A comprehensive model for the kyr and Myr time scales of Earth's axial magnetic dipole field – response to Reviewers' comments

Matthias Morzfeld, Bruce A. Buffett

We thank both Reviewers for their careful reviews and good suggestions. We incorporated most of the Reviewers' comments into our revised manuscript. A detailed, point-by-point response is provided below. We repeat the Reviewers' comments in italic and our responses appear in standard font. We also provide a latex-diff that can be used to track all changes from our original submission to this revised version. The line numbers below (and in the reviews) refer to the original submission which we also provide.

In addition to the changes described below, we revised the entire manuscript based on the Reviewers' comments. This has led us to remove Figure 2 (it was not really a necessary figure). We split Figure 3 into two figures and included additional information about our error models in a revised Figure 7 (all Figure numbers refer to the original submission). We further included several new figures. We revised Sections 3 and 4 for clarity and brevity, and expanded many of our explanations (as requested by the Reviewers), which causes the revised manuscript to be longer than the original submission. We included a new section that touches on the usefulness of our approach. All these changes are inspired by the Reviewers' comments, but not all of them are directly related to the Reviewers' major or minor points.

We further included a correction factor for the approximate standard deviation (see Equation (13) of the revised manuscript). The correction factor is (for most relevant parameter values) near one and, for that reason, including the correction causes only minor changes in the numerical (posterior) estimates and does not change our conclusions or overall results.

Response to Reviewer 1 (Johannes Wicht)

The geomagnetic field varies on time scales from about a year to several tens of million years. The vastly different time scales have been revealed by various data sources, all having their specific inherent problems and limitations. In their pioneering work from 2001, Hoyng and Schmitt suggest that at least the axial dipole variations on time scales from millennia to some millions of years can be describe by the Langevin equation, a simple stochastic differential equation used, for example, to model Brownian motion. In recent years, the respective model has been refined to include the effects of correlated noise, random errors, or the limited time resolution of sedimentary data. This manuscript introduces another refinement, a Bayesian approach that allows to incorporate different types of "data" in a probabilistic manor to constrain the (five) parameters of the stochastic model. The paper is interesting and well written, but requires a few additional clarifications here and there (see below). The authors already mention that their analysis reveals difficulties and/or inconsistencies, but the paper remains too vague at this point. There problem could result from the different (and inconsistent) treatment of time-domain and frequency domain data, but this is hard to judge from the manuscript. A generally more critical discussion of the approach also seems in order. In addition, it remains unclear whether the results reveal anything new about geomagnetic field variations. A few respective additional sentences, for example in the conclusion, would certainly strengthen the paper.

We thank the Reviewer for the suggestions. We revised the manuscript in several places to tighten our statements about model and data inconsistencies (see also below). We further extended our discussion and provide more details about the limitations and advantages of our overall approach. We have added a new section to indicate why and how our approach is useful for studying the geomagnetic field. In short, a model is not a reliable scientific tool unless the model parameters are carefully chosen (based on "data") and unless limitations and uncertainties of the model are understood. We describe here a framework for doing just that for a stochastic model of the geomagnetic field.

Response to major points

1. *Data: the stochastic model is constraint based on paleomagnetic and archeomagnetic data, which both have many problems. The dating uncertainties and the smoothing due to the lock-in-time in sediments are mentioned in the text, but there are more. While there is little the authors can do about this, some more critical discussion and application seems in order. For example, the fact that the two paleomagnetic models show sizable differences suggests using a large uncertainty when modelling the respective data. The reversal rate is another example. How well can one determine the reversal rate based on a 30 Myr record when the underlying process is Poissonian?*

   The Reviewer is right in that there is a long list of problems. We did our best to bring up the ones that are most severe and most relevant for our purposes, but we do not claim that our list of problems is complete. We do use a large uncertainty when modeling the two paleomagnetic data sets. The uncertainties that arise from differences of the two data sets, however, are small in comparison to uncertainties that arise due to the fact that the time series is "only" 2 Myrs long. We describe this in detail in Section 4.2.2, see in particular Figure 4 (of the new submission). We have added explanations that the reversal rate cannot be accurately derived form a 30 Myr record. As is the case with the paleomagnetic data, the main source of this uncertainty is the "shortness" of the geomagnetic record. We have revised the manuscript throughout to highlight this point. How well one can determine the reversal rate based on a 30 Myr record when the underlying process is Poissonian depends on the rate of the Poissonian process and on how certain one is about one's knowledge of that rate.

2. *Geomagnetic power spectra: The power spectra and the respective models are not discussed in any detail. Also missing is an explanation how the spectra are incorporated in the Bayesian approach. Combining hundreds of "frequency data points" with only three ?time domain data points? may not be the best way to proceed and certainly required some care. A parametrization or identification of the most important spectral features seems the way to go here. See Baerenzung et al. (2018) for such an alternative approach. Figure 3 and 7 suggest that neither the spectra from the stochastic model runs nor the theoretical spectra do a very convincing job in replicating the spectra from the data, at least when it comes to the location of "knees" or of typical slopes. Some additional text seems in order here. How much can we trust the spectra and in which frequency range? What are the limitation of the stochastic model when it comes to the spectra? How much agreement can we reasonably expect? Please also remind the reader why the Sint-2000/PADM2M and CALS10K.2 spectra are so different where they overlap in frequencies. Why does figure 7 show a different range than figure 3? Do you really think that the stochastic model can capture the high frequency part shown in figure 7?*

   The power spectral densities are computed from the time series (Sint-2000, PADM2M, CALS10k.2) using the multi-taper technique of Constable and Johnson (2005). Sections 4.2.2 and 4.2.3 describe how the PSDs are incorporated in the Bayesian approach (via feature-based likelihoods). We do not agree that a parametrization or identification of the most important spectral features is obviously "the way to go here" as it may not be straightforward to extract these features from the data. We believe both approaches have advantages and disadvantages, but decided to stick to our approach of tuning the error variances. We also did not find Baerenzung et al. (2018) particularly useful for learning about how to parametrize or identify the most important spectral features. Nonetheless, we added this paper to our references. We have added additional text to discuss in more detail the "quality" of the fit. We have also included additional figures to make our points more clear. We have overlooked this before and thank the Reviewer for brining up that the original submission indeed lacked a lot of required explanations. We now use the same frequency range for (what used to be) Figures 3 and 7.

3. *Critical discussion: The authors already point out some problems or "inconsistencies" in the sense that their model cannot capture all "data" convincingly. This should be discussed in more detail. Is this a problem in the data or is the model too simple? The deficiencies in describing the spectra seems to imply the latter. It seems to me that the model is doing OK for describing the long-term variations where any complexity due to the convection and the internal dynamo process may not matter so much. Implementing a archeomagnetic (or even historic model) seem then overambitious. The authors should also discuss whether we can we learn anything about the geodynamo from this approach?*

   We have added explanations and clarifications and discuss the inconsistencies in model and data in more detail. It is difficult to come to a indisputable conclusion as to whether the inconsistencies are (mainly) due to the data or due to the model being too simple. Clearly, the model is simple, but also the data are known to be inconsistent. We explain these difficulties, but ultimately are not able to judge if the model is too simple or the data too inconsistent. We added a new section to discuss how our approach can be useful for studying the geomagnetic field (Section 7). We believe that our main finding about modeling of the geomagnetic field is that uncertainties in the data are dominated by errors that arise from the shortness of the record. Errors that arise due to the specifics of how the data are obtained (e.g., differences between PADM2M and Sint-2000) seem minor in comparison.

Response to minor points

1. *Please check the way you cite. You seem to mix up citep and citet.*

   We thank the Reviewer for bringing this to our attention. This was a careless mistake and we fixed it.

2. *Abstract, last sentence: Bayesian reasoning is frequently used for combining different data. What exactly is new in your approach?*

   We revised this sentence and hope that our revision is acceptable for the Reviewer. While it is true that Bayesian methods are used to combine different data, it is not well-understood what to do in a situation when several data describe the same quantity, but not necessarily in a consistent way. It is difficult to explain in a sentence or two why our approach is useful in this situation, which is why we decided not to include a (vague) description in the abstract.

3. *Same sentence: ... data sets, which is particularly ...*

   We fixed this typo.

4. *Page 1, line 23: What do you mean by: ... even basic analytical calculations are often intractable?*

We dropped any mention of analytical calculations.

5. *Page 2, line 5 and following: To my knowledge, the basic concept has been introduced by Hoyng, Schmitt and Ossendrijver in a series of papers in 2001 and 2002 (see below). Please give them credit. What exactly are the new ingredient in the B13 model? Note that Meduri & Wicht (2016) claim that the linearization used later in the paper (and which may also be a component of model B13) may only be of limited use.*

   We included the suggested references and also discussed Meduri & Wicht (2016) in more detail.

6. *Page 2, line 54: Define SDE.*

   We fixed this issue.

7. *Page 2, line 19: Buffett and Puranam (2017) try to mimic the effects of sedimentation*

   We use the suggested formulation in the revised manuscript.

8. *Page 3, line 29: . . . to reduce the influence of non-dipole components and various error sources . . .*

   This sentence disappeared in our revision.

9. *Page 4, line 3: CALSK10k.2 sampled at an interval of 1 year? Is this really the model's resolution?*

   We have added clarification of this issue. CALS10k.2 can be sampled at any rate (suggested is 1 yr to 200 yrs) and the nominal resolution is about 100 yrs.

10. *Page 5, line 14: At least set "Brownian motion" in quotation marks.*

    We do not understand this comment. In our revision, we decided to bring up the term "Brownian motion" only when we define the stochastic process, commonly known as Brownian motion (top of page 6 of revised manuscript).

11. *Page 5, line 18: Concerning a constant D, see Meduri & Wicht (2016).*

    We have added clarification of this issue.

12. *Page 5, line 22: . . . well potential potential .*

    We fixed this typo (and many more).

13. *Caption of fig. 2: ... and potential U(x), with U?(x)=- v(x) ... Or use an integral formulation.*

    This figure has been removed in the revision because it was not essential.

14. *Page 6, line 4: Could you explain iid in a few words for the non-experts?*

    We added a definition of iid.

15. *Page 6, line 9: ... are affected by affected by ...*

    We fixed this typo (and many more).

16. *Caption of fig. 3: There is something wrong with the sentences.*

    We fixed the issues in the figure caption. Figure 3, however, has changed during our revision (we made it into two separate figures).

17. *Page 8: Because an SDE is noisy ... Well, no surprise there. Could you be more specific? Wouldn't this also depend on the noise parameters? How expensive is an SDE integration? How long do you have to integrate?*

    We added clarification of this issue. The Reviewer is right in that the required simulation length depends on the noise parameters (small noise will not be an issue). With nominal parameters we decided that even a simulation of 10 billion years is not sufficiently accurate (in the resulting PSD) for our purposes. Our sampling approach requires repeated simulations. One MCMC run requires about 1 million simulations. We perform six different MCMC runs to check the validity of error models, the model's limitations and the impact of the various data sources on parameter estimates. Even with a 10 billion year simulation time, this would require a little more than 1 month or so in computation time (not using parallelism in the MCMC, timing a 10 billion year simulation at 3.6 seconds and assuming 10^6 simulations for the MCMC). The code that uses the approximations runs in less than a day. We have also validated our approximations against long simulations (using only nominal parameters).

18. *Page 8, line 10: The comparison does not look too good. Please discuss.*

    We added a discussion of the model-data fit and its quality.

19. *Page 8, line 14: ... the time averaged value of the absolute value of x(t) ..*

    This sentence does not appear in the revised manuscript.

20. *Page 8, line 23: approximating → approximate*

    This sentence does not appear in the revised manuscript.

21. *The SD is not very close to the observation. Please discuss.*

    We added a discussion of the standard deviation not being very close to the observation.

22. *Sampled once per year . . . (see above).*

    See above.

23. *Discuss the comparison.*

    We added a discussion of the comparison.

24. *... the bound may be overly pessimistic ... Well, the problem is not solved yet, but it looks like electron-electron interaction could at best only have a mild effect. Anyway, there is no need to dive into this topic in the paper and I would simply drop this sentence.*

    We have revised our derivation of the bounds according to the Reviewer's suggestion.

25. *Note that the stochastic model represents longer statistical time scales. Arguing with flow velocities is of limited use here.*

    The only place we mention flow velocity is where we put bounds on the parameter γ. We simply require the effects of diffusion to be small relative to the induction term, which depends on velocity. We make no claim that the stochastic model provides an estimate of flow velocity.

26. *Sections 4.2.2 and 4.2.3: Please provide more details here. (See comment above)*

    We have added new explanations and supplied more details in our revision.

27. *Figure 6: Provide colorbar. Discuss a bit more. How well are the parameters constrained, for example compared to the priors?*

    The use of a color-bar is unusual in the context of corner plots. We added explanations of how the posterior distributions compares to the uniform prior based on the bounds in Section 3.4.

28. *Page 17, line 2: The "however" seems out of place.*

    We revised this sentence based on the Reviewer's suggestion.

29. *Page 17, line 5: ... the impact of each ...*

    We fixed this typo.

30. *Page 21, line 1: ... in the context of our simplistic model CALS10k.2 mostly constraints ...*

    We revised the sentence based on the Reviewer's suggestion.

31. *Page 21, line 30: Every geomagnetic data point is indeed the result of hard work, but why is this a challenge for the model described here?*

    We agree, this is not an "issue" here and we removed this part of the sentence. The hard work required for each data point contributes to the fact that the amount of data we have is rather limited.

32. *Page 22, top: Incorporating different type of "data" in a Bayesian approach is a standard application. Please point out the specific novelty in your approach.*

    We revised our explanations to emphasize why our overall approach is useful. To be sure, we do not "invent" new algorithms in this paper, we apply known numerical techniques to an interesting and important problem and discuss their uses and, to some extent, their implications.

33. *Page 22, line 4: "We use the full paleomagnetic record"? The years of hard work have resulted in more than just Sint-2000 and PADM2M.*

    We agree, we revised this statement according to the Reviewer's suggestion.

References

Baerenzung, J., Holschneider, M., Wicht, J. Sanchez, S. and Lesur, V., Modeling and Predicting the Short-Term Evolution of the Geomagnetic Field, J. Geophys Res. (solid Earth), 123, pp.4539-4560, 2018.

Hoyng, P., Ossendrijver, M.A.J.H. and Schmitt, D., The geodynamo as a bistable oscillator,Geosphys. Astrophys. Fluid. Dyn., 94, 2001.

Hoyng, P., Schmitt, D. and Ossendrijver, M.A.J.H., A theoretical analysis of the observed variability of the geomagnetic dipole field, Phys. Earth Planet. Int., 130, pp. 143-157, 2002.

Schmitt, D., Ossendrijver, M.A.J.H. and Hoyng, P., Magnetic field reversals and secular variation in a bistable geodynamo model, Phys. Earth Planet. Int., 125, pp. 119-124,

Response to Reviewer 2

The authors take an interesting Bayesian approach, but there is a number of issues which require an improvement of the manuscript.

We thank the Reviewer for their comments. We have thoroughly revised the manuscript throughout to address the concerns. We have expanded explanations and (hopefully) clarified all the issues that are brought up.

The developed stochastic model uses of paleomagnetic and archeomagnetic data. How do the authors treat uncertainties in dating, smooth the data and carry out other data massages?

The treatment of uncertainties is discussed in Section 4.2. The validity of the assumptions about our error models (uncertainties) are assessed by a suite of numerical experiments in Section 6. Uncertainties in dating are not important for our purposes (assuming that the overall length of the Sint-2000 and PADM2M time series is about 2 Myrs and the overall length of CALS10k.2 is about 10 kyrs). We do not smooth the data or carry out any other data "massages". We use the Sint-2000, PADM2M and CALS10k.2 data sets directly, without modifications.

How can one combine many frequency-domain data points with a few time-domain data points?

We discuss this issue in Section 4.2.1 and in Section 6. We "artificially" decrease the error covariances of the few time-domain points to increase their impact on the parameter estimates. We clearly spell out the consequences this action has on the resulting uncertainties of posterior estimates. A related issues was brought up by Reviewer 1, who also suggested an alternative. We explained in our response to Reviewer 1 that we do not anticipate that the suggested alternative will lead to improved results.

Please, discuss the geomagnetic power spectra and the respective models in some more detail. How do they incorporate them in the Bayesian approach?

This issue was also brought up by Reviewer 1. The power spectral densities are computed from the time series (Sint-2000, PADM2M, CALS10k.2) using the multi-taper technique of Constable and Johnson (2005). Sections 4.2.2 and 4.2.3 describe how the PSDs are incorporated in the Bayesian approach (via feature-based likelihoods).

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. Our goal is to estimate the model's parameters using diverse and independent data sources that had previously been treated separately, so that the model is a valid representation of an expanded paleomagnetic record on kyr to Myr time scales. We formulate the estimation problem within the Bayesian framework and define a feature-based posterior distribution that describes probabilities of model parameters given a set of features, derived from the data. Numerically, we use Markov chain Monte Carlo (MCMC) to obtain a sample-based representation of the posterior distribution. The Bayesian problem formulation and its MCMC solution allow us to study the model's limitations and remaining posterior uncertainties. Our approach thus results in a reliable stochastic model for selected aspects of the long term behavior of the geomagnetic dipole field whose limitations and errors are well understood. We believe that such a model is useful for hypothesis testing and give a few examples of how the model can be used in this context. Another important aspect of our overall approach is that it can reveal inconsistencies between model and data, or within the various data sets. Identifying these shortcomings is a first and necessary step towards building more sophisticated models or towards resolving inconsistencies within the data.

1 Introduction

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 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).

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 and more tractable mean-field models require questionable parameterizations. 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). Other models have been derived by Rikitake (1958); Pétrélis and Fauve (2008).

Following Schmitt et al. (2001); Hoyng et al. (2001, 2002), we consider a stochastic differential equation (SDE) model for Earth's axial dipole. The basic idea is to model Earth's dipole field analogous to the motion of a particle in a double well potential. Time variations of the dipole field and dipole reversals then occur 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. The pushes and pulls by the noise process lead to variations of the dipole field around a typical 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 drift and diffusion coefficients 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. 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 account directly for the effects of sedimentation. Buffett and Puranam (2017) try to mimic the effects of sedimentation by sending the solution of the SDE through a low-pass filter. With this extension, the B13 model is more suitable to be compared to the data record of Earth's dipole field on a Myr time-scale.

A basic assumption of an SDE model is 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 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) and by 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 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 (Ziegler et al., 2011; Valet et al., 2005);

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

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

The approach ultimately leads to a family of SDE models, valid over Myr and kyr time scales, whose parameters are informed by a comprehensive paleomagnetic record composed of the above three sources of data. The results we obtain here are thus markedly different from previous work were data at 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). Priors are often assumed to be "uninformative", i.e., only conservative bounds for all parameters are known, and likelihoods describe point-wise model-data mismatch. Such likelihoods, however, are difficult to use when a variety of diverse data sources are to be combined. For example, a point-wise mismatch of model and data is difficult to enforce when two different data sets report two different values for the same quantity (see, e.g., Figure 1). Assumed error models in the data can control the effects each data set has on the parameter estimates. Since error models describe what "*we do not know*", good error models are notoriously difficult to come by. In this context, we discover that the "shortness of the paleomagnetic record," i.e., the limited amount of data available, is the main source of uncertainty. For example, PADM2M or Sint-2000 provide a time series of 2,000 consecutive "data points" (2 Myr sampled once per kyr). Errors in power spectral densities, computed from such a short time series, dominate the expected errors in these data. Similarly, errors in the reversal rate statistics are likely dominated by the fact that only a small number of reversals, e.g., those that occurred over the past 30 Myr, are useful for computing reversal rates. Reliable error models should thus reflect errors due to the shortness of the paleomagnetic record, rather than building error models on assumed errors in the data.

To address these issues we 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 unifying contributions from several independent data sources in a well-defined sense even if the various data may not be entirely self-consistent and further allows us to construct error models that reflect uncertainties induced by the shortness of the paleomagnetic record. In addition, we perform a suite of numerical experiments to check, in hind-sight, our a priori assumptions about the error models.

2 Description of the data

Variations in the virtual axial dipole moment (VADM) over the past 2 Myr can be derived from stacks of marine sediment.Two different compilations are considered in this study: Sint-2000, (Valet et al., 2005) and PAMD2M (Ziegler et al., 2011). Both

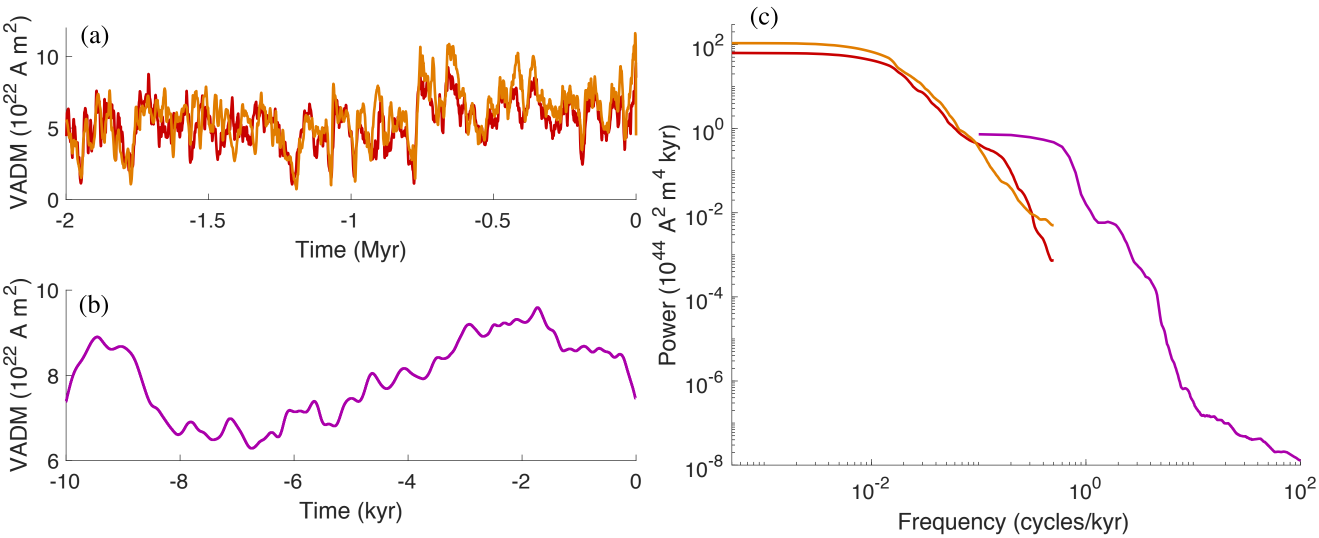

Figure 1. Data used in this paper. (a) Sint-2000 (orange) and PADM2M (red): VADM as a function of time over the past 2 Myr. (b) CALS10k.2: VADM as a function of time over the past 10 kyr. (c) Power spectral densities of the data in (a) and (b), computed by the multi-taper spectral estimation technique of Constable and Johnson (2005). Orange: Sint-2000. Red: PADM2M. Purple: CALS10k.2

of these data sets are sampled every 1 kyr and, thus, provide a time series of 2,000 consecutive VADM values. The PADM2M and Sint-2000 data sets are shown in the upper left panel of Figure 1. The CALS10k.2 data set, plotted in the lower left panel of Figure 1, describes variations of VADM over the past 10 kyr (Constable et al., 2016). The time dependence of CALS10k2 is represented using B-splines, so that the model can be sampled at arbitrary time intervals. We sample CALS10k.2 at an interval of one year, although the resolution of CALS10k.2 is nominally 100 years (Constable et al., 2016).

Below we use features derived from power spectral densities (PSD) of the Sint-2000, PADM2M and CALS10k.2 data. The PSDs are computed by the multi-taper spectral estimation technique of Constable and Johnson (2005). A restricted range of frequencies are retained in the estimation to account for data resolution and other complications (see below). We show the resulting PSDs of the three data sets in the right panel of Figure 1. During parameter estimation we further make use of the time average VADM and the standard deviation of VADM over time of the Sint-2000 and PADM2M data sets, listed in Table 1.

Data source	Time avg. VADM (10^{22} Am^2)	Std. dev. of VADM (10^{22} 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.

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 Myr $-$ 5 Myr, 5 Myr $-$ 10 Myr, $\ldots$, 25 Myr $-$ 30 Myr. This leads to the average reversal rate and standard deviation listed in Table 2. Increasing the 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 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 the PSDs from CALS10k.2 and those from the lower resolution data sets (SINT-2000 and PADM2M) at the overlapping frequencies. Dating uncertainties, smoothing due to sedimentary processes and the finite duration of the records all contribute to these discrepancies. We do not attempt to identify the source of these discrepancies. Instead, we seek to recover parameter values for a stochastic model by combining a feature-based approach with realistic estimates of the data uncertainty (see Section 4).

We further note that the amount of data is rather limited: we have 2 Myrs of VADM sampled at 1/ kyr, 10 kyr of high frequency VADM and use a 30 Myr record to compute reversal rates. The limited amount of data directly affects how the accuracy of the data should be interpreted. As an example, the mean and standard deviation of the reversal rate, based on a 30 Myr record may not be accurate; errors in the PSDs of PADM2M, Sint-2000 or CALS10k.2 are dominated by the fact that these are computed from relatively short time series. We address these issues by using the feature-based approach that allows us to build error models that reflect uncertainties due to the shortness of the paleomagnetic record. We further perform extensive numerical tests that allow us to check, in hind-sight, the validity our assumptions about errors (see Section 6).

3 Description of the model

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, e.g., Schmitt et al. (2001); Hoyng et al. (2001, 2002); 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 the following 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., $D(x) = D$. Modest variations in D have been reported on the basis of geodynamo simulations (Buffett and Matsui, 2015; Meduri and Wicht, 2016). Representative variations in D, however, have a small influence on the statistical properties of solutions of the SDE (1), see Buffett and Matsui (2015).

The function v is called the "drift" and is derived from a double-well potential, $U'(x) = -v(x)$. Here, we consider drift coefficients of the form

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

where $\bar{x}$ and γ are parameters. The parameter $\bar{x}$ defines where the drift coefficient vanishes and also corresponds to the time average of the associated linear model

$$\mathrm{d}x^l = -\gamma(x^l - \bar{x})\mathrm{d}t + \sqrt{D}\mathrm{d}W, \tag{3}$$

which is obtained by Taylor expanding $v(x)$ at $\bar{x}$. It is now clear that the parameter γ defines a relaxation time.

Nominal values of the parameters $\bar{x}$, γ and D are listed in Table 3. 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.

	$\bar{x}$ (10^{22} Am^2)	D (10^{44} A^{2}m^4 kyr^{-1})	γ (kyr^{-1})	T_s (kyr)	a (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

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

$$x_k = f(x_{k-1}, \Delta t) + \sqrt{2D\Delta t}\, w_k, \quad w_k \sim \mathcal{N}(0,1), \text{ iid} \tag{4}$$

where Δt is the time step, $\sqrt{\Delta t}\, w_k$ is the discretization of Brownian motion W in (1) and where $f(x_{k-1}, \Delta t)$ is the RK4 step. Here, iid stands for "independent and identically distributed", i.e., each random variable w_k, for $k > 0$, has the same Gaussian probability distribution, $\mathcal{N}(0,1)$, and w_i and w_j are independent for all $i \neq j$. 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 (4).

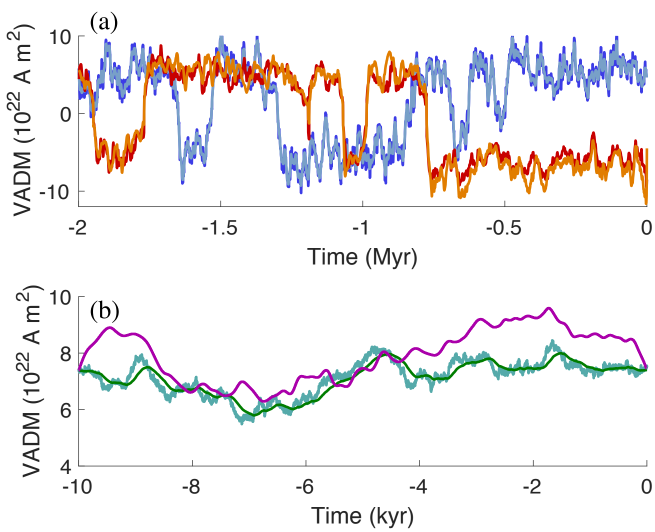

Figure 2. Simulations with nominal parameter values and data on the Myr and kyr scales. (a) VADM as a function of time on the Myr time scale. The output of the Myr model, x_j^{Myr}, is shown in dark blue (often hidden). The smoothed output, $x_j^{\text{Myr,s}}$, is shown in light blue. The signed Sint-2000 and PADM2M are shown in orange and red with signs (reversal timings) taken from Cande and Kent (1995). (b) VADM as a function of time on the kyr time scale. The output of the kyr-model with uncorrelated noise is shown in turquoise. The output of the kyr-model with correlated noise is shown in green. VADM of CALS10k.2 is shown in purple.

3.1 Models for the Myr amd kyr time scale

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 source of paleomagnetic data in Sint-2000 and PADM2M are affected by gradual acquisition of magnetization due to sedimentation processes, which amount 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 effects of sedimentation in the model by convolving the solution of (4) by a Gaussian filter

$$g(t) = \sqrt{\frac{6}{\pi T_s^2}} \cdot \exp\left(-\frac{6t^2}{T_s^2}\right), \tag{5}$$

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 whose state is denoted by $x^{\text{Myr,s}}$. Simulations with the “Myr model” and the nominal parameters of Table 3 are shown in Figure 2(a) where we plot the model output x_j^{Myr} in dark blue and the smoothed model output, $x_j^{\text{Myr,s}}$, in a lighter blue over a period of 2 Myr.

On a Myr scale, the assumption that the noise is uncorrelated in time is reasonable because one focuses on low frequencies and large sample intervals of the dipole, as in Sint-2000 and PADM2M, whose sampling interval is $1/$kyr. On a shorter time scale, as in CALS10k.2, 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 (4) 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

$$y_k = (1 - a\Delta t)y_{k-1} + \sqrt{2a\Delta t}\,w_k, \quad w_k \sim \mathcal{N}(0,1), \text{ iid} \tag{6}$$

$$x_k = f(x_{k-1}, \Delta t) + \sqrt{Da}\,\Delta t\,y_k, \tag{7}$$

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 2(b) along with the CALS10k.2 data.

3.2 Approximate power spectral densities

Accurate computation of the power spectral density (PSD) from the time-domain solution of the B13 model requires extremely long simulations. For example, the PSDs of two (independent) 1 billion year simulations with the Myr model are still surprisingly different. In fact, errors that arise due to "short" simulations substantially outweigh errors due to linearization. Recall that the PSD of the linear model (3) is easily calculated to be

$$\hat{x}^l(f) = \frac{2D}{\gamma^2 + 4\pi^2 f^2}, \tag{8}$$

where f is the frequency (in $1/\text{kyr}$). Since the Fourier transform of the Gaussian filter is known analytically, the PSD of the smoothed linear model is also easy to calculate:

$$\hat{x}^{l,s}(f) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \exp\left(-\frac{4\pi^2 f^2 T_s^2}{12}\right). \tag{9}$$

Similarly, an analytic expression for the PSD of the kyr-model with correlated noise in equations (6)-(7) can be obtained by taking the limit of continuous time ($\Delta t \to 0$):

$$\hat{x}^{l,kyr}(f) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \frac{a^2}{a^2 + 4\pi^2 f^2}. \tag{10}$$

Here, the first term is as in (8) and the second term appears because of the correlated noise.

Figure 3 illustrates a comparison of the PSDs obtained from simulations of the nonlinear models and their linear approximations. Specifically, the PSDs of the (smoothed) Myr scale nonlinear model, computed from a 50 Myr simulation, are shown in comparison to the approximate PSDs in equations (8)-(9). Note that the PSD of the smoothed model output, $x^{\text{Myr,s}}$, 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). The PSD of the kyr model with correlated noise, computed from a 10 kyr simulation, is also shown in Figure 3 in comparison with the linear PSD in equation (10). The good agreement between the theoretical spectra of the linear models and the spectra of the nonlinear models justifies the use of the linear approximation. We have further noted in numerical experiments that the agreement between the nonlinear and linear spectra increases with increasing simulation time, however a "perfect" match requires extremely long simulations of the nonlinear model (hundreds

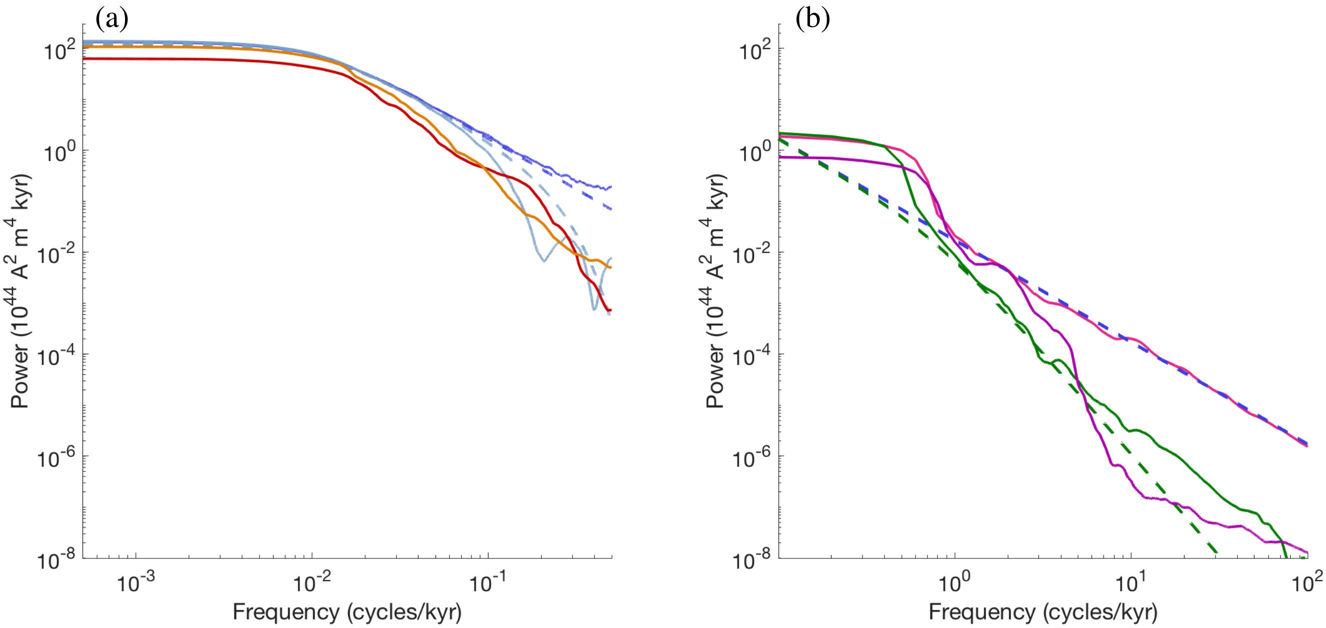

Figure 3. Power spectral densities of the model with nominal parameter values. (a) *Myr-model*: A PSD of a 50 Myr simulation with the Myr model is shown as a solid dark blue line. The corresponding theoretical PSD of the linear model is shown as a dashed dark blue line. A PSD of a 50 Myr simulation with the Myr model and high-frequency roll-off is shown as a solid light blue line. The corresponding theoretical PSD of the linear model is shown as a dashed light blue line. The PDSs of Sint-2000 and PADM2M are shown in orange and red. (b) *kyr model*: A PSD of a 10 kyr simulation of the kyr model with uncorrelated noise is shown as a solid pink line. The corresponding theoretical spectrum of the linear model is shown as a dashed blue line. A PSD of a 10 kyr simulation of the kyr model with correlated noise is shown as a solid green line. The corresponding theoretical spectrum is shown as a dashed green line. The PDSs of CALS10k.2 is shown in purple. All PSDs are computed by the multi-taper spectral estimation technique of Constable and Johnson (2005).

of billions of years). The approximate PSDs, based on the linear models, will prove useful in the construction of likelihoods in Section 4.2.

In addition to a good match of the PSDs of the nonlinear and linear models, we note that the PSDs of the model match, at least to some extent, the PSDs of the data (Sint-2000, PADM2M and CALS10k.2). This means that our choice for the nominal values is “reasonable” because this choice leads to a reasonable fit to the data. The goal of using a Bayesian approach to parameter estimation, described in Section 4, is to improve this fit and to equip the (nominal) parameter values with an error estimate, i.e., to define and compute a distribution over the model parameters. This will lead to an improved fit, along with an improved understanding of model uncertainties.

3.3 Approximate reversal rate, VADM time average and VADM standard deviation

The nonlinear SDE model (1) and its discretization (4) exhibit reversals, i.e., a 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 average strength of the dipole field. In principle, these quantities (reversal rate, time average and standard deviation) can be computed from simulations of the Myr and kyr model in the time domain. Similarly to what we found in the context of PSD computations and approximations, we find that estimates of the reversal rate, time average and standard deviation are subject to large errors unless the simulation time is very long (hundreds of millions of years). Using the linear model and Kramers formula, however, one can approximate the time average, reversal rate and standard deviation without simulating the nonlinear model (see below). Computing the approximate values is instantaneous (evaluation of simple formulas) and the approximations are comparable to what we obtain from very long simulations with the nonlinear model. As is the case with the PSD approximations based on linear models, the below approximations of the reversal rate, time average and standard deviation based will prove useful for formulating likelihoods in Section 4.2.

Specifically, the parameter $\bar{x}$ defines the time average of the linear model (3) and it also defines where the drift term (2) vanishes. These values coincide quite closely with the time average of the nonlinear model which suggests the approximation

$$E(x) \approx \bar{x} \cdot 10^{22} \text{ Am}^2. \tag{11}$$

The reversal rate can approximated by Kramers formula (Buffett and Puranam, 2017; Risken, 1996)

$$r \approx \frac{\gamma}{2\pi} \cdot \exp\left(-\frac{\gamma \bar{x}^2}{6D}\right) \cdot 10^3 \text{ Myr}^{-1}. \tag{12}$$

The standard deviation is the square root of the area under the PSD. Using the linear model that incorporates the effects of smoothing (due to sedimentation), one can approximate the standard deviation by computing the integral of the PSD in equation (9):

$$\sigma \approx \left(\frac{D}{\gamma} \exp\left(\frac{(\gamma T_s)^2}{12}\right) \operatorname{erfc}\left(\frac{\gamma T_s}{2\sqrt{3}}\right)\right)^{1/2} \cdot 10^{22} \text{ Am}^2, \tag{13}$$

where $\operatorname{erfc}(\cdot)$ is the (Gauss) error function. Without incorporating the smoothing, the standard deviation based on the linear model would be the integral of the PDS in (8), which is $\sigma \approx \sqrt{D/\gamma}$. The exponential and error function terms in (13) can thus be interpreted as a correction factor that accounts for the effects of sedimentation. It is easy to check that this correction factor is always smaller than one, i.e., the (approximate) standard deviation accounting for sedimentation effects is smaller than the (approximate) standard deviation that does not account for these effects.

For the nominal parameter values in Table 3 we calculate a time average of $\bar{x} = 5.23 \cdot 10^{22}$ Am^2, a standard deviation of $\sigma \approx 2.07 \cdot 10^{22}$ 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. Similar to what we observed of the model-data fit in terms of (approximate) PSDs, we find that the nominal parameter values lead to a "reasonable" fit of the model's reversal rate, time average and standard deviation. The Bayesian parameter estimation in Section 4 will improve this fit and lead to a better understanding of model uncertainties.

3.4 Parameter bounds

The Bayesian parameter estimation, described in Section 4, makes use of "prior" information about the model parameters. We formulate prior information in terms of parameter bounds and construct uniform prior distributions with these bounds. 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 (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 dipole may relax on timescales shorter than the slowest decay mode and a recent theoretical calculation (Pourovskii et al., 2017) suggests that the magnetic diffusivity may be slightly larger than $0.8\text{m}^2/\text{s}$. Both of these changes would cause the lower bound for γ to increase. 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 $\gamma \leq 0.7\ \text{kyr}^{-1}$.

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

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, $\bar{x} \in [0, 10] \cdot 10^{22}$ 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 3.

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 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 shorthand notation for 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

$$p(\theta|y) \propto p_0(\theta) p_l(y|\theta). \tag{14}$$

The posterior distribution combines the prior information with the 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 model of the Earth's dipole in view of the data we use.

On the other hand, the posterior distribution depends on several assumptions: since we define the prior and likelihood, we also implicitly define the posterior distribution. In particular, formulations of the likelihood require that one be able to describe anticipated errors in the data as well as anticipated model error. Such error models are difficult to come by in general, but even more so when the amount of data is limited. We address this issue by first formulating "reasonable" error models, followed by a set of numerical tests that confirm (or disprove) our choices of error models (see Section 6). In our formulation of error models, we focus on errors that arise due to the shortness of the paleomagnetic record because these errors dominate.

We solve the Bayesian parameter estimation problem numerically by using a Markov chain Monte Carlo (MCMC) method. An MCMC method generates a (Markov) chain of parameter values whose stationary distribution is the posterior distribution. The chain is constructed by proposing a new parameter vector and then accepting or rejecting this proposal with a specified probability that takes the posterior probability of the proposed parameter vector into account. A numerical solution via MCMC thus requires that the likelihood be evaluated for every proposed parameter vector. Below we formulate a likelihood that involves computing the PSDs of the Myr and kyr model, as well as reversal rates, time averages and standard deviations. As explained in Section 3, obtaining these quantities from simulations with the nonlinear models requires extremely long simulations. Long simulations, however, require more substantial computations. This perhaps would not be an issue if we were to compute the PDSs, reversal rate and other quantities once, but the MCMC approach we take requires repeated computing. For example, we consider Markov chains of length 10^6, which requires 10^6 computations of PSDs, reversal rates etc. Moreover, we will repeat these computations in a variety of settings to assess the validity of our error models (see Section 6). To keep the computations feasible (fast), we thus decided to use the approximations of the PSDs, reversal rate, time average and standard deviation, based on the linear models (see Section 3), to define the likelihood. Evaluation of the likelihood is then instantaneous because simulations with the nonlinear models are replaced by formulas that are simple to evaluate. Using the approximation is further justified by the fact that the approximate PDSs, reversal rates, time averages and standard are comparable to what we obtain from very long simulations with the nonlinear model.

4.1 Prior distribution

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 parameters 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", 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 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 parameters.

4.2 Feature-based likelihoods

In "data assimilation" or, more generally, in Bayesian estimation, information from a mathematical or numerical model is merged with information from data (sometimes called "observations") by combining a prior distribution with a likelihood to define a posterior distribution. The likelihood describes the probability of the data given the numerical model and its parameters and, therefore, connects the model and the data. Likelihoods are often based on a point-wise mismatch of the model outputs and the data.

We wish to use to use a collection of paleomagnetic observations 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). The various data sets are not consistent and, for example, Sint-2000 and PADM2M report different VADM values at the same time instant (see Figure 1). Likelihoods that are defined in terms of a point-wise mismatch of model and data balance the effects of each data set via (assumed) error covariances: the data set with smaller error covariances has a stronger effect on the parameter estimates. Accurate error models, however, are hard to come by. For this reason, we use an alternative approach called "feature-based data assimilation" (see Morzfeld et al. (2018); Maclean et al. (2017)). The idea is to extract "features" from the data and to subsequently define likelihoods that are based on the mismatch of the features derived from the data and the model. Below, we formulate features that account for discrepancies across the various data sources and derive error models for the features. The error models are built to reflect uncertainties that arise due to the shortness of the paleomagnetic record. The resulting feature-based posterior distribution describes the probability of model parameters in view of the features. Thus, model parameters with a large feature-based posterior probability lead to model features that are comparable to the features derived from the data, within the assumed uncertainties due to the shortness of the paleomagnetic record.

Specifically, we define likelihoods based on features derived from PSDs of the Sint-2000, PADM2M and CALS10k.2 data sets, as well as the reversal rate, time average VADM and VADM standard deviation. 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);

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

In the Bayesian approach, and assuming that errors are independent, this means that the likelihood, $p_l(y|\theta)$ in equation (14) can be written as the product of three terms

$$p_l(y|\theta) \propto p_{l,\text{td}}(y|\theta)\, p_{l,\text{lf}}(y|\theta) p_{l,\text{hf}}(y|\theta), \tag{15}$$

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

We define the likelihood component of the time domain data based on the equations

$$y_{\text{rr}} = h_{\text{rr}}(\theta) + \varepsilon_{\text{rr}}, \tag{16}$$

$$y_{\bar{x}} = h_{\bar{x}}(\theta) + \varepsilon_{\bar{x}}, \tag{17}$$

$$y_{\sigma} = h_{\sigma}(\theta) + \varepsilon_{\sigma}, \tag{18}$$

where y_{rr}, $y_{\bar{x}}$ y_{σ} are features derived from the time domain data, $h_{\text{rr}}(\theta)$, $h_{\bar{x}}(\theta)$ and $h_{\sigma}(\theta)$ are functions that connect the model parameters to the features, based on the approximations described in Section 3.3, and where ε_{rr}, $\varepsilon_{\bar{x}}$ and ε_{σ} are independent Gaussian error models with mean zero and variances σ_{rr}^2, $\sigma_{\bar{x}}^2$, and σ_{σ}^2. Taken all together, the likelihood term $p_{l,\text{td}}(y|\theta)$ in (15) is then given by the product of the three likelihoods defined by equations (16), (17) and (18):

$$p_{l,\text{td}}(y|\theta) \propto \exp\left(-\frac{1}{2}\left(\left(\frac{y_{\text{rr}} - h_{\text{rr}}(\theta)}{\sigma_{\text{rr}}}\right)^2 + \left(\frac{y_{\bar{x}} - h_{\bar{x}}(\theta)}{\sigma_{\bar{x}}}\right)^2 + \left(\frac{y_{\sigma} - h_{\sigma}(\theta)}{\sigma_{\sigma}}\right)^2\right)\right). \tag{19}$$

The reversal rate feature is simply the average reversal rate we computed from the chronology of Ogg (2012) (see Section 2), i.e., $y_{\text{rr}} = 4.23$ reversals/Myr. The function $h_{\text{rr}}(\theta)$ is based on the approximation using Kramers formula in equation (20):

$$h_{\text{rr}}(\theta) = \frac{\gamma}{2\pi} \exp\left(-\frac{\gamma \bar{x}^2}{6D}\right) \cdot 10^3 \text{ reversals/Myr.} \tag{20}$$

The time average feature is the mean of the time averages of PADM2M and Sint-2000: $y_{\bar{x}} = 5.56 \cdot 10^{22}$ Am^2. The function $h_{\bar{x}}(\theta)$ is based on the linear approximation discussed in Section 3.3, i.e., $h_{\bar{x}}(\theta) = \bar{x}$. The feature for the VADM standard deviation is the average of the VADM standard deviations of PADM2M and Sint-2000: $y_\sigma = 1.66 \cdot 10^{22}$ Am^2. The function $h_\sigma(\theta)$ uses the linear approximation of the standard deviation (13):

$$h_\sigma(\theta) = \left(\frac{D}{\gamma} \exp\left(\frac{(\gamma T_s)^2}{12} \right) \text{erfc}\left(\gamma T_s/2/\sqrt{(3)} \right) \right)^{1/2} \cdot 10^{22} \text{ Am}^2 \text{ kyr}^2 \quad (21)$$

Candidate values for these error variances are as follows. The error variance of the reversal rate, σ_{rr}^2, can be based on the standard deviations we computed from the Ogg (2012) chronology in Table 2. Thus, we might use the standard deviation of the 10-Myr average and take $\sigma_{\text{rr}} = 0.5$. One can also use the model with nominal parameter values (see Table 3) to compute candidate values of the standard deviation σ_{rr}. We perform 1,000 independent 10 Myr simulations and, for each simulation, determine the reversal rate. The standard deviation of the reversal rate based on these simulations is 0.69 reversal per Myr, which is comparable to the 0.5 reversals per Myr we computed from the Ogg (2012) chronology, using an interval length of 10 Myr. Similarly, the standard deviation of the reversal rate of 1,000 independent 5 Myr simulations is 0.97, which is also comparable to the standard deviation of 1.01 reversals per Myr, suggested by the Ogg (2012) chronology, using an interval length of 5 Myr.

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. We can also derive error covariances using the model with nominal parameters and perform 1,000 independent 2 Myr simulations. For each simulation, we compute the time average and the VADM standard deviations, which then allows us to compute standard deviations of these quantities. Specifically, we found a value of $0.26 \cdot 10^{22}$ Am^2 for the standard deviation of the time average and $0.11 \cdot 10^{22}$ Am^2 for standard deviation of the VADM standard deviation. These values are comparable to what we obtained from the data, especially if we base the standard deviations on *half* of the difference of the PADM2M and Sint-2000 values, i.e., assuming that the data sets are within two standard deviations (rather than within one, which we assumed above).

A difficulty with these error covariances is that we have few time domain observations compared with the large number of spectral data in the power spectra (see below). This vast difference in the number of time domain and spectral data means that the spectral data can 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, derived from the data by a factor of 100 and set

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

We discuss this choice and its consequences on parameter estimates and associated uncertainty in more detail in Section 6. An alternative approach would be to reduce the number of spectral data, e.g., by condensing the spectra into “a few” characteristics such as corner frequencies or slopes, see also (Bärenzung et al., 2018). The difficulty with such an approach, however, is that corner frequencies etc. are not easy to extract from the PDSs of the data. For these reasons, we pursue in this paper a “tuning” of the error variances to control the impact of the time domain data on the parameter estimates, but alternative strategies are straightforward to implement within our overall feature-based Bayesian estimation approach.

4.2.2 Low frequencies

The component $p_{l,\text{lf}}(y|\theta)$ of the feature-based likelihood (15) 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

$$y_{\text{lf}} = h_{\text{lf}}(\theta) + \varepsilon_{\text{lf}}, \tag{23}$$

where y_{lf} is a feature that represents the PSD of the Earth's dipole field at low frequencies, $h_{\text{lf}}(\theta)$ maps the model parameters to the data y_{lf} and where ε_{lf} represents the errors we expect.

We define y_{lf} to be the mean of the PSDs of Sint-2000 and PADM2M. The function $h_{\text{lf}}(\theta)$ maps the model parameters to the feature y_{lf} and is based on the PSD of the linear model (3). To account for the smoothing introduced by sedimentation processes we define $h_{\text{lf}}(\theta)$ to be a function that computes the PSD of the Myr model by using the "un-smoothed" spectrum of equation (8) for frequencies less than 0.05 cycles/kyr, and uses the "smoothed" spectrum of equation (9) for frequencies between $0.05 - 0.5$ cycles/kyr:

$$h_{\text{lf}}(\theta) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \begin{cases} 1 & \text{if } f \leq 0.05 \\ \exp\left(-(4\pi^2 f^2 T_s^2)/12\right) & \text{if } 0.05 < f \leq 0.5 \end{cases} \tag{24}$$

Note that $h_{\text{lf}}(\theta)$ 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 uncertainty introduced by sampling the VADM once per kyr for only 2 Myrs is the dominant source of error in the power spectral densities. For a Gaussian error model ε_{lf} with zero mean, this means that the error covariance should describe uncertainties that are induced by the limited amount of data. We construct such a covariance as follows. We perform 10^4 simulations, each of 2 Myr, with the nonlinear Myr model (4) 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 $5 \cdot 10^3$ samples of ε_{lf} added to y_{lf} (grey). 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.

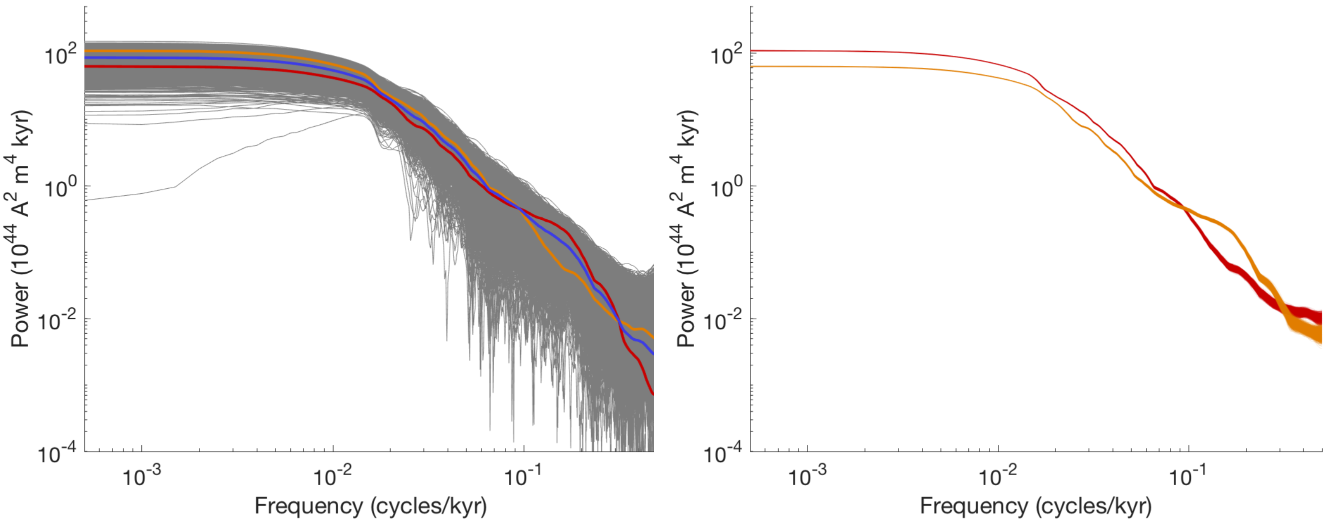

Figure 4. Left: low frequency data and error model due to shortness of record. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Blue: mean of PSDs of Sint-2000 and PADM2M (y_{lf}). Grey: $5 \cdot 10^3$ samples of the error model ε_{lf} added to y_{lf}. Right: error model based on errors in Sint-2000. Orange: 10^3 samples of the PSDs computed from “perturbed” Sint-2000 VADMs. Red: 10^3 samples of the PSDs computed from “perturbed” PADM2M VADMs.

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 $0.9 - 9.9$ cycles/kyr, where the upper limit is set by the resolution of the CALS10k.2 data. The lower limit is chosen 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 may be less reliable that the PSD of CALS10k.2 for these frequencies. As above, we construct the likelihood $p_{l,\text{hf}}(y|\theta)$ from an equation similar to (23):

$$y_{\text{hf}} = h_{\text{hf}}(\theta) + \varepsilon_{\text{hf}}, \tag{25}$$

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

We base $h_{\text{hf}}(\theta)$ on the PSD of the linear model (see equation (10)) and set

$$h_{\text{hf}}(\theta) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \frac{a^2}{a^2 + 4\pi^2 f^2}, \tag{26}$$

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 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 (6)-(7) with nominal parameter values. We

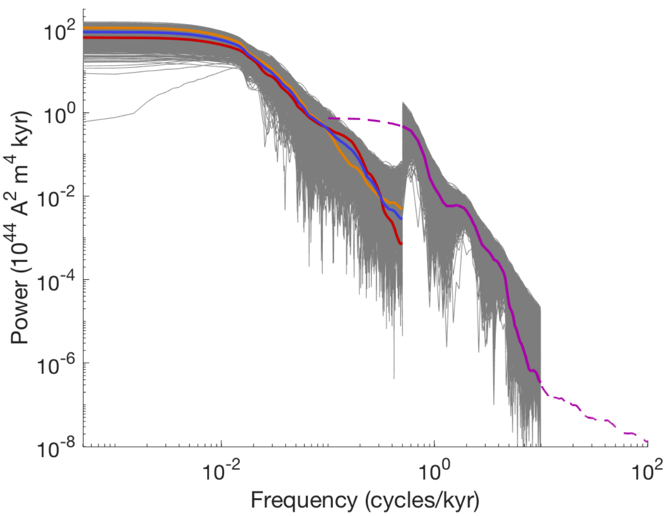

Figure 5. Data and error models for low and high frequencies. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Blue: mean of PSDs of Sint-2000 and PADM2M ($y_{\rm lf}$). Grey (low frequencies): $5 \cdot 10^3$ samples of the error model $\varepsilon_{\rm lf}$ added to $y_{\rm lf}$. Dashed purple: PSD of CALS10k.2. Solid purple: PSD of CALS10k.2 at frequencies we consider ($y_{\rm hf}$). Grey (high frequencies): $5 \cdot 10^3$ samples of the error model $\varepsilon_{\rm hf}$ added to $y_{\rm hf}$.

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 $\varepsilon_{\rm hf}$, which is illustrated along with the low frequency 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 (14).

4.3 Numerical solution by MCMC

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

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 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 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 samples, but the number of "effective" samples is $10^6/\text{IACT}$. For all MCMC runs we perform (see Sections 5 and 6), the IACT of the Markov chain is about 100. We discard the first $10 \cdot \text{IACT}$ 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 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) MCMC sampler to generate a long chain (10^6 samples). Thus, the precise choice of MCMC sampler is not so important for our purposes. We found that the MCMC Hammer solves the problem with sufficient efficiency for our purposes.

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 that are missing some of the factors that comprise the overall feature-based likelihood (15). This allows us to study the impact of 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 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) which define the "uninformative", uniform prior. Thus, the posterior probability, which synthesizes the information from the data via the definition of the features, is concentrated over a smaller subset of parameters than the prior probability. In this way, the Bayesian parameter estimation has sharpened the knowledge about the parameters by incorporating the data.

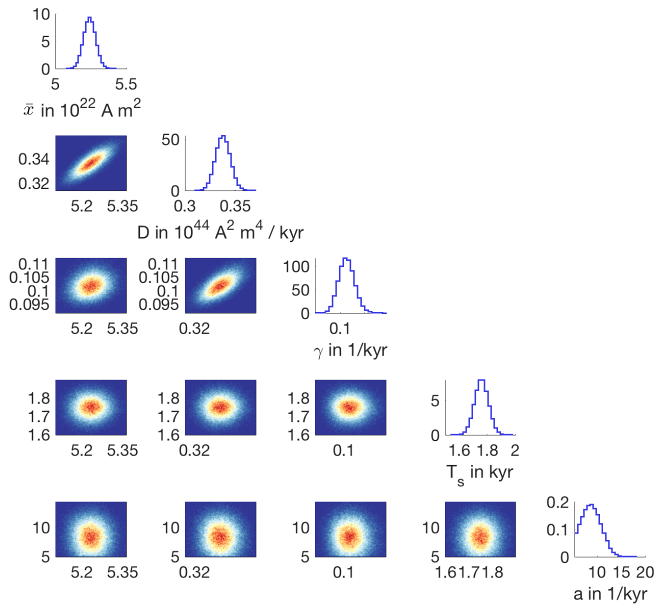

Figure 6. 1- and 2-dimensional histograms of the posterior distribution.

$\bar{x}$ in 10^{22} Am^2	D in 10^{44} A^{2}m^4 kyr^{-1}	γ in kyr^{-1}	T_s in kyr	a in kyr^{-1}	σ in 10^{22} Am^2	Rev. rate in reversals/Myr
5.23 (0.043)	0.34 (0.0072)	0.10 (0.0033)	1.75 (0.050)	8.56 (1.93)	1.77 (0.024)	4.06 (0.049)

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

The 2-dimensional histograms indicate correlations among the parameters $\theta = (\bar{x}, D, \gamma, T_s, a)^T$, with strong correlations between $\bar{x}$, D and γ. These correlations can also be described by the correlation coefficients

	$\bar{x}$	D	γ	T_s	a
$\bar{x}$	1.00	0.78	0.20	0.02	−0.03
D	0.78	1.00	0.64	0.02	−0.03
γ	0.20	0.64	1.00	−0.01	−0.02
T_s	0.02	0.02	−0.01	1.00	0.00
a	−0.03	−0.03	−0.02	0.00	1.00

(27)

The strong correlation between $\bar{x}$ and D and γ, is due to the contribution of the reversal rate to the overall likelihood (see Equation (20)) and the dependence of the spectral data on D and γ (see Equations (9) and (10)). From the samples, we 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 2,000 samples of the posterior distribution followed by evaluation of Equations (20) and (13) 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 deviation (mean value of $1.77 \cdot 10^{22}$ Am^2) is also far, as measured by the posterior standard deviation, from the value we use as data (1.6610^{22} Am^2). These large deviations indicate an inconsistency between the VADM standard deviation and the reversal rate. A higher reversal rate could be achieved with a higher VADM standard deviation. The reason is that the reversal rate in Equation (20) can be re-written as

$$r \approx \frac{\gamma}{2\pi} \cdot \exp\left(-\frac{\bar{x}^2}{6\sigma^2}\right) \cdot 10^3 \text{ Myr}^{-1}, \tag{28}$$

using $\sigma \approx \sqrt{D/\gamma}$, i.e., neglecting the correction factor due to sedimentation, which has only a minor effect. Using a time average of $\bar{x} = 5.23 \cdot 10^{22}$ Am^2, a reversal rate $r = 4.2$ reversals / Myr, setting $\gamma = 0.1\,\text{kyr}^{-1}$ (posterior mean value), and solving for the VADM standard deviation results in $\sigma \approx 1.86 \cdot 10^{22}$ Am^2, which is not compatible with the SINT-2000 and PADM2M data sets (where $\sigma \approx 1.66 \cdot 10^{22}$ Am^2). One possible source of discrepancy is that the low-frequency data sets underestimate the standard deviation and also the time average. For example, Ziegler et al. (2008) report a time average VADM of $7.64 \cdot 10^{22}$ Am^2 and a standard deviation $\sigma = 2.72 \cdot 10^{22}$ Am^2 for paleointensity measurements from the past 0.55 Myr. These measurements are unable to provide any constraint on the temporal evolution of the VADM (in contrast to the SINT-2000 and PADM2M models). Instead, these measurements represent a sampling of the steady-state probability distribution for the dipole moment. The results thus suggest that a larger mean and standard deviation are permitted by paleointensity observations. Using the larger values for the time average and VADM standard deviation, but keeping $\gamma = 0.1\,\text{kyr}^{-1}$ (posterior mean value), leads to a reversal rate of $r \approx 4.27$ reversals per Myr, which is compatible with the reversal rates based on the past 30 Myr in Table 2. It is, however, also possible that the model for the reversal rate has shortcomings. Identifying these shortcomings is a first step in making model improvements and the Bayesian parameter estimation framework we describe is a mathematically and computationally sound tool for discovering such inconsistencies.

The model fit to the spectral data is illustrated in the left panel of Figure 7. Here, we plot 100 PSDs, 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. For comparison, the figure also shows the PADM2M, Sint-2000 and CALS10k.2 data as well as $5 \cdot 10^3$ realizations of the low- and high-frequency error models. We note that the overall uncertainty is reduced by the Bayesian parameter estimation. The reduction in uncertainty is apparent from the expected errors generating a “wider” cloud of PSDs (in grey) than the posterior estimates (in blue and turquoise). We further note that the PSDs of the models, with parameters drawn from the posterior distribution, fall largely within the expected errors (illustrated in grey). In particular the high-frequency PSDs (the CALS10k.2 range) are well within the errors we imposed by the likelihoods. The low frequencies of Sint-2000/PADM2M are also within the expected errors and so are the high-frequencies beyond the second roll-off due to the sedimentation effects. At intermediate frequencies, some of the PDSs of the model are outside of the expected errors. This indicates a model inconsistency because it

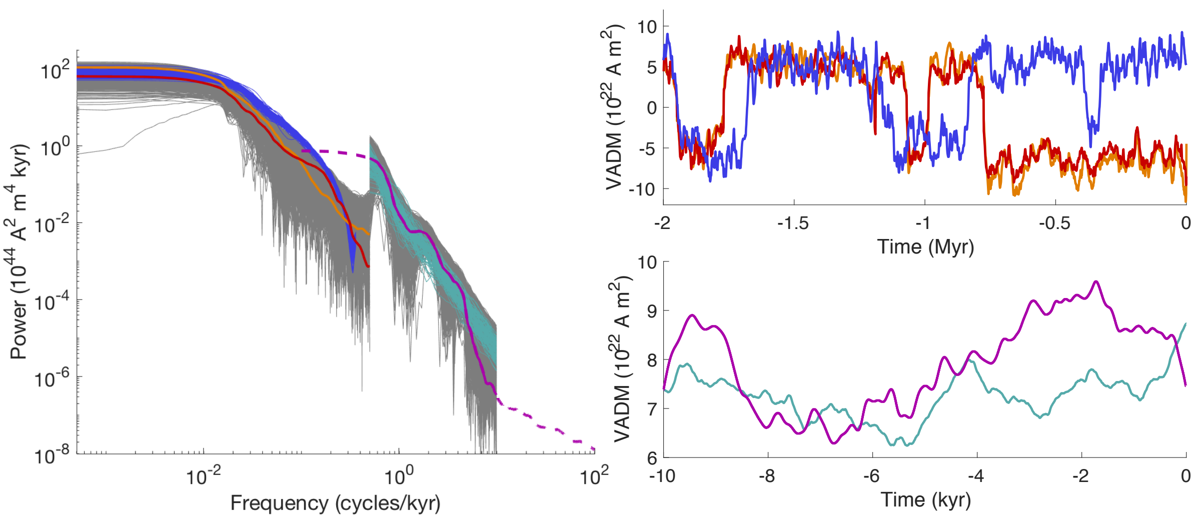

Figure 7. Parameter estimation results. Left: PSDs of data and model. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Dashed purple: PSD of CALS10k.2. Solid purple: PSD of CALS10k.2 at frequencies we consider (y_{hf}). Grey (low frequencies): $5 \cdot 10^3$ samples of the error model $\varepsilon_{\mathrm{lf}}$ added to y_{lf}. Grey (high frequencies): $5 \cdot 10^3$ samples of the error model $\varepsilon_{\mathrm{hf}}$ added to y_{hf}. Dark blue: PSDs of 100 posterior samples of Myr model (with smoothing). Turquoise: PSDs of 100 posterior samples of kyr model with uncorrelated noise. Right, top: Sint-2000 (orange), PADM2M (red) and a realization of the Myr model with smoothing and with posterior mean parameters (blue). Right, bottom: CALS10k.2 (purple) and a realization of the kyr model with correlated noise and with posterior mean parameters kyr model (turquoise).

is difficult to account for the intermediate frequencies with model parameters that fit the other data (spectral and time domain) within the assumed error models. We investigate this issue further in Section 6.

The right panels of Figure 7 show a Myr model run (top) and kyr model run (bottom) using the posterior mean values for the parameters. We note that the model with posterior mean parameter values exhibits qualitatively similar characteristics as the Sint-2000, PADM2M and CALS10k.2 data. The figure thus illustrates that the feature-based Bayesian parameter estimation, which is based solely on PSD, reversal rates, time average VADM and VADM standard deviation, translates into model parameters that also appear reasonable when a single simulation in the time domain is considered.

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 (as compared to the uniform prior). The posterior distribution describes a set of model parameters that yield model outputs that are comparable with the data in the feature-based 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 paleomagnetic 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

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

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 model with posterior parameters. As indicated above, this discrepancy could be due to 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 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.

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, $p_{l,\text{hf}}(y|\theta)$ in the feature-based likelihood (15). 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. The differences and similarities are apparent when we compare the corner plots of the posterior distributions of configurations (a), shown in Figure 6, and of configuration (b), shown in Figure 8. The corner plots are nearly identical except for the bottom row of plots which illustrates marginals of the posterior related to a. We note that the posterior distribution over a is nearly identical to its prior distribution. Thus, the parameter a is not constrained by the data used in configuration (b), which is perhaps not surprising because a only appears in the Bayesian parameter estimation problem via the high-frequency likelihood $p_{l,\text{hf}}(y|\theta)$. Moreover, since $p_{l,\text{lf}}(y|\theta)$ and $p_{l,\text{td}}(y|\theta)$ 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 is self-consistent and consistent with the

Configuration	(a)	(b)	(c)
$\bar{x}$ in 10^{22} Am^2	5.23 (0.043)	5.23 (0.042)	3.56 (0.26)
D in 10^{44} A^{2}m^4 kyr^{-1}	0.34 (0.0072)	0.34 (0.0072)	0.13 (0.014)
γ in kyr^{-1}	0.10 (0.0033)	0.10 (0.0033)	0.081 (0.0052)
T_s in kyr	1.75 (0.050)	1.74 (0.050)	1.68 (0.14)
a in kyr^{-1}	8.56 (1.93)	22.45 (10.01)	11.69 (3.91)
σ in 10^{22} Am^2	1.77 (0.024)	1.77 (0.023)	1.22 (0.060)
Rev. rate in reversals/Myr	4.06 (0.049)	4.06 (0.023)	3.34 (0.52)
Configuration	(d)	(e)	(f)
$\bar{x}$ in 10^{22} Am^2	5.04 (2.91)	5.56 (0.048)	5.04 (2.88)
D in 10^{44} A^{2}m^4 kyr^{-1}	0.094 (0.015)	0.44 (0.021)	0.093 (0.015)
γ in kyr^{-1}	0.078 (0.0063)	0.14 (0.014)	0.077 (0.0064)
T_s in kyr	1.64 (0.19)	2.98 (1.15)	1.64 (0.19)
a in kyr^{-1}	12.92 (4.79)	22.59 (10.08)	22.37 (10.13)
σ in 10^{22} Am^2	1.08 (0.074)	1.66 (0.036)	1.07 (0.075)
Rev. rate in reversals/Myr	2.89 (4.11)	4.23 (0.050)	2.83 (4.10)

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

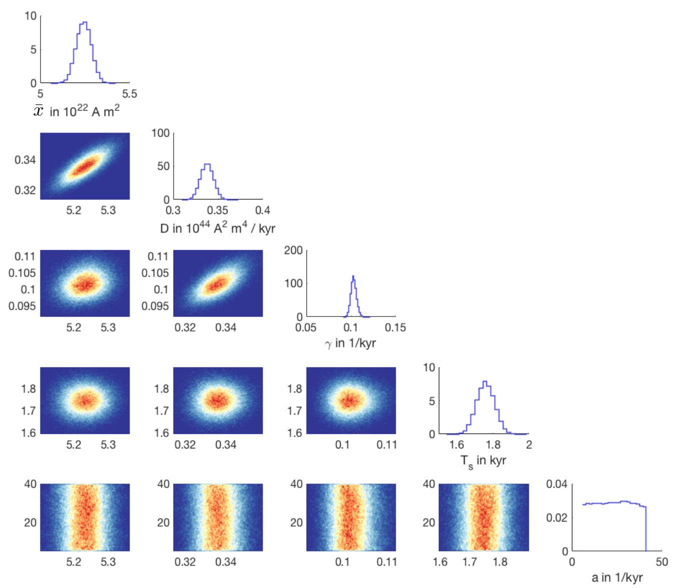

Figure 8. 1- and 2-dimensional histograms of the posterior distribution of configuration (b).

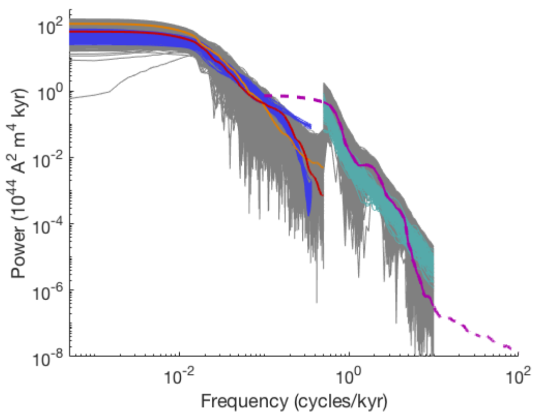

Figure 9. PSDs of data and model with parameters drawn from the posterior distribution of configuration (c). Orange: PSD of Sint-2000. Red: PSD of PADM2M. Dashed purple: PSD of CALS10k.2. Solid purple: PSD of CALS10k.2 at frequencies we consider ($y_{\rm hf}$). Grey (low frequencies): $5 \cdot 10^3$ samples of the error model $\varepsilon_{\rm lf}$ added to $y_{\rm lf}$. Grey (high frequencies): $5 \cdot 10^3$ samples of the error model $\varepsilon_{\rm hf}$ added to $y_{\rm hf}$. Dark blue: PSDs of 100 posterior samples of Myr model (with smoothing). Turquoise: PSDs of 100 posterior samples of kyr model with uncorrelated noise.

data on the Myr and kyr time scales; in the context of our simple stochastic model the data from CALS10k.2 mostly constraints 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 configuration (c), the spectral data are emphasized during the Bayesian estimation which also leads to an overall better fit of the spectra. This is illustrated in Figure 9, where we plot the 100 PSDs generated by 100 (independent) simulations with the model with parameters drawn from the posterior distribution of configuration (c). For comparison, we also plot the PSDs of PADM2M, Sint-2000, CALS10k.2 and $5 \cdot 10^3$ realizations of the high- and low-frequency error model. In contrast to configuration (a) (see Figure 7), we find that the PSDs of the model of configuration (c) are all well within the expected errors. One the other hand, the reversal rate drops to about 3 reversals/Myr, and the time average VADM and VADM standard deviation also decrease significantly as compared to configuration (a). This is caused by the posterior mean of D being decreased by more than 50%, while γ and T_s are comparable for configurations (a)-(c). The fact that the improved fit of the PDSs comes at the cost of a poor fit of the reversal rate, time average and standard deviation is another indication of an inconsistency between the reversal rate and the VADM data sets. As indicated above, one of the strengths of the Bayesian parameter estimation framework we describe here is to be able to identify such inconsistencies. Once identified, one can try to fix the model. For example, we can envision a modification of the functional form of the drift term in (2). A nearly linear dependence of the drift term on x near $x = \bar{x}$ is supported by the VADM data sets, but the behavior near $x = 0$ is largely unconstrained. Symmetry of the underlying governing equations suggests that the drift term should vanish and the functional form adopted in (2) is just one way that a linear trend can be extrapolated to $x = 0$. Other functional forms that lower the barrier between the potential wells would have the effect of increasing the reversal rate. This simple change to the model could bring the reversal rate into better agreement with the time average and standard deviation of the VADM data sets.

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), where the error variances of the time domain data are "large", and (d), where the error variance of the time domain data are "infinite". Thus, the impact of the time domain data is minimal if the error variances of the time domain data are large. 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. We further note that the parameter estimates of configurations (c) or (d) are quite different from the parameter estimates of configuration (a) (see above). For an overall good fit of the model to the spectral and time domain data, the error variances for the time domain data must by small, as in configuration (a). Otherwise, the reversal rates are too low. Small error variances, however, imply (relatively) large deviations between the time domain data and the model predictions. Small error variances also comes at the cost of not necessarily realistic posterior variances.

Comparing configurations (d) and (e), we note that if only the spectral data are used, the reversal rates are unrealistically small (nominally 1 reversal per Myr). 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 either the model has some inconsistencies, or that the reversal rate and the VADM standard deviation are not consistent. Specifically, our experiments suggest that a good match to spectral data requires a set of model parameters that is quite different from the set of model parameters that lead to a good fit to the reversal rate, time average VADM and VADM standard deviation. Experimenting with different functional forms for the drift term is one strategy for achieving better agreement between the reversal rate, the time average VADM and VADM standard deviation.

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 configurations (a) and (b) above). The results, shown in Table 6, indicate that the parameter estimates based on configurations (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 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.

Finally, note that the data used in configuration (d) does not inform the parameter $\bar{x}$, and configuration (f) does not inform $\bar{x}$ or a. If the data do not inform the parameters, then the posterior distribution over these parameters is essentially equal to the prior distribution, which is uniform. This is illustrated in Figure 10, where we show the corner plot of the posterior distribution of configuration (f). We can clearly identify the uniform prior in the marginals over the parameters $\bar{x}$ and a. This means that the Sint-2000 and PADM2M data only constrain the parameters D, γ and T_s.

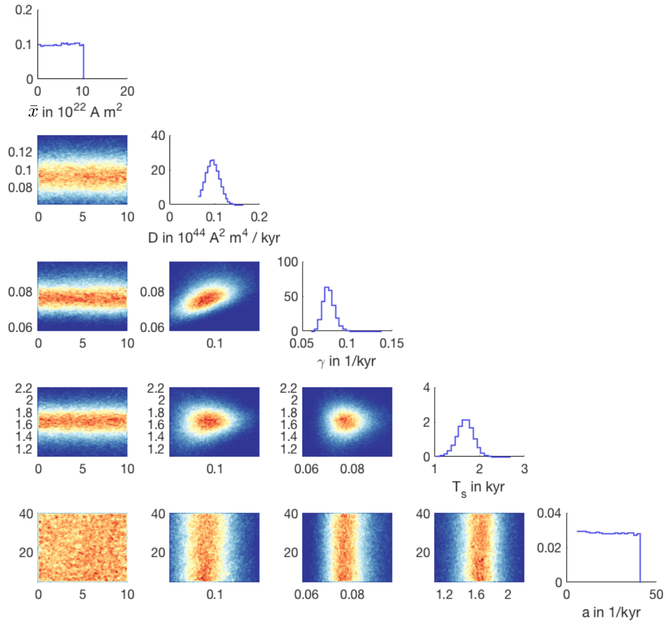

Figure 10. 1- and 2-dimensional histograms of the posterior distribution of configuration (f).

7 Examples of applications of the model

The Bayesian estimation technique we describe leads to a model with stochastic parameters whose distributions are informed by the paleomagnetic data. Moreover, we ran a large number of numerical experiments to understand the limitations of the model, to discover inconsistencies between the model and the data and to check our assumptions about error modeling. This process results in a well-understood and well-founded stochastic model for selected aspects of the long term behavior of the geomagnetic dipole field. We believe that such a model can be useful for a variety of purposes, including testing hypotheses about selected long-term aspects of the geomagnetic dipole.

For example, it was noted by Ziegler et al. (2011) that the VADM (time averaged) amplitude during the past chron was slightly lower than during the previous chron. Specifically, the time averaged VADM for $-0.78 < t < 0$ Myr is $E(x) = 6.2 \cdot 10^{22}$ Am^2 but for $-2 < t < -0.78$ Myr the time average is $E(x) = 4.8 \cdot 10^{22}$ Am^2. A natural question is: is this increase in the time average significant or is it due due to random variability? We investigate this question using the model whose parameters are the posterior mean values of configuration (a) (the configuration that leads to an overall good fit to all data). Specifically, we perform 10,000 simulations of duration 0.78 Myr and 10,000 independent simulations of duration 1.22 Myr. For each simulation, we compute the time average, which allows us to estimate the standard deviation of the difference in means (assuming no correlation between the two time intervals). We found that this standard deviation is about $0.46 \cdot 10^{22}$ Am^2, which is much smaller than the differences in VADM time averages of $1.4 \cdot 10^{22}$ Am^2. This suggests that the increase in time averaged VADM is likely not due to random variability.

A similar approach can be applied to the question of changes in the reversal rate over geological time. The observed reversal rate over the past 30 Myr is approximately 4.26 Myr^{-1}. When the record is divided into 10-Myr intervals, the reversal rate varies about the average, with a standard deviation of about 0.49 Myr^{-1} (see Table 2). These variations are within the expected fluctuations for the stochastic model. Specifically, we can use an ensemble of 10^5 simulations, each of duration 10 Myr to compute the average and standard deviation of the reversal rate The results, obtained by using nominal parameters and posterior mean parameters of configurations (a) (overall good fit to all data) and (e) (emphasis on reversal rate data) are shown in Table 7. As already indicated in Section 4.2.1, the standard deviation from the geomagnetic polarity time scale is comparable to the

	Nominal parameter values	Configuration (a)	Configuration (e)
Average reversal rate (in Myr^{-1})	4.50	3.76	3.56
Standard deviation (in Myr^{-1})	0.68	0.63	0.59

Table 7. Average reversal rate and and standard deviation of an ensemble of 10^5 simulations of duration10 Myr, filtered to a resolution of 30 kyr. The simulations are done with nominal parameter values (Table 3), or the posterior mean values of configurations (a) and (e).

standard deviation we compute via the model. The observed reversal rate for the 10-Myr interval between 30 and 40 Myr, however, is approximately 2.0 Myr^{-1}, which departs from the 0 – 30 Myr average by more than three standard deviations. This suggests that the reversal rate between 30 and 40 Ma cannot be explained by natural variability in the model. Instead, it suggests that model parameters were different before 30 Ma, implying that there was a change in the operation of the geodynamo.

8 Summary and conclusions

We consider parameter estimation for a model of Earth's axial magnetic dipole field. The idea is to estimate the model parameters using data that describe Earth's dipole field over kyr and Myr time scales. The resulting model, with calibrated parameters, is thus a representation of Earth's dipole field on these time scales. We formulated a Bayesian estimation problem in terms of "features" that we derived from the model and data. The data include two time series (Sint-2000 and PADM2M) that describe the strengths of Earth's dipole over the past 2 Myr, a shorter record (CALS10k.2) that describes dipole strength over the past 10 kyr, as well as reversal rates derived from the geomagnetic polarity time scale. The features are used to synthesize information from these data sources (that had previously been treated separately).

Formulating the Bayesian estimation problem requires defining anticipated model error. We found that the main source of uncertainty is the shortness of the paleomagnetic record and constructed error models to incorporate this uncertainty. Numerical solution of the feature-based estimation problem is done via conventional Markov chain Monte Carlo (an affine invariant ensemble sampler). With suitable error models, our numerical results indicate that the paleomagnetic data constrain all model parameters in the sense that the posterior probability mass is concentrated on a smaller subset of parameters than the prior probability. Moreover, the posterior parameter values yield model outputs that fit the data in a precise, feature-based sense, which also translates into a good fit by other, more intuitive measures.

A main advantage of our approach (Bayesian estimation with an MCMC solution) is that it allows us to understand the limitations and remaining (posterior) uncertainties of the model. After parameter estimation, we thus have produced a reliable, stochastic model for selected aspects of the long term behavior of the geomagnetic dipole field whose limitations and errors are well-understood. We believe that such a model is useful for hypothesis testing and have given several examples of how the model can be used in this context. Another important aspect of our overall approach is that it can reveal inconsistencies between model and data. For example, we ran a suite of numerical experiments to assess the internal consistency of the data and the underlying model. We found that the model is self-consistent on the Myr and kyr time scales, but we discovered inconsistencies that make it difficult to achieve a good fit to all data simultaneously. It is also possible that the data themselves are not entirely self consistent in this regard. Our methodology does not resolve these questions, but once inconsistencies are identified, several strategies can be pursued to resolve them, e.g., improving the model or resolving consistency issues of the data themselves. Our conceptual and numerical framework can also be used to reveal the impact that some of the individual data sets have on parameter estimates and associated posterior uncertainties. In this paper, however, we focused on describing the mathematical and numerical framework and only briefly mention some of the implications.

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.

Acknowledgements. We thank Dr. Johannes Wicht of Max Planck Institute for Solar System Research, and an anonymous reviewer for careful comments that helped improving our paper. We also thank Dr. Andrew Tangborn of NASA GSFC for interesting discussion and useful comments. MM gratefully acknowledges support by the National Science Foundation under grant DMS-1619630 and by the Alfred P. Sloan Foundation. B.B. acknowledges support from grant EAR-164464 from the National Science Foundation.

References

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

Bärenzung, J., Holschneider, M., Wicht, J., Sanchez, S., and Lesur, V.: Modeling and Predicting the Short-Term Evolution of the Geomagnetic Field, Journal of Geophysical Research: Solid Earth, 123, 4539–4560, https://doi.org/10.1029/2017JB015115, 2018.

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, 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, 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 Geophysical Research: Solid Earth, 100, 6093–6095, 1995.

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 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, 263–314, 2001.

Hoyng, P., Schmitt, D., and Ossendrijver, M.: A theoretical analysis of the observed variability of the geomagnetic dipole field, Physics of Earth and Planetary Interiors, 130, 143–157, 2002.

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.

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.

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.

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.

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

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.

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.

Schmitt, D., Ossendrijver, M., and Hoyng, P.: Magnetic field reversals and secular variation in a bistable geodynamo model, Physics of Earth and Planetary Interiors, 125, 119–124, 2001.

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.

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

Ziegler, L., Constable, C., and Johnson, C.: Testing the robustness and limitations of 0-1 Ma absolute paleointensity data, Physics of Earth and Planetary Interiors, 170, 34–45, 2008.

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.

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~~ an expanded 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 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~~ We discuss these issues and describe that inconsistencies can result from a combination of the model deficiencies and 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~~ , possibly inconsistent data sets, which is particularly useful when the amount of data is limited.

1 Introduction

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 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).

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. Bruce, can you clarify this?

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~~ Other models have been derived by ~~Hoyng et al. (2001); Rikitake (1958); Pétrélis and Fauve (2008)~~Rikitake (1958); Pétrélis and Fauve (2008).

~~We focus on a simpler stochastic model that can potentially describe~~ Following Schmitt et al. (2001); Hoyng et al. (2001, 2002), we consider a stochastic differential equation (SDE) model for 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)~~axial dipole. The basic idea is to model Earth's dipole field ~~as being analogous to~~ analogous to the motion of a particle in a double well potential. Time variations of the dipole field and dipole reversals ~~are~~ then occur 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 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 drift and diffusion coefficients that define the B13 model are derived from the PADM2M data ~~Ziegler et al. (2011)~~ (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~~ account directly for the effects of sedimentation. Buffett and Puranam (2017) ~~added sedimentation processes to the model~~ try to mimic the effects of sedimentation by sending 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~~ record of Earth's dipole field on a Myr time-scale.

~~The B13 model relies on the assumption~~ A basic assumption of an SDE model is 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 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)~~Morzfeld et al. (2017) and by 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 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 ~~, see Ziegler et al. (2011); Valet et al. (2005)~~(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).~~ (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~~ a comprehensive paleomagnetic record composed of the above three sources of data. The results we obtain here are thus markedly different from previous work were data at 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). Priors are often assumed to be "uninformative", i.e., only conservative bounds for all parameters are known, and likelihoods describe point-wise 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.~~

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

The Bayesian approach further requires that we make assumptions about the errors in the model and in the data. Such "error models" describe 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

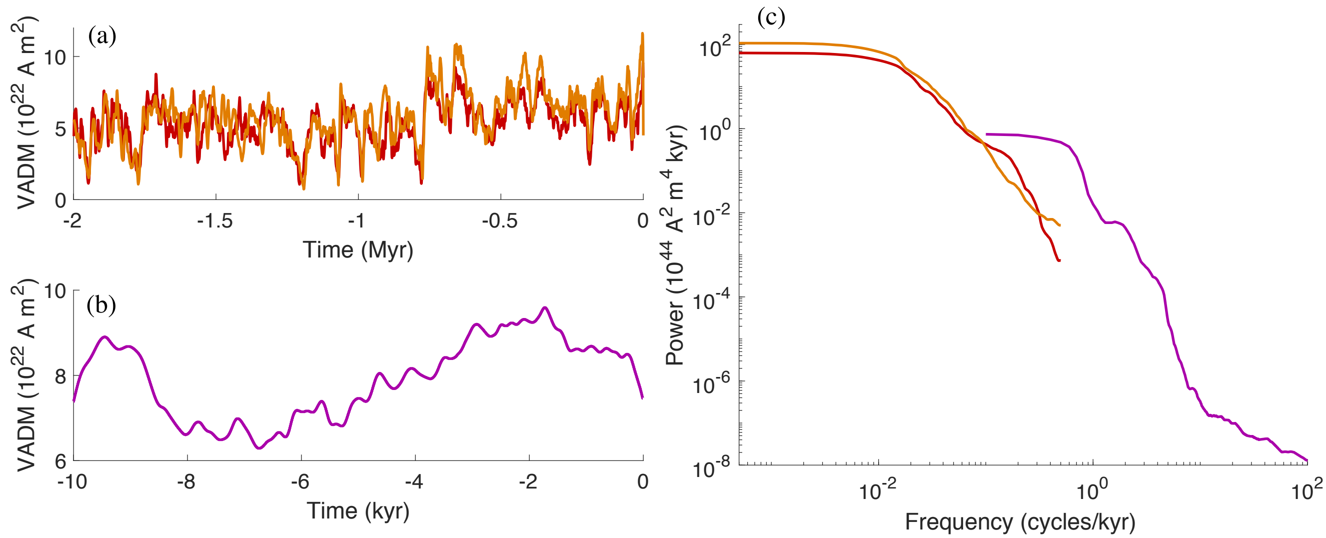

Figure 1. Data used in this paper. (a) Sint-2000 (orange) and PADM2M (red): VADM as a function of time over the past 2 Myr. (b) CALS10k.2: VADM as a function of time over the past 10 kyr. (c) Power spectral densities of the data in (a) and (b), computed by the multi-taper spectral estimation technique of Constable and Johnson (2005). Orange: Sint-2000. Red: PADM2M. Purple: CALS10k.2

assumptions about error models. In this context, we discover that the "shortness of the record," i.e., the limited amount of data available, is

2 Description of the data

~~We describe variations~~ Variations in the virtual axial dipole moment (VADM) over the past 2 Myr ~~using~~ can be derived from stacks of marine sediment~~data to reduce the influence of non-dipole components of the geomagnetic field.~~. Two different compilations are considered in this study: Sint-2000, ~~Valet et al. (2005)~~ (Valet et al., 2005) and PAMD2M ~~Ziegler et al. (2011)~~(Ziegler et al., 201 Both of these data sets are sampled every 1 kyr ~~. The results~~ and, thus, provide a time series of 2,000 consecutive VADM values. The PADM2M and Sint-2000 data sets 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. Time avg. VADM Std. dev. of VADM Data source (10^{22} Am^2) (10^{22} Am^2)PADM2M 5.23 1.48Sint-2000 5.81 1.84 Time average VADM and VADM standard deviation of PADM2M and Sint-2000.~~

~~We use the~~ The CALS10k.2 data set~~to describe variations in the~~ , plotted in the lower left panel of Figure 1, describes variations of VADM over the past 10 kyr (Constable et al., 2016). The time dependence of CALS10k2 is represented using B-splines, so that the model can be sampled at arbitrary time intervals. ~~The~~ We sample CALS10k.2 ~~data are shown in the lower left panel of Figure 1, sampled~~ at an interval of ~~1 year.~~ one year, although the resolution of CALS10k.2 is nominally 100 years (Constable et al., 2016).

Below we use features derived from power spectral densities (PSD) of the Sint-2000, PADM2M and CALS10k.2 data. The PSDs are computed by the multi-taper spectral estimation technique of Constable and Johnson (2005). A restricted range of frequencies are retained in the estimation to account for data resolution and other complications (see below). We show the resulting PSDs of the three data sets in the ~~left~~ right panel of Figure 1. During parameter estimation we further make use of the time average VADM and the standard deviation of VADM over time of the Sint-2000 and PADM2M data sets, listed in Table 1.

Data source	Time avg. VADM (10^{22} Am^2)	Std. dev. of VADM (10^{22} 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.

Lastly, we make use of reversal rates of the Earth's dipole computed from the geomagnetic polarity time scale ~~Cande and Kent (1995); Lo~~ 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 Myr – 5 Myr, 5 Myr – 10 Myr, ..., 25 Myr – 30 Myr. This leads to the average reversal rate and standard deviation listed in Table 2. ~~Increasing the interval to 10 Myr leads to the same mean but decreases the standard deviation (see Table 2).~~ Increasing the 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 the PSDs from CALS10k.2 and those from the lower resolution data sets (~~Sint-2000~~ SINT-2000 and PADM2M) ~~over the last 10 kyr (see Figure 1). These discrepancies may reflect inherent uncertaintiesthat 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~~ at the overlapping frequencies. Dating uncertainties, smoothing due to sedimentary processes, and the finite duration of the records all contribute to these discrepancies. We do not attempt to identify the source of these discrepancies. Instead, we seek to recover parameter values for a stochastic model by combining a feature-based approach with realistic estimates of the data uncertainty (see Section 4).

We further note that the amount of data is rather limited: we have 2 Myrs of VADM sampled at 1/ kyr, 10 kyr of high frequency VADM

3 Description of the ~~models on multiple time scales~~model

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).~~(see, e.g., Schmitt et al. (2001); Hoyng 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~~ the following 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"~~diffusion, $D(x)$, is constant, i.e., $D(x) = D$~~, as is suggested by the previous literature on this model~~. Modest variations in D have been reported on the basis of geodynamo simulations (Buffett et al., 2014; Buffett and Matsui, 2015; Meduri and Wicht, 2016). Are these the correct refe Representative variations in D, however, have a small influence on the statistical properties of solutions of the SDE (1).

The function v is called the "drift" and is derived from a double-well potential~~:~~, $U'(x) = -v(x)$. Here, we consider drift coefficients of the form

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

where $\bar{x}$ and γ are parameters. The parameter $\bar{x}$ defines where the drift coefficient vanishes and also corresponds to the time average of the associated linear model

$$\mathrm{d}x^l = -\gamma(x^l - \bar{x})\mathrm{d}t + \sqrt{D}\mathrm{d}W, \tag{3}$$

which is obtained by Taylor expanding $v(x)$ at $\bar{x}$. It is now clear that the parameter γ defines a relaxation time.

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 **??**. 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.~~ 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.

	$\bar{x}$ $(10^{22}\ \text{Am}^2)$	D $(10^{44}\ \text{A}^2\text{m}^4\ \text{kyr}^{-1})$	γ (kyr^{-1})	T_s (kyr)	a (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

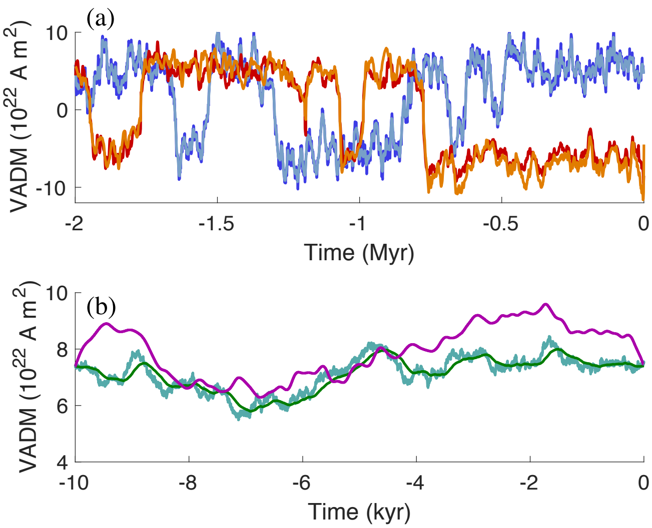

Figure 2. Drift coefficient $v(x)$ Simulations with nominal parameter values and data on the Myr and kyr scales. (a) VADM as a function of time on the Myr time scale. The output of the Myr model, x_j^{Myr}, is shown in dark blue (often hidden). The smoothed output, $x_j^{\text{Myr,s}}$, is shown in light blue. The signed Sint-2000 and potential $U'(x) = -v(x)$ PADM2M are shown in orange and red with signs (turquoisereversal timings) taken from Cande and Kent (1995). (b) VADM as a function of time on the kyr time scale. The output of the kyr-model with uncorrelated noise is shown in turquoise. The output of the kyr-model with correlated noise is shown in green. VADM of CALS10k.2 is shown in purple. Simulations and data. (a) Output, x_j^{Myr} (dark blue, often hidden) and smoothed model output, $x_j^{\text{Myr,s}}$ (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).

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

$$x_k = f(x_{k-1}, \Delta t) + \sqrt{2D\Delta t}\, w_k, \quad w_k \sim \mathcal{N}(0,1), \text{ iid} \tag{4}$$

where Δt is the time step, $\sqrt{\Delta t}\, w_k$ is the discretization of Brownian motion W in (1) and where $f(x_{k-1}, \Delta t)$ is the RK4 step. Here, iid means stands for "independent and identically distributed", and where $f(x_{k-1}, \Delta t)$ is the RK4 stepi.e., each random variable w_k, for $k > 0$, has the same Gaussian probability distribution, $\mathcal{N}(0,1)$, and w_i and w_j are independent for all $i \neq j$. 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 (4).

3.1 Models for the Myr amd kyr time scaleand 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 source of paleomagnetic data in Sint-2000 and PADM2M are affected by affected by gradual acquisition of magnetization due to sedimentation process which amounts processes, which amount 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 effects of sedimentation in the model by convolving the solution of (4) by a Gaussian filter

$$g(t) = \sqrt{\frac{6}{\pi T_s^2}} \cdot \exp\left(-\frac{6t^2}{T_s^2}\right), \tag{5}$$

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 whose state is denoted by $x^{\text{Myr,s}}$. Simulations with the "Myr model" using and the nominal parameters of Table 3 are shown in Figure 2(a) where we plot the model output x_j^{Myr} in dark blue and the smoothed model output, $x_j^{\text{Myr,s}}$, in a lighter blue over a period of 2 Myr. The PSDs of simulations corresponding to 50 Myr simulations are shown in Figure 2(c)

On a Myr scale, the assumption that the noise is uncorrelated in time is reasonable because one focuses on low frequencies and large sample intervals of the dipole, as in Sint-2000 and PADM2M, whose sampling interval is 1/kyr. On a shorter time scale, as in CALS10k.2, 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 (4) 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

$$y_k = (1 - a\Delta t)y_{k-1} + \sqrt{2a\Delta t}\, w_k, \quad w_k \sim \mathcal{N}(0,1), \text{ iid} \tag{6}$$

$$x_k = f(x_{k-1}, \Delta t) + \sqrt{Da}\, \Delta t\, y_k, \tag{7}$$

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 2(b) along with the CALS10k.2 data.

3.2 Approximate power spectral densities

Accurate computation of the power spectral density (PSD) from the time-domain solution of the B13 model requires extremely long simulations. For example, the PSDs of two (independent) 1 billion year simulations with the Myr model are still quite different. In fact, errors that arise due to "short" simulations outweigh errors due to linearization. Recall that the PSD of the linear model (3) is easily calculated to be

$$\hat{x}^l(f) = \frac{2D}{\gamma^2 + 4\pi^2 f^2}, \tag{8}$$

where f is the frequency (in $1/\text{kyr}$). Since the Fourier transform of the Gaussian filter is known analytically, the PSD of the smoothed linear model is also easy to calculate:

$$\hat{x}^{l,s}(f) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \exp\left(-\frac{4\pi^2 f^2 T_s^2}{12}\right). \tag{9}$$

Similarly, an analytic expression for the PSD of the kyr-model with correlated noise in equations (14)-(14) can be obtained by taking the limit of continuous time ($\Delta t \to 0$):

$$\hat{x}^{l,kyr}(f) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \frac{a^2}{a^2 + 4\pi^2 f^2}. \tag{10}$$

Here, the first term is as in (8) and the second term appears because of the correlated noise.

Figure 3 illustrates a comparison of the PSDs obtained from simulations of the nonlinear models and their linear approximations. Specifically, the PSDs of the (smoothed) Myr scale nonlinear model, computed from a 50 Myr simulation, are shown in comparison to the approximate PSDs in equations (8)-(9). Note that the PSD of the smoothed model output, $x^{\text{Myr,s}}$, 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)and Figure 2.

Because an SDE is noisy (Brownian motion), long simulations are required to obtain accuracy in . The PSD of the kyr model with correlated noise, computed from a 10 kyr simulation, is also shown in Figure 3 in comparison with the linear PSD in equation (10). The good agreement between the theoretical spectra of 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* linear models and the spectra of the nonlinear models justifies the use of the linear approximation. We have further noted in numerical experiments that the agreement between the nonlinear and linear spectra increases with increasing simulation time, however a "perfect" match requires extremely long simulations of the nonlinear model (hundreds of billions of years). The approximate PSDs, based on the linear models, will prove useful in the construction of likelihoods in Section 4.2.

In addition to a good match of the PSDs of the nonlinear and linear models, we note that the PSDs of the model

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

whose PSD is easy to calculate analytically as

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

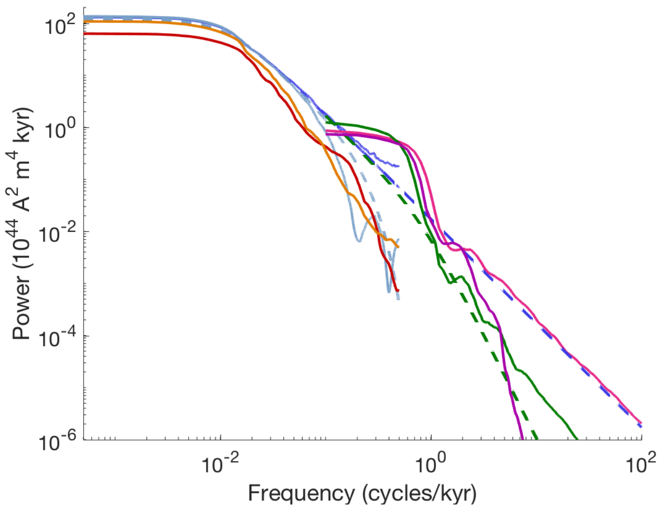

Figure 3. Power spectral densities of the model with nominal parameter values. *Myr-model*: A PSD of a 50 Myr simulation with the Myr model is shown as a solid dark blue line. The corresponding theoretical PSD of the linear model is shown as a dashed dark blue line. A PSD of a 50 Myr simulation with the Myr model and high-frequency roll-off is shown as a solid light blue line. The corresponding theoretical PSD of the linear model is shown as a dashed light blue line. *kyr model*: A PSD of a 10 kyr simulation of the kyr model with uncorrelated noise is shown as a solid pink line. The corresponding theoretical spectrum of the linear model is shown as a dashed blue line. A PSD of a 10 kyr simulation of the kyr model with correlated noise is shown as a solid green line. The corresponding theoretical spectrum is shown as a dashed green line. *Data*: The PDSs of Sint-2000, PADM2M and CALS10k.2 are shown in orange, red and purple. The PSDs of the simulations and data are computed by the multi-taper spectral estimation technique of Constable and Johnson (2005).

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

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

The approximate PSDs $\hat{x}^l(f)$ and $\hat{x}^{l,s}(f)$, computed by (8) and (9), are plotted in Figure 2(a) in comparison to the PSDs computed from a 50 Myr simulation with the discrete time model (4). match, at least to some extent, the PSDs of the data (Sint-2000, PADM2M and CALS10k.2). This means that our choice for the nominal values is “reasonable” because this choice leads to a reasonable fit to the data. The goal of using a Bayesian approach to parameter estimation, described in Section 4, is to improve this fit and to equip the (nominal) parameter values with an error estimate, i.e., to define and compute a distribution over the model parameters. This will lead to an improved fit, along with an improved understanding of model uncertainties.

3.3 ~~Reversal~~ Approximate reversal rate, VADM time average and VADM standard deviation

The nonlinear SDE model (1) and its discretization (4) exhibit reversals, i.e., a 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 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~~ In principle, these quantities (reversal rate, time ~~domain simulations with the discrete time model (4) are dominated by errors that arise from limitations of~~ average and standard deviation) can be computed from simulations of the Myr and kyr model in the time domain. Similarly to what we found in the context of PSD computations and approximations, we fund that estimates of the reversal rate, time average and standard deviation are subject to large errors unless the simulation time ~~. Only long simulations~~ is very long (hundreds of millions of years)~~lead to accurate estimates .~~. Using the linear model and Kramers formula, however, one can approximate the time average, reversal rate and standard deviation without simulating the nonlinear model (see below). Computing the approximate values is instantaneous (evaluation of simple formulas) and the approximations are comparable to what we obtain from very long simulations with the nonlinear model. As is the case with the PSD approximations based on linear models, approximations of the reversal rate, time average ~~VADM and VADM standard deviation . Long simulations, however, require more substantial computations. This would not be an issue 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~~

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

~~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.~~and standard deviation based on the linear models will prove useful for formulating likelihoods in Section 4.2.

Specifically, the parameter $\bar{x}$ defines the time average of the linear model (3) and it also defines where the drift term (2) vanishes. These values coincide quite closely with the time average of the nonlinear model which suggests the approximation

$$E(x) \approx \bar{x} \cdot 10^{22} \text{ Am}^2. \tag{11}$$

The reversal rate can approximated by Kramers formula (Buffett and Puranam, 2017; Risken, 1996)

$$r \approx \frac{\gamma}{2\pi} \cdot \exp\left(-\frac{\gamma \bar{x}^2}{6D}\right) \cdot 10^3 \text{ Myr}^{-1}. \tag{12}$$

The standard deviation is the square root of the area under the PSD. Using the linear model that incorporates the effects of smoothing (due to sedimentation), one can approximate the standard deviation by computing the integral of the PSD in equation (9):

$$\sigma \approx \left(\frac{D}{\gamma}\exp\left(\frac{(\gamma T_s)^2}{12}\right)\mathrm{erfc}\left(\frac{\gamma T_s}{2\sqrt{3}}\right)\right)^{1/2}\cdot 10^{22}\ \mathrm{Am}^2\ \mathrm{kyr}^2, \tag{13}$$

where erfc(·) is the (Gauss) error function. Without incorporating the smoothing, the standard deviation based on the linear model would be the integral of the PDS in (8), which is $\sigma \approx \sqrt{D/\gamma}$. The exponential and error function terms in (13) can thus be inerpreted as a correction factor that accounts for the effects of sedimentation. It is easy to check that this correction factor is always smaller than one, i.e., the (approximate) standard deviation accounting for sedimentation effects is smaller than the (approximate) standard deviation that does not account for these effects.

For the nominal values we thus parameter values in Table 3 we 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 $\sigma \approx 2.07 \cdot 10^{22}$ 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.

3.4 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 (4) 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

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

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 2(b) along with the CALS10k.2 data.

As with the Myr model, we approximate the PSD of the kyr model (14) and (14) 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

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

where the first term is as in (8) 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 2(c) in comparison to the PSD computed from a 10 kyr run and the PSD of the CALS10k.2 dataSimilar to what we observed of the model-data fit in terms of (approximate) PSDs, we find that the nominal parameter values lead to a "reasonable" fit of the model's reversal rate, time average and standard deviation. The Bayesian parameter estimation in Section 4 will improve this fit and lead to a better understanding of model uncertainties.

3.4 Parameter bounds

~~We have~~ The Bayesian parameter estimation, described in Section 4, makes use of "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 2). 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~~ We formulate prior information in terms of parameter bounds and construct uniform prior distributions ~~in Section 4~~with these bounds. 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)~~(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 $\gamma \geq 0.0205$ 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,~~ dipole may relax on timescales shorter than the slowest decay mode and a recent theoretical calculation (Pourovskii et al., 2017) suggests that the magnetic diffusivity may be slightly larger than $0.8\text{m}^2/\text{s}$. Both of these changes would cause 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 $\gamma \leq 0.7$ kyr^{-1}.

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

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, $\bar{x} \in [0,10] \cdot 10^{22}$ 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.~~ 3.

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 shorthand notation for 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

$$p(\theta|y) \propto p_0(\theta) p_l(y|\theta). \tag{14}$$

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~~ model of the Earth's dipole in view of the data we ~~have.~~ use.

On the other hand, the posterior distribution depends on several assumptions: since we define the prior and likelihood~~and, therefore,~~, we also implicitly define the posterior distribution. ~~In the following we describe how we formulate the prior and the likelihood.~~ In particular, formulations of the likelihood require that one be able to describe anticipated errors in the data as well as anticipa

We solve the Bayesian parameter estimation problem numerically by using a Markov chain Monte Carlo (MCMC) method. An MCMC

4.1 Prior distribution

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~~ parameters 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", 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 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 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-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).

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);

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

In the Bayesian approach, and assuming that errors are independent, this means that the likelihood, $p_l(y|\theta)$ in equation (14) can be written as the product of three terms

$$p_l(y|\theta) \propto p_{l,\text{td}}(y|\theta)\, p_{l,\text{lf}}(y|\theta) p_{l,\text{hf}}(y|\theta), \tag{15}$$

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

We define the likelihood component that addresses the reversal rate of the time domain data based on the equation

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

equations

$$y_{\mathrm{rr}} = h_{\mathrm{rr}}(\theta) + \varepsilon_{\mathrm{rr}}, \tag{16}$$

$$y_{\bar{x}} = h_{\bar{x}}(\theta) + \varepsilon_{\bar{x}}, \tag{17}$$

$$y_{\sigma} = h_{\sigma}(\theta) + \varepsilon_{\sigma}, \tag{18}$$

where y_{rr}~~is the reversal rate~~, $y_{\bar{x}}$ y_{σ} are features derived from the time domain data, $h_{\mathrm{rr}}(\theta)$~~is a function that maps~~, $h_{\bar{x}}(\theta)$ and $h_{\sigma}(\theta)$ are functions that connect the model parameters to the ~~reversal rate data,~~ features, based on the approximations described in Section 3.3, and where $\varepsilon_{\mathrm{rr}}$~~is an error model. As described in Section 2, we use the chronology from Ogg (2012) and set~~, $\varepsilon_{\bar{x}}$ and ε_{σ} are independent Gaussian error models with mean zero and variances σ^2_{rr}, $\sigma^2_{\bar{x}}$, and σ^2_{σ}. Taken all together, the likelihood term $p_{l,\mathrm{td}}(y|\theta)$ in (15) is then given by the product of the three likelihoods defined by equations (16), (17) and (18):

$$p_{l,\mathrm{td}}(y|\theta) \propto \exp\left(-\frac{1}{2}\left(\left(\frac{y_{\mathrm{rr}} - h_{\mathrm{rr}}(\theta)}{\sigma_{\mathrm{rr}}}\right)^2 + \left(\frac{y_{\bar{x}} - h_{\bar{x}}(\theta)}{\sigma_{\bar{x}}}\right)^2 + \left(\frac{y_{\sigma} - h_{\sigma}(\theta)}{\sigma_{\sigma}}\right)^2\right)\right). \tag{19}$$

The reversal rate feature is simply the average reversal rate we computed from the chronology of Ogg (2012) (see Section 2), i.e., $y_{\mathrm{rr}} = 4.23$ reversals/Myr. The function $h_{\mathrm{rr}}(\theta)$ is based on the ~~nonlinear model (2) but avoids the need for long simulations by~~ approximation using Kramers formula ~~Risken (1996) to define the reversal rate. Following Buffett and Puranam (2017) (see also equation (13)), we thus approximate the reversal rate of the model by~~ in equation (20):

$$h_{\mathrm{rr}}(\theta) = \frac{\gamma}{2\pi}\exp\left(-\frac{\gamma\bar{x}^2}{6D}\right)\cdot 10^3 \text{ reversals/Myr.} \tag{20}$$

~~We use a Gaussian error model with mean zero, i.e., $\varepsilon_{\mathrm{rr}} \sim \mathcal{N}(0,\sigma^2_{\mathrm{rr}})$. We discuss the error variance σ^2_{rr} below.~~

~~We define the likelihood for the time averaged VADM by~~

~~$y_{\bar{x}} = h_{\bar{x}}(\theta) + \varepsilon_{\bar{x}}$.~~

~~We use the average~~ The time average feature is the mean of the time averages of PADM2M and Sint-2000~~as the datum and set~~: $y_{\bar{x}} = 5.56 \cdot 10^{22}$ Am^2. The function ~~$h_{\bar{x}}(\theta) = \bar{x}$~~$h_{\bar{x}}(\theta)$ is based on the linear approximation discussed in Section 3.3, 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, $\varepsilon_{\bar{x}} \sim \mathcal{N}(0,\sigma^2_{\bar{x}})$, and the error model variance $\sigma^2_{\bar{x}}$ is discussed below.~~

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

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

~~We use~~ $h_{\bar{x}}(\theta) = \bar{x}$. The feature for the VADM standard deviation is the average of the VADM standard deviations of PADM2M and Sint-2000~~as the datum~~: $y_{\sigma} = 1.66 \cdot 10^{22}$ Am^2. The ~~standard deviation over time of the linear model (3) 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 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 (15) is then given by the product of the three likelihoods defined by equations (16), (17) and (18): function $h_\sigma(\theta)$ uses the linear approximation of the standard deviation (13):

$$p_{l,\mathrm{td}} h_\sigma(y\theta) \propto = \left(\frac{D}{\gamma}\exp\left(-\frac{1}{2}\frac{y_{\mathrm{rr}}-h_{\mathrm{rr}}(\theta)}{\sigma_{\mathrm{rr}}}\frac{(\gamma T_s)^2}{12}\right)^2 + \mathrm{erfc}\left(\frac{y_{\bar{x}}-\bar{x}}{\sigma_{\bar{x}}}\gamma T_s/2/\sqrt{(3)}\right)^2 + \frac{y_\sigma - h_\sigma(\theta)}{\sigma_\sigma}\right)^{1/2} \cdot 10^{22}\ \mathrm{Am}^2 . \mathrm{kyr}^2 \qquad (21)$$

Candidate values for the error variances σ_{rr}^2, $\sigma_{\bar{x}}^2$ and σ_σ^2 these error variances are as follows. The error variance of the reversal rate, $\sigma_{\bar{x}}^2 \sigma_{\mathrm{rr}}^2$, can be based on the standard deviations we computed from the Ogg (2012) chronology in Table 2. Thus, we might use $\sigma_{\bar{x}} = 0.5$the standard deviation of the 10-Myr average and take $\sigma_{\mathrm{rr}} = 0.5$. 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 these error covariances is that we have few time domain observations compared with the large number of spectral data in the power spectra (see below). This vast difference in the number of time domain and spectral data means that the spectral data completely can 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

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

We discuss this choice and its consequences on parameter estimates and associated uncertainty in more detail in Section 6. An alternative approach would be to reduce the number of spectral data, e.g., by condensing the spectra into "a few" characteristics such as corner frequencies or slopes, see also (Bärenzung et al., 2018). The difficulty with such an approach, however, is that corner frequencies etc. are not easy to extract from the PDSs of the data. For these reasons, we pursue in this paper a "tuning" of the error variances to control the impact of the time domain data on the parameter estimates, but alternative strategies are straightforward to implement within our overall feature-based Bayesian estimation approach.

4.2.2 Low frequencies

The component $p_{l,\mathrm{lf}}(y|\theta)$ of the feature-based likelihood (15) 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

$$y_{\mathrm{lf}} = h_{\mathrm{lf}}(\theta) + \varepsilon_{\mathrm{lf}}, \qquad (23)$$

where y_{lf} is a feature that 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~~feature 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 these computations by approximating the~~ and is based on the PSD of the ~~nonlinear model by the PSD of the~~ linear model (3). ~~This can be done simply by evaluating (8).~~ 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 (8) for frequencies less than 0.05 cycles/kyr, and uses the "smoothed" spectrum of equation (9) for frequencies between $0.05-0.5$ cycles/kyr:

$$h_{\mathrm{lf}}(\theta) = \frac{2D}{\gamma^2+4\pi^2 f^2}\cdot\begin{cases}1 & \text{if } f\le 0.05\\ \exp\left(-(4\pi^2 f^2 T_s^2)/12\right) & \text{if } 0.05 < f \le 0.5\end{cases} \tag{24}$$

Note that $h_{\mathrm{lf}}(\theta)$ 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 $\varepsilon_{\mathrm{lf}}$ 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~~

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~~ spectral densities. For a Gaussian error model $\varepsilon_{\mathrm{lf}}$ with zero mean, this means that the error covariance ~~is computed~~ should describe uncertainties that are induced by the limited amount of data. We construct such a covariance as follows. We perform 10^4 simulations, each of 2 Myr, with the nonlinear Myr model (4) 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~~ $5\cdot 10^3$ samples of $\varepsilon_{\mathrm{lf}}$ added to y_{lf} (~~turquoise~~grey). 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 sets would not change the conclusion that the largest source of uncertainty in the power spectra is due to the shortness of the record.

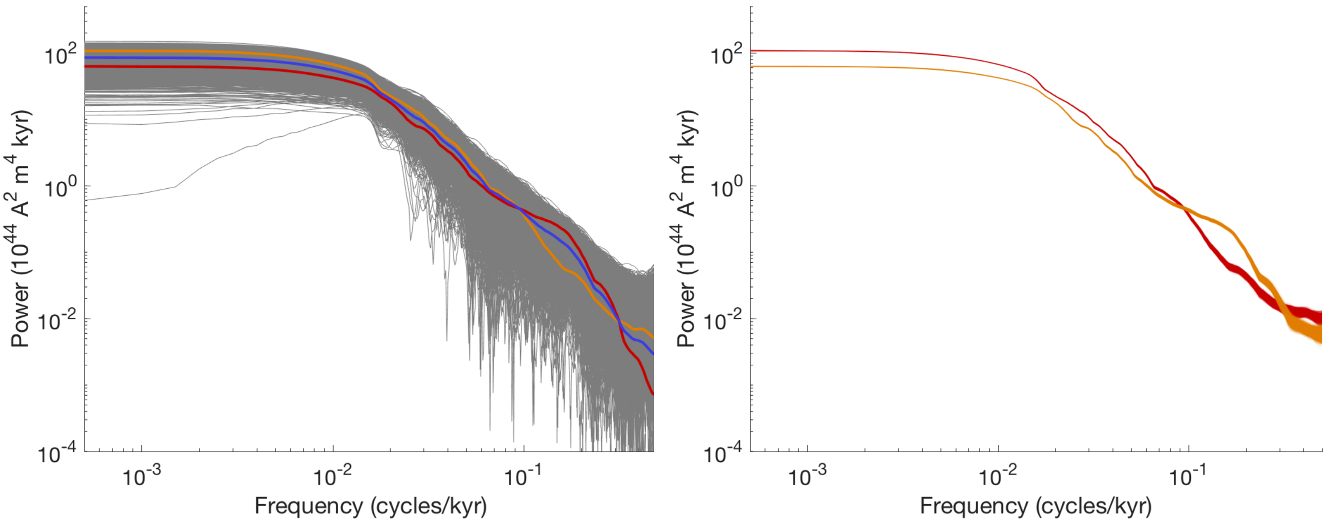

Figure 4. Left: low frequency data and error model due to shortness of record. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Blue: mean of PSDs of Sint-2000 and PADM2M ($y_{\rm lf}$). ~~Turquoise~~Grey: ~~10^3~~ $5\cdot 10^3$ samples of the error model $\varepsilon_{\rm lf}$ added to $y_{\rm lf}$. Right: 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.

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 $0.9-9.9$ cycles/kyr~~. We chose this window~~ , where the upper limit is set by the resolution of the CALS10k.2 data. The lower limit is chosen 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~~ may be less reliable that the PSD of CALS10k.2 for these frequencies. As above, we construct the likelihood $p_{l,\rm hf}(y|\theta)$ from an equation similar to (23):

$$y_{\rm hf} = h_{\rm hf}(\theta) + \varepsilon_{\rm hf}, \tag{25}$$

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

~~As above, we~~ We base $h_{\rm hf}(\theta)$ on the PSD of the linear model (see equation (10)) to avoid errors due to computations in the time domain and set

$$h_{\rm hf}(\theta) = \frac{2D}{\gamma^2+4\pi^2 f^2}\cdot\frac{a^2}{a^2+4\pi^2 f^2}, \tag{26}$$

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 model.

The error model $\varepsilon_{\rm hf}$ 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 (14)-(14) with nominal parameter values.

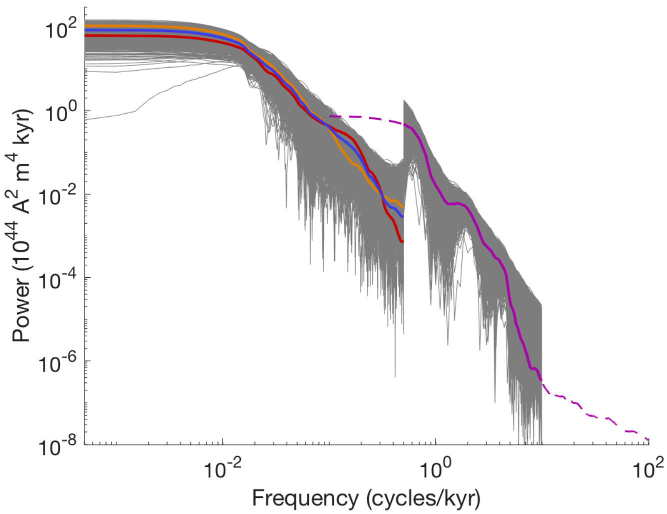

Figure 5. Data and error models for low and high frequencies. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Blue: mean of PSDs of Sint-2000 and PADM2M (y_{lf}). ~~Turquoise~~ Grey (low frequencies): ~~10^3~~ $5 \cdot 10^3$ samples of the error model $\varepsilon_{\mathrm{lf}}$ added to y_{lf}. Dashed purple: PSD of CALS10k.2. Solid purple: PSD of CALS10k.2 at frequencies we consider (y_{hf}). ~~Turquoise~~ Grey (~~low~~ high frequencies): ~~10^3~~ $5 \cdot 10^3$ samples of the error model $\varepsilon_{\mathrm{hf}}$ added to y_{hf}.

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 $\varepsilon_{\mathrm{hf}}$, which is illustrated along with the low frequency 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 (14).

4.3 Numerical solution by MCMC

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

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 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 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 samples, but the number of "effective" samples is $10^6/\text{IACT}$. For all MCMC runs we perform (see Sections 5 and 6), the IACT of the Markov chain is about 100. We discard the first $10 \cdot \text{IACT}$ 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 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) 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, .~~. We found that the MCMC Hammer solves the problem ~~efficiently~~with sufficient efficiency for our purposes.

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 that are missing some of the factors that comprise the overall feature-based likelihood (15). This allows us to study the impact of 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 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.~~ which define the "uninformative", uniform prior. Thus, the posterior probability, which synthesizes the information from the data via the definition of the features, is concentrated over a smaller subset of parameters than the prior probability. In this way, the Bayesian parameter estimation has sharpened the knowledge about the parameters by incorporating the data.

The 2-dimensional histograms indicate correlations among the parameters $\theta = (\bar{x}, D, \gamma, T_s, a)^T$, with strong correlations between $\bar{x}$, D and γ. These correlations can also be described by the correlation coefficients ~~between the five parameters which~~

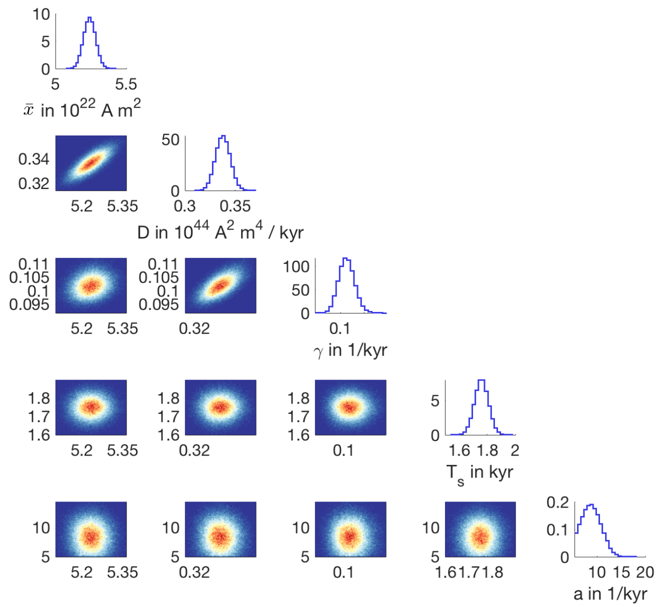

Figure 6. 1- and 2-dimensional histograms of the posterior distribution.

$\bar{x}$ (in 10^{22} Am^2)	D (in 10^{44} A^{2}m^4 kyr^{-1})	γ (in kyr^{-1})	T_s (in kyr)	a (in kyr^{-1})	σ (in 10^{22} Am^2)	Rev. rate (in reversals/Myr)
~~5.21 (0.041~~5.23 (0.043)	0.34 (~~0.0074~~0.0072)	0.10 (~~0.0037~~0.0033)	~~1.74~~ 1.75 (0.050)	~~8.57 (1.95~~8.56 (1.93)	~~1.80 (0.023~~1.77 (0.024)	4.06 (~~0.048~~0.049)

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

~~we computed as~~

$$\rho = \begin{array}{c|ccccc} & \bar{x} & D & \gamma & T_s & a \\ \hline \bar{x} & 1.00 & 0.78 & 0.20 & 0.02 & -0.03 \\ D & 0.78 & 1.00 & 0.64 & 0.02 & -0.03 \\ \gamma & 0.20 & 0.64 & 1.00 & -0.01 & -0.02 \\ T_s & 0.02 & 0.02 & -0.01 & 1.00 & 0.00 \\ a & -0.03 & -0.03 & -0.02 & 0.00 & 1.00 \end{array} \quad (27)$$

~~We note~~ The 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 (20)) and the dependence of the spectral data on D and γ (see Equations (9) and (10)). From the samples, we 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~~ 2,000 samples of the posterior distribution ~~(using Equation~~ followed by evaluation of Equations (20) and (13) 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 deviation (mean value of ~~$1.80 \cdot 10^{22}$ Am^2~~$1.77 \cdot 10^{22}$ Am^2) is also far~~(in terms of posterior standard deviations)~~ , as measured by the posterior standard deviation, from the value we use as data (1.6610^{22} Am^2). These large deviations indicate an inconsistency between the VADM standard deviation and the reversal rate. A higher reversal rate could be achieved with a higher VADM standard deviation. The reason is that the reversal rate in Equation (20) can be re-written as

$$r \approx \frac{\gamma}{2\pi} \cdot \exp\left(-\frac{\bar{x}^2}{6\sigma^2}\right) \cdot 10^3 \text{ Myr}^{-1}, \tag{28}$$

using $\sigma \approx \sqrt{D/\gamma}$, i.e., neglecting the correction factor due to sedimentation, which has only a minor effect. Using a time average of $\bar{x} = 5.23 \cdot 10^{22}$ Am^2, a reversal rate $r = 4.2$ reversals / Myr, setting $\gamma = 0.1\,\text{kyr}^{-1}$ (posterior mean value), and solving for the VADM standard deviation results in $\sigma \approx 4.53 \cdot 10^{22}$ Am^2, which is not compatible with the SINT-2000 and PADM2M data sets (where $\sigma \approx 1.66 \cdot 10^{22}$ Am^2). It is possible that the low-frequency data sets underestimate the standard deviation and also the time average. For example, Ziegler et al. (2008) report a time average VADM of $7.64 \cdot 10^{22}$ Am^2 and a standard dev

The model fit to the spectral data ~~, however, is good, as is~~ 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. For comparison, the figure also shows the PADM2M, Sint-2000 and CALS10k.2 data as well as $5 \cdot 10^3$ realizations of the low- and high-frequency error models. We note that the overall uncertainty is reduced by the Bayesian parameter estimation. The reduction in uncertainty is apparent from the expected errors generating a "wider" cloud of PSDs (in grey) than the posterior estimates (in blue and turquoise). We further note that the PSDs of the models, with parameters drawn from the posterior distribution, fall largely within the expected errors (illustrated in grey). In particular the high-frequency PSDs (the CALS10k.2 range) are well within the errors we imposed by the likelihoods. The low frequencies of Sint-2000/PADM2M are also within the expected errors and so are the high-frequencies beyond the second roll-off due to the sedimentation effects. At intermediate frequencies, some of the PDSs of the model are outside of the expected errors. This indicates a model inconsistency in the frequency domain: based on the data (spectral *and* time domain) and the assumed error models, it is difficult to find model parameters that lead to a good spectral fit at all relevant frequencies. We investigate this issue further in Section 6.

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,~~ We note that the model with posterior mean parameter values exhibits qualitatively similar characteristics as the Sint-2000, PADM2M and CALS10k.2 data. The figure thus illustrates that the feature-based Bayesian parameter estimation, which is based solely on PSD, reversal rates, time average VADM and VADM standard deviation, translates into model parameters that also appear reasonable when a single simulation in the time domain is considered.

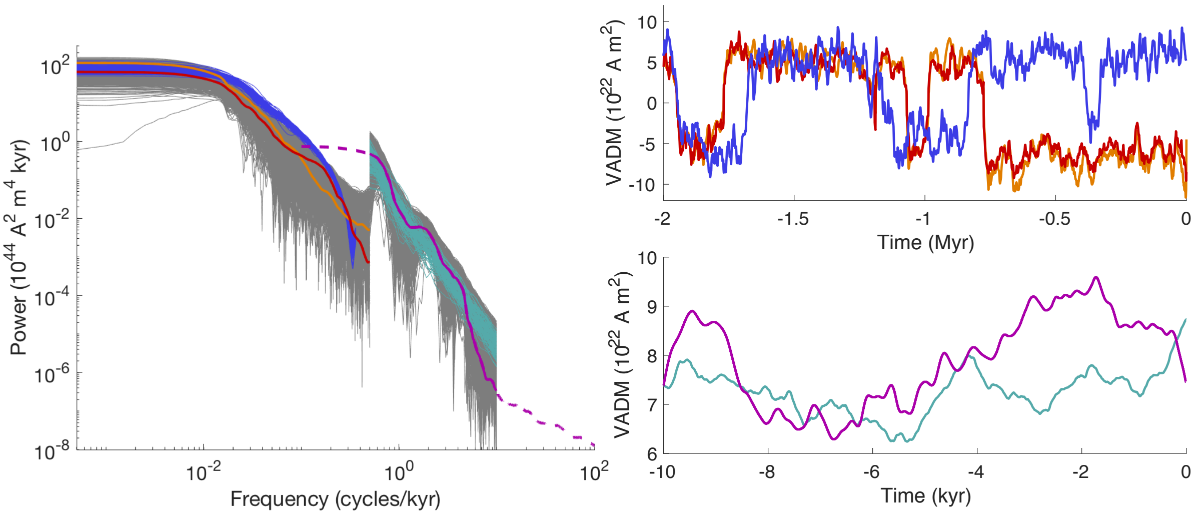

Figure 7. Parameter estimation results. Left: PSDs of data and model. Orange: PSD of Sint-2000. Red: PSD of PADM2M. ~~Purple~~Dashed purple: PSD of CALS10k.2. ~~Dark blue~~Solid purple: ~~PSDs~~ PSD of ~~100 posterior~~ CALS10k.2 at frequencies we consider (y_{hf}). Grey (low frequencies): $5 \cdot 10^3$ samples of ~~Myr~~ the error model $\varepsilon_{\mathrm{lf}}$ added to y_{lf}. Grey (~~with and without smoothing~~high frequencies): $5 \cdot 10^3$ samples of the error model $\varepsilon_{\mathrm{hf}}$ added to y_{hf}. ~~Light~~ Dark blue: PSDs of 100 posterior samples of ~~kyr~~ Myr model (with ~~uncorrelated noise~~smoothing). Turquoise: PSDs of 100 posterior samples of kyr model with ~~correlated~~ uncorrelated noise. Right, top: Sint-2000 (orange), PADM2M (red) and a realization of the Myr model with smoothing and with posterior mean parameters (blue). Right, bottom: CALS10k.2 (purple) ~~,~~and a realization of the kyr model with correlated noise and with posterior mean parameters kyr model (turquoise).

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 (as compared to the uniform prior). The posterior distribution describes a set of model parameters that yield model outputs that are comparable with the data in the ~~feature-base~~ feature-based 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 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 model with posterior parameters. ~~This is a first~~ As indicated above, this is a hint at that there are inconsistencies between spectral data and time domain data, which we will study in more detail in the next section.

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.

6 Discussion

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

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, $p_{l,\text{hf}}(y|\theta)$ in the feature-based likelihood (15). 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~~The differences and similarities are apparent when we compare the corner plots of the posterior distributions of configurations (a), shown in Figure 6, and of configuration (b), shown in Figure 8. The corner plots are nearly identical except for the bottom row of plots which illustrates marginals of 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~~ posterior related to a. We note that the posterior distribution over a is nearly identical to its prior distribution~~, we conclude that~~ . Thus, the parameter a is not ~~well~~ constrained by the data ~~of~~ used in configuration (b)~~. The fact that~~ , which is perhaps not surprising because 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 $p_{l,\text{hf}}(y|\theta)$~~;~~ . Moreover, since $p_{l,\text{lf}}(y|\theta)$ and $p_{l,\text{td}}(y|\theta)$ 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 is self-consistent and consistent with the data on the Myr and kyr time scales; in the context of our simplistic model CALS10k.2 ~~is mostly useful for constraining~~ mostly constraints the noise correlation parameter a.

Configuration	(a)	(b)	(c)
$\bar{x}$ (10^{22} Am^2)	~~5.21 (0.041~~5.23 (0.043)	~~5.21~~ 5.23 (0.042)	3.56 (0.26)
D (10^{44} A^{2}m^4 kyr^{-1})	0.34 (~~0.0074~~0.0072)	0.34 (~~0.0073~~0.0072)	0.13 (0.014)
γ (kyr^{-1})	0.10 (~~0.0037~~0.0033)	0.10 (~~0.0035~~0.0033)	0.081 (0.0052)
T_s (kyr)	~~1.74~~ 1.75 (0.050)	1.74 (0.050)	1.68 (~~0.13~~0.14)
a (kyr^{-1})	~~8.57 (1.95~~8.56 (1.93)	22.45 (~~10.17~~10.01)	~~11.66 (3.93~~11.69 (3.91)
σ (10^{22} Am^2)	~~1.80 (0.023~~1.77 (0.024)	~~1.80~~ 1.77 (0.023)	~~1.25 (0.062~~1.22 (0.060)
Rev. rate (reversals/Myr)	4.06 (~~0.048~~0.049)	~~4.05 (0.046~~4.06 (0.023)	~~3.33~~ 3.34 (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~~0.44 (0.021)	0.093 (0.015)
γ (kyr^{-1})	0.078 (0.0063)	~~0.18 (0.016~~0.14 (0.014)	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~~22.59 (10.08)	22.37 (10.13)
σ (10^{22} Am^2)	~~1.10 (0.077~~1.08 (0.074)	~~1.65~~ 1.66 (0.036)	~~1.09 (0.077~~1.07 (0.075)
Rev. rate (reversals/Myr)	~~2.93 (4.08~~2.89 (4.11)	4.23 (~~0.054~~0.050)	~~2.80 (3.90~~2.83 (4.10)

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).

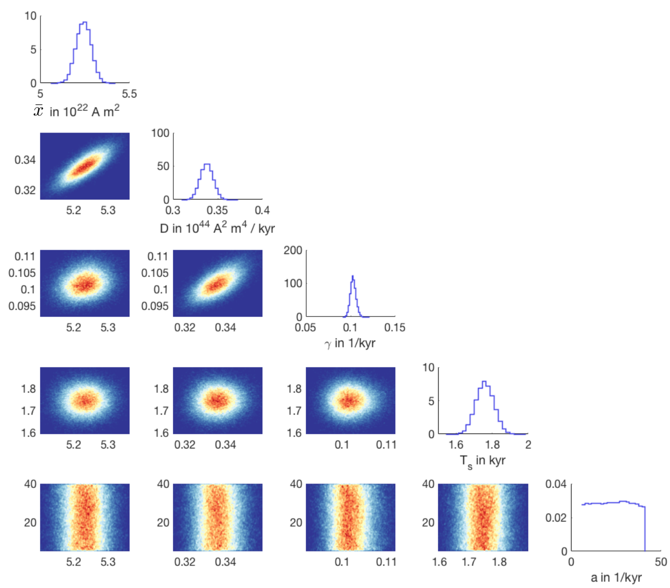

Figure 8. 1- and 2-dimensional histograms of the posterior distribution of configuration (b).

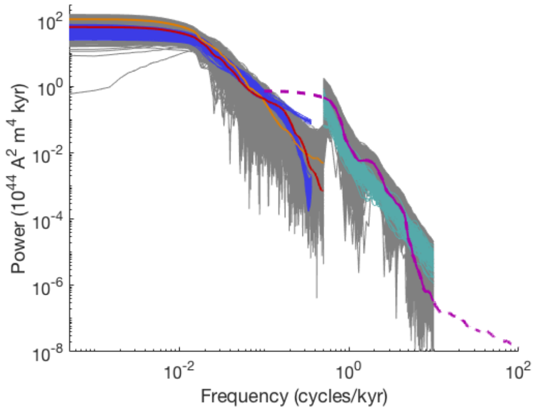

Figure 9. PSDs of data and model with parameters drawn from the posterior distribution of configuration (c). Orange: PSD of Sint-2000. Red: PSD of PADM2M. Dashed purple: PSD of CALS10k.2. Solid purple: PSD of CALS10k.2 at frequencies we consider (y_{hf}). Grey (low frequencies): $5 \cdot 10^3$ samples of the error model $\varepsilon_{\mathrm{lf}}$ added to y_{lf}. Grey (high frequencies): $5 \cdot 10^3$ samples of the error model $\varepsilon_{\mathrm{hf}}$ added to y_{hf}. Dark blue: PSDs of 100 posterior samples of Myr model (with smoothing). Turquoise: PSDs of 100 posterior samples of kyr model with uncorrelated noise.

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 configuration (c), the spectral data are emphasized during the Bayesian estimation which also leads to an overall better fit of the spectra. This is illustrated in Figure 9, where we plot the 100 PSDs generated by 100 (independent) simulations with the model with parameters drawn from the posterior distribution of configuration (c). For comparison, we also plot the PSDs of PADM2M, Sint-2000, CALS10k.2 and $5 \cdot 10^3$ realizations of the high- and low-frequency error model. In contrast to configuration (a) (see Figure 7), we find that the PSDs of the model of configuration (c) are all well within the expected errors. One the other hand, the reversal rate drops to about 3 reversals/Myr, and the time average VADM and VADM standard deviation also decrease significantly as compared to configuration (a). ~~The~~ This is caused by the posterior mean of D ~~decreases~~ being decreased by more than 50%~~;~~, while γ and T_s are comparable for configurations (a)-(c). The fact that the improved fit of the PDSs comes at the cost of a poor fit of the reversal rate, time average and standard deviation is another indication of an inconsistency between the reversal rate and the VADM data sets. As indicated above, one of the strengths of the Bayesian parameter estimation framework we describe here is to be able to identify such inconsistencies. Once identified, one can try to fix the model For example, we can envision a modification of the functional form of the drift term in (2). A nearly linear dependence of the drift term on x near $x = \bar{x}$ is supported by the VADM data sets, but the behavior near $x = 0$ is largely unconstrained. Symmetry of the underlying governing equations suggests that the drift term should vanish and the functional form adopted in (2) is just one way that a linear trend can be extrapolated to $x = 0$. Other functional forms that lower the barrier between the potential wells would have the effect of increasing the reversal rate. This simple change to the model could bring the reversal rate into better agreement with the time average and standard deviation of the VADM data sets.

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), where the error variances of the time domain data are “large”, and (d)~~but are quite different from the parameter estimates of configuration (a). Thus, we conclude that if the error variances~~ , where the error variance of the time domain data are ~~large, then~~ “infinite”. Thus, the impact of ~~these~~ the time domain data is minimal if the error variances of the time domain data are large. 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. We further note that the parameter estimates of configurations (c) or (d) are quite different from the parameter estimates of configuration (a) (see above). For an overall good fit of ~~model and *all* datasmall~~ the model to the spectral and time domain data, the error variances for the time domain data ~~are required or else the model yields reversal rates that~~ must by small, as in configuration (a). Otherwise, the reversal rates are too low. ~~Using small~~ Small error variances, however, imply (relatively) large deviations between the time domain data and the model predictions. Small error variances also comes at the cost of not necessarily realistic posterior variances.

Comparing configurations (d) and (e), we note that if only the spectral data are used, the reversal rates are unrealistically small (nominally 1 reversal per Myr). 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~~:~~. Specifically, our experiments suggest that a good match to spectral data requires a set of model parameters that is quite different from the set of model parameters that lead to a good fit to the reversal rate, time average VADM and VADM standard deviation. Experimenting with different functional forms for the drift term is one strategy for achieving better agreement between the reversal rate, the time average VADM and VADM standard deviation.

~~Finally, comparing~~ 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 configurations (a) and (b) above). The results, shown in Table 6, indicate that the parameter estimates based on configurations (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 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.

Finally, note that the data used in configuration (d) does not inform the parameter $\bar{x}$, and configuration (f) does not inform $\bar{x}$ or a. If the data do not inform the parameters, then the posterior distribution, over these parameters is essentially equal to the prior distribution, which is uniform. This is illustrated in Figure 10, where we show the corner plot of the posterior distribution of configuration (f). We can clearly identify the uniform prior in the marginals over the parameters $\bar{x}$ and a. This means that the Sint-2000 and PADM2M data only constrain the parameters D, γ and T_s.

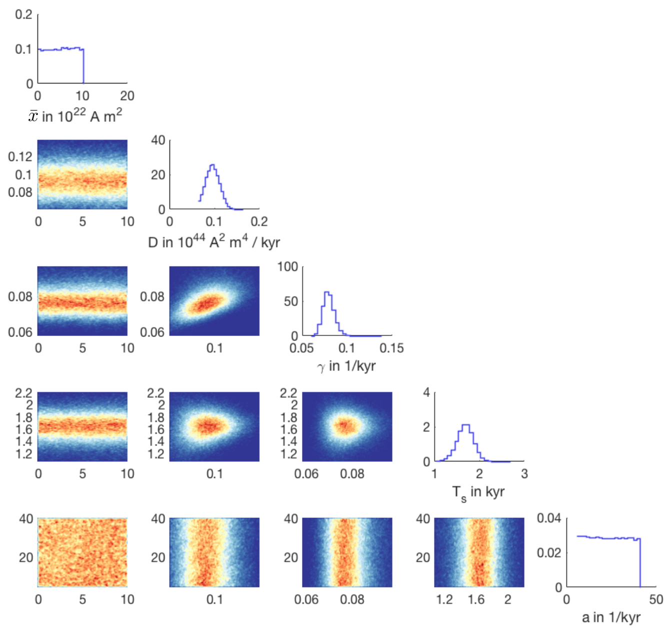

Figure 10. 1- and 2-dimensional histograms of the posterior distribution of configuration (f).

7 Examples of applications of the model

The Bayesian estimation technique we describe leads to a model with stochastic parameters whose distributions are informed by the geomagnetic data. Moreover, we ran a large number of numerical experiments to understand the limitations of the model, to discover inconsistencies of the model and to check our assumptions about error modeling. This process results in a well-understood and well-founded stochastic model for selected aspects of the long term behavior of the geomagnetic dipole field. We believe that such a model can be useful to test hypotheses about selected long-term aspects of the geomagnetic dipole.

For example, it was noted by Ziegler et al. (2011) that the VADM (time averaged) amplitude during the past chron was slightly lower than during the previous chron. Specifically, the time averaged VADM for $-0.78 < t < 0$ Myr is $E(x) = 6.2 \cdot 10^{22}$ Am^2 but for $-2 < t < -0.78$ Myr the time average is $E(x) = 4.8 \cdot 10^{22}$ Am^2. A natural question is: is this increase in the time average significant or is it due due to random variability? We investigate this question using the model whose parameters are the posterior mean values of configuration (a) (the configuration that leads to an overall good fit to all data). Specifically, we perform 10,000 simulations of duration 0.78 Myr and 10,000 independent simulations of duration 1.22 Myr. For each simulation, we compute the time average, which allows us to estimate the standard deviation of the difference in means (assuming no correlation between the two time intervals). We found that this standard deviation is about $0.46 \cdot 10^{22}$ Am^2, which is much smaller than the differences in VADM time averages of $1.4 \cdot 10^{22}$ Am^2. This suggests that increase in time averaged VADM is likely not due to random variability.

A good understanding of model limitations and inconsistencies further allows for selecting a different set of parameters for different "tasks" of the model. To study hypotheses about the reversal record, for example, one may want to de-emphasize the model fit to the spectral data (similar to configuration (a) or (d)). Any conclusions based on the model should then incorporate the fact that the model may not match other aspects of the field.

8 Summary and conclusions

We designed a Bayesian estimation problem for the parameters of a family of stochastic models that can describe the Earth's magnetic dipole over kyr and Myr time scales. The main challenge here is that the dataare 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 datasources for parameter estimation. The main tool A main advantage of our approach is that it can be used to discover inconsistencies between model and data. Once these inconsistencies are identified, several strategies can pursued to resolve these inconsistencies, e.g., improving the model or resolving consistency issues of the data themselves. The framework can also be used to reveal the impact that some of the individual data sets have on parameter estimates. The Bayesian approach we describe here leads to a stochastic SDE model with stochastic parameters whose distributions are informed by the geomagnetic data. Moreover, we can use our techniques to study the limitations and (remaining) uncertainties of the model. The result is thus a reliable, well-understood stochastic model for selected aspects of the long term behavior of the geomagnetic dipole field. We believe that such a model is useful for hypothesis testing and have given several examples of how the model can be used in this contextare . In this paper, we focus on describing the mathematical and numerical framework and only briefly mention some of the implications.

We formulated the Bayesian estimation problem in terms of "features" that we 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 modelThe features allow for synthesizing information from several, independent data sources that had previously been treated separately. Formulating the Bayesian estimation problem further requires defining model error, in particular defining model error variances. We found that the main source of uncertainty is the shortness of the geomagnetic record and constructed error models to incorporate this uncertainty. 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

With suitable error models, our numerical results indicate that these the data constrain all parameters of the modelmodel parameters in the sense that the posterior probability mass is concentrated on a smaller subset of parameters than the prior probability. Moreover, the posterior parameter values yield model outputs that fit the data in a precise, feature-based sense, which also translates into a good fit by other, more intuitive measures.

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 We ran a suite of numerical experiments . Further numerical experiments revealed the impact individual data sets have on parameter estimates. Our numerical experiments also suggest to assess the internal consistency of the data and the underlying model. We found that the model has a deficiency in that there are inconsistencies between the model 's spectra and its reversal rates on the Myr and kyr time scale is self-consistent, but we discovered inconsistencies that make it difficult to achieve a good fit to all data simultaneously. In particular, the model has difficulties to simultaneously match reversal rates and spectral data. 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.

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.

Acknowledgements. MM gratefully acknowledges support by the National Science Foundation under grant DMS-1619630 and by the Alfred P. Sloan Foundation.

References

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

Bärenzung, J., Holschneider, M., Wicht, J., Sanchez, S., and Lesur, V.: Modeling and Predicting the Short-Term Evolution of the Geomagnetic Field, Journal of Geophysical Research: Solid Earth, 123, 4539–4560, https://doi.org/10.1029/2017JB015115, 2018.

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, 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, 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 Geophysical Research: Solid Earth, 100, 6093–6095, 1995.

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 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, 263–314, 2001.

Hoyng, P., Schmitt, D., and Ossendrijver, M.: A theoretical analysis of the observed variability of the geomagnetic dipole field, Physics of Earth and Planetary Interiors, 130, 143–157, 2002.

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.

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.

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.

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.

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 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.

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.

Schmitt, D., Ossendrijver, M., and Hoyng, P.: Magnetic field reversals and secular variation in a bistable geodynamo model, Physics of Earth and Planetary Interiors, 125, 119–124, 2001.

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.

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

Ziegler, L., Constable, C., and Johnson, C.: Testing the robustness and limitations of 0-1 Ma absolute paleointensity data, Physics of Earth and Planetary Interiors, 170, 34–45, 2008.

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.

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 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 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

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 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).

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).

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 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 coefficients 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 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 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 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, 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 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). 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.

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 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 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 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.

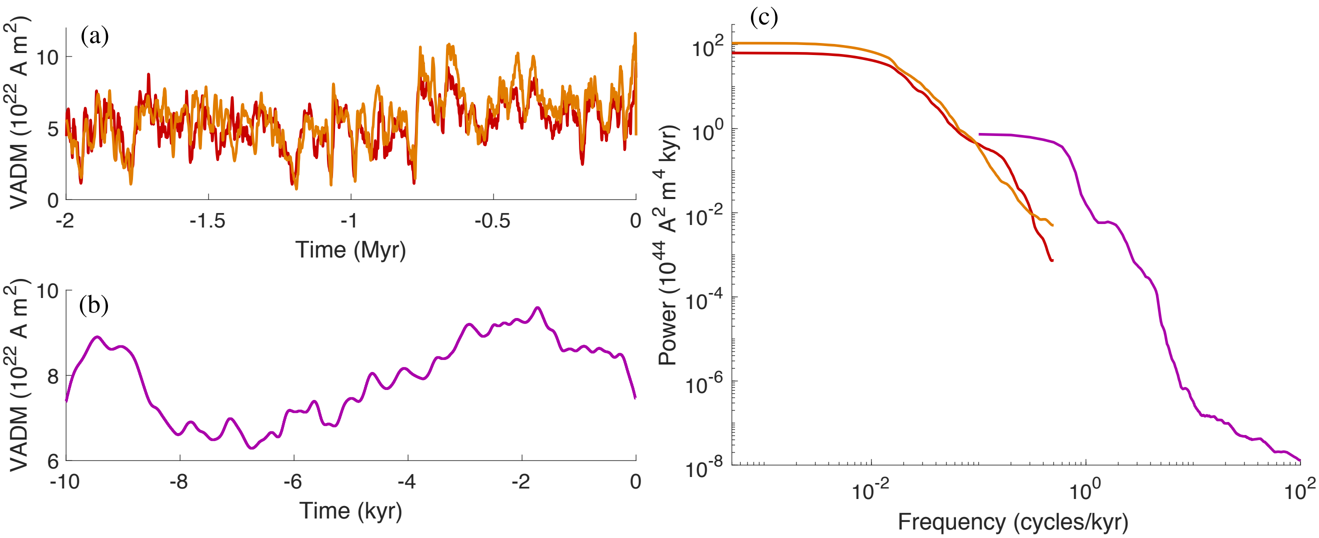

Figure 1. Data used in this paper. (a) Sint-2000 (orange) and PADM2M (red): VADM as a function of time over the past 2 Myr. (b) CALS10k.2: VADM as a function of time over the past 10 kyr. (c) Power spectral densities of the data in (a) and (b), computed by the multi-taper spectral estimation technique of Constable and Johnson (2005). Orange: Sint-2000. Red: PADM2M. Purple: CALS10k.2

Data source	Time avg. VADM (10^{22} Am^2)	Std. dev. of VADM (10^{22} 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). 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 Myr $-$ 5 Myr, 5 Myr $-$ 10 Myr, $\ldots$, 25 Myr $-$ 30 Myr. This leads to the average reversal rate and standard deviation listed in Table 2. Increasing the 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.

	$\bar{x}$ (10^{22} Am^2)	D (10^{44} A^{2}m^4 kyr^{-1})	γ (kyr^{-1})	T_s (kyr)	a (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

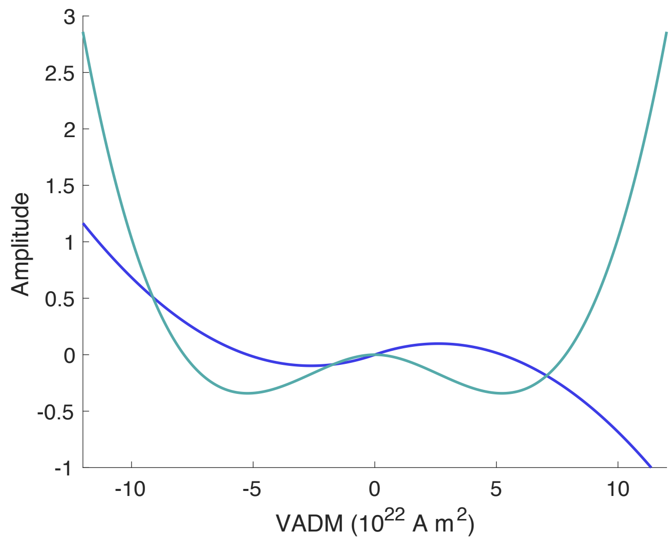

Figure 2. Drift coefficient $v(x)$ (blue) and potential $U'(x) = -v(x)$ (turquoise).

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

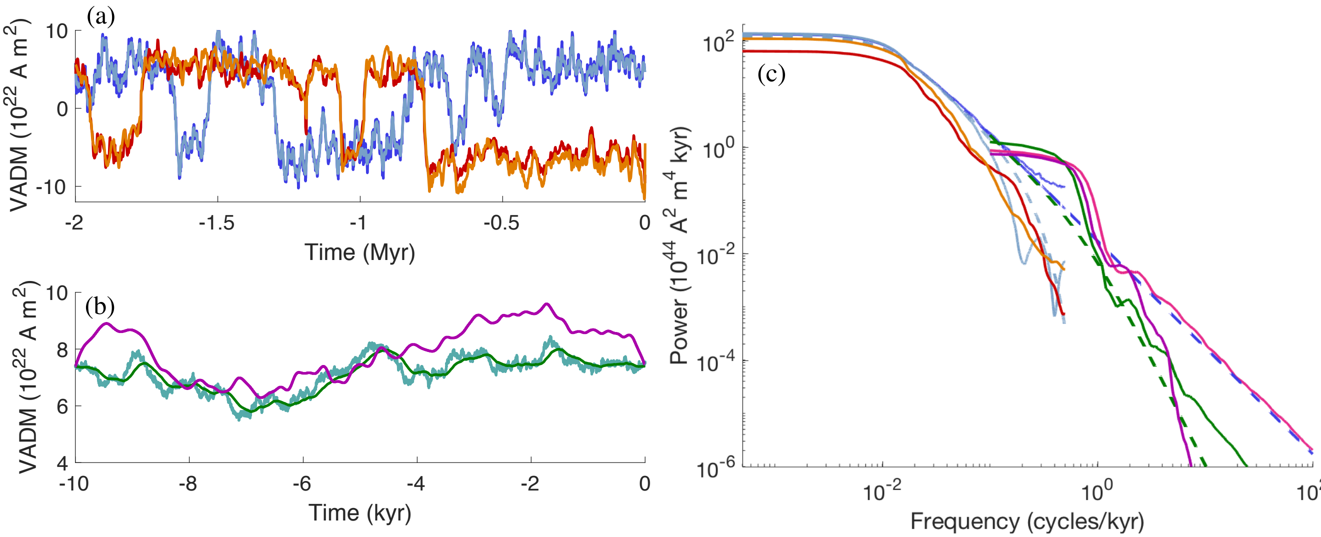

Figure 3. Simulations and data. (a) Output, x_j^{Myr} (dark blue, often hidden) and smoothed model output, $x_j^{\mathrm{Myr,s}}$ (light blue). Signed Sint-2000 (orange) and PADM2M (red), signs taken from Cande and Kent (1995). (b) Output of the kyr-model with uncorrelated noise (turquoise) and correlated 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 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) 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

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

where f is the frequency (in $1/\text{kyr}$). 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 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 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 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

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

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 $1/\text{kyr}$. 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 $\gamma \leq 0.7\ \text{kyr}^{-1}$.

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 $0.062 \cdot 10^{44}\text{A}^2\text{m}^4\text{kyr}^{-1} \leq D \leq 2.1 \cdot 10^{44}\text{A}^2\text{m}^4\text{kyr}^{-1}$.

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

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", 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 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-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).

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);

(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

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 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

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

We use the average of the time averages of PADM2M and Sint-2000 as the datum and set $y_{\bar{x}} = 5.56 \cdot 10^{22}$ 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.

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 $y_{\sigma} = 1.66 \cdot 10^{22}$ 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 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 $p_{l,\text{td}}(y|\theta)$ in (13) is then given by the product of the three likelihoods defined by equations (14), (16) and (17):

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

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 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 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

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

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

4.2.2 Low frequencies

The component $p_{l,\text{lf}}(y|\theta)$ 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

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

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

We define y_{lf} to be the mean of the PSDs of Sint-2000 and PADM2M. The function $h_{\text{lf}}(\theta)$ 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 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 $h_{\text{lf}}(\theta)$ 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:

$$h_{\text{lf}}(\theta) = \frac{2D}{\gamma^2 + 4\pi^2 f^2} \cdot \begin{cases} 1 & \text{if } f \leq 0.05 \\ \exp\left(-(4\pi^2 f^2 T_s^2)/12\right) & \text{if } 0.05 < f \leq 0.5 \end{cases} \tag{21}$$

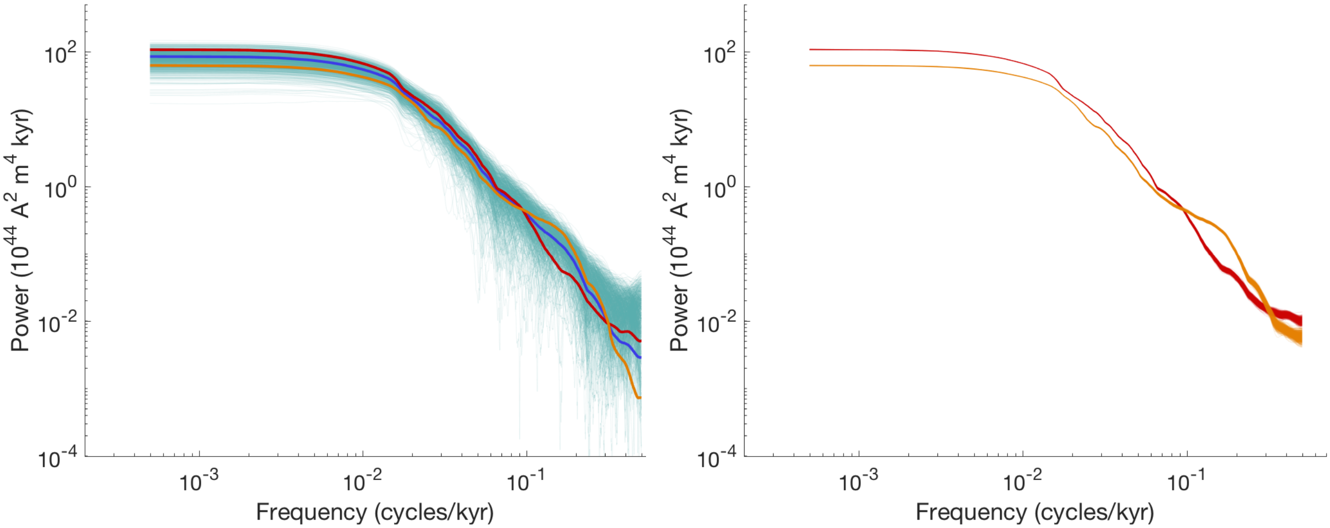

Figure 4. Left: low frequency data and error model due to shortness of record. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Blue: mean of PSDs of Sint-2000 and PADM2M (y_{lf}). Turquoise: 10^3 samples of the error model $\varepsilon_{\mathrm{lf}}$ added to y_{lf}. Right: error model based on errors 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 $h_{\mathrm{lf}}(\theta)$ 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 $\varepsilon_{\mathrm{lf}}$ 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 $\varepsilon_{\mathrm{lf}}$ 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 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 $0.9-9.9$ 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 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 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 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 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.

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

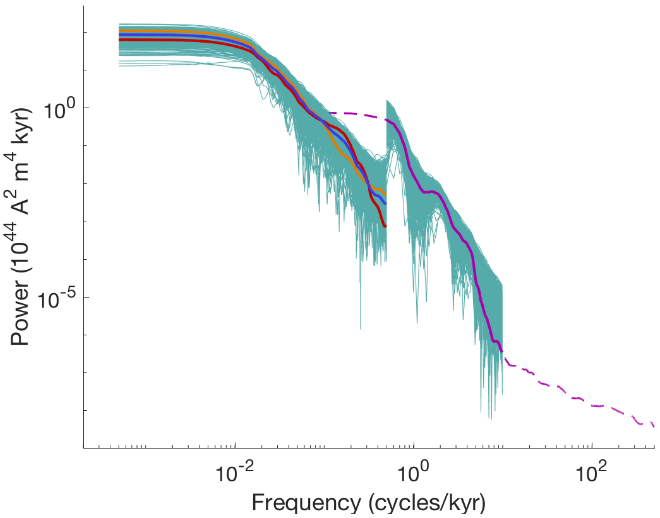

Figure 5. Data and error models for low and high frequencies. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Blue: mean of PSDs of Sint-2000 and PADM2M ($y_{\rm lf}$). Turquoise (low frequencies): 10^3 samples of the error model $\varepsilon_{\rm lf}$ added to $y_{\rm lf}$. Dashed purple: PSD of CALS10k.2. Solid purple: PSD of CALS10k.2 at frequencies we consider ($y_{\rm hf}$). Turquoise (low frequencies): 10^3 samples of the error model $\varepsilon_{\rm hf}$ added to $y_{\rm hf}$.

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 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 samples, but the number of "effective" samples is $10^6/\text{IACT}$. For all MCMC runs we perform (see Sections 5 and 6), the IACT of the Markov chain is about 100. We discard the first $10 \cdot \text{IACT}$ 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 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)

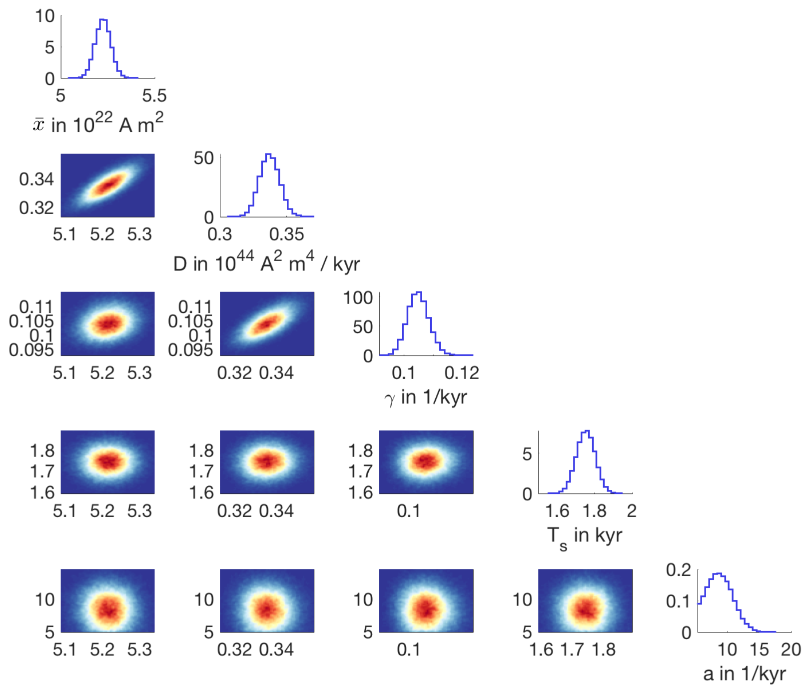

Figure 6. 1- and 2-dimensional histograms of the posterior distribution.

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 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 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

$\bar{x}$ (in 10^{22} Am^2)	D (in 10^{44} A^{2}m^4 kyr^{-1})	γ (in kyr^{-1})	T_s (in kyr)	a (in kyr^{-1})	σ (in 10^{22} Am^2)	Rev. rate (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 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 deviation (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.

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 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

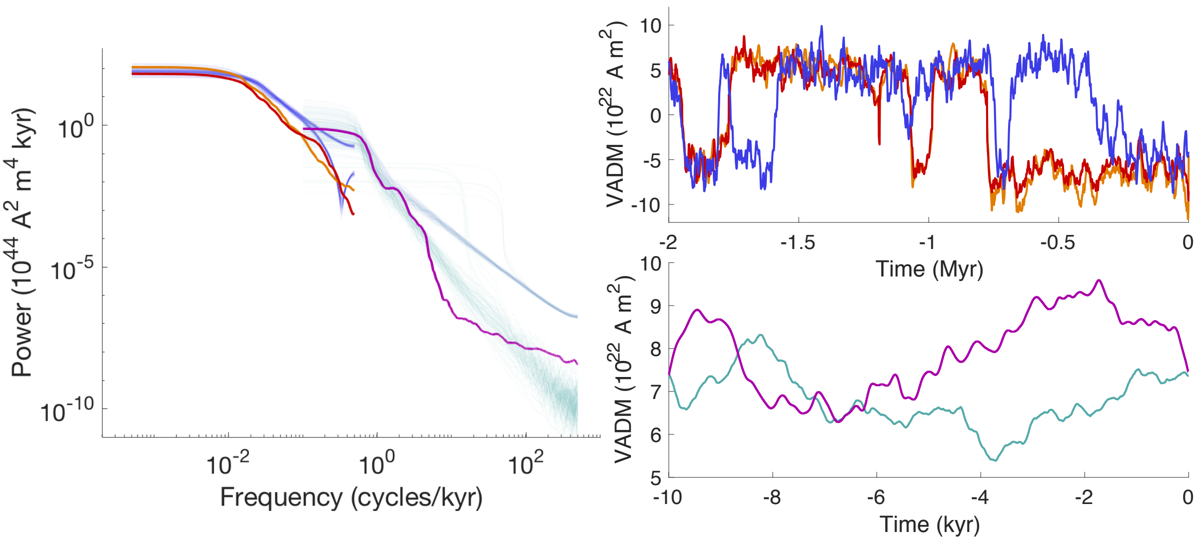

Figure 7. Parameter estimation results. Left: PSDs of data and model. Orange: PSD of Sint-2000. Red: PSD of PADM2M. Purple: PSD of CALS10k.2. Dark blue: PSDs of 100 posterior samples of Myr model (with and without smoothing). Light blue: PSDs of 100 posterior samples of kyr model with uncorrelated noise. 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 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.

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

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 $p_{l,\text{hf}}(y|\theta)$; since $p_{l,\text{lf}}(y|\theta)$ and $p_{l,\text{td}}(y|\theta)$ 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 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, 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 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 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 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 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 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 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 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.

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.

Acknowledgements. MM gratefully acknowledges support by the National Science Foundation under grant DMS-1619630 and by the Alfred P. Sloan Foundation.

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, 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, 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 Geophysical Research: Solid Earth, 100, 6093–6095, 1995.

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 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, 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.

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.

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.

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.

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 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.

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.

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.