AGU PUBLICATIONS

Paleoceanography

RESEARCH ARTICLE
10.1002/2016PA002963

Key Points:
- Confidence levels of ≥99% should be used to validate the existence of astronomical cycles in strata when only power spectra are investigated
- The 90%–99% confidence levels that are commonly employed will lead to the occurrence of excessive type I (false positive) errors
- Fixed value confidence levels are likely ill suited to verifying astronomical forcing in strata in the absence of additional evidence

Supporting Information:
- Supporting Information S1

Correspondence to:
D. B. Kemp,
david.kemp@abdn.ac.uk

Citation:
Kemp, D. B. (2016), Optimizing significance testing of astronomical forcing in cyclostratigraphy, *Paleoceanography*, *31*, 1516–1531, doi:10.1002/2016PA002963.

Received 13 APR 2016
Accepted 8 NOV 2016
Accepted article online 15 NOV 2016
Published online 5 DEC 2016

Optimizing significance testing of astronomical forcing in cyclostratigraphy

David B. Kemp[1]

[1]School of Geosciences, University of Aberdeen, Aberdeen, UK

Abstract The recognition of astronomically forced (Milankovitch) climate cycles in geological archives marked a major advance in Earth science, revealing a heartbeat within the climate system of general importance and key utility. Power spectral analysis is the primary tool used to facilitate identification of astronomical cycles in stratigraphic data, but commonly employed methods for testing the statistical significance of relatively high narrow-band variance of potential astronomical origin in spectra have been criticized for inadequately balancing the respective probabilities of type I (false positive) and type II (false negative) errors. This has led to suggestions that the importance of astronomical forcing in Earth history is overstated. It can be readily demonstrated, however, that the imperfect nature of the stratigraphic record and the quasiperiodicity of astronomical cycles sets an upper limit on the attainable significance of astronomical signals. Optimized significance testing is that which minimizes the combined probability of type I and type II errors. Numerical simulations of stratigraphically preserved astronomical signals suggest that optimum significance levels at which to reject a null hypothesis of no astronomical forcing are between 0.01 and 0.001 (i.e., 99–99.9% confidence level). This is lower than commonly employed in the literature (90–99% confidence levels). Nevertheless, in consonance with the emergent view from other scientific disciplines, fixed-value null hypothesis significance testing of power spectra is implicitly ill suited to demonstrating astronomical forcing, and the use of spectral analysis remains a difficult and subjective endeavor in the absence of additional supporting evidence.

1. Introduction

Astronomically forced (Milankovitch) climate cycles have been implicated as a significant driver of past climate variability, and their recognition in strata has proven invaluable for the development of high-resolution chronologies [e.g., *Hays et al*., 1976; *Zachos et al*., 2001; *Pälike et al*., 2006; *Gradstein et al*., 2012]. Spectral analysis is the primary method for assessing the presence of periodic astronomical cycles in (cyclo)stratigraphic data [e.g., *Weedon*, 2003; *Vaughan et al*., 2011; *Meyers*, 2012]. Crucial to the utility of the method is the application of an appropriate statistical test that measures the significance of narrow-band peaks in power spectra ostensibly attributable to astronomical forcing against the wide-band spectrum of nonperiodic variance ubiquitous in cyclostratigraphic data. Significance testing methods rely on fitting an appropriate function to the spectrum that provides a good match to the morphology of this wide-band "noise." The significance of spectral peaks is then determined based on their power relative to the expected (chi-square) distribution of power around this idealized noise continuum [*Mann and Lees*, 1996; *Meyers*, 2012]. The choice of continuum model to fit is, ideally, motivated by the known spectral properties of climate variability [*Hasselmann*, 1976; *Mann and Lees*, 1996; *Meyers*, 2012] and has led to widespread application of first-order autoregressive (AR1) models that in many cases (though by no means all) approximate rather well the statistical characteristics of cyclostratigraphic time series of climate proxies [*Mann and Lees*, 1996; *Weedon*, 2003].

As in other areas of science, rejection of a null hypothesis of no astronomical forcing is dependent on a spectral peak exceeding a critical value drawn from the theoretical probability distribution of the continuum model. Typically, in cyclostratigraphy, critical significance levels (α) are set at ≥0.01 (i.e., ≤99% confidence level). In other words, the critical value chosen has a probability of α of being matched or exceeded by a spectral peak. This approach has been criticized because in spectra calculated from data sets of length N (and assuming no smoothing or zero padding of the spectrum) there are $N_f = N/2$ independent frequencies and hence an expectation of $\sim \alpha N_f$ significant peaks. This essentially guarantees rejection of the null hypothesis in time series of ≥200 points [*Vaughan et al*., 2011]. In the limiting case of small α and large N_f, an appropriate correction to α that accounts for the fact that significance is effectively being tested multiple (i.e., N_f) times is to

employ a critical significance value of $\alpha' = \alpha/N_f$, commonly termed the Bonferroni correction [*Vaughan et al.*, 2011; *Mudelsee*, 2010]. In widely employed methods of spectral analysis the calculated frequencies are not independent, but the argument stands that inadequately stringent α will lead to erroneous rejection of the null hypothesis, i.e., a type I error (false positive). This led *Vaughan et al.* [2011] to suggest that the importance of astronomical forcing in Earth history has been overstated. *Hilgen et al.* [2014] countered that such stringent significance testing applied to what is demonstrably a noisy and imperfect stratigraphic record risked erroneous dismissal of astronomical forcing when it is, in fact, present, i.e., type II errors (false negatives). They showed how high statistical significance of astronomical cyclicity is not always attained in real data sets, despite strong independent evidence for its presence and importance.

Running parallel to this debate, fixed value null hypothesis significance testing in science has itself been criticized for the arbitrary way in which critical significance levels are chosen, and for the lack of information such tests provide on either the importance or ultimate meaning of phenomena they are designed to validate [e.g., *Carver*, 1978, 1993; see also *Wasserstein and Lazar*, 2016]. Such criticisms apply to significance testing of astronomical forcing. The goal of significance testing should be to minimize the probability (or where appropriate, the cost) of making mistakes; hence, an optimized significance level is one that minimizes the combined probability of both type I and type II errors [*Mudge et al.*, 2012]. In this study, numerical simulations of stratigraphically preserved astronomical signals are used to illustrate this point. Optimized critical significance levels ($\alpha_{optimal}$) and associated confidence levels for stratigraphically preserved astronomical signals are quantified, elucidating the caveats inherent in fixed value null hypothesis significance testing and consequent validation of astronomical forcing.

2. Modeling Approach

One of the primary issues surrounding the validation of astronomical forcing in the stratigraphic record is uncertainty in the expected character and expression of that forcing. Quantifying significance in mathematical terms is moot if we are unsure of what "strong" evidence of astronomical forcing should actually look like. Beyond ~50 Ma, the problem is perhaps more acute owing to the lack of an accurate astronomical solution [*Laskar et al.*, 2004; see also *Waltham*, 2015]. A major source of uncertainty with a known capability for limiting the power and significance of astronomical cycles arises from the imperfect nature of the stratigraphic record. Sedimentation is typically a discrete, unsteady process, and rates of sedimentation in a succession vary [*Sadler*, 1981; *Tipper*, 1983; *Sadler and Strauss*, 1990; *Herbert*, 1994; *Weedon,* 2003; *Huybers and Wunsch*, 2004; *Meyers*, 2012; *Kemp and Sexton*, 2014]. Preserved astronomical cycles of constant (or at least near-constant) duration hence vary in thickness, and the signal is "jittered" [e.g., *Huybers and Wunsch*, 2004]. The consequence of this jittering is that spectral peaks pertaining to astronomical cyclicity are smeared in a manner dependent on the strength of jitter and the statistical characteristics of the process [*Moore and Thomson*, 1991; *Huybers and Wunsch*, 2004; *Mudelsee et al.*, 2009; *Taylor Perron and Huybers*, 2009; *Rhines and Huybers*, 2011]. *Huybers and Wunsch* [2004] derived from empirical observation a general model of marine sedimentation rate variability in oxic environments that approximates a stochastic first-order autoregressive process with a power spectrum with the general form:

$$\Phi(f) = \frac{1}{f^2 + f_0^2} \tag{1}$$

where f is frequency and $1/f_0$ is the decorrelation time (i.e., memory) of the process [*Huybers and Wunsch*, 2004] (Figure 1a). A sedimentation rate record, R, can be reconstructed from this via

$$R(t) = \Im^{-1}\left\{\hat{\omega}\cdot\sqrt{\Phi}\right\} \tag{2}$$

where t is the time step, $\Im^{-1}$ is the inverse Fourier transform, and $\hat{\omega}$ is the Fourier transform of a Gaussian white noise process [see also *Kemp and Sexton*, 2014]. Analysis of sedimentation rates in Pleistocene ocean cores by *Huybers and Wunsch* [2004], determined independently of any astronomical time control, suggests that $1/f_0$ is close to ~100 kyr (Figure 1a), and that in a majority of records sedimentation rates are significantly not normally distributed (Figure 1b) [*Kemp and Sexton*, 2014].

These equations are used in this study to model unsteady sedimentation and the consequent effects this has on the statistical significance of stratigraphically preserved astronomical cycles (Figure 2a). Skew can be

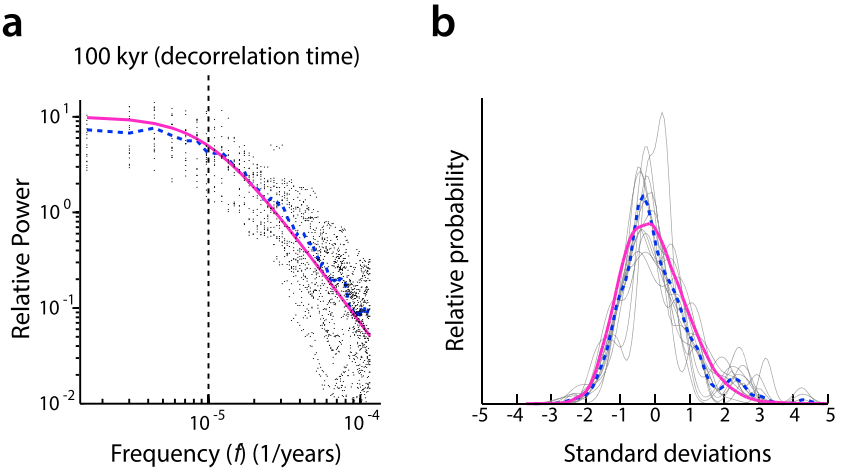


Figure 1. (a) Power spectrum of sedimentation rates from Pleistocene ocean sediment cores. Blue dashed line is the mean spectrum. Purple line is the fit to these estimates using equation (1) of main text, with decorrelation time ($1/f_0$) 100 kyr. Redrawn from *Huybers and Wunsch* [2004]. (b) Normalized (unit standard deviation and zero mean) kernel density plots showing distribution of sedimentation rates in Pleistocene records (grey curves). Mean skewness (Pearson second coefficient of skewness, calculated as $3 \times$ ((mean rate − median rate)/standard deviation)) in these records is 0.35. Blue dashed curve shows average kernel density for the records. Purple curve is the average kernel density plot of 100 sedimentation rate models constructed using equations (1) and (2) of main text, with skewness of 0.3 ± 0.3 (1σ). Pleistocene records analyzed are as follows: Deep Sea Drilling Project Site 607, Cores md900963, PC18, PC72, and Ocean Drilling Program Sites 663, 664, 677, 846, 849, 925, 980, and 982 (data from *Huybers* [2007]).

introduced in a modeled sedimentation rate record by transforming R with a prescribed exponent, implemented here to provide sedimentation rates with a skewness matching that observed in the studied records (Figure 1b). With some further overall normalization, the mean rate and absolute unsteadiness of the process can be set, with jitter (J) defined here as the coefficient of variation at a particular timescale (i.e., standard deviation of rates divided by the mean rate measured over a particular time span; see *Kemp and Sexton* [2014] and Figure 2a). The analysis of *Kemp and Sexton* [2014] suggests that marine (oceanic and epicontinental) strata typically have J values within the range 0.22 to 0.52 (mean 0.33; $N = 19$) when measured at 20–40 kyr time spans in records of ~1 million year duration. Integrating a record of sedimentation rates produced in this way provides the accumulated stratigraphic profile with increasing time. An astronomical insolation signal [*Laskar et al*., 2004] (Figure 2c) can be mapped to the stratigraphic profile to provide the stratigraphically preserved (jittered) record of the signal (Figure 2e). In this study, astronomical insolation solutions dominated by ~21 kyr precession cycles are used (Figures 2c and 2d). Finally, the standard cyclostratigraphic workflow of sampling the record at evenly spaced intervals can be replicated by linear interpolation of the preserved signal at equally spaced stratigraphic height increments. For the purposes of this study, sedimentation rate defines only the expected frequency of a periodic signal in a given record, dependent on the chosen sampling resolution. Negative rates were not included to ensure that all records were stratigraphically complete at the scale of the sampling resolution.

Variance in cyclostratigraphic data (and indeed time series of many other natural processes) is typically weighted toward low frequencies, generating a red noise-like spectrum [e.g., *Mann and Lees*, 1996]. Such a continuum of variance in proxy records spanning millennial to million year scales is predictable from knowledge of the behavior of climate over these scales [*Pelletier*, 1998; *Wunsch*, 2003; *Huybers and Curry*, 2006; *Meyers*, 2012]. The theoretically perfect recording of a climate signal in the stratigraphic record would reflect this variability and hence generate a red noise-like spectrum, with superimposed periodicity at astronomical scales [*Huybers and Curry*, 2006]. In reality, analytical errors, proxy limitations, and imperfect stratigraphic recording of climate complicate the spectrum, and thus the clarity of both the continuum and periodic forcing. Thus, the challenge of cyclostratigraphy is to resolve a periodic signal (or set of signals) attributable to astronomical forcing from a complex noise spectrum of uncertain relative variance that likely only tenuously reflects the underlying dynamics of climate. A large amount of previous research indicates that the continuum in cyclostratigraphic records commonly approximates an AR1 (red noise) process. For the

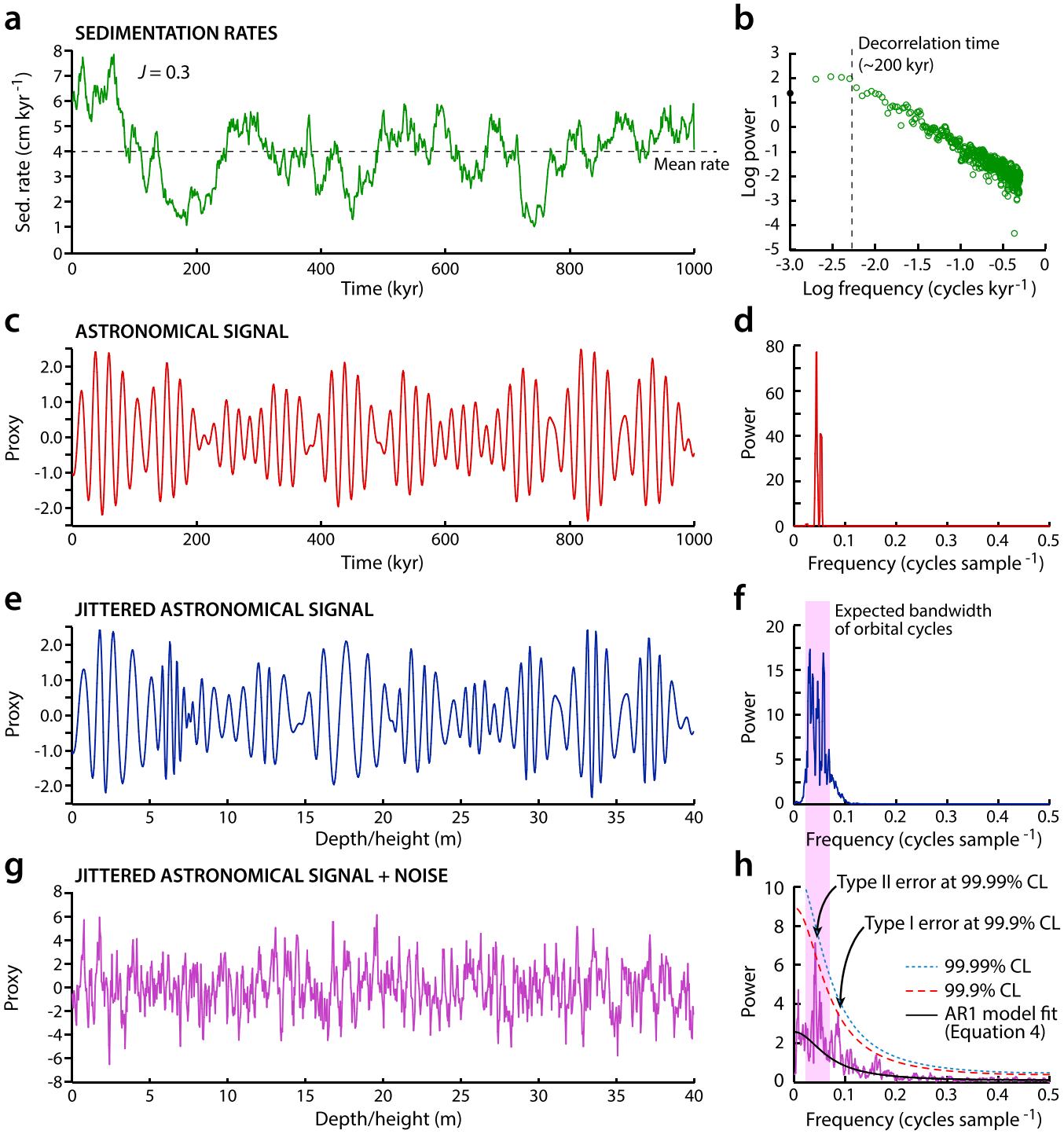


Figure 2. (a) Sedimentation rate record produced from equations (1) and (2) of main text, with $1/f_0 = 200$ kyr and $J = 0.3$ at the 25 kyr scale. The jitter value is calculated as the standard deviation of averaged nonoverlapping 25 kyr intervals divided by the mean sedimentation rate (in this case 4 cm kyr^{-1}; see also *Kemp and Sexton* [2014]). (b) Power spectrum of the sedimentation rate record highlighting the decorrelation time of ~200 kyr, which is approximately the period at which the spectrum flattens. (c) Astronomical insolation signal from 3 to 4 Ma at 40°N latitude from *Laskar et al.* [2004] and (d) accompanying power spectrum showing characteristic dual-peak precession cyclicity (~23 and ~19 kyr period; mean ~21 kyr). (e) Astronomical signal in Figure 2c translated to stratigraphic depth/height using the sedimentation rate model in Figure 2a. The signal is jittered: note how cycles become stretched and squeezed. (f) Power spectrum of the jittered astronomical signal in Figure 2e, emphasizing how maximum power is weakened and how the power smeared across a wider frequency range. Note also how the dual peak response of precession (Figure 2d) is lost. (g) Jittered astronomical signal as in e with red noise added ($\rho = 0.6$). (h) Power spectrum of signal in Figure 2g showing how the jittered spectral peak related to astronomical forcing is still resolvable but with significance that does not exceed 99.99% confidence level (CL). Note also how peaks related to the red noise can reach significance levels >99.9% CL. See main text for details of signal construction and significance testing.

purposes of this study, therefore, the noise continuum can be replicated by adding to the insolation signal an AR1 process constructed using

$$r(t) = \rho r(t-1) + \omega_t \quad (3)$$

where r is the AR1 noise, $0 < \rho < 1$ is the lag-1 autocorrelation coefficient, and ω is a Gaussian white noise process (Figure 2g). *Wunsch* [2004] argued that the fraction of variance within the spectrum of Quaternary climate change attributable to astronomical forcing is unlikely to be >20%. *Meyers et al.* [2008] contested that individually, periodic components related to astronomical forcing comprise 28–41% of total signal variance in the Vostok ice core record of climate change spanning the past ~400 ka. In pre-Pleistocene sedimentary records, where proxy and stratigraphic uncertainties are greater, the figure is uncertain.

Spectral analysis and significance testing was conducted following standard methods. Multitaper spectral analysis (time-bandwidth product = 2.5 and four tapers) was used throughout, and significance testing was based on fitting a model to the spectrum that characterizes the gross morphology of the underlying AR1 noise spectrum [*Mann and Lees*, 1996]:

$$S_{AR1}(f) = S_0 \frac{1-\rho^2}{1-2\rho\ \cos\pi\left(\frac{f}{f_N}\right)+\rho^2} \quad (4)$$

where f_N is the Nyquist frequency (i.e., the highest resolvable frequency in the spectrum) and S_0 is average power of the spectrum. Values for S_0 and ρ were evaluated using least squares fitting (Figure 2h). Median smoothing of spectra prior to model fitting was advocated by *Mann and Lees* [1996] as a way of minimizing the bias that results from fitting a noise model in the presence of strong spectral peaks, which would otherwise have the effect of raising the fitted continuum above the actual noise background (however, see *Meyers* [2012]). Median smoothing was implemented in this study following the method of *Mann and Lees* [1996]. The smoothing window length was set by the expected bandwidth of the spectral peak assessed from modeling of red noise-free jittered astronomical signals (e.g., Figure 2f), up to a maximum width of $f_N/4$ [*Mann and Lees*, 1996]. This approach ensured that the best possible model fits were used for each simulation.

3. Results

3.1. Spectral Effects of Unsteady Sedimentation

Using the above modeling and analysis approach, the spectral effects of jitter on the preservation of astronomical forcing in strata can be investigated. Figure 3 highlights the effects of jitter on the spectra of astronomical signals encoded stratigraphically for a variety of jitter strengths and sampling resolutions, with sampling resolution defined as the resolution at which cycles are sampled, from ~160 samples cycle^{-1} to ~5 samples cycle^{-1}. Astronomical signals are from the *Laskar et al.* [2004] astronomical insolation solution (0–5 Ma; mean insolation at 40°N), with the precise duration of the record dependent on the desired sampling resolution. Each simulation was 1000 samples long. As noted above, the astronomical signals used in the modeling are dominated by ~21 kyr precession cycles (Figures 2c and 2d), with negligible contribution from eccentricity and obliquity. Thus, testing was carried out here and throughout this study solely on precession cycles. To provide statistically stable data, each result in Figure 3 is based on the analysis of 5000 numerical simulations of red noise-free jittered astronomical signals. In these simulations, the decorrelation time term in equation (1) ($1/f_0$) was uniformly randomly chosen between 40 kyr and 400 kyr, and skewness (Pearson's second skewness coefficient) was $\sim 0.3 \pm 0.3$ (1σ) (note that there is no correlation between α_{optimal} and $1/f_0$ or skewness).

The results in Figure 3 emphasize how the departure from a linear time-depth relationship causes spectral power attributable to astronomical precession cycles to be smeared over a wider frequency range (compare also Figures 2d and 2f). The greater the jitter, the more the power is smeared (Figure 3b). The variance of the signal is conserved, and hence, the power of the maximum spectral peak will decrease with increasing jitter (Figure 3a). The effects are also frequency dependent, with power reduced by a greater amount for low-resolution (i.e., high-frequency) cycles (Figure 3a). For example, for a weakly jittered astronomical signal ($J = 0.1$) with a sampling resolution of ~5 samples cycle^{-1}, the average power of the largest spectral peak associated with this cyclicity (from 5000 simulations) is <50% of the power of the largest peak in an unjittered signal (Figure 3a). With higher sampling resolution (i.e., more samples per cycle), the effects of jitter are

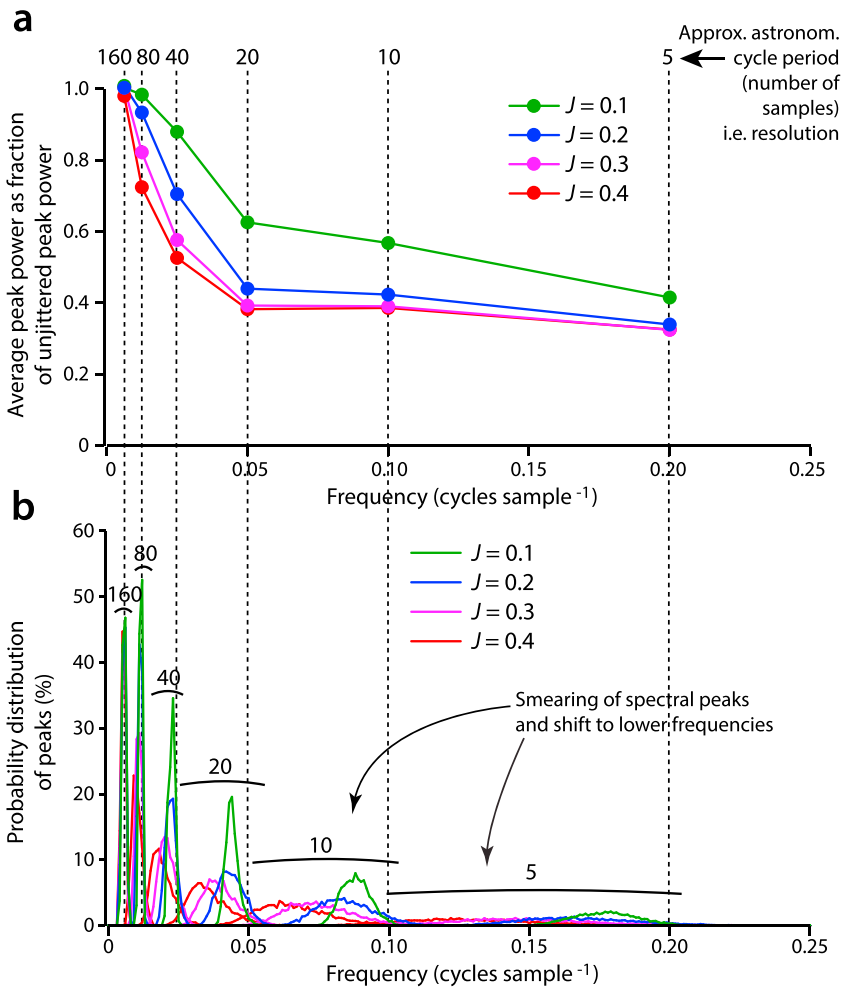


Figure 3. (a) Plot highlighting the effects of jitter on the power attained by spectral peaks related to astronomical precession cycles. Values plotted are the average maximum peak power values in 5000 simulations of jittered, red noise-free astronomical signals, expressed as a fraction of the maximum peak power attained by an unjittered astronomical signal (see main text for details of parameters used). Densely sampled (i.e., high-resolution) astronomical signals are more robust to the effects of jitter. Equally, even weak jitter (i.e., $J = 0.1$) has a significant impact on the attainable power of an astronomical signal at lower sampling resolutions (particularly ~20 samples cycle^{-1} and lower). Increasing jitter has progressively less effect on attainable power, with the largest drop in power occurring between $J = 0.1$ and $J = 0.2$. (b) Plot showing the probability distribution of the strongest spectral peaks in the signals analyzed in Figure 3a. The plot shows how increasing jitter causes peak maxima to shift toward lower frequencies and become distributed over a wider frequency range.

reduced. If the astronomical signal is sampled at a resolution of ~160 samples cycle^{-1}, then the peak power is very close to that expected of an unjittered signal, even if jitter is high (i.e., $J = 0.4$; Figure 3a). Another effect of jitter is a shift in the position of expected peaks to lower frequencies, with the magnitude of the shift frequency dependent (Figure 3b). Ostensibly, this might be expected owing to the use of a sedimentation rate model with nonnormally distributed (skewed) rates. Analysis of models with normally distributed sedimentation rates indicates, however, that the effect arises largely from the reduced resolution (and hence increased peak stability) of the spectrum at lower frequencies. Thus, in a jittered record variance is shifted to both lower and higher frequencies, but variance shifted to higher frequencies will be more smeared. Peaks occurring at higher frequencies will therefore always tend to be weaker than peaks occurring at lower frequencies, and Figure 3b specifically compiles the location of the strongest peaks in the spectrum.

3.2. Quantifying Type I and Type II Errors and Optimized Critical Significance Levels ($\alpha_{optimal}$)

Monte Carlo style analysis using multiple numerical simulations of stratigraphically preserved noisy astronomical signals provides a viable way of determining the probability that a stratigraphic record will preserve

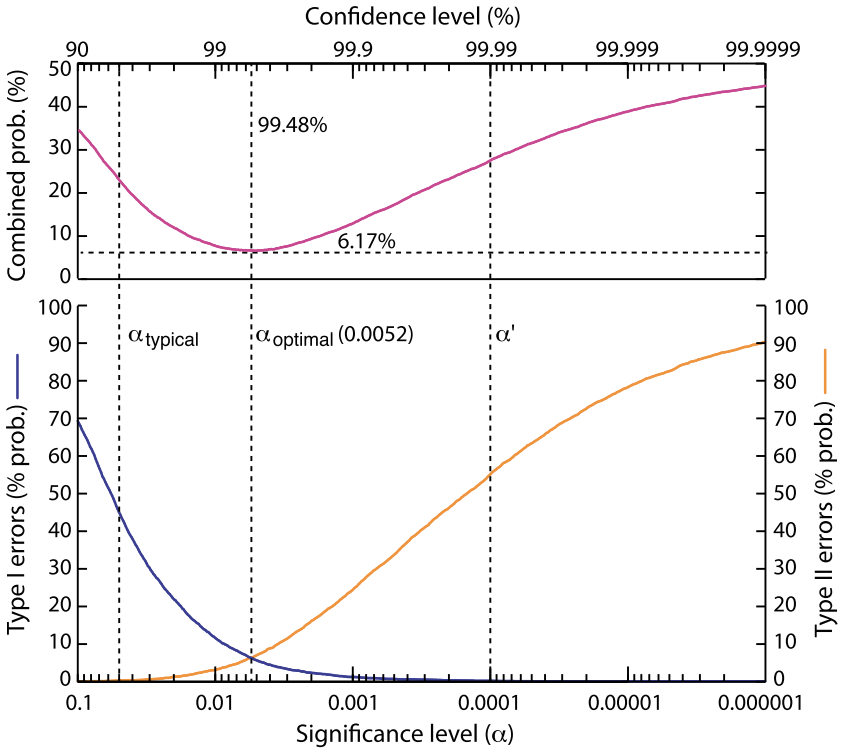


Figure 4. Graph showing the type I, type II, and combined error probabilities in simulations of stratigraphically preserved astronomical signals. Type II probabilities were calculated from 5000 simulations of jittered astronomical signals ($J = 0.2$, data length is 1000 points, and sampling resolution is ~20 samples per precession cycle), with red noise added to each signal using $\rho = 0.6$ and with variance that is 75% of the total variance of the signal (i.e., astronomical signal is 25% of the total signal variance). Type I errors were calculated from 5000 simulations of pure red noise. See main text for further details. The combined probability of making an error (purple line in upper plot) is the average of the type I and type II error probabilities. The graph highlights how at a critical significance level typically used for cyclostratigraphic studies ($\alpha_{typical} = 0.05$), the probability of a type I error is high (>40%), but at higher critical significance levels (i.e., $\alpha < 0.01$) the probability of type II errors increases sharply (note the log scale). The optimum critical significance level to employ ($\alpha_{optimal}$) is the value at which the combined error is minimized, i.e., 0.0052 (99.48% confidence level). The minimum combined error probability is 6.17%.

statistically significant cycles when deposited under a given set of conditions (i.e., specific noise level, jitter level, etc.). To exemplify the approach, Figure 4 shows the key statistical results from an analysis of 5000 jittered stratigraphic records (each 1000 samples long) that encode an astronomical insolation signal sampled at ~20 samples cycle^{-1}. J was set at 0.2, a conservative value relative to that expected in deep-sea successions (see section 2) [*Kemp and Sexton*, 2014]. Unique realizations of red noise ($\rho = 0.6$; equation (3)) were added to each jittered astronomical signal. The astronomical signal variance in each simulation was set as 25% of the total variance of the combined signal and noise: broadly similar to the median expected strength of astronomical forcing determined empirically by *Wunsch* [2004] and *Meyers et al.* [2008] (see section 2). The choice of sampling interval (~20 samples cycle^{-1}) is close to the typical resolution of cyclostratigraphic studies. For instance, in deep-sea and epicontinental marine successions compacted sedimentation rates typically vary between 1 and 10 cm kyr^{-1} [*Huybers and Wunsch*, 2004] meaning that an astronomical insolation signal dominated by ~21 kyr precession cycles would preserve cycles ~0.2 to ~2 m thick. Sampling rates of such successions commonly vary between 1 and 10 cm, hence yielding a sampling resolution of ~1 sample kyr^{-1} or ~20 samples cycle^{-1}. As in Figure 3, $1/f_0$ was uniformly randomly chosen between 40 and 400 kyr and skewness was ~0.3 ± 0.3 (1σ).

Type II errors quantified in these simulations are instances when no spectral peak within the expected bandwidth of the astronomical precession cyclicity matches or exceeds the significance level, α (see also Figure 2h). The bandwidth within which a peak is expected for a given jitter strength and sampling resolution is determined as the frequency range over which 99% of peak maxima are contained in 5000 simulations of

jittered red noise-free astronomical signals (Figure 3b). The type II error probabilities shown in Figure 4 (orange line) are simply the percentage of the 5000 noisy jittered simulations that have type II errors at a given significance level. Type I error probabilities are quantified in the same way from analysis of 5000 simulations of pure red noise with the same ρ value (blue line; Figure 4). A type I error is recorded if at least one spectral peak not related to the astronomical forcing occurs that matches or exceeds the given significance level. Type I and type II error probabilities are quantified at significance levels between 0.1 (90% confidence level) and 0.000001 (99.9999% confidence level) following standard methods (i.e., as implemented in the widely used SSA-MTM toolkit; http://www.atmos.ucla.edu/tcd/ssa/). Assuming an equal a priori probability of rejecting or accepting the null hypothesis of no astronomical forcing, the combined error probability is the average of the type I and type II error probabilities [*Mudge et al.,* 2012] (purple line; Figure 4).

In Figure 4, the minimum combined error probability (6.17%) occurs at a significance level of 0.0052; i.e., 99.48% confidence level (Figure 4). Hence, this can be regarded as the optimized critical significance level ($\alpha_{optimal}$) to employ to minimize the chance of making a mistake in hypothesis testing. Notably, the combined error probability is considerably greater than $\alpha_{optimal}$ (0.0617 compared to 0.0052). At significance levels $>$0.01, the probability of a type I error rises sharply (Figure 4). Thus, in this modeling scenario, setting a critical significance level at 0.05, as is common in cyclostratigraphy ($\alpha_{typical}$; Figure 4), results in a 44.66% chance of a type I error (22.44% combined type I and type II error; Figure 4). Equally, however, $\alpha_{optimal}$ is higher than the raw conservative multiple-test-corrected significance level of 0.05 (see section 1). This raw level (α') is at 0.0001 since the spectra contain 500 frequencies (0.05/500; see section 1). Setting the critical significance level at this value means that there is a 45.36% chance that a peak associated with the astronomical signal is significant (i.e., 54.64% type II error probability; Figure 4), despite the corresponding type I error probability being 0.14% (Figure 4). Indeed, between $\alpha_{optimal}$ and α', the type I error probability reduces by just 5.17%, but the type II error probability increases by nearly 50%. It is also noteworthy that $\alpha_{optimal}$ is greater than the false alarm level advocated by *Thompson* [1990]; calculated simply as 1/*N* (i.e., 0.001 in the 1000 point data set analyzed; Figure 4).

Repetition of the modeling demonstrates that the method yields results that are statistically stable, with repetition of the above simulations yielding the same $\alpha_{optimal}$ on each occasion. A further key advantage of modeling type I and type II errors in this way is that results are largely independent of the precise characteristics of the noise used and the model fit quality, since any overestimation/underestimation of type I errors will be largely balanced by a corresponding underestimation/overestimation of type II errors. However, the use of noisy signals to quantify type II errors analytically means that type II error probabilities are potentially biased by the presence of type I errors. In other words, a type II error will be avoided if a false positive peak, not related to the cyclicity in the signal, exceeds the significance level within the expected frequency range of the astronomical cyclicity, even if no peaks attributable to the astronomical signal exceeds this level. Analysis of the power attained by both pure red noise and noisy astronomically forced signals, however, suggests that such bias is unlikely. This is because for noisy jittered astronomical signals, the strongest peak in the spectrum is likely to occur within the frequency range of the astronomical signal, and this is likely to be related to the astronomical signal and not noise because the power of this peak typically exceeds the strongest peak in a spectrum of pure red noise in the same frequency range.

3.3. Effects of Jitter on $\alpha_{optimal}$

To explore further the effects of jitter on the results presented in Figure 4, Figure 5 highlights the type I, type II, and combined error probabilities for simulations with *J* varied from 0 to 0.4. Other model parameters remain the same as those used to generate Figure 4. The results show that $\alpha_{optimal}$ is relatively invariant at jitter strengths $\geq$0.1, with $\alpha_{optimal}$ clustering between 0.01 and 0.001 (99%–99.9% confidence level; Figure 5). The results also demonstrate that a perfectly recorded astronomical signal with no sedimentation rate unsteadiness ($J = 0$) does not have infinitely high $\alpha_{optimal}$ (i.e., zero type II error probability) but is instead constrained. The minimum combined error probability for $J = 0$ is 0.07%, and this occurs at a significance level of ~0.00012 (99.987% confidence level; Figure 5). This is because the insolation signal being tested is quasiperiodic. The precession cycles in astronomical insolation have a multiple peak spectral signature, with the strongest peaks at periods of ~23 kyr and ~19 kyr (Figure 2d). In effect, therefore, the tested astronomical signal is already "jittered" in the sense that precession cycles are not of constant period.

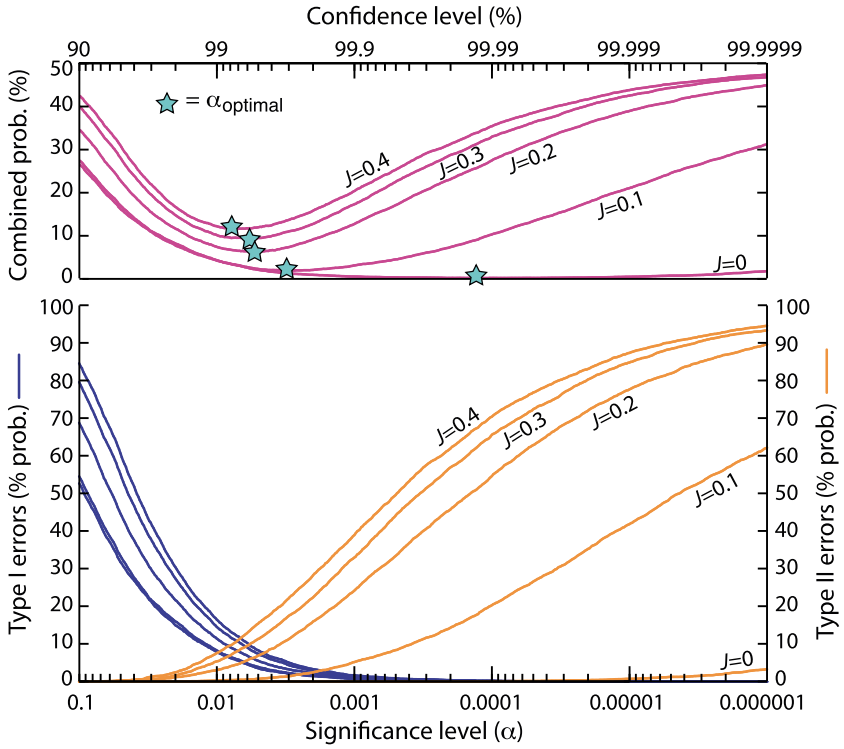


Figure 5. Graph showing the type I, type II, and combined probability errors in simulations of stratigraphically preserved astronomical signals with varying jitter. Note how even under conditions of no jitter ($J = 0$), $\alpha_{optimal}$ is constrained to >0.0001 (<99.99% confidence level). This is because the astronomical precession cycles being tested are quasiperiodic (i.e., they are not of constant period; see main text for details). Note also how at jitter strengths ≥0.1 there is relatively little effect on type II error probability and $\alpha_{optimal}$, with $\alpha_{optimal}$ clustered between 0.01 and 0.001 (99% and 99.9% confidence level).

3.4. Effects of Noise Strength on $\alpha_{optimal}$

The experiments in Figures 4 and 5 consider a specific scenario where the insolation signal represents 25% of the total variance of the signal, and clearly, a key limitation on the ability to resolve astronomical cycles in cyclostratigraphic data is the strength of the astronomical signal relative to the background noise. To explore the effects of signal to noise ratio, experiments have been conducted using astronomical signals with variance that is between 5% and 45% of the total signal variance (Figure 6). This range is in-line with the range defined by *Wunsch* [2004] and *Meyers et al.* [2008]. The very nature of cyclostratigraphic investigations means that the precise variance of signal and noise is not cognizable from a record a priori and represents a key unknown in any cyclostratigraphic study. At a signal strength of 5%, cyclicity is only weakly resolvable, and type II error probabilities are correspondingly high, with $\alpha_{optimal} \sim 0.032$ (96.8% confidence level). The combined error probability at $\alpha_{optimal}$ is also high at 33.3%. Combined error probabilities fall as the signal strength is increased, and $\alpha_{optimal}$ decreases uniformly to a minimum of 0.0005 (~99.95%) when the signal level is 45%.

3.5. Effects of Nonrandom Jitter on $\alpha_{optimal}$

Analysis of sedimentation rate variability in deep-sea successions highlights how astronomical forcing can itself sometimes control sedimentation rates [e.g., *Mix et al.*, 1995; *Guyodo and Channell*, 2002]. Thus, jitter may not be purely random as modeled above but exhibit cyclicity caused by the same astronomical cycle (s) preserved in proxy data. This makes sense given the possibility that sediments with different physical characteristics and sources can be deposited under the varying climatic conditions within an astronomical cycle. Such "systematic," rather than random, jitter may improve the prediction of both the strength and location of spectral peaks. Figure 7 shows the results of experiments using cyclic rather than random jitter in the modeling scenario of Figure 4. The same overall jitter level ($J = 0.2$) is used, allowing direct comparison of the two models. Using a purely cyclic model of sedimentation rates results in $\alpha_{optimal}$ of 0.0148 (98.52% confidence

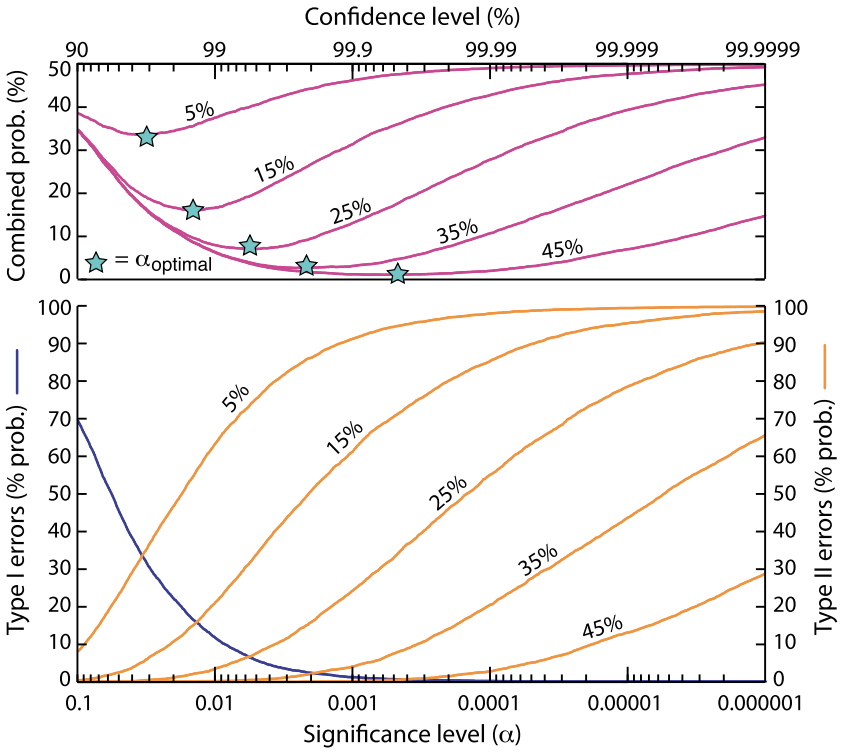


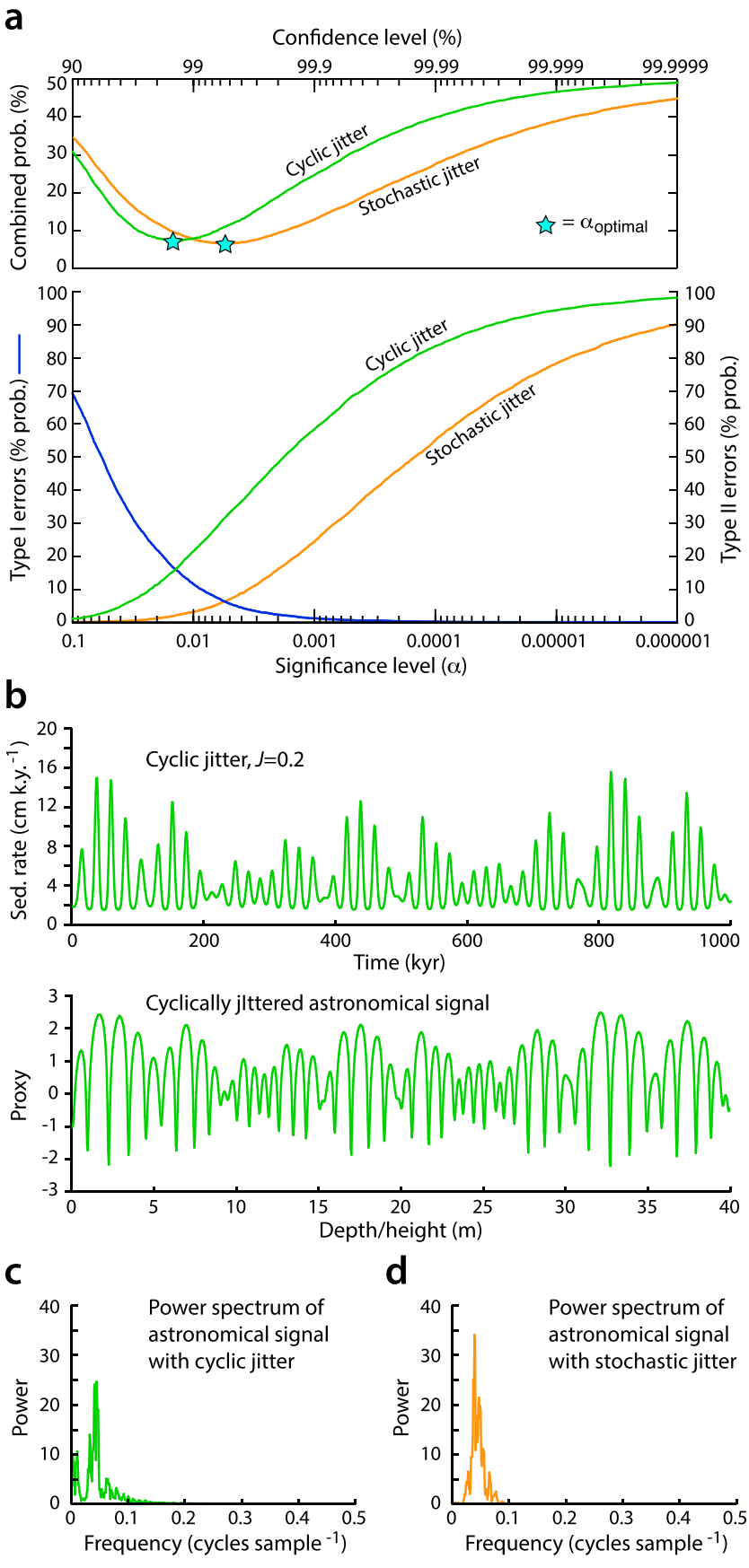
Figure 6. Graph showing the type I, type II, and combined probability errors in simulations of stratigraphically preserved astronomical signals with varying signal strength levels (expressed as a percentage of the total variance of the signal) from 5% to 45%. Note how a stronger astronomical signal will reduce the probability of a type II error and hence reduce the overall combined error probability and increase $\alpha_{optimal}$.

level; Figure 7a). This is higher than the $\alpha_{optimal}$ attained in the stochastic model (0.0052, 99.48% confidence level; see also Figure 4).

For this particular modeling scenario, therefore, the attainable statistical significance of astronomical precession cycles is lower under conditions of cyclic jitter. This is because cyclic variations in sedimentation rates generate less sinusoidal, and more cuspate shaped cycles (Figure 7b), and this leads to significant leakage of spectral power to higher-frequency harmonics (Figure 7c) [*Schiffelbein and Dorman*, 1986; *Meyers et al*., 2001; *Weedon*, 2003]. To ensure the same overall jitter ($J = 0.2$) between the cyclic and stochastic models, the range in sedimentation rates is also higher in the cyclic model. Equally, because the amplitude of the insolation signal is modulated by ~100 kyr eccentricity cycles, some power is also transferred to lower frequencies (this is known as rectification; see *Weedon* [2003] for details; Figure 7c). For comparison, the spectrum of an astronomical signal jittered using the random sedimentation rate model is shown in Figure 7d. This spectrum shows higher maximum power and relatively less leakage of power to lower and higher frequencies (Figure 7d). The result in Figure 7 is dependent in part on the precise nature of the signal distortion caused by cyclic sedimentation rate variation, and further appraisal of this is beyond the scope of this study. Nevertheless, the experiments help to emphasize that any departure from a linear time-depth relationship will result in the statistical significance of spectral peaks being limited.

3.6. Likely Bounds on $\alpha_{optimal}$ in Geologic Data

To provide a more holistic assessment of the precise effects of jitter, sampling resolution, and also noise type, Figure 8 shows the $\alpha_{optimal}$ values and combined error probabilities calculated from a range of different simulations. These were designed to mimic geologic conditions and hence real cyclostratigraphic data, encompassing a wide gamut of jitter levels, noise characteristics, and cycle frequencies. Sampling resolution is varied from ~5 to ~160 samples cycle^{-1}, and random jitter models are used, with J varied from 0 to 0.4. Three different styles of red noise are considered, with $\rho = 0.3$, 0.6, and 0.9. These ρ values reflect a range


a
Confidence level (%)
90
99
99.9
99.99
99.999
99.9999
Combined prob. (%)
Cyclic jitter
Stochastic jitter
= αoptimal
Type I errors (% prob.)
Type II errors (% prob.)
Significance level (α)
0.1
0.01
0.001
0.0001
0.00001
0.000001
b
Cyclic jitter, J=0.2
Sed. rate (cm k.y.-1)
Time (kyr)
Cyclically jittered astronomical signal
Proxy
Depth/height (m)
c
Power spectrum of astronomical signal with cyclic jitter
Power
Frequency (cycles sample-1)
d
Power spectrum of astronomical signal with stochastic jitter
Power
Frequency (cycles sample-1)

of noise spectrum morphologies. Red noise with $\rho = 0.3$ has a relatively flat spectrum, with variance distributed fairly uniformly across all frequencies. In contrast, red noise with $\rho = 0.9$ has strongly nonuniformly weighted variance, with variance concentrated at the lowest frequencies.

The results demonstrate how the style of red noise exerts a key influence on $\alpha_{optimal}$ (Figure 8). This is because the significance of a spectral peak at a given frequency is strongly dependent on the local signal to noise ratio at that frequency. Thus, in the modeling scenarios with $\rho = 0.9$, the significance of spectral peaks at long wavelengths (i.e., 40–160 samples cycle$^{-1)}$ is reduced (and $\alpha_{optimal}$ is correspondingly high) since it is at these low frequencies where the majority of the noise variance is concentrated (Figure 8). This is despite the fact that such high-resolution cycles are typically better resolved and more resilient to jitter (Figure 3). In contrast, in scenarios with $\rho = 0.3$, $\alpha_{optimal}$ is lower for low-frequency (i.e., high-resolution) cycles compared to high-frequency cycles (Figure 8).

Taken together, $\alpha_{optimal}$ is between 0.01 and 0.0001 (99% to 99.99% confidence level) in the majority (64%) of modeling scenarios (Figure 8). $\alpha_{optimal}$ is only >0.01 ($<$99% confidence level) in 11% of scenarios. For signals with jitter strengths that are likely most typical of real successions (i.e., $\geq$0.2) [*Kemp and Sexton*, 2014], $\alpha_{optimal}$ clusters between 0.01 and 0.001 (99% to 99.9% confidence level) in 59% of modeling scenarios (Figure 7). Combined error probabilities are typically higher in scenarios with higher jitter, and range from 0 to 17.69% across all the simulations (Figure 7). For the geologically most plausible modeling scenarios with $J \geq 0.2$, average combined error probabilities are 3.87%, 5.28%, and 7.12% for $\rho = 0.3$, 0.6, and 0.9, respectively. Note that in a few simulations (typically those with $J = 0$ or 0.1), no type II errors are recorded (unfilled circles in Figure 8). In these cases, $\alpha_{optimal}$ is the significance level at which the type I error probability reaches zero. In other words, it is the minimum significance level required to ensure no error is made in hypothesis testing (i.e., zero combined error probability; Figure 8).

As noted in section 2, the modeling approach fits AR1 noise models to median smoothed spectra in order to minimize bias in model fitting caused by the presence of spectral peaks (see *Mann and Lees* [1996] for full details). *Meyers* [2012], however, recently drew attention to the fact that median smoothing can bias significance testing of spectra owing to the tendency for the smoothing to result in a poor model fit at low frequencies. Repetition of the modeling shown in Figure 8 without using median smoothing leads typically to slight increases in $\alpha_{optimal}$ (see supporting information), with $\alpha_{optimal}$ values between 0.01 and 0.0001 (99% and 99.99% confidence level) in 57% of the modeled scenarios and $\alpha_{optimal} > 0.01$ (i.e., $<$99%) in 19% of scenarios. For signals with jitter strengths deemed most typical of real successions (i.e., $\geq$0.2) [*Kemp and Sexton*, 2014], $\alpha_{optimal}$ values are between 0.01 and 0.001 in 50% of modeling scenarios. In these scenarios (i.e., $J \geq 0.2$), average combined error probabilities are 5.51%, 8.03%, and 10.94% for $\rho = 0.3$, 0.6, and 0.9, respectively (see supporting information).

4. Discussion

The modeling results demonstrate how the balance between type I and type II errors in cyclostratigraphic analyses will vary with the choice of significance level. A low significance level (i.e., $\alpha < 0.001$) will clearly guard against type I errors, but the perceived gain in robustness is likely to be negated by a significant increase in type II error probability and reduction in statistical power (e.g., Figure 4). The precise balance is largely a function of the unsteadiness of sedimentation, which limits the attainable power of spectral peaks

Figure 7. (a) Graphs showing the type I, type II, and combined error probabilities from simulations of stratigraphically preserved astronomical signals that are jittered using a cyclic sedimentation rate model (green lines). These data are compared to data from a stochastic sedimentation rate modeling scenario (the same model as presented in Figure 4, orange lines). The results demonstrate that under conditions of cyclic sedimentation rates, the attainable statistical significance of astronomical cycles is lower than under conditions of random sedimentation rates (i.e., $\alpha_{optimal}$ is higher). (b) Plots showing the cyclic sedimentation rate variations used in the modeling and a noise-free version of the cyclically jittered astronomical signal. Note how the cyclically jittered precession cycles are nonsinusoidal and have a cuspate shape. (c) Power spectrum of the cyclically jittered astronomical signal presented in Figure 7b. Note how power is leaked to adjacent frequencies. (d) Example spectrum of a noise-free astronomical signal jittered using a stochastic sedimentation rate model with the same jitter ($J = 0.2$). Note how the peak power in this spectrum is higher than in the cyclically jittered spectrum and how there is less leakage to adjacent frequencies. These plots help to explain why $\alpha_{optimal}$ in the cyclic jitter model is typically higher compared to signals jittered using the stochastic model.

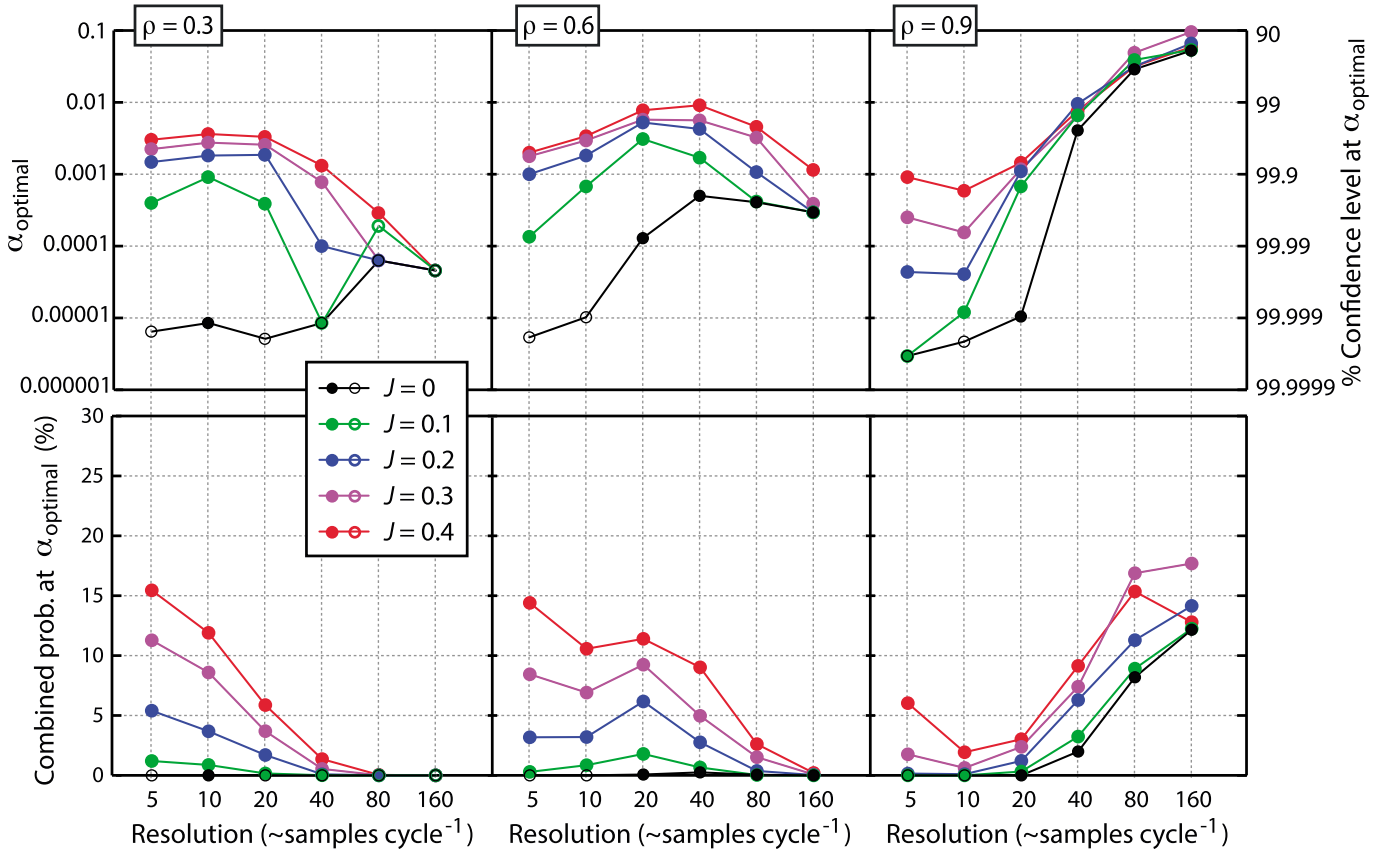


Figure 8. Summary of $\alpha_{optimal}$ and combined error probability values for a wide range of jitter levels, sampling resolutions and red noise styles (other parameters as in Figure 4). The parameters used encapsulate the wide range expected in real stratigraphic successions. Note how for a majority of scenarios, $\alpha_{optimal}$ is <0.01 and >0.0001 (i.e., between 99% and 99.99%), and for the most geologically likely scenarios (i.e., $J \geq 0.2$), $\alpha_{optimal}$ is typically >0.001 (<99.9%). Unfilled circles represent simulations with no type II errors recorded. In these instances, $\alpha_{optimal}$ is the significance level at which the type I error probability reaches zero; i.e., $\alpha_{optimal}$ is the minimum significance level required to ensure that combined error probability is zero. See main text for further details.

associated with astronomical cycles. The power of spectral peaks is also constrained by the quasiperiodic nature of astronomical insolation, regardless of any sedimentation rate unsteadiness (Figures 5 and 8). The findings help emphasize how the critical significance level α defines strictly only the probability of a type I error—because the value is formulated from the properties of the null model—i.e., the continuum red noise. The probability of a type II error is not the inverse of this, because the phenomenon we want to identify is distinct from this null model. A high critical significance reduces only the probability of a type I error but tells us nothing about the probability of a type II error. Moreover, the combined error probabilities calculated in this study emphasize that significance levels do not define the probability of making a mistake in hypothesis testing, and this "mistake probability" may be rather high. Across the modeling scenarios with geologically typical jitter ($J \geq 0.2$) shown in Figure 8, the average combined error probability at $\alpha_{optimal}$ is 5.42%.

5. Implications for Cyclostratigraphy

The numerical experiments suggest that significance testing at 0.1–0.01 significance levels (90–99% confidence levels) will rarely optimize the balance between type I and type II errors. Instead, the results advocate that significance testing should be conducted at 0.01–0.001 significance levels (99–99.9% confidence levels). This is higher than has typically been employed in the literature to date (90–99% confidence levels) but potentially lower than the corresponding multiple test corrected significance levels advocated by *Vaughan et al.* [2011], who did not explicitly consider the power/significance limiting effects of unsteady sedimentation. Nevertheless, the results are conservative, and it is unlikely that $\alpha_{optimal}$ values calculated in the models overestimates the optimized significance levels to use in real cyclostratigraphic data. This

is because noise-free, linearly translated astronomical signals are used as input signals, and no further sources of temporal or proxy uncertainty (e.g., analytical errors) are factored in to the models despite the certainty that such errors compromise the fidelity of real stratigraphically preserved astronomical signals [*Meyers et al*., 2008]. Astronomical cycles preserved stratigraphically also undergo a complex translation process from insolation to proxy [*Meyers et al.,* 2008]. Equally, a specific astronomical insolation signal is used throughout, dominated by ~21 kyr precession cycles. The nature of astronomical forcing is dependent on latitude, however, and was variable through time [*Waltham*, 2015].

Taking all these findings together, the likely conservative nature of the modeling means that a spectral peak exceeding the 99% confidence level in real data should be considered strongly suggestive of astronomical forcing. Such high confidence levels can be achieved by astronomical cycles in real data, including from pre-Mesozoic strata [e.g., *Weedon*, 2003; *Vaughan et al*., 2011; *Wu et al*., 2013; *Hilgen et al*., 2014]. Equally, however, *Hilgen et al.* [2014] showed with geologic examples how a 99% confidence level might be "unrealistically high when analyzing cyclostratigraphic records." Similarly, *Meyers* [2012] noted (though not through quantitative modeling) that given typical sedimentation rate instability associated with pelagic and hemipelagic environments, astronomical cycles with peaks only just exceeding 90% confidence levels (i.e., $\alpha \sim 0.1$) may be "the best that can be hoped for" [*Meyers*, 2012]. Importantly, the modeling here demonstrates that if geologic conditions are such that astronomical cycles cannot achieve significance of <0.01 ($>$99% confidence level), the associated likelihood of numerous type I errors represents, ostensibly at least, a significant limitation to the robustness and utility of spectral analysis. Under these circumstances, if cyclostratigraphy and spectral analysis is to be robust and have genuine utility (and respectability), spectral analysis needs to be supported with additional evidence. Such supporting evidence can come from multiple independent proxies and/or analysis of subsections of time series that produce similar spectral results [*Weedon*, 2003]. Equally, more advanced testing of astronomical signal characteristics beyond simple periodicity testing can be employed (i.e., AM and FM analysis [*Hinnov and Park*, 1998; *Zeeden et al.,* 2015; *Meyers*, 2015; *Laurin et al.,* 2016]). Independent or at least semiindependent verification of astronomical frequencies can also be established from radiometric dating, integrated stratigraphy approaches, and techniques such as the average spectral misfit approach [*Hilgen et al*., 2014; *Meyers and Sageman*, 2007].

The significance of spectral peaks and hence α_{optimal} is highly sensitive to the signal-to-noise ratio (Figure 6). The signal-to-noise ratio is a frequency-dependent quantity because the red noise that typifies cyclostratigraphic data has frequency-dependent variance (Figure 8). The wide range of α_{optimal} values calculated in Figure 8 for different types of background red noise attests to this. Equally, the deleterious effects of jitter on the attainable power of spectral peaks are also frequency dependent (Figures 3 and 8). In real strata, the choice of noise model to serve as the null model for hypothesis testing adds a further level of subjectivity in the assessment of statistical significance. *Vaughan et al.* [2011] in particular noted how in many situations AR1 models do not provide an optimal fit to the spectra of cyclostratigraphic data. Together, these observations, coupled at least in part with uncertainties in the strength and nature of astronomical forcing through Earth history, emphasize how using a fixed critical significance value is implicitly ill suited to testing for astronomical forcing in strata, regardless of what this value might be.

As a final point to note, one common method used to support inferences of astronomical forcing in the absence of independent time control utilizes the frequency ratio of spectral peaks [e.g. *Hays et al.,* 1976; *Weedon*, 2003; *Suan et al.,* 2008; *Meyers and Sageman*, 2007]. This has also been used to help match spectral peaks to astronomical parameters (e.g., the ~1:5 ratio of ~21 kyr precession and ~100 kyr eccentricity). The numerical experiments presented here, however, demonstrate that under conditions of unsteady sedimentation the frequencies of spectral peaks become unstable, and the strongest peaks are liable to occur at lower frequencies (Figure 3b; see also *Weedon* [2003]). The degree to which peaks shift is dependent on the jitter level and the frequency (Figure 3b). This phenomenon may have an impact on the use of cycle frequency ratios as a tool to identify astronomical cycles and validate astronomical forcing. Equally, the results also emphasize how the significance attained by peaks of different frequencies will vary within the same spectrum, even if the actual variance of the cyclicity associated with those peaks is equal. Indeed, the strong reduction in maximum power caused by moderate jitter at high frequencies emphasizes how jittered astronomical cycles sampled at <5 samples cycle^{-1} are unlikely to yield statistically significant peaks at all. Thus, in broad agreement with *Herbert* [1994], a sampling rate of 5 samples cycle^{-1} (8 samples cycle^{-1} in *Herbert*, 1994] should be considered the minimum required to resolve astronomical cycles.

6. Conclusions

The results presented here advocate that critical confidence levels of 99–99.9% should be used to validate astronomical forcing in strata. Cyclostratigraphic studies have hitherto typically employed confidence levels of 90–99%. In line with the findings of *Vaughan et al.* [2011], the modeling results demonstrate that confidence levels of $<$99% are unlikely to adequately balance the respective probabilities of type I and type II errors in cyclostratigraphy but instead lead to unacceptably high probability of type I errors. Nevertheless, using critical confidence levels of $>$99.9% risk a high likelihood of type II errors, and unsteadiness of sedimentation sets an upper limit on the attainable significance of spectral peaks. The 99–99.9% confidence levels advocated here represent an attempt to balance the probability of falsely rejecting and falsely accepting a null hypothesis of no astronomical forcing. Even so, confidence levels $>$99% may in many cases exceed the significance achievable by genuine stratigraphically preserved astronomical signals [*Meyers*, 2012; *Hilgen et al.,* 2014]. This issue underlines the importance of seeking additional evidence for astronomical forcing in strata, beyond that which spectral analysis alone provides. Indeed, the range in optimized significance levels determined in the experiments, coupled with the inability to know a priori the true sedimentation history and nature of astronomical forcing in most records, means that fixed value null hypothesis significance testing of spectra remains a subjective endeavor and implicitly poorly suited to validating astronomical forcing.

Acknowledgments
The manuscript benefited greatly from the comments of Stephen Meyers and Christian Zeeden. Data, model details, and code are available on request from the author at david.kemp@abdn.ac.uk.

References

Carver, R. (1978), The case against statistical significance testing, *Harvard Educ. Rev.*, *48*, 378–399.
Carver, R. (1993), The case against statistical significance testing, revisited, *J. Exp. Educ.*, *61*, 287–292.
Gradstein, F., J. Ogg, M. Schmitz, and G. Ogg (2012), *The Geologic Time Scale 2012*, Elsevier, Amsterdam.
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.
Hasselmann, K. (1976), Stochastic climate models: Part I. Theory, *Tellus*, *6*, 473–485, doi:10.1111/j.2153-3490.1976.tb00696.x.
Hays, J. D., J. Imbrie, and N. J. Shackleton (1976), Variations in the Earth's orbit: Pacemaker of the ice ages, *Science*, *194*, 1121–1132, doi:10.1126/science.194.4270.1121.
Herbert, T. D. (1994), Reading orbital signals distorted by sedimentation: Models and examples, in *Orbital Forcing and Cyclic Sequences*, *Spec. Publ. 19 of the IAS (International Association Of Sedimentologists Series)*, edited by P. L. DeBoer and D. G. Smith, pp. 576, Wiley-Blackwell, Oxford, U. K.
Hilgen, F. J., et al. (2014), Stratigraphic continuity and fragmentary sedimentation: The success of cyclostratigraphy as part of integrated stratigraphy, in *Strata and Time: Probing the Gaps in Our Understanding*, *Geol. Soc., London, Spec. Publ.*, *404*, doi:10.1144/SP404.12.
Hinnov, L., and J. Park (1998), Detection of astronomical cycles in the stratigraphic record by frequency modulation (FM) analysis, *J. Sediment. Res.*, *68*(4), 524–539.
Huybers, P. (2007), Glacial variability over the last two million years: An extended depth-derived age model, continuous obliquity pacing, and the Pleistocene progression, *Quat. Sci. Rev.*, *26*, 37–55, doi:10.1016/j.quascirev.2006.07.013.
Huybers, P., and W. Curry (2006), Links between annual, Milankovitch and continuum temperature variability, *Nature*, *441*, 329–332, doi:10.1038/nature04745.
Huybers, P., and C. Wunsch (2004), A depth-derived Pleistocene age model: Uncertainty estimates, sedimentation variability, and nonlinear climate change, *Paleoceanography*, *19*, PA1028, doi:10.1029/2002PA000857.
Kemp, D. B., and P. F. Sexton (2014), Time-scale uncertainty of abrupt events in the geologic record arising from unsteady sedimentation, *Geology*, *42*, 891–894, doi:10.1130/G35783.1.
Laskar, J., P. Robutel, F. Joutel, M. Gastineau, A. Correia, and B. Levrard (2004), A long-term numerical solution for the insolation quantities of the Earth, *Astron. Astrophys.*, *428*(1), 261–285, doi:10.1051/0004-6361:20041335.
Laurin, J., S. R. Meyers, S. Galeotti, and L. Lanci (2016), Frequency modulation reveals the phasing of orbital eccentricity during Cretaceous Oceanic Anoxic Event II and the Eocene hyperthermals, *Earth Planet. Sci. Lett.*, *442*, 143–156.
Mann, M. E., and J. M. Lees (1996), Robust estimation of background noise and signal detection in climatic time series, *Clim. Change*, *33*, 409–445, doi:10.1007/BF00142586.
Meyers, S. R. (2012), Seeing red in cyclic stratigraphy: Spectral noise estimation for astrochronology, *Paleoceanography*, *27*, PA3228, doi:10.1029/2012PA002307.
Meyers, S. R. (2015), The evaluation of eccentricity-related amplitude modulation and bundling in paleoclimate data: An inverse approach for astrochronologic testing and time scale optimization, *Paleoceanography*, *30*, 1625–1640, doi:10.1002/2015PA002850.
Meyers, S. R., and B. B. Sageman (2007), Quantification of deep-time orbital forcing by average spectral misfit, *Am. J. Sci.*, *307*, 773–792.
Meyers, S. R., B. B. Sageman, and L. Hinnov (2001), Integrated quantitative stratigraphy of the Cenomanian-Turonian Bridge Creek Limestone Member using evolutive harmonic analysis and stratigraphic modeling, *J. Sediment. Res.*, *71*, 628–644.
Meyers, S. R., B. B. Sageman, and M. Pagani (2008), Resolving Milankovitch: Consideration of signal and noise, *Am. J. Sci.*, *308*, 770–786.
Mix, A. C., J. Lee, and N. J. Shackleton (1995), Benthic foraminiferal stable isotope stratigraphy of Site 846: 0–1.8 Ma, in *Proceedings of the Ocean Drilling Program, Scientific results*, vol. 138, edited by N. G. Pisias et al., pp. 839–854, Ocean Drilling Program, College Station, Tex., doi:10.2973/odp.proc.sr.138.160.1995.
Moore, M. I., and P. J. Thomson (1991), Impact of jittered sampling on conventional spectral estimates, *J. Geophys. Res.*, *96*, 18,519–18,526, doi:10.1029/91JC01623.
Mudelsee, M. (2010), *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*, Springer, London.
Mudelsee, M., D. Scholz, R. Rathlisberger, D. Fleitmann, A. Mangini, and E. W. Wolff (2009), Climate spectrum estimation in the presence of timescale errors, *Nonlinear Process Geophys.*, *16*(1), 43–56, doi:10.5194/npg-16-43-2009.

Mudge, J. F., L. F. Baker, C. B. Edge, and J. E. Houlahan (2012), Setting an optimal α that minimizes errors in null hypothesis significance tests, *PLoS One*, *7*, e32734, doi:10.1371/journal.pone.0032734.

Pälike, H., R. Norris, J. Herrle, P. Wilson, H. Coxall, C. Lear, N. Shackleton, A. Tripati, and B. Wade (2006), The heartbeat of the Oligocene climate system, *Science*, *314*, 1894–1898, doi:10.1126/science.1133822.

Pelletier, J. (1998), The power-spectral density of atmospheric temperature from time scales of 10^2 to 10^6 yr, *Earth Planet. Sci. Lett.*, *158*, 157–164.

Rhines, A., and P. Huybers (2011), Estimation of spectral power laws in time uncertain series of data with application to the Greenland Ice Sheet Project 2 $d^{18}O$ record, *J. Geophys. Res.*, *116*, D01103, doi:10.1029/2010JD014764.

Sadler, P. M. (1981), Sediment accumulation rates and the completeness of stratigraphic sections, *J. Geol.*, *89*, 569–584.

Sadler, P. M., and D. J. Strauss (1990), Estimation of completeness of stratigraphical sections using empirical data and theoretical models, *J. Geol. Soc.*, *147*, 471–485, doi:10.1144/gsjgs.147.3.0471.

Schiffelbein, P. and L. Dorman (1986), Spectral effects of time–depth nonlinearities in deep sea sediment records: A demodulation technique for realigning time and depth scales, *J. Geophys. Res.*, *91*, 3821–3835.

Suan, G., B. Pittet, I. Bour, E. Mattioli, L. V. Duarte, and S. Mailliot (2008), Duration of the Early Toarcian carbon isotope excursion deduced from spectral analysis: Consequence for its possible causes, *Earth Planet. Sci. Lett.*, *267*, 666–679.

Taylor Perron, J., and P. Huybers (2009), Is there an orbital signal in the polar layered deposits on Mars, *Geology*, *37*, 155–158, doi:10.1130/G25143A.1.

Thompson, D. J. (1990), Time series analysis of Holocene climate data, *Philos. Trans. R. Soc. London Ser. A*, *330*, 601–616.

Tipper, J. C. (1983), Rates of sedimentation and stratigraphical completeness, *Nature*, *302*, 696–698.

Vaughan, S., R. J. Bailey, and D. G. Smith (2011), Detecting cycles in stratigraphic data: Spectral analysis in the presence of red noise, *Paleoceanography*, *26*, PA4211, doi:10.1029/2011PA002195.

Waltham, D. (2015), Milankovitch period uncertainties and their impact on cyclostratigraphy, *J. Sediment. Res.*, *85*, 990–998.

Wasserstein, R. L., and N. A. Lazar (2016), The ASA's statement on *p*-values: Context, process, and purpose, *Am. Stat.*, *70*, 129–133, doi:10.1080/00031305.2016.1154108.

Weedon, G. P. (2003), *Time-Series Analysis and Cyclostratigraphy*, 259 pp., Cambridge Univ. Press, Cambridge, U. K., doi:10.1017/CBO9780511535482.

Wu, H., S. Zhang, L. A. Hinnov, G. Jiang, Q. Feng, H. Li, and T. Yang (2013), Time-calibrated Milankovitch cycles for the late Permian, *Nat. Commun.*, doi:10.1038/ncomms3452.

Wunsch, C. (2003), The spectral description of climate change including the 100 ky energy, *Clim. Dyn.*, *20*, 353–363.

Wunsch, C. (2004), Quantitative estimate of the Milankovitch-forced contribution to observed Quaternary climate change, *Quat. Sci. Rev.*, *23*(9-10), 1001–1012, doi:10.1016/j.quascirev.2004.02.014.

Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups (2001), Trends, rhythms, and aberrations in global climate 65 Ma to present, *Science*, *292*, 686–693.

Zeeden, C., S. R. Meyers, L. J. Lourens, and F. J. Hilgen (2015), Testing astronomically tuned age models, *Paleoceanography*, *30*, 369–383, doi:10.1002/2014PA002762.