Bayesian estimation of past astronomical frequencies, lunar distance, and length of day from sediment cycles

Alberto Malinverno[1] and Stephen R. Meyers[2]

[1]Lamont-Doherty Earth Observatory of Columbia University
[2]University of Wisconsin - Madison

December 1, 2023

Abstract

Astronomical cycles recorded in stratigraphic sequences offer a powerful data source to estimate Earth's axial precession frequency k, as well as the frequency of rotation of the planetary perihelia (gi) and of the ascending nodes of their orbital planes (si). Together, these frequencies control the insolation cycles (eccentricity, obliquity and climatic precession) that affect climate and sedimentation, providing a geologic record of ancient Solar system behavior spanning billions of years. Here we introduce two Bayesian methods that harness stratigraphic data to quantitatively estimate ancient astronomical frequencies and their uncertainties. The first method (TimeOptB) calculates the posterior probability density function (PDF) of the axial precession frequency k and of the sedimentation rate u for a given cyclostratigraphic data set, while setting the Solar system frequencies gi and si to fixed values. The second method (TimeOptBMCMC) applies an adaptive Markov chain Monte Carlo algorithm to efficiently sample the posterior PDF of all the parameters that affect astronomical cycles recorded in stratigraphy: five gi, five si, k, and u. We also include an approach to assess the significance of detecting astronomical cycles in cyclostratigraphic records. The methods provide an extension of current approaches that is computationally efficient and well suited to recover the history of astronomical cycles, Earth-Moon history, and the evolution of the Solar system from geological records. As case studies, data from the Xiamaling Formation (N. China, 1.4 Ga) and ODP Site 1262 (S. Atlantic, 55 Ma) are evaluated, providing updated estimates of astronomical frequencies, Earth-Moon history, and secular resonance terms.

Bayesian estimation of past astronomical frequencies, lunar distance, and length of day from sediment cycles

A. Malinverno[1] and S. R. Meyers[2]

[1]Lamont-Doherty Earth Observatory of Columbia university, Palisades, NY 10964.

[2] Department of Geoscience, University of Wisconsin–Madison, Madison, WI 53706.

Corresponding author: Alberto Malinverno (alberto@ldeo.columbia.edu)

Key Points:

- We present two updated methods for Bayesian astrochronology: TimeOptB and TimeOptBMCMC
- TimeOptB simultaneously estimates the Earth's axial precession frequency and the sedimentation rate from cyclostratigraphic data
- In addition, TimeOptBMCMC simultaneously estimates Solar system *g*-frequencies and *s*-frequencies from cyclostratigraphic data

Abstract

Astronomical cycles recorded in stratigraphic sequences offer a powerful data source to estimate Earth's axial precession frequency k, as well as the frequency of rotation of the planetary perihelia (g_i) and of the ascending nodes of their orbital planes (s_i). Together, these frequencies control the insolation cycles (eccentricity, obliquity and climatic precession) that affect climate and sedimentation, providing a geologic record of ancient Solar system behavior spanning billions of years. Here we introduce two Bayesian methods that harness stratigraphic data to quantitatively estimate ancient astronomical frequencies and their uncertainties. The first method (TimeOptB) calculates the posterior probability density function (PDF) of the axial precession frequency k and of the sedimentation rate u for a given cyclostratigraphic data set, while setting the Solar system frequencies g_i and s_i to fixed values. The second method (TimeOptBMCMC) applies an adaptive Markov chain Monte Carlo algorithm to efficiently sample the posterior PDF of all the parameters that affect astronomical cycles recorded in stratigraphy: five g_i, five s_i, k, and u. We also include an approach to assess the significance of detecting astronomical cycles in cyclostratigraphic records. The methods provide an extension of current approaches that is computationally efficient and well suited to recover the history of astronomical cycles, Earth-Moon history, and the evolution of the Solar system from geological records. As case studies, data from the Xiamaling Formation (N. China, 1.4 Ga) and ODP Site 1262 (S. Atlantic, 55 Ma) are evaluated, providing updated estimates of astronomical frequencies, Earth-Moon history, and secular resonance terms.

Plain Language Summary

Earth's transit through our Solar system is ever evolving, and so are such seemingly unwavering planetary characteristics as the number of hours in a day. For example, it is well known that the length of the day generally increases with time as Earth's rotation rate decreases from tidal interactions with our orbiting moon. But the ability to chart out this evolution over the history of the Solar system has been hampered by limitations of both data and theoretical models. This study presents a computational approach to map out the history of Solar system motions and the history of the Earth-Moon system, including the length of a day, by leveraging geological data and astronomical theory within a statistical framework that fully accounts for uncertainties. As such, the approach provides a means to use the geological archive as an astronomical observatory, allowing us to explore Solar system and Earth-Moon dynamics throughout their long history.

1 Introduction

Quasiperiodic variations in Earth's orbit and axis of rotation influence the amount of solar radiation received at the Earth's surface, causing climate variations and corresponding changes in sediment deposition, and resulting in cyclic sediment sequences that provide a geologic archive of the astronomical rhythms. Following the groundbreaking discovery that astronomical cycles, or "Milankovitch cycles" (Milanković, 1941), pace the Pleistocene ice ages (Hays et al., 1976), there has been growing interest in the use of astrochronology to date stratigraphic sequences and constrain the geological time scale, as well as their use to evaluate Earth System and Solar System evolution (Hinnov, 2013; Ma et al., 2017; Meyers, 2019; Meyers

& Malinverno, 2018; Olsen et al., 2019; Pälike et al., 2004). This study presents a Bayesian inversion approach to quantitatively reconstruct ancient astronomical cycles by linking astronomical theory with geologic observation, building on the framework of Meyers (2015; M15 hereafter) and Meyers and Malinverno (2018; MM18 hereafter).

The periods of the most prominent Milankovitch cycles (eccentricity, obliquity and climatic precession) are controlled by fundamental Solar system secular frequencies that describe the frequency of rotation of the planetary orbital perihelia (g_i) and the frequency of rotation of the ascending nodes of their ecliptic planes (s_i), combined with the precession frequency of the Earth's spin axis (k). The periods of eccentricity cycles in the Earth's orbit are determined by differences $g_i - g_j$, while those of the obliquity of the Earth's axis by sums $s_i + k$, and those of climatic precession (precession modulated by eccentricity) by sums $g_i + k$. We list in Table 1 the most important cycles used in the present study. The frequencies g_i and s_i are mostly controlled by the corresponding planet (i = 1 for Mercury, 2 for Venus, etc.). The eccentricity and climatic precession cycles in Table 1 depend on the g_i for the five innermost planets, and the obliquity frequencies are a function of the s_i for the four innermost planets and Saturn (i = 6); s_5 for Jupiter is zero as a consequence of angular momentum conservation (Fitzpatrick, 2012, p. 180).

In principle, ancient sediment records that record Milankovitch cycles can be used to estimate past variations in climatic precession, obliquity and eccentricity, as well as the fundamental frequencies (g_i, s_i) and the axial precession frequency (k) from which they derive. This provides a powerful means to peer into the early history of the Solar System and Earth-Moon system, analogous to a telescope imaging distant stars and galaxies to reconstruct the history of the universe (Meyers & Peters, 2022).

It has long been known that tidal friction results in a torque that progressively slows down the Earth rotation and accelerates the Moon, sending it into a higher orbit (e.g., Darwin, 1898). In turn, the slowing of the Earth's spin and increasing lunar distance result in an increase in the period of the precession of the Earth's axis and a decrease in the axial precession frequency k. This is a large effect over geologic time scales: models and data indicate that k decreased from ~86 arcsec/yr at 1.4 Ga to a present value of ~50.5 arcsec/yr (MM18; Farhat, Auclair-Desrotour, et al., 2022). In contrast to k, long-term Solar system calculations show that the fundamental frequencies g_i and s_i did not vary greatly over geologic time (Hoang et al., 2021). The value of k can therefore be estimated from sedimentary records by comparing eccentricity frequencies, which do not depend on k, with climatic precession or obliquity frequencies, which depend on k (see Table 1; MM18; Lantink et al., 2022). Estimates of past values of k can constrain the past history of the Earth's length of day (LOD) and lunar distance, informing models for the evolution of tidal dissipation over geological time scales (e.g., Farhat, Auclair-Desrotour, et al., 2022), and better defining the past values of climatic precession and obliquity frequencies for astronomical timescale development.

Sediment records can also give information on past values of the fundamental Solar system frequencies g_i and s_i. For example, Olsen et al. (2019) used a long Newark basin Triassic record (~210 Ma) to estimate a period of 1.75 Myr for the $g_4 - g_3$ cycle, compared to its present period of ~2.4 Myr. Zeebe and Lourens (2019) calculated a Solar system solution that best fitted the Walvis Ridge Site 1262 record, and noted that their solution contains a shift in the $g_4 - g_3$ cycle from a period of ~1.5 Myr before 50 Ma to ~2.4 Myr (near the present value) afterwards. A similar shift of the $g_4 - g_3$ cycle was observed by MM18, through the analysis of a segment of the Walvis Ridge Site 1262 cyclostratigraphic record around 55 Ma. Because of chaotic dynamics, Solar system solutions calculated starting from the present state diverge considerably at ages beyond ~50 Ma (Laskar, 2020), and at earlier ages the period of $g_4 - g_3$ in these model results fluctuates in a broad range of 1.5-2.6 Myr (Figure 7 of Olsen et al., 2019). Astronomical cycles recorded in sediments can constrain the value of this long-term periodicity and identify which computed solutions are consistent with the past Solar system history.

Table 1. Fundamental frequencies of the Solar system (g_i and s_i), axial precession frequency (k), and astronomical cycle frequencies (eccentricity, obliquity and climatic precession) used in this study and their present day values. Present day values of g_i and s_i after Hoang et al. (2021); present day value of k after Farhat, Auclair-Desrotour, et al. (2022).

Astronomical frequencies			
	Frequency (arcsec/yr)	Frequency (cycles/kyr)	Period (kyr)
g_1	5.759	0.0044	225.0
g_2	7.448	0.0057	174.0
g_3	17.269	0.0133	75.0
g_4	17.896	0.0138	72.4
g_5	4.257	0.0033	304.4
s_1	–5.652	–0.0044	229.3
s_2	–6.709	–0.0052	193.2
s_3	–18.773	–0.0145	69.0
s_4	–17.707	–0.0137	73.2
s_6	–26.348	–0.0203	49.2
k	50.468	0.0389	25.7
Eccentricity			
	Frequency (arcsec/yr)	Frequency (cycles/kyr)	Period (kyr)
$g_2 - g_5$	3.191	0.0025	406.2
$g_3 - g_2$	9.821	0.0076	132.0
$g_4 - g_2$	10.448	0.0081	124.0
$g_3 - g_5$	13.012	0.0100	99.6
$g_4 - g_5$	13.639	0.0105	95.0
Obliquity			
	Frequency (arcsec/yr)	Frequency (cycles/kyr)	Period (kyr)
$s_6 + k$	24.120	0.0186	53.7
$s_3 + k$	31.695	0.0245	40.9
$s_4 + k$	32.761	0.0253	39.6
$s_2 + k$	43.759	0.0338	29.6
$s_1 + k$	44.816	0.0346	28.9
Climatic precession			
	Frequency (arcsec/yr)	Frequency (cycles/kyr)	Period (kyr)
$g_5 + k$	54.725	0.0422	23.7
$g_1 + k$	56.227	0.0434	23.0
$g_2 + k$	57.916	0.0447	22.4
$g_3 + k$	67.737	0.0523	19.1
$g_4 + k$	68.364	0.0527	19.0

Astronomical signals recorded by sediment sequences are superimposed on a sizable background of other variability, due to fluctuations in sediment characteristics that are not related to astronomically-driven climatic cycles (e.g., tectonic, geochemical, or ocean circulation changes that influence sedimentation, diagenetic processes). Unrecognized variations in sedimentation rate and hiatuses in sedimentation will also distort the astronomical signals. Many approaches have been developed by the cyclostratigraphic community to recognize astronomical signals, from visual correlation to elaborate quantitative analyses (for an overview, see Sinnesael et al., 2019). The method we present here focuses on 1) simultaneously quantifying uncertainties in the estimated sedimentation rate and astronomical frequencies, and 2) providing a measure of significance of the results to avoid the false detection of astronomical signals in records that do not contain them (Type I errors; Meyers, 2019; Weedon, 2022).

In our previous work, M15 established the TimeOpt method, based on how closely stratigraphic data matched Milankovitch periodicities and the expected eccentricity modulation of climatic precession. The method determined a best-fit value for sedimentation rate for prescribed values of five eccentricity frequencies and three climatic precession frequencies (Table 1 of M15). TimeOpt also assessed the statistical significance of the results by comparing the fit obtained for the stratigraphic data to that calculated for random time series of similar statistical characteristics.

To extend the methodology and determine from cyclostratigraphic data past values and uncertainties of the astronomical frequencies, MM18 then developed TimeOptMCMC, a Markov chain Monte Carlo method that performs a random walk in the space of the parameters of interest and samples a posterior probability density function (PDF) of sedimentation rate u, of five Solar system frequencies g_i, and of the axial precession frequency k. The posterior PDF combines a prior PDF of the parameters (from information other than that provided by stratigraphic data) and a likelihood function that quantifies how closely data predicted by the parameters fit the stratigraphic data. However, a drawback of TimeOptMCMC is that it typically requires a computationally expensive initial experimentation phase to set up a proposal distribution for the random walk steps that appropriately samples the posterior PDF of the parameters. Once the proposal PDF is properly ‘tuned’, the method is still computationally expensive in its original implementation, typically requiring days to weeks of simulation for each cyclostratigraphic data set.

In the present study, we introduce two modified methods that offer significant improvements over the original M15 and MM18 approaches. TimeOptB (‘B’ for Bayesian) extends the TimeOpt methodology of M15 to calculate the posterior PDFs of both sedimentation rate and axial precession frequency, keeping the Solar system fundamental frequencies fixed to characteristic prior values. The statistical significance (‘p-value’) of the fit of astronomical cycles to the data is also evaluated. TimeOptBMCMC provides a more complete solution by sampling the posterior PDF of sedimentation rate and of all the astronomical parameters of interest: ten Solar system fundamental frequencies (five g_i and five s_i) and the axial precession frequency k. Compared to the previous version, TimeOptBMCMC implements an adaptive sampling strategy that requires no preliminary set up and is orders of magnitude faster in obtaining a useful sample of the posterior PDF. Both methods also account for the possible presence of obliquity cycles (which were not considered in M15 and MM18; however, see Meyers (2019) for TimeOpt applications that include obliquity), implement updated Bayesian priors for the Solar system fundamental frequencies and the axial precession frequency based on astronomical calculations (Farhat, Auclair-Desrotour, et al., 2022; Hoang et al., 2021), and include improvements in the approach used for likelihood estimation.

The main goal of this paper is to present in detail the TimeOptB and TimeOptBMCMC methods, applying them to a synthetic data set and to the two data sets previously studied in MM18 for example demonstrations. The focus of this contribution is on the methodology used to estimate sedimentation rate and astronomical frequencies, not on the implications of the results for the history of tidal dissipation and for improvements in astrochronology. Applications of the methods to the the analysis of a number of records throughout geologic time are currently in development and will be published in the near future (Ajibade et al., 2023; Wu et al., 2023)

In the rest of this paper, we first describe the Bayesian formulation to compute the value of the posterior PDF for any value of the astronomical parameters of interest. We then explain in detail the two new methods, compare their results for the two data sets examined by MM18 (Xiamaling Formation, N. China, 1.4 Ga and ODP Site 1262, S. Atlantic, 55 Ma), and describe how to obtain estimates of lunar distance and length of day and their uncertainties from the posterior PDF of the axial precession frequency. We conclude by discussing strengths and limitations of our approach and future improvements.

2 Bayesian Formulation

The vector **m** of the parameters of interest consists of the sedimentation rate u, five values of g_i, five values of s_i, and the precession frequency k as in

$$\mathbf{m} = [g_1, g_2, g_3, g_4, g_5, s_1, s_2, s_3, s_4, s_6, k, u]. \quad (1)$$

The posterior PDF of **m** is defined from Bayes rule as

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{m})\, p(\mathbf{d}|\mathbf{m})}{p(\mathbf{d})}, \quad (2)$$

where the vector **d** consists of N sediment property values (e.g., sedimentologic or geochemical proxy data) measured at constant increments of stratigraphic depth. The two key terms in Equation (2) are the prior PDF $p(\mathbf{m})$ and the likelihood function $p(\mathbf{d}|\mathbf{m})$. (The denominator $p(\mathbf{d})$ does not depend on **m** and is a normalizing constant that is not relevant for the methods presented here.) The symbols and acronyms used in this paper are listed in Table 2.

2.1 The Prior PDF

The role of the prior PDF is to limit the space of possible parameters to values that agree with information other than that provided by the stratigraphic data in **d**. As there is no information on prior correlations between the parameters they are taken as independent, so the prior PDF of **m** is simply the product of the prior PDFs of each parameter as in

$$p(\mathbf{m}) = p(g_1)\, p(g_2) \ldots p(g_5)\, p(s_1) \ldots p(s_6)\, p(k)\, p(u).$$

The prior PDF of sedimentation rate u is defined as a uniform distribution between a minimum and maximum value. These bounds on a realistic value of u can be based on independent chronostratigraphic information (e.g., radioisotopic dating, bio- or magnetostratigraphy) or on the environment of deposition (e.g., from the range of sedimentation rates determined in similar modern and ancient depositional settings).

The prior PDFs for the fundamental Solar system frequencies g_i and s_i are the distributions obtained by Hoang et al. (2021), determined by running a large number of long-term astronomical solutions starting from slightly different initial conditions. The PDFs of g_i and s_i are skew Gaussians with some secondary modes, and their parameters are listed in Table 2 of Hoang et al. (2021) as a function of geologic time. The parameters of the prior PDFs were obtained from frequencies obtained over intervals of 20 Myr (inner planets, Mercury to Mars) or 50 Myr (outer planets). The prior PDFs of the frequencies g_i and s_i are illustrated in Figure 1 for ages between the present and 3.3 Ga, a time interval that includes most stratigraphic records available for astronomical cycle analysis.

Table 2. Symbols and acronyms used in this study.

Symbols	
a	Semi-major axis of lunar orbit
$\mathbf{C}_e$	Covariance matrix of residuals $\mathbf{e}$
$\mathbf{d}$	Vector of measured sediment property data
$\mathbf{d}_{\mathrm{pred}}$	Vector of data predicted by parameters in $\mathbf{m}$
$\mathbf{e}$	Vector of residuals $\mathbf{d} - \mathbf{d}_{\mathrm{pred}}$
g_i	Fundamental Solar system frequencies for the rotation of the planetary perihelia
k	Earth's axial precession frequency
$\mathbf{m}$	Vector of parameters (g_i, s_i, k, u)
N	Number of data points in vector $\mathbf{d}$
N_{eff}	Effective number of independent observations in vector $\mathbf{d}$
N_{sim}	Number of simulated random data sets in significance testing
R^2	Squared correlation coefficient
$\mathbf{R}_e$	Correlation matrix of residuals $\mathbf{e}$
r_i	Autocorrelation coefficient of residuals $\mathbf{e}$ at lag i
s_i	Fundamental Solar system frequencies for the rotation of the ascending nodes of the orbital planes
u	Sedimentation rate
σ_e^2	Variance of residuals $\mathbf{e}$
τ	Lag where the autocorrelation of $\mathbf{e}$ reaches zero
ϕ_i	Coefficient of an AR(P) process
ω	Earth's spin rate
Acronyms	
AR(P)	Autoregressive process of order P
ETP	Eccentricity, tilt, and precession
LOD	Length of day
M15	Meyers (2015)
MM18	Meyers and Malinverno (2018)
MAP	Maximum a posteriori
MCMC	Markov chain Monte Carlo
PDF	Probability density function

The changes in the mean value of the prior PDF and uncertainties of the fundamental Solar system frequencies in the past are relatively small, a few percent at most. The frequencies associated with the outer planets (g_5 and s_6) vary the least, followed by g_2. For example, the Earth eccentricity frequency $g_2 - g_5$ (period ~405 ka) has remained nearly constant through geologic time and has been proposed as a stable anchoring cycle in astrochronology (e.g., Hinnov, 2013; Laskar, 2020; Laskar et al., 2004; Olsen et al., 2019).

In contrast, the Earth precession frequency k decreased systematically through time due to tidal energy dissipation. The general trend of k in time can be estimated by modeling tidal effects and/or by interpolating past geological estimates of k (e.g., Berger & Loutre, 1994; Laskar et al., 2004). The most recent study is by Farhat, Auclair-Desrotour, et al. (2022), who calculate past precession frequency from a tidal dissipation model that accounts for changes in the overall continental distribution and Earth spin rate. The actual history of tidal dissipation, however, is not accurately known, and estimated past values of k have large uncertainties (e.g., Waltham, 2015).

We set the prior PDF of k to a normal distribution with a time-dependent mean $\mu_k(t)$ and standard deviation $\sigma_k(t)$. The prior mean is from a polynomial fit to the past variation of k calculated by Farhat, Auclair-Desrotour, et al. (2022; see their Figure 6) for ages 0-3.3 Ga, which is

$$\mu_k(t) = 50.4677 + 23.1305\, t + 13.0658\, t^2 - 11.2346\, t^3 + 2.4322\, t^4,$$

where age t is in Ga. This polynomial accounts for the long-term expected variation of k in the past, excluding some shorter-term fluctuations at ages < 600 Ma; these shorter-term variations should be confirmed (or not) by cyclostratigraphic data and not imposed a priori. The tidal dissipation model of Farhat, Auclair-Desrotour, et al. (2022, p. 4) was deliberately not fitted to geological data, and it is appropriate to use the trend they computed as the prior mean of k.

Farhat, Auclair-Desrotour, et al. (2022) also calculate an uncertainty of the value of k obtained from their tidal dissipation model, and these uncertainties are small compared to the uncertainties of k estimated from cyclostratigraphy (see their Figure 6). Our goal is to estimate k from cyclostratigraphy in a way that is generally consistent with the effects of tidal dissipation, but the prior standard deviation should be large enough so that the posterior PDF of k we obtain is not unduly influenced by and provides a test of the tidal modeling results. We therefore set the prior standard deviation of k using the large uncertainties in the past precession period given in Waltham (2015), allowing the opportunity for deviations from the Farhat, Auclair-Desrotour, et al. (2022) tidal model. These uncertainties are a conservative estimate based on substantially different assumptions about the past history of tidal dissipation, and we assume that they correspond to ± two standard deviations. By fitting a polynomial to the fractional uncertainty (uncertainty divided by the mean) of the precession period given by the JavaScript calculator of Waltham (2015) between the present and 3.3 Ga, we obtained an expression for the prior standard deviation of k:

$$\sigma_k(t) = (0.0962\, t - 0.0262\, t^2 + 0.0030\, t^3)\, \mu_k(t).$$

The resulting prior PDF of k is shown in Figure 1.

2.2 The Likelihood Function

The likelihood function quantifies how probable it is to observe the measured stratigraphic data **d** when the parameters have the values in **m**, and depends on the difference between **d** and a vector $\mathbf{d}_{\text{pred}}$ of data predicted by **m**. We define an error or residual vector that is

$$\mathbf{e} = \mathbf{d} - \mathbf{d}_{\text{pred}}. \tag{3}$$

The value of the likelihood function depends on the overall size of the residuals **e**; the likelihood of having observed the data **d** if the model parameters equal the values in **m** will be greater if the residuals are smaller. Following general practice, the residual vector is assumed to have a normal distribution and the likelihood is the multivariate normal PDF of the vector of residuals **e** with a mean of zero and a $N \times N$ covariance matrix $\mathbf{C}_e$. We consider here the general case where the residuals can be assumed to be second-order stationary, so that their covariance does not change with position (in our case, stratigraphic depth) and the covariance matrix can be written as

$$\mathbf{C}_e = \sigma_e^2 \mathbf{R}_e,$$

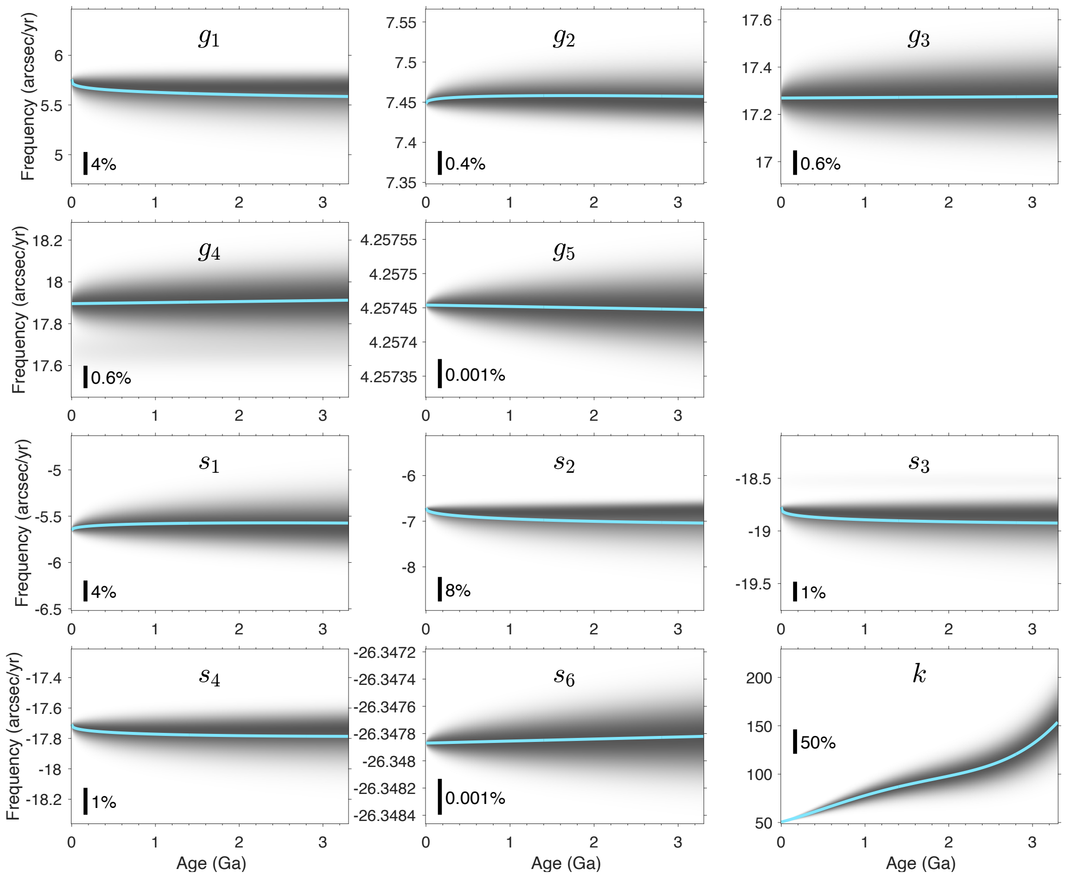

Figure 1. Prior PDFs of astronomical frequencies shown as gray scale images as a function of age (0-3.3 Ga; see the text for details). The blue continuous line shows the prior mean and the black vertical bars display the scale of the overall variations as a percentage of the present value. The Solar system fundamental frequencies g_1 to g_5, s_1 to s_4, and s_6 display a much lower variability compared to the systematic decrease with time of the axial precession frequency k.

where $\mathbf{R}_e$ is a symmetric Toeplitz matrix of correlation coefficients with a unit diagonal and constant off-diagonal entries as in

$$\mathbf{R}_e = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{N-1} \\ r_1 & 1 & r_1 & \ddots & \vdots \\ r_2 & r_1 & 1 & \ddots & r_2 \\ \vdots & \ddots & \ddots & \ddots & r_1 \\ r_{N-1} & \cdots & r_2 & r_1 & 1 \end{bmatrix},$$

and r_i is the autocorrelation function of of **e** at lag i ($-1 < r_i < 1$). If the residuals were uncorrelated, $\mathbf{R}_e$ would equal the identity matrix.

TimeOptB and TimeOptBMCMC use two likelihood functions that measure the fit to two kinds of predicted data. The first ("spectral fit" of M15 and MM18) is based on predicted data $\mathbf{d}_{\text{pred}}$ obtained by fitting to the observed stratigraphic data cycles of eccentricity, obliquity, and climatic precession given by the astronomical frequencies and sedimentation rate in **m**. The second ("envelope fit" of M15 and MM18) is based on predicting the envelope of a bandpass-filtered climatic precession signal by fitting a combination of cosine and sine functions with the eccentricity frequencies derived from **m**. For both the "spectral" and "envelope" evaluation, fitting cosine and sine terms at each astronomical frequency allows estimation of their amplitudes and phases, as in a standard Fourier transform. Details on the calculation of the predicted data in the spectral and envelope fit are in the Supporting Information.

In both the spectral and envelope fit, the residuals in **e** are positively correlated. For example, it is well known that stratigraphic data have a "red noise" character and can be modeled as autoregressive processes with positive correlations of nearby values (e.g., Mann & Lees, 1996). It is important to account for these correlations in the likelihood function because they affect the posterior uncertainties of the parameters. Consider a simple case where the parameter of interest is the mean of the observations, estimated from a sample mean μ as in

$$\mu = \frac{1}{N}\sum_{i=1}^{N} d_i.$$

If the residuals $\mathbf{e} = \mathbf{d} - \mu$ have a variance σ_e^2 and are uncorrelated, the likelihood function of the sample mean would have a variance equal to σ_e^2/N. However, if the residuals are positively correlated there are fewer than N independent observations. For example, if the autocorrelation function of the residuals **e** decreased from unity at zero lag to a value near zero at a lag τ, the effective number of independent observations would approximately be

$$N_{\text{eff}} \approx N/\tau$$

(e.g., Neal, 1993; Priestley, 1981; Zięba, 2010; Zięba & Ramza, 2011). As $N_{\text{eff}} < N$, the sample mean would have a variance $\sigma_e^2/N_{\text{eff}}$ that is greater than in the case where the residuals were uncorrelated. If correlations in the data residuals were ignored, the likelihood function would be artificially concentrated around its mode, causing an underestimation of the uncertainties in the parameters. This could be a substantial bias; in the example of the sample mean, if the data were correlated up to a lag $\tau = 9$, ignoring these correlations would underestimate the posterior uncertainty by a factor of three (measured from the standard deviation).

Moreover, when the likelihoods for several data fits are combined, it is important to account for differences in the correlations of the residuals. In our application, the residuals in the spectral fit are clearly less correlated than the much smoother residuals in the envelope fit. Ignoring this difference in the correlations would not properly weigh the importance of each data fit in constraining the parameters.

An outstanding problem in defining the likelihood function in Bayesian inference is that the variance and autocorrelation of the residuals **e** are typically unknown and cannot be confidently set a priori. On the other hand, the data may be informative about the statistical properties of the residuals. For example, fitting a few harmonic components as in the spectral fit will always result in non-zero residuals, and the statistics of these residuals may be used to infer the residual variance and autocorrelation.

One way to extract this information is to follow a hierarchical Bayes strategy (Gelman et al., 2004; Malinverno & Briggs, 2004) by adding σ_e^2 and parameters that define the correlation matrix $\mathbf{R}_e$ to the unknowns of the problem as "hyperparameters." The original TimeOptMCMC of MM18 implemented this strategy by adding to the parameter vector two hyperparameters for each of the spectral and envelope fit: the variance of the data residuals σ_e^2 and the coefficient ϕ_1 of an autoregressive process of order 1 that defined their autocorrelation. These four hyperparameters were then sampled by MCMC, and the sampled values were used to define the covariance matrix $\mathbf{C}_e$ when calculating the likelihood at each iteration. The final histogram of the sampled σ_e^2 and ϕ_1 described their posterior PDFs (Fig. S4, S7, and S10 of MM18).

In the updated methodology presented here, we apply an empirical Bayes strategy, where values of the hyperparameters are estimated from the data, e.g., by choosing their maximum likelihood value (Carlin & Louis, 2000; Casella, 1985). While hierarchical Bayes fully accounts for the posterior uncertainty of the hyperparameters, empirical Bayes simplifies the calculations, speeds up the inversion, and can return a posterior PDF for the parameters in **m** that is close to that obtained by hierarchical Bayes (see the discussion of Figure 9 in Malinverno & Briggs, 2004).

The rest of this section describes the form of the likelihood function for the spectral and envelope fits. Assuming that there are no correlations between the residuals obtained in the two fits, the total likelihood is simply the product of the spectral and envelope likelihoods.

2.2.1 Likelihood for the Spectral Fit

The spectral fit likelihood is based on modeling the residuals **e** in Equation 3 as an autoregressive process of order 2, or AR(2), as in

$$e_i = \phi_1 e_{i-1} + \phi_2 e_{i-1} + w_i, \tag{4}$$

where the vector **w** is white noise, a sequence of uncorrelated normally distributed values that have zero mean and a variance σ_w^2. The AR process exploits the correlations in the vector **e** to predict the i-th value e_i with a linear combination of nearby values, while the driving noise term w_i accounts for unpredictable random effects. If the time series in **e** is adequately modeled by an AR(2) process, the resulting **w** (which can be obtained by solving Equation 4 for w_i) should be white noise. This can be verified by computing the sample autocorrelation of the estimated **w** and checking that the autocorrelation values are not significantly different from zero for nonzero lags. Whereas cyclostratigraphic analyses often assume that stratigraphic records can be modeled as an AR(1) process (e.g., MM18; Mann & Lees, 1996), we found that in several cases an AR(2) process is necessary to produce a vector **w** that is close to white noise. A general description of AR processes can be found in treatments of time series analysis (Chatfield, 1989; Cox & Miller, 1965; Priestley, 1981).

Dettmer at al. (2012) proposed a way to simplify the evaluation of a multivariate normal likelihood if the residuals **e** can be modeled as an AR process. In the AR(2) process (Equation 4), values e_i can be predicted by e_{i-1} and e_{i-2} plus a driving noise w_i that is uncorrelated. Therefore, the residuals **e** contain a predictable component and a random component **w**; if we subtract the predictable component of **e**, the likelihood function can then be written as the PDF of the uncorrelated driving noise **w**. This simplifies considerably the calculation of the likelihood because the covariance matrix of **w** is diagonal. To complete the calculation of the spectral fit likelihood, we apply an empirical Bayes strategy and estimate the AR(2) coefficients ϕ_1 and ϕ_2 and the variance σ_w^2 of the driving noise **w** from the residuals **e** (Andersen, 1974; Burg, 1967; Ulrych & Bishop, 1975). Details on the estimation of ϕ_1, ϕ_2, and σ_w^2 and on the equation for the spectral fit likelihood are in the Supporting Information.

2.2.2 Likelihood for the Envelope Fit

It seems reasonable to apply the same methodology to the evaluation of the likelihood of the envelope fit. However, an AR(P) model is not a good representation of the residuals of the envelope fit, even if the order P is high. The reason is that these residuals **e** are the difference of two low-frequency band-limited signals: the envelope of a filtered climatic precession signal in the data (**d** in Equation 3) minus the sum of harmonic components with the periods of eccentricity ($\mathbf{d}_{\mathrm{pred}}$). Therefore, the residuals in the envelope fit are very smooth and cannot be well reproduced by an autoregressive process driven by uncorrelated noise.

The likelihood of the envelope fit instead uses an effective number of independent observations $N_{\mathrm{eff}} = N/\tau < N$, based on an estimate of the lag τ where the autocorrelation of the envelope fit residuals reaches zero (Zięba, 2010; Zięba & Ramza, 2011). Details on the bandpass filtering to extract the climatic precession signal in the data (Zeeden et al., 2018), on the calculation of the predicted precession envelope, on the estimation of the lag τ, and on the equation for the envelope fit likelihood are in the Supporting Information.

3 TimeOptB Methodology

As the Solar system frequencies g_i and s_i do not vary greatly throughout geologic time (Figure 1), in TimeOptB we fix these frequencies to their prior mean value at the time of sediment deposition, so that the only variable parameters in **m** are the sedimentation rate u and the axial precession frequency k. The value of the likelihood, prior PDF, and posterior PDF can then be calculated over a 2-D grid of u and k. The boundaries of this grid can be initially set to span the range of the prior PDF and can then be narrowed to resolve details of the posterior PDF.

Compared to the original TimeOpt of M15, the major enhancements in TimeOptB are that 1) the axial precession frequency k is not fixed but is a variable that is estimated from the data and 2) that the Bayesian formulation provides a measure of uncertainty in the values of u and k consistent with the data.

The significance of astronomical cycles inferred from noisy stratigraphic data is an outstanding issue, and it has been claimed that false detection of such cycles is likely widespread in existing studies (Smith, 2023; Weedon, 2022). As done in TimeOpt, we implemented in TimeOptB a simple Monte Carlo procedure to investigate the statistical significance of the detected astronomical cycles. The procedure is based on generating a large sample of N_{sim} random simulated data series that are AR(2) processes with coefficients ϕ_1 and ϕ_2 equal to those estimated from the observed data for the maximum a posteriori value (MAP) values of u and k. In each of these N_{sim} data sets, we repeat the TimeOptB procedure for the spectral fit over the range of u and k explored with the measured data and retain the maximum value of the Pearson R^2 correlation coefficient (the ratio of the variances of the data predicted by fitting astronomical cycles over the total variance). We then compare the maximum spectral fit R^2 values obtained in each of the N_{sim} simulated data sets to the R^2 obtained for the actual data at the MAP values of u and k. (It should be noted that in each of the simulated data series the maximum R^2 will be obtained for values of u and k that are different than the MAP values in the measured data.)

Following the general philosophy of significance testing (Hacking, 2001) we define a p-value as the fraction of N_{sim} cases where the R^2 of the simulated data sets is as large or larger than the R^2 in the measured data. If the data contain significant astronomical cycles, a comparable fit should only occur rarely in the random simulated data sets and the p-value should be small. To further investigate astronomical cycle significance, we also repeat the same Monte Carlo procedure separately for cycles of eccentricity, obliquity, and climatic precession. Whereas the critical significance test is for all the astronomical cycles, the results of the Monte Carlo experiment when only one set of cycles is considered will highlight which cycles are most informative in a particular cyclostratigraphic data set.

The accuracy of the p-value estimated in this Monte Carlo procedure will obviously improve as N_{sim} grows; we suggest $N_{sim} \geq 1{,}000$. Even if N_{sim} is large, the estimated p-value is not assured to be the same in different runs of N_{sim} Monte Carlo simulations. In practice, it may be the case that no simulated data set reaches the fit level observed for the measured data; in that case, all that can be concluded from the Monte Carlo experiment is that the p-value is $< 1/ N_{sim}$.

4 TimeOptB example results

4.1 ETP curve (45 Ma)

To evaluate the efficacy of the TimeOptB approach, we test it against a synthetic data set that consists of known astronomical signals plus random noise. An ETP astronomical signal is constructed as the sum of eccentricity, obliquity (tilt), and climatic precession from the solution of Laskar et al. (2004). The synthetic record consists of 1000 data points spanning a 1 Myr interval centered on an age of 45 Ma and was converted to depth assuming a sedimentation rate of 1 cm/kyr. Each of the three astronomical signals was normalized to zero mean and unit variance before their summation. A time series of AR(1) correlated noise ($\phi_1 = 0.8$) was added to the astronomical signals to obtain the final synthetic data set. The noise variance was adjusted so that the variance of the astronomical signals was 0.44 times the total variance (a value of $R^2 = 0.44$ is close to that obtained for the stratigraphic data sets that will be shown later).

Images of the log-posterior and posterior PDFs as a function of sedimentation rate u and axial precession frequency k are shown in Figure 2. The posterior PDF images display a strong positive correlation between u and k, which is intrinsic to the estimation of astronomical periods from stratigraphic data. If the stratigraphic data contain a cycle with a distinct spatial wavelength attributed to an astronomical cycle, the temporal frequency of that cycle will be a function of the sedimentation rate; if the sedimentation rate were higher, the frequency of the astronomical cycle will increase correspondingly (see also the discussion).

The marginal posterior PDFs of u and k in Figure 2 are obtained by integrating the images in the vertical and horizontal directions, respectively. The posterior means of u and k

(0.995 cm/kyr and 51.367 arcsec/yr) are very close to the sedimentation rate used in the synthetic example and to the axial precession frequency at 45 Ma in the Laskar et al. (2004) calculations (which is 51.707 arcsec/yr, from their Equation 40). Notably, if the sedimentation rate were increased by 0.5% to the exact value of 1 cm/yr, the posterior mean axial precession would increase by the same amount to 51.624 arcsec/yr, getting even closer to the value expected in the ETP signal.

The fit to the data and to the precession envelope for the maximum a posteriori (MAP) value of u and k is shown in Figure 3. The R^2 for the data fit is 0.61, which is greater than the 0.44 value used to construct the synthetic data set. This is due to a small amount of variance in the added noise being attributed to astronomical cycles. The periodogram of the ETP data (Figure 3c) shows a close correspondence with the spectral lines of the astronomical cycles.

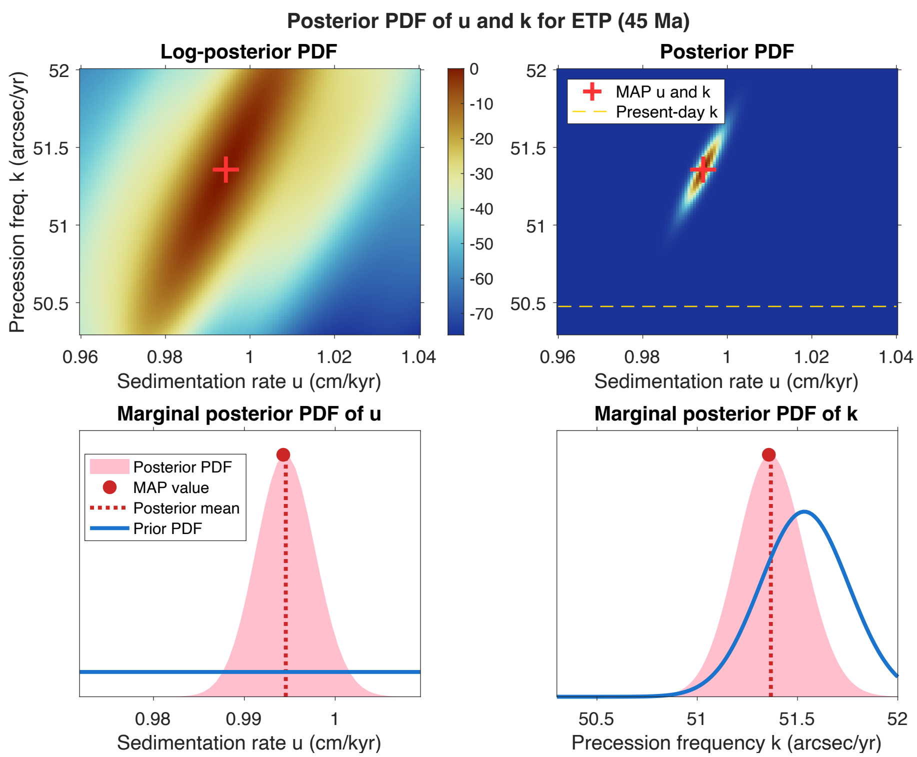

Figure 2. Posterior PDFs of sedimentation rate u and axial precession frequency k obtained by TimeOptB from the synthetic ETP test data set. In the PDF images, the log-posterior PDF is normalized to a MAP value of zero and the posterior PDF to a MAP value of 1. The horizontal dashed line in the posterior PDF image shows the present value of k. The parameters used to construct the synthetic ETP data set were u = 1 cm/yr and k = 51.707 arcsec/yr.

To check the significance of the estimated astronomical signals, we generated N_{sim} = 1,000 AR(2) time series with coefficients $\phi_1 = 0.84$ and $\phi_2 = -0.07$, equal to those estimated for the MAP value of u and k. The value of these coefficients are close to those of the AR noise that was added to the data ($\phi_1 = 0.8$, $\phi_2 = 0$). Figure 4 shows that the fit to all the astronomical cycles and to each individual set of cycles (climatic precession, obliquity, or eccentricity) is highly significant. Finally, the fit of an AR(2) spectrum to the periodogram of the ETP data, and the sample autocorrelation of the driving noise of the AR(2) process in the residuals $\mathbf{e}$ of the spectral fit, are shown in Figure S1. The sample autocorrelation of the driving noise is close to that of white noise, as expected. In conclusion, TimeOptB is successful in recovering the sedimentation rate and axial precession frequency in a synthetic data set contaminated by a realistic amount of correlated noise.

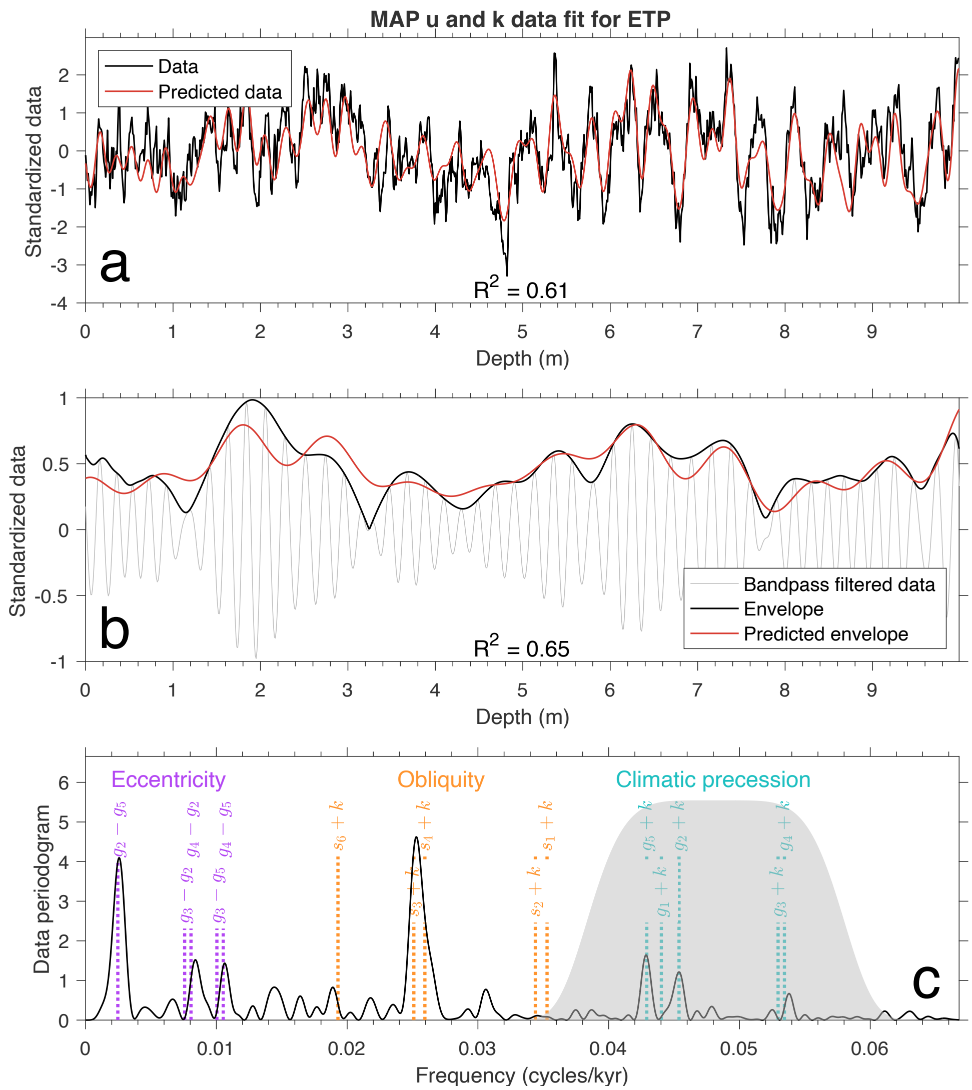

Figure 3. Fit to the synthetic ETP test stratigraphic data for the TimeOptB-derived MAP value of sedimentation rate u (0.994 cm/kyr) and axial precession frequency k (51.357 arcsec/yr). (a) Fit between measured and predicted stratigraphic data (spectral fit). (b) Fit between the envelope of the bandpassed climatic precession signal and the envelope predicted by the eccentricity frequencies (envelope fit). (c) Data periodogram (black continuous line) and frequencies of astronomical cycles (dotted vertical lines). The gray shaded area shows the frequency response of the filter used to compute the bandpassed climatic precession signal in the data (gray curve in (b)).

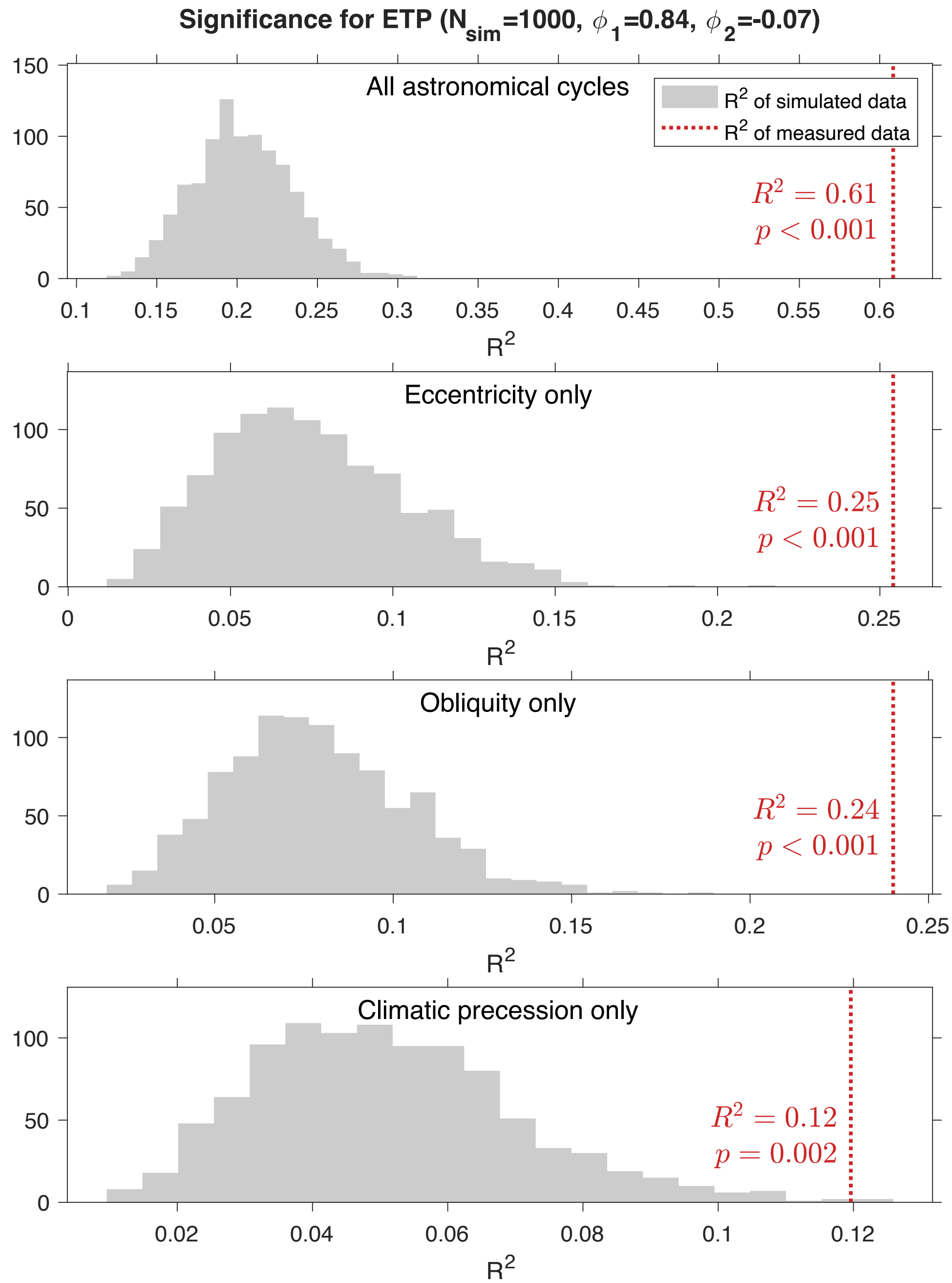

Figure 4. TimeOptB Monte Carlo significance testing for the synthetic ETP data set. The gray histograms show the distribution of TimeOptB R^2 values in N_{sim} = 1000 random AR(2) time series. The R^2 in the random time series matches or exceeds the value obtained for the synthetic ETP data set (red dotted line) at most two times out of 1000 when evaluating climatic precession alone, and does not exceed any of the simulated R^2 values when evaluating obliquity only, eccentricity only, or all of the astronomical cycles together.

4.2 Xiamaling Formation (1.4 Ga)

We apply the TimeOptB methodology to a published Cu/Al record from the 1.4 Ga Mesoproterozoic Xiamaling Formation, North China craton (Zhang et al., 2015), one of the data sets studied by MM18. The data interval is 2 m-thick and spans about 570 kyr (for the posterior mean sedimentation rate determined below).

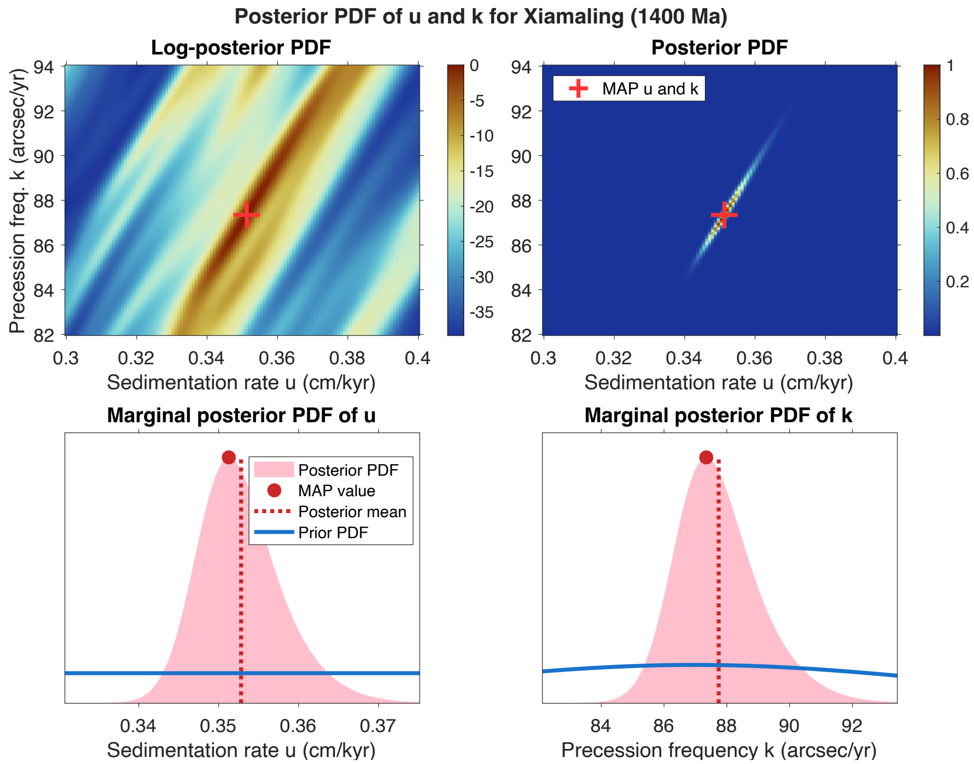

Figure 5. Posterior PDFs of sedimentation rate *u* and axial precession frequency *k* obtained by TimeOptB from the Xiamaling formation data set. In the PDF images, the log-posterior PDF is normalized to a MAP value of zero and the posterior PDF to a MAP value of 1.

The posterior PDFs of sedimentation rate *u* and axial precession frequency *k* are shown in Figure 5. The prior PDF of *k* is very broad, reflecting a large uncertainty about *k* at 1.4 Ga, but the data are informative and result in a much narrower posterior PDF. The MAP value of *u* and *k* predict data that match closely the precession-modulated climatic precession cycles in the measured Cu/Al data, and prominent peaks in the data periodogram are near the predicted frequencies of eccentricity and climatic precession (Figure 6). The Monte Carlo significance experiments in Figure 7 support the presence of astronomical cycles, with low *p*-values of 0.001 when all the astronomical cycles are considered or when only climatic precession is tested. The fit of an AR(2) process to the Xiamaling data is illustrated in Figure S2, and it confirms that the driving noise of the AR(2) process is nearly white noise.

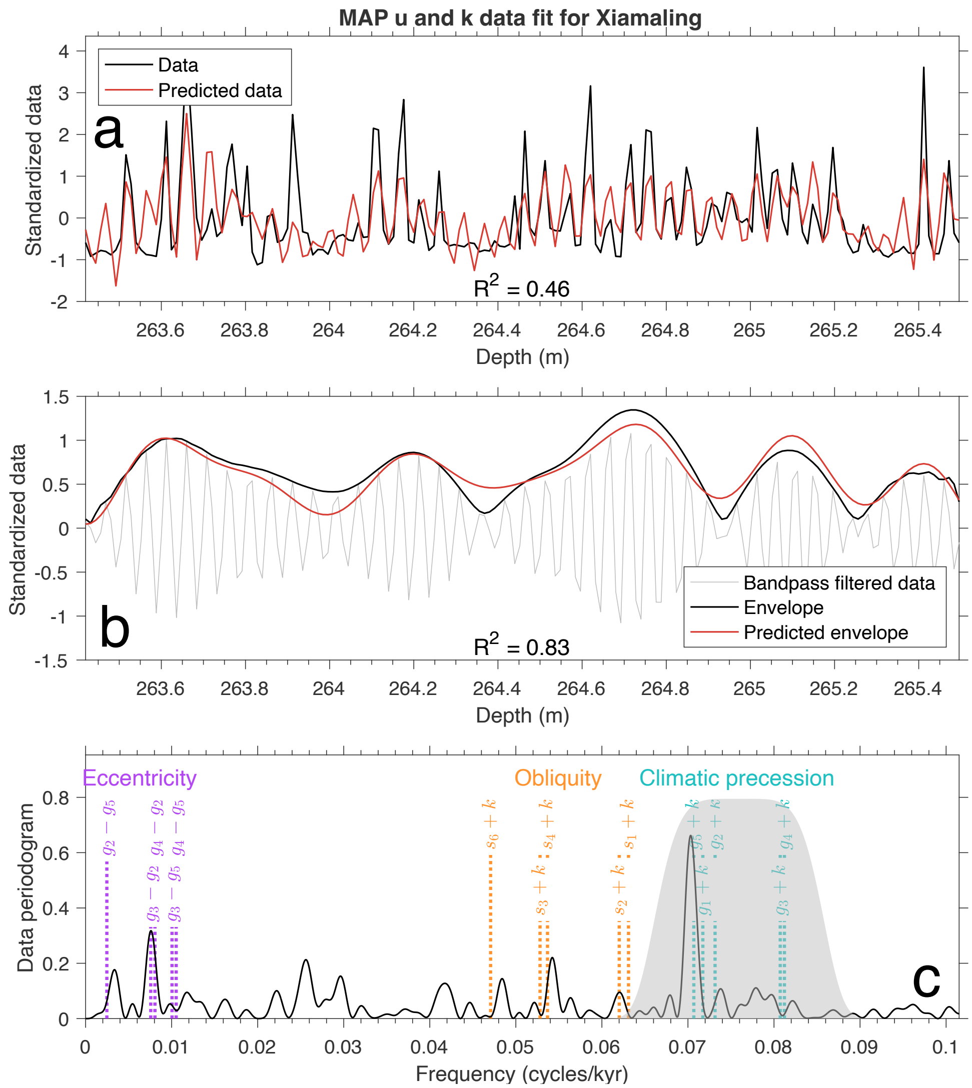

Figure 6. Fit to the Xiamaling formation Cu/Al data obtained by TimeOptB for the MAP value of sedimentation rate *u* and axial precession frequency *k* (see Table 3). (a) Fit between measured and predicted stratigraphic data (spectral fit). (b) Fit between the envelope of the bandpassed climatic precession signal and the envelope predicted by the eccentricity frequencies (envelope fit). (c) Data periodogram (black continuous line) and frequencies of astronomical cycles (dotted vertical lines). The gray shaded area shows the frequency response of the filter used to compute the bandpassed climatic precession signal in the data (gray curve in (b)).

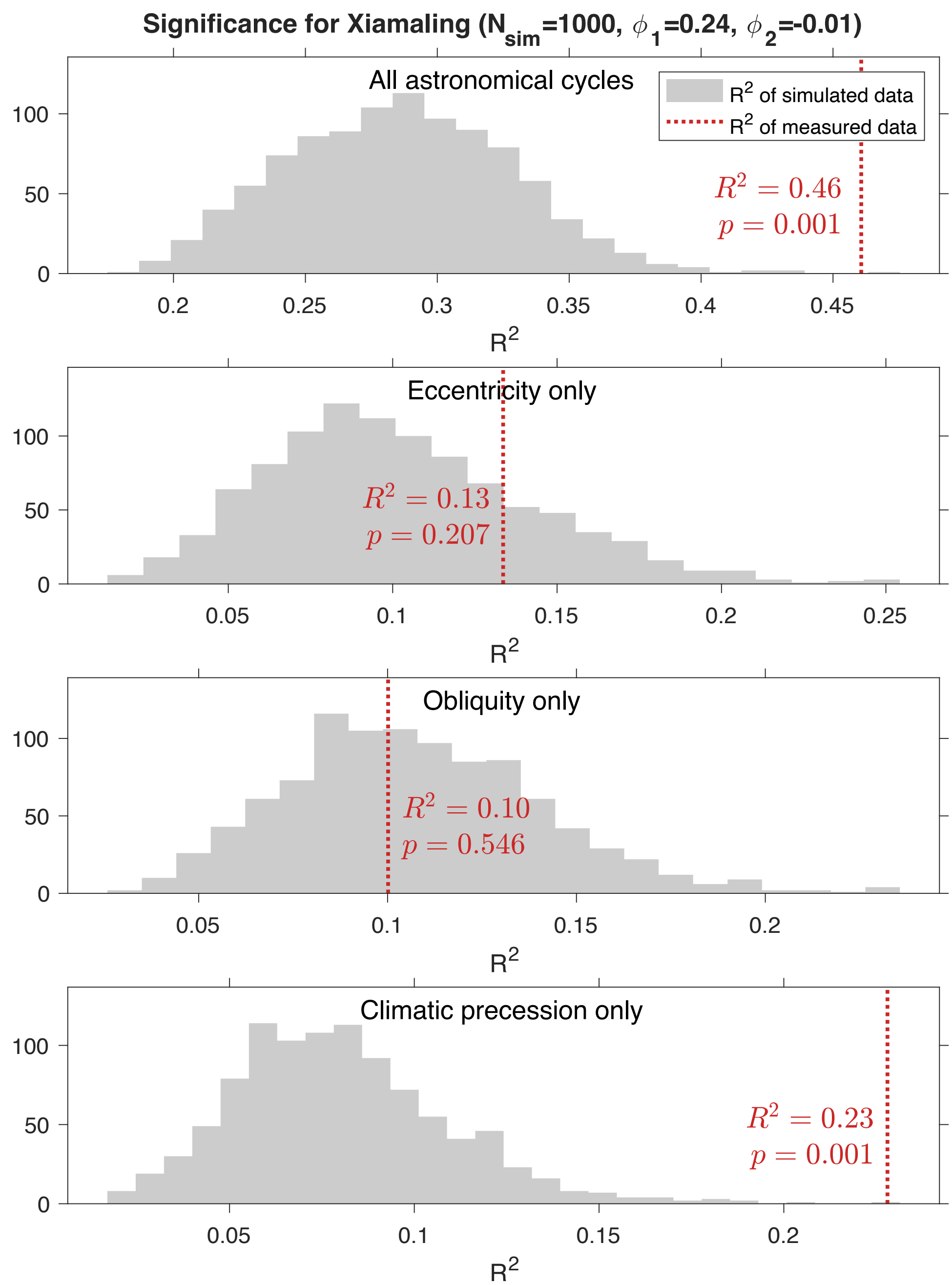

Figure 7. TimeOptB Monte Carlo significance testing for the Xiamaling formation data set. The gray histograms show the distribution of TimeOptB R^2 values in $N_{sim} = 1000$ random AR(2) time series. The R^2 values in the random time series are all clearly lower than the value obtained for the measured data (red dotted line) when considering all the astronomical cycles or the climatic precession cycles only.

Table 3. Results of TimeOptB and TimeOptBMCMC for the Xiamaling Formation Cu/Al and Walvis Ridge a* data. MAP = Maximum a posteriori, value of the parameter at the mode of the posterior PDF.

Xiamaling Formation (1.4 Ga)					
	MAP value	Posterior mean	Posterior σ	95% credible interval	Method
Sedimentation rate u	0.351	0.353	0.00540	0.343-0.365	TimeOptB
(cm/kyr)	0.353	0.352	0.00541	0.343-0.364	TimeOptBMCMC
Axial precession frequency k	87.34	87.74	1.38	85.37-90.81	TimeOptB
(arcsec/yr)	87.82	87.49	1.38	85.21-90.61	TimeOptBMCMC
Semi-major axis of lunar orbit a (Earth radii)		53.08	0.25		Based on TimeOptB posterior mean and σ
LOD (hrs)		18.47	0.22		
Walvis Ridge (55 Ma)					
	MAP value	Posterior mean	Posterior σ	95% credible interval	Method
Sedimentation rate u	1.308	1.309	0.00605	1.297-1.320	TimeOptB
(cm/kyr)	1.311	1.310	0.00620	1.299-1.322	TimeOptBMCMC
Axial precession frequency k	51.15	51.25	0.29	50.70-51.81	TimeOptB
(arcsec/yr)	51.31	51.29	0.29	50.75-51.85	TimeOptBMCMC
Semi-major axis of lunar orbit a (Earth radii)		60.07	0.12		Based on TimeOptB posterior mean and σ
LOD (hrs)		23.75	0.11		

4.3 Walvis Ridge ODP Site 1262 (55 Ma)

Another case study for the TimeOptB methodology uses a record of reflectivity data (a*, red/green) measured on Eocene-age sediments cored at ODP Site 1262, Walvis Ridge, South Atlantic (Zachos et al., 2004), which was also studied by MM18. We refer to that study and Zachos et al. (2004) for details about the a* data set. The data interval is 21 m-thick and spans about 1.6 Myr (for the posterior mean sedimentation rate determined below).

The posterior PDFs of sedimentation rate u and axial precession frequency k are illustrated in Figure 8. At 55 Ma, the prior PDF of k is much narrower than in the Mesoproterozoic example; the Walvis Ridge data point to values of k that are somewhat lower than those in the prior PDF. As in the previous example, the MAP values of u and k result in predicted data that closely reproduce the observed precession-modulated climatic precession cycles, and the predicted frequencies of eccentricity and climatic precession coincide with the highest peaks in the data periodogram (Figure 9). The periodogram of the Walvis record shows very little power at the expected frequencies of obliquity, and the Monte Carlo significance experiments show high significance for all astronomical cycles, for eccentricity only, and for climatic precession only (Figure 10). In contrast, the power of cycles at the obliquity frequencies in the random simulated data is always greater than in the measured data; the reason is that the obliquity frequency band (0.019-0.035 cycles/kyr) of the periodogram of the Walvis data has markedly lower power than that of the fitted AR(2) process (Figure S3a). Figure S3b shows that the driving noise of the fitted AR(2) process is nearly white noise.

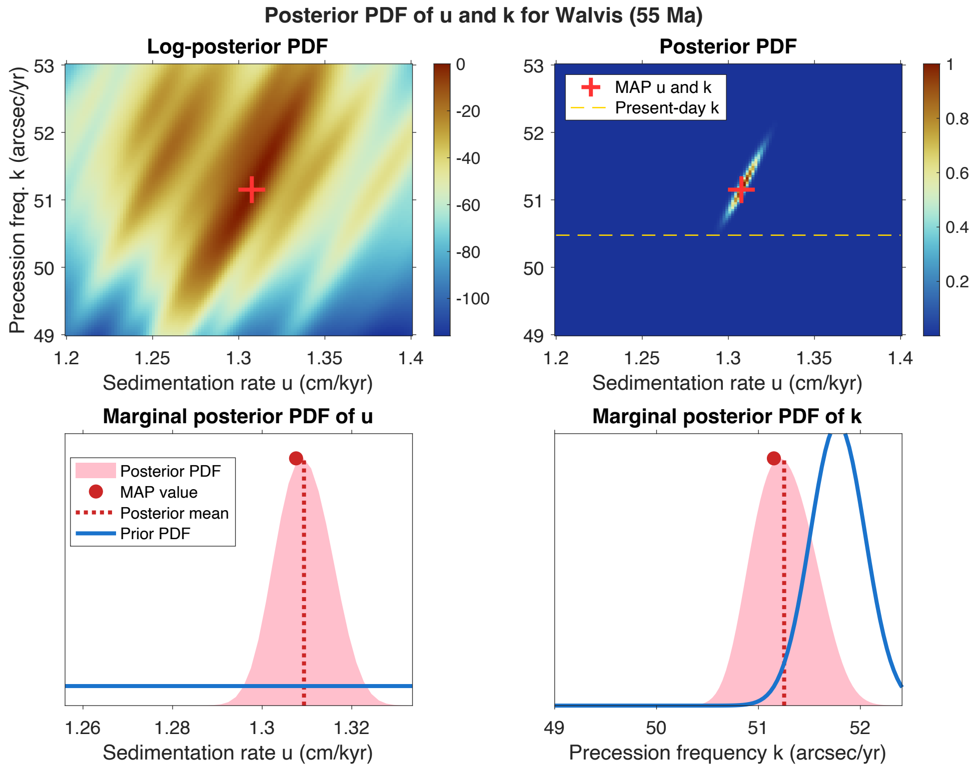

Figure 8. Posterior PDFs of sedimentation rate *u* and axial precession frequency *k* obtained by TimeOptB from the Walvis Ridge a* data set. In the PDF images, the log-posterior PDF is normalized to a MAP value of zero and the posterior PDF to a MAP value of 1. The horizontal dashed line in the posterior PDF image shows the present value of *k*.

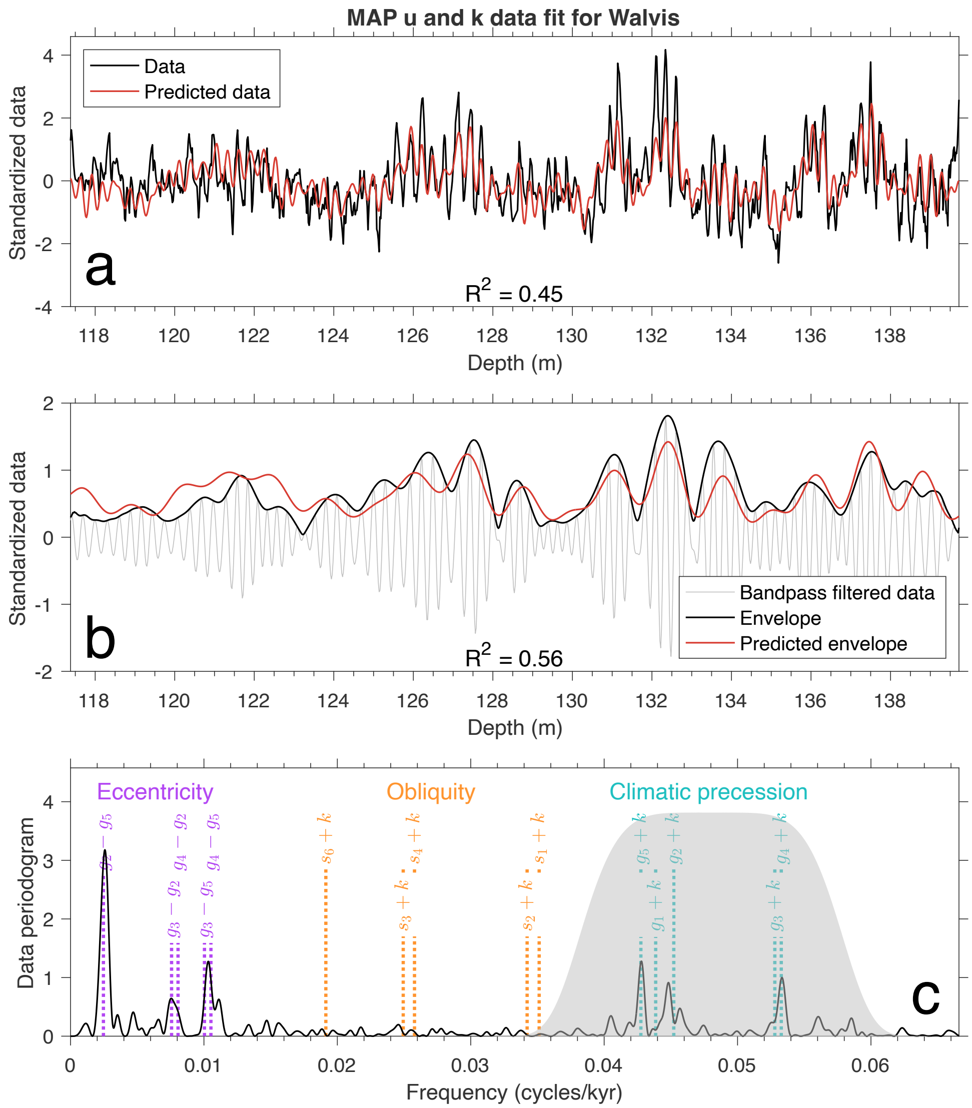

Figure 9. Fit to the Walvis Ridge a* data obtained by TimeOptB for the MAP value of sedimentation rate *u* and axial precession frequency *k* (see Table 3). (a) Fit between measured and predicted stratigraphic data (spectral fit). (b) Fit between the envelope of the bandpassed climatic precession signal and the envelope predicted by the eccentricity frequencies (envelope fit). (c) Data periodogram (black continuous line) and frequencies of astronomical cycles (dotted vertical lines). The gray shaded area shows the frequency response of the filter used to compute the bandpassed climatic precession signal in the data (gray curve in (b)).

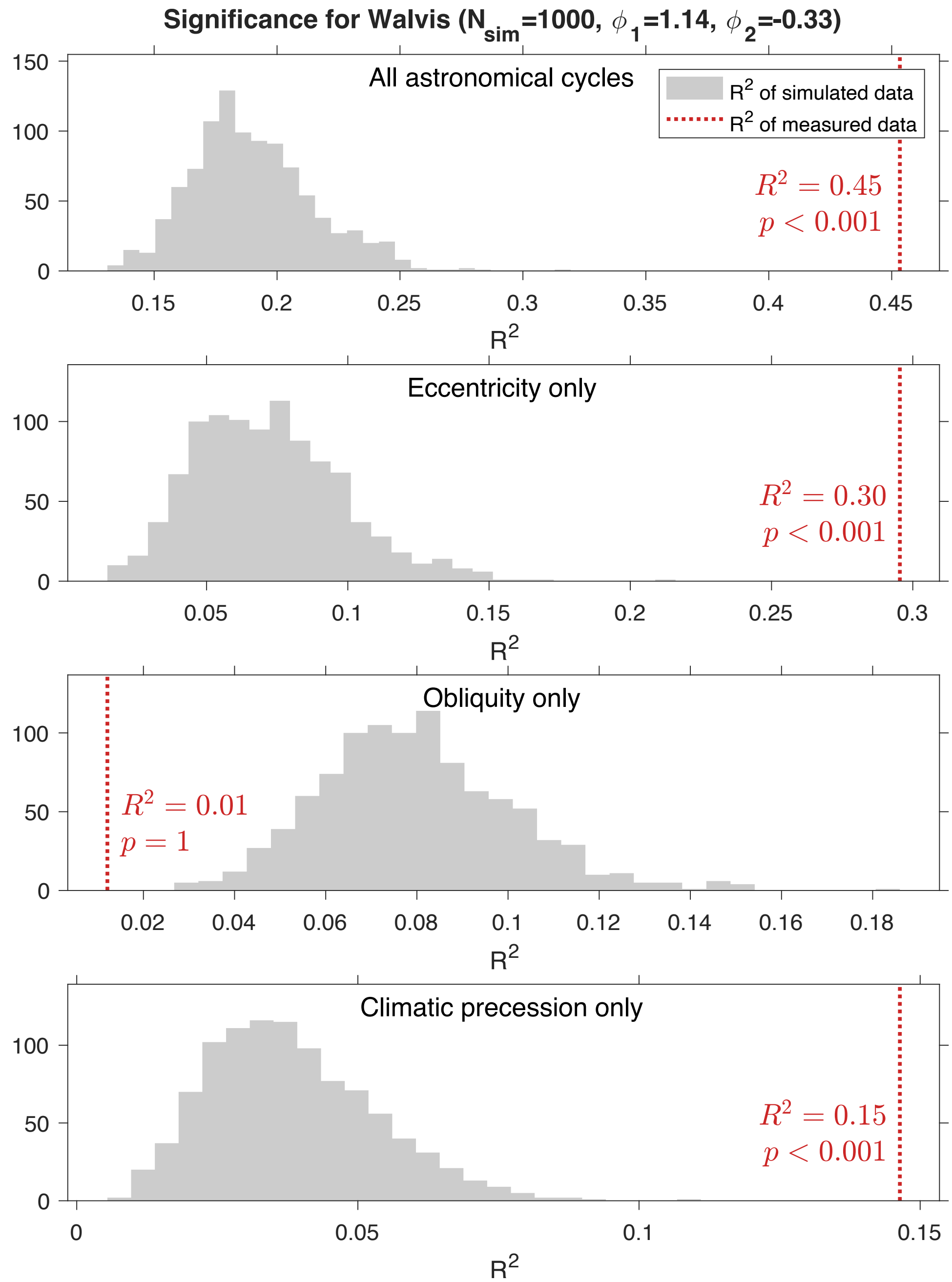

Figure 10. TimeOptB Monte Carlo significance testing for the Walvis Ridge data set. The gray histograms show the distribution of TimeOptB R^2 values in $N_{sim} = 1000$ random AR(2) time series. The R^2 values in the random time series are clearly lower than the value obtained for the measured data (red dotted line) when considering all the astronomical cycles, the eccentricity cycles only, or the climatic precession cycles only.

5 TimeOptBMCMC methodology

In the TimeOptB method, the only variable parameters are the sedimentation rate u and axial precession frequency k, while the fundamental Solar system frequencies g_i and s_i are kept fixed at their prior mean values. As noted in the Introduction, however, we also aim to use stratigraphic records to constrain the history of variation in the frequencies g_i and s_i and in long-term astronomical periodicities such as the $g_4 - g_3$ cycle. The method we present here is an offshoot of the TimeOptMCMC procedure of MM18, which sampled the posterior PDF of five fundamental Solar system frequencies g_i, axial precession frequency k, and sedimentation rate u (plus some hyperparameters, discussed below). In TimeOptBMCMC we add the five Solar system frequencies s_i to determine the posterior PDF of the full twelve-parameter vector in Equation 1.

Whereas in TimeOptB the value of the posterior PDF was calculated systematically over a grid of two parameters (u and k), the same strategy cannot be used for twelve parameters. Evaluating the PDF over a grid of M points for each parameter (say, $M = 100$) would require M^{12} calculations, which is entirely impractical. In contrast, MCMC algorithms perform a random walk that concentrates on the high-posterior probability region of the parameter space and are designed to return a sample distributed as in the posterior PDF. General treatments of MCMC in the statistical literature can be found in Gilks et al. (1996) and Brooks et al. (2011); examples of applications to geophysical inverse problems are in Malinverno (2002), Sambridge & Mosegaard (2002), Piana Agostinetti & Malinverno (2010), and Sen & Stoffa (2013).

TimeOptBMCMC uses a Metropolis-within-Gibbs algorithm (originally described by Metropolis et al., 1953): in each step of the random walk, a candidate parameter vector is obtained by adding to one of the parameters a random value chosen from a proposal PDF (e.g., a zero-mean normal PDF). The candidate is then accepted with a probability that depends on the ratio of the posterior PDFs of the candidate and the current parameter vector. This simple strategy will asymptotically return a sample of parameter vectors distributed as in the posterior PDF.

An outstanding issue in implementing a Metropolis algorithm is how to choose the scale parameter of the proposal PDF (e.g., the standard deviation of a normal PDF). If this scale is set too large, most candidates will not be accepted; if too small, the probability of acceptance will be large but the random walker will diffuse too slowly through the parameter space. In both cases, it will take a long time to explore the high-posterior probability region. In TimeOptBMCMC, we apply an adaptive Metropolis-within-Gibbs algorithm (Haario et al., 2001; Roberts & Rosenthal, 2009), where parameters are changed one at a time and the standard deviation of the normal proposal PDF of each parameter is progressively adjusted from a starting value to maintain a target rate of acceptance of 0.44, which has been shown to be optimal in this case (Roberts & Rosenthal, 2001). This is a significant improvement over TimeOptMCMC (MM18), which required a laborious initial experimentation, running a number of MCMC sampling chains to adjust the scale parameters of the proposal PDFs, often resulting in acceptance rates that were not ideal, which increased the computation time.

Another key difference is that TimeOptBMCMC applies an empirical Bayes approach to estimate directly from the data a best value of the hyperparameters that control the form of the covariance matrix of the residuals in the likelihood function for the spectral fit (two AR process coefficients and the residual variance; see the Supporting Information). TimeOptMCMC instead characterized the residuals with an AR(1) process, which is not always appropriate (e.g., the spectral fit residuals in the Walvis Ridge data required a nonzero coefficient ϕ_2; see Figure S3), and included the AR process coefficient and the variance of the residuals as variable hyperparameters in the inversion, which adds to the computational cost of the MM18 procedure.

6 TimeOptBMCMC example results

6.1 Xiamaling Formation (1.4 Ga)

The progress of TimeOptBMCMC in sampling the posterior PDF of all the parameters for the Xiamaling Cu/Al data is illustrated in Figure 11. The chain is started from the prior mean value of the Solar system frequencies g_i and s_i, axial precession frequency k, and sedimentation rate u (whose prior PDF is a uniform distribution between 0.3 and 0.4 cm/kyr) and proceeds for 50,000 iterations. The value of the posterior PDF rises very quickly at the start of the MCMC sampling chain and then fluctuates within the high-probability region (Figure 11a). The initial values of the standard deviation of each proposal PDF are set to the prior standard deviation of the astronomical frequencies (with an upper limit for the proposal standard deviation of k, where the prior standard deviation can be very large) and to a small fraction of the prior mean of u. The progressive adjustment of the proposal PDF standard deviations (Figure 11b-d) and the corresponding change in the frequency of acceptance (Figure 11e-g) show that after about 5,000 iterations the proposal standard deviations and the frequency of acceptance for each parameter fluctuate around a constant value, with an average frequency of acceptance around the optimal value of 0.44.

Figure 12 compares the prior PDFs of each parameter to the histograms of the values sampled by TimeOptBMCMC, which approximate each posterior PDF. The prior and posterior PDFs of the g_i and s_i frequencies are very similar, whereas the data clearly constrain the posterior values of k and u to a much narrower interval than in the prior PDF. Figure 12 also shows the posterior histogram of the period corresponding to the $g_4 - g_3$ frequency, which has a sizable posterior uncertainty (the central 95% interval of the posterior PDF is 1.47-3.78 Myr). The $g_4 - g_3$ frequency has a large uncertainty because it is estimated from a relatively short record that only spans a ~570 kyr interval. The present day value of the $g_4 - g_3$ period (2.4 Myr) is within the range consistent with the Xiamaling data at 1.4 Ga.

The posterior correlations between the parameters are generally small, with the exception of a strong positive correlation between u and k (Figure S4), which is the same positive correlation obtained in the TimeOptB results for the Xiamaling formation Cu/Al data (Figure 5). In contrast, the g_i and s_i frequencies are not correlated a posteriori with the sedimentation rate u because their prior variances are much smaller than that of k (Figure 1) so that they cannot vary over an interval large enough to relate to differences in $\underline{u}$. In fact, the sedimentation rate is mostly constrained by eccentricity frequencies $g_i - g_j$ in the spectral and envelope fit, which have a small prior variability. The marginal posterior PDFs of u and k obtained by TimeOptB (g_i and s_i fixed to their prior mean values) and TimeOptBMCMC (g_i and s_i variable) are also very similar (compare Figures 5 and 12 and the posterior PDF statistics in Table 3). Finally, the MAP value of the parameters sampled by TimeOptBMCMC results in predicted data that are essentially the same as those obtained by TimeOptB (compare Figures 6 and S5).

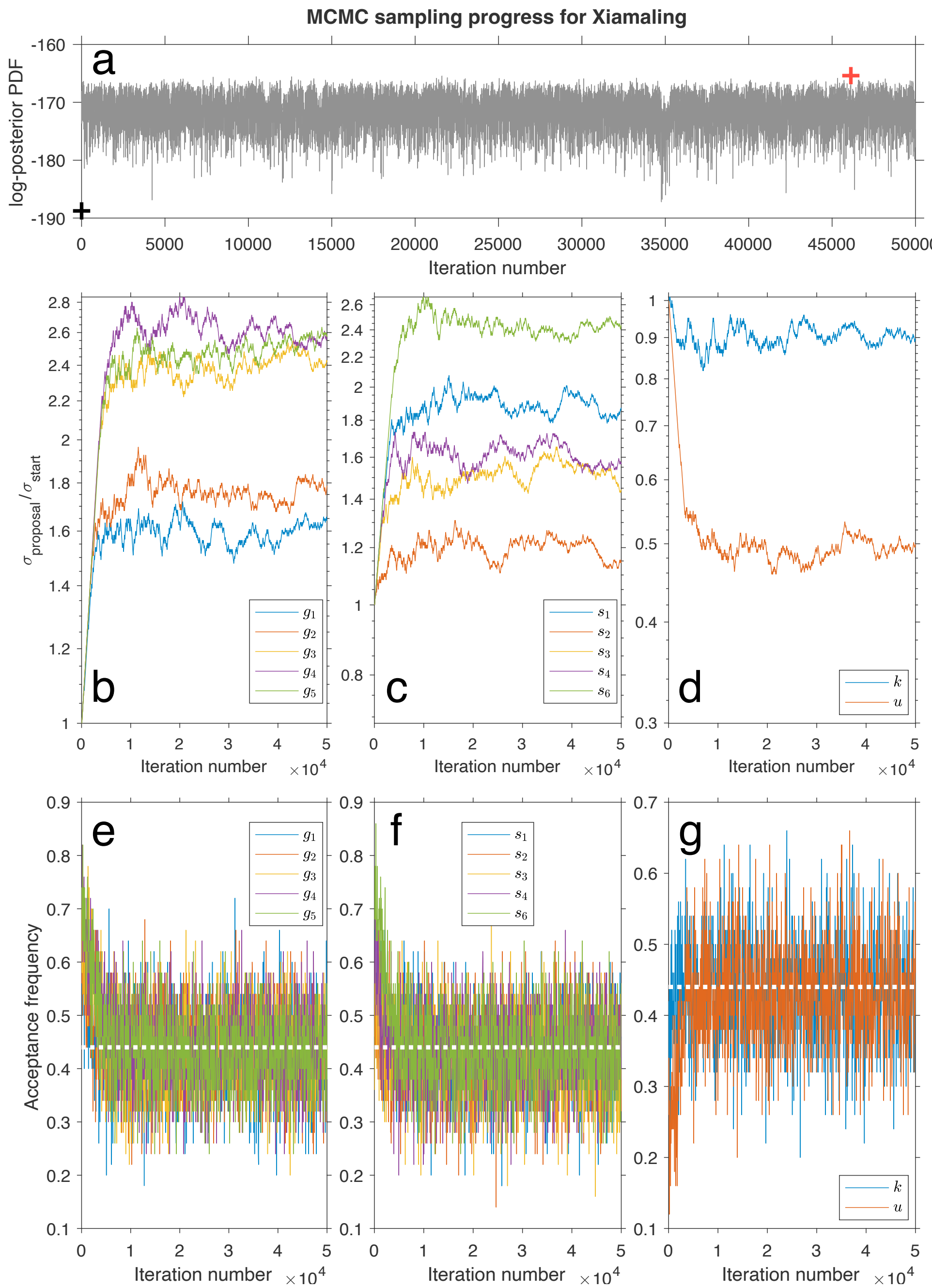

Figure 11. Progress of TimeOptBMCMC sampling for the Xiamaling formation Cu/Al data set over 50,000 iterations. (a) Value of the log-posterior PDF for the sampled model parameter vectors. The black cross is the starting value and the red cross the MAP. (b, c, d) Standard deviation of the proposal PDF (as a ratio over the starting value) for each model parameter. (e, f, g) Frequency of acceptance of the proposed steps in the MCMC random walk. The adaptive Metropolis algorithm used in TimeOptBMCMC adjusts the standard deviations of the proposal PDF to keep the frequency of acceptance around the optimal value of 0.44 for all model parameters (white horizontal dotted line).

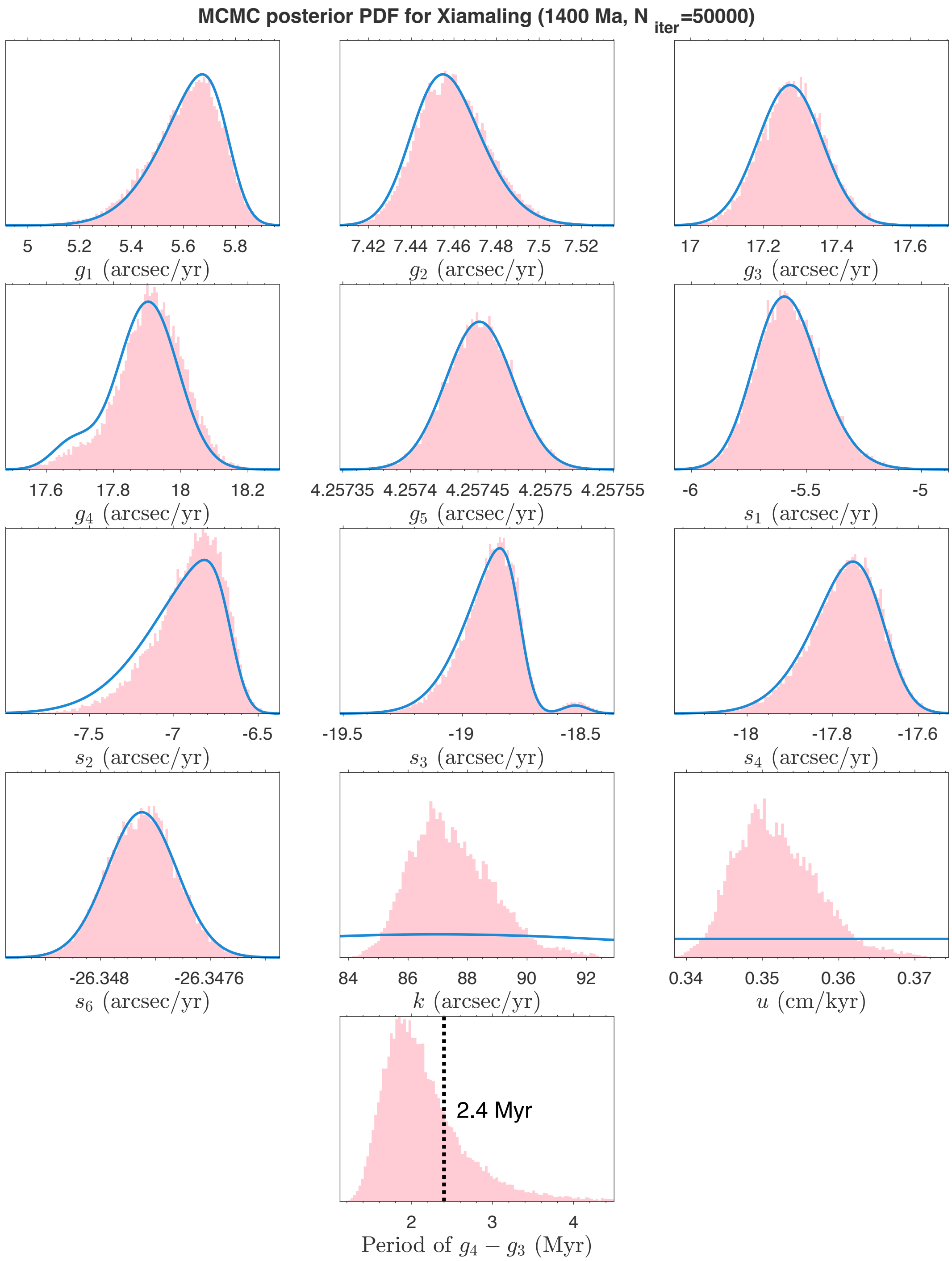

Figure 12. Histograms of posterior model parameter values sampled by TimeOptBMCMC for the Xiamaling formation Cu/Al data set over 50,000 iterations (light red) compared to the prior PDFs (blue curves). The bottom panel shows the posterior distribution of sampled $g_4 - g_3$ periods compared to the present day value of 2.4 Myr (vertical dotted black line).

6.2 Walvis Ridge ODP Site 1262 (55 Ma)

The progress of TimeOptBMCMC in sampling the posterior PDF for the Walvis Ridge a* data (Figure S6) is very similar to that seen for the Xiamaling formation Cu/Al data (Figure 11). The prior and posterior PDFs of g_i and s_i are also similar, with the exception of g_4, whose posterior PDF is shifted towards higher frequencies (Figure 13). As a result, the posterior PDF of the $g_4 - g_3$ period is shifted toward shorter periods, and the present day value of 2.4 Myr is in the tail of the posterior PDF (the central 95% interval of the posterior PDF is 1.69-2.38 Myr).

As seen for the Xiamaling formation Cu/Al data set, the posterior correlations in the Walvis Ridge a* results are generally small, except for the strong positive correlation between u and k (Figure S7) that was also seen in the TimeOptB results (Figure 8). Again, the marginal posterior PDFs of u and k obtained by TimeOptB (g_i and s_i fixed) and TimeOptBMCMC (g_i and s_i variable) are very similar (compare Figures 8 and 13 and posterior statistics in Table 3). The data predicted by the MAP value obtained by TimeOptB and TimeOptBMCMC are also essentially identical (compare Figures 9 and S8).

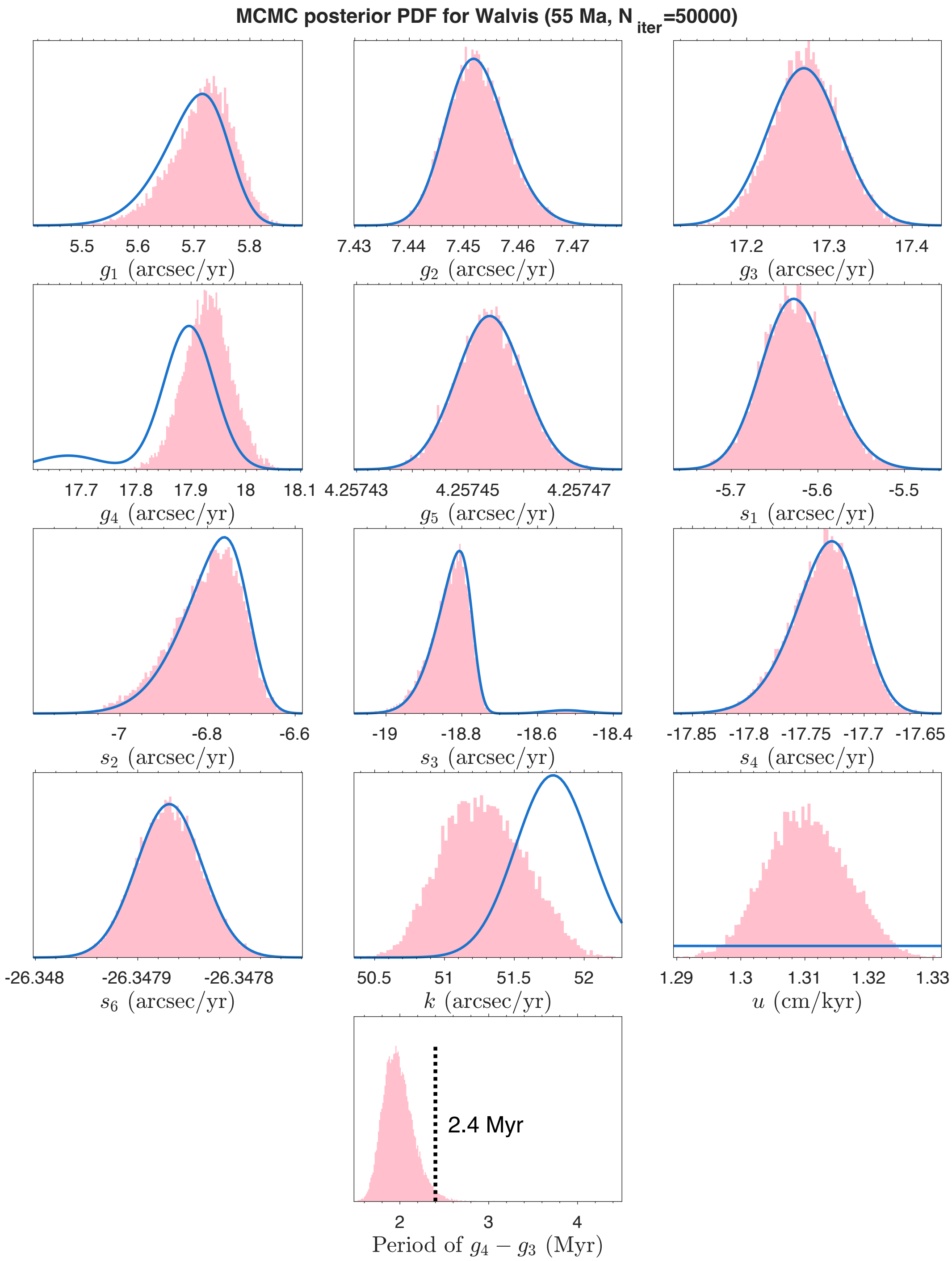

Figure 13. Histograms of posterior model parameter values sampled by TimeOptBMCMC for the Walvis Ridge formation a* data set over 50,000 iterations (light red) compared to the prior PDFs (blue curves). The bottom panel shows the posterior distribution of sampled $g_4 - g_3$ periods compared to the present day value of 2.4 Myr (vertical dotted black line).

7 Lunar distance and LOD from an estimate of axial precession frequency k

The axial precession frequency *k* depends on both the lunar distance *a* (semi-major axis of the Moon orbit) and the Earth spin rate ω (or equivalently, LOD); e.g., see Equation 7 of Berger & Loutre (1994) or Equation 4.14 of Laskar (2020). Therefore, obtaining values of lunar distance and LOD from an estimate of *k* requires an additional constraint, which is usually provided by the conservation of angular momentum in the Earth-Moon system (e.g., MM18; Lantink et al., 2022).

To estimate lunar distance and LOD from *k*, we apply two equations derived from Equations 6 and 7 of Walker & Zahnle (1986). The first equation gives the relationship between the axial precession frequency *k*, lunar distance *a*, and Earth spin rate ω as

$$\frac{\omega(t)}{\omega(0)} = \frac{k(t)}{k(0)} \frac{K+1}{K+\left[\frac{a(t)}{a(0)}\right]^{-3}} , \tag{5}$$

where $k(t)/k(0)$, $a(t)/a(0)$, and $\omega(t)/\omega(0)$ are ratios between values at age *t* and present day values, and *K* is a dimensionless constant. Equation 5 can be derived from the fundamental equation for the precession frequency *k* (e.g., Equation 7 of Berger & Loutre, 1994) for circular, coplanar orbits and a constant obliquity ε. The precession frequency equation also contains the dynamic ellipticity $H = (C - A)/C$, where *C* and *A* are the Earth's moments of inertia about polar and equatorial axes, respectively. *H* will change as the Earth changes its shape due to variations in spin rate ω, and Equation 5 is derived assuming that over long time scales the Earth deforms as a fluid in hydrostatic equilibrium so that *H* is proportional to the square of ω (Equation 5.3.2 of Munk & MacDonald, 1960). The dynamic ellipticity *H* can also change with changes in mass distribution within the Earth, e.g., because of the effects of glaciations or mantle convection. The resulting changes in *H*, however, are relatively small; they do not exceed about ~0.25% for glaciation effects in the last 47 Ma (Figure 3 of Farhat, Laskar, et al., 2022) or for mantle convection in the last 50 Ma (Figure 1A of Ghelichkhan et al., 2021). In contrast, the progressive decrease in the Earth spin rate due to tidal dissipation had much greater effects on *H* over geologic time scales. For example, a simple calculation of the effect of tidal dissipation (Equation 4.19 of Laskar, 2020) gives a ω^2 that was 3.2% greater than the present at 50 Ma, 6.6% greater at 100 Ma, and 22.5% greater at 300 Ma. This substantial systematic change justifies assuming that over time scales of tens of Myr the dynamic ellipticity *H* is primarily controlled by the progressive slowing down of the Earth spin rate.

The second equation is the relationship between lunar distance and LOD that conserves angular momentum in the Earth-Moon system:

$$\frac{\omega(t)}{\omega(0)} = 1 + A - A\left[\frac{a(t)}{a(0)}\right]^{1/2} . \tag{6}$$

The values of the dimensionless constants in Equations 5 and 6 were originally given by Walker & Zahnle (1986) as $K = 0.465$ and $A = 4.87$. Here we adjust the values of these dimensionless constants to account for effects that were originally neglected in Walker & Zahnle (1986): the systematic increase of obliquity ε during geologic time and the effect of solar ocean tides on the slowdown of the Earth spin rate. These adjustments were done by comparing the predictions of Equations 5 and 6 with the values of *a*, ω, and *k* calculated over the last 3.3 Ga by Farhat, Auclair-Desrotour, et al. (2022). The updated values of the constants are $K = 0.358$ and $A = 4.81$; details are in the Supporting Information.

The two relationships above define two curves of $\omega(t)/\omega(0)$ as a function of $a(t)/a(0)$: a "K-curve" that corresponds to a given value of $k(t)/k(0)$ (Equation 5) and an "AM-curve" that conserves angular momentum (Equation 6). The intersection of these two curves, illustrated in Figure 14a, gives the values of past lunar distance a, Earth spin rate ω, and LOD. The Supporting Information describes a simple way to obtain the intersection from a polynomial fit.

An uncertainty in the estimates of a and LOD can be calculated on the basis of the uncertainties in the K-curve and AM-curve and of the uncertainty in the value of k estimated from cyclostratigraphic data; see Figure 14b for an illustration and the Supporting Information for details of the calculation. The approach outlined above provides an accurate and quick means to obtain a, LOD, and their uncertainties from an estimate of k, and the results for the examples evaluated here are listed in Table 3.

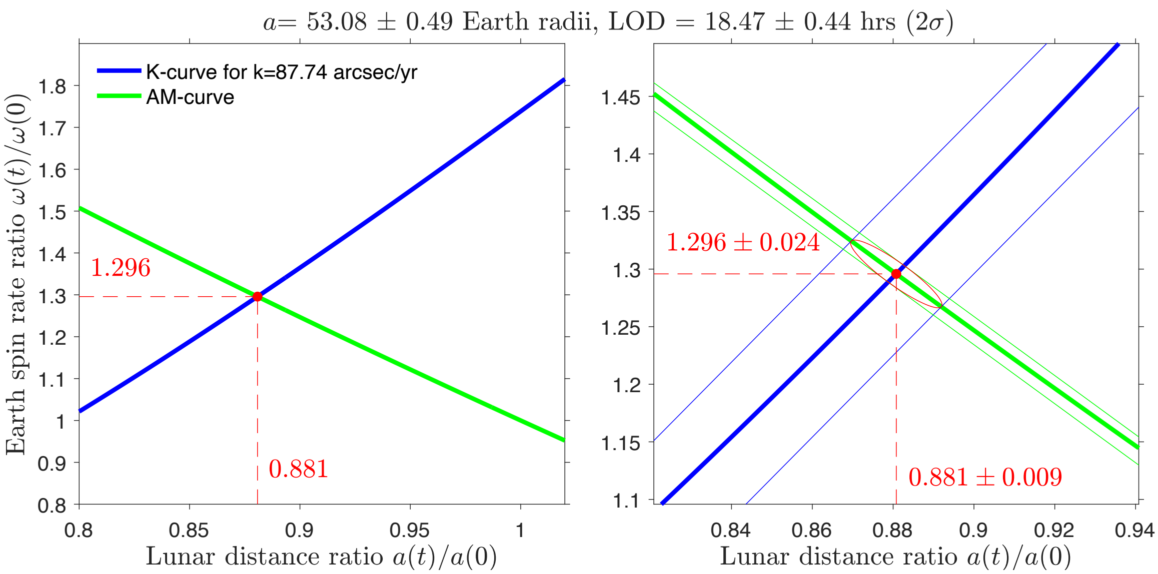

Figure 14. Ratio $\omega(t)/\omega(0)$ between the Earth's spin rate at age t and the present day value as a function of the ratio of the lunar distances $a(t)/a(0)$. (a) The blue curve (K-curve; Equation 5) shows the relationship for the axial precession frequency $k(t)$ estimated from the Xiamaling formation Cu/Al data set (t = 1.4 Ga) and the green curve (AM-curve; Equation 6) the relationship that conserves the Earth-Moon angular momentum. The red dot at the intersection of the two curves gives the values of $a(t)/a(0)$ and $\omega(t)/\omega(0)$ at age t. (b) Thin blue and green lines are the 95% contours of normal distributions that describe the uncertainties of the K-curve and AM-curve, respectively, and the red ellipse is the 95% contour that defines the uncertainty of the intersection (see the Supporting Information for details).

8 Discussion

8.1 Estimating axial precession frequency k

The case studies evaluated here show that TimeOptB and TimeOptBMCMC are effective in estimating a value for the precession frequency k from stratigraphic data. The key requirement is that the data should display clear eccentricity cycles (which do not depend on k) and clear climatic precession and/or obliquity cycles (which depend on k). The difference in the observed frequencies of eccentricity and those of climatic precession/obliquity allows for estimating k.

Thus, to estimate k it is important that sizable astronomical cycles are observed in the periodogram plots for eccentricity and either precession or obliquity (Figures 3, 6, and 9). The TimeOptB significance test should also weed out cases where stratigraphic sequences do not contain significant astronomical cycles (Figures 4, 7, and 10). It should be noted that the methods presented here will not work appropriately if sedimentation rate is not relatively constant within the analyzed stratigraphic interval, which requires careful selection of cyclostratigraphic data sets or portions thereof (more on the sedimentation rate assumption below). Considering these limitations, there should be many stratigraphic data sets that can return valid estimates of the past axial precession frequency. In addition to providing valuable information on the evolution of lunar distance, LOD, and tidal dissipation, past estimates of k will improve the accuracy of astrochronologies based on climatic precession and obliquity cycles in data.

The results in Table 3 supersede those obtained for the Xiamaling formation and Walvis Ridge in MM18. The differences are minor, and in both cases the posterior PDFs of k and u in this study overlap with those in MM18. The posterior PDFs of k in the Xiamaling formation are not identical with MM18 because TimeOptB and TimeOptBMCMC include the fit to obliquity components, which results in a small increase in k (the posterior mean changes from 85.79 arcsec/yr in MM18 to 87.74 arcsec/yr in Table 3). The posterior PDF of k in the Walvis Ridge data is very close to that in MM18 even though the prior PDF was different between the studies.

The posterior PDFs of u and k obtained by TimeOptB and TimeOptBMCMC are similar in both the Xiamaling Formation Cu/Al and Walvis Ridge a* data sets (Figures 5, 8, 12, 13, and Table 3). Thus, in these two case studies, letting the g_i and s_i frequencies be variable parameters does not lead to different estimates of u and k or to a substantially improved fit of the results (Figures 6, 9, S5, S8).

8.2 Estimating Solar system fundamental frequencies g_i and s_i

The posterior PDFs of the g_i and s_i frequencies sampled by TimeOptBMCMC are generally close to the respective priors, with the exception of g_4 in the Walvis Ridge a* data set. When astronomical cycles are well expressed in the data, this result shows that TimeOptBMCMC can constrain the values of Solar system fundamental frequencies. The past Solar system frequencies inferred from stratigraphic data will have inherent uncertainties. In practice, long-period cycles such as $g_4 - g_3$ will not be reconstructed with high accuracy from stratigraphic records of relatively short duration, but nonetheless the range of their possible values can be estimated by TimeOptBMCMC. For example, although the posterior PDF of the $g_4 - g_3$ period in the Walvis Ridge record (55 Ma) spans a broad interval, the results in Figure 13 suggest a $g_4 - g_3$ period that is shorter than the present 2.4 Myr. For comparison, Zeebe & Lourens (2019) also found that the Solar system solution that best fit the Walvis Ridge data displayed a decrease in the $g_4 - g_3$ period to ~1.5 Myr at ages older than 50 Ma (though 1.5 Myr is at the very low end of the posterior PDF of the $g_4 - g_3$ period in Figure 13).

A suggested practical procedure is to run TimeOptB first on a data set, including the Monte Carlo significance experiments to support the presence of astronomical cycles in the data. If there is evidence for astronomical cycles in the data, a TimeOptBMCMC run can show whether the sampled values of the Solar system fundamental frequencies are distributed as in the prior PDF, meaning that the data are not informative (as in the case of the Xiamaling formation Cu/Al data set) or whether there are differences from the prior that highlight past variations (as for g_4 and $g_4 - g_3$ in the Walvis Ridge a* data set).

8.3 Assumption: constant sedimentation rate

A key assumption in TimeOptB and TimeOptBMCMC is that the sedimentation rate was constant in the studied stratigraphic interval. A preliminary moving window power spectral analysis or wavelet-based analysis can indicate whether prominent cycles have nearly constant spatial frequencies as predicted by a constant sedimentation rate. This means that suitable data sets will likely span a relatively short time interval, and there will be a tradeoff between the need to have a long enough record of eccentricity cycles and the requirement of a constant sedimentation rate. Also, the strategy presented here will not be reliable if astronomical signals are distorted by large cyclic changes in sedimentation rate driven by the effects of particular astronomical cycles (e.g., Herbert, 1994).

Even when sedimentation rate is nearly constant over the interval studied, the examples presented here show that any error in estimating sedimentation rate u will result in the same error in axial precession frequency k: a sedimentation rate overestimated by 1% means k will be overestimated by 1% (see the discussion of the ETP data set results in Figure 2). There is no way to know k within a small fraction of its value unless the sedimentation rate, or more generally the time-stratigraphic depth relationship, is also known within that same small fraction. As stratigraphic data invariably contain variations unrelated to astronomical forcing ("geological noise;" Meyers, 2019), the time-depth relationship can be determined only approximately. This is a fundamental issue at the root of cyclostratigraphy and astrochronology applications, and it cannot be solved by methodological improvements. On the other hand, methods such as those presented here can quantify the resulting uncertainty and highlight the value and the limitations of conclusions drawn from the analysis of astronomical cycles in stratigraphic records.

8.4 Assumption: constant Earth-Moon angular momentum

As noted earlier, the axial precession frequency k depends on both lunar distance a and LOD. Estimating both a and LOD on the basis of k therefore requires an additional constraint, which we impose by using the common assumption that the Earth-Moon angular momentum remained constant throughout Earth's history (with a correction due to the small effect of Solar ocean tides in slowing down the Earth's rotation).

In contrast, Zahnle & Walker (1987) and Bartlett & Stevenson (2016) proposed that when LOD decreased to ~21 hrs in the Proterozoic, a solar atmospheric tide became resonant with the Earth's spin rate and counteracted the effect of the lunar ocean tide, maintaining a constant Earth spin rate for a prolonged duration (between ~2 Ga and ~1 Ga; Bartlett & Stevenson, 2016). During this interval, the lunar ocean tide would still have resulted in a torque that moved the Moon to a higher orbit, so that the total angular momentum of the Earth-Moon system would have increased through time by as much as 10-20%, extracting angular momentum from the Earth's orbit around the Sun (Zahnle & Walker, 1987). Our results give some information on the possible size of the change in the Earth-Moon angular momentum if this were the case: taking the value of $k(t)$ estimated from the Xiamaling formation Cu/Al record and assuming that LOD was 21 hours rather than keeping the Earth-Moon angular momentum to its present value, Equation 5 gives a ratio $a(t)/a(0) = 0.834$. If the Earth was spinning with a LOD of 21 hrs and the lunar distance was 83.4% of the present value, the Earth-Moon angular momentum at 1.4 Ga would have been approximately 95% of the present value.

By themselves, estimates of the past axial precession frequency from cyclostratigraphy will only constrain a combination of lunar distance and LOD, and an additional independent constraint is needed to determine both parameters. In particular, estimates of LOD from cyclostratigraphy (e.g., as used by Mitchell & Kirscher, 2023) assume conservation of the present Earth-Moon angular momentum and cannot provide a test of the hypothesis of a constant LOD in the Proterozoic, because if LOD remained constant during an extended period the Earth-Moon angular momentum had to increase with time.

9 Conclusions

We presented here two methods, TimeOptB and TimeOptBMCMC, to determine the frequencies of astronomical cycles in the geologic past recorded by stratigraphic sequences. The results show a decrease in the Earth's axial precession frequency from about 88.2 arcsec/year (a period of 14.7 kyr) in the Mesoproterozoic (1.4 Ga) to 51.2 arcsec/year (25.3 kyr) in the Eocene (55 Ma). Our results imply that at 1.4 Ga Earth days were ~18.4 hours long and that the Moon was 12% closer to the Earth compared to the present (assuming that the angular momentum of the Earth-Moon system was conserved).

Stratigraphic data invariably contain "geological noise" unrelated to astronomical forcing, and resultant estimates of astronomical frequencies are inevitably uncertain. By applying a Bayesian formulation, we determine posterior probability distributions that describe how much each astronomical frequency can vary while fitting the observed data. We also describe a Monte Carlo procedure to test whether astronomical cycles are significant over a noisy background of sediment property variations.

A key assumption of our methods is that sedimentation rate remains constant in the studied interval. This conservative requirement keeps the analysis simple and ensures that recovered astronomical cycles are not the result of overfitting due to arbitrary changes in sedimentation rate. We plan to investigate relaxing this assumption in future developments, for example using "sedimentation templates" (Meyers, 2019) or age models defined by a number of age-depth tie points (e.g., Haslett & Parnell, 2008). Variations in the age-depth relationship should be kept as small as possible to avoid overfitting, and a sound significance analysis should be performed to help guard against artificially identifying astronomical cycles.

While the constant sedimentation rate assumption restricts the range of suitable cyclostratigraphic records, the examples shown here demonstrate that relatively short stratigraphic intervals (spanning as little as ~600 kyr) provide valid estimates of past astronomical frequencies. The methods we presented are well suited to recover from the geological record the history of variation in the Earth's axial precession frequency, the fundamental Solar system frequencies, and the periods of the resultant astronomical insolation rhythms. The results will be useful to constrain the past history of the Earth-Moon and Solar system, to inform models of past tidal dissipation, and to improve astrochronology estimates, especially those based on climatic precession and obliquity cycles.

Acknowledgments

The authors declare no conflict of interest. This work was funded by the Heising-Simons Foundation under grant # 2021-2798 (Malinverno) and grant # 2021-2797 (Meyers). The authors thank the members of the CycloAstro project for valuable feedback throughout the development of this study.

Open Research

The Xiamaling Formation Cu/Al data (Zhang et al., 2015) and Walvis Ridge a* data (Zachos et al., 2004) used as examples in this study have been previously published and are also accessible with the function ‘getData’ of the ‘Astrochron’ package for R (Meyers, 2014). The prototype code for the TimeOptB and TimeOptBMCMC analyses presented in this work was created in MATLAB and is being used and tested on additional data by CycloAstro graduate students advised by the authors, whose results have not been yet published. The algorithm will be made available in the ‘Astrochron’ package following publication of additional results and translation into the free statistical software R.

References

Ajibade, R., Meyers, S. R., Lantink, M. L., Malinverno, A., Hinnov, L. A., & Carroll, A. R. (2023). Earth-Moon History and Astronomical Parameters: Constraints from the Permian and Proterozoic. Presented at the 2023 AGU Fall Meeting, Abstract PP21D-1339, San Francisco, CA.

Andersen, N. (1974). On the calculation of filter coefficients for maximum entropy spectral analysis. *Geophysics*, *39*(1), 69–72. https://doi.org/10.1190/1.1440413

Bartlett, B. C., & Stevenson, D. J. (2016). Analysis of a Precambrian resonance-stabilized day length. *Geophys. Res. Lett.*, *43*, 5716–5724, doi:10.1002/2016GL068912.

Berger, A., & Loutre, M. F. (1994). Astronomical forcing through geological time. *Spec. Publ. Int. Ass. Sediment.*, *19*, 15–24.

Brooks, S., Gelman, A., Jones, G. L., & Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Boca Raton, Florida: Chapman & Hall/CRC.

Burg, J. P. (1967). Maximum entropy spectral analysis. Presented at the 37th Annual International Meeting, Society of Exploration Geophysicists, Oklahoma City, Oklahoma.

Carlin, B. P., & Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton, Florida: Chapman and Hall/CRC.

Casella, G. (1985). An Introduction to Empirical Bayes Data Analysis. *The American Statistician*, *39*(2), 83–87. https://doi.org/10.2307/2682801

Chatfield, C. (1989). *The Analysis of Time Series: An Introduction*. London: Chapman and Hall.

Cox, D. R., & Miller, H. D. (1965). *The Theory of Stochastic Processes*. London: Chapman and Hall.

Darwin, S. G. H. (1898). *The Tides and Kindred Phenomena in the Solar System: The Substance of Lectures Delivered in 1897 at the Lowell Institute, Boston, Massachusetts*. London: John Murray.

Dettmer, J., Molnar, S., Steininger, G., Dosso, S. E., & Cassidy, J. F. (2012). Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models. *Geophysical Journal International*, *188*(2), 719–734. https://doi.org/10.1111/j.1365-246X.2011.05302.x

Farhat, M., Laskar, J., & Boué, G. (2022). Constraining the Earth's Dynamical Ellipticity From Ice Age Dynamics. *Journal of Geophysical Research: Solid Earth*, *127*(5), e2021JB023323. https://doi.org/10.1029/2021JB023323

Farhat, M., Auclair-Desrotour, P., Boué, G., & Laskar, J. (2022). The resonant tidal evolution of the Earth-Moon distance. *Astron. Astrophys.*, *665*, L1. https://doi.org/10.1051/0004-6361/202243445

Fitzpatrick, R. (2012). *An Introduction to Celestial Mechanics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139152310

Gelman, A. B., Carlin, J. S., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis* (2nd ed.). Boca Raton, Florida: Chapman and Hall/CRC.

Ghelichkhan, S., Fuentes, J. J., Hoggard, M. J., Richards, F. D., & Mitrovica, J. X. (2021). The precession constant and its long-term variation. *Icarus*, *358*, 114172. https://doi.org/10.1016/j.icarus.2020.114172

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London: Chapman and Hall.

Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, *7*(2), 223–242.

Hacking, I. (2001). *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press.

Haslett, J., & Parnell, A. (2008). A simple monotone process with application to radiocarbon-dated depth chronologies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *57*(4), 399–418. https://doi.org/10.1111/j.1467-9876.2008.00623.x

Hays, J. D., Imbrie, J., & Shackleton, N. J. (1976). Variations in the Earth's orbit: Pacemaker of the Ice Ages. *Science*, *194*, 1121–1132.

Herbert, T. D. (1994). Reading orbital signals distorted by sedimentation: Models and examples. *Spec. Publ. Int. Ass. Sediment.*, *19*, 483–507.

Hinnov, L. A. (2013). Cyclostratigraphy and its revolutionizing applications in the earth and planetary sciences. *GSA Bulletin*, *125*(11–12), 1703–1734. https://doi.org/10.1130/B30934.1

Hoang, N. H., Mogavero, F., & Laskar, J. (2021). Chaotic diffusion of the fundamental frequencies in the Solar System. *Astronomy & Astrophysics*, *654*, A156. https://doi.org/10.1051/0004-6361/202140989

Lantink, M. L., Davies, J. H. F. L., Ovtcharova, M., & Hilgen, F. J. (2022). Milankovitch cycles in banded iron formations constrain the Earth–Moon system 2.46 billion years ago. *Proceedings of the National Academy of Sciences*, *119*(40), e2117146119. https://doi.org/10.1073/pnas.2117146119

Laskar, J. (2020). Chapter 4 - Astrochronology. In F. M. Gradstein, J. G. Ogg, M. D. Schmitz, & G. M. Ogg (Eds.), *Geologic Time Scale 2020* (pp. 139–158). Elsevier. https://doi.org/10.1016/B978-0-12-824360-2.00004-8

Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A. C. M., & Levrard, B. (2004). A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.*, *428*, 261–285.

Ma, C., Meyers, S. R., & Sageman, B. B. (2017). Theory of chaotic orbital variations confirmed by Cretaceous geological evidence. *Nature*, *542*(7642), 468–470. https://doi.org/10.1038/nature21402

Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophys. J. Int.*, *151*, 675–688. https://doi.org/10.1046/j.1365-246X.2002.01847.x

Malinverno, A., & Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, *69*, 1005–1016. https://doi.org/10.1190/1.1778243

Mann, M. E., & Lees, J. M. (1996). Robust estimation of background noise and signal detection in climatic time series. *Climatic Change*, *33*, 409–445.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.*, *21*, 1087–1092.

Meyers, S. R. (2014). Astrochron: An R Package for Astrochronology. Retrieved from https://cran.r-project.org/package=astrochron

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*(12), 1625–1640. https://doi.org/10.1002/2015PA002850

Meyers, S. R. (2019). Cyclostratigraphy and the problem of astrochronologic testing. *Earth-Science Reviews*, *190*, 190–223. https://doi.org/10.1016/j.earscirev.2018.11.015

Meyers, S. R., & Malinverno, A. (2018). Proterozoic Milankovitch cycles and the history of the solar system. *Proc. Nat. Acad. Sci.*, *115*, 6363–6368. https://doi.org/10.1073/pnas.1717689115

Meyers, S. R., & Peters, S. E. (2022). Exploring the depths of Solar System evolution. *Proceedings of the National Academy of Sciences*, *119*(43), e2216309119. https://doi.org/10.1073/pnas.2216309119

Milanković, M. (1941). *Canon of insolation and the ice-age problem: (Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem) Belgrade, 1941*. Jerusalem: Israel Program for Scientific Translations; [available from U.S. Department of Commerce, Clearinghouse for Federal Scientific and Technical Information, Springfield, Va.].

Mitchell, R. N., & Kirscher, U. (2023). Mid-Proterozoic day length stalled by tidal resonance. *Nature Geoscience*, *16*(7), 567–569. https://doi.org/10.1038/s41561-023-01202-6

Munk, W. H., & MacDonald, G. J. F. (1960). *The Rotation of the Earth: A Geophysical Discussion*. Cambridge, UK: Cambridge University Press.

Neal, R. N. (1993). *Probabilistic inference using Markov chain methods* (Technical Report CRG-TR-93-1). Department of Computer Science, University of Toronto. Retrieved from https://www.cs.toronto.edu/~radford/ftp/review.pdf

Olsen, P. E., Laskar, J., Kent, D. V., Kinney, S. T., Reynolds, D. J., Sha, J., & Whiteside, J. H. (2019). Mapping Solar System chaos with the Geological Orrery. *Proceedings of the National Academy of Sciences*, *116*(22), 10664–10673. https://doi.org/10.1073/pnas.1813901116

Pälike, H., Laskar, J., & Shackleton, N. J. (2004). Geologic constraints on the chaotic diffusion of the Solar System. *Geology*, *32*, 929–932. https://doi.org/10.1130/G20750.1

Piana Agostinetti, N., & Malinverno, A. (2010). Receiver function inversion by trans-dimensional Monte Carlo sampling. *Geophys. J. Int.*, *181*, 858–872. https://doi.org/10.1111/j.1365-246X.2010.04530.x

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. New York: Academic Press.

Roberts, G. O., & Rosenthal, J. S. (2001). Optimal Scaling for Various Metropolis-Hastings Algorithms. *Statistical Science*, *16*(4), 351–367.

Roberts, G. O., & Rosenthal, J. S. (2009). Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics*, *18*(2), 349–367. https://doi.org/10.1198/jcgs.2009.06134

Sambridge, M., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Rev. Geophys.*, *40(3)*, 1009. https://doi.org/10.1029/2000RG000089

Sen, M. K., & Stoffa, P. L. (2013). *Global Optimization Methods in Geophysical Inversion* (2nd ed.). Cambridge, New York: Cambridge University Press.

Sinnesael, M., De Vleeschouwer, D., Zeeden, C., Batenburg, S. J., Da Silva, A.-C., de Winter, N. J., et al. (2019). The Cyclostratigraphy Intercomparison Project (CIP): consistency, merits and pitfalls. *Earth-Science Reviews*, *199*, 102965. https://doi.org/10.1016/j.earscirev.2019.102965

Smith, D. G. (2023). The Orbital Cycle Factory: Sixty cyclostratigraphic spectra in need of re-evaluation. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *628*, 111744. https://doi.org/10.1016/j.palaeo.2023.111744

Ulrych, T. J., & Bishop, T. N. (1975). Maximum entropy spectral analysis and autoregressive decomposition. *Reviews of Geophysics*, *13*(1), 183–200. https://doi.org/10.1029/RG013i001p00183

Walker, J. C. G., & Zahnle, K. J. (1986). Lunar nodal tide and distance to the Moon during the Precambrian. *Nature*, *320*, 600–602.

Waltham, D. (2015). Milankovitch Period Uncertainties and Their Impact On Cyclostratigraphy. *Journal of Sedimentary Research*, *85*(8), 990–998. https://doi.org/10.2110/jsr.2015.66

Weedon, G. P. (2022). Problems with the current practice of spectral analysis in cyclostratigraphy: Avoiding false detection of regular cyclicity. *Earth-Science Reviews*, *235*, 104261. https://doi.org/10.1016/j.earscirev.2022.104261

Wu, Y., Malinverno, A., Meyers, S. R., & Hinnov, L. A. (2023). Earth's axial precession frequency, length of day, lunar distance, and insolation periodicities over the past 650 Myr from cyclostratigraphy. Presented at the 2023 AGU Fall Meeting, Abstract ID 1340823, San Francisco, CA.

Zachos, J. C., Kroon, D., Blum, P., & others. (2004). *Proc. ODP, Init. Repts., 208*. College Station, TX: Ocean Drilling Program.

Zahnle, K., & Walker, J. C. G. (1987). A constant daylength during the Precambrian era? *Precambrian Res.*, *37*, 95–105.

Zeebe, R. E., & Lourens, L. J. (2019). Solar System chaos and the Paleocene–Eocene boundary age constrained by geology and astronomy. *Science*, *365*(6456), 926–929. https://doi.org/10.1126/science.aax0612

Zeeden, C., Kaboth, S., Hilgen, F. J., & Laskar, J. (2018). Taner filter settings and automatic correlation optimisation for cyclostratigraphic studies. *Computers & Geosciences*, *119*, 18–28. https://doi.org/10.1016/j.cageo.2018.06.005

Zhang, S., Wang, X., Hammarlund, E. U., Wang, H., Costa, M. M., Bjerrum, C. J., et al. (2015). Orbital forcing of climate 1.4 billion years ago. *Proceedings of the National Academy of Sciences*, *112*(12), E1406–E1413. https://doi.org/10.1073/pnas.1502239112

Zięba, A. (2010). Effective Number of Observations and Unbiased Estimators of Variance for Autocorrelated Data - an Overview. *Metrology and Measurement Systems*, *17*(1), 3–16. https://doi.org/10.2478/v10178-010-0001-0

Zięba, A., & Ramza, P. (2011). Standard Deviation of the Mean of Autocorrelated Observations Estimated with the Use of the Autocorrelation Function Estimated From the Data. *Metrology and Measurement Systems*, *18*(4), 529–542. https://doi.org/10.2478/v10178-011-0052-x

Geochemistry, Geophysics, Geosystems

Supporting Information for

Bayesian estimation of past astronomical frequencies, lunar distance, and length of day from sediment cycles

A. Malinverno[1] and S. R. Meyers[2]

[1]Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York, USA

[2]Department of Geoscience, University of Wisconsin-Madison, Madison, Wisconsin, USA

Contents of this file

Text S1 to S4
Table S1
Figures S1 to S8

Text S1. Calculation of the Likelihood for the Spectral Fit

To calculate the data $\mathbf{d}_{\mathsf{pred}}$ predicted by parameters in $\mathbf{m}$, the first step is to transform the stratigraphic depth in the data to an age vector $\mathbf{t}$ (which is zero at the top) using the sedimentation rate u in $\mathbf{m}$. (Symbols and acronyms used here are listed in Table S1.) To compute the predicted data, we then construct a matrix $\mathbf{G}$ whose columns contain sine and cosine signals as a function of ages in $\mathbf{t}$ with the frequencies of eccentricity, obliquity, and climatic precession listed in Table 1 of the main text calculated for the values of g_i, s_i, and k in $\mathbf{m}$.

We then obtain by least squares the amplitudes of the sine and cosine terms in a vector $\mathbf{y}$ fitted to the data as

$$\mathbf{y} = (\mathbf{G}^{\mathsf{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathsf{T}}\mathbf{d}, \tag{S1}$$

where the data in $\mathbf{d}$ were linearly detrended and standardized to zero mean and unit variance. The data predicted by the astronomical frequencies are

$$\mathbf{d}_{\mathsf{pred}} = \mathbf{G}\,\mathbf{y}. \tag{S2}$$

The sum of a sine and a cosine function of the same frequency and amplitudes in the corresponding elements of $\mathbf{y}$ results in a fitted sinusoidal function with the same frequency and a phase determined by the data in $\mathbf{d}$. This is a "spectral fit" in the sense that sinusoidal functions of arbitrary phase are fitted to the data, but we stress that it is not based on an estimate of the power spectrum.

The calculation of the spectral likelihood is based on fitting an AR(2) process to the vector $\mathbf{e}$ of residuals that are the difference between observed and predicted data as in

$$\mathbf{e} = \mathbf{d} - \mathbf{d}_{\mathsf{pred}}. \tag{S3}$$

Following an empirical Bayes strategy, we estimate the AR(2) coefficients ϕ_1 and ϕ_2 from $\mathbf{e}$ with a method originally due to Burg (1967) that is based on minimizing the prediction error variance computed both in the forward and backward direction (Andersen, 1974; Ulrych & Bishop, 1975). Using the fitted AR(2) coefficients, we then compute the vector $\mathbf{w}$ of the driving noise from

$$w_i = e_i - \phi_1 e_{i-1} - \phi_2 e_{i-2}. \tag{S4}$$

If $\mathbf{e}$ can be successfully modeled as an AR(2) process, the vector $\mathbf{w}$ should be close to uncorrelated white noise. This can be checked from the sample autocorrelation of $\mathbf{w}$, shown in Figures S1b to S3b for the data sets examined here.

If residuals $\mathbf{e}$ are modeled as an AR process, the spectral fit likelihood can be written as the multivariate normal PDF of the noise $\mathbf{w}$ (Dettmer et al., 2012), which is

$$p(\mathbf{d} \mid \mathbf{m}) = p(\mathbf{w} \mid \mathbf{m}) = \frac{1}{(2\pi)^{N/2}\,\sigma_w^N} \exp\left[-\frac{\mathbf{w}^\mathsf{T}\mathbf{w}}{2\sigma_w^2}\right], \tag{S5}$$

where N is the number of data points in $\mathbf{d}$ and $\mathbf{w}$. Applying again empirical Bayes, we estimate the variance σ_w^2 from $\mathbf{w}$ as

$$\sigma_w^2 = \frac{\mathbf{w}^\mathsf{T}\mathbf{w}}{N}. \tag{S6}$$

Substituting this estimate of σ_w^2 in Equation S5 we obtain a final expression for the spectral fit likelihood that is

$$p(\mathbf{d} \mid \mathbf{m}) = p(\mathbf{w} \mid \mathbf{m}) = \frac{1}{(2\pi)^{N/2}\,\sigma_w^N} \exp\left[-\frac{N}{2}\right], \tag{S7}$$

Text S2. Calculation of the Likelihood for the Envelope Fit

In the envelope fit, the observed data vector $\mathbf{d}$ is the amplitude envelope of the climatic precession signal in the data and $\mathbf{d}_{\mathsf{pred}}$ is the envelope predicted by eccentricity signals with the frequencies given by the values of g_i in $\mathbf{m}$. The precession envelope of the data is calculated by first extracting the signal in the climatic precession frequency band applying a Taner bandpass filter with a roll-off rate of 10^7 (Zeeden et al., 2018). The cutoff frequencies equal the minimum climatic precession frequency minus 0.005 cycles/kyr and the maximum climatic precession frequency plus 0.005 cycles/kyr. The envelope of the climatic precession signal is then computed from its Hilbert transform and is standardized to have zero mean and unit variance.

The predicted precession envelope is calculated by fitting the eccentricity frequencies given by the g_i values in $\mathbf{m}$ to the precession envelope of the data; the result is then standardized to zero mean and unit variance. The procedure is the same as that described for the spectral likelihood, except that the matrix $\mathbf{G}$ contains only sine and cosine terms for the eccentricity frequencies. The residual vector $\mathbf{e}$ in the envelope fit is the difference between the precession envelope extracted from the data by bandpass filtering and that predicted by the eccentricity frequencies defined by the values in $\mathbf{m}$.

The envelope likelihood calculation is based on an effective number of observations

$$N_{\text{eff}} = N/\tau < N, \tag{S8}$$

where $\tau > 1$ is the lag where the autocorrelation of the envelope residuals decays to zero (see Section 2.2 in the main text). Consequently, if we were to subsample the vector of residuals $\mathbf{e}$ by taking one every τ values, we would obtain a vector of uncorrelated residuals $\mathbf{e}_{\text{sub}}$ of length N_{eff}. The likelihood for these subsampled uncorrelated residuals would be a multivariate normal PDF as in

$$p(\mathbf{d} \mid \mathbf{m}) = p(\mathbf{e}_{\text{sub}} \mid \mathbf{m}) = \frac{1}{(2\pi)^{N_{\text{eff}}/2}\,\sigma_e^{N_{\text{eff}}}} \exp\left[-\frac{\mathbf{e}_{\text{sub}}^{\mathsf{T}}\mathbf{e}_{\text{sub}}}{2\sigma_e^2}\right]. \tag{S9}$$

As the vector $\mathbf{e}_{\text{sub}}$ consists of N_{eff} elements, the sum of their squared values will be approximately the same as the sum of squares of $\mathbf{e}$ times the ratio N_{eff}/N as in

$$\frac{\mathbf{e}_{\text{sub}}^{\mathsf{T}}\mathbf{e}_{\text{sub}}}{\sigma_e^2} \approx \frac{\mathbf{e}^{\mathsf{T}}\mathbf{e}}{\sigma_e^2}\,\frac{N_{\text{eff}}}{N} = \frac{\mathbf{e}^{\mathsf{T}}\mathbf{e}}{\tau\sigma_e^2}, \tag{S10}$$

which shows that correlations in the vector $\mathbf{e}$ result in an increase of the variance σ_e^2 by a factor τ. If we follow an empirical Bayes strategy and estimate σ_e^2 from the sample variance of $\mathbf{e}$ as in

$$\sigma_e^2 = \frac{\mathbf{e}^{\mathsf{T}}\mathbf{e}}{N}, \tag{S11}$$

the likelihood of the subsampled vector of residuals can be written as

$$p(\mathbf{d} \mid \mathbf{m}) = p(\mathbf{e}_{\text{sub}} \mid \mathbf{m}) = \frac{1}{(2\pi)^{N_{\text{eff}}/2}\,\sigma_e^{N_{\text{eff}}}} \exp\left[-\frac{N_{\text{eff}}}{2}\right]. \tag{S12}$$

(Note that the calculation of the envelope fit likelihood never requires the hypothetical vector $\mathbf{e}_{\text{sub}}$.)

We estimate the lag τ from a simple model of the autocorrelation of envelope fit residuals that contain periodic components with eccentricity frequencies. The autocorrelation will first cross zero at a lag that is approximately a quarter of the wavelength λ_{ecc} of the shortest eccentricity cycle in the data, so that

$$\tau \approx \frac{\lambda_{\text{ecc}}}{4\Delta z}, \tag{S13}$$

where Δz is the data sampling interval and λ_{ecc} will equal uT_{ecc}, where u is the sedimentation rate and T_{ecc} is the period of the shortest eccentricity cycle (e.g., 100 kyr), so that

$$\tau = \frac{uT_{\text{ecc}}}{4\Delta z}. \tag{S14}$$

However, making τ proportional to a variable sedimentation rate has the effect of inducing a systematic bias that makes the envelope likelihood in Equation S12 substantially larger at lower sedimentation rates where τ is lower and N_{eff} is larger. (The measured and predicted envelope data are standardized to unit variance, so that the residual $\sigma_e < 1$; for a given value of σ_e, the likelihood will be greater for a greater value of N_{eff}.) Numerical experiments show that this sedimentation rate bias overwhelms the effect of differences in σ_e, and the highest likelihood for the envelope fit is invariably at the lowest sedimentation rate considered. To avoid this bias, the value of τ is calculated from Equation S14 at a reference value of u, set to the average sedimentation rate in the prior range considered.

Text S3. Lunar Distance and LOD from the Axial Precession Frequency

In this section, we describe our approach and provide easy-to-use expressions to obtain lunar distance and LOD from estimates of the axial precession frequency. To simplify the notation, henceforth we denote the ratios between past and present values of axial precession frequency k, lunar distance a, and Earth spin rate ω as follows:

$$k_r = \frac{k(t)}{k(0)}, \qquad a_r = \frac{a(t)}{a(0)}, \qquad \omega_r = \frac{\omega(t)}{\omega(0)}, \tag{S15}$$

where t is age and 0 denotes present day. With this notation, Equation 5 in the main text for the K-curve is

$$\omega_r = k_r \frac{K+1}{K+a_r^{-3}} \tag{S16}$$

and Equation 6 for the AM-curve is

$$\omega_r = 1 + A - A\sqrt{a_r} \tag{S17}$$

The values of the constants K and A in Equation S16 and S17 were adjusted from the original values in Walker & Zahnle (1986) to better fit the relationship between lunar distance, Earth spin rate, and axial precession frequency in the tidal dissipation model results of Farhat et al. (2022), which take into account the long-term increase in the obliquity ε of the Earth's axis and the effect of solar ocean tides in slowing down Earth's rotation (see Section 7 in the the main text).

The value of k that can be predicted for given values of a and ω by rearranging Equation S16 and the original value of $K = 0.465$ in Walker & Zahnle (1986) results in a difference with the k computed by Farhat et al. (2002) that increases with increasing age, reaching 3.3% at 3.3 Ga. This misfit matches the predicted effect of the secular trend in obliquity: as the full expression for the axial precession frequency contains a $\cos\varepsilon$ term (e.g., Berger & Loutre, 1994; Laskar, 2020), the ratio $k(t)/k(0)$ at age t will be multiplied by a factor $\cos\varepsilon(t)/\cos\varepsilon(0)$, which equals 1.032 at 3.3 Ga in the results of Farhat et al. (2022). By adjusting the value of K to 0.358, the difference between the k predicted from Equation S16 and that computed by Farhat et al. (2022) remains within ±0.14% between the present and 3 Ga, increasing to 0.28% at 3.3 Ga.

We also compared the LOD predicted by Equation S17 for a given lunar distance a to the values computed by Farhat et al. (2022). The original value of $A = 4.87$ in Walker & Zahnle (1986) results in a difference in LOD with the Farhat et al. (2022) values that increases with increasing age, reaching 0.55% at 3.3 Ga. With an adjusted value of $A = 4.81$, this difference remains between ±0.03% from the present to 3 Ga, reaching −0.07% at 3.3 Ga.

At the intersection of the K-curve and of the AM-curve, the value of ω_r in Equation S16 and S17 must be the same, and we obtain an expression for the ratio k_r that conserves Earth-Moon angular momentum and is a function of a_r only:

$$k_r = \left(1 + A - A\sqrt{a_r}\right)\frac{K + a_r^{-3}}{K+1}. \tag{S18}$$

This nonlinear equation cannot be solved directly for a_r. However, an accurate value can be obtained by computing the value of k_r from Equation S18 for a_r between 1 and 0.77 (the value at 3.3 Ga in the results of Farhat et al. 2022) and fitting a polynomial to a_r as a function of k_r. The relationship between a_r and the natural logarithm of k_r is almost linear, and the third order polynomial in the following equation fits the values of a_r to within $\pm 3 \times 10^{-5}$:

$$a_r = 1 - 0.217194 \log(k_r) - 0.00060922 [\log(k_r)]^2 + 0.00621404 [\log(k_r)]^3. \tag{S19}$$

Once the ratio a_r is obtained from k_r using Equation S19, ω_r can be computed from Equation S17.

Text S4. Uncertainty in Estimated Lunar Distance and LOD

The K-curve and AM-curve will intersect at a point of coordinates $\hat{a}_r$ and $\hat{\omega}_r$, which define the lunar distance and Earth spin rate at a past time from an estimate of the axial precession frequency (Figure 14a in the main text). This section describes how to obtain the uncertainty in $\hat{a}_r$ and $\hat{\omega}_r$ from the uncertainty in an estimated value of axial precession frequency ratio $\hat{k}_r$. There are three sources of uncertainty to take into account in this problem:

- Uncertainty in the precession frequency ratio $\hat{k}_r$ estimated from stratigraphic data, quantified by a standard deviation $\sigma_{\hat{k}_r}$;
- Uncertainty in the value of ω_r predicted by the K-curve in Equation S16 for a given a_r and k_r, quantified by a standard deviation σ_{K};
- Uncertainty in the value of ω_r predicted by the AM-curve in Equation S17 for a given a_r, quantified by a standard deviation σ_{AM}.

The standard deviation $\sigma_{\hat{k}_r}$ equals the posterior standard deviation of k divided by the present day axial precession frequency $k(0)$. A thorough determination of the standard deviations σ_{K} and σ_{AM} requires quantifying and propagating uncertainties in the constants that define the K-curve and AM-curve (in Equation S16 and S17), and is beyond the scope of the present study. We use here a value of 0.005 for both σ_{K} and σ_{AM}, meaning that the K-curve and AM-curve predict the value of the ratio ω_r with a standard deviation of 0.5%. This value is well above the misfit in fitting the values calculated by Farhat et al. (2022) with the adjusted values of K in Equation S16 and of A in Equation S17 (see Text S3), thus we take it as providing a conservative measure of the the uncertainty.

The uncertainty of the AM-curve in the region around the intersection at $\hat{a}_r$ and $\hat{\omega}_r$, where we approximate the AM-curve by a straight line, can be represented by a bivariate normal PDF that is centered on the intersection and has a covariance matrix

$$\mathbf{C}_{\mathsf{AM}} = \sigma_0^2 \begin{bmatrix} 1 & b_{\mathrm{AM}} \\ b_{\mathrm{AM}} & b_{\mathrm{AM}}^2 + \frac{\sigma_{\mathrm{AM}}^2}{\sigma_0^2} \end{bmatrix}, \tag{S20}$$

where b_{AM} is the slope of the AM-curve (Equation S17) at the intersection

$$b_{\mathrm{AM}} = \left.\frac{d\omega_r}{da_r}\right|_{a_r=\hat{a}_r} = -\frac{A}{2\sqrt{\hat{a}_r}} \tag{S21}$$

and σ_0 spans a range of a_r that is large compared to the overlap between the uncertain AM-curve and K-curve; see Figure 14b in the main text for an illustration. (As shown below, σ_0 will be eliminated from our final expressions.)

The bivariate normal PDF that represents the uncertainty of the K-curve in the region around the intersection is

$$\mathbf{C}_{\mathsf{K}} = \sigma_0^2 \begin{bmatrix} 1 & b_{\mathrm{K}} \\ b_{\mathrm{K}} & b_{\mathrm{K}}^2 + \frac{\sigma_{\mathrm{K}}^2 + \sigma_{\hat{k}_r}^2}{\sigma_0^2} \end{bmatrix}, \tag{S22}$$

where b_{K} is the slope of the K-curve (Equation S16) at the intersection

$$b_{\mathrm{K}} = \left.\frac{d\omega_r}{da_r}\right|_{a_r=\hat{a}_r} = \hat{k}_r \frac{3\,(K+1)\,\hat{a}_r^2}{(K\hat{a}_r^3+1)^2} \tag{S23}$$

and the sum $\sigma_{\mathrm{K}}^2 + \sigma_{\hat{k}_r}^2$ accounts for the uncertainty of both the K-curve and of the estimated k_r. (Equation S16 shows that the uncertainty in the value of ω_r predicted by k_r is the same as the uncertainty of k_r.)

The uncertainty of the intersection will be defined by the product of the two PDFs that quantify the uncertainty of the AM-curve and of the K-curve. This product is another bivariate normal PDF that is centered on the intersection point and that has a covariance matrix

$$\mathbf{C} = \left[\mathbf{C}_{\mathsf{K}}^{-1} + \mathbf{C}_{\mathsf{AM}}^{-1}\right]^{-1}, \tag{S24}$$

and an expression for $\mathbf{C}$ can be obtained letting $\sigma_0^2 \to \infty$ as

$$\mathbf{C} = \frac{1}{(b_{\mathrm{K}} - b_{\mathrm{AM}})^2} \begin{bmatrix} \sigma_{\mathrm{AM}}^2 + \sigma_{\mathrm{K}}^2 + \sigma_{\hat{k}_r}^2 & b_{\mathrm{AM}}(\sigma_{\mathrm{K}}^2 + \sigma_{\hat{k}_r}^2)^2 + b_{\mathrm{K}}\sigma_{\mathrm{AM}}^2 \\ b_{\mathrm{AM}}(\sigma_{\mathrm{K}}^2 + \sigma_{\hat{k}_r}^2) + b_{\mathrm{K}}\sigma_{\mathrm{AM}}^2 & b_{\mathrm{AM}}^2(\sigma_{\mathrm{K}}^2 + \sigma_{\hat{k}_r}^2) + b_{\mathrm{K}}^2\sigma_{\mathrm{AM}}^2 \end{bmatrix}. \tag{S25}$$

The diagonal elements of $\mathbf{C}$ contain the variances of the values of $\hat{a}_r$ (C_{11}) and $\hat{\omega}_r$ (C_{22}) at the intersection of the AM-curve and K-curve.

Table S1. List of symbols and acronyms.

a	Semi-major axis of lunar orbit
a_r	Ratio $a(t)/a(0)$ of a at age t over the present day value
A	Constant in equation for conservation of angular momentum
$\mathbf{C}$	Covariance matrix
$\mathbf{d}$	Vector of observed data
$\mathbf{d}_{\mathsf{pred}}$	Vector of data predicted by a given value of $\mathbf{m}$
$\mathbf{e}$	Vector of data residuals $\mathbf{d} - \mathbf{d}_{\mathsf{pred}}$
$\mathbf{e}_{\mathsf{sub}}$	Subsampled vector of data residuals
$\mathbf{G}$	Matrix of sine and cosine terms with astronomical frequencies
g_i	Fundamental Solar system frequencies for the rotation of the planetary perihelia
k	Precession frequency of the Earth's spin axis
k_r	Ratio $k(t)/k(0)$ of k at age t over the present day value
$\mathbf{m}$	Vector of parameters (g_i, s_i, k, u)
N	Number of data points in $\mathbf{d}$
N_{eff}	Effective number of independent observations in $\mathbf{d}$
$p(x \mid y)$	Probability density function (PDF) of x given y
P	Order of an AR(P) process
r_i	Autocorrelation of data residuals $\mathbf{e}$ at lag i
s_i	Fundamental Solar system frequencies for the rotation of the ascending nodes of the orbital planes
$\mathbf{t}$	Vector of ages in the stratigraphic data
T_{ecc}	Period of shortest eccentricity cycle in the data
u	Sedimentation rate
$\mathbf{w}$	Vector of uncorrelated (white) noise driving an autoregressive (AR) process
λ_{ecc}	Wavelength of shortest eccentricity cycle in the data
ϕ_i	i-th coefficient of an AR(P) process ($1 \leq i \leq P$)
σ_e^2	Variance of data residuals $\mathbf{e}$
σ_w^2	Variance of uncorrelated noise $\mathbf{w}$
τ	Minimum lag where the autocorrelation function crosses zero
ω	Earth spin rate
ω_r	Ratio $\omega(t)/\omega(0)$ of ω at age t over the present day value
AR	Autoregressive process
LOD	Length of day
MAP	Maximum a posteriori
MCMC	Markov chain Monte Carlo
PDF	Probability density function

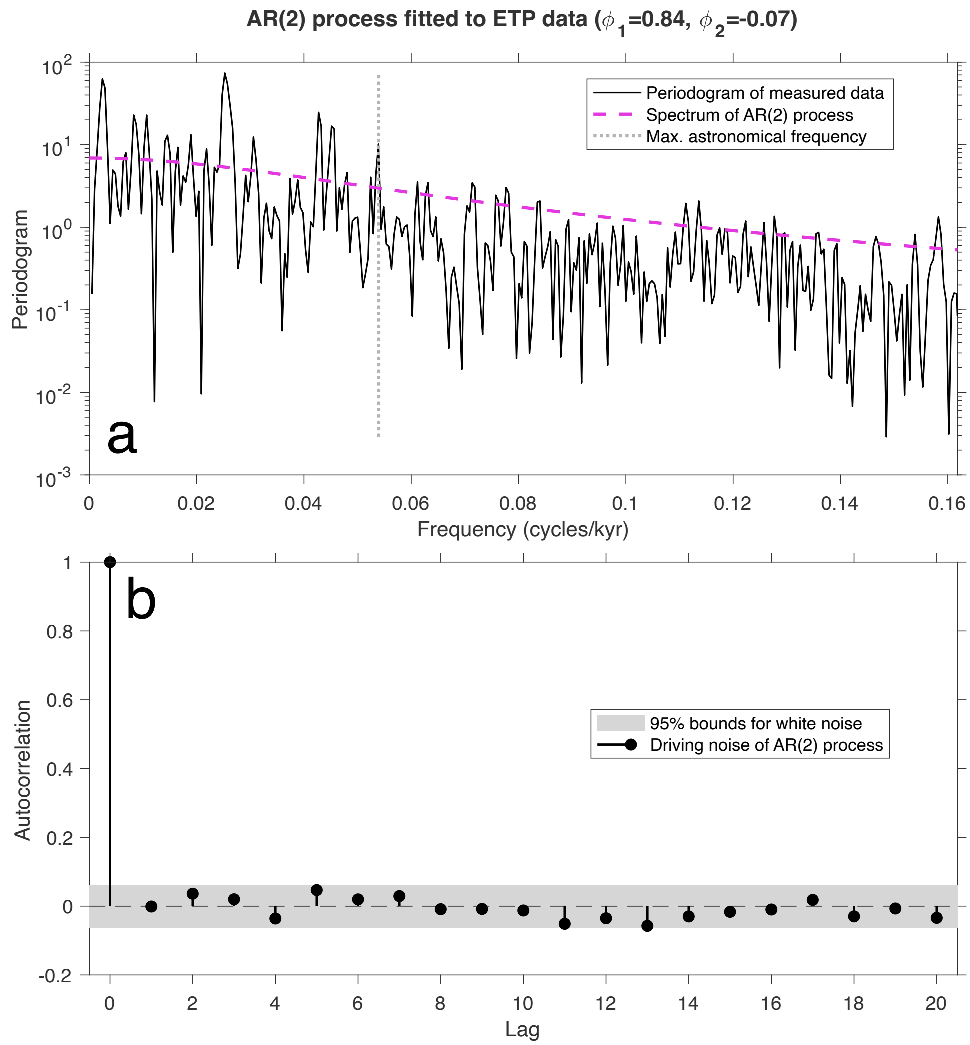

Figure S1. Fit of an AR(2) process to the residuals for the synthetic ETP test data set. (a) Comparison of data periodogram with the spectrum of an AR(2) process with the coefficients ϕ_1 and ϕ_2 fitted for the MAP value of the parameters. The vertical dotted line marks the maximum climatic precession frequency in the data. (b) Sample autocorrelation of the driving noise $\mathbf{w}$ of the residuals obtained for the MAP values of the parameters.

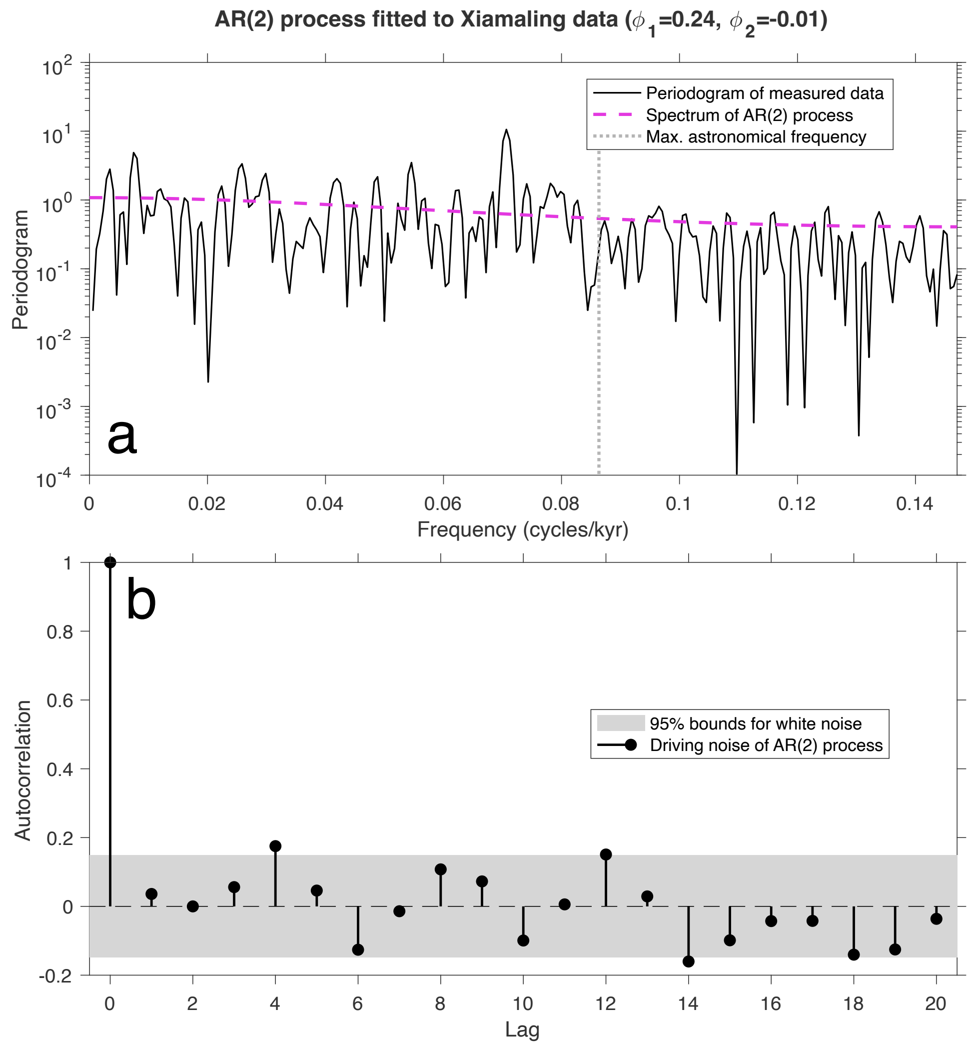

Figure S2. Fit of an AR(2) process to the residuals for the Xiamaling formation Cu/Al data set. (a) Comparison of data periodogram with the spectrum of an AR(2) process with the coefficients ϕ_1 and ϕ_2 fitted for the MAP value of the parameters. The vertical dotted line marks the maximum climatic precession frequency in the data. (b) Sample autocorrelation of the driving noise **w** of the residuals obtained for the MAP values of the parameters.

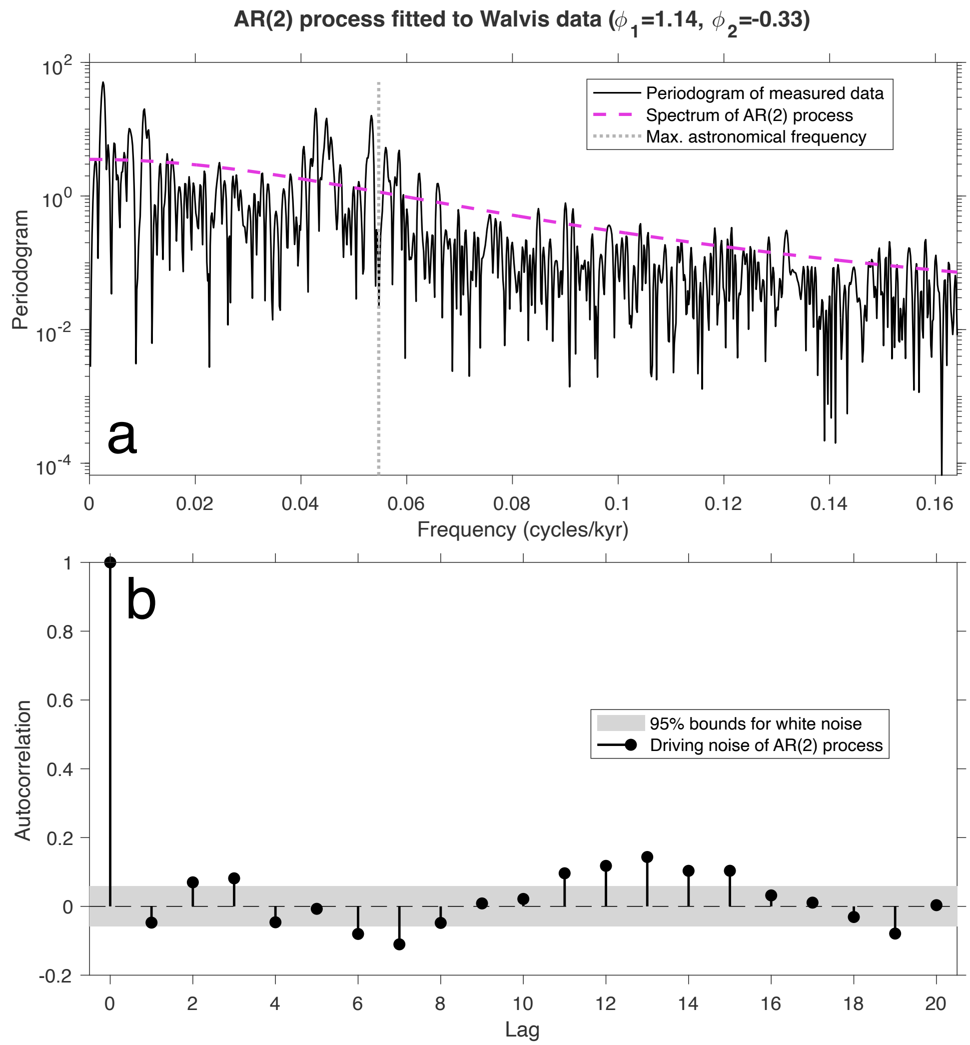

Figure S3. Fit of an AR(2) process to the residuals for the Walvis Ridge a* data set. (a) Comparison of data periodogram with the spectrum of an AR(2) process with the coefficients ϕ_1 and ϕ_2 fitted for the MAP value of the parameters. The vertical dotted line marks the maximum climatic precession frequency in the data. (b) Sample autocorrelation of the driving noise $\mathbf{w}$ of the residuals obtained for the MAP values of the parameters.

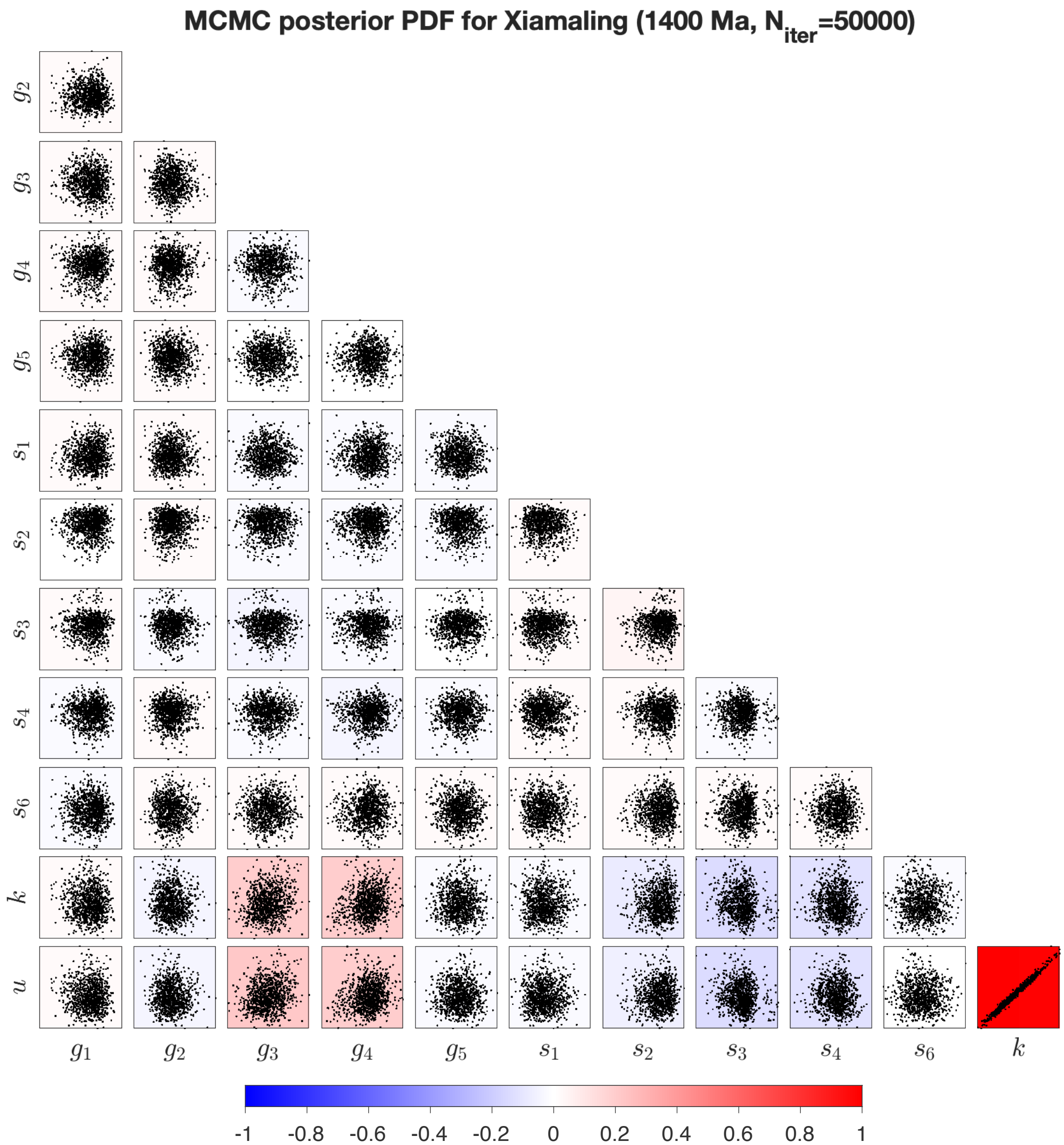

Figure S4. Posterior correlation of parameters sampled by TimeOptBMCMC for the Xiamaling formation Cu/Al data set. The color background for each pair of parameters is proportional to the correlation coefficient (as shown by the color bar at the bottom of the figure). Posterior correlations are near zero, with the exception of a strong positive correlation between *u* and *k*.

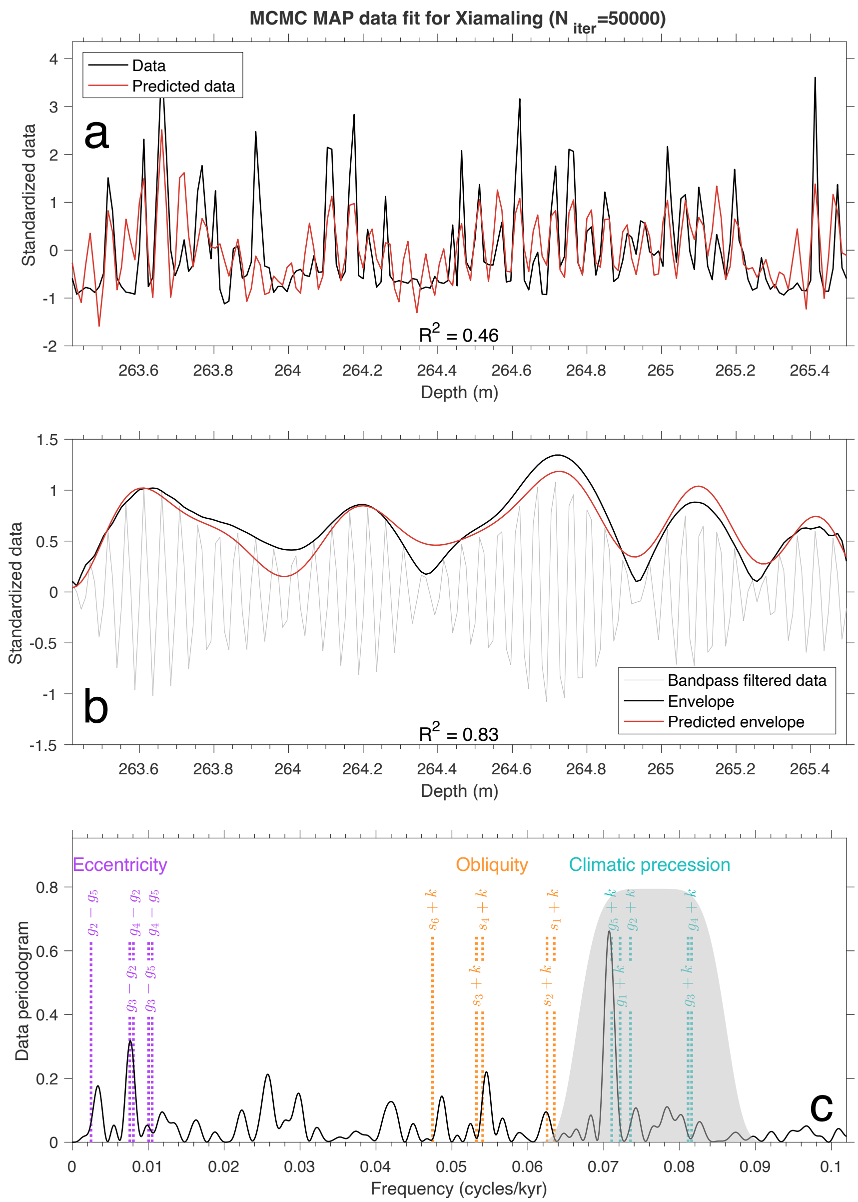

Figure S5. Fit to the Xiamaling formation Cu/Al data obtained by TimeOptBMCMC for the MAP value of sedimentation rate u and astronomical frequencies g_i, s_i, and k. (a) Fit between measured and predicted stratigraphic data (spectral fit). (b) Fit between the envelope of the bandpassed climatic precession signal and the envelope predicted by the eccentricity frequencies (envelope fit). (c) Data periodogram (black continuous line) and frequencies of the reconstructed astronomical cycles in the data (dotted vertical lines). The gray shaded area shows the frequency response of the filter used to compute the bandpassed climatic precession signal in the data (gray curve in (b)).

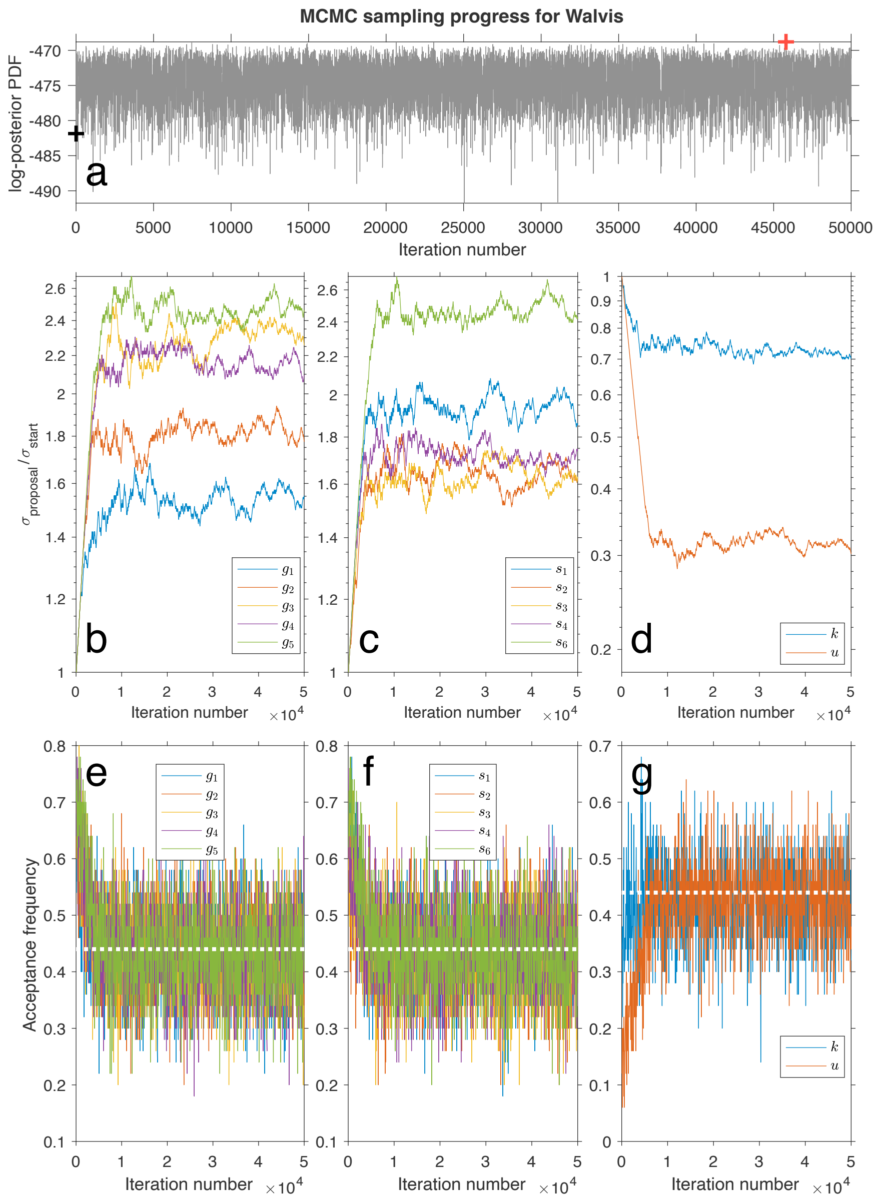

Figure S6. Progress of TimeOptBMCMC sampling for the Walvis Ridge a* data set over 50,000 iterations. (a) Value of the log-posterior PDF for the sampled model parameter vectors. The black cross is the starting value and the red cross the MAP. (b, c, d) Standard deviation of the proposal PDF (as a ratio over the starting value) for each model parameter. (e, f, g) Frequency of acceptance of the proposed steps in the MCMC random walk. The adaptive Metropolis algorithm used in TimeOptBMCMC adjusts the standard deviations of the proposal PDF to keep the frequency of acceptance around the optimal value of 0.44 for all model parameters (white horizontal dotted line).

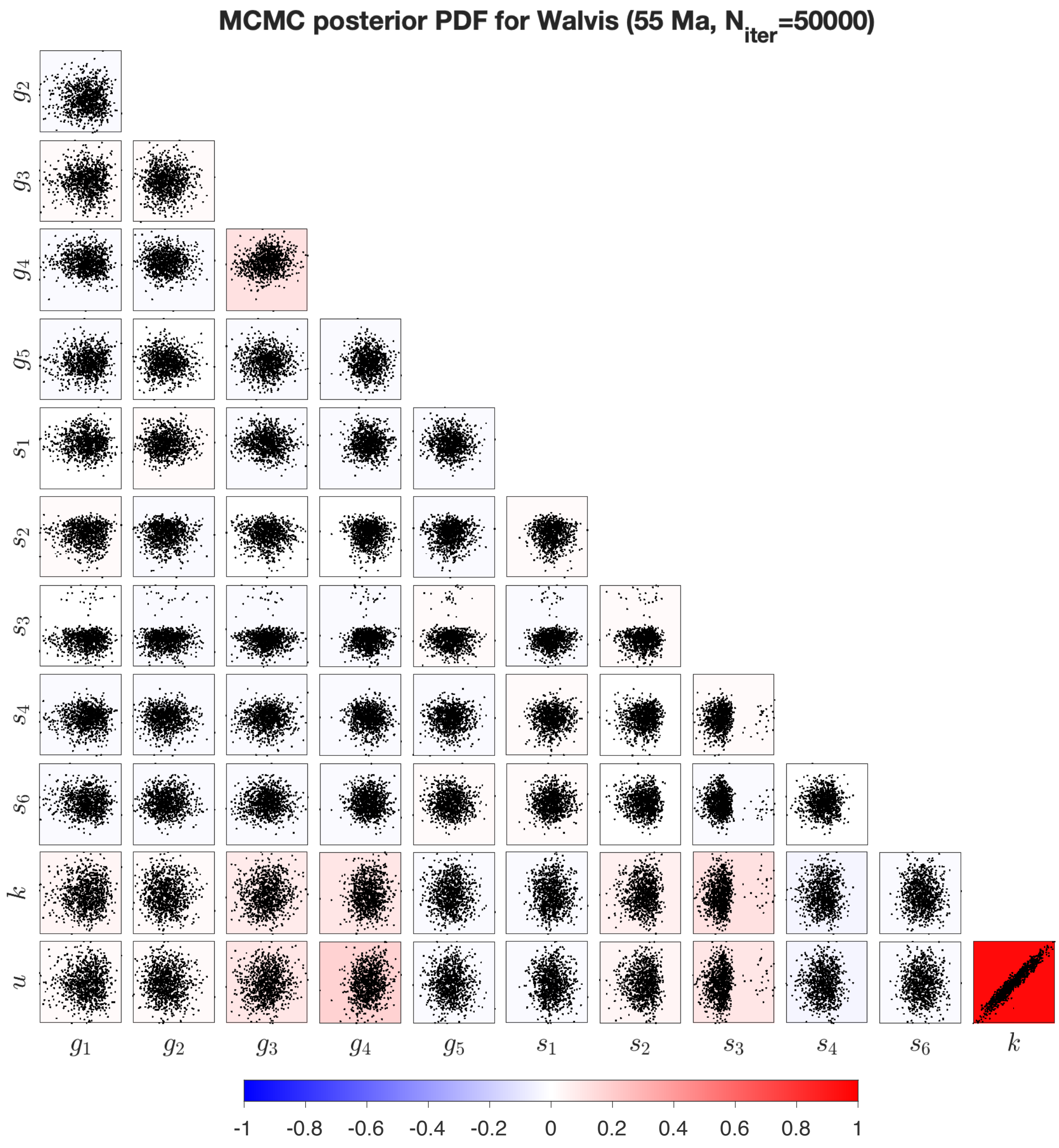

Figure S7. Posterior correlation of parameters sampled by TimeOptBMCMC for the Walvis Ridge a* data set. The color background for each pair of parameters is proportional to the correlation coefficient (as shown by the color bar at the bottom of the figure). Posterior correlations are near zero, with the exception of a strong positive correlation between *u* and *k*.

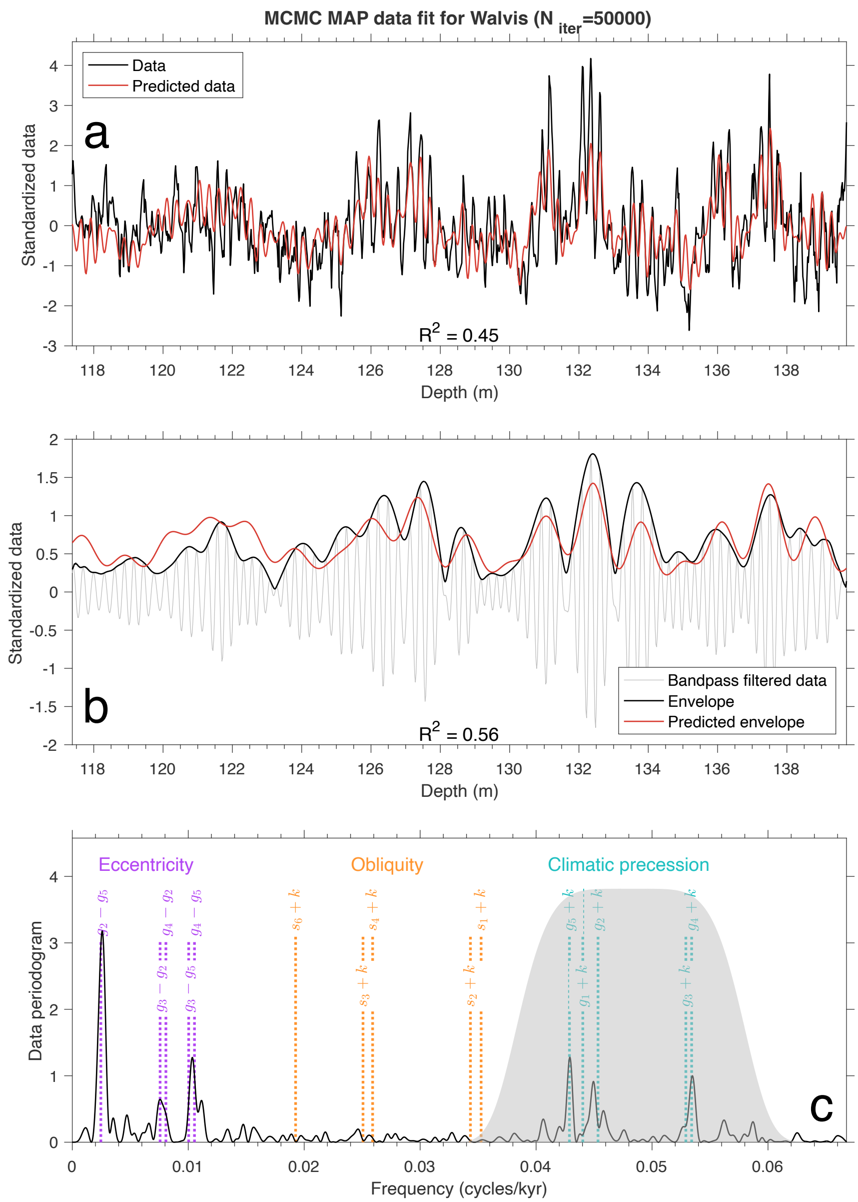

Figure S8. Fit to the Walvis Ridge a* data obtained by TimeOptBMCMC for the MAP value of sedimentation rate u and astronomical frequencies g_i, s_i, and k. (a) Fit between measured and predicted stratigraphic data (spectral fit). (b) Fit between the envelope of the bandpassed climatic precession signal and the envelope predicted by the eccentricity frequencies (envelope fit). (c) Data periodogram (black continuous line) and frequencies of the reconstructed astronomical cycles in the data (dotted vertical lines). The gray shaded area shows the frequency response of the filter used to compute the bandpassed climatic precession signal in the data (gray curve in (b)).