R_1+R_2)}{2(N-R_1-R_2)} = \left(\frac{1}{p}\right)^{1/N-2} - 1, \tag{14}$$

with $p=0.05$. Using equation (5) this gives the critical angle between the two mean directions as

$$\cos\gamma_c = 1 - \frac{(N-R_1-R_2)(R_1+R_2)}{R_1R_2}\left[\left(\frac{1}{p}\right)^{1/N-2} - 1\right]. \tag{15}$$

If the observed angle γ_o exceeds γ_c then the hypothesis of a common mean direction may be rejected at the 95 per cent confidence level. Otherwise the reversal test is considered to be positive, but is classified according to the angle γ_c.

It should be noted that if $N_2 = R_2 = 1$ then equation (15) reduces to equation (11).

3.4 Test with several observations per polarity using simulation

Watson (1983) suggested

$$V = 2(S_r - R_w) \tag{16}$$

as a test statistic that may be used regardless of the populations sharing a common precision. As a point of interest, if $k_1 = k_2 = k$ then

$$V = 2k(R_1 + R_2 - R). \tag{17}$$

Clearly V is zero if the site-mean directions are the same, and increases with increasing dispersion of the site-mean directions. Thus the null hypothesis of a common mean direction may be rejected if V is too large. Unfortunately, V does not have a convenient distribution in small samples (Watson 1984), but it is a simple matter to simulate the distribution of V (see McFadden 1990) as follows.

(1) Calculate the observed value V_0 of V.

(2) Assuming Fisher distributions, under the null hypothesis of a common mean direction, simulate a new set of observations (x_{ij}, y_{ij}, z_{ij}) by choosing $\tau_{\theta ij}$ and $\tau_{\varphi ij}$ from a set of pseudo-random numbers uniformly distributed in the interval [0, 1] and then calculating

$$\Lambda_{ij} = \frac{-\ln[\tau_{\theta ij}(1-e^{-2k_i}) + e^{-2k_i}]}{2k_i},$$

$$\theta_{ij} = 2 \arcsin\sqrt{\Lambda_{ij}}$$

[there are several similar equations for generating pseudo-random variates from a Fisher distribution, but this particular form gives the best performance, see Fisher, Lewis & Willcox (1981)],

$$\varphi_{ij} = 2\pi\tau_{\varphi ij},$$

and

$$x_{ij} = \cos\theta_{ij}\cos\varphi_{ij},$$

$$y_{ij} = \cos\theta_{ij}\sin\varphi_{ij},$$

$$z_{ij} = \sin\theta_{ij}.$$

(3) Calculate the simulated value V_1 of V.

(4) Repeat steps (2) and (3) to obtain 1000 simulated values $V_1, \ldots, V_{1000}$.

(5) Order the simulated values $V_1, \ldots, V_{1000}$ from smallest to largest as $v_1, \ldots, v_{1000}$ (in practice this ordering is performed by creating a linked list as the values are simulated).

(6) v_λ is then the critical value for V_o at the $100(1-p)$ per cent level, where λ is the largest integer not exceeding $[1000(1-p)+1]$. As before, $p=0.05$ for the 95 per cent confidence level.

It follows from equation (16) that the critical value R_{wc} of R_w is given by

$$R_{wc} = S_r - \frac{V_\lambda}{2}, \tag{18}$$

and from equation (10) that

$$\cos\gamma_c = \frac{R_{wc}^2 - (k_1R_1)^2 - (k_2R_2)^2}{2k_1R_1k_2R_2}. \tag{19}$$

Again, if the observed angle γ_o exceeds γ_c then the hypothesis of a common mean direction may be rejected at the 95 per cent confidence level. Otherwise the reversal test is considered to be positive, but is classified according to the angle γ_c.

4 USE OF SUFFICIENT STATISTICS

In order to perform each of the tests, all that is required is N_1, R_1, N_2, R_2, and the angle γ_o between the two mean directions. Thus it is a simple matter to perform any of the tests even if only the mean directions, the numbers of observations, and the sufficient statistics R_1 and R_2 are available. Similarly, if only the mean directions, the numbers of observations, and either of the sufficient statistics k_i or $(\alpha_{95})_i$ is available for each polarity, then the tests can be performed by inverting either equation (6) or

$$\cos(\alpha_{95})_i = 1 - \frac{N_i - R_i}{R_i}\left[\left(\frac{1}{p}\right)^{1/N_i-1} - 1\right] \tag{20}$$

to obtain R_i. This is particularly useful if the observations have been obtained from several rock types having different within-site precisions and an analysis has been performed to determine the between-site precisions; it is then preferable to perform the test using the estimate of the between-site precision.

5 APPLICATION OF TEST

Application of the suggested classification to 535 results from the literature indicates that it provides a realistic method of appraising the reversal test. The data were in fact extracted from the global palaeomagnetic database (McElhinny & Lock 1990a,b) by searching for all those entries on which it was possible to perform a reversal test. There are 2444 results in the database, which covers all data available in the literature for the years from approximately 1975/76 to 1988.

A histogram of the antipodal angles between the two polarities is shown in Fig. 1, the width of each cell being 5°. If a particular test is to receive the classification 'A', then the supplement of its antipodal angle must be less than the critical angle at the 95 per cent level of confidence (in order to pass the test), and this critical angle must not exceed 5° (in order to classify the available information as 'A'). Thus a prerequisite for the 'A' classification is that the antipodal angle exceeds 175°. Similarly, a prerequisite for the 'B' classification is that the antipodal angle lies between 170° and 175°, and for the 'C' classification is that the antipodal angle lies between 160° and 170°. This maximum possible classification is indicated in Fig. 1.

Naturally one first has to decide which test to use, based on the available observations. If one of the polarities has only one observation then the isolated-observation test must be used, otherwise one of the multiple-observation tests is used. If there are multiple observations for each polarity then the hypothesis of a common kappa is tested. If this hypothesis is rejected then the alternative hypothesis of different kappas is accepted and the 'Different Kappa' test is performed (using the simulation given in Section 3.4). If the null hypothesis of a common kappa has not been rejected, but either of the polarities has fewer than five observations then a 'Simulation' test (Section 3.4) is performed anyway. Finally, if the hypothesis of a common kappa has not been rejected and each polarity has at least five observations then the 'Distribution' test (Section 3.3) is performed (on the assumption that the distribution of the test statistic is known). The classifications resulting from these tests are shown in Table 1 and Fig. 2.

Table 1. Reversal test classifications for data from the global palaeomagnetic database.

Classification	Type of test: Isolated observation	Multiple observations for each polarity: Common Kappa: Distribution	Common Kappa: Simulation	Different Kappa	Total
F	14	42	26	39	121
I	96	10	82	18	206
C	19	58	43	33	153
B	2	26	8	12	48
A	0	4	1	2	7
Total	131	140	160	104	535

With the isolated-observation test we know *a priori* that there is very little information available, so it is no surprise to see that the results are totally dominated by 'INDETERMINATEs' and that there are no 'A' classifications. With the 'Distribution' test, there are at least 10 observations with five for each polarity, so there is a reasonable amount of information. The domination by 'C' and 'F' classifications is therefore as expected. With the 'Simulation' test, at least one of the polarities has fewer than five observations, so there is not a lot of information (but more than in the isolated-observation tests). Again the results are dominated by 'INDETERMINATEs', but not as strongly as for the isolated-observation tests. With the 'Different Kappa' test, there is sufficient information to reject the null hypothesis of a common kappa, and, again as expected, the results are dominated by 'C' and 'F' classifications.

A histogram of the classifications, irrespective of

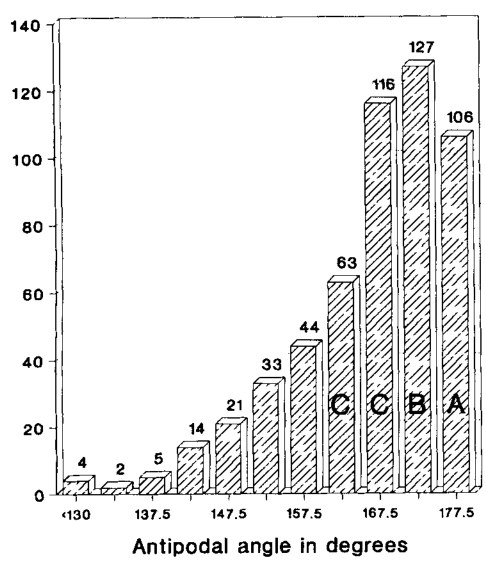

Figure 1. Histogram of antipodal angles for data extracted from the global database.

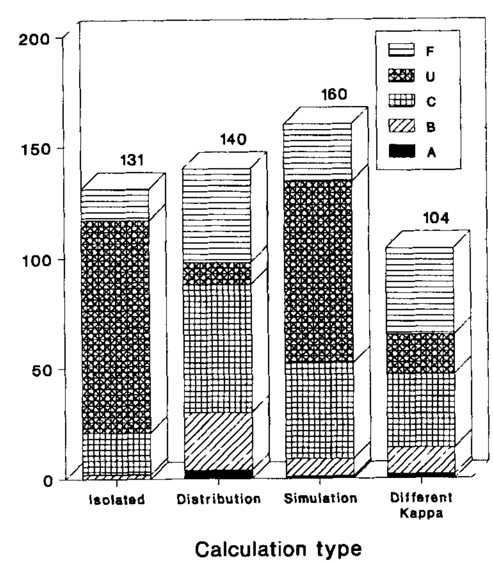

Figure 2. Histogram of the reversal test classifications, according to calculation type, for data from the global database.

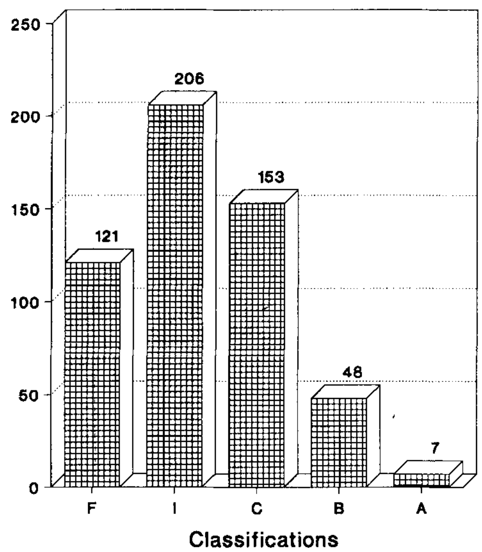

Figure 3. Histogram of the reversal test classifications for data from the global database.

calculation type, is shown in Fig. 3, and a comparison of this with Fig. 1 clearly illustrates the value of performing the classification. For example, of the 94 results for which the antipodal angle was within 5° of the ideal 180°, in only seven cases was the information sufficient to classify the test as 'A'.

6 AVAILABILITY OF SOFTWARE

A program for IBM-compatibles to perform the tests and classification in this paper may be obtained from BMR via PLM.

ACKNOWLEDGMENT

This paper is published with the permission of the Director, Bureau of Mineral Resources, Geology & Geophysics.

REFERENCES

Cox, A. & Doell, R. R., 1960. Review of paleomagnetism, *Geol. Soc. Am. Bull.*, **71**, 647–768.

Fisher, R. A., 1953. Dispersion on a sphere, *Proc. R. Soc. Lond.* A, **217**, 295–305.

Fisher, N. I., Lewis, T. & Willcox, M. E., 1981. Tests of discordancy for samples from Fisher's distribution on the sphere, *Appl. Stat.*, **30**, 230–237.

McElhinny, M. W. & Lock, J., 1990a. IAGA global palaeomagnetic database, *Geophys. J. Int.*, **101**, 763–766.

McElhinny, M. W. & Lock, J., 1990b. The global palaeomagnetic database project, *Phys. Earth planet. Inter.*, in press.

McFadden, P. L., 1982. Rejection of palaeomagnetic observations, *Earth planet Sci. Lett.*, **61**, 392–395.

McFadden, P. L., 1990. A new fold test for palaeomagnetic studies, *Geophys. J. Int.*, **103**, 163–169.

McFadden, P. L. & Lowes, F. J., 1981. The discrimination of mean directions drawn from Fisher distributions, *Geophys. J. R. astr. Soc.*, **67**, 19–33.

Van der Voo, R. & McElhinny, M. W., 1989. The GP Compass, *EOS, Trans. Am. geophys. Un.*, **31**, 748, 758.

Watson, G. S., 1956. Analysis of dispersion on a sphere, *Mon. Not. R. astr. Soc., Geophys. Suppl.*, **7**, 153–159.

Watson, G. S., 1983. Large sample theory of the Langevin distributions, *J. stat. Planning Inference*, **8**, 245–256.

Watson, G. S., 1984. The theory of concentrated Langevin distributions, *J. Multivar. Anal.*, **14**, 74–82.