Geochemistry Geophysics Geosystems G3
AN ELECTRONIC JOURNAL OF THE EARTH SCIENCES
Published by AGU and the Geochemical Society

Article
Volume 7, Number 9
26 September 2006
Q09009, doi:10.1029/2006GC001350
ISSN: 1525-2027

Limitations in correlation of regional relative geomagnetic paleointensity

D. G. McMillan
Department of Earth and Space Science and Engineering, York University, Toronto, Ontario, Canada M3J 1P3 (dgm@yorku.ca)

C. G. Constable
Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0225, USA (cconstable@ucsd.edu)

[1] Time domain correlations of common features among relative paleointensity records from sedimentary cores are invaluable to paleomagnetism and paleoclimatology. Sediments with high accumulation rates might now provide millennial scale correlations of temporal variations in the geomagnetic dipole moment. Errors in the ages of paleomagnetic data samples, however, can make such correlations difficult and unreliable. We use spectral methods to assess the level of coherence expected among individual and stacked high-resolution simulated paleointensity records for the time interval 0–75 ka. Correlations between individual paleointensity records are systematically degraded with decreased sedimentation rate and increased magnitude of age errors. We find that with optimistic age errors and interpolation of depth sampled data to evenly spaced time series, only short period signal in high-resolution relative paleointensity is corrupted. For currently available methods of establishing chronologies, we estimate the minimum characteristic timescale of correlative features between pairs of regional stacked records at about 4.5 kyr. From an analysis of NAPIS-75 and SAPIS data, it appears that the limit is inherent to the regional stacks and not a consequence of comparison of distant, independent data sets. A detailed comparison of the NAPIS-75 and SAPIS stacks shows that this limit is likely larger, perhaps 6 kyr. At long periods the two regional stacks are more poorly correlated than those from our simulations, suggesting somewhat larger age errors in the individual paleointensity records.

Components: 7325 words, 5 figures, 2 tables.

Keywords: relative geomagnetic paleointensity; high-resolution magnetostratigraphy; regional stacking; interhemispheric correlation; age errors; spectral analysis.

Index Terms: 1560 Geomagnetism and Paleomagnetism: Time variations: secular and longer; 1530 Geomagnetism and Paleomagnetism: Rapid time variations; 3270 Mathematical Geophysics: Time series analysis (1872, 4277, 4475).

Received 25 April 2006; **Accepted** 22 June 2006; **Published** 26 September 2006.

McMillan, D. G., and C. G. Constable (2006), Limitations in correlation of regional relative geomagnetic paleointensity, *Geochem. Geophys. Geosyst.*, *7*, Q09009, doi:10.1029/2006GC001350.

1. Introduction

[2] Time series of paleomagnetic directions and relative intensity from lacustrine and marine sediment cores provide the most complete and precise picture of long period geomagnetic variations. Relative paleointensity variations, with characteristic timescales ranging from centuries to millions of

years, are vital for understanding the geodynamo and can provide high-resolution magnetostratigraphic age constraints for other geoscientific disciplines, particularly climate science. Realizing this goal is dependent upon a thorough understanding of the degree to which geomagnetic signal is disrupted by a number of physical factors and data processing methods. These disruptions act to attenuate, shift in age, or limit the reliability of temporal features that may be used to establish age constraints.

[3] In paleomagnetism, we expect that, in the absence of any interference, time series of measurements from nearby marine sediment cores will be almost identical, whereas those from progressively more distance locations will become less similar, and eventually have only global (dipolar) signal in common. In reality though, a number of effects will cause a reduction in the degree of correlation between two relative paleointensity time series. These include (1) magnetization acquisition (lock-in) processes; (2) depth sampling rate; (3) errors in the measurement of NRM; (4) improper normalization; (5) resolution of age model; (6) age errors due to unknown variations in sedimentation rate, uncertainties in tie point ages, and misidentification of tie points (human error); and (7) interpolation to a common age sequence.

[4] *Teanby and Gubbins* [2000] show that (1) can act as a low pass filter on the time series and (2), or more accurately time sampling rate, can cause significant aliasing. *Guyodo and Channell* [2002] did a systematic study of the effects of (5) and age errors due to sedimentation rate variations, one of the three contributions to (6). Several important effects are neglected in both studies, leading to potential over-estimation of data quality. Uniform depth sampling becomes non-uniform time sampling after transformation from depth to age; jumps in sampling rate occur at tie points, where lower interval sedimentation rates correspond to lower resolution time sampling. Interpolation to a common age sequence is thus less accurate in intervals with lower sedimentation rate, effectively filtering those sections more severely than others with larger sedimentation rate [*McMillan et al.*, 2004]. Age errors due to uncertainties in tie point ages are generally larger in magnitude than those due to unknown variations in sedimentation rate and cause more severe degradation of paleomagnetic signal [*McMillan et al.*, 2002]. It is often the case that a pair of paleointensity time series that are to be compared or combined are derived from quite different age models, unlike the comparisons of *Guyodo and Channell* [2002]. When age models are different there is a substantial reduction in correlation between otherwise identical time series [*McMillan et al.*, 2004], but this effect could be masked by large depositional errors.

[5] Stacking, or averaging, globally and regionally distributed paleomagnetic time series is a common analysis technique in sedimentary paleomagnetism. It is widely thought that stacking multiple records of relative geomagnetic paleointensity will enhance the common signal originating in geomagnetic dipole moment variations and reduce the influence of noise and local geomagnetic variations, and this has been shown to be a plausible assumption at long periods [*McMillan et al.*, 2002, 2004]. Well known global stacks have been used to infer geomagnetic dipole moment variations throughout the Brunhes polarity interval [*Guyodo and Valet*, 1996, 1999; *Laj et al.*, 2004], and the past two million years [*Valet et al.*, 2005]. These variations are of interest in their own right, but they also affect production of radiogenic isotopes in the upper atmosphere [*Elsasser et al.*, 1956] complicating attempts to understand the influence of production on climate [e.g., *Beer et al.*, 2002]. Regional paleointensity [*Channell et al.*, 2000; *Laj et al.*, 2000; *Stoner et al.*, 2002] stacks have also been used in a number of contexts, including estimating the duration of the Laschamp excursion [*Laj et al.*, 2004] and detailed magnetostratigraphic applications. *Stoner et al.* [2002] conclude that NAPIS-75 and SAPIS, two regional paleomagnetic stacks with independent chronologies from the northern and southern hemispheres reveal similar features with characteristic timescales on the order of 1 kyr. This result has important implications for paleoclimatology because it may indicate that constraints derived from paleointensity stacks on the relative timing of rapid climate change events in the two hemispheres are better than can be achieved with conventional isotope stratigraphy methods. The NAPIS-75 [*Laj et al.*, 2000] and SAPIS [*Stoner et al.*, 2002] stacks are commonly referred to as "high resolution," indicating that mean sedimentation rates of the contributing records exceed 10 cm/kyr and measurements are made in pass-through magnetometers, which provide pseudo-continuous relative intensity time series.

[6] To date there has been no systematic analysis to document the reliability or statistical significance of the correlation suggested by *Stoner et al.* [2002]. The main goal of this paper is to conduct such an analysis and to discover how the disruptive ele-

ments described above affect the ability to correlate regional stacks. We first report results of a comprehensive investigation of items 2, 5, 6 and 7 (listed above). We neglect (1) because lock-in effects are not expected to be a factor with high sedimentation rate cores. In low accumulation rate cores (~2 cm/kyr), lock-in processes act as a low pass filter with a cut-off at a period of about 4 kyr, which is equivalent to f_c = 0.25 kyr^{-1} [*Teanby and Gubbins*, 2000]. The cut-off period decreases (f_c increases) with increasing accumulation rate. Age errors typically disrupt signals at periods less than about 5 kyr (f_c = 0.20 kyr^{-1}) depending on the type and their magnitude [*McMillan et al.*, 2002]. In the statistical model considered here [*McMillan et al.*, 2002], errors of realistic magnitude in NRM and improper normalization (items 3 and 4 above) are modeled simultaneously as a Gaussian process, based on the prevailing model for acquisition of detrital remanent magnetism [*Tauxe*, 1993; *Valet*, 2003]. Model errors with a magnitude of 30% of the relative intensity determination are overwhelmed by age errors in the disruption of paleomagnetic signal, even at relatively low sedimentation rates. For this reason and because we are primarily interested in isolating the effects of age errors, we neglect items 3 and 4 above. Most normalization methods involve a measure of bulk magnetization, such as magnetic susceptibility or anhysteretic remanent magnetization (ARM) [*King et al.*, 1983; *Tauxe*, 1993; *Tauxe et al.*, 1995], but in the event that the standard model of magnetization acquisition is inadequate, these errors may be substantially larger [*Tauxe et al.*, 2006]. We compare simulated regional stacks from the northern and southern hemispheres, whose contributing time series are individually subject to resampling, age errors and interpolation. Lastly, a detailed analysis of the spectral characteristics of the NAPIS-75 and SAPIS data sets is presented.

2. Simulating and Analyzing Data

2.1. Geomagnetic Field

[7] We generate time series of magnetic intensity from the Gauss coefficients of simulation g by Glatzmaier and Roberts [*Glatzmaier et al.*, 1999]. This particular model was chosen for its favorable characteristics: reasonable secular variation that is neither too strong nor too weak as compared to Earth's field; and spectral characteristics that are intermediate between alternate simulations e and h. All three simulations have been scrutinized using a variety of analysis methods [*Glatzmaier et al.*, 1999; *Coe et al.*, 2000; *Bouligand et al.*, 2005]. *McMillan et al.* [2004] showed that when analyzing stacked time series, spectral results are robust among the three simulations cited above and are thus applicable to paleomagnetic data in a quantitative way.

2.2. Spectral Methods for Data Analysis

[8] We follow the methods of *McMillan* [2006], which are briefly summarized here for completeness. Our data analyses are based primarily on the coherence between pairs of paleomagnetic time series, either individual records or regional stacks. The coherence spectrum $\gamma^2(f)$ illustrates how correlation varies as a function of frequency and indicates the frequency bands over which we can expect reliable correlations. We compute a confidence level $\mathcal{C}_\alpha$, below which two white noise time series will be uncorrelated $\alpha \times 100\%$ of the time. Conversely, strong coherence indicates that the two time series are linearly related, and have statistically similar features at those frequencies. The phase gives the relative timing of correlated features.

[9] The critical frequency f_c and coherence at zero frequency $\gamma_0^2 = \gamma^2(f = 0)$ characterize two important features of coherence spectra in the analysis of paleomagnetic time series: the uppermost frequency at which we cannot expect the two time series to be correlated; and the quality of the coherence where the two are correlated. f_c is defined to be the frequency at which the coherence first drops below the $\mathcal{C}_{.99}$ confidence level and marks the upper limit of the coherent band. The critical period T_c is the reciprocal of f_c. Large values of γ_0^2 indicate reliable correlation at long periods.

[10] In summary, we compute coherence and phase spectra between individual or stacked pairs of time series, then find f_c and γ_0^2. These give a simple and qualitative means of assessing our ability to correlate features between two similar time series: features with characteristic lengths smaller than T_c cannot be reliably correlated. This does not mean that the features themselves are unreliable, but simply that they cannot be reliably compared between the two time series. As the strength of the coherent band decreases (lower γ_0^2), the correlation of features with characteristic lengths larger than T_c becomes less reliable.

2.3. Sampling, Sedimentation, and Tie Points

[11] The ability to delineate temporal paleomagnetic variations and identify contemporaneous fea-

tures in different cores depends on a number of data processing techniques, the most important of which is the transformation of core depth to age. Most relative intensity records are provided with a first order age scale by comparison of corresponding climate proxy measurements, such as δ^{18}O or magnetic susceptibility from the same core (for example, NAPIS-75: *Channell et al.* [1997]; *Kissel et al.* [1997]; SAPIS: *Channell et al.* [2000]) with an accepted time series of climate change based on δ^{18}O variations [*Martinson et al.*, 1987]. Although stacking paleointensity time series can reinforce long period signal, inevitable uncertainties in inferred age scales of individual records may degrade the quality of short period signals and affect the correlation of fine structure among a group of records. Potentially important medium and short period features can be significantly attenuated and shifted in age [*McMillan et al.*, 2004], making identification of common medium and short period features in stacked records difficult or statistically unreliable.

[12] We simulate intensity records at geographic locations corresponding to the cores in the NAPIS-75 and SAPIS data compilations. Core lengths, mean sedimentation rates (also abbreviated by MSR) and time series lengths are similar to those of the original data. We neglect the smoothing effects of magnetometer transfer functions and fix the depth interval between measurements at Δz = 1 cm. This reflects a typical measurement interval in a pass through magnetometer, where the level of smoothing depends on the magnetometer aperture [*Constable and Parker*, 1991]. The response function width at half peak amplitude is of the order of 5 cm for the U-channel samples used in these high-resolution studies [*Guyodo et al.*, 2002]. For sedimentation rates exceeding 10 cm/kyr this will only result in smoothing of records at periods shorter than about 1000 years (f_c = 1 kyr^{-1}).

[13] Each intensity time series is subject to the following procedures: temporal resampling according to the prescribed depth sampling and a unique age model; contamination with age errors; and interpolation to a common age interval. We use Δa = 0.1 kyr, the interpolation interval of the NAPIS-75 stack.

[14] The age models are determined by specifying random collections of depths, ages and age uncertainties. We investigate three decreasingly accurate sets of tie point ages, which we refer to as levels 1, 2 and 3. *Guyodo and Channell* [2002] used the same approach in their assessment of the effects of depositional errors. Level 1 age models have an average temporal density of one about every 5 kyr, and standard deviations in ages that increase approximately linearly with age to a maximum of 2.6 kyr. The density of level 2 age models is one per 20 kyr on average, with uncertainties that are consistent with those encountered when using ages of oxygen isotope events. The largest level 2 standard deviation (SD) in age in our simulations is 5.0 kyr, whereas the largest standard error for δ^{18}O events for the past 75 kyr is 7.74 kyr [*Martinson et al.*, 1987]. For level 3, there are only two tie points (TP) in the age model, one at the top and one at the bottom of the core segment, whose uncertainties are similar to those of level 2. Level 1 and 3 data are designed to represent the maximum and minimum density and accuracy of tie points available in this type of age model. In some cases, we also refer to error-free results, in which we consider the tie point ages to be exact and the sedimentation rates to be constant between tie points. Individual age models are randomly generated for every intensity time series and automatically checked for extreme changes in sedimentation rate or very large sedimentation rates between TPs. Age models that are deemed unphysical are rejected.

[15] We consider a combination of three types of age errors: event errors, resulting from uncertainties in the ages assigned to tie points; misidentification errors, incurred when ages are assigned to the wrong depth; and depositional errors, which arise from the assumption of a constant sedimentation rate between TPs, when in fact, accumulation varies by some unknown stochastic process. Since our age models and age errors are generated randomly, we examine 100 independent realizations for each set of parameters in order to illustrate statistically the strengths and weaknesses of each case. In general, we accomplish this by presenting mean values and SDs of f_c and γ_0^2 over those 100 realizations. Multiple realizations also provide estimates of the probability of certain examples of undesirable behavior, such as poor coherence at long periods (i.e., small γ_0^2).

3. Paleointensity From Nearby Cores

[16] We now look at the effects of age errors (independent of age model and MSR) on the coherence between two paleointensity time series from nearby cores. We adopt the locations of cores PS2644-5 and SU90-24 from the NAPIS-75 data set, which have MSRs between 15 and 20 cm/kyr, durations of about 75 kyr and a geographical

Table 1. Experimental Parameters for Simulated Nearby Pairs[a]

Index	MSRs, cm/kyr	Min/Max Event Error SD, kyr	Δz, cm	Δa, kyr
1	20/20	0.5/2.5	1.0	0.1
2	20/20	1.0/5.0	1.0	0.1
3	6/6	0.5/2.5	1.0	0.1
4	6/6	1.0/5.0	1.0	0.1
5	20/20	0.5/2.5	1.0	0.5
6	20/20	1.0/5.0	1.0	0.5
7	6/6	0.5/2.5	1.0	0.5
8	6/6	1.0/5.0	1.0	0.5
9	20/20	0.5/2.5	4.0	0.1
10	20/20	1.0/5.0	4.0	0.1
11	6/6	0.5/2.5	4.0	0.1
12	6/6	1.0/5.0	4.0	0.1

[a] Mean sedimentation rates (MSRs) are quoted for both cores in comparison. Event error standard deviations (SD) are for level 2 and 3 tie point data only (see section 2.3 for explanation about level 1 tie point data); all other parameters are valid for level 1 results. Δz is the depth sampling interval, and Δa is the interpolation interval for final time series.

separation of 14°. The two simulated paleointensity time series have MSRs of 20 cm/kyr and each is assigned its own age model. We take one of the data sets as the reference and duplicate the second 100 times. All 101 data sets are then subject to independent realizations of age errors. We compute 100 coherence spectra, critical frequencies and coherences at zero frequency. The mean values $\bar{f}_c$ and $\bar{\gamma}_0^2$ and SDs based on each set of realizations are the basis for our discussion below. By counting the number of occurrences in each of our 100 realizations, we estimate the probability that the coherence is less than $\mathcal{C}_{.99}$ for all frequencies and the probability that the coherence at zero frequency is less than $\mathcal{C}_{.99}$.

[17] We repeat the experiment described above for each set of parameters in Table 1, which includes a reduction in MSR to 6 cm/kyr because the smallest MSR of the NAPIS-75 compilation is somewhat less than 10 cm/kyr and because we wish to explore in detail the effects of MSR later. We do not apply age errors of two different magnitudes to level 1 age models because this represents the best achievable accuracy. Since our main goal is to assess the limits and reliability of correlation between regional stacks and because these experiments produce numerous results, we limit our presentation here to a summary of our findings on nearby individual relative paleointensity time series.

[18] We first recall that, because we are examining comparisons between individual time series, we can make only qualitative observations about trends related to changes in parameters. The width of the coherent band is not significantly reduced with decreasing accuracy in age control, but it becomes less reliable as the age scale becomes less certain. The highest probability that we find no coherence in any of these experiments is 1%, however, the existence of a truncated coherent band becomes substantially more probable with decreasing age control; maximum probabilities are in the range of 30–50% when considering level 3 age models. We see minor decreases in both $\bar{f}_c$ and $\bar{\gamma}_0^2$ when there is an increase in the magnitude of age errors with level 2 and 3 TP data, while increasing the interpolation time step to 0.5 kyr or increasing the depth sampling interval to 4 cm does not significantly reduce $\bar{f}_c$ or $\bar{\gamma}_0^2$.

[19] We follow a similar procedure to establish the effects of MSR in the absence of age errors. In this case, we begin with 7 groups of 100 time series, identical in every respect except each group has a different MSR: 2, 6, 10, 15, 20, 30 and 40 cm/kyr. Within each group, there are 100 different realizations of level 2 depth-age data giving a total of 700 independent age models. Each group of 100 time series is compared to each of the other groups pairwise, resulting in 100 coherence spectra for each combination of two MSRs, of which there are 21. For each of these combinations, we compute $\bar{f}_c$ and $\bar{\gamma}_0^2$ over the 100 coherence estimates.

[20] Every comparison involving a core with low sedimentation rates has a comparatively low $\bar{f}_c$. There is no improvement in coherence when one of the two cores has $\bar{s}$ = 2 cm/kyr. We find an upper limit in $\bar{f}_c$ that increases with the smaller of the two MSRs. In reality though the situation is not so

simple. The experiments described above, which include the effects of age errors, show that mean critical frequencies are not statistically different when increasing the common MSR among pairs of time series from 6 to 20 cm/kyr. Nevertheless, we are able to reach an important conclusion: we must minimize the effects of age errors in order to reap the full benefits of high sedimentation rates.

[21] We turn now to real paleointensity data and compute f_c and γ_0^2 for each pair of relative intensity time series in each of the NAPIS-75 and SAPIS compilations (Figures 1 and 2, respectively). The core indices are in order of increasing MSR, which ranges from 7.9 to 37 cm/kyr (NAPIS-75) and 14.9 to 25 cm/kyr (SAPIS). Zero values of f_c indicate that there is no significant coherence between the two time series anywhere in the relevant frequency range.

[22] For the NAPIS-75 data (Figure 1), we find that $f_c = 0$ in three of five comparisons with core 1 (SU90-33). There is also a truncated coherent band in the coherence of SU90-33 with core 6 (MD95-2034, which has the largest MSR). These two observations are supported by γ_0^2 results, which are below $\mathcal{C}_{.99}$ for all of the comparisons noted above. Overall, there is no systematic dependence on MSR, although it is true that the poorest performance is associated with core SU90-33, the core with the lowest MSR (a point also noted by *Laj et al.* [2000]). We cannot be certain this is due only to the low MSR because we have seen $\bar{s}$ = 7.9 cm/kyr time series ought to be reasonably coherent with other time series from cores with equal or larger MSRs. Furthermore the existence of the truncated coherent band is a good indicator that core SU90-33 is likely the victim of unusually large age errors [*McMillan et al.*, 2002]. Apart from age errors, *Laj et al.* [2000] suggested that problems with correlation between individual records may be due to different lock-in depths. For all but one of the remaining NAPIS-75 comparisons, we find that the critical frequency has a distinct upper bound of about 0.2 kyr^{-1}, which corresponds to a critical period T_c = 5 kyr. Since only one pair out of 15 demonstrates strong coherence at frequencies larger than 0.2 kyr^{-1}, it is not likely that a stacked result will perform any better than the upper bound on f_c. Accordingly, we do not expect there to be reliable signal with characteristic timescales smaller than 5 kyr.

[23] The SAPIS data (Figure 2) yields a rather different outcome: there is a much higher incidence of a truncated coherent band (6 out of 10 pairs); and critical frequencies are substantially larger than those for NAPIS-75. We see again that age errors can potentially swamp gains made by higher MSRs by noting that core PC4, with the largest MSR (25 cm/kyr) shows no significant

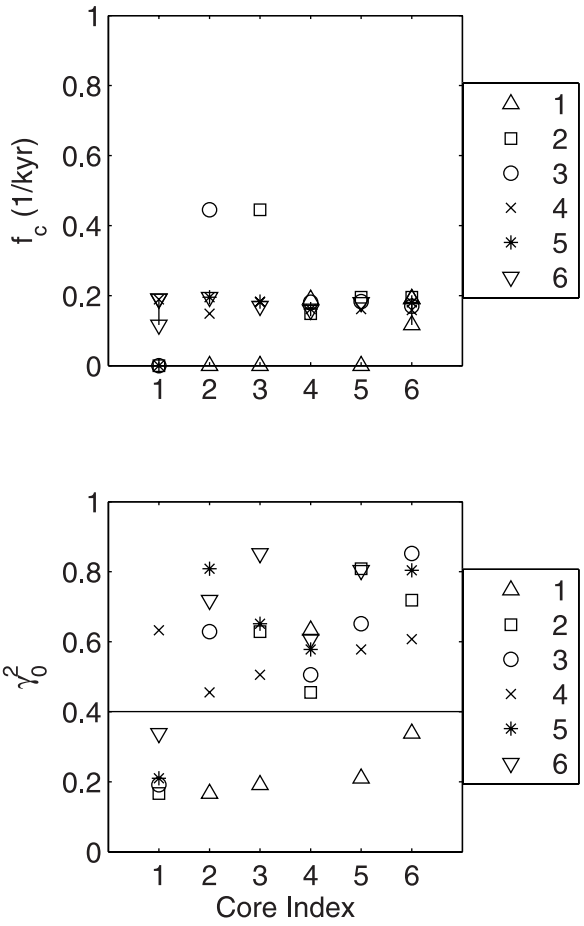

Figure 1. Comparison of relative paleointensity time series from the NAPIS-75 compilation. (top) Critical frequencies and (bottom) coherences at zero frequency and the 99% confidence level for coherence (bottom, horizontal line) are shown for time series from the NAPIS-75 compilation [*Laj et al.*, 2000]. The core indices on the x axis and in the legend represent the individual cores arranged in order of increasing mean sedimentation rate: 1, SU90-33; 2, ODP983; 3, PS2644-5; 4, SU990-24; 5, MD95-2009; 6, MD95-2034. Each symbol is a direct comparison of a particular pair of time series from the NAPIS-75 data set; for example, the squares show the critical frequencies f_c (or mean coherence at zero frequency γ_0^2) for core ODP983 as compared with every other core as indexed on the x axis. A zero value for f_c indicates that there is no significant coherence over the full frequency range. The vertical lines connecting pairs of like symbols indicate that the coherence is above the 99% confidence level only in the frequency range between the connected points.

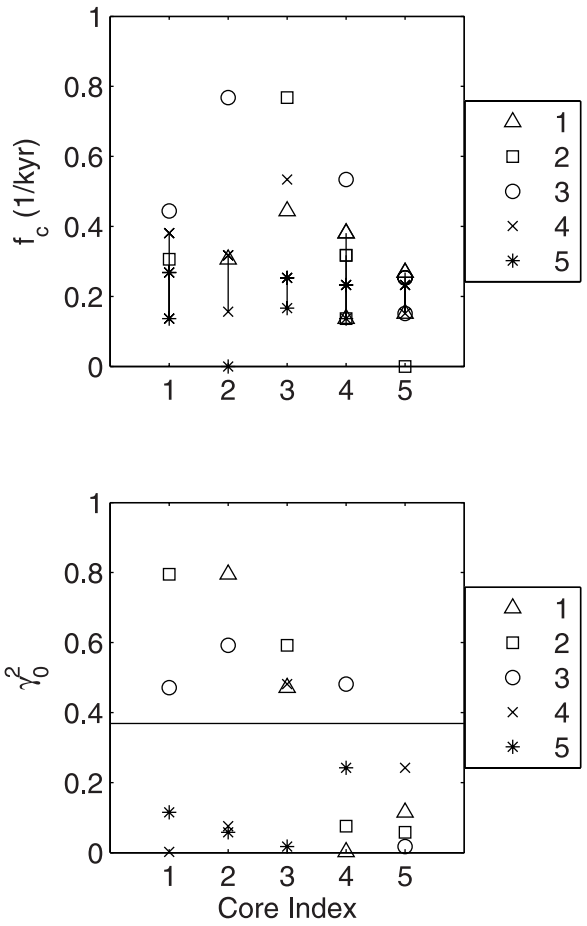

Figure 2. Comparison of relative paleointensity time series from the SAPIS compilation. (top) Critical frequencies and (bottom) coherences at zero frequency and the 99% confidence level for coherence (bottom, horizontal line) are shown for time series from the SAPIS compilation [*Stoner et al.*, 2002]. The core indices on the x axis and in the legend represent the individual cores arranged in order of increasing mean sedimentation rate: 1, PC21; 2, PC10; 3, 1089; 4, PC5; 5, PC4. Each symbol is a direct comparison of a particular pair of time series from the SAPIS data set; for example, the squares show the critical frequencies f_c (or mean coherence at zero frequency γ_0^2) for core PC10 as compared with every other core as indexed on the x axis. A zero value for f_c indicates that there is no significant coherence over the full frequency range. The vertical lines connecting pairs of like symbols indicate that the coherence is above the 99% confidence level only in the frequency range between the connected points.

coherence with PC10 and no significant coherence with any of the others at frequencies less than $f_c \approx 0.15$ kyr^{-1}. The lower limits of the truncated coherent bands are clustered just above 0.15 kyr^{-1}.

[24] SAPIS critical frequencies are well above the 0.2 kyr^{-1} we saw for NAPIS-75, and exceed 0.4 kyr^{-1} in three comparisons. With critical periods of 5 (NAPIS-75) and 2.5 kyr (SAPIS), it is possible that the SAPIS stack contains reliable signal at smaller critical periods (perhaps as low as 2.5 kyr), but the overall lack of consistency and the large percentage of truncated coherent bands suggest that there may be substantial discrepancies in the individual age models, likely due to overtuning to other paleomagnetic reference curves and rendering the stack less reliable than the NAPIS-75 result at periods in excess of about 6.5 kyr.

[25] Stacking does indeed reinforce long period signal [*McMillan et al.*, 2004] and therefore we do not recommend discarding poor performers such as SU90-33 and PC4; they likely do not adversely affect the result. In the SAPIS stack, the numerous truncated coherent bands are also likely vanquished by the strong and wide coherent bands of the remaining pairs.

4. Interhemispheric Correlation of Regional Stacks

[26] We first consider an assessment of errors that might be expected in the interhemispheric correlation of regional stacks like NAPIS-75 and SAPIS. We begin with six simulated intensity time series corresponding to the NAPIS-75 locations and five corresponding to the SAPIS data. Each is subject to: resampling and interpolation according to a randomly generated age model; age errors; and pre-stack interpolation with age step $\Delta a = 0.1$ kyr. The two groups are stacked and a frequency domain comparison performed as above. Again taking the statistical approach, we repeat this process 100 times with independent realizations of age models and age errors for all 9 time series.

[27] Results for the parameter values of Table 2 are shown in Figure 3. Critical frequencies are clustered around 0.2 kyr^{-1} and for level 2 and 3 TP data, larger errors lead to smaller f_c. Although the standard deviations of f_c for level 3 are larger than those for level 2, there is some indication that critical frequencies are, on average, larger and the coherent band wider. This is likely due to the fact that we do not impose age errors on the uppermost tie points in cores when their ages are less than 1 kyr, leading to a more reliable stack and better coherence in the more recent part of the record. The resulting wider coherent band suggests that it

Table 2. Experimental Parameters for Figure 3[a]

Index	Tie Point Level	Min/Max Event Error SD, kyr	Δz, cm	Δa, kyr
1	1	–	1.0	0.1
2	2	0.5/2.5	1.0	0.1
3	2	1.0/5.0	1.0	0.1
4	3	0.5/2.5	1.0	0.1
5	3	1.0/5.0	1.0	0.1

[a] Event error standard deviations (SD) are for level 2 and 3 tie point data only (see section 2.3 for explanation about level 1 tie point data). Δz is the depth sampling interval, and Δa is the interpolation interval for final time series.

may sometimes be wiser to choose only a few tie points with small uncertainties over many with potentially larger uncertainties.

[28] γ_0^2 and probabilities results (Figure 3) indicate that, for this class of stacked records, correlation of long period signal is exceptionally reliable. The net effect of larger age errors is not a significant reduction in critical frequency, but rather an increase in the likelihood that γ_0^2 decreases. In other words, larger age errors do not make the coherent band narrower, just potentially less reliable. For this class of paleointensity time series and age errors of reasonable magnitude (although they are perhaps a bit on the optimistic side), there appears to be a distinct limit to the width of the coherent band. The best age control (level 1 TP data) yields $f_c \approx 0.21$ kyr^{-1}, corresponding to $T_c \approx 4.5$ kyr.

[29] In Figure 4, we compare the NAPIS-75 and SAPIS stacks in the time and frequency domains. The coherence reveals a truncated coherent band, while the phase indicates a slight shift in roughly the same frequency band. The maximum phase shift is about 0.62 radians at $f = 0.1$ kyr^{-1}, which corresponds to a time shift of about 1 kyr. The peak coherence is $\gamma^2 = 0.45$, which unfortunately means that correlating features with characteristic lengths between 6 and 19 kyr will not be terribly reliable. f_c and γ_0^2 for this comparison are reproduced in Figure 5 above the asterisk. For interhemispheric correlation of regional stacks, we are primarily interested in features of short duration, which are typically about 2 kyr, but as large as 7 kyr [see *Stoner et al.*, 2002, Figure 8]. Thus we ignore the low coherence at timescales larger than 19 kyr. Although f_c is comparable to that of the simulations described above, γ_0^2 is considerably lower, barely above $\mathcal{C}_{.99}$. As noted above, larger age errors have a more dramatic effect on γ_0^2 than on f_c, which given the value for NAPIS/SAPIS, suggests larger age errors than those used in the simulations, i.e., maximum event error standard deviation of 5 kyr.

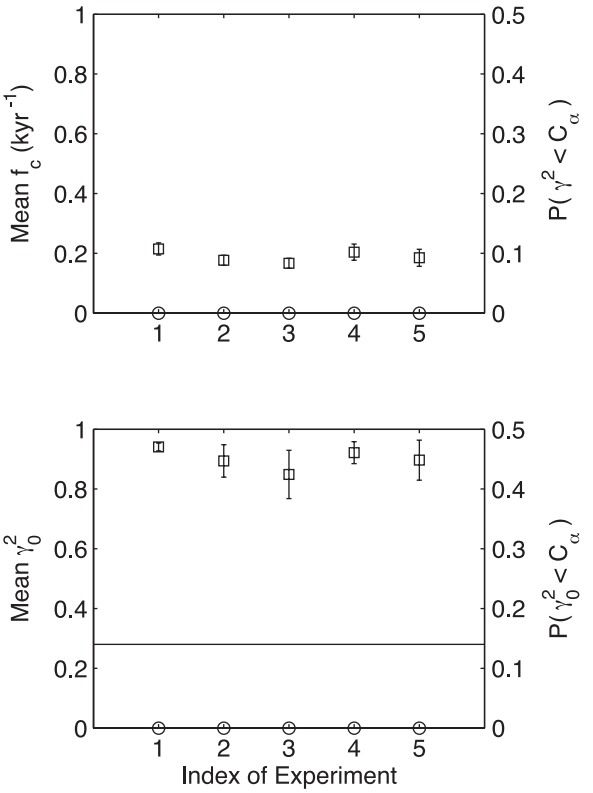

Figure 3. Interhemipsheric correlation of simulated regional stacks. (top) Mean critical frequencies (squares) and (bottom) mean coherences at zero frequency (squares), with their standard deviations (error bars), and the 99% confidence level for coherence (bottom, horizontal line) are shown for simulated regional stacks from the northern and southern hemispheres. Also shown are the estimated probabilities that the coherence is less than the 99% confidence level at all frequencies (top, circles) and the estimated probabilities that the coherence at zero frequency is less than the 99% confidence level (bottom, circles). These statistics were computed from coherence spectra for 100 pairs of stacked time series, whose contributing time series were subject to independent realizations of tie point data and age errors. The index of experiment, detailed in Table 2, indicates the different parameters for the individual time series.

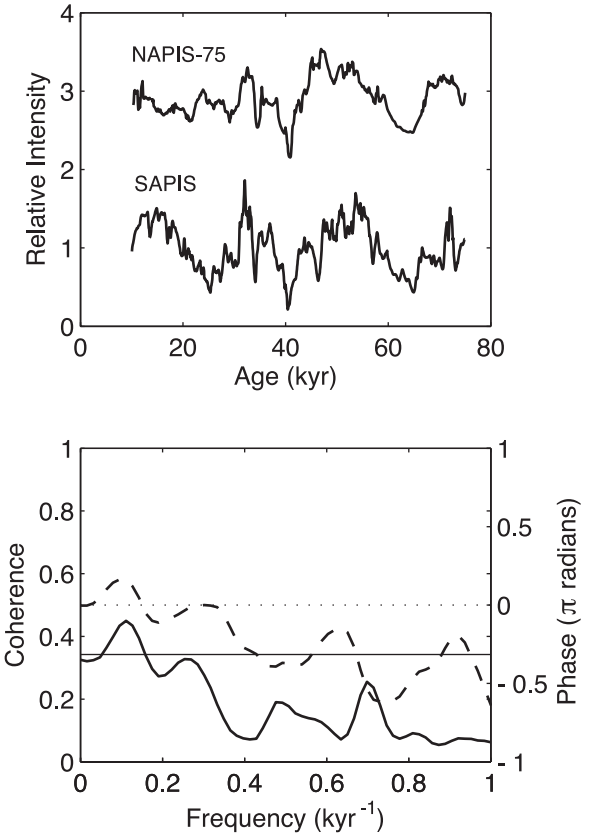

Figure 4. Comparison of NAPIS-75 and SAPIS regional stacks. (top) In the time domain, the two stacks are shown over their common age interval. (bottom) Coherence and phase spectra (solid and dashed, respectively), with the 99% confidence level (solid line) and the zero-phase line (dotted).

[30] On the basis of the above discussion, we should not only be interested in the bulk properties of the coherence, but also its behavior on shorter intervals. We therefore perform an identical spectral analysis of five segments in the common age interval of ~10–75 ka (Figure 5). The segments are equally spaced and each has a duration of 20 kyr. Clearly there is great variability in both f_c and γ_0^2. In the first part of the stacks, at least up to 21 ka, there is no chance of any correlation. The next segment (~21–42 ka) shows reasonable reliability in correlations at timescales larger than 2.6 kyr. From ~32–53 ka, the critical period is increased to 4.0 kyr and the reliability is only slightly above the significance level. Segment 4, from about 43 to 64 ka experiences a truncated coherent band at quite high frequencies, with reliable correlation in the range 1.8 to 2.2 kyr. Lastly, on the interval from ~54 to 75 ka, we find the most reliable correlations down to 4.8 kyr.

[31] In general, γ_0^2 is lower than those in our simulation results, again suggesting somewhat larger, but not devastating, age errors. And despite the fact that segment 4 has an unusually high frequency truncated coherent band, it is excellent for correlating temporal features 2 kyr in duration. In general, we should be looking for correlative features somewhat wider than the lower bound provided by T_c. Many of the features in correlations proposed by *Stoner et al.* [2002] in their Figure 8 do not meet this criteria; they are too narrow to be reliably correlated. In some cases, this is remedied by centering a correlation on the larger surrounding

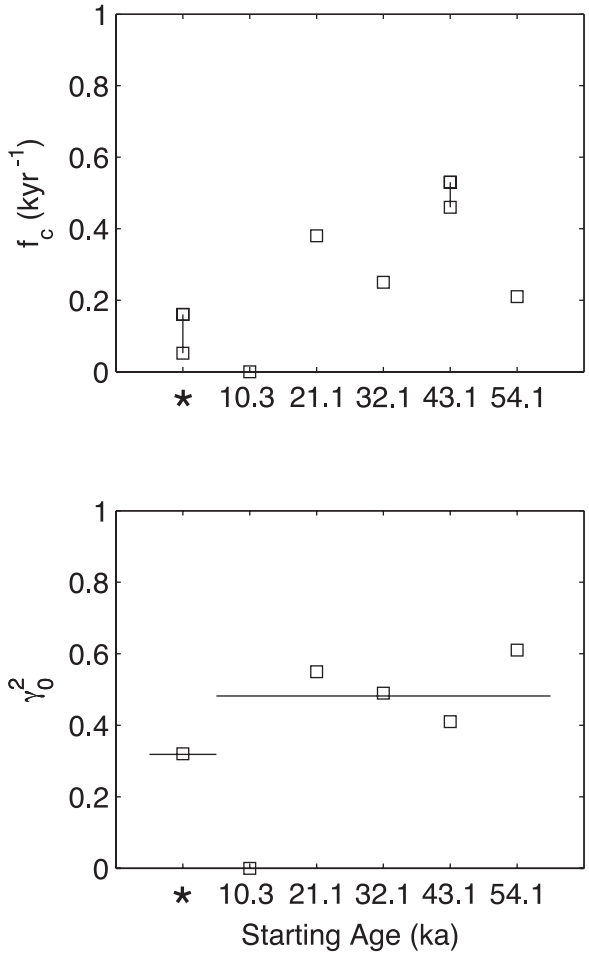

Figure 5. Sliding window analysis of NAPIS-75/SAPIS coherence. Horizontal axis labels give starting ages for every 20 kyr segment of the two regional stacks shown in Figure 4. Results for the full common interval (~10–75 ka) are given at the left and indicated by the asterisk. The horizontal lines represent the 99% confidence levels, which are different due to the different time series lengths and our requirement that the frequency resolution be the same for all comparisons.

low (or high), thus increasing its width into the reliable range. The "resolution" of the correlation is then decreased, but for NAPIS-75/SAPIS critical periods (maximum of 4.8 kyr) millennial scale correlations are still possible (i.e., in the range 4.8–10 kyr). The remaining correlations are generally unreliable by this criteria and do not have sufficiently large maxima or minima surrounding them or they are difficult to identify. Since we are unlikely to find reliable correlations below 4.8 kyr, a low pass filtered version could facilitate the identification of useful features. We note finally that this sliding window analysis indicates the intervals over which we can correlate at the maximum (or minimum) resolution.

5. Summary and Conclusions

[32] The experiments described above provide important insights into the behavior of paleointensity time series from high-resolution marine sediment cores. Here, high resolution implies a mean sedimentation rate (MSR) larger than about 10 cm/kyr and high-density depth sampling, accomplished with a pass-through magnetometer.

[33] We have assumed that the cores are sampled discretely and point-wise at a depth interval of 1 cm in most cases, and have neglected spatial smoothing effects of the response function of a pass-through magnetometer. For pass-through measurements, this depth interval cannot be considered sufficiently large to provide independent paleomagnetic samples, but will have minimal effect on the spectral content of the resulting time series at periods greater than 1 kyr. We have attempted to isolate the effects of age errors and have not considered lock-in processes or uncertainties associated with measurement error and improper normalization. Details of normalization effects, especially the potential for leakage of periodic signals from climate-controlled lithological variations are certainly worth investigation. In principle, it is possible to incorporate each of these effects into our model, and distinguish their effects from those due to age errors and sampling.

[34] Our statistical results are based on the output of simulation g of *Glatzmaier et al.* [1999], and since results for stacked records are similar among statistically diverse simulations, the results can be extended quantitatively to paleomagnetic data [*McMillan et al.*, 2004]. When comparing individual time series, however, simulation results can only be extended in a qualitative way. Certain details, such as the critical frequency for a particular set of parameters, will differ among simulations because the spectral properties are only similar in an average sense. General trends, on the other hand, are robust; for example, increasing the magnitudes of age errors increases the probability of poor coherence between two time series at long periods for any simulation, and correspondingly for actual data.

[35] We examine the effects of three different representative sets of tie point age models on the simulated intensity time series: first, one tie point (TP) at each end of the core section (level 3); secondly, with a temporal TP density of one every 20 kyr, on average (level 2); and lastly, a density of one every 5 kyr, on average (level 1). Level 3 TPs represent the minimum depth-age information that can be applied to a time series and provide lower bounds on our results. Level 2 TPs can be regarded as representative of correlations based on oxygen isotope events and are typical for current paleomagnetic methodology. The uncertainties in age are consistent with isotopic events [*Martinson et al.*, 1987] and have a maximum standard deviation of 5.0 kyr. The more dense and accurate level 1 TPs, with a maximum standard deviation in age of 2.6 kyr, are intended to approximately model the best accuracy achievable in depth-age information. In addition to the age errors associated with uncertainties in TP age, which are designated event errors, we include age errors due to improper identification of tie point depths (misidentification errors) and random variations in sedimentation rate (depositional errors). The latter are typically smaller in magnitude [*McMillan et al.*, 2002] than event errors, but may have significant effects in the case of level 1 TPs.

[36] The critical frequency f_c (equivalently, critical period T_c) and coherence at zero frequency γ_0^2 can be used as simple and effective quality control criteria for time series that are used in paleointensity stacks *McMillan* [2006]. In most cases, f_c gives the width of the coherent band, which in turn determines the timescales at which we can expect reliable correlation between two time series. A truncated coherent band refers to statistically insignificant coherence at the longest periods. In this case, the coherent band is reduced in width at low frequency (long period) rather than by a decrease in f_c. γ_0^2 provides a measure of the quality of the coherent band; larger γ_0^2 indicates a more reliable correlation between two time series within the coherent band and small γ_0^2 (less than the 99%

confidence level) reveals a truncated coherent band. With multiple realizations of TP data, we can estimate the probability of two events: the coherence is less than the 99% confidence level for all frequencies (virtually no chance of correlating anything between two time series); and the coherence at zero frequency is less than the 99% confidence level (truncated coherent band). In both cases, larger probabilities signify less reliability in correlation of the two time series.

[37] In comparisons of individual paleointensity time series subject to age errors, we find that more dense and accurate age models produce more reliable correlations between individual time series, but do not reduce the minimum timescale of correlative features. Conversely, simply going for quantity is not likely to be beneficial. In many instances, rejecting tie points with large uncertainties may improve the ability to correlate two time series. When considering the most dense and accurate age models, we find very strong coherence at long periods with little variance among realizations of the error model, suggesting very reliable correlations. Larger age errors for level 2 and 3 TP data substantially increase the probability of a truncated coherent band. Modest increases in the interpolation time step or the depth sampling interval do not cause significant reduction of the critical frequency or coherence at zero frequency.

[38] Critical frequencies and coherences at zero frequency for the NAPIS-75 and SAPIS compilations show no dependence on MSR. This reflects the unavoidable fact that these data are affected by age errors. With the NAPIS-75 data, there is a distinct upper limit to f_c, which suggests that temporal features with timescales less than about 5 kyr might not be reliable in a stack of this data. Reasonably large values of γ_0^2, however, indicate that the stacking process can indeed reinforce signal at timescales longer than 5 kyr. When examining the SAPIS data, we find critical frequencies that are typically larger and more scattered than those for the NAPIS-75 data and there is a much higher incidence of truncated coherent bands. These results suggest that, although it is possible the SAPIS stack has reliable geomagnetic content at slightly higher frequencies, it is probably less reliable than the NAPIS-75 stack at long periods. Thus the NAPIS-75 stack is more likely to limit the timescales of interhemispheric correlations, while the SAPIS stack will make such correlations less reliable.

[39] When comparing simulated regional stacks of high-resolution relative paleointensity time series, there is very little difference in critical frequency or its standard deviation as age control becomes less accurate. At best, $\bar{f}_c \approx 0.21$ kyr^{-1}, or $\bar{T}_c \approx 4.5$ kyr, implying that, with these types of regional stacks, we cannot expect reliable correlations of features with characteristic timescales less than 4.5 kyr. This lower limit could rise to 5.6 kyr if age errors are larger and tie point density is lower. The strong results for $\bar{\gamma}_0^2$, indicate very good reliability in correlation of features with timescales larger than 4.5 kyr.

[40] Comparisons of NAPIS-75 and SAPIS data yield similar values of f_c critical frequency, but generally lower values of γ_0^2, and hence reliability of correlations. This is an indicator of age errors with standard deviations slightly larger than the 5 kyr maximum used in the simulations. Sliding window analysis of NAPIS-75 and SAPIS data shows that the lower limit on the width of correlative features is quite variable, suggesting that we may relax our temporal width requirements in certain intervals of these stacks. This also gives a good estimate of the age resolution involved in correlations between regional stacks.

Acknowledgments

[41] We thank Carlo Laj and Joe Stoner for providing us with the NAPIS-75 and SAPIS data, respectively, and for a number of thoughtful discussions. Keith Aldridge supplied a comprehensive review of the original draft. D.G.M. acknowledges partial support for this work from the UK Engineering and Physical Sciences Research Council and from the Natural Science and Engineering Research Council of Canada, and C.G.C. acknowledges support from U.S. NSF grant EAR-0337712.

References

Beer, J., R. Muscheler, G. Wagner, C. Laj, C. Kissel, P. Kubik, and H.-A. Synal (2002), Cosmogenic nuclides during Isotope Stages 2 and 3, *Quat. Sci. Rev.*, *21*, 1129–1139.

Bouligand, C., G. Hulot, A. Khokhlov, and G. Glatzmaier (2005), Statistical palaeomagnetic field modelling and dynamo numerical simulation, *Geophys. J. Int.*, *161*, 603–626.

Channell, J., D. Hodell, and B. Lehman (1997), Relative geomagnetic paleointensity and δ^{18}O at Site 983 (Garder Drift, North Atlantic) since 350 ka, *Earth Planet. Sci. Lett.*, *153*, 103–118.

Channell, J., J. Stoner, D. Hodell, and C. Charles (2000), Geomagnetic paleointensity for the last 100 kyr from the sub-Antarctic South Atlantic: A tool for interhemispheric correlation, *Earth Planet. Sci. Lett.*, *175*, 145–160.

Coe, R., L. Hongre, and G. Glatzmaier (2000), An examination of simulated geomagnetic reversals from a paleomagnetic

perspective, *Philos. Trans. R. Soc. London, Ser. A*, *358*, 1141–1170.

Constable, C., and R. Parker (1991), Deconvolution of long-core paleomagnetic measurements: Spline therapy for the linear problem, *Geophys. J. Int.*, *104*, 453–468.

Elsasser, W., E. Ney, and J. Winckler (1956), Cosmic ray intensity and geomagnetism, *Nature*, *178*, 1226–1227.

Glatzmaier, G., R. Coe, L. Hongre, and P. Roberts (1999), The role of Earth's mantle in controlling the frequency of geomagnetic reversals, *Nature*, *401*, 885–890.

Guyodo, Y., and J. E. T. Channell (2002), Effects of variable sedimentation rates and age errors on the resolution of sedimentary paleointensity records, *Geochem. Geophys. Geosyst.*, *3*(8), 1048, doi:10.1029/2001GC000211.

Guyodo, Y., and J.-P. Valet (1996), Relative variations in geomagnetic intensity from sedimentary cores: The past 200,000 years, *Earth Planet. Sci. Lett.*, *143*, 23–36.

Guyodo, Y., and J.-P. Valet (1999), Global changes in intensity of the Earth's magnetic field during the past 800 kyr, *Nature*, *399*, 249–252.

Guyodo, Y., J. E. T. Channell, and R. G. Thomas (2002), Deconvolution of u-channel paleomagnetic data near geomagnetic reversals and short events, *Geophys. Res. Lett.*, *29*(17), 1845, doi:10.1029/2002GL014927.

King, J., S. Banerjee, and J. Marvin (1983), A new rock magnetic approach to selecting sediments for geomagnetic paleointensity studies: Application to paleointensity for the last 4000 years, *J. Geophys. Res.*, *88*, 5911–5921.

Kissel, C., C. Laj, B. Lehman, L. Labeyrie, and V. Bout-Roumazeilles (1997), Changes in the strength of the Iceland-Scotland overflow water in the last 200,000 years: Evidence from magnetic anisotropy analysis of core su90-33, *Earth Planet. Sci. Lett.*, *152*, 25–36.

Laj, C., C. Kissel, A. Mazaud, J. Channell, and J. Beer (2000), North Atlantic palaeointensity stack since 75 ka (NAPIS-75) and the duration of the Laschamp event, *Philos. Trans. R. Soc. London, Ser. A*, *358*, 1009–1025.

Laj, C., C. Kissel, and J. Beer (2004), High resolution global paleointensity stack since 75 kyr (GLOPIS-75) calibrated to absolute values, in *Timescales of the Paleomagnetic Field*, *Geophys. Monogr. Ser.*, vol. 145, edited by J. Channell et al., pp. 255–265, AGU, Washington, D. C.

Martinson, D., N. Pisias, J. Hays, J. Imbrie, T. J. Moore, and N. Shackleton (1987), Age dating and orbital theory of ice ages: Development of a high resolution 0 to 300,000-year chronostratigraphy, *Quat. Res.*, *27*, 1–29.

McMillan, D. (2006), Assessing the quality and limitations of geophysical time series, in *Proceedings of the 20th International Synmposium on High Performance Computing and Applications*, edited by R. Deupree, p. 31, Inst. of Electr. and Electron. Eng., New York.

McMillan, D., C. Constable, and R. Parker (2002), Limitations on stratigraphic analyses due to incomplete age control and their relevance to sedimentary paleomagnetism, *Earth Planet. Sci. Lett.*, *201*, 509–523.

McMillan, D., C. Constable, and R. Parker (2004), Assessing the dipolar signal in stacked paleointensity records using a statistical error model and geodynamo simulations, *Phys. Earth Planet. Inter.*, *145*, 37–54.

Stoner, J., C. Laj, J. Channell, and C. Kissel (2002), South Atlantic and North Atlantic geomagnetic paleointensity stacks (0–80 ka): Implications for interhemispheric correlation, *Quat. Sci. Rev.*, *21*, 1141–1151.

Tauxe, L. (1993), Sedimentary records of relative paleointensity of the geomagnetic field: Theory and practice, *Rev. Geophys.*, *31*, 319–354.

Tauxe, L., T. Pick, and Y. Kok (1995), Relative paleointensity in sediments: A pseudo-Thellier approach, *Geophys. Res. Lett.*, *22*, 2885–2888.

Tauxe, L., J. Steindorf, and A. J. Harris (2006), Depositional remanent magnetization: Toward an improved theoretical and experimental foundation, *Earth Planet. Sci. Lett.*, *244*, 515–529.

Teanby, N., and D. Gubbins (2000), The effects of aliasing and lock-in processes on palaeosecular variation records from sediments, *Geophys. J. Int.*, *142*, 563–570.

Valet, J.-P. (2003), Time variations in geomagnetic intensity, *Rev. Geophys.*, *41*(1), 1004, doi:10.1029/2001RG000104.

Valet, J.-P., L. Meynadier, and Y. Guyodo (2005), Geomagnetic dipole strength and reversal rate over the past two million years, *Nature*, *435*, 802–805.