An assessment of long duration geodynamo simulations using new paleomagnetic modeling criteria (Q_{PM})

Courtney J. Sprain[1*§], Andrew J. Biggin[1], Christopher J. Davies[2], Richard K. Bono[1], and Domenico G. Meduri[1]

[1] Geomagnetism Laboratory, Department of Earth, Ocean and Ecological Sciences, University of Liverpool, Liverpool L69 7ZE, UK

[2] School of Earth & Environment, University of Leeds, Leeds LS2 9JT, UK.

*Corresponding Author: csprain@ufl.edu

[§] Now at Department of Geological Sciences, University of Florida, Gainesville, Florida 32611, USA

Abstract:

Long-term temporal variations of the magnetic field (timescales > 10 Myr), characterized from paleomagnetic data, have been hypothesized to reflect the evolution of Earth's deep interior and couplings between the core and mantle. By tying observed changes in the paleomagnetic record to mechanisms predicted from numerical geodynamo simulations, we have a unique tool for assessing changes in the deep interior back in time. However, numerical simulations are not run in an Earth-like parameter regime and assessing how well they reproduce the geomagnetic field is difficult. Criteria have been proposed to determine the level of spatial and temporal agreement between simulations and observations spanning historical and Holocene timescales, but no such criteria exist for longer timescales.

Here we present a new set of five criteria (Quality of Paleomagnetic Modeling criteria, Q_{PM}) that assess the degree of semblance between a simulated dynamo and the temporal and spatial variations of the long-term (~ 10 Myr) paleomagnetic field. These criteria measure inclination anomaly, virtual geomagnetic pole dispersion at the equator, latitudinal variation in virtual geomagnetic pole dispersion, normalized width of virtual dipole moment distribution, and dipole field reversals. We have assessed 46 geodynamo simulations using the Q_{PM} criteria. The simulations have each been run for the equivalent of at least ~300 kyr, span reversing and non-reversing regimes, and include either homogeneous or heterogeneous heat flux boundary conditions. We find that none of our simulations reproduce all salient aspects of the long-term paleomagnetic field behavior for the past 10 Myr. Nevertheless, our simulations bracket Earth values, suggesting that an Earth-like simulation is feasible within the available computationally accessible parameter space. This new set of criteria can inform future simulations that aim to reproduce all aspects of Earth's long-term magnetic field behavior.

Keywords: Numerical geodynamo simulations, paleosecular variation, time average magnetic field, paleomagnetism

1. Introduction

The geomagnetic field has been a fundamental feature of the Earth for the past 3.5 billion years (Biggin et al., 2011, 2015; Tarduno et al., 2010) and may have been active since the Hadean (Tarduno et al., 2015). Originating in Earth's core and extending into space, the magnetic field shields the atmosphere from erosion by the solar wind, allowing for the preservation of liquid water on the surface and ultimately habitability (Tarduno et al., 2014). Thus, the geomagnetic field is a link between surface, interior, and exterior processes back through geologic time. The magnetic field observed at Earth's surface contains contributions primarily from internal sources. Temporal variations (secular variation) and spatial variations in the internally generated field can be characterized through direct observations using surface, satellite, and aeromagnetic measurements for the last few hundred years (historical record) and indirectly via paleomagnetic/archeomagnetic measurements going further back in time. These variations provide insight into the magnetohydrodynamic processes that occur in the outer core, and into how these processes may be affected by boundary conditions imparted at the mantle and inner core interfaces (Aubert et al., 2010; Olson et al., 2013). The similarity between the timescales observed for long-term variations in the magnetic field, e.g. the timescale for reversal frequency variability, to those of convective overturn in the mantle (200 Myr), has led to the hypothesis that long-term magnetic field variations are a result of external forcing mechanisms and reflect the evolution of Earth's deep interior (Biggin et al., 2012; Jones, 1977; McFadden and Merrill, 1984). If observed variations in the long-term magnetic field can be tied to mechanisms predicted from numerical geodynamo simulations, it would then be possible to evaluate changes in the deep interior going back in geologic time, adding a crucial dimension to our understanding of Earth's evolution.

In the last three decades, significant advances have been made in the field of numerical geodynamo modeling. These simulations have succeeded in capturing the main features of the Earth's magnetic field, such as a dipole dominated field and polarity reversals (e.g. Christensen and Wicht, 2015; Glatzmaier and Coe, 2015; Glatzmaier and Roberts, 1995), in addition to aspects of historical secular variation (Bloxham, 2000; McMillan et al., 2001) such as westward drift (Christensen and Olson, 2003) and weak activity in the Pacific hemisphere (e.g. Aubert et al., 2013; Davies et al., 2008; Gubbins et al., 2007; Mound et al., 2015). Furthermore, simulations have been used to make predictions about magnetic field behavior including estimates of internal field strength and core flow speed (Christensen et al., 2009; Christensen and

Aubert, 2006), relations between field strength and reversal frequency (Olson, 2007), the role of core-mantle boundary heat flow in affecting field behavior and flow dynamics (Amit et al., 2015; Amit and Olson, 2015; Olson and Christensen, 2002; Olson et al., 2010) and time-average field morphology (Amit et al., 2015; Amit and Choblet, 2009; Davies et al., 2008; Gubbins et al., 2007). However, due to computational limitations, numerical dynamo simulations cannot yet run with the small diffusion coefficients that characterize the core fluid. In terms of non-dimensional numbers, the Ekman number E (the ratio of viscous to Coriolis forces) and the magnetic Prandtl number Pm (the ratio of viscous to magnetic diffusion) are many orders of magnitude larger than those estimated for Earth. Lowering E to geophysical values of 10^{-15} is the main challenge. Recent simulations that utilized millions of CPU hours have reached $E = 10^{-7}$ (Schaeffer et al., 2017) or 10^{-8} by parameterizing the smallest scales of the turbulence (Aubert et al., 2017), but were only run for short time periods and do not reverse. A key issue is then to determine to what degree a given simulation can be said to exhibit 'Earth-like' properties.

To assess whether a numerical dynamo simulation produces an Earth-like magnetic field, past studies have utilized observed behavior of the recent geomagnetic field derived from global time-dependent field models spanning historical and Holocene timescales, to develop criteria that can be used to assess the similarity between numerical simulations and Earth (Amit et al., 2015; Christensen et al., 2010; Davies and Constable, 2014; Mound et al., 2015). These global field models are constructed from satellite, observatory and survey magnetic field observations over the historical period 1590-1990 AD (gufm1; Jackson et al., 2000) and from archeomagnetic and paleomagnetic data, collected from archeological artifacts, sediments, and volcanic rocks for the past 10 to 100 kyr (e.g. Korte and Constable, 2011; Panovska et al., 2018). Existing criteria utilize large-scale properties of the field morphology (Christensen et al., 2010; Mound et al., 2015; Amit et al., 2015) or the frequency content of the dipole moment time-series (Davies and Constable, 2014) derived from these global field models as the basis for assessing whether a numerical simulation reproduces Earth's magnetic field behavior. In practice, these criteria have been used to assess the compliance of both short (< 10^5 yr) and long duration (~10^5-10^7 yr) dynamo simulations with Earth-like behavior (e.g. Driscoll and Wilson, 2018), despite being based on features of the recent geomagnetic field. While it has been suggested that modern secular variation, as captured by global field models, is representative of expected variations over the entire history of the geodynamo, this is fundamentally uncertain (Johnson and McFadden, 2015). Furthermore, current criteria based on time-dependent field models do not include aspects of Earth's long-term magnetic field behavior not observed in the Holocene, such as polarity reversals. To properly assess whether simulations behave like Earth on longer time scales (>10^5-10^7 yrs), we need to define a new set of criteria which can be used to assess how well numerical simulations reproduce paleomagnetic field behavior.

Here we present a new set of criteria to compare long-term behavior of numerical dynamo simulations with paleomagnetic observations: the Quality of Paleomagnetic Modeling (Q_{PM}) criteria. Criteria are assessed using a two-fold approach: 1) the calculation of a non-parametric misfit score (ΔQ_{PM}^{i}) between simulated and Earth data, inspired by the approach used in Christensen et al. (2010), and 2) the assignment of a binary score (Q_{PM}^{i}), inspired by the paleomagnetic Q (Van der Voo, 1990) and Q_{PI} (Biggin and Paterson, 2014) approaches commonly used in the assessment of paleodirectional and paleointensity studies, respectively. Total misfit values, ΔQ_{PM}, and total Q_{PM} scores are evaluated over all criteria, where for each criterion that is met the total Q_{PM} score increases by 1, to a maximum score of five. The utility of a two-fold method is as follows. First, this approach helps bring all simulations, regardless of the parameter space in which they were run, to the same baseline, easing comparison between them. Second, the ΔQ_{PM} helps to quantify overall how close a simulation is to reproducing Earth's paleomagnetic behavior, while the Q_{PM} score highlights which specific aspects of the paleomagnetic field a simulation is reproducing well. Finally, the Q_{PM} approach does not prescribe a strict threshold below which a simulation is deemed incompatible with paleomagnetic observations, which allows users to assess which of the paleo-field properties are most important to reproduce for their study. This permits users to get the most out of their simulations, which for timescales on the order of 1 Myr may have taken tens of thousands of CPU hours to run.

The chosen five criteria represent a range of commonly reported paleomagnetic observables that reflect temporal and spatial variations in the long-term magnetic field. Global time-dependent field models are not available for the timescales of interest here and our new criteria reflect the available data in the paleomagnetic record. For the purpose of this study, their Earth-like values are derived for the past 10 Myr as reported in the recent compilation of paleomagnetic directional data, PSV10 (Cromwell et al., 2018), and the paleointensity (PINT) database (Biggin et al., 2009, 2015) (Table 1). These criteria are assessed at Earth's surface, requiring conversion of Gauss coefficients from geodynamo simulations into pseudo-paleomagnetic data. The five criteria address different aspects of the time-average and time-varying field and are as follows: inclination anomaly, virtual geomagnetic pole (VGP) dispersion at the equator, latitudinal variation of VGP dispersion, normalized width of virtual dipole moment (VDM) distribution, and dipole field reversals. We have assessed the compliance of our criteria with a large number of published (Davies et al., 2008; Davies and Constable, 2014; Davies and Gubbins, 2011; Gubbins et al., 2007), and new long-duration geodynamo simulations. These simulations span a wide parameter space that was chosen to best capture a broad range of simulation behavior and serves to demonstrate the Q_{PM} approach. Because we cannot predict *a priori* which simulations will reproduce Earth's paleomagnetic field, we have chosen to explore this parameter space systematically.

In the following sections we will first review the observable properties of the paleomagnetic field that will be used as the foundation of the Q_{PM} criteria and introduce the five Q_{PM} criteria. We then outline how compliance with these criteria is met. Next, we use these criteria to assess how well our suite of 46 long-duration geodynamo simulations reproduce the Earth's paleomagnetic field. We close with a discussion of the implications of our results, and how we foresee the utilization of these criteria in the future.

2. Paleomagnetic Modeling Criteria for Geodynamo Simulations (Q_{PM})

In order to be effective, the criteria for assessing a numerical simulation should be objective and quantifiable (Christensen et al., 2010), and address a well-established property of the paleomagnetic field. We base our criteria solely on paleomagnetic observables and not on global time-dependent models (e.g. Panovska et al., 2018), which do not cover the time-frame of interest (>100 kyr), or statistical field models (e.g. Tauxe and Kent, 2004), which fail to reproduce paleomagnetic observations of paleosecular variation (PSV) and time-average field (TAF) behavior, or TAF models (e.g. Cromwell et al., 2018), which do not represent PSV. Observations made directly from paleomagnetic datasets provide the most reliable representation of long-term magnetic field behavior and we therefore use them as the foundation of our criteria.

Another constraint on viable criteria arises from limitations inherent in current dynamo simulations. Simulations spanning paleomagnetic timescales must run for long periods, which increases the computational cost and further limits the parameter range that can be accessed. We therefore focus on large-scale features of the field as has been done in previous studies (Christensen et al., 2010; Davies and Constable, 2014; Mound et al., 2015; Wicht and Meduri, 2016).

2.1 Paleomagnetic Basis for Q_{PM} Criteria

To assess the behavior of the magnetic field on long time scales we are interested in both the geometry of the TAF in addition to temporal variations about the long term average (PSV). An overview of standard paleomagnetic observables is presented in the Supplementary Materials. Full vector records of the paleomagnetic field are sparse and unevenly reported. Therefore, in defining the Q_{PM} criteria, magnetic directions and intensity are treated separately. The five criteria chosen for Q_{PM} analysis discussed below are as follows: inclination anomaly (IncAnom), VGP dispersion (VGPa and VGPb), normalized width of VDM distribution (VDMVar), and reversals (Rev).

2.1.1 Criterion Based on Time-Average Field Behavior

A fundamental assumption in paleomagnetic studies is that over a sufficiently long time period the field can be best approximated by a geocentric axial dipole (GAD), where inclination (I_{GAD}) is predicted to vary with latitude (λ) via the axial dipole equation

$$\tan I_{GAD} = 2\tan\lambda. \qquad (1)$$

However, paleomagnetic records show small, yet persistent, deviations from GAD (Cox, 1975; Cromwell et al., 2018; Johnson et al., 2008). The two parameters used to represent this offset in paleomagnetic studies are inclination anomaly and declination anomaly, defined as

$$\Delta I = \bar{I} - I_{GAD}\ , \qquad (2)$$

$$\Delta D = \bar{D}. \qquad (3)$$

Here, $\overline{I}$ and $\overline{D}$ are the calculated Fisher mean (Fisher, 1953) inclination and declination values from measured samples. Note, the declination predicted from a GAD field is zero. Due to large gaps in spatial coverage in long-term paleomagnetic records, investigations are restricted to latitudinal structure only. In observational datasets, an accurate measure of ΔD is much harder to capture from paleomagnetic data than ΔI, due to error in or absence of sample orientation, unrecognized tectonic rotation, and the expected long-term behavior of the longitudinal variation of the non-GAD field as captured by declination. Therefore, for our criteria we utilize ΔI (**IncAnom**), as one of our measures of TAF behavior.

The IncAnom criterion utilizes the maximum absolute median ΔI (calculated from 10° latitude bins) and its 95% confidence intervals as a measure of TAF behavior. We do not require that simulations match the observed latitudinal geometry of ΔI, which we believe is justified since the latitudinal variation of ΔI is not well-constrained in the long-term magnetic field (Cromwell et al., 2018).

A measure of the mean field intensity or mean VDM (see equation 8 below) is most often used as a metric of TAF intensity. Dynamo simulations solve dimensionless equations and so scaling the results into a dimensional field strength is non-unique. Davies and Constable (2018) found that estimates of the local field intensity varied by a factor of 2-3 between two different magnetic field scalings within a given geodynamo simulation. Additionally, the ratio between the Elsasser and Lehnert number scalings commonly used in the literature is $\left(\frac{E}{Pm}\right)^{1/2}$ (Olson and Christensen, 2006), which at $E = 10^{-4}$ and $Pm = 1$ could produce a factor of 100 or more difference in the field strength. Due to these complexities, we do not include a direct measure of TAF behavior in regard to intensity in our Q_{PM} criteria.

2.1.2 Criteria Based on Paleosecular Variation Behavior

VGP angular dispersion (S) is a commonly used metric to quantify paleosecular variation in the long-term paleomagnetic field. Using VGP dispersion allows for the estimation of paleosecular variation when detailed age control and time series data are unavailable. To mitigate the latitudinal dependence of magnetic field direction, a standard approach in paleomagnetism is to calculate the geocentric dipole that would give rise to the observed site directions, where the VGP is the position that the dipole pierces Earth's surface (cf. Butler, 1992). Here we use the paleomagnetic definition of sites, which are assumed to capture individual snapshots of the magnetic field, i.e., a single cooling unit. The dispersion about a mean pole S, from a set of n VGPs contained in a locality or latitude band, can then be determined as an estimate of paleosecular variation, where

$$S = \left[\frac{1}{n-1}\sum_{i=1}^{n} \Delta_i^2\right]^{\frac{1}{2}}. \tag{5}$$

Here, Δ_i is the angular distance of the ith VGP from the geographic pole or mean VGP. Note, for paleomagnetic data, S would further be corrected to remove within site dispersion due to random errors in measuring and sampling (e.g., Cromwell et al., 2018).

It has been observed that S varies as a function of latitude, for which various explanations have been hypothesized (cf. Merrill et al., 1996). The phenomenological Model G of McFadden et al. (1988) is often used to approximate the latitudinal variation of VGP dispersion where S is described as a function of (paleo)latitude and two parameters, a and b,

$$S^2 = a^2 + (b\lambda)^2 . \tag{6}$$

Here, a and b are argued to represent variations in the equatorially symmetric and equatorially anti-symmetric spherical harmonic decomposition of the field, respectively. The a and b parameters are calculated by a least squares fit between the measured VGP dispersion curve and that determined by Model G.

For our criteria we have chosen to apply the quadratic fit, as defined by Model G, as a metric of PSV behavior, with a and b parameters defining separate criteria (**VGPa** and **VGPb)**. We treat the compliance with the minimum (equatorial) dispersion, a, and the latitude dependence, b, separately in our framework, since these characterize different aspects of field variability.

The input simulated data for VGP dispersion are S values calculated after using a Vandamme cutoff for both Earth data and simulated outputs (S_{VD}; Vandamme, 1994). The Vandamme cutoff helps to exclude anomalous VGP data, with the intention of preventing bias in the dispersion estimate from magnetic excursions or reversals. The Vandamme cutoff is not constant, but instead is allowed to vary as follows

$$\lambda_{cut} = 90° - (1.8S + 5°), \tag{7}$$

where S is calculated from the simulated data. Sites with VGP latitudes less than λ_{cut} are excluded, S is recomputed, and the procedure is repeated until all remaining VGPs are within the cutoff angle. The final S value is then noted as S_{VD}.

Like magnetic directions, the intensity of the magnetic field is also latitudinally dependent. To remove this dependence, a VDM is calculated, which is the strength of the geocentric dipole that produces the observed field intensity F, at a given paleolatitude

$$VDM = \frac{4\pi {r_e}^3}{\mu_o} F(1 + 3\cos^2\theta_m)^{-\frac{1}{2}}, \tag{8}$$

where r_e is radius of Earth’s surface, θ_m is the magnetic colatitude calculated using the mean inclination and the axial dipole equation (1), and μ_0 is the permeability of free space.

To provide an estimate of temporal variation in magnetic intensity, in this study we chose to measure the variability of a distribution of VDMs (**VDMVar**), through

$$V\% = \widehat{VDM}/VDM_{med}, \tag{9}$$

where $\widehat{VDM}$ is the interquartile range of a distribution of VDM values, and VDM_{med} is the corresponding median. The VDMVar criterion is passed if the $V\%$ calculated from simulated data falls within the range estimated for Earth.

2.1.3 Criteria Based on Other Paleomagnetic Observables

The final criterion assesses dipole field reversals (**Rev**). The Rev criterion is met if a simulation reverses in an Earth-like manner. While reversals are a fundamental feature of Earth's magnetic field, an agreed formal description remains elusive. Here, we define a set of standards that we think faithfully represent the fundamental characteristics of geomagnetic field reversals. To pass this criterion a simulation must: a) exhibit at least one reversal in the dipole field after the initial transient period, b) result in a new stable direction, and c) the proportion of time spent in a transitional state is within the range calculated for Earth. For our simulations, we estimated the first two standards by first calculating τ_n, the relative proportion of time spent with a normal polarity (i.e., the time spent with true dipole latitudes >45° divided by the total simulation time), τ_r, the relative proportion of time with a reverse polarity (i.e. the time spent with true dipole latitudes <45° divided by the total simulation time), and τ_t, the relative proportion of time spent in transitional periods (i.e. the time spent with true dipole latitudes between 45° and -45° divided by the total simulation time). A simulation passes the first two requirements if both τ_n and τ_r are greater than τ_t. Finally, if the calculated τ_t for a simulation falls within the range estimated for Earth, the simulation passes the Rev criterion.

2.2 Acceptance Thresholds Based on Earth Values for the Past 10 Myr

Establishing acceptance thresholds for Q_{PM} criteria that are representative of Earth's long-term magnetic field behavior is non-trivial. Ideally, the values for the established criteria should be representative of the paleomagnetic field for all of Earth's history. However, it has been hypothesized that PSV and the TAF structure are dependent on conditions at the core-mantle boundary (CMB), and therefore are expected to be variable throughout geologic time (Jones et al., 1977). For the purpose of this study, we consequently chose to focus on the PSV and TAF structure of the paleomagnetic field for the last 10 Myr. This time period was chosen because paleomagnetic data for the last 10 Myr provide sufficient temporal and spatial coverage to enable global analysis, and are additionally young enough to not be strongly affected by plate motion and changing CMB conditions. For the assessment of TAF and PSV behavior for the past 10 Myr we utilized two datasets, PSV10 (Cromwell et al., 2018) and the PINT database (Biggin et al., 2009). For an assessment of reversal behavior for Earth for the past 10 Myr we utilized the 2012 Geomagnetic Polarity timescale (Ogg, 2012). Acceptance thresholds based on Earth values for our chosen criteria as measured are reported in Table 1. A description of the datasets and how specific criteria were estimated is presented in the Supplementary materials.

2.3 Rating Compliance with the Paleomagnetic Field

To rate the compliance of the numerical simulation output with long-term magnetic field behavior we first define a misfit parameter for each criterion, ΔQ_{PM}^{i}, where *i* denotes the five criteria VGPa, VGPb, Rev, VDMVar, and IncAnom. We chose to use this method because it is non-parametric, as the distribution of paleomagnetic data is not well-constrained. Here, ΔQ_{PM}^{i} is calculated by

$$\Delta Q_{PM}^{i} = \frac{|m_{Earth}^{i} - m_{Sim}^{i}|}{\sigma_{Earth}^{i} + \sigma_{Sim}^{i}} . \qquad (10)$$

This parameter is the ratio of the absolute distance between the median Earth value (m_{Earth}^{i}) for a given criterion, i, and the median value estimated from the simulated data (m_{Sim}^{i})

to the total distance covered by the uncertainty bounds (measured as 95% confidence intervals) that lie between Earth and the simulated data ($\sigma^i_{Earth} + \sigma^i_{Sim}$). E.g., if $|m^i_{Earth}| > |m^i_{Sim}|$, then σ^i_{Earth} would be the lower 95% confidence bound for Earth and σ^i_{Sim} would be the upper 95% confidence bound for the simulated data, and vice versa when $|m^i_{Earth}| < |m^i_{Sim}|$. If $\Delta Q^i_{PM} \leq 1$, then the simulation passes the criterion and the Q^i_{PM} score for that criterion is set to 1, otherwise the Q^i_{PM} score is set to 0.

Once each criterion is assessed, the total misfit ΔQ_{PM} and the total Q_{PM} score can be calculated as

$$\Delta Q_{PM} = \sum_{i=1}^{5} \Delta Q^i_{PM.} \tag{11}$$

and

$$Q_{PM} = \sum_{i=1}^{5} Q^i_{PM,} \tag{12}$$

respectively. If $\Delta Q_{PM} \leq 5$ and Q_{PM}= 5, then a simulation meets all set criteria.

3. Methods

3.1 Geodynamo Simulations

The geodynamo simulations parametrization and solution methods used in this study have been extensively documented elsewhere (Davies and Constable, 2014; Davies and Gubbins, 2011; Willis et al., 2007) and so only a brief description is given here. An incompressible Boussinesq fluid is confined within a spherical shell of width $d = r_o - r_i$, where r_i and r_o are the inner and outer boundary radii respectively, rotating about the vertical direction at an angular frequency Ω. The system is thermally driven and the Boussinesq approximation is employed so that density variations are accounted for only in the buoyancy force. The fluid has a constant kinematic viscosity ν, thermal diffusivity κ, thermal expansivity α, and magnetic diffusivity $\eta =$

$(\sigma\mu_0)^{-1}$, where σ is the electrically conductivity. The shell aspect ratio is fixed to r_i/r_o= 0.35 in this study and Prandtl number ($Pr = \frac{\nu}{\kappa}$) is set to 1. The following parameters control the system;

$$E = \frac{\nu}{2\Omega d^2}\,, \tag{13}$$

$$Pm = \frac{\nu}{\eta}\,, \tag{14}$$

$$Ra = \frac{\alpha g \beta d^2}{2\Omega\kappa}\,. \tag{15}$$

Here, g is gravity, Ra is the modified Rayleigh number, and β/d is the amplitude of the prescribed temperature gradient at the outer boundary. The solution consists of the magnetic field $\boldsymbol{B}$, fluid velocity $\boldsymbol{u}$, and temperature T throughout the spherical shell and at each time point.

All simulations employ no-slip boundary conditions, that is $\boldsymbol{u} = \boldsymbol{0}$ at r_i and r_o. For the magnetic field, the top and bottom boundaries are insulating. Therefore, above the core region the magnetic field is represented by a potential field that matches to the dynamo solution at r_o. Fixed heat flux is prescribed at r_o in all simulations (denoted FF), while FF or fixed temperature (FT) conditions are applied at r_i. Some simulations additionally employ lateral variations in heat flow at r_o. Here the pattern is either derived from the seismic shear-wave velocity model of Masters et al. (1996) or a recumbent Y_2^0 heat flux pattern is used as an approximation to the observed shear-wave structures (Dziewonski et al., 2010). The amplitude of the heat flow anomalies is defined by the parameter $\epsilon = (q^{max} - q^{min})/q^{ave}$, where q^{max}, q^{min} and q^{ave} are the maximum, minimum and average heat flow on the outer boundary. We consider values of $\epsilon = 0.3 - 1.5$ (Table 2) and note that the largest values do not conflict with the Boussinesq approximation (see Mound and Davies, 2017).

In our suite of simulations, 10 have been reported in previous studies (Davies et al., 2008; Davies and Constable, 2014; Davies and Gubbins, 2011; Gubbins et al., 2007) (Table 2). Three of these simulations were integrated further here [Model 2 (B2), Model 3 (B4), Model 8 (B3)] in addition to 36 new simulations. The parameter regime explored in these simulations is as follows: $E = 10^{-3} - 1.2 \times 10^{-4}$, Rayleigh numbers ranging from 20-450 corresponding to roughly 1-100 times the critical value for onset of non-magnetic convection, and magnetic Prandtl numbers ranging between 2 and 20 (Table 2). All simulations were run for ~3-30 outer core magnetic diffusion times, or the equivalent of a minimum of about 300 kyr – 3 Myr using the electrical conductivity value of 3 x 10^5 $\mathrm{S/m}$ from Stacey and Loper (2007).

3.2 Q_{PM} Criteria Calculation Protocol

For the assessment of Q_{PM} criteria, Gauss coefficients up to spherical harmonic degree l_{max}= 10 were calculated at Earth's surface for each simulation. From the truncated data, we generated simulated values of declination (D), inclination (I), and intensity (F) using a spherical harmonic expansion, where V is the magnetic scalar potential and $\boldsymbol{B} = -\nabla V$, defined according to

$$V(r,\theta,\varphi) = r_e \sum_{l=1}^{l_{max}} \sum_{m=0}^{l} \left(\frac{r_e}{r}\right)^{l+1} (g_l^m \cos m\varphi + h_l^m \sin m\varphi) P_l^m(\cos\theta), \qquad (16)$$

$$I = \tan^{-1}\left(\frac{-B_r}{\left({B_\theta}^2 + {B_\varphi}^2\right)^{1/2}}\right), \qquad (17)$$

$$D = \tan^{-1}\left(\frac{B_\varphi}{-B_\theta}\right), \qquad (18)$$

$$F = \sqrt{{B_r}^2 + {B_\theta}^2 + {B_\varphi}^2}\ . \qquad (19)$$

Here, r, θ, and φ are spherical coordinates (radius, colatitude, and longitude), P_l^m are the Schmidt-normalized associated Legendre functions of degree l and order m, and g_l^m and h_l^m are the Gauss coefficients.

For the assessment of PSV and the TAF behavior, we chose to downsample our simulations to mimic the spatial and temporal coverage of real data present within PSV10, thereby mitigating against potential biases due to uneven spatial and temporal sampling. To do this, simulations were downsampled to each of the 51 modified PSV10 localities (see Supplementary materials, Table S1). At each locality, N random time-steps were chosen from the simulation, where N is equal to the number of sites at that locality. Values for D, I, F, VGP latitude, VGP longitude, and VDM for that time-step at that locality were then calculated as per standard paleomagnetic methods (Eqns. 16-19 and 8, respectively, for VGP latitude and longitude see Butler, 1992). Simulated data were normalized to the same polarity.

From these parameters, ΔI, a, b, and $V\%$ were calculated as described in section 2. To address the potential for statistical variation we repeated the downsampling procedure 10,000 times, from which 95% confidence intervals were estimated for each calculated parameter.

4. Results

In our Q_{PM} assessment of 46 geodynamo simulations we find that no simulation successfully reproduces all observed features of the paleomagnetic field. Total Q_{PM} scores for the 46 geodynamo simulations are in the range from 0 to 3, out of a maximum score of five, with a median score of 1 (Fig. 1a). The VGPb criterion had the highest pass rate of 63%, followed by the IncAnom criterion at 35%, VDMVar at 20%, VGPa at 7%, and ending with Rev at 4% (Fig. 1b). Of the 46 simulations assessed, 22 reversed, but only two had τ_t values within the range for Earth, thus passing Rev. The VGPa criterion was only met by three simulations (Fig. 1b), and none of these simulations reversed (Table 3). Of the four simulations that had Q_{PM} = 3, all passed VGPb and IncAnom, three passed VDMVar, one passed Rev, and none passed VGPa.

Representative examples of simulations that pass or fail each criterion are presented in Fig. 2 (Rev), Fig. 3 (IncAnom), Fig. 4 (VGPa and VGPb), and Fig. 5 (VDMVar). All assessed simulation results are presented in Supplemental Figures S1-3. Values for all calculated Q_{PM} parameters are given in Supplemental Table S2 and Q_{PM} results are in Table 3.

Total misfit values, ΔQ_{PM}, for all 46 simulations range from 5.6 to 22.2, with a median value of 10.5. VGPa had the highest median misfit value of 3.4 (Fig. 1c). In a majority (74%) of simulations, misfit values for VGPa were higher than for any other criterion (Fig. 1c). The distribution of ΔQ_{PM} reveals no correlation between total Q_{PM} score and ΔQ_{PM} (Fig. 1d). This lack of correlation clearly highlights that none of our simulations are simultaneously reproducing all aspects of Earth's long-term field behaviour; if a simulation is reproducing some aspects of the paleomagnetic field behavior (highlighted by Q_{PM} scores of 2 or 3), often it is very far from reproducing a different aspect (evidenced by high ΔQ_{PM} values). In the majority of cases with high Q_{PM} scores and high ΔQ_{PM} (74%), the parameter with the highest misfit ($\gg 1$) is VGPa.

The distributions of simulated values for each Q_{PM} criterion generally display two peaks that fall to either side of Earth values for the last 10 Myr (Fig. 6). For most criteria, the simulations fail to pass because simulated values were higher than Earth (except for VGPb, where the latitude dependence of VGP dispersion is equally under or over represented relative to Earth). Furthermore, for each criterion, simulations showing reversals had higher simulated values than those that did not reverse, with the highest values obtained for simulations with $\tau_t > 0.15$. Reversing simulations show high VGP dispersion and ΔI, but Earth-like $V\%$ values. In general, non-reversing simulations have lower VGP dispersion and high ΔI, and often insufficient variation in field strength to pass the VDMVar criterion (Fig. 6). No reversing simulations passed the VGPa criterion, with calculated values higher than those observed for

Earth (Fig. 6). In general, positive correlations are observed between calculated values for $V\%$ – ΔI, $V\%$– b, $a - b$, $\tau_t - a$, and $\Delta I - a$ (Supp Fig. S4), forming a quasi-linear trend that contains Earth.

No universal trends between Q_{PM} or ΔQ_{PM} and input parameters for the simulations assessed in this study were identified. In general, the application of inhomogeneous boundary conditions pushed simulations further from Earth, as reflected in increased ΔQ_{PM} values as ϵ increases (Table 3). However, this trend only applies when the application of an inhomogeneous boundary condition resulted in a reversing simulation with $\tau_t > 0.15$. In the case where a simulation with an inhomogeneous boundary condition remained non-reversing, there are small changes in the calculated parameters (Table S2), which results in a lower misfit score with increasing ϵ. Future work will need to be conducted to further determine the effects of heterogeneous boundary conditions on long-term field behavior. In general, there is a positive trend between the magnetic Reynolds number ($Rm = \ Ud/\eta$, where U is the time-averaged RMS flow amplitude) and all calculated parameters utilized for Q_{PM} assessment (Supp. Fig. S5).

Plotting our simulation results as a function of magnetic Ekman number ($E_\eta = E/Pm$) and Rm shows that many of our simulations fall within the wedge-shaped region of Christensen et al. (2010) for simulations with FF boundary conditions (Fig. 7). However, conformance with Earth's long-term field behavior for simulations that fall within the wedge is not assured as Q_{PM} scores within the wedge range from 0 to 3 and ΔQ_{PM} values range from ~6 to 22. Furthermore, many of our simulations that performed relatively well, with ΔQ_{PM} less than 10, fall outside of the wedge.

5. Discussion

5.1. Limitations of Q_{PM} Approach

One limitation of the presented Q_{PM} criteria is that only data from the past 10 Myr are used to calculate values for Earth's TAF and PSV behavior and are not necessarily representative of all periods of Earth history. As stated previously, we utilize paleomagnetic records for the past 10 Myr because this time period represents the most comprehensive record of TAF and PSV behavior. However, the Q_{PM} framework can be used for any interval of Earth history where a sufficient quantity of robust paleomagnetic data are available, but the relative importance of each criterion and associated acceptance regions will need to be updated to reflect paleomagnetic behavior for that time period. We also acknowledge, as discussed in section 2, that alternative paleomagnetic observables exist which are not used here. Notwithstanding, the parameters chosen for Q_{PM} criteria are based on well-established and commonly employed measures in paleomagnetic studies. We are confident that they appropriately describe the paleomagnetic field and are suitable to assess the degree to which geodynamo simulations are accurately replicating Earth's long-term magnetic field behavior.

A caveat to the Q_{PM} framework, and to any other study that uses the observed field to assess dynamo simulations, is that reproducing these paleomagnetic observables does not inherently demonstrate that a simulation is Earth-like. Magnetohydrodynamic theory suggests that the magnetic, Coriolis and buoyancy (Archimedian) forces are dominant in the momentum equation, termed MAC balance (e.g. Aubert et al., 2017; Starchenko and Jones, 2002). However, it is currently unclear whether the core is in a global MAC balance (Aurnou and King, 2017) and the issue cannot be resolved by current observations. It appears that MAC balance emerges in simulations as E and Pm are reduced towards geophysically relevant values (Aubert et al., 2017; Schaeffer et al., 2017), though some simulations at relatively high $E(\sim 10^{-4})$ may display MAC balance at leading order with non-negligible secondary contributions from viscous and inertial effects (Aubert et al., 2017; Dormy, 2016). As stated previously, low E and Pm values have not been achieved in simulations that span long timescales. In view of these limitations, here we chose to focus on criteria that can be derived from paleomagnetic observations and do not consider those based on the internal dynamics of the simulations.

5.2 Implications of Simulation Assessment

An unexpected outcome from our assessment of 46 simulations using the Q_{PM} criteria is that none are simultaneously reproducing all aspects of Earth's paleomagnetic field and that there are no obvious combinations of control parameters which will yield a simulation that reproduces Earth's long-term field behavior. This result contrasts with the findings in Christensen et al. (2010), who showed that geodynamo simulations within a certain E_{η}/Rm space can reproduce properties of the historical field. A potential explanation for the discrepancy between our results and those of Christensen et al. (2010) is that the uncertainty estimated for Earth parameters in the two studies were constructed following different approaches. Because the different time-dependent models of magnetic observations utilize direct and indirect observations, and span different time intervals, Christensen et al. (2010) assigned generalized 1-σ error bounds ranging from a factor of 1.75 - 2.5 times the magnitude of the observation to their Earth parameters. For our criteria, we instead utilized 95% confidence bounds calculated directly from paleomagnetic data. Our most restrictive criterion is VGPa, but it is arguably one of the best constrained Earth parameters for the last 10 Myr. The determination of a is dependent upon S_{VD} values for localities near the equator. In the PSV10 dataset, there are eight localities with latitudes between 10° and -10° ranging across all longitudes, with a minimum number of sites at each locality of at least 33. The maximum S_{VD} values estimated from these localities is ~15° (including 95% confidence intervals) and the minimum is ~ 6°, which is the absolute range that a can fall within. Even if we use these estimates for our range for Earth a values, our simulations are still well outside this range with a minimum a value of ~27° for simulations that reverse. Furthermore, a recent compilation of directional data for the Cretaceous and Middle Jurassic suggests that a values were between ~8° and 13° for these time periods, respectively, similar to our estimates for the past 10 Myr (Doubrovine et al., 2019). If we use the same approach as Christensen et al. (2010) for estimating uncertainty bounds, it would extend the values for a from 0° to 36°, at 1σ; such a range is inconsistent with estimates determined by paleomagnetic data.

An additional potential cause of the discrepancy between our findings and those of Christensen et al. (2010) could simply be that the field morphology observed for the historical field is not the one expected for long time scales and that secular variation of the recent field does not accurately reflect the behavior of the long-term paleomagnetic field. This is quite plausible given that spontaneous variations in field behaviour appear, from e.g. the PADM2M dipole model (Ziegler et al., 2011), to be active on timescales far longer than those captured in time-dependent field models.

The relatively low total Q_{PM} scores achieved by our simulations appears to be related to a tendency for many simulations (particularly those which reverse) to produce strong and/or strongly variable non-g_1^0 components. Generally, simulations that reverse have higher ΔI, a, b, and $V\%$ values (falling significantly outside the range of Earth), suggesting that these high/more variable non-g_1^0 components are more prevalent in reversing and multipolar simulations, as known from previous dynamo studies (Christensen and Aubert, 2006; Kutzner and Christensen,

2002) (Fig. 6). This trend may not hold true for all reversing simulations, as only two simulations passed Rev in this study, and more simulations should be assessed in the future to test this trend. To find a simulation that better captures Earth's paleomagnetic field, the non-g_1^0components must be reduced while the g_1^0 term remains capable of spontaneously changing its sign. Simulations that reverse and maintain a larger degree of dipole dominance have been produced in previous studies (e.g., Driscoll and Olson, 2009; Lhuillier et al., 2013; Wicht et al., 2009; Wicht and Meduri, 2016) and in future work it would be valuable to assess how these simulations perform using the Q_{PM} criteria. In our study, the only simulations with ΔQ_{PM} values approaching the Earth-like regime are those that do not reverse, suggesting that we currently cannot exclude non-reversing simulations in our quest for an Earth-like simulation.

We can conclude that simulations which fall within the 'wedge' of Christensen et al. (2010) are not guaranteed to reproduce Earth's paleomagnetic field behavior, and that compliance with the long-term magnetic field should be assessed separately, similar to findings of Davies and Constable (2014) for the Holocene. This is especially pertinent for studies that use the output from numerical geodynamo simulations to formulate corrections to paleomagnetic data [e.g., Driscoll and Wilson (2018), Lhuillier and Gilder (2013)], as these corrections may include non-Earth-like TAF and PSV behavior.

In this study we did not find long-duration simulations which simultaneously reproduce all aspects of Earth's paleomagnetic field behavior. However, our exploration of the possible parameter space is not exhaustive, and the fact that our simulations bracket Earth values suggest that a simulation reproducing Earth's paleo-field behavior should exist within a computationally accessible parameter regime. Overall, more long-duration simulations need to be assessed using the Q_{PM} criteria in the future.

6. Conclusions

We developed a framework for assessing the compliance between numerical geodynamo simulations and long-term magnetic field behavior (Q_{PM} criteria). Using Q_{PM} criteria, the compliance of 46 simulations with magnetic field behavior for the past 10 Myr was considered. We found that our simulations achieved a maximum Q_{PM} score of 3 out of 5, with most simulations scoring much lower, and with median ΔQ_{PM} misfit values of ~10, where less than 5 indicates compliance with Earth behavior. Low Q_{PM} scores appear to be partly due to enhanced non-g_1^0 components relative to those observed for the last 10 Myr on Earth. There appears to be no specific combination of E_η/Rm parameters in which simulations reliably replicate Earth's long-term field behavior. Furthermore, we find that compliance with the criteria set by Christensen et al. (2010) does not guarantee that a simulation reproduces Earth-like TAF and PSV behavior.

The Q_{PM} framework can provide a path towards developing simulations which can reproduce Earth's long-term magnetic field behavior in the future. This framework can be modified to represent periods of different geodynamo behavior in Earth's past, e.g. the Cretaceous or Middle Jurassic, allowing for a more robust characterization of the evolution of the deep interior through Earth's history, provided a sufficient quantity of robust paleomagnetic data are available.

Acknowledgements

CJS, AJB, and CJD acknowledge support from the Natural Environment research council (standard grant, NE/P00170X/1); AJB, RKB, and DGM acknowledge support from The

Leverhulme Trust (Research Leadership Award, RL-2016-080); CJD acknowledges a Natural Environment Research Council personal fellowship, reference NE/H01571X/1. A portion of the geodynamo simulations were performed on the UK National service ARCHER (via allocation through the Mineral Physics Consortium). We would like to thank Hagay Amit for his constructive and thorough reviews which helped improve this manuscript.

References

Amit, H., Choblet, G., 2009. Mantle-driven geodynamo features — effects of post-Perovskite phase transition. Earth Planets Sp. 61, 1255–1268.

Amit, H., Deschamps, F., Choblet, G., 2015. Numerical dynamos with outer boundary heat flux inferred from probabilistic tomography—consequences for latitudinal distribution of magnetic flux. Geophys. J. Int. 203, 840–855. https://doi.org/10.1093/gji/ggv332

Amit, H., Olson, P., 2015. Lower mantle superplume growth excites geomagnetic reversals. Earth Planet. Sci. Lett. 414, 68–76. https://doi.org/10.1016/j.epsl.2015.01.013

Aubert, J., Finlay, C.C., Fournier, A., 2013. Bottom-up control of geomagnetic secular variation by the Earth’s inner core. Nature 502, 219–223. https://doi.org/10.1038/nature12574

Aubert, J., Gastine, T., Fournier, A., 2017. Spherical convective dynamos in the rapidly rotating asymptotic regime. J. Fluid Mech. 813, 558–593. https://doi.org/10.1017/jfm.2016.789

Aubert, J., Tarduno, J.A., Johnson, C.L., 2010. Observations and models of the long-term evolution of earth’s magnetic field. Space Sci. Rev. 155, 337–370. https://doi.org/10.1007/s11214-010-9684-5

Aurnou, J.M., King, E.M., 2017. The cross-over to magnetostrophic convection in planetary dynamo systems. Proc. R. Soc. A Math. Phys. Eng. Sci. 473.

https://doi.org/10.1098/rspa.2016.0731

Biggin, A.J., de Wit, M.J., Langereis, C.G., Zegers, T.E., Voûte, S., Dekkers, M.J., Drost, K., 2011. Palaeomagnetism of Archaean rocks of the Onverwacht Group, Barberton Greenstone Belt (southern Africa): Evidence for a stable and potentially reversing geomagnetic field at ca. 3.5Ga. Earth Planet. Sci. Lett. 302, 314–328. https://doi.org/10.1016/j.epsl.2010.12.024

Biggin, A.J., Paterson, G.A., 2014. A new set of qualitative reliability criteria to aid inferences on palaeomagnetic dipole moment variations through geological time. Front. Earth Sci. 2, 1–9. https://doi.org/10.3389/feart.2014.00024

Biggin, A.J., Piispa, E.J., Pesonen, L.J., Holme, R., Paterson, G.A., Veikkolainen, T., Tauxe, L., 2015. Palaeomagnetic field intensity variations suggest Mesoproterozoic inner-core nucleation. Nature 526, 245–248. https://doi.org/10.1038/nature15523

Biggin, A.J., Steinberger, B., Aubert, J., Suttie, N., Holme, R., Torsvik, T.H., Van Der Meer, D.G., Van Hinsbergen, D.J.J., 2012. Possible links between long-term geomagnetic variations and whole-mantle convection processes. Nat. Geosci. 5, 526–533. https://doi.org/10.1038/ngeo1521

Biggin, A.J., Strik, G.H.M.A., Langereis, C.G., 2009. The intensity of the geomagnetic field in the late-Archaean : new measurements and an analysis of the updated IAGA palaeointensity database. Earth Planets Sp. 61, 9–22.

Biggin, A.J., Strik, G.H.M.A., Langereis, C.G., 2008. Evidence for a very-long-term trend in geomagnetic secular variation. Nat. Geosci. 1, 395–398. https://doi.org/10.1038/ngeo181

Bloxham, J., 2000. Sensitivity of the geomagnetic axial dipole to thermal core-mantle interactions. Nature 405, 63–65. https://doi.org/10.1038/35011045

Butler, R.F., 1992. Paleomagnetism: magnetic domains to geologic terranes. Blackwell Scientific

Publications Boston.

Christensen, U.R., Aubert, J., 2006. Scaling properties of convection-driven dynamos in rotating spherical shells and application to planetary magnetic fields. Geophys. J. Int. 166, 97–114. https://doi.org/10.1111/j.1365-246X.2006.03009.x

Christensen, U.R., Aubert, J., Hulot, G., 2010. Conditions for Earth-like geodynamo models. Earth Planet. Sci. Lett. 296, 487–496. https://doi.org/10.1016/j.epsl.2010.06.009

Christensen, U.R., Holzwarth, V., Reiners, A., 2009. Energy flux determines magnetic field strength of planets and stars. Nature 457, 167–169. https://doi.org/10.1038/nature07626

Christensen, U.R., Olson, P., 2003. Secular variation in numerical geodynamo models with lateral variations of boundary heat flow. Phys. Earth Planet. Inter. 138, 39–54. https://doi.org/10.1016/S0031-9201(03)00064-5

Christensen, U.R., Wicht, J., 2015. Numerical dynamo simulations. Treatise Geophys. (Second Ed. 8, 245–277.

Cox, A., 1975. The Frequency of Geomagnetic Reversals and the Symmetry of the Nondipole Field. Rev. Geophys. Sp. Phys. 13, 35–51.

Cromwell, G., Johnson, C.L., Tauxe, L., Constable, C.G., Jarboe, N.A., 2018. PSV10: A Global Data Set for 0–10 Ma Time-Averaged Field and Paleosecular Variation Studies. Geochemistry, Geophys. Geosystems 19, 1533–1558. https://doi.org/10.1002/2017GC007318

Davies, C.J., Constable, C.G., 2018. Searching for geomagnetic spikes in numerical dynamo simulations. Earth Planet. Sci. Lett. 504, 72–83. https://doi.org/10.1016/j.epsl.2018.09.037

Davies, C.J., Constable, C.G., 2014. Insights from geodynamo simulations into long-term geomagnetic field behaviour. Earth Planet. Sci. Lett. 404, 238–249.

https://doi.org/10.1016/j.epsl.2014.07.042

Davies, C.J., Gubbins, D., 2011. A buoyancy profile for the Earth's core. Geophys. J. Int. 187, 549–563. https://doi.org/10.1111/j.1365-246X.2011.05144.x

Davies, C.J., Gubbins, D., Willis, A.P., Jimack, P.K., 2008. Time-averaged paleomagnetic field and secular variation: Predictions from dynamo solutions based on lower mantle seismic tomography. Phys. Earth Planet. Inter. 169, 194–203. https://doi.org/10.1016/j.pepi.2008.07.021

Dormy, E., 2016. Strong-field spherical dynamos. J. Fluid Mech. 789, 500–513. https://doi.org/10.1017/jfm.2015.747

Doubrovine, P. V, Veikkolainen, T., Pesonen, L.J., Piispa, E., 2019. Latitude Dependence of Geomagnetic Paleosecular Variation and its Relation to the Frequency of Magnetic Reversals : Observations From the Cretaceous and Jurassic Geochemistry , Geophysics , Geosystems. Geochemistry, Geophys. Geosystems 20, 1240–1279. https://doi.org/https://doi.org/10.1029/2018GC007863

Driscoll, P., Olson, P., 2009. Effects of buoyancy and rotation on the polarity reversal frequency of gravitationally driven numerical dynamos. Geophys. J. Int. 178, 1337–1350. https://doi.org/10.1111/j.1365-246X.2009.04234.x

Driscoll, P.E., Wilson, C., 2018. Paleomagnetic Biases Inferred From Numerical Dynamos and the Search for Geodynamo Evolution. Front. Earth Sci. 6, 1–18. https://doi.org/10.3389/feart.2018.00113

Dziewonski, A.M., Lekic, V., Romanowicz, B.A., 2010. Mantle Anchor Structure: An argument for bottom up tectonics. Earth Planet. Sci. Lett. 299, 69–79. https://doi.org/10.1016/j.epsl.2010.08.013

Fisher, R., 1953. Dispersion on a Sphere. Proc. R. Soc. A Math. Phys. Eng. Sci. 217, 295–305. https://doi.org/10.1098/rspa.1953.0064

Glatzmaier, G.A., Coe, R.S., 2015. Magnetic Polarity Reversals in the Core, Treatise on Geophysics. Elsevier B.V. https://doi.org/10.1016/B978-0-444-53802-4.00146-9

Glatzmaier, G.A., Roberts, P.H., 1995. A three-dimensional self-consistent simulation of a geomagnetic field reversal. Nature 377, 203–209.

Gubbins, D., Willis, A.P., Sreenivasan, B., 2007. Correlation of Earth's magnetic field with lower mantle thermal and seismic structure. Phys. Earth Planet. Inter. 162, 256–260. https://doi.org/10.1016/j.pepi.2007.04.014

Jackson, A., Jonkers, A.R.T., Walker, M.R., 2000. Four centuries of geomagnetic secular variation from historical records. Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci. 358, 957–990.

Johnson, C.L., Constable, C.G., Tauxe, L., Barendregt, R., Brown, L.L., Coe, R.S., Layer, P., Mejia, V., Opdyke, N.D., Singer, B.S., Staudigel, H., Stone, D.B., 2008. Recent investigations of the 0-5 Ma geomagnetic field recorded by lava flows. Geochemistry, Geophys. Geosystems 9. https://doi.org/10.1029/2007GC001696

Johnson, C.L., McFadden, P.L., 2015. The Time-Averaged Field and Paleosecular Variation, Treatise on Geophysics. Published by Elsevier Inc. https://doi.org/10.1016/B978-0-444-53802-4.00105-6

Jones, G.M., 1977. Thermal Interactions of the Core and the Mantle and Long-Term Behavior of the Geomagnetic Field. J. Geophys. Res. 82, 1703–1709.

Korte, M., Constable, C., 2011. Improving geomagnetic field reconstructions for 0-3ka. Phys. Earth Planet. Inter. 188, 247–259. https://doi.org/10.1016/j.pepi.2011.06.017

Kutzner, C., Christensen, U.R., 2002. From stable dipolar towards reversing numerical dynamos 131, 29–45. https://doi.org/10.1016/S0031-9201(02)00016-X

Lhuillier, F., Gilder, S.A., 2013. Quantifying paleosecular variation: Insights from numerical dynamo simulations. Earth Planet. Sci. Lett. 382, 87–97. https://doi.org/10.1016/j.epsl.2013.08.048

Lhuillier, F., Hulot, G., Gallet, Y., 2013. Statistical properties of reversals and chrons in numerical dynamos and implications for the geodynamo. Phys. Earth Planet. Inter. 220, 19–36. https://doi.org/10.1016/j.pepi.2013.04.005

Masters, G., Johnson, S., Laske, G., Bolton, H., 1996. A shear-velocity model of the mantle. Philos. Trans. R. Soc. A 354, 1385–1411.

McFadden, P.L., Merrill, R.T., 1984. Lower mantle convection and geomagnetism. J. Geophys. Res. 89, 3354–3362.

McFadden, P.L., Merrill, R.T., McElhinny, M.W., 1988. Dipole / Quadrupole Family Modeling of Paleosecular Variation. J. Geophys. Res. 93, 11,583-11,588.

McMillan, D., Constable, C., Parker, R., Glatzmaier, G., 2001. A statistical analysis of magnetic fields from some geodynamo simulations. Geochem. Geophys. Geosyst 2, 2000GC000130. https://doi.org/10.1029/2000GC000130

Merrill, R.T., McElhinny, M.W., McFadden, P.L., 1996. The Magnetic Field of the Earth: Paleomagnetism, the Core, and the Deep Mantle Academic Press. San Diego, CA 531.

Mound, J., Davies, C., Silva, L., 2015. Inner core translation and the hemispheric balance of the geomagnetic field. Earth Planet. Sci. Lett. 424, 148–157. https://doi.org/10.1016/j.epsl.2015.05.028

Mound, J.E., Davies, C.J., 2017. Heat transfer in rapidly rotating convection with heterogeneous thermal boundary conditions. J. Fluid Mech. 828, 601–629. https://doi.org/10.1017/jfm.2017.539

Ogg, J.G., 2012. Geomagnetic Polarity Time Scale, in: The Geologic Time Scale 2012. Elsevier, pp. 85–113. https://doi.org/10.1016/B978-0-444-59425-9.00005-6

Olson, P., 2007. Gravitational dynamos and the low-frequency geomagnetic secular variation. Proc. Natl. Acad. Sci. 104, 20159–20166. https://doi.org/10.1073/pnas.0709081104

Olson, P., Christensen, U.R., 2006. Dipole moment scaling for convection-driven planetary dynamos 250, 561–571. https://doi.org/10.1016/j.epsl.2006.08.008

Olson, P., Christensen, U.R., 2002. The time-averaged magnetic field in numerical dynamos with non-uniform boundary heat flow. Geophys. J. Int. 151, 809–823. https://doi.org/10.1046/j.1365-246X.2002.01818.x

Olson, P., Deguen, R., Hinnov, L.A., Zhong, S., 2013. Controls on geomagnetic reversals and core evolution by mantle convection in the Phanerozoic. Phys. Earth Planet. Inter. 214, 87–103. https://doi.org/10.1016/j.pepi.2012.10.003

Olson, P.L., Coe, R.S., Driscoll, P.E., Glatzmaier, G.A., Roberts, P.H., 2010. Geodynamo reversal frequency and heterogeneous core-mantle boundary heat flow. Phys. Earth Planet. Inter. 180, 66–79. https://doi.org/10.1016/j.pepi.2010.02.010

Panovska, S., Constable, C.G., Korte, M., 2018. Extending Global Continuous Geomagnetic Field Reconstructions on Timescales Beyond Human Civilization. Geochemistry, Geophys. Geosystems 19, 4757–4772. https://doi.org/10.1029/2018GC007966

Schaeffer, N., Jault, D., Nataf, H.C., Fournier, A., 2017. Turbulent geodynamo simulations: A leap towards Earth's core. Geophys. J. Int. 211, 1–29. https://doi.org/10.1093/gji/ggx265

Stacey, F.D., Loper, D.E., 2007. A revised estimate of the conductivity of iron alloy at high pressure and implications for the core energy balance. Phys. Earth Planet. Inter. 161, 13–18. https://doi.org/10.1016/j.pepi.2006.12.001

Starchenko, S. V., Jones, C.A., 2002. Typical velocities and magnetic field strengths in planetary interiors. Icarus 157, 426–435. https://doi.org/10.1006/icar.2002.6842

Tarduno, J.A., Blackman, E.G., Mamajek, E.E., 2014. Detecting the oldest geodynamo and attendant shielding from the solar wind: Implications for habitability. Phys. Earth Planet. Inter. 233, 68–87. https://doi.org/10.1016/j.pepi.2014.05.007

Tarduno, J.A., Cottrell, R.D., Davis, W.J., Nimmo, F., Bono, R.K., 2015. A Hadean to Paleoarchean geodynamo recorded by single zircon crystals. Paleomagnetism 349, 521–524.

Tarduno, J.A., Cottrell, R.D., Watkeys, M.K., Hofmann, A., Doubrovine, P. V., Mamajek, E.E., Liu, D., Sibeck, D.G., Neukirch, L.P., Usui, Y., 2010. Geodynamo, solar wind, and magnetopause 3.4 to 3.45 billion years ago. Science (80-.). 327, 1238–1240. https://doi.org/10.1126/science.1183445

Van der Voo, R., 1990. The reliability of paleomagnetic data. Tectonophysics 184, 1–9. https://doi.org/10.1016/0040-1951(90)90116-P

Vandamme, D., 1994. A new method to determine paleosecular variation. Phys. Earth Planet. Inter. 85, 131–142. https://doi.org/10.1016/0031-9201(94)90012-4

Wicht, J., Meduri, D.G., 2016. A gaussian model for simulated geomagnetic field reversals. Phys. Earth Planet. Inter. 259, 45–60. https://doi.org/10.1016/j.pepi.2016.07.007

Wicht, J., Stellmach, S., Harder, H., 2009. Numerical Models of the Geodynamo: From Fundamental Cartesian Models to 3D Simulations of Field Reversals BT - Geomagnetic Field Variations, in: Glaβmeier, K.-H., Soffel, H., Negendank, J.F.W. (Eds.), . Springer

Berlin Heidelberg, Berlin, Heidelberg, pp. 107–158. https://doi.org/10.1007/978-3-540-76939-2_4

Willis, A.P., Sreenivasan, B., Gubbins, D., 2007. Thermal core-mantle interaction: Exploring regimes for “locked” dynamo action. Phys. Earth Planet. Inter. 165, 83–92. https://doi.org/10.1016/j.pepi.2007.08.002

Ziegler, L.B., Constable, C.G., Johnson, C.L., Tauxe, L., 2011. PADM2M: a penalized maximum likelihood model of the 0–2 Ma palaeomagnetic axial dipole moment. Geophys. J. Int. 184, 1069–1089. https://doi.org/10.1111/j.1365-246X.2010.04905.x

Figure Captions:

Figure 1. A. Bar graph showing the number of simulations that received scores of 0, 1, 2, and 3, respectively. Note, no simulation received Q_{PM} scores of 4 or 5. B. Bar graph showing the number of simulations that passed (failed) in blue (red) for each criterion. The percentage marks the percent of simulations that passed. C. Box plot of ΔQ_{PM}^{i} values over all simulations for each criterion. Horizontal lines mark median values, boxes outline the interquartile range (IQR), and error bars show full range excluding outliers (diamonds) which are defined as being more than 1.51 IQR outside the box. The dashed line indicates a target value of 1, and data below this line pass the respective criterion. D. Histogram of ΔQ_{PM} values for all simulations. Colors within each ΔQ_{PM} bin indicate total Q_{PM} score. For color see online version.

Figure 2. Representative reversal behavior for three end-member behaviors observed from the evaluated simulations: 1) Simulations that failed to reverse, 2) Simulations that passed the Rev criterion, and 3) Simulations that reversed but had $\tau_t > 0.15$ and failed Rev. In each subplot, the

figure plots calculated true dipole latitude versus time in years, calculated using the diffusion timescale and the electrical conductivity value of 3 x 10^5 S/m from Stacey and Loper (2007). Dipole latitude is reported in degrees.

Figure 3. Representative ΔI vs. latitude curves showing three end-member behaviors observed from the evaluated simulations: 1) Simulations that failed the IncAnom criterion due to low values, 2) Simulations that passed the IncAnom criterion, and 3) Simulations that failed the IncAnom criterion because values were too high. In each plot, data points mark the median ΔI values and 95% confidence bounds estimated from the repeated 10,000 downsampling routines, for each 10° latitude band. The star indicates the maximum median ΔI value used to evaluate the Q_{PM} criterion. The dashed blue lines mark the 95% confidence bounds for Earth and the negative equivalent. Units are in degrees.

Figure 4. Representative VGP dispersion (using the Vandamme cutoff) vs. latitude curves for four end-members behaviors observed in the evaluated simulations: 1) Simulations that failed because a was too high but b passed, 2) Simulations that passed both VGPa and VGPb, 3) Simulations that failed because both a and b values were too high, and 4) Simulations that failed because both a and b values were too low. The red solid line marks the Model G curve plotted using median a and b parameters and the light red envelope marks the 95% confidence interval. The solid blue line in each figure is the Model G curve calculated from median a and b parameters for Earth and the light blue envelope marks the 95% confidence interval (for color see online version). Units are in degrees.

Figure 5. A. Representative $V\%$ values and dipole moment distributions (calculated without downsampling, units are non-dimensionalized) for three end-member behaviors observed from the evaluated simulations: Model 30) Simulation that failed because the $V\%$ value was too low, Model 6) Simulation that passed, and Model 1) Simulation that failed because the $V\%$ value was too high. The dashed lines in A mark Earth range. Insets plot the distribution of virtual dipole moments for Earth between 0-1 Myr (B) and 1-10 Myr (C), units are in ZAm^2 (10^{21}).

Figure 6. Histograms of calculated values from each simulation, shown for each criterion, colored by the proportion of data from simulations that are in the locked regime of convection (lock e.g. Davies et al., 2008; light blue), did not reverse (Non; blue), reversed (Rev; dark blue), and reversed but had $\tau_t > 0.15$ and did not pass Rev (MP; darkest blue). Pink boxes mark the range for Earth values in each subplot. For color see online version.

Figure 7. Evaluated dynamo simulations plotted as a function of magnetic Ekman number (E_η) vs. Magnetic Reynolds number (Rm) following Christensen et al. (2010). Circle size denotes total Q_{PM} score, with the largest circles having scores of 3 and the smallest circles having scores of 0. Color denotes total misfit value, ΔQ_{PM}. The dashed line marks the wedge-shaped region that contain simulations with Earth-like misfit scores ($\chi^2 < 4$) and FF boundary conditions in Christensen et al. (2010).

Table 1. Summary of Earth time average field and paleosecular variation values. Med indicates median values, high indicates the upper 95% confidence bound, and low indicates lower 95% confidence bound. Values for a, b, and ΔI were calculated from data presented in Cromwell et

al. (2018), $V\%$ values for the 0-1 Ma interval and 1-10 Ma interval were calculated from data presented within the PINT15 database, and τ_t values were estimated from Ogg (2012).

Table 2. Summary of input and output parameters for assessed geodynamo simulations. Sim. Name = Simulation Name, Pr = Prandtl number, Pm = Magnetic Prandtl number, E = Ekman number, Ra = Rayleigh number, ε = amplitude of prescribed outer boundary heat flux heterogeneity, BC = boundary condition (inner-outer), Rm = Magnetic Reynolds number, and Rev indicates reversing regime defined using τ_t. non=non-reversing, rev=reversing with $0.0375< \tau_t <0.15$, and multi= $\tau_t >0.15$. Note, parentheses in Sim. Name indicate previously published or further integrated simulations, where the name in parentheses corresponds to the name used in Davies and Constable (2014). * indicate simulations that utilized an inhomogeneous boundary condition after Masters et al. (1996). Unless noted, all other inhomogeneous boundary conditions utilized a recumbent Y_2^0 heat flux pattern (Dziewonski et al., 2010).

Table 3. Summary of ΔQ_{PM} and Q_{PM} scores for assessed geodynamo simulations. Rev indicates reversing regime as defined in Table 2. Sim. Name = simulation name (see Table 2 caption for details), τ_t = proportion transitional, ΔQ_{PM} = ΔQ_{PM}^i misfit values for each respective criterion, and Q_{PM} = Q_{PM}^i score for each respective criterion.