A comprehensive model for the kyr and Myr time scales of Earth's axial magnetic dipole field

Matthias Morzfeld[1] and Bruce A. Buffett[2]

[1]Department of Mathematics, University of Arizona, 617 N. Santa Rita Ave., P.O. Box 210089, Tucson, Arizona 85721, USA
[2]University of California, Department of Earth & Planetary Science, 307 McCone Hall, Berkeley, CA 94720, USA

Correspondence: Matthias Morzfeld, mmo@math.arizona.edu

Abstract. We consider a stochastic differential equation model for Earth's axial magnetic dipole field. The model's parameters are estimated using diverse and independent data sources that had previously been treated separately. The result is a numerical model that is informed by the full paleomagnetic record on kyr to Myr time scales and whose outputs match data of Earth's dipole in a precisely defined "feature-based" sense. Specifically, we compute model parameters and associated uncertainties
5   that lead to model outputs that match spectral data of Earth's axial magnetic dipole field but our approach also reveals difficulties with simultaneously matching spectral data and reversal rates. This could be due to model deficiencies or inaccuracies in the limited amount of data. More generally, the approach we describe can be seen as an example of an effective strategy for combining diverse data sets that is particularly useful when the amount of data is limited.

1   Introduction

10   Earth possesses a time-varying magnetic field which is generated by the turbulent flow of liquid metal alloy in the core. The field can be approximated as a dipole with north and south magnetic poles slightly misaligned with the geographic poles. The dipole field changes over a wide range of timescales, from years to millions of years and these changes are documented by several different sources of data, see, e.g., Hulot et al. (2010). Satellite observations reveal changes of the dipole field over years to decades Finlay et al. (2016), while changes on time scales of thousands of years are described by paleomagnetic
15   data, including observations of the dipole field derived from archeological artifacts, young volcanics, and lacustrine sediments Constable et al. (2016). Variations on even longer time scales of millions of years are recorded by marine sediments Valet et al. (2005); Ziegler et al. (2011) and by magnetic anomalies in the oceanic crust Ogg (2012); Cande and Kent (1995); Lowrie and Kent (2004). On such long time scales, we can observe the intriguing feature of Earth's axial magnetic dipole field to reverse its polarity (magnetic north pole becomes the magnetic south pole and vice versa).

20   Understanding Earth's dipole field, at any time scale, is difficult because the underlying magnetohydrodynamic problem is highly nonlinear. For example, many numerical simulations are far from Earth-like due to severe computational constraints. Besides difficult numerics and computational barriers, even basic analytical calculations are often intractable. An alternative approach is to use "low-dimensional models" which aim at providing a simplified but meaningful representation of some aspects of Earth's geo-dynamo. Several such models have been proposed over the past years. The model of Gissinger (2012),

for example, describes the Earth's dipole over millions of years by a set of three ordinary differential equations, one for the dipole, one for the non-dipole field and one for velocity variations at the core. A stochastic model for Earth's dipole over millions of years was proposed by Pétrélis et al. (2009). Several other models have been derived by Hoyng et al. (2001); Rikitake (1958); Pétrélis and Fauve (2008).

5   We focus on a simpler stochastic model that can potentially describe Earth's dipole field over several time scales from thousands of years to millions of years (see Section 3 and below for references and model details). The basic idea is to model Earth's dipole field as being analogous to a particle in a double well potential. Time variations of the dipole field and dipole reversals are as follows. The state of the SDE is within one of the two wells of the double well potential and is pushed round by noise (Brownian motion). The pushes and pulls by the noise process lead to variations of the dipole field around a typical

10   value. Occasionally, however, the noise builds up to push the state over the potential well which causes a change of its sign. A transition from one well to the other represents a reversal of Earth's dipole. The state of the SDE then remains, for a while, within the opposite well, and the noise leads to time variations of the dipole field around the negative of the typical value. Then, the reverse of this process may occur.

A basic version of this model, which we call the "B13 model" for short, was discussed by Buffett et al. (2013). The coef-

15   ficients that define the B13 model are derived from the PADM2M data Ziegler et al. (2011) which describe variations in the strengths of Earth's axial magnetic dipole field over the past 2 Myr. Thus, the basic B13 model is valid for Myr time scales and in particular for the the past 2 Myr. The PADM2M data are derived from marine sediments which means that the data are smoothed by sedimentation processes, see, e.g., Roberts and Winkhofer (2004). The B13 model, however, does not directly account for the effects of sedimentation. Buffett and Puranam (2017) added sedimentation processes to the model by sending

20   the solution of the SDE through a suitable low-pass filter. With this extension, the B13 model is more suitable to be compared to the data about Earth's dipole field on a Myr time-scale.

The B13 model relies on the assumption that the noise process within the SDE is uncorrelated in time. This assumption is reasonable when describing the dipole field on the Myr time scale but is not valid on a shorter time scale of thousands of years. Buffett and Matsui (2015) derived an extension of the B13 model to extend it to time scales of thousands of years, by

25   adding a time-correlated noise process. An extension of B13 to represent changes in reversal *rates* over the past 150 Myrs is considered by Morzfeld et al. (2018). Its use for predicting the probability of an imminent reversal of Earth's dipole is described by Morzfeld et al. (2017); Buffett and Davis (2018). The B13 model is also discussed by Meduri and Wicht (2016); Buffett et al. (2014); Buffett (2015).

The B13 model and its extensions are constructed with several data sets in mind that document Earth's axial dipole field

30   over the kyr and Myr time scales. The data, however, are not considered simultaneously: the B13 model is based on one data source (paleomagnetic data on the Myr time scale) and some of its modifications are based on other data sources (the shorter record over the past 10 kyrs). Our goal is to construct a comprehensive model for Earth's axial dipole field by calibrating the B13 model to several independent data sources *simultaneously*, including

(i) observations of the strength of the dipole over the past 2 Myrs as documented by the PADM2M and Sint-2000 data sets,

35   see Ziegler et al. (2011); Valet et al. (2005);

(ii) observations of the dipole over the past 10 kyr as documented by CALS10k.2, see Constable et al. (2016);

(iii) reversals and reversal rates derived from magnetic anomalies in the oceanic crust see, e.g., Ogg (2012).

The approach ultimately leads to a family of SDE models, valid over Myr and kyr time scales, whose parameters are informed by the complete paleomagnetic record. The results we obtain here are thus markedly different from previous work were data at

5   different time scales are considered separately. We also use our framework to assess the effects of the various data sources on parameter estimates and to discover inconsistencies between model and data.

At the core of our model calibration is the Bayesian paradigm in which uncertainties in data are converted into uncertainties in model parameters. The basic idea is to merge prior information about the model and its parameters, represented by a prior distribution, with new information from data, represented by a likelihood, see, e.g., Reich and Cotter (2015); Asch et al. (2017).

10   Priors are often assumed to be "uninformative", i.e., only conservative bounds for all parameters are known, and likelihoods describe model-data mismatch. The Bayesian approach, however, requires that we make assumptions about the errors in the model and in the data. Such "error models" are describing what "*we do not know*" and, for that reason, good error models are notoriously difficult to come by. We address this difficulty by performing a suite of numerical experiments that allows us to check, in hind-sight, our a priori assumptions about error models.

15   More generally we present a numerical and computational framework that can address a typical challenge in geophysical modeling, where data are limited with each datum being the result of years of careful work, but with uncertainties in the data that are large and not well understood. To address this challenge we adapt the usual Bayesian approach for model calibration so that we can make use of *all* available data sources simultaneously. Priors are not critical in this context. We thus focus on formulations of likelihoods. Recall that likelihoods are typically defined in terms of a point-wise mismatch between model

20   outputs and data. Such likelihoods, however, are not meaningful when a variety of diverse data sources are used. For example, a point-wise mismatch of model and data is not useful when two different data sets report two different values for the same quantity (see, e.g., Figure 1). We thus substitute likelihoods based on point-wise mismatch of model and data by a "feature-based" likelihood, as discussed by Maclean et al. (2017); Morzfeld et al. (2018). A feature-based likelihood is based on error in "features" extracted from model outputs and data rather than the usual point-wise error. The feature-based approach enables

25   unifying contributions from several independent data sources in a well-defined sense even if the various data may not be entirely self-consistent.

2   Description of the data

We describe variations in the virtual axial dipole moment (VADM) over the past 2 Myr using stacks of marine sediment data to reduce the influence of non-dipole components of the geomagnetic field. Two different compilations are considered in this

30   study: Sint-2000, Valet et al. (2005) and PAMD2M Ziegler et al. (2011). Both of these data sets are sampled every 1 kyr. The results are shown in the upper left panel of Figure 1. During parameter estimation we will make use of the time average VADM and the standard deviation of VADM over time, listed in Table 1.
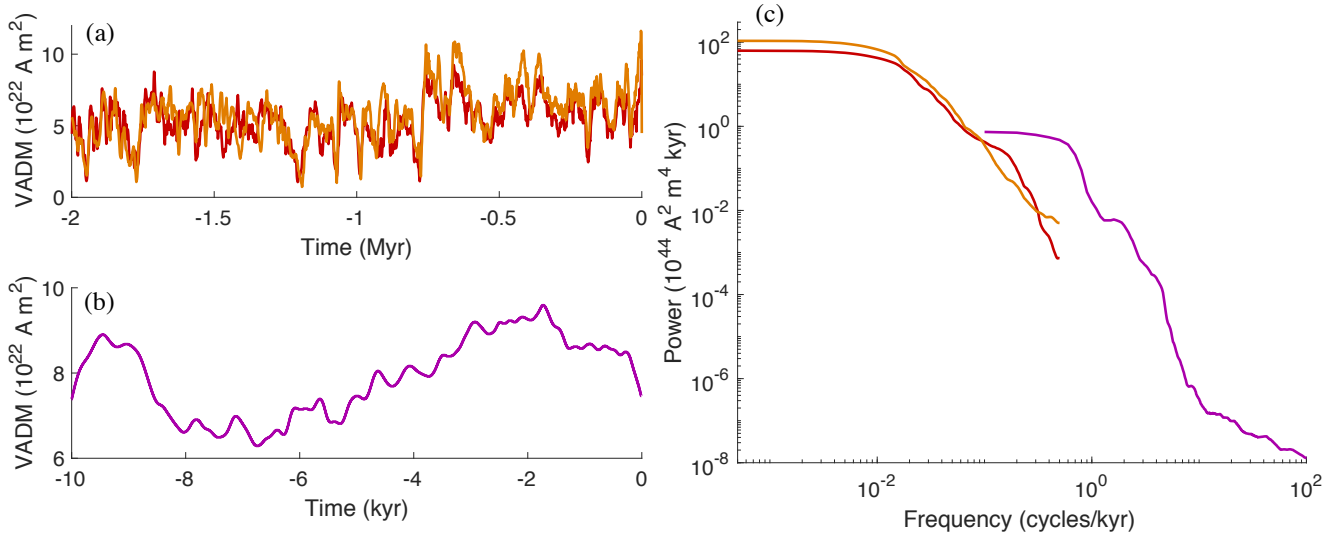
Figure 1. Data used in this paper. (a) Sint-2000 (orange) and PADM2M (red): VADM as a function of time over the past 2 Myr. (b) CALS10k.2: VADM as a function of time over the past 10 kyr. (c) Power spectral densities of the data in (a) and (b), computed by the multi-taper spectral estimation technique of Constable and Johnson (2005). Orange: Sint-2000. Red: PADM2M. Purple: CALS10k.2

Data source	Time avg. VADM $(10^{22}\ \text{Am}^2)$	Std. dev. of VADM $(10^{22}\ \text{Am}^2)$
PADM2M	5.23	1.48
Sint-2000	5.81	1.84

Table 1. Time average VADM and VADM standard deviation of PADM2M and Sint-2000.

We use the CALS10k.2 data set to describe variations in the VADM over the past 10 kyr. The time dependence of CALS10k2 is represented using B-splines, so the model can be sampled at arbitrary time intervals. The CALS10k.2 data are shown in the lower left panel of Figure 1, sampled at an interval of 1 year. Below we use features derived from power spectral densities (PSD) of the data. The PSDs are computed by the multi-taper spectral estimation technique of Constable and Johnson (2005).

5 We show the PSDs of the three data sets in the left panel of Figure 1.

Lastly, we make use of reversal rates of the Earth's dipole computed from the geomagnetic polarity time scale Cande and Kent (1995); Lowrie and Kent (2004); Ogg (2012). Using the chronology of Ogg (2012), we compute reversal rates for 5 Myr intervals from today up to 30 Myr ago. That is, we compute the reversal rates for the intervals $0\ \text{Myr} - 5\ \text{Myr}, 5\ \text{Myr} - 10\ \text{Myr}, \ldots, 25\ \text{Myr} - 30\ \text{Myr}$. This leads to the average reversal rate and standard deviation listed in Table 2. Increasing the

10 interval to 10 Myr leads to the same mean but decreases the standard deviation (see Table 2).

Interval length	Average reversal rate (reversals/Myr)	Std. dev. (reversals/Myr)
5 Myr	4.23	1.01
10 Myr	4.23	0.49

Table 2. Average reversal rate and standard deviation computed over the past 30 Myr using the chronology of Ogg (2012).

Note that the various data we use are not all consistent. For example, visual inspection of VADM (Figure 1), as well as comparison of the time average and standard deviation (Table 1) indicate that the PADM2M and Sint-2000 data sets report different VADM. These differences can be attributed, at least in part, to differences in the calibration of the marine sediment measurements and to differences in the way the measurements are stacked to recover the dipole component of the field. There are also notable differences between CALS10k.2 and the lower resolution data sets (Sint-2000 and PADM2M) over the last 10 kyr (see Figure 1). These discrepancies may reflect inherent uncertainties that arise from "observing" Earth's magnetic axial dipole thousand to millions of years into the past. We address discrepancies between the various data sets we use by using features, rather then the "raw" data (see Section 4).

3   Description of the models on multiple time scales

Our models for variations in the dipole moment on Myr and kyr time scales are based on a scalar stochastic differential equation (SDE)

$$\mathrm{d}x = v(x)\mathrm{d}t + \sqrt{2D(x)}\mathrm{d}W, \tag{1}$$

where t is time and where x represents the VADM and polarity of the dipole, see Buffett et al. (2013). A negative sign of $x(t)$ corresponds to the current polarity, a positive sign means reversed polarity. W is Brownian motion, a stochastic process with well-defined properties: $W(0) = 0$, $W(t) - W(t + \Delta T) \sim \mathcal{N}(0, \Delta t)$, $W(t)$ is almost surely continuous for all $t \geq 0$, see, e.g., Chorin and Hald (2013). Here and below, $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian random variable with mean μ, standard deviation σ and variance σ^2. Throughout this paper, we assume that the "diffusion", $D(x)$, is constant, i.e., $D(x) = D$, as is suggested by the previous literature on this model.

The function v is called the "drift" and is derived from a double-well potential:

$$v(x) = \gamma \frac{x}{\bar{x}} \cdot \begin{cases} (\bar{x} - x), & \text{if } x \geq 0 \\ (x + \bar{x}), & \text{if } x < 0 \end{cases}, \tag{2}$$

where $\bar{x}$ and γ are parameters. Nominal values of the parameters $\bar{x}$, γ and D are listed in Table 3. The drift coefficient, $v(x)$, for the nominal parameter values and the corresponding double well potential potential, $v(x) = -U'(x)$, are shown in Figure 2. With the nominal values, the model exhibits "dipole reversals", which are represented by a change of the sign of x. This is the "basic" B13 model.

	$\bar{x}$	D	γ	T_s	a
	$(10^{22}$ Am$^2)$	$(10^{44}$ A^{2}m^4 kyr$^{-1})$	(kyr^{-1})	(kyr)	(kyr^{-1})
Nominal value:	5.23	0.3403	0.075	2.5	5
Lower bound:	0	0.0615	0.0205	1	5
Upper bound:	10	2.1	0.7	5	40

Table 3. Nominal parameter values and parameter bounds
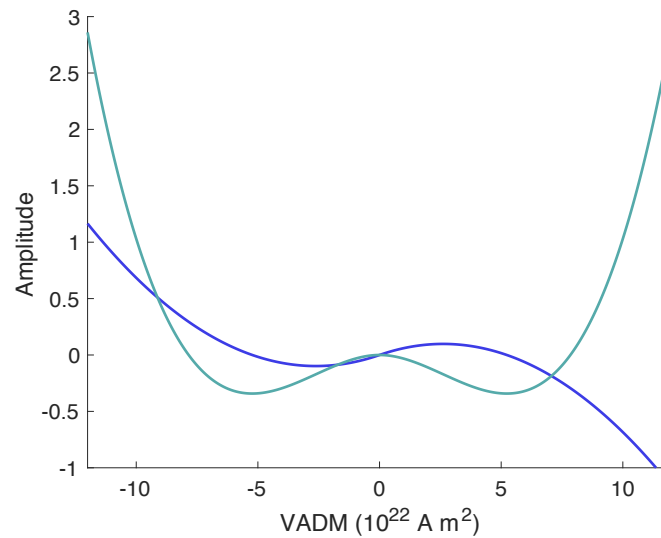

Figure 2. Drift coefficient $v(x)$ (blue) and potential $U'(x) = -v(x)$ (turquoise).

For computations, we discretize the SDE using a 4th order Runge-Kutta (RK4) method for the drift and an Euler-Maruyama method for the diffusion. This results in the discrete time B13 model

$$x_k = f(x_{k-1}, \Delta t) + \sqrt{2D\Delta t}\, w_k, \quad w_k \sim \mathcal{N}(0,1), \text{ iid} \tag{3}$$

where Δt is the time step, $\sqrt{\Delta t}\, w_k$ is the discretization of Brownian motion W in (1), iid means "independent and identically
distributed", and where $f(x_{k-1}, \Delta t)$ is the RK4 step. We distinguish between variations in the Earth's dipole over kyr to Myr
time scales and, for that reason, present modifications of the basic B13 model (3).

3.1 Models for the Myr time scale and their power spectral densities

For simulations over Myr time scales we chose a time-step $\Delta t = 1$ kyr, corresponding to the sampling time of the Sint-2000
and PADM2M data. On a Myr time scale, the primary type of paleomagnetic data in Sint-2000 and PADM2M are affected by
affected by gradual acquisition of magnetization due to sedimentation process which amounts to an averaging over a (short)
time interval, see, e.g., Roberts and Winkhofer (2004). We follow Buffett and Puranam (2017) and include the smoothing
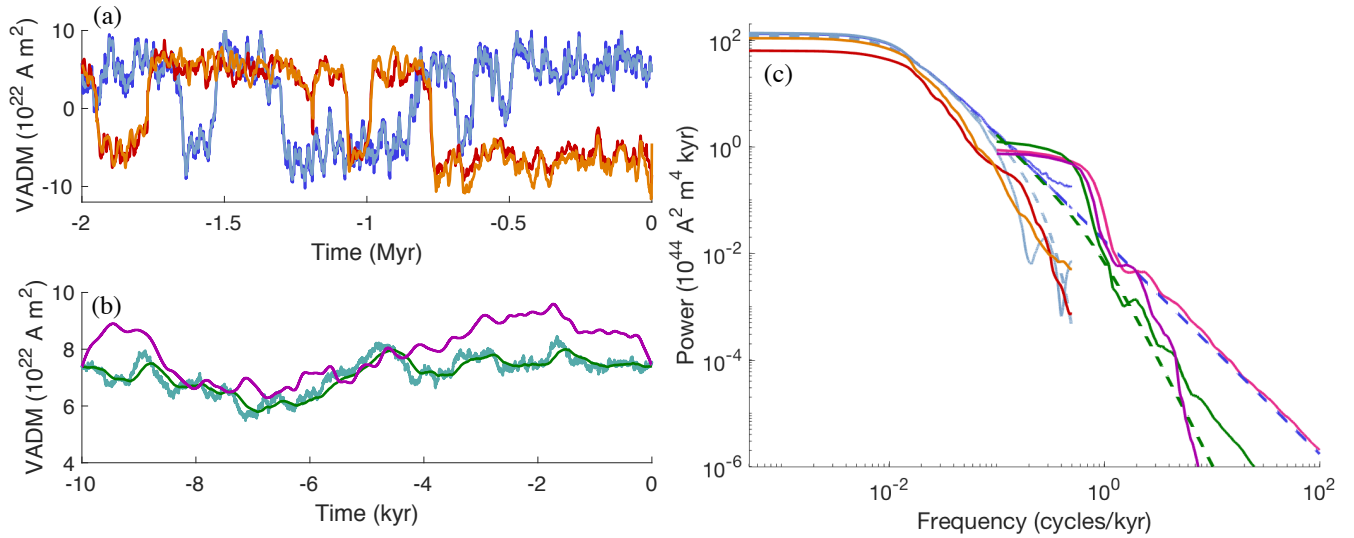
Figure 3. Simulations and data. (a) Output, x_j^{Myr} (dark blue, often hidden) and smoothed model output, $x_j^{\mathrm{Myr,s}}$ (light blue). Signed Sint-2000 (orange) and PADM2M (red), signs taken from Cande and Kent (1995). (b) Output of the kyr-model with uncorrelated noise (turquoise) and correlated noise (green), along with CALS10k.2 data (purple). (c) Power spectral densities. *Myr-model*: PSD of Myr model based on 50 Myr simulation (solid dark blue) and theoretical PSD of linear model (dashed dark blue); PSD of Myr model based on 50 Myr simulation and high-frequency roll-off (solid light blue) and corresponding theoretical PSD of linear model (dashed light blue) *kyr model*: PSD of kyr model based on 10 kyr simulation with uncorrelated noise (solid pink) and corresponding theoretical spectrum (dashed blue). PSD of kyr model based on 10 kyr simulation with correlated noise (solid green) and corresponding theoretical spectrum (dashed green). *Data*: Sint-2000 (orange), PADM2M (red) and CALS10k.2 data (purple).

effects of sedimentation in the model by convolving the solution of (3) by a Gaussian filter

$$g(t) = \sqrt{\frac{6}{\pi T_s^2}} \cdot \exp\left(-\frac{6t^2}{T_s^2}\right), \tag{4}$$

where T_s defines the duration of smoothing, i.e., the width of a time window over which we average. The nominal value for T_s is given in Table 3. The result is a smoothed Myr model $x^{\mathrm{Myr,s}}$. Simulations with the "Myr model" using the nominal parameters

5   of Table 3 are shown in Figure 3(a) where we plot the model output x_j^{Myr} in dark blue and the smoothed model output, $x_j^{\mathrm{Myr,s}}$, in a lighter blue over a period of 2 Myr. The PSDs of simulations corresponding to 50 Myr simulations are shown in Figure 3(c). Note that the PSD of the smoothed model output, $x^{\mathrm{Myr,s}}$, taking into account sedimentation processes, rolls-off quicker than the PSD of x_j^{Myr}. For that reason, the PSD of the smoothed model seems to fit the PSDs of the Sint-2000 and PADM2M data "better", i.e., we observe a similarly quick roll-off at high frequencies in model and data; see also Buffett and Puranam (2017)

10   and Figure 3.

Because an SDE is noisy (Brownian motion), long simulations are required to obtain accuracy in the PSD of a numerical SDE solution. To avoid large errors due to short simulations we approximate the PSD of (1) by the PSD of a *linear* model

$$\mathrm{d}x^l = -\gamma(x^l - \bar{x})\mathrm{d}t + \sqrt{D}\mathrm{d}W, \tag{5}$$

whose PSD is easy to calculate analytically as

$$5 \quad \hat{x}^l(f) = \frac{2D}{\gamma^2 + 4\pi^2 f^2}, \tag{6}$$

where f is the frequency (in 1/kyr). Since the Fourier transform of the Gaussian filter is known analytically, the PSD of the smoothed model can be approximated by

$$\hat{x}^{l,s}(f) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \exp\left(-\frac{4\pi^2 f^2 T_s^2}{12}\right). \tag{7}$$

The approximate PSDs $\hat{x}^l(f)$ and $\hat{x}^{l,s}(f)$, computed by (6) and (7), are plotted in Figure 3(a) in comparison to the PSDs
10   computed from a 50 Myr simulation with the discrete time model (3).

3.2 Reversal rate, VADM time average and VADM standard deviation

The nonlinear SDE model (1) and its discretization (3) exhibit reversals, i.e., change in the sign of x. Moreover, the overall "power", i.e., the area under the PSD curve, is given by the standard deviation of the absolute value of $x(t)$ over time. Another important quantity of interest is the time averaged value of the absolute value of $x(t)$, which describes the overall average
15   strength of the dipole field.

Errors in the computation of these quantities, the reversal rate r, the time average VADM and VADM standard deviation, from time domain simulations with the discrete time model (3) are dominated by errors that arise from limitations of the simulation time. Only long simulations (hundreds of millions of years) lead to accurate estimates of the reversal rate, time average VADM and VADM standard deviation. Long simulations, however, require more substantial computations. This would not be an issue
20   if we were to compute reversal rate and other quantities once, but below we use sampling techniques which require repeated computing of reversal rates etc. for different parameters. Moreover, we will perform sampling in a variety of settings (see Section 6). We thus streamline computations by following Buffett and Puranam (2017) and approximating the time average VADM, the VADM standard deviation σ and the reversal rate by

$$E(x) \approx \bar{x} \cdot 10^{22} \text{ Am}^2, \quad \sigma \approx \sqrt{D/\gamma} \cdot 10^{22} \text{ Am}^2 \text{ kyr}^2, \quad r \approx \frac{\gamma}{2\pi} \cdot \exp\left(-\frac{\bar{x}^2}{6\sigma^2}\right) \cdot 10^3 \text{ Myr}^{-1}. \tag{8}$$

25   These approximations do not require solving the SDE over long time periods and the computations, for a single evaluation of reversal rate etc., are instantaneous. For the nominal values we thus calculate a time average of $\bar{x} = 5.23 \cdot 10^{22}$ Am^2, a standard deviation of $\sigma \approx 2.13 \cdot 10^{22}$ Am^2 and a reversal rate $r \approx 4.37$ Myr^{-1}. These should be compared to the corresponding values of PADM2M and Sint-2000 in Table 1 and to the reversal rate from the geomagnetic polarity time scale in Table 2.

3.3 Models for the kyr time scale and their power spectral densities

In the basic SDE model in (1) we assume that the noise is uncorrelated in time (Brownian motion). This assumption is reasonable when one focuses on low frequencies and large sample intervals of the dipole, as in Sint-2000 and PADM2M, whose sampling interval is 1/kyr. When the sampling interval is shorter, as in CALS10k.2 which is sampled once per year, this assumption is not valid and a correlated noise is more appropriate Buffett and Matsui (2015). Computationally, this means that we swap the uncorrelated, iid, noise in (3) for a noise that has a short but finite correlation time. This can be done by "filtering" Brownian motion. The resulting discrete time model for the kyr time scale is

$$y_k = (1 - a\Delta t)y_{k-1} + \sqrt{2a\Delta t}\,w_k, \quad w_k \sim \mathcal{N}(0,1),\ \text{iid} \tag{9}$$

$$x_k = f(x_{k-1}, \Delta t) + \sqrt{Da}\,\Delta t\, y_k, \tag{10}$$

where a is the model parameter that defines the correlation time $T_c = 1/a$ of the noise and $\Delta t = 1$ yr. A 10 kyr simulation of the kyr models with uncorrelated and correlated noise using the nominal parameters of Table 3, are shown in Figure 3(b) along with the CALS10k.2 data.

As with the Myr model, we approximate the PSD of the kyr model (9) and (10) by the PSD of a corresponding linear model, driven by correlated noise. In the limit of continuous time ($\Delta t \to 0$), the PSD of the kyr-model with correlated noise is

$$\hat{x}^{l,kyr}(f) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \frac{a^2}{a^2 + 4\pi^2 f^2}, \tag{11}$$

where the first term is as in (6) and the second term appears because of the correlated noise. The PSD, $\hat{x}^{l,kyr}(f)$, of the linear kyr model with correlated noise is plotted in Figure 3(c) in comparison to the PSD computed from a 10 kyr run and the PSD of the CALS10k.2 data.

3.4 Parameter bounds

We have "prior" information about the model parameters. For example, we could come up with nominal values that lead to model outputs (spectra, reversal rate etc.) that are in rough agreement with the data (see Figure 3). These nominal values are based on inferences from Buffett and Puranam (2017). In addition, we can find lower and upper bounds for the model parameters which we will use to construct uniform prior distributions in Section 4. The parameter bounds we use are quite wide, i.e., the upper bounds are probably too large and the lower bound are probably too small, but this is not critical for our purposes as we explain in more detail in Section 4.

The parameter γ is defined by the inverse of the dipole decay time, see Buffett et al. (2013). An upper bound on the dipole decay time τ_{dec} is given by the slowest decay mode $\tau_{\text{dec}} \leq R^2/(\pi^2 \eta)$, where R is the radius of the Earth and $\eta = 0.8\,\text{m}^2/\text{s}$ is the magnetic diffusivity. Thus, $\tau_{\text{dec}} \leq 48.6$ kyr, which means that $\gamma \geq 0.0205\,\text{kyr}^{-1}$. This is a fairly strict lower bound because the value of diffusivity corresponds to high electrical conductivities, recently favored by theoretical calculations, see, e.g., Pozzo et al. (2012). In fact, the bound may be overly pessimistic because neglected effects, e.g., electron-electron scattering, would likely cause η to increase and, therefore, the lower bound for γ to increase, see Pourovskii et al. (2017). To obtain an upper

bound for γ, we note that if γ is large, the magnetic decay is short, which means that it becomes increasingly difficult for convection in the core to maintain the magnetic field. The ratio of dipole decay time τ_{dec} to advection time $\tau_{\text{adv}} = L/V$, where $L = 2259$ km is the width of the fluid shell and $V = 0.5$ mm/s, needs to be 10:1 or (much) larger. This leads to the upper bound $\gamma \leq 0.7\,\text{kyr}^{-1}$.

5 Bounds for the parameter D can be found by considering the linear Myr time scale model in equation (5), which suggests that the variance of the dipole moment is $\text{var}(x) = D/\gamma$, see Buffett et al. (2013). Thus, we may require that $D \sim \text{var}(x)\gamma$. The average of the variance of Sint-2000 ($\text{var}(x) = 3.37 \cdot 10^{44}\,\text{A}^2\text{m}^4$) and PADM2M ($\text{var}(x) = 2.19 \cdot 10^{44}\,\text{A}^2\text{m}^4$) is $\text{var}(x) \approx 2.78 \cdot 10^{44}\,\text{A}^2\text{m}^4$. We use the rounded up value $\text{var}(x) \approx 3 \cdot 10^{44}\,\text{A}^2\text{m}^4$ and, together with the lower and upper bounds on γ, this leads to the lower and upper bounds $0.062 \cdot 10^{44}\text{A}^2\text{m}^4\text{kyr}^{-1} \leq D \leq 2.1 \cdot 10^{44}\text{A}^2\text{m}^4\text{kyr}^{-1}$.

10 The smoothing time, T_s, due to sedimentation and the correlation parameter for the noise, a, define the roll-off frequency of the power spectra for the Myr and kyr models, respectively. We assume that T_s is within the interval $[1,5]$ kyr, and that the correlation time a^{-1} is within $[0.025, 0.2]$ kyr (i.e. a within [5, 40] kyr^{-1}). These choices enforce that T_s controls roll-off at lower frequencies (Myr model) and a controls the roll-off at higher frequencies (kyr model). Bounds for the parameter $\bar{x}$ are not easy to come by and we assume wide bounds, $\bar{x} \in [0,10] \cdot 10^{22}\,\text{Am}^2$. Here, $\bar{x} = 0$ is the lowest lower bound we can think of since the average value of the field is always normalized to be positive. The value of the upper bound of $\bar{x} \leq 10$ is chosen to be excessively large – the average field strength over the last 2 Myr is $\bar{x} \approx 5$. Lower and upper bounds for all five model parameters are summarized in Table 1.

4 Formulation of the Bayesian parameter estimation problem and numerical solution

The family of models, describing kyr and Myr time-scales and accounting for sedimentation processes and correlations in the noise process, has five unknown parameters, $\bar{x}, D, \gamma, T_s, a$. We summarize the unknown parameters in a "parameter vector" $\theta = (\bar{x}, D, \gamma, T_s, a)^T$. Our goal is to estimate the parameter vector θ using a Bayesian approach, i.e., to sharpen prior knowledge about the parameters by using the geomagnetic data described in Section 2. This is done by expressing prior information about the parameters in a prior probability distribution $p_0(\theta)$, and by defining a likelihood $p_l(y|\theta)$, y being the data of Section 2. The prior distribution describes information we have about the parameters independently of the data. The likelihood describes the probability of the data given the parameters θ and therefore connects model output and data. The prior and likelihood define the posterior distribution

$$p(\theta|y) \propto p_0(\theta)p_l(y|\theta). \tag{12}$$

The posterior distributions combines the prior information with information we extract from the data. In particular, we can estimate parameters based on the posterior distribution. For example, we can compute the posterior mean and posterior standard deviation for the various parameters and we can also compute correlations between the parameters. The posterior distribution contains all information we have about the model parameters, given prior knowledge and information extracted from the data. Thus, the SDE model with random parameters whose distribution is the posterior distribution represents a comprehensive and

complete model about the Earth's dipole in view of the data we have. On the other hand, the posterior distribution depends on several assumptions since we define the prior and likelihood and, therefore, implicitly define the posterior distribution. In the following we describe how we formulate the prior and the likelihood.

4.1 Prior distribution

5   The prior distribution describes knowledge about the model parameters we have *before* we consider the data. In Section 3.4, we discussed lower and upper bounds for the model parameter and we use these bounds to construct the prior distribution. This can be achieved by assuming a uniform prior over a five-dimensional hyper-cube whose corners are defined by the parameter bounds in Table 1. Note that the bounds we derived in Section 3.4 are fairly wide. Wide bounds are preferable for our purposes, because wide bounds implement minimal prior knowledge about the parameters. With such "uninformative priors",

10   the posterior distribution, which contains information from the data, reveals how well the parameter values are constrained by data. More specifically, if the uniform prior distribution is morphed into a posterior distribution that describes a well-defined "bump" of posterior probability mass in parameter space, then the model parameters are constrained by the data (to be within the bump of posterior probability "mass"). If, one the other hand, the posterior distribution is nearly equal to the prior distribution, then the data have nearly no effect on the parameter estimates and, therefore, the data do not constrain the

15   parameters.

4.2 Feature-based likelihoods

We wish to use the recent geomagnetic record to calibrate and constrain all five model parameters. For this purpose, we use the data sets Sint-2000, PADM2M and CALS10k.2, as well as information about the reversal rate based on the geomagnetic polarity time scale (see Section 2). It is not feasible, and not desirable, to try to find model parameters that lead to a "point-by-

20   point match" of the model and all these data. Specifically, matching data point-wise is hindered by the fact that we have data sets that lead to different VADM values at the same time instant (see Figure 1). Rather, we extract "features" from the data and find model parameters such that the model produces comparable features. The features are based on PSDs of the Sint-2000, PADM2M and CALS10k.2 data sets as well as the reversal rate, time average VADM and VADM standard deviation (see Morzfeld et al. (2018) for a more in depth explanation of feature-based approaches to data assimilation).

25   The overall likelihood consists of three factors:

  (i) one factor corresponds to the contributions from the reversal rate, time average VADM and VADM standard deviation data, which we summarize as "time domain data" from now on for brevity;

 (ii) one factor describes the contributions from data at low frequencies of $10^{-4} - 0.5$ cycles per kyr (PADM2M and Sint-2000);

30 (iii) one factor describes contribution of data at high frequencies of $0.9 - 9.9$ cycles/kyr (CALS10k.2)

In the Bayesian approach, this means that the likelihood, $p_l(y|\theta)$ in equation (12) can be written as the product of three terms

$$p_l(y|\theta) \propto p_{l,\text{td}}(y|\theta)\, p_{l,\text{lf}}(y|\theta) p_{l,\text{hf}}(y|\theta), \tag{13}$$

where $p_{l,\text{td}}(y|\theta)$, $p_{l,\text{lf}}(y|\theta)$ and $p_{l,\text{hf}}(y|\theta)$ represent the contributions from the time domain data (reversal rate, time average VADM and VADM standard deviation), the low frequencies and the high frequencies; recall that y is shorthand notation for all the data we use. We now describe how each component of the overall likelihood is constructed.

4.2.1 Reversal rates, time average VADM and VADM standard deviation

5   We define the likelihood component that addresses the reversal rate based on the equation

$$y_{\text{rr}} = h_{\text{rr}}(\theta) + \varepsilon_{\text{rr}}, \tag{14}$$

where y_{rr} is the reversal rate data, $h_{\text{rr}}(\theta)$ is a function that maps the model parameters to the reversal rate data, and where ε_{rr} is an error model. As described in Section 2, we use the chronology from Ogg (2012) and set $y_{\text{rr}} = 4.23$ reversals/Myr. The function $h_{\text{rr}}(\theta)$ is based on the nonlinear model (2) but avoids the need for long simulations by using Kramers formula

10   Risken (1996) to define the reversal rate. Following Buffett and Puranam (2017) (see also equation (8)), we thus approximate the reversal rate of the model by

$$h_{\text{rr}}(\theta) = \frac{\gamma}{2\pi} \exp\left(-\frac{\gamma \bar{x}^2}{6D}\right) \cdot 10^3 \text{ reversals/Myr.} \tag{15}$$

We use a Gaussian error model with mean zero, i.e., $\varepsilon_{\text{rr}} \sim \mathcal{N}(0, \sigma_{\text{rr}}^2)$. We discuss the error variance σ_{rr}^2 below.

   We define the likelihood for the time averaged VADM by

15    $$y_{\bar{x}} = h_{\bar{x}}(\theta) + \varepsilon_{\bar{x}}. \tag{16}$$

We use the average of the time averages of PADM2M and Sint-2000 as the datum and set $y_{\bar{x}} = 5.56 \cdot 10^{22} \text{ Am}^2$. The function $h_{\bar{x}}(\theta) = \bar{x}$, i.e., we approximate the time average of the absolute value of the solution of the nonlinear model by the time average of the linear model. The error model is again Gaussian, $\varepsilon_{\bar{x}} \sim \mathcal{N}(0, \sigma_{\bar{x}}^2)$, and the error model variance $\sigma_{\bar{x}}^2$ is discussed below.

20   The likelihood for the standard deviation of the VADM is defined by

$$y_{\sigma} = h_{\sigma}(\theta) + \varepsilon_{\sigma}. \tag{17}$$

We use the average of the VADM standard deviations of PADM2M and Sint-2000 as the datum $y_{\sigma} = 1.66 \cdot 10^{22} \text{ Am}^2$. The standard deviation over time of the linear model (5) is used to define the function $h_{\sigma}(\theta) = \sqrt{D/\gamma}$. We use this approximation here for the same reason we adopt Kramers formula for computing the reversal rates. The nonlinear model requires long

25   simulations to estimate the time average or standard deviation of the VADM because simple analytical expressions are not available. Using the linear approximation avoids this requirement and gives results that are comparable to estimates from long realizations of the nonlinear model. As above, the error model $\varepsilon_{\sigma} \sim \mathcal{N}(0, \sigma_{\sigma}^2)$ is Gaussian and we discuss the error variance σ_{σ}^2 below.

Taken all together, the likelihood term $p_{l,\mathrm{td}}(y|\theta)$ in (13) is then given by the product of the three likelihoods defined by equations (14), (16) and (17):

$$p_{l,\mathrm{td}}(y|\theta) \propto \exp\left(-\frac{1}{2}\left(\left(\frac{y_{\mathrm{rr}} - h_{\mathrm{rr}}(\theta)}{\sigma_{\mathrm{rr}}}\right)^2 + \left(\frac{y_{\bar{x}} - \bar{x}}{\sigma_{\bar{x}}}\right)^2 + \left(\frac{y_\sigma - h_\sigma(\theta)}{\sigma_\sigma}\right)^2\right)\right). \tag{18}$$

Candidate values for the error variances σ_{rr}^2, $\sigma_{\bar{x}}^2$ and σ_σ^2 are as follows. The error variance of the reversal rate, $\sigma_{\bar{x}}^2$, can be based

5 on the standard deviations we computed from the Ogg (2012) chronology in Table 2. Thus we might use $\sigma_{\bar{x}} = 0.5$. A candidate for the standard deviation of the time average VADM is the difference of the time averages of Sint-2000 and PADM2M, which gives $\sigma_{\bar{x}} = 0.48 \cdot 10^{22}$ Am^2. Similarly, one can define the standard deviation σ_σ by the difference of VADM standard deviations (over time) derived from Sint-2000 and PADM2M. This gives $\sigma_\sigma = 0.36 \cdot 10^{22}$ Am^2.

The difficulty with this choice is that there are relatively few time domain observations compared with the large number of

10 spectral data in the power spectra. This vast difference in the number of time domain and spectral data means that the spectral data completely overwhelm the recovery of model parameters. To ensure that time domain observations also contribute to the parameter estimates we lower the error variances σ_{rr}, $\sigma_{\bar{x}}$ and σ_σ by a factor of 100 and set

$$\sigma_{\mathrm{rr}} = 0.05 \text{ reversals/Myr}, \quad \sigma_{\bar{x}} = 0.048 \cdot 10^{22} \text{ Am}^2 \quad \sigma_\sigma = 0.036 \cdot 10^{22} \text{ Am}^2. \tag{19}$$

We discuss this choice and its consequences on parameter estimates and associated uncertainty in more detail in Section 6.

15 ### 4.2.2 Low frequencies

The component $p_{l,\mathrm{lf}}(y|\theta)$ of the feature-based likelihood (13) addresses the behavior of the dipole at low frequencies of $10^{-4} - 0.5$ cycles per kyr, and is based on the PSDs of the Sint-2000 and PADM2M data sets. We construct the likelihood using the equation

$$y_{\mathrm{lf}} = h_{\mathrm{lf}}(\theta) + \varepsilon_{\mathrm{lf}}, \tag{20}$$

20 where y_{lf} represents the PSD of the Earth's dipole field at low frequencies, $h_{\mathrm{lf}}(\theta)$ maps the model parameters to the data y_{lf} and where $\varepsilon_{\mathrm{lf}}$ represents the errors we expect.

We define y_{lf} to be the mean of the PSDs of Sint-2000 and PADM2M. The function $h_{\mathrm{lf}}(\theta)$ maps the model parameters to the data y_{lf}. This could be done by performing simulations with the nonlinear Myr model and computing corresponding PSDs. We found, however, that this requires long simulations or else we have no accuracy in the computed PSDs. We thus simplify

25 these computations by approximating the PSD of the nonlinear model by the PSD of the linear model (5). This can be done simply by evaluating (6). To account for the smoothing introduced by sedimentation processes we define $h_{\mathrm{lf}}(\theta)$ to be a function that computes the PSD of the Myr model by using the "un-smoothed" spectrum of equation (6) for frequencies less than 0.05 cycles/kyr, and uses the "smoothed" spectrum of equation (7) for frequencies between $0.05 - 0.5$ cycles/kyr:

$$h_{\mathrm{lf}}(\theta) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \begin{cases} 1 & \text{if } f \leq 0.05 \\ \exp\left(-(4\pi^2 f^2 T_s^2)/12\right) & \text{if } 0.05 < f \leq 0.5 \end{cases} \tag{21}$$
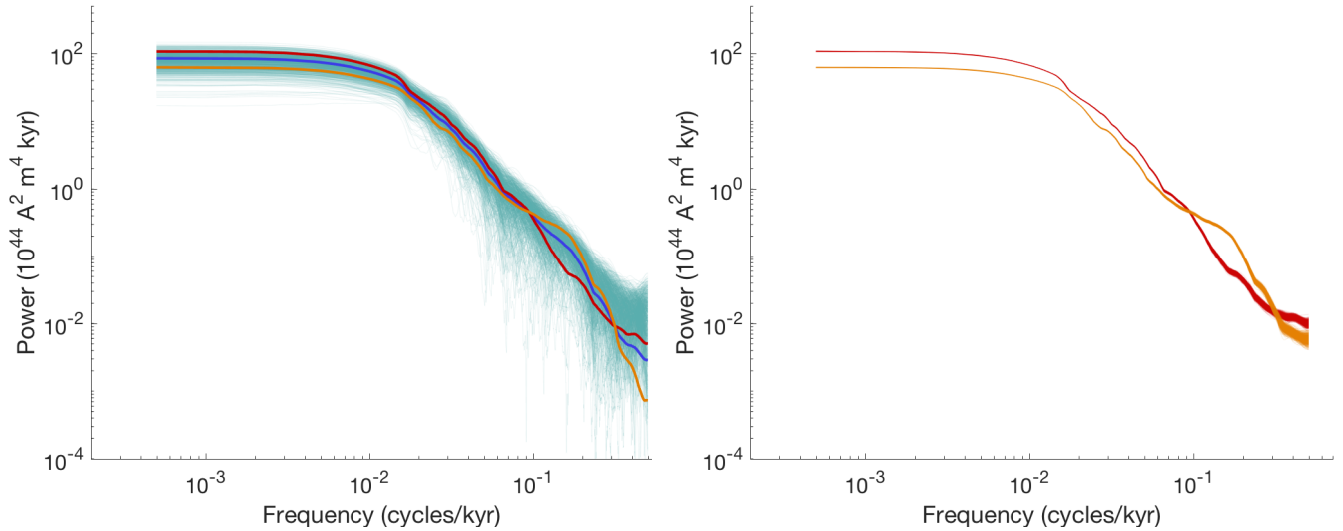
Figure 4. Left: low frequency data and error model due to shortness of record. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Blue: mean of PSDs of Sint-2000 and PADM2M (y_{lf}). Turquoise: 10^3 samples of the error model ε_{lf} added to y_{lf}. Right: error model based on errors in Sint-2000. Orange: 10^3 samples of the PSDs computed from "perturbed" Sint-2000 VADMs. Red: 10^3 samples of the PSDs computed from "perturbed" PADM2M VADMs.

Note that $h_{lf}(\theta)$ does not depend on $\bar{x}$ or a. This also means that the data regarding low frequencies are not useful for determining these two parameters (see Section 6).

The error model ε_{lf} is Gaussian with zero mean. The error covariance is constructed in a way to represent errors due to the shortness of the geomagnetic record. The uncertainty introduced by sampling the VADM once per kyr for only 2 Myrs is the dominant source of error in the power spectrum. An estimate for the error covariance is computed as follows. We perform 10^4 simulations, each of 2 Myr, with the nonlinear Myr model (3) and its nominal parameters (see Table 1). We compute the PSD of each simulation and build the covariance matrix of the 10^4 PSDs. In the left panel of Figure 4 we illustrate the error model by plotting the PSDs of PADM2M (red), Sint-2000 (orange), their mean, y_{lf}, (dark blue), and 10^3 samples of ε_{lf} added to y_{lf} (turquoise). Since the PSDs of Sint-2000 and PADM2M are well within the cloud of PSDs we generated with the error model, this choice for modeling the expected errors in low frequency PSDs seems reasonable to us.

For comparison, we also plot 10^3 samples of an error model that only accounts for the reported errors in Sint-2000. This is done by adding independent Gaussian noise, whose standard deviation is given by the Sint-2000 data set every kyr, to the VADM of Sint-2000 and PADM2M. This results in 10^3 "perturbed" versions of Sint-2000 or PADM2M. For each one, we compute the PSD and plot the result in the right panel of Figure 4. The resulting errors are smaller than the errors induced by the shortness of the record. In fact, the reported error does not account for the difference in the Sint-2000 and PADM2M data sets. This suggests that the reported error is too small. However, adopting larger errors for the Sint-2000 and PADM2M data

14

sets would not change the conclusion that the largest source of uncertainty in the power spectra is due to the shortness of the record.

4.2.3 High frequencies

We now consider the high frequency behavior of the model and use the CALS10k.2 data. We focus on frequencies between

5 $0.9 - 9.9$ cycles/kyr. We chose this window to avoid overlap between the PSDs of CALS10k.2 and Sint-2000/PADM2M. Our choice also acknowledges that the high-frequency part of the PSD for Sint-2000/PADM2M is less reliable. As above, we construct the likelihood $p_{l,\text{hf}}(y|\theta)$ from an equation similar to (20):

$$y_{\text{hf}} = h_{\text{hf}}(\theta) + \varepsilon_{\text{hf}}, \tag{22}$$

where y_{hf} is the PSD of CALS10k.2 in the frequency range we consider, $h_{\text{hf}}(\theta)$ is a function that maps model parameters to

10 the data and where ε_{hf} is the error model.

As above, we base $h_{\text{hf}}(\theta)$ on the PSD of the linear model (see equation (11)) to avoid errors due to computations in the time domain and set

$$h_{\text{hf}}(\theta) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \frac{a^2}{a^2 + 4\pi^2 f^2}, \tag{23}$$

where f is the frequency in the range we consider here. Recall that a^{-1} defines the correlation time of the noise in the kyr

15 model.

The error model ε_{hf} is Gaussian with mean zero and the covariance is designed to represent errors due to the shortness of the record. This is done, as above, by using 10 kyr simulations of the nonlinear model (9)-(10) with nominal parameter values. We perform 5000 simulations and for each one compute the PSD over the frequency range we consider $(0.9 - 9.9$ cycles/kyr). The covariance matrix computed from these PSDs defines the error model ε_{hf}, which is illustrated along with the low frequency

20 error model and the data in Figure 5.

This concludes the construction of the likelihood and, together with the prior (see Section 4.1) we have now formulated the Bayesian formulation of this problem in terms of the posterior distribution (12).

4.3 Numerical solution by MCMC

We solve the Bayesian parameter estimation problem numerically by Markov Chain Monte Carlo (MCMC). This means that

25 we use a "MCMC sampler" that generates samples from the posterior distribution in the sense that averages computed over the samples are equal to expected values computed over the posterior distribution in the limit of infinitely many samples. A (Metropolis-Hastings) MCMC sampler works as follows: the sampler proposes a sample by drawing from a proposal distribution, and the sample is accepted with a probability to ensure that the stationary distribution of the Markov chain is the targeted posterior distribution.

30 We use the affine invariant ensemble sampler, called MCMC Hammer, of Goodman and Weare (2010), implemented in Matlab by Grinsted (2018). The MCMC Hammer is a general purpose ensemble sampler that is particularly effective if there
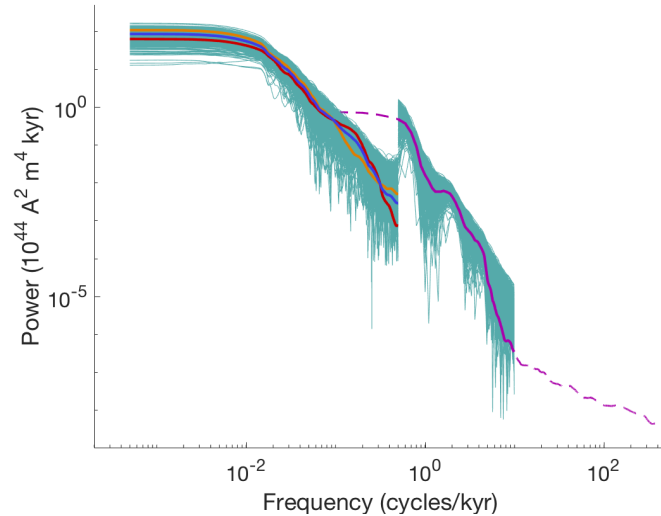
Figure 5. Data and error models for low and high frequencies. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Blue: mean of PSDs of Sint-2000 and PADM2M (y_{lf}). Turquoise (low frequencies): 10^3 samples of the error model $\varepsilon_{\mathrm{lf}}$ added to y_{lf}. Dashed purple: PSD of CALS10k.2. Solid purple: PSD of CALS10k.2 at frequencies we consider (y_{hf}). Turquoise (low frequencies): 10^3 samples of the error model $\varepsilon_{\mathrm{hf}}$ added to y_{hf}.

are strong correlations among the various parameters. The Matlab implementation of the method is easy to use, and requires that we provide the sampler with functions that evaluate the prior distribution and the likelihood, as described above.

In addition, the sampler requires that we define an initial ensemble of ten walkers (two per parameter). This is done as follows. We draw the initial ensemble from a Gaussian whose mean is given by the nominal parameters in Table 3, and whose

5 covariance matrix is a diagonal matrix whose diagonal elements are 50% of the nominal values. The Gaussian is constrained by the upper and lower bounds in Table 3. The precise choice of the initial ensemble, however, is not so important as the ensemble generated by the MCMC hammer quickly spreads out to search the parameter space.

We assess the numerical results by computing integrated auto correlation time (IACT) using the definitions and methods described by Wolff (2004). The IACT is a measure of how effective the sampler is. We generate an overall number of 10^6

10 samples, but the number of "effective" samples is $10^6/$IACT. For all MCMC runs we perform (see Sections 5 and 6), the IACT of the Markov chain is about 100. We discard the first $10 \cdot$ IACT samples as "burn in", further reducing the impact of the distribution of the initial ensemble. We also ran shorter chains with 10^5 samples and obtained similar results, indicating that the chains of length 10^6 are well resolved.

Recall that all MCMC samplers yield the posterior distribution as their stationary distribution, but the specific choice of

15 MCMC sampler defines "how fast" one approaches the stationary distribution and how effective the sampling is (Burn-in time and IACT). In view of the fact that likelihood evaluations are, by our design, computationally inexpensive, we may run (any)
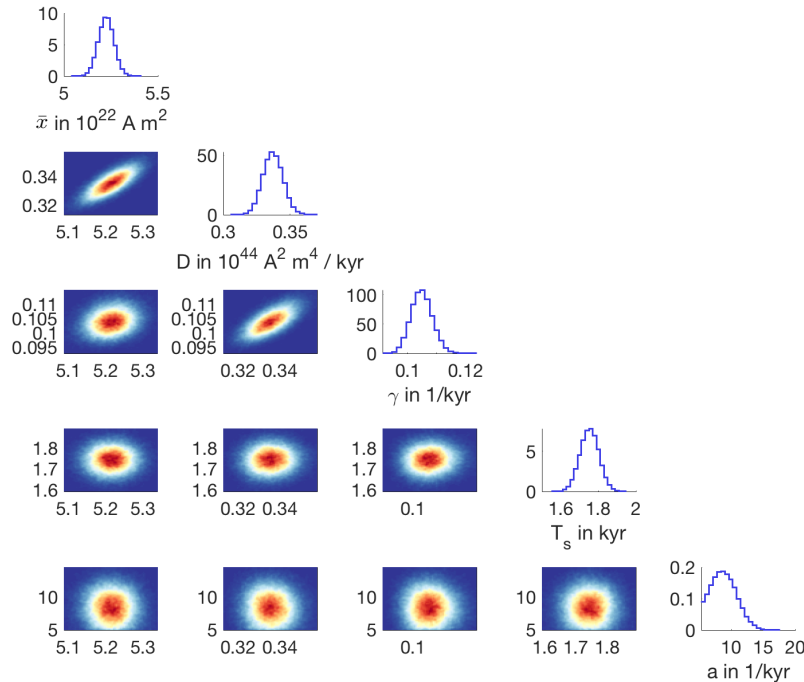
Figure 6. 1- and 2-dimensional histograms of the posterior distribution.

MCMC sampler to generate a long chain (10^6 samples). Thus, the precise choice of MCMC sampler is not so important for our purposes, however, the MCMC Hammer solves the problem efficiently.

The code we wrote is available on github: https://github.com/mattimorzfeld/ It can be used to generate 100,000 samples in a few hours and 10^6 samples in less than a day. For this reason, we can run the code in several configurations and with likelihoods 5 that are missing some of the factors that comprise the overall feature-based likelihood (13). This allows us to study the impact each individual data set has on the parameter estimates and it also allows us to assess the validity of some of our modeling choices, in particular with respect to error variances which are notoriously difficult to come by (see Section 6).

5   Results

We run the MCMC sampler to generate 10^6 samples, approximately distributed according to the posterior distribution. We 10 illustrate the posterior distribution by a corner plot in Figure 6. The corner plot shows all 1- and 2-dimensional histograms of the posterior samples. We observe that the four 1-dimensional histograms are well-defined "bumps" whose width is considerably smaller than the assumed parameter bounds (see Table 3). This means that the data constrain the parameters well. The 2-dimensional histograms indicate correlations among the parameters. These correlations can also be described by the correlation

17

$\bar{x}$	D	γ	T_s	a	σ	Rev. rate
(in 10^{22} Am^2)	(in 10^{44} A^{2}m^4 kyr^{-1})	(in kyr^{-1})	(in kyr)	(in kyr^{-1})	(in 10^{22} Am^2)	(in reversals/Myr)
5.21 (0.041)	0.34 (0.0074)	0.10 (0.0037)	1.74 (0.050)	8.57 (1.95)	1.80 (0.023)	4.06 (0.048)

Table 4. Posterior mean and standard deviation (in brackets) of the model parameters and estimates of reversal rate and VADM standard deviation

coefficients between the five parameters which we computed as

$$\rho = \begin{pmatrix} 1.00 & 0.76 & 0.20 & 0.01 & -0.02 \\ 0.76 & 1.00 & 0.67 & 0.04 & -0.04 \\ 0.20 & 0.67 & 1.00 & 0.05 & -0.02 \\ 0.01 & 0.04 & 0.05 & 1.00 & 0.003 \\ -0.02 & -0.04 & -0.02 & 0.003 & 1.00 \end{pmatrix}. \tag{24}$$

We note strong correlation between $\bar{x}$ and D and γ, which is due to the contribution of the reversal rate to the overall likelihood (see Equation (15)) and the dependence of the spectral data on D and γ (see Equations (7) and (11)). From the samples, we

5 can also compute means and standard deviations of all five parameters and we show these values in Table 4.

The table also shows the reversal rate and VADM standard deviation that we compute from 200 samples of the posterior distribution (using Equation (8) for each sample). We note that the reversal rate (4.06 reversals/Myr) is lower than the reversal rate we used in the likelihood (4.23 reversals/Myr). Since the posterior standard deviation is 0.049 reversals/Myr, the reversal rate data are about four standard deviations away from the mean we compute. Similarly, the posterior VADM standard de-

10 viation (mean value of $1.80 \cdot 10^{22}$ Am^2) is also far (in terms of posterior standard deviations) from the value we use as data (1.6610^{22} Am^2). The model fit to the spectral data, however, is good, as is illustrated in the left panel of Figure 7. Here we plot 100 PSDs we computed from 2 Myr and 10 kyr model runs and where each model run uses a parameter set drawn at random from the posterior distribution. The right panels of Figure 7 show a Myr model run (top) and kyr model run using the posterior mean values for the parameters. For comparison, the Figure also shows the PADM2M, Sint-2000 and CALS10k.2 data.

15 In summary, we conclude that the likelihoods we constructed and the assumptions about errors we made lead to a posterior distribution that constrains the model parameters tightly. The posterior distribution describes a set of model parameters that yield model outputs that are comparable with the data in the feature-base sense. The estimates of the uncertainty in the parameters, e.g., posterior standard deviations, however should be used with the understanding that error variances are not easy to define. For the spectral data, we constructed error models that reflect uncertainty induced by the shortness of the geomagnetic

20 record. For the time domain data (reversal rate, time average VADM and VADM standard deviation) we used error variances that are smaller than intuitive error variances to account for the fact that the number of spectral data points (hundreds) is much larger than the number of time domain data points (three data points). Moreover, the reversal rate and VADM standard deviation data are far (as measured by posterior standard deviations) from the reversal rate and VADM standard deviation of the
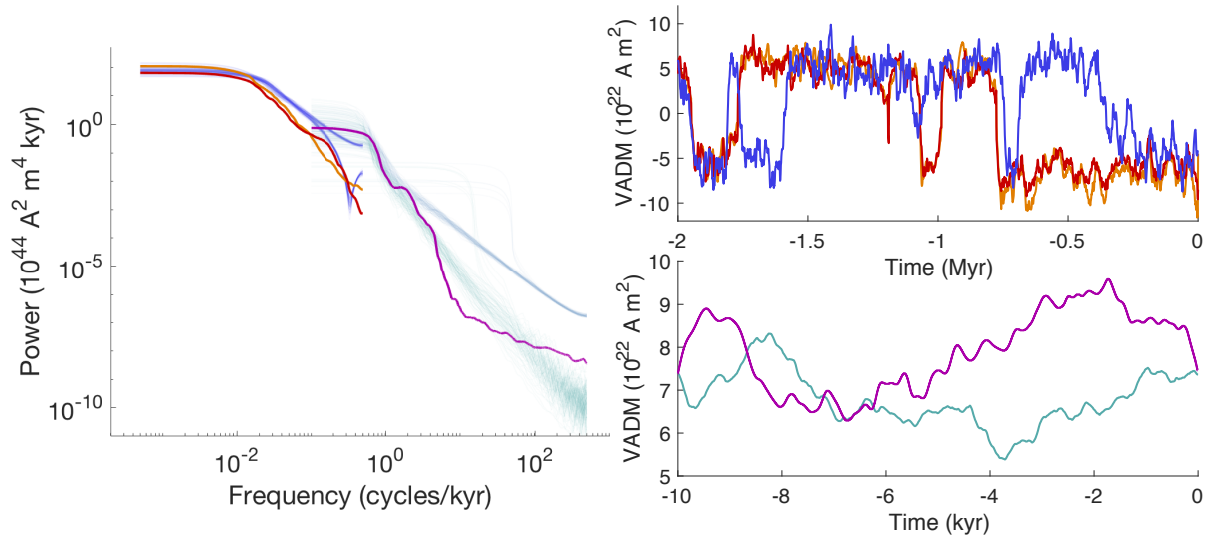
Figure 7. Parameter estimation results. Left: PSDs of data and model. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Purple: PSD of CALS10k.2. Dark blue: PSDs of 100 posterior samples of Myr model (with and without smoothing). Light blue: PSDs of 100 posterior samples of kyr model with uncorrelated noise. Turquoise: PSDs of 100 posterior samples of kyr model with correlated noise. Right, top: Sint-2000 (orange), PADM2M (red) and Myr model (blue). Right, bottom: CALS10k.2 (purple), and kyr model with correlated noise (turquoise).

model with posterior parameters. This is a first hint at that there are inconsistencies between spectral data and time domain data, which we will study in more detail in the next section.

6 Discussion

We study the effects the independent data sets have on the parameter estimates and also study the effects of different choices

5   for error variances for the time domain data (reversal rate, time average VADM and VADM standard deviation). We do so by running the MCMC code in several configurations. Each configuration corresponds to a posterior distribution and, therefore, to a set of parameter estimates. The configurations we consider are summarized in Table 4 and the corresponding parameter estimates are reported in Table 5. Configuration (a) is the default configuration described in the previous sections. We now discuss the other configurations in relation to (a) and in relation to each other.

10   Configuration (b) differs from configuration (a) in that the CALS10k.2 data are not used, i.e., we do not include the high-frequency component, $p_{l,\mathrm{hf}}(y|\theta)$ in the feature-based likelihood (13). Configurations (a) and (b) lead to nearly identical posterior distributions and, hence, nearly identical parameter estimates with the exception of the parameter a, which controls the correlation of the noise on the kyr time scale. In configuration (b), the estimate (posterior mean) of a is nearly equal to its prior mean (average of lower and upper bound) and the posterior standard deviation is large. Because the posterior distribution

15   over a is nearly identical to its prior distribution, we conclude that a is not well constrained by the data of configuration (b).

Nonlinear Processes
in Geophysics
Discussions

Configuration	(a)	(b)	(c)	(d)	(e)	(f)
PADM2M & Sint-2000	✓	✓	✓	✓	✗	✓
CALS10k.2	✓	✗	✓	✓	✗	✗
Rev. Rate, time avg., std. dev.	✓	✓	✓	✗	✓	✗
σ_{rr} (reversals/Myr)	0.05	0.05	0.5	N/A	0.05	N/A
$\sigma_{\bar{x}}$ (10^{22} Am^2)	0.048	0.048	0.48	N/A	0.048	N/A
σ_σ 10^{22} Am^2	0.036	0.036	0.36	N/A	0.036	N/A

Table 5. Configurations for several Bayesian problem formulations. A checkmark means that the data set is used; a cross means it is not used in the overall likelihood construction. The standard deviations (σ) define the Gaussian error models for the reversal rate, time average VADM and VADM standard deviation.

Configuration	(a)	(b)	(c)
$\bar{x}$ (10^{22} Am^2)	5.21 (0.041)	5.21 (0.042)	3.56 (0.26)
D (10^{44} A^{2}m^4 kyr^{-1})	0.34 (0.0074)	0.34 (0.0073)	0.13 (0.014)
γ (kyr^{-1})	0.10 (0.0037)	0.10 (0.0035)	0.081(0.0052)
T_s (kyr)	1.74 (0.050)	1.74 (0.050)	1.68 (0.13)
a (kyr^{-1})	8.57 (1.95)	22.45 (10.17)	11.66 (3.93)
σ (10^{22} Am^2)	1.80 (0.023)	1.80 (0.023)	1.25 (0.062)
Rev. rate (reversals/Myr)	4.06 (0.048)	4.05 (0.046)	3.33 (0.52)
Configuration	**(d)**	**(e)**	**(f)**
$\bar{x}$ (10^{22} Am^2)	5.04 (2.91)	5.56 (0.048)	5.04 (2.88)
D (10^{44} A^{2}m^4 kyr^{-1})	0.094 (0.015)	0.48 (0.025)	0.093 (0.015)
γ (kyr^{-1})	0.078 (0.0063)	0.18 (0.016)	0.077 (0.0064)
T_s (kyr)	1.64 (0.19)	2.98 (1.15)	1.64 (0.19)
a (kyr^{-1})	12.92 (4.79)	22.45 (10.12)	22.37 (10.13)
σ (10^{22} Am^2)	1.10 (0.077)	1.65 (0.036)	1.09 (0.077)
Rev. rate (reversals/Myr)	2.93 (4.08)	4.23 (0.054)	2.80 (3.90)

Table 6. Posterior parameter estimates (mean and standard deviation) and corresponding VADM standard deviation (σ) and reversal rates for five different set ups (see Table 5).

The fact that a is not constrained by the Sint-2000, PADM2M and time domain data is not surprising since a only appears in the Bayesian parameter estimation problem via the high-frequency likelihood $p_{l,\mathrm{hf}}(y|\theta)$; since $p_{l,\mathrm{lf}}(y|\theta)$ and $p_{l,\mathrm{td}}(y|\theta)$ are independent of a, the marginal of the posterior distribution of configuration (b) over the parameter a is independent of the data. More interestingly, however, we find that all other model parameters are estimated to have nearly the same values, independently of whether CALS10k.2 being used during parameter estimation or not. This latter observation indicates that the model

5

is self-consistent and consistent with the data on the Myr and kyr time scales; CALS10k.2 is mostly useful for constraining the noise correlation parameter a.

Configuration (c) differs from configuration (a) in the error variances for the time domain data (reversal rate, time average VADM and VADM standard deviation). With the larger values used in (c), the reversal rate drops to about 3 reversals/Myr,

5   and the time average VADM and VADM standard deviation also decrease significantly as compared to configuration (a). The posterior mean of D decreases by more than 50%; γ and T_s are comparable for configurations (a)-(c).

In configuration (d), the spectral data are used, but the time domain data are not used (which corresponds to infinite σ_{rr}, σ_σ and $\sigma_{\bar{x}}$). We note that the posterior means and variances of all parameters are comparable for configurations (c) and (d) but are quite different from the parameter estimates of configuration (a). Thus, we conclude that if the error variances of the time

10   domain data are large, then the impact of these data is minimal. The reason is that the number of spectral data points is larger (hundreds) than the number of time domain data (three data points: reversal rate, time average VADM and VADM standard deviation). When the error variances of the time domain data decrease, the impact these data have on the parameter estimates increases. For an overall good fit of model and *all* data small error variances for the time domain data are required or else the model yields reversal rates that are too low. Using small error variances, however, comes at the cost of not necessarily realistic

15   posterior variances.

Comparing configurations (d) and (e), we note that if only the spectral data are used, the reversal rates are unrealistically small. Moreover, the parameter estimates based on the spectral data are quite different from the estimates we obtain when we use the time domain data (reversal rate, time average VADM and VADM standard deviation). This is further evidence that the model has some inconsistencies: a good match to spectral data requires a set of model parameters that is different from the set

20   of model parameters that lead to a good fit to the reversal rate, time average VADM and VADM standard deviation.

Finally, comparing configurations (d) and (f), we can further study the effects that the CALS10k.2 data have on parameter estimates (similarly to how we compared (a) and (b) above). The results shown in Table 6, indicate that the parameter estimates based on (d) and (f) are nearly identical, except in the parameter a that controls the time correlation of the noise on the kyr time scale. This confirms what we already found by comparing configurations (a) and (b): the CALS10k.2 data are mostly

25   useful for constraining a. These results, along with configurations (a) and (b), suggest that the model is self consistent with the independent data on the Myr scale (Sint-2000 and PADM2M) and on the kyr scale (CALS10k.2). Our experiments, however, also suggest that the model has difficulties to reconcile the spectral and time domain data.

7   Summary and conclusions

We designed a Bayesian estimation problem for the parameters of a family of stochastic models that can describe the Earth's

30   magnetic dipole over kyr and Myr time scales. The main challenge here is that the data are limited, that each datum is the result of years of hard work, and that the data have large uncertainties and unknown errors. For that reason, we adapted the usual Bayesian approach to parameter estimation to be more suitable for using a *collection* of diverse data sources for parameter estimation. The main tool in this context are "features" derived from the models and data. Likelihoods for the Bayesian problem

are then defined in terms of the features rather than the usual point-wise errors in model outputs and data. The feature-based approach enables fusing different types of data and to assess the internal consistency of the data and the underlying model. Numerical solution of the feature-based estimation problem is done via conventional MCMC (an affine invariant ensemble sampler). We used the full paleomagnetic record to estimate parameters and our numerical results indicate that these data

5 constrain all parameters of the model. Moreover, the posterior parameter values yield model outputs that fit the data in a precise, feature-based sense.

Formulating the parameter estimation problem requires estimating model error and in particular defining model error variances. We have carefully investigated the validity of our error model choices by a set of numerical experiments. Further numerical experiments revealed the impact individual data sets have on parameter estimates. Our numerical experiments also

10 suggest that the model has a deficiency in that there are inconsistencies between the model's spectra and its reversal rates. It is also possible that the data themselves are not entirely self consistent in this regard. Our methodology does not resolve these questions, but it does provide an effective strategy for combining diverse data sets that had previously been treated separately. This gives us an opportunity to expose inconsistencies between the data and models, which is an important step for making progress in data-limited field.

15 *Code and data availability.* The code and data used in this paper is available on github: https://github.com/mattimorzfeld

Competing interests. No competing interests are present.

References

Asch, M., Bocquet, M., and Nodet, M.: Data assimilation: methods, algorithms and applications, SIAM, 2017.

Buffett, B.: Dipole fluctuations and the duration of geomagnetic polarity transitions, Geophysical Research Letters, 42, 7444–7451, 2015.

Buffett, B. and Davis, W.: A Probabilistic Assessment of the Next Geomagnetic Reversal, Geophysical Research Letters, 45, 1845–1850,
5   https://doi.org/10.1002/2018GL077061, 2018.

Buffett, B. and Matsui, H.: A power spectrum for the geomagnetic dipole moment, Earth and Planetary Science Letters, 411, 20–26, 2015.

Buffett, B. and Puranam, A.: Constructing stochastic models for dipole fluctuations from paleomagnetic observations, Physics of the Earth
      and Planetary Interiors, 272, 68 – 77, 2017.

Buffett, B., Ziegler, L., and Constable, C.: A stochastic model for paleomagnetic field variations, Geophysical Journal International, 195,
10   86–97, 2013.

Buffett, B. A., King, E. M., and Matsui, H.: A physical interpretation of stochastic models for fluctuations in the Earth's dipole field,
      Geophysical Journal International, 198, 597–608, 2014.

Cande, S. and Kent, D.: Revised calibration of the geomagnetic polarity timescale for the late Cretaceous and Cenozoic, Journal of Geophys-
      ical Research: Solid Earth, 100, 6093–6095, 1995.

15 Chorin, A. and Hald, O.: Stochastic tools in mathematics and science, Springer, third edn., 2013.

Constable, C. and Johnson, C.: A paleomagnetic power spectrum, Physics of Earth and Planetary Interiors, 153, 61–73, 2005.

Constable, C., Korte, M., and Panovska, S.: Persistent high paleosecular variation activity in southern hemisphere for at least 10000 years,
      Earth and Planetary Science Letters, 453, 78 – 86, 2016.

Finlay, C., Olsen, N., Kotsiaros, S., Gillet, N., and Lars, T.: Recent geomagnetic secular variation from Swarm and ground observatories as
20   estimated in the CHAOS-6 geomagnetic field model, Earth Planets and Space, 68, 1–18, 2016.

Gissinger, C.: A new deterministic model for chaotic reversals, The European physical Journal B, 85, 137, 2012.

Goodman, J. and Weare, J.: Ensemble samplers with affine invariance, Comm. App. Math. Com. Sc., 5, 65–80, 2010.

Grinsted, A.: GWMCMC, https://github.com/grinsted/, 2018.

Hoyng, P., Ossendrijver, M., and Schmitt, D.: The geodynamo as a bistable oscillator, Geophysical and Astrophysical Fluid Dynamics, 94,
25   263–314, 2001.

Hulot, G., Finlay, C. C., Constable, C. G., Olsen, N., and Mandea, M.: The magnetic field of planet Earth, Space Science Reviews, 152,
      159–222, 2010.

Lowrie, W. and Kent, D.: Geomagnetic polarity time scale and reversal frequency regimes, Timescales of the paleomagnetic field, 145,
      117–129, 2004.

30 Maclean, J., Santitissadeekorn, N., and Jones, C. K.: A coherent structure approach for parameter estimation in Lagrangian Data Assimilation,
      Physica D: Nonlinear Phenomena, 360, 36 – 45, https://doi.org/https://doi.org/10.1016/j.physd.2017.08.007, http://www.sciencedirect.
      com/science/article/pii/S0167278916304572, 2017.

Meduri, D. and Wicht, J.: A simple stochastic model for dipole moment fluctuations in numerical dynamo simulations, Frontiers in Earth
      Science, 4, 2016.

35 Morzfeld, M., Fournier, A., and Hulot, G.: Coarse predictions of dipole reversals by low-dimensional modeling and data assimilation, Physics
      of the Earth and Planetary Interiors, 262, 8 – 27, https://doi.org/https://doi.org/10.1016/j.pepi.2016.10.007, http://www.sciencedirect.com/
      science/article/pii/S0031920116301431, 2017.

Nonlinear Processes
in Geophysics
Discussions

Open Access

EGU

Morzfeld, M., Adams, J., Lunderman, S., and Orozco, R.: Feature-based data assimilation in geophysics, Nonlinear Processes in Geophysics, 25, 355–374, 2018.

Ogg, J.: Geomagnetic polarity time scale, in: The geological time scale 2012, edited by Gradstein, F., Ogg, J., Schmitz, M., and Ogg, G., chap. 5, pp. 85–1130, Elsevier Science, Boston, 2012.

5 Pétrélis, F. and Fauve, S.: Chaotic dynamics of the magnetic field generated by dynamo action in a turbulent flow, Journal of Physics: Condensed Matter, 20, 494 203, 2008.

Pétrélis, F., Fauve, S., Dormy, E., and Valet, J.-P.: Simple mechanism for reversals of Earth's magnetic field, Physical Review Letters, 102, 144 503, 2009.

Pourovskii, L., Mravlje, J., Georges, A., Simak, S., and Abrikosov, I.: Electron-electron scattering and thermal conductivity of epsilon-iron

10 at Earth's core conditions, New Journal of Physics, 19, 073 022, 2017.

Pozzo, M., Davies, C., Gubbins, D., and Alfe, D.: Thermal and electrical conductivity at Earth's core conditions, Nature, 485, 355–358, 2012.

Reich, S. and Cotter, C.: Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, 2015.

Rikitake, T.: Oscillations of a system of disk dynamos, Mathematical Proceedings of the Cambridge Philosophical Society, 54, 89–105, 1958.

15 Risken, H.: The Fokker-Planck equation: Methods of solution and applications, Springer, 1996.

Roberts, A. and Winkhofer, M.: Why are geomagnetic excursions not always recorded in sediments? Constraints from post-depositional remanent magnetization lock-in modeling, Earth and Planetary Science Letters, 227, 345–359, 2004.

Valet, J.-P., Meynadier, L., and Guyodo, Y.: Geomagnetic field strength and reversal rate over the past 2 million years, Nature, 435, 802–805, 2005.

20 Wolff, U.: Monte Carlo errors with less errors, Comput. Phys. Commun., 156, 143–153, 2004.

Ziegler, L. B., Constable, C. G., Johnson, C. L., and Tauxe, L.: PADM2M: a penalized maximum likelihood model of the 0-2 Ma paleomagnetic axial dipole model, Geophysical Journal International, 184, 1069–1089, 2011.