Available online at www.sciencedirect.com
SCIENCE DIRECT
EPSL
Earth and Planetary Science Letters 6786 (2003) 1–14
www.elsevier.com/locate/epsl

Non-dipole fields and inclination bias: insights from a random walk analysis

Joseph G. Meert *, Endale Tamrat, John Spearman

Department of Geological Sciences, University of Florida, 241 Williamson Hall, Gainesville, FL 32611, USA

Received 25 April 2003; received in revised form 9 July 2003; accepted 18 July 2003

Abstract

The paleomagnetic assumption that the Earth's magnetic field is reduced to a geocentric axial dipole (GAD) when sufficiently sampled has been called into question for Mesozoic and earlier times. It has been suggested, for example, that modest contributions from axial quadrupolar (10%) and octupolar (25%) fields are resolvable using inclination only data from paleomagnetic studies. The underlying assumption in inclination only studies is that considerable continental drift has occurred over a sufficiently long period of time to render paleomagnetic sampling random in a paleogeographic sense. This assumption was stated in all previous studies dating back to 1976, but was never tested. We have developed a random walk model designed to test this assumption. Our model uses three different configurations for the continents in the random walk and allows the user to vary parameters such as maximum velocity, sampling distribution, sampling frequency and frequency of directional change. The model generates large sample sizes that cannot be adequately evaluated using the standard χ^2 statistical test and therefore we introduce two statistical parameters used in structural equation models. Our models indicate that the 'random paleogeographic sampling' assumption used in the previous studies is not valid due primarily to the lack of an adequate sample size and temporal distribution. We show, for example, that even the most robust dataset compiled in 1998 is severely undersampled. A series of model runs on a GAD earth with sampling over a 600 Myr period demonstrates that detailed sampling will, on average, produce a GAD-like distribution only 30% of the time. Other model runs demonstrate that inadequate sampling can produce false quadrupolar and octupolar effects. It is our conclusion that time-averaged inclination only studies using the extant paleomagnetic database should be viewed with extreme caution.

Keywords: geocentric axial dipole; octupole; quadrupole; paleomagnetism; inclination

1. Introduction

One of the fundamental working assumptions in paleomagnetic studies is that the Earth's magnetic field averages to a geocentric axial dipole (GAD) when sufficiently sampled. While the assumption seems to hold, within error, for the past 5 million years [1,2], it has been recently questioned for pre-Mesozoic times [3–5]. It has been suggested, for example, that contributions from axial quadrupolar (10%) and octupolar (25%)

* Corresponding author.
E-mail address: jmeert@geology.ufl.edu (J.G. Meert).

0012-821X/03/$ – see front matter © 2003 Published by Elsevier B.V.
doi:10.1016/S0012-821X(03)00417-5

fields are resolvable using inclination only data from paleomagnetic studies on Precambrian and Paleozoic datasets [3]. Others [4,5] have analyzed perceived misfits between continental reconstructions and paleomagnetic data and suggested a modest contribution from a persistent (e.g. 10%) octupolar field during Paleozoic and Mesozoic time. Because persistent non-dipole fields can cause errors in paleomagnetically based reconstructions, it is important to develop several reliable independent tests for the GAD assumption.

Evans [6] proposed one such test that relies solely on the frequency of inclination data through time. He argued that a ∼600 Myr random cycling of continents around the globe should produce a distribution of magnetic inclinations that follows a probability function based on the surface area occupied by the sampling sites over time. This probability function is easily solved by estimating the surface area of the globe corresponding to any set of inclination classes (typically 10° intervals). In an effort to minimize sample bias problems, the inclination data are binned in 10°×10° grids for a given time interval [3,6]. The surface area of the globe corresponding to each inclination class can be calculated by:

$$\partial\Omega = 2\pi\int_{\phi_1}^{\phi_2}\sin\phi\,\partial\phi \qquad (1)$$

where ϕ is the co-latitudinal range corresponding to each inclination class. Assuming a GAD field, Eq. 1 should produce the inclination frequency distribution shown in Fig. 1. The dependence of the absolute value of the inclination $|I|$ on co-latitude (ϕ) for axial multipoles is given by:

$$\tan I_l = \frac{-(1+l)P_l}{(\partial P_l/\partial\phi)} \qquad (2)$$

where P_l is the Legendre polynomial of degree l and I_l is the resulting inclination from the axial multipole of degree l. The solution to Eq. 2 for axial quadrupolar ($G2 = g_0^2/g_1^0$) and octupolar ($G3 = g_0^3/g_1^0$) fields is given by [6] as:

$$\tan I' = \frac{2\cos\phi + 1.5\times G2(3\cos^2\phi - 1) + 2\times G3(5\cos^3\phi - 3\cos\phi)}{\sin\phi + G2(3\sin\phi\cos\phi) + 1.5\times G3(5\sin\phi\cos^2\phi - \sin\phi)} \qquad (3)$$

Eq. 3 can be used to compare observed inclination frequencies with expected inclination frequencies based on various contributions of non-dipole fields (Fig. 1). The original analysis found that inclination only data were compatible with the GAD hypothesis at least for the last 600 million years [6]. Subsequent studies [3,7] reached different conclusions. Piper and Grant [7], using unbinned inclination data, concluded that the observed inclination frequencies were compatible with the GAD assumption for the time interval from 600 to 3000 Ma, but were significantly different than GAD for the time interval from 300 to 600 Ma. Kent and Smethurst [3], using binned data, concluded that the GAD hypothesis could not be rejected for the Mesozoic and Cenozoic, but the observed distributions were significantly different (at the 95% confidence level) from GAD during the Paleozoic and Precambrian.

There were a number of explanations offered for this departure from GAD. For example, [3,5,8] suggested that stabilization of the inner core, via growth, might stabilize the geodynamo and result in diminishing contributions from higher order harmonics through time. Inclination shallowing was also proposed to explain the apparent bias towards low inclinations [3,5]. This explanation appears less likely because the inclination bias is also observed in igneous rocks [3,5]. Other explanations include preferential cycling of continents to low latitudes via true polar wander [3,7]; the indiscriminate use of poorly resolved paleomagnetic data [9] or that the underlying assumption of random sampling of the globe is invalid [9].

Evans [6], in his original analysis of inclination only data, asserted that the last 500–600 million years of continental drift rendered the paleomagnetic sampling random in a paleogeographic sense. Subsequent studies [3,7] made a similar ‘random sampling’ assertion, but provided no evidence to support the assumption. In addition, both [3] and [7] made a fundamental error in their calculation of the χ^2 value. In this paper, we examine the underlying assumption of random sampling and provide a more sensitive test of significance for the observed distributions.

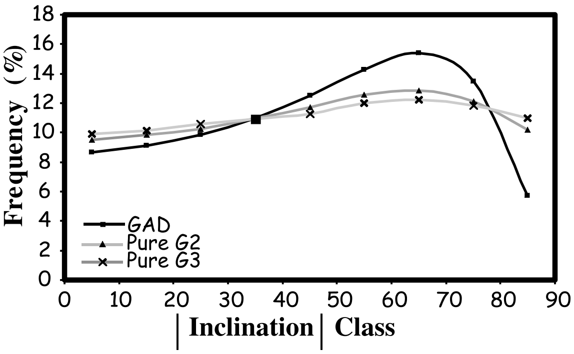

Fig. 1. Theoretical inclination class frequency curves for a GAD field, a pure G2 (quadrupole) field and a pure G3 (octupole field).

2. Previous studies

Table 1 shows the results of previous studies on inclination only data from the paleomagnetic database. Each of these studies applied the χ^2 test (see Appendix 1a) to the observed and theoretical distributions. We note, however, that the studies of [3,7] incorrectly applied the χ^2 test resulting in an underestimate of the χ^2 value (see Table 1 for corrected values). This underestimate significantly affects the conclusions of [7] regarding the Phanerozoic and Precambrian datasets. The original paper concluded that the observed distributions were indistinguishable from GAD; however, the corrected χ^2 values indicate that the two distributions are significantly different from GAD above the 99% confidence interval (99% $\chi^2_{crit} = 20.09$). In the case of [3], the corrected χ^2 values render the Mesozoic distribution different than the GAD above the 99% confidence interval (Table 1). The other χ^2 values cited in [3] also increase, but the conclusions remain identical to the previous study, namely that the Paleozoic and Precambrian data are significantly non-GAD.

We were initially interested in extracting the temporal development of the non-GAD field from the paleomagnetic inclination database. Therefore, we developed an algorithm that can take the measured inclination distribution and invert Eq. 3 to generate a family of 'best-fit' solutions based on varying contributions of the G2 (quadrupolar) and G3 (octupolar) fields and the calculated χ^2 value. Thus, if the assumption of random sampling has been met and the χ^2 test statistic is a valid measure of goodness of fit, then these families of G2 and G3 contributions might provide information regarding the evolutionary development of the magnetic field through time. Table 1 lists the results of this analysis for the previous studies. Although the relative contributions of G2 and G3 vary over time according to the analysis conducted by [3], there appears to be no systematic change evident in their relative contributions on a temporal scale. This led us to seek

Table 1
Previous results

Period	Bins or observations	Reference	χ^2_1	χ^2_2	N_{crit}	RMSEA	G2	G3
Cambrian–Tertiary	430	[6]	13.49	13.49	494.2	0.07	NC	NC
Phanerozoic (0–600 Ma)	3888[a]	[7]	4.61	293.9	206.1	0.10	0.21	0.18
Precambrian (600–3000 Ma)	899[a]	[7]	8.32	136.4	103.1	0.15	0.16	0.22
Cenozoic (0–65 Ma)	253	[3]	3.63	9.19	426.3	0.07	NC	NC
Mesozoic (65–250 Ma)	342	[3]	7.18	24.50	216.9	0.10	0.28[b]	0.14[b]
Paleozoic (250–550 Ma)	352	[3]	32.23	113.48	48.9	0.21	0.11	0.28
Precambrian (550–3500 Ma)	531	[3]	20.39	108.76	76.6	0.17	0.14	0.23
All (0–3500 Ma)	1478	[3]	–	135.75	169.8	0.11	0.16	0.17
Phanerozoic	947	[3]	–	54.33	270.8	0.091	0.18	0.14
Mesozoic+Cenozoic	595	[3]	–	28.72	321	0.083	0.31	0.16

χ^2_1, as calculated by the original authors; χ^2_2 as calculated by Meert et al. (this study).
N_{crit} = critical N-index or Hoelter index; RMSEA = root mean square error of approximation; G2 and G3 are best-fit calculation to the observed binned distribution; NC = not calculated since the results are indistinguishable from GAD.
[a] Not binned.
[b] Best fit is still significantly different than both GAD and the observed distribution at the 95% confidence level using the χ^2 test.

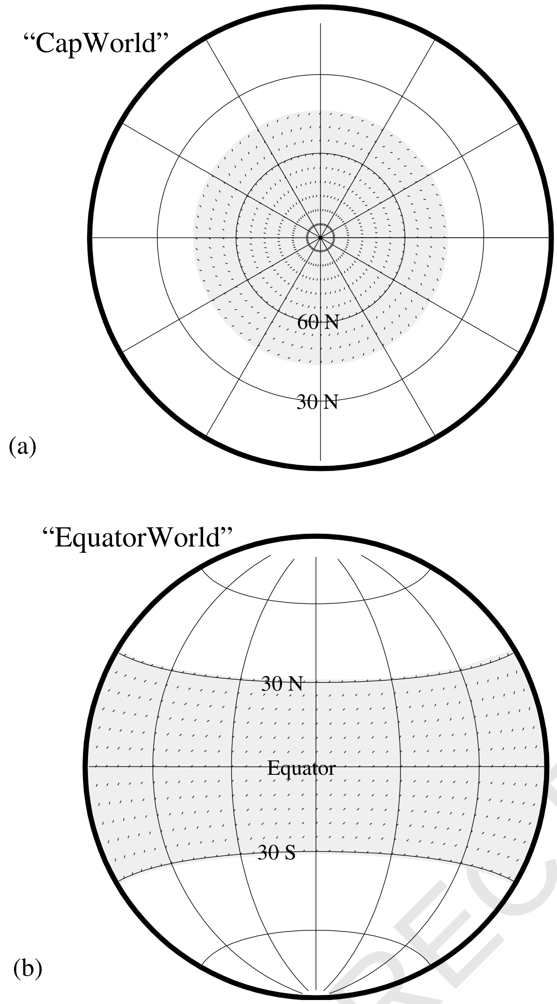

Fig. 2. (a) ‘Cap world’ continent representing ~11% of the Earth’s surface area. Sampling sites (dots) are located at 5° intervals. (b) ‘Equator world’ continent representing ~25% of the Earth’s surface area. Sampling sites (dots) are located at 5° intervals.

an alternative explanation for the non-GAD distributions cited by [3].

3. A random walk model

The basic assumption inherent in each of the three previous inclination only studies [3,6,7] was that 500–600 million years is enough time to randomly sample the globe. This assumption has never been tested. We formulated an algorithm for generating a continental-size random walk on a sphere. In an effort to accurately simulate continental drift, we have generated different continental configurations. One configuration is based on the present-day geography of the earth (major continents only) with ‘sampling’ locations positioned at 5° intervals. This is equivalent to binning the observations as suggested by [3,6] using 5° bins instead of 10° bins. Other configurations used in this analysis were an ‘equator world’ starting configuration (Fig. 2b) and a ‘cap world’ starting configuration (Fig. 2a). In order to simulate both the rotational and translational components of motion, each continent is rotated about a computer-generated random Euler pole (see Appendix 1b). Since the maximum velocity of any point on the continent is dependent on its distance from the Euler pole, maximum drift velocities were constrained to between 2 and 8 cm/yr via a random number generator. This method of controlling velocity has the practical effect of producing continental motion with variable drift rates. We also analyzed the major directional changes evident in the Phanerozoic apparent polar wander path for Laurentia. The path indicates a major directional change approximately every 75 Myr on average. For most of our models, we used a 75 Myr time period as the interval of constant motion. The algorithm thus generated a new Euler pole and velocity maximum every 75 Myr. Each of these parameters is user controlled. Lastly, the algorithm allows the user to ‘sample’ the GAD field at a given frequency. Most of our runs collected an inclination sample at each site for every 25–50 million years. The total duration of each run lasted 575 Myr, 600 Myr or 3500 Myr. The computer program calculates the χ^2 test statistic, the Hoelter index (or N_{crit} value) and the root mean square error of approximation (or RMSEA; see Appendix 1a) after each sampling period. We consider a good fit to the theoretical distribution reached when two of the three statistical measures fall within the critical values ([11,14]; χ^2 value < 15.51; RMSEA < 0.05; $N_{crit} > 200$).

We ran a number of sensitivity tests on the model as shown in Fig. 3. In general, the models are relatively insensitive to speed (S) and sampling

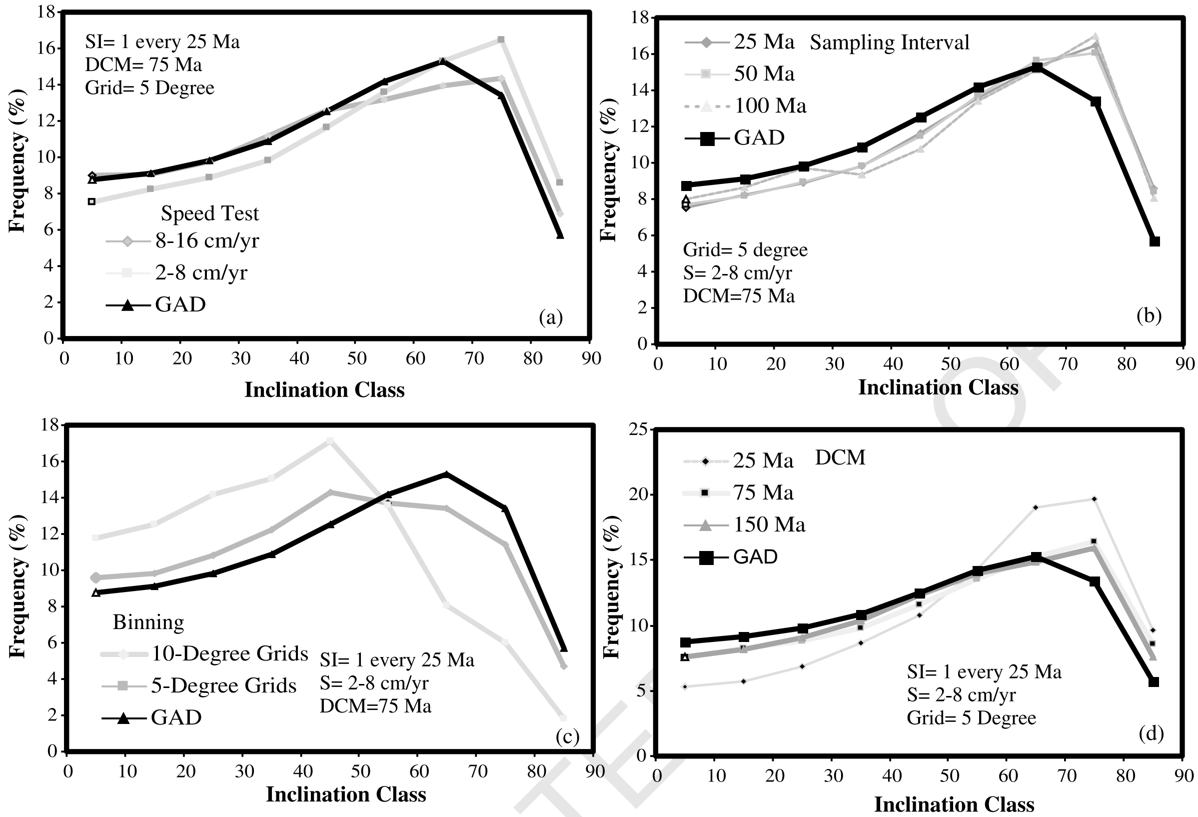

Fig. 3. Model sensitivity tests showing (a) a 575 Myr run of 'today world'. Each of these runs started with the same random seed value. (a) Sensitivity of the model to changing drift rates. After 575 Myr, the 'faster' drift rates results in a slightly improved fit, but the two curves are not statistically different. (b) A 575 Myr run of 'today world' using different sampling intervals (1 sample every 25 Myr, 1 sample every 50 Myr or 1 sample every 100 Myr). The 1/25 and 1/50 Myr runs are almost identical. The 1/100 Myr sampling interval results in only a slightly more uneven distribution of inclination values. (c) A 575 Myr run of 'equator world' using a 5°×5° grid and a 10°×10° grid. The 10°×10° grid results in a significant loss of signal with respect to the 5°×5° grid. (d) A 575 Myr run of 'today world' demonstrating the sensitivity of the model to changing the direction of plate motion. The model appears to be more sensitive to rapid directional changes (e.g. 1 every 25 Myr) as compared to slower plate motion changes (e.g. 1 every 75–150 Myr).

interval (SI; within the ranges shown in Fig. 3a,b) and most sensitive to the gridding interval (Fig. 3c) and the duration of constant motion (DCM; Fig. 3d). It is possible to choose extreme values for any of these parameters and drastically change the results; however, we have attempted to run the model within the confines of reasonable geodynamic conditions as described above.

Since it is difficult to illustrate 3500 million years of drift on the globe, we chose to illustrate the mechanics of the model using a synthetic drift history of Australia over a 600 million year period (Fig. 4A,B). The model begins with Australia in its present-day position. In this run, the assigned model parameters included maximum drift rates of 2–8 cm/yr, sample collection every 25 million years and directional change every 75 Myr. The results of this particular model run show that Australia was restricted to latitudes between 30°N and 30°S. Most of the sampled units should therefore generate inclinations of less than 50°. The observed frequency distribution is shown in Fig. 4B in agreement with the modeled drift history.

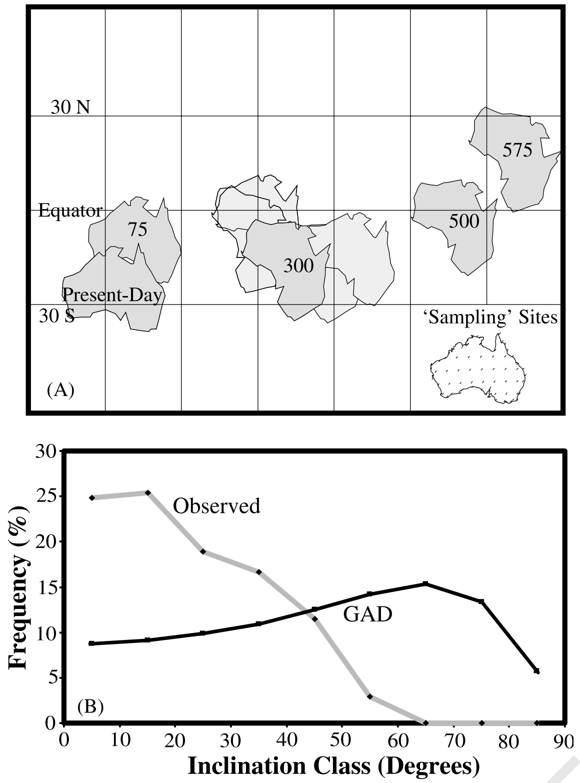

Fig. 4. (A) Synthetic drift history of Australia during a 575 Myr model run. The starting point was the present-day position of Australia. The inset shows the sampling locations within Australia. (B) Observed inclination class–frequency diagram based on the drift in A.

4. Results

4.1. Cap world

This model was run using a continental spherical cap. The SI was 25 Ma, S varied between 2 and 8 cm/yr and DCM was 75 Myr. This model, representing 11.7% of the surface area on the Earth, illustrates several key observations regarding the random drift assumption and the sensitivity of the χ^2 test statistic. The statistical results of this run are shown in Fig. 5. Fig. 5a shows the value for each of the three test statistics after each 25 Myr sampling interval. Although N_{crit} rises above 200 after 875 Ma, it is not until 1475 Ma that both the RMSEA and N_{crit} reach critical values (Fig. 5b, Table 2). The 'best fit' is achieved after 2050 Myr.

4.2. Equator world

This model was run using a hemispheric equatorial continent. The SI, S and directional change DCM were equivalent to the cap world model described above. The model represents 25% of the Earth and thus serves as a reasonable proxy for the current land surface of the Earth. The statistical results of this run are shown in Fig. 6. In this case, a 'good fit' is first reached at 600 Myr (Fig. 6a), and much better statistical fits were achieved at 750 Myr and again at 1200 Myr. The latter two results fell within the critical intervals for all three statistical measures (Fig. 6a, Table 2). The model run at 750 Myr shows the effect of calculating the Hoelter index when the χ^2 test statistic is not significant. In this case, N_{crit} has a value exceeding the number of samples in the population (Table 2).

4.3. Today world

This model uses the present-day distribution of the major continents as a starting point. The SI, S and DCM were equivalent to the cap world and equator world models described above. The statistical results of this run are shown in Fig. 7 and listed in Table 2. The model reached a good fit after 725 Myr and a best fit at 1125 Myr.

4.4. Runs at 600 Myr

Evans [6] argued that 500–600 Myr of continental motion is sufficient to produce a random geographic sampling of paleomagnetic data. In order to test this assertion, we conducted numerous tests of the random walk covering 600 Myr (see Table 2). The models showed a highly variable response under identical sampling conditions. In these runs, the sampling interval was taken as 1 sample/25 Myr, the speed was allowed to vary between maxima of 2 and 8 cm/yr and a directional change took place every 75 Myr. Fig. 8a,b shows the results for cap world. In this run, the model never reaches a statistically valid fit to the GAD model (Fig. 8a) and the 'best fit', obtained after 600 Myr shows a high inclination bias (Fig. 8b, Table 2).

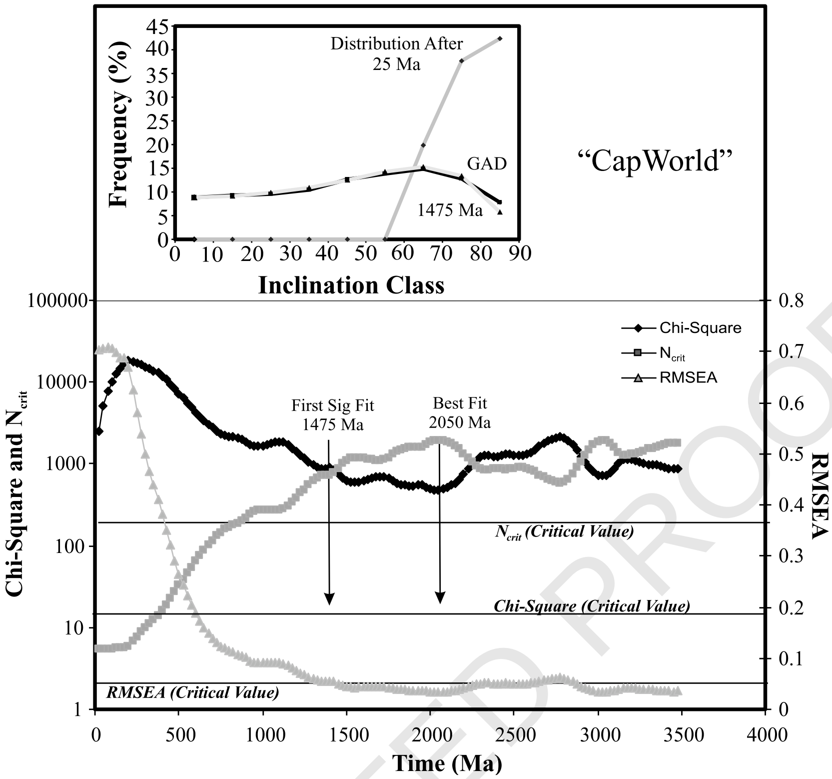

Fig. 5. Statistical results of a 3500 Myr run of 'cap world'. In this run, the model reaches a GAD fit after 1475 Myr and a best fit to the GAD model at 3025 Myr. The critical value for each statistical index is indicated by a labeled line. (Inset) The inclination class–frequency curves for 'cap world' after 25 Myr and 1475 Myr compared to the expected GAD frequency.

Table 2
Results from selected random walks

Run	Time (Ma)	χ^2 value	N_{crit}	RMSEA	Number of samples
Cap world-3500	1 475	668	1 004	0.0471	43 070
Cap world-3500	2 050	479	1 944	0.0338	59 860
Cap world-575	**600**	**2 077**	**153**	**0.137**	**17 520**
Equator world-3500	600	144	931	0.0476	8 640
Equator world-3500	*750*	*14.2*	*11 796*	*0.0134*	*10 800*
Equator world-3500	1 200	18.2	14 693	0.012	17 280
Equator world-575	600	660	743	0.062	8 640
Today world-3500	725	486	943	0.0485	29 522
Today world-3500	1 125	39.9	17 807	0.011	45 810
Today world-575a	**575**	**3 403**	**108**	**0.144**	**23 414**
Today world-575b	*325*	*11.5 < /C >*	*18 940*	*0.011*	*14 044*

Number of samples is equivalent to the number of 5° bins sampled in the modeled run. Bold indicates that no significant result was achieved in the run. Italics indicates all three measures of significance were achieved in the run.

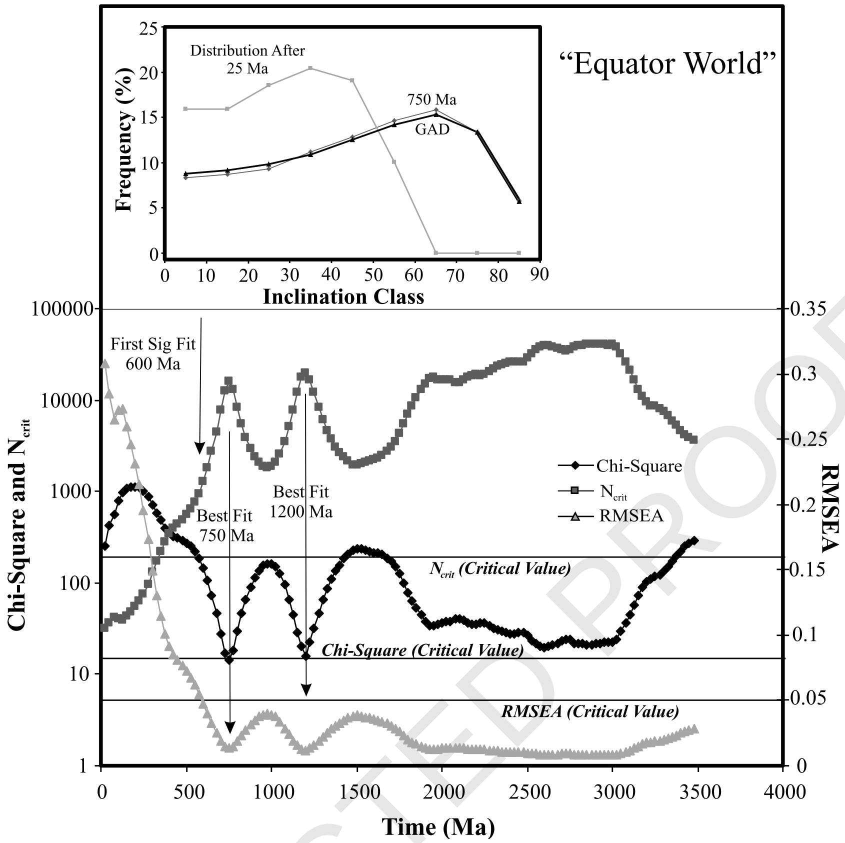

Fig. 6. Statistical results of a 3500 Myr run of 'equator world'. In this run, the model reaches a GAD fit after 600 Myr and a best fit to the GAD model at 750 and 1200 Myr. The critical value for each statistical index is indicated by a labeled line. (Inset) The inclination class–frequency curves for 'cap world' after 25 Myr and 750 Myr compared to the expected GAD frequency.

Fig. 8c,d shows the results of a 600 Myr run using the equator world starting configuration. The conditions were identical to the cap world run discussed above. The model does not generate a statistically valid fit to the GAD model (Fig. 8c) and the 'best fit' obtained after 600 Myr shows a low inclination bias (Fig. 8d, Table 2).

Fig. 9 shows two runs of 'today world' using identical conditions to those cited above. The only difference between the two runs was the choice of the initial random seed value and hence Euler pole sequence (see Appendix 1b). In the first example, the model does not reach a statistically valid fit after 600 Myr (Fig. 9a) and the best fit shows a high inclination bias (Fig. 9b, Table 2). In the second run, the model reaches a statistically valid fit after only 325 Myr (Fig. 9c,d; Table 2).

We note that in the case of the 600 Myr runs, the models appear sensitive to the starting conditions in that the endpoint distribution at 600 Myr most often reflects the initial bias. For example, both 'cap world' and 'today world' tend to show a bias towards high inclinations consistent with their initial biases towards high inclinations (see Figs. 5b and 7b). Similarly, 'equator world' tends to exhibit a low-latitude bias consistent with its starting distribution (Fig. 8c). However, this is not always the case. For example, Fig. 10a shows a cap world run most closely resembling a pure g_2^0 or pure g_3^0 contribution. Fig. 10b shows a run of today world with a low inclination bias best fit to a 6% +G3 contribution and Fig. 10c shows an equator world run with a high inclination bias.

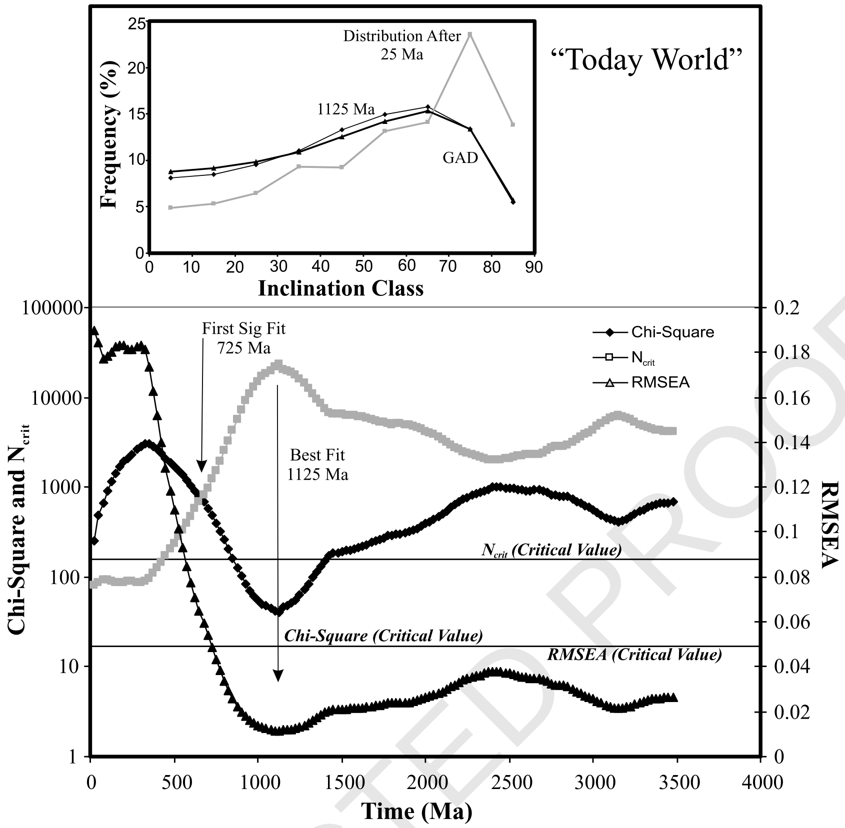

Fig. 7. Statistical results of a 3500 Myr run of 'today world'. In this run, the model reaches a GAD fit after 725 Myr and a best fit to the GAD model at 1125 Myr. The critical value for each statistical index is indicated by a labeled line. (Inset) The inclination class–frequency curves for 'cap world' after 25 Myr and 1000 Myr compared to the expected GAD frequency.

4.5. Small sample sizes

The previously cited analyses generate much larger sample sizes than are available in the paleomagnetic database. We wished to examine whether or not a much smaller random sample size would be capable of testing the GAD hypothesis. For example, in the analysis conducted by [3], the Cenozoic database ($n = 253$ bins) yielded an inclination distribution that is statistically indistinguishable from GAD as did the original study of the Phanerozoic by [6]. In the first case, the Cenozoic compilation of [3] represents the effects of distributed sampling sites across the globe rather than a randomization process resulting from continental motion. In the second case [6], the compiled database represents results through 1971. A majority of those results were not subject to detailed demagnetization such that the influence of younger overprints produced additional bias in that analysis. The maximum bin size for Mesozoic and Paleozoic times can be calculated using the 10 time periods (Cambrian–Cretaceous) multiplied by the number of $10° \times 10°$ bins. This calculation shows that [3] sampled only $\sim$11% of the potential bins. We therefore created a set of synthetic continents covering 11% of the Earth and conducted sampling at 25 Myr intervals for 485 Myr in order to mimic the Mesozoic–Paleozoic conditions of [3]. The model never approached a GAD fit in 50+ runs, but showed a wide variety of inclination bias.

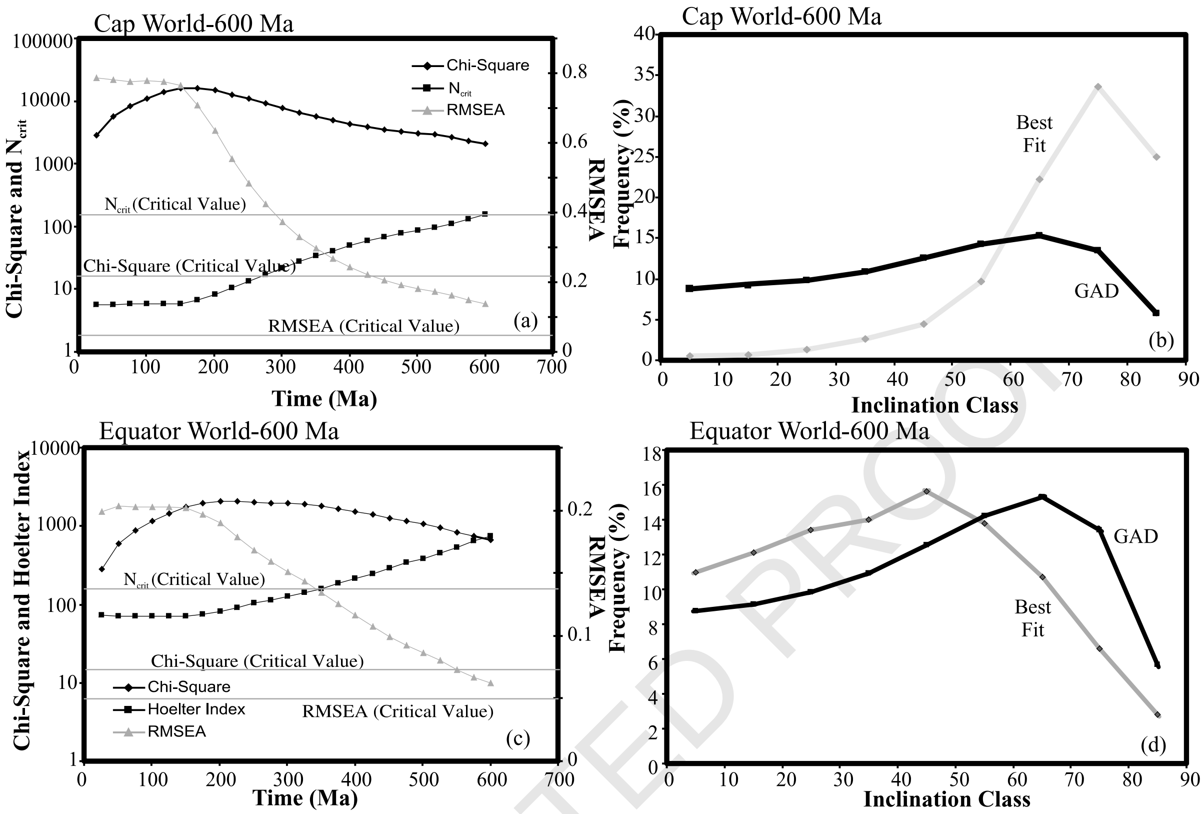

Fig. 8. (a) Statistical results of a 600 Myr run of 'cap world'. In this run, the model never reaches a statistically significant GAD-like fit. (b) The best fit reached after 600 Myr shows a high inclination bias for the 'cap world' run. (c) Statistical results of a 600 Myr run of 'equator world'. In this run, the model never reaches a statistically significant GAD-like fit. (d) The best fit reached after 600 Myr shows a low inclination bias for the 'equator world' run.

5. Discussion

Of the myriad explanations given for the low inclination bias observed in previous studies of inclination only data [3,7], the possibility of incomplete sampling has not been rigorously explored. Our random walk model, discussed above, suggests that the previous inclination only studies are fundamentally flawed. Further, we suggest that all conclusions drawn from these earlier studies should be viewed with extreme caution. We offer two principal reasons for rejecting this type of study. In an effort to determine the minimum number of studies necessary to confidently test the dipole nature of the field using inclination only studies, we ran the 'today world' model through 1000 iterations. The model parameters were identical to those described in Section 3.4. In the 600 Myr time period, the model produced inclination distributions that were statistically indistinguishable from the GAD 29.6% of the time. To first order, this means that a 600 Myr sampling interval has less than a 30% chance of adequately testing the GAD model when properly sampled. Furthermore, Fig. 11 shows a histogram based on the number of bins needed to achieve a GAD-like fit. The minimum number of sampling bins needed to achieve a GAD-like fit (5×5°) was 5500. The average was 15 270 bins. These numbers are particularly striking when compared to previous studies. Kent and Smethurst [3] used 531 bins to evaluate the Precambrian (550–3500 Ma). The bins were based on 10°×10° regions and 50 Myr intervals. If we consider that continental regions composed, on average, 25% of the surface area of the Earth during the Precambrian [16],

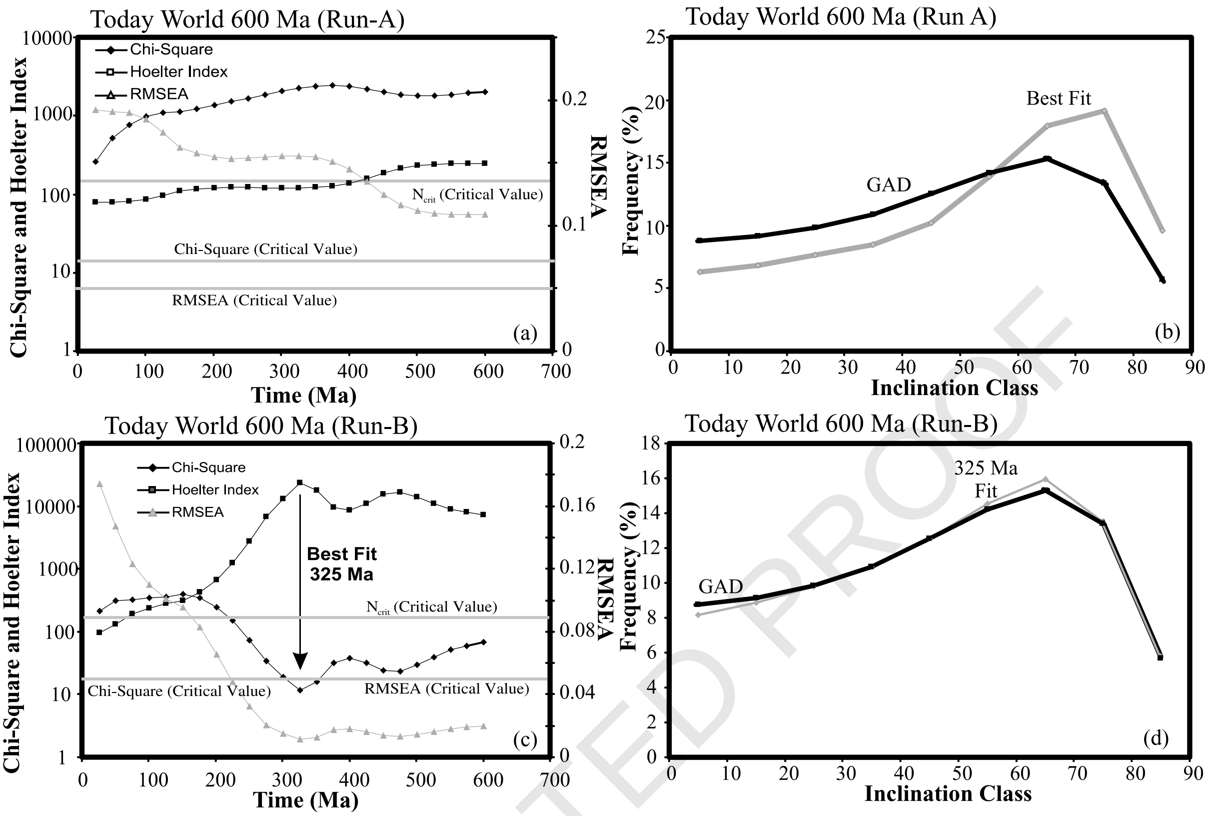

Fig. 9. (a) Statistical results of a 600 Myr run of 'today world'. In this run, the model never reaches a statistically significant GAD-like fit. (b) The best fit reached after 600 Myr shows a high inclination bias for the 'today world' run. (c) Statistical results of a 600 Myr run of 'today world' using a different random seed. In this run, the model reaches a GAD-like fit after 225 Myr and a best fit after 325 Myr. (d) Comparison of the 325 Myr inclination–frequency curve to the GAD curve for the 'today world' run in panel c.

then the theoretical maximum sample size would total 10 450 bins. Thus, the study by [3] would optimistically have sampled only ~5% of the potential bins and therefore is unlikely to provide a good test of the GAD hypothesis. The Paleozoic dataset of [3] contained 352 bins averaged over 300 Myr. When we ran our synthetic drift model (using 12 216 bins) at 300 Myr intervals, we produced a GAD-like result less than 8% of the time. The entire Phanerozoic data from [3] are more robust, with 947 total bins. Nevertheless, we estimate that this represents only ~13% of the potential sample size and therefore is unlikely to provide a strong test of the GAD hypothesis due to both the relatively small sample size and the short interval of time available for sampling.

It is more difficult to statistically evaluate the approach taken by [7] since our model is based on the premise of binned data. However, we note that when the corrected χ^2 values were calculated along with the N_{crit} and RMSEA statistical parameters, the results were all statistically different from the GAD model. In essence, [7] was operating with a slightly smaller dataset than [3]. Given that we have shown the analysis of [3] is faulty, it is difficult to imagine a scenario whereby an unbinned analysis is a superior model.

6. Conclusions

Previous studies that have attempted to look at inclination only distributions to test the GAD hypothesis were based on the assumption that 600

Myr was an adequate time period to produce a random sampling of the globe. That assumption was never tested and recent studies [3] suggested that the Paleozoic and Precambrian were dominated by modest contributions from persistent G2 (quadrupolar) and G3 (octupolar) fields. We tested the 'random paleogeographic sampling' assumption using a random walk model under a variety of conditions and configurations. Our results indicate that the sampling density required to adequately test the GAD model is not attainable using the extant paleomagnetic database. Our random walk models show that many non-GAD-like distributions can be produced via insufficient sampling. Thus, previous conclusions regarding persistent non-dipole fields based on inclination only studies should be viewed with extreme caution. Unfortunately, the model cannot a priori test whether or not such non-dipole fields existed in the past. Indeed, other modes of analysis have made the case for persistent non-dipole fields in Paleozoic–Mesozoic time [4,5] and these and other methods of GAD analysis should be pursued.

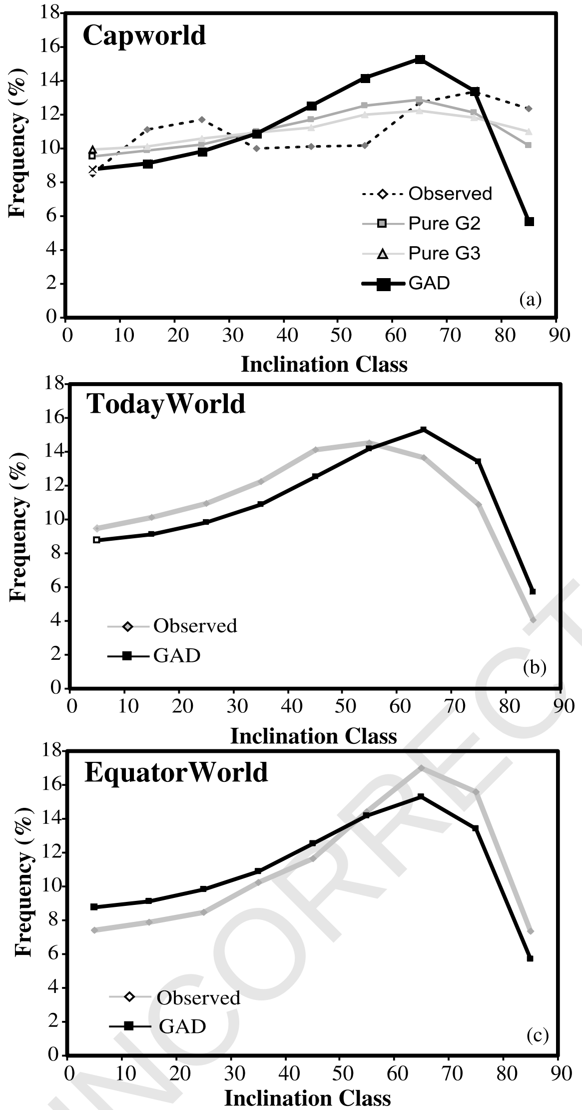

Fig. 10. (a) Inclination class–frequency curve from a 575 Myr run of 'cap world'. In this run, the statistical best fit after 575 Myr resembles either a pure G2 or pure G3 curve. (b) Inclination class–frequency curve from a 575 Myr run of 'today world'. In this run, the statistical best fit after 575 Myr yields a low inclination bias best fit to a 6% G3 (octupolar) field. (c) Inclination class–frequency curve from a 575 Myr run of 'equator world'. In this run, the statistical best fit after 575 Myr shows a high inclination bias.

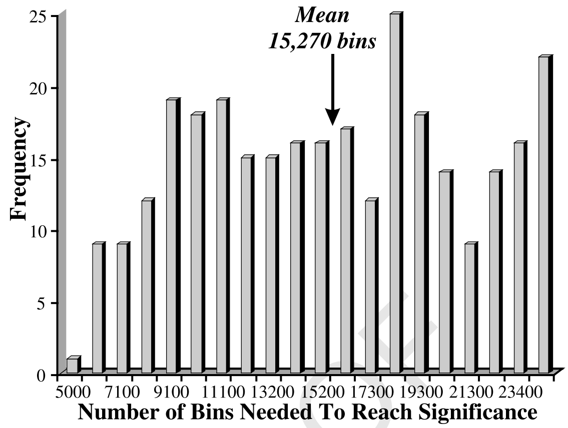

Fig. 11. Histogram representing the significant results of 1000 runs of 'today world' as sample bin size. The duration of each of these runs was 600 Myr using the parameters cited in the text. In this test, the model reached a GAD-like distribution 29.6% of the time. The average number of bins needed to reach GAD was 15 270 (5°×5° bins).

Acknowledgements

The authors wish to thank Dennis Kent for supplying us with his binned dataset and Larry Winner (University of Florida Statistics) for interesting discussions on the random walk model and the statistical analysis described in this paper. Trond Torsvik and Mike McElhinny are thanked for their extremely positive reviews of this paper. We also thank an anonymous reviewer for valuable comments.***[SK]***

Appendix 1a. Statistical tests

The χ^2 test compares the difference between the observed distributions and the theoretical GAD distribution and generates a χ^2 value. This value is then compared to the critical χ^2 value at the desired level of significance. A value of χ^2 which is less than the critical value signifies that the observed distribution cannot be distinguished from the theoretical distribution [10]. The formula for the χ^2 test is given as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \text{(A)}$$

where O_i = observed frequency in counts and E_i is the expected frequency in counts. The analysis conducted by [3,7] mistakenly used frequency in percent which has the effect of underestimating the χ^2 value by a factor of:

$$O_i/100 \quad \text{(B)}$$

Corrected χ^2 values are given in Table 1. Unfortunately, the χ^2 statistic is extremely sensitive to the number of observations. In a practical sense, this means that distributions based on a large N will almost always cause the χ^2 value to exceed the critical value even when the distribution functions are nearly identical (see Fig. 5b; Table 2). Statisticians have therefore developed a number of additional statistical tests that are designed to account for the sample size used in the χ^2 analysis [11,12].

The Hoelter index [11] is used to evaluate the observed distribution when the χ^2 value is deemed significant. This critical N index is sensitive to the number of observations used in the analysis. The Index is calculated as follows:

$$N_{\mathrm{crit}} = \frac{(N-1) \times \chi^2_{\mathrm{crit}}}{\chi^2} + 1 \quad \text{(C)}$$

where N = number of observations, χ^2_{crit} is the critical value of the χ^2 analysis and χ^2 is the calculated χ^2 value from Eq. A above. N_{crit} is therefore the maximum sample size at which the χ^2 value would not be significant. Hoelter [11] recommends values of at least 200 and states that values of less than 75 indicate very poor model fit. As originally formulated, the Hoelter index should only be used to evaluate models where the χ^2 values are significant. However, we note that when the analysis is conducted on a sample distribution where the χ^2 value is not significant then N_{crit} exceeds the true N used in the analysis. Thus, the Hoelter index generates useful information even when applied to statistically insignificant χ^2 distributions.

The RMSEA [12–14] is based on the non-centrality parameter and can be evaluated as follows:

$$\mathrm{RMSEA} = \sqrt{\left[\left(\frac{\chi^2}{\mathrm{df}-1}\right)/(N-1)\right]} \quad \text{(D)}$$

where df = degrees of freedom, N = number of observations and χ^2 is the calculated χ^2 value from Eq. A above. Values of RMSEA < 0.05 are considered an excellent fit although values of RMSEA < 0.08 are considered adequate [12,14].

In our analysis, we examined each of our runs and looked for statistically good fits based on at least two of the above mentioned parameters. The Hoelter index was deemed useful even when the calculated χ^2 value was not significant because it produces a peak in the N_{crit} parameter. We therefore look for minima in both the χ^2 value and the RMSEA and maxima in the Hoelter index to assess our goodness of fit.

Appendix 1b. Randomization of Euler poles

We have modified the standard linear congruential (SLC) random number generator [15] in order to produce 'true' randomizing effect. Each run begins with the input of a number between 0 and 655 (the range is arbitrary) to 'seed' the random number generator. The 'seed' number will generate the same sequence of 'random numbers' in most computer systems using the SLC formula. Our algorithm begins the run with this seed and when a new Euler pole is called for, the program seeks the next sequential value in the series of random numbers. Following this, the program is 're-seeded' using a random choice from one of the first 50 longitude values in the model. This has the practical effect of generating a sequence of truly random drift models. The utility of this method is that the algorithm can be 'reset' by quitting the program and starting over with the same random seed. This allows us to test the sensitivity of the model to changing input parameters.

References

[1] M.W. McElhinny, P.L. McFadden, R.T. Merrill, The time-averaged paleomagnetic field 0–5 Ma, J. Geophys. Res. 101 (1996) 25007–25027.

[2] T. Hatakeyama, M. Kono, Geomagnetic field model for the last 5 myr: time averaged field and secular variation, Phys. Earth Planet. Inter. 133 (2002) 181–215.

[3] D.V. Kent, M.A. Smethurst, Shallow bias of paleomagnetic inclinations in the Paleozoic and Precambrian, Earth Planet. Sci. Lett. 160 (1998) 391–402.

[4] T.H. Torsvik, R. Van der Voo, Refining Gondwana and Pangea palaeogeography; estimates of Phanerozoic non-dipole (octupole) fields, Geophys. J. Int. 151 (2002) 771–794.

[5] R. Van der Voo, T.H. Torsvik, Evidence for Permian and Mesozoic non-dipole fields provides an explanation for Pangea reconstruction problems, Earth Planet. Sci. Lett. 187 (2001) 71–81.

[6] M.E. Evans, Test of the non-dipole nature of the geomagnetic field throughout Phanerozoic time, Nature 262 (1976) 676–677.

[7] J.D.A. Piper, S. Grant, A paleomagnetic test of the axial dipole assumption and implications for continental distribution through geologic time, Phys. Earth Planet. Inter. 55 (1989) 37–53.

[8] R. Hollerbach, C.A. Jones, On the magnetically stabilizing role of the Earth's inner core, Phys. Earth Planet. Inter. 87 (1995) 171–181.

[9] M.W. McElhinny, P.L. McFadden, Paleomagnetism Continents and Oceans, Academic Press International Geophysics Series 73, Boston, MA, 2000,386 pp.

[10] J.R. Carr, Numerical Analysis for the Geological Sciences, Prentice Hall, Englewood Cliffs, NJ, 1995, pp. 104–106.

[11] J.W. Hoelter, The analysis of covariance structures, Soc. Methods Res. 11 (1983) 325–344.

[12] R. Cudeck, M.W. Browne, Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value, Psychometrika (1992) 357–369.

[13] L. Hu, P.M. Bentler, Evaluating model fit, in: R.H. Hoyle (Ed.), Structural Equation Modeling: Concepts, Issues and Applications, Sage Publishing, Thousand Oaks, CA, 1995, pp. 76–99.

[14] J. Nevitt, G.R. Hancock, Improving the root mean square error of approximation for nonnormal conditions in structural equation modeling, J. Exp. Educ. 68 (2000) 251–268.

[15] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vettering, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, Cambridge, MA, 1988, pp. 191–225.

[16] A. Reymer, G. Schubert, Phanerozoic addition rates to the continental crust and crustal growth, Tectonics 3 (1984) 63–77.