Physics of the Earth and Planetary Interiors 366 (2025) 107399

Contents lists available at ScienceDirect

Physics of the Earth and Planetary Interiors

journal homepage: www.elsevier.com/locate/pepi

Evaluation of archeo- and paleomagnetic field models and their common features[☆]

Ingo Wardinski [a,*], Filipe Terra-Nova [b], Erwan Thébault [c]

[a] Institut Terre et Environnement Strasbourg, Ecole et Observatoire des Sciences de La Terre, Université de Strasbourg, CNRS UMR 7063, Strasbourg, France
[b] Laboratoire de Planétologie et de Géodynamique, Université de Nantes, Nantes Atlantiques Universités, CNRS UMR 6112 Nantes, France
[c] Laboratoire Magmas et Volcans, Université Clermont Auvergne, Clermont-Ferrand, France

ARTICLE INFO

Keywords:
Archeo- and paleomagnetic field models
Millennial variation of the geomagnetic field
Earth-likeness of numerical dynamo simulations

ABSTRACT

Archeo- and paleomagnetic field models show a wide range of temporal variability and of spatial content. While the temporal variability may reflect true geomagnetic field variation, the different spatial content of individual models could be explained by different modeling strategies and data sources, but mostly by data uncertainties. To overcome these data uncertainties, we derive averaged models over the last 100 kyrs from a large suite of different archeo- and paleomagnetic field models using different average techniques. The averaged models allow us to evaluate the robustness and the significance of spatial features of these models throughout time. It is utilized to compute structural criteria that quantify the axial dipole dominance, the equator-symmetry of magnetic field, its zonality and other important measures of weak field regions and control of the geodynamo by heterogeneous heat flux at the lowermost mantle. These criteria are used to quantify the Earth-likeness of numerical dynamo simulations. Over 100 kyrs the criteria show larger fluctuations than previously assumed, which implies a wider range of numerical dynamo simulations to be considered Earth-like.

1. Introduction

Magnetic field models that are based on archeo- and paleomagnetic data allowed studies of the past morphology of Earth's magnetic field (Korte and Holme, 2010; Terra-Nova et al., 2015; Terra-Nova et al., 2013), its temporal evolution (Amit et al., 2011; Dumberry and Finlay, 2007a) and the underlying processes (Dumberry and Bloxham, 2006; Wardinski and Korte, 2008). Though, a spatially limited and uneven distribution of data and uncertainties in their precise dating limit the spatial and temporal resolution of the models. However, the large scale field morphology including the South Atlantic Anomaly (SAA) and the decay of the dipole field can be studied. Observed long-term variations allow to infer magnetohydrodynamic processes in the outer core and the influence of boundary conditions at the base of the mantle and at the inner core surface on these processes (Amit and Choblet, 2009; Aubert et al., 2013; Aubert et al., 2010; Olson et al., 2010). These models also allow comparisons with results of numerical simulations of the geodynamo that are based on first principles (see for instance (Christensen and Wicht, 2007)).

Ab initio calculations of the geodynamo effect in Earth's outer core provide different kinds of observations. These numerical simulations mimic the generation of Earth's magnetic field that depends on different characteristic parameters which control the strength of the forces in a force balance, i.e. the Navier-Stokes equation. However, these characteristic parameters in numerical dynamo simulations have to take unrealistic values, as the current computing power does not allow simulations to run with Earth-like values. In order to compare results of numerical dynamo simulations to geomagnetic field observations, (Christensen et al., 2010) introduced criteria to assess the Earth-likeness of numerical dynamo simulations. These criteria are model-based and require a description of the geomagnetic field by a spherical harmonic model. We refer to these criteria as structural criteria, as they measure different aspects of the large scale morphology of the geomagnetic field and numerical dynamo simulations. But, so far, criteria were derived under the premises that Earth's magnetic field is rather steady and that magnetic field excursion and reversals do not occur so frequently. In fact, one of these criteria, the dipolarity, has a value that is derived from the *gufm1* (Jackson et al., 2000) covering the interval 1590 to 1990 and would never permit field excursions and reversals to occur in numerical dynamo simulations. Recently, (Sprain et al., 2019) proposed another

---

[☆] This article is part of a Special issue entitled: 'SEDI 2024' published in Physics of the Earth and Planetary Interiors.

[*] Corresponding author.
E-mail address: wardinski@unistra.fr (I. Wardinski).

https://doi.org/10.1016/j.pepi.2025.107399
Received 12 August 2024; Received in revised form 19 May 2025; Accepted 4 June 2025

set of criteria which reflects paleomagnetic observables of the paleosecular variation and the time-averaged field on timescales $>$ 10 Myrs. Their observation-based criteria use a compilation of paleomagnetic directional data (Cromwell et al., 2018) and the PINT paleointensity database (Biggin et al., 2015; Biggin et al., 2009). The criteria quantify properties of the virtual geomagnetic dipole, inclination anomaly, the virtual dipole moment and the time spent by the dipole at latitude lower than 45 degrees during magnetic field reversals. (Sprain et al., 2019) found accordingly that most numerical simulations fail to reverse in an Earth-like manner. However, (Meduri et al., 2021) find a set of numerical simulations that passes these criteria and that can reproduce paleosecular variation and time-averaged field behavior over 10 Myrs.

In this study we derive structural criteria of Earth's magnetic field for different periods to obtain insights to the variability of its morphology over the past 100 kyrs. The studied periods are: the archeo- and historical period between 1100 BCE - 1800 CE, the Holocene between 10,000 BCE - 1000 BCE, and the upper Pleistocene (100 kyr BCE - 10 kyr BCE) that includes two known magnetic field excursions. The Laschamp excursion (Bonhommet and Babkine, 1967) that centered at 39000 BCE and the Mono-Lake excursion (Denham and Cox, 1971; Liddicoat and Coe, 1979) at 31000 BCE, respectively. A description of the analyzed archeo- and paleomagnetic field models is given in section 2,. In the third section, we also outline a derivation of a representative model for these periods. Sections 4 and 5 are dedicated to quantify differences between individual models and a mean model in order to highlight common features. We derive criteria of the geomagnetic field morphology in section 6. The last section contains a final discussion of our result and their conclusions.

2. Archeo- and paleomagnetic field models

Geomagnetic field models that cover the last 400 years use direct observations of Earth's magnetic field taken by satellites and at ground-based magnetic observatories, or are based on nautical charts taken during ship cruises between 1600 and 1900 CE (Jonkers et al., 2003). Indirect observations of Earth's magnetic field derived from archaeological artifacts, volcanic lava flows, lake and ocean-sediments are used to provide a description of the Archeo and Paleomagnetic field prior to the first man-made observations around 1600 CE. These mathematical descriptions (field models) are built on representations of Earth's internal magnetic field as a potential field in a spherical coordinate system (see for instance (Alken et al., 2021a)), i.e.

$$V(r,\theta,phi,t) = a \sum_{\ell=1}^{L_{\text{int}}} \sum_{m=0}^{\ell} \left\{ \left(g(t)_\ell^m \cos(m\phi) + h(t)_\ell^m \sin(m\phi)\right) \left(\frac{a}{r}\right)^{\ell+1} P_\ell^m(\cos\theta) \right\} \tag{1}$$

where V is the magnetic potential of internal origin that can be decomposed in spherical harmonics. The coefficients time-varying $\{g(t)_\ell^m, h(t)_\ell^m\}$ are the Gauss coefficients, a is the mean radius of Earth's surface and r is the radius of interest. θ and ϕ are colatitude and east longitude, respectively. The $P_\ell^m(\cos\theta)$ are the Schmidt semi-normalized associated Legendre functions, where l is the degree and m the order. $\ell_{\max}$ is the truncation degrees of the spherical harmonic expansion.

In this study, we evaluate 13 models that describe the magnetic field over the archeo- and historical period and the Holocene, namely:

- A_FM, ASD_FM and ASDI_FM (Licht et al., 2013) - past 3 kyr.
- ArchKalmag14k.r (Schanner et al., 2022) past 14 kyr,
- BIGMUDI4k (Arneitz et al., 2019) - past 4 kyr,
- CALS3k.4b (Korte and Constable, 2011) - past 3 kyr,
- CALS10k.2 (Constable et al., 2016) - past 10 kyr,
- COV-ARCH and COV-LAKE (Hellio and Gillet, 2018) - past 3 kyr,
- HFM.OL1.A1 (Constable et al., 2016) - past 9 kyr,
- pfm9k.2 (Nilsson and Suttie, 2021) - past 9 kyr,
- SHA.DIF.14 k (Pavón-Carrasco et al., 2014) - past 14 kyr,
- SHAWQ2K (Campuzano et al., 2019) - past 3 kyr.

And three paleomagnetic field models that cover the upper Pleistocene:

- LSMOD.2 (Brown et al., 2018) - 30–50 BCE kyr,
- GGFSS70 (Panovska et al., 2021) - 15-75 BCE kyr,
- GGF100k (Panovska et al., 2018) - past 100 kyr.

There are models, which are not considered in this study, namely pfm9k.1 by (Nilsson et al., 2014) which is replaced by its successor the pfm9k.2 (Nilsson and Suttie, 2021) and ARCH3k.1 (Korte et al., 2009) which is only suitable for northern hemisphere studies. Furthermore, models that are designed to capture the dipole field behavior during the Laschamp excursion, the IMODE (Leonhardt and Fabian, 2007) and earlier field excursions in the Pleistocene, the IMIBE (Lanci et al., 2008) are also not considered. Mainly because, these models only describe the behavior of the first spherical harmonic degree robustly and do not allow inferences of the non-dipole field in these excursion periods.

All models evaluated in this study are derived from archeomagnetic, volcanic and sediment data sets provided by the *GEOMAGIA* database (Brown et al., 2015a; Brown et al., 2015b).

These models can be grouped into two different sets according to their modeling strategies. One set applies iterative stochastic inversion (SI) (Gubbins, 1983; Jackson, 1979), the other set is based on iterative Bayesian inference (BI). With direct observations of the magnetic field (satellite and observatory data), both methods yield similar results, see for instance (Alken et al., 2021b) where results of these different modeling strategies are evaluated. (Backus, 1988a; Backus, 1988b) provide a more detailed discussion of these techniques and their differences.

All SI-models use cubic B-splines to parameterize the temporal variability of the geomagnetic field, whereas the BI-models use stochastic processes that are approximated by linear functions between model epochs. Most of the models provide uncertainty estimates of their Gauss coefficients. The SI-models provide these estimates by using a boot-strap technique (Korte et al., 2009), the BI-models evaluate the posterior distribution of the model parameters. Both techniques of uncertainty estimation are equivalent (Efron, 2013). However, these estimations may not fully quantify the total error budget, which also has to contain timing uncertainties, uncertainties of the sedimentation process, acquisition errors, etc. Equally important are modeling uncertainties introduced by implicit or explicit prior information. These information are designed either to minimize the energy of parameters or to assign a prior probability distribution to parameters during the inverse problem that may not be well constrained due to data sparsity at Earth's surface. For different models based on more or less the same data, these latter error types may be accounted differently by the individuals studies and the possibly different total error budgets should be kept in mind when comparing different models.

The archeomagnetic field models of the CALSx & HFM series (Constable et al., 2016) and SHA.DIF.14 k commonly use stochastic inversion, cubic B-splines as temporal representations and a similar boot-strap method to evaluate model uncertainties (Korte et al., 2009). They differ in the usage of sediment data, CALS10k.2 and HFM.OL1.A1 are dominated by sediment data extending to 10 kyrs BCE, whereas SHA.DIF.14 k avoids using sediment data. Undoubtedly, large timing uncertainties are associated with sediment data, but for some epochs these are the only available data during some periods. CALS10k.2 is, furthermore constrained to fit *gufm1* from 1840 to 1950.

These models are expanded in spherical harmonics up to harmonic degree 10 $\ell_{max} = 10$ (1) and uses cubic B-spline to parameterize the temporal variability. Their spatial regularization scheme is set to find the models with minimum ohmic dissipation at the core surface

(Gubbins and Bloxham, 1985) that fit the data to a desired accuracy. This norm limits each Gauss coefficient only by requiring it not to produce by itself more ohmic heat in the core than the observed heat flow at Earth's surface. The application of the heat flux norm may not provide a strong constraint for the large-scale, but for the small to intermediate scales of the magnetic field at the core surface (Wardinski and Thébault, 2019).

Ensemble and Bayesian approaches are used to derive COV-ARCH and COV-LAKE, BIGMUDI4k, ArchKalmag14k.r and the pfm9k.2. For the pfm9k.2 magnetic field and data ages are co-estimated to overcome chronological uncertainties in the data, particularly those in sediment data.

A third group of models consists of model that are also built on stochastic inversion and cover periods earlier than the Holocene, i.e. the upper Pleistocene, like the GGF100k, the GGFSS70 and the LSMOD.2.

While it is generally believed that the temporal resolution of archeomagnetic field models is not shorter than 20 years and may reach only 100 years for paleomagnetic models, the spatial resolution of these models is still a matter of debate. In this study, we use a temporal sampling of 50 years during the Archaeological and Holocene periods, as the individual models provide model coefficient every 50 years. During the Pleistocene, we set the sampling to 100 years, which is also determined by the analyzed models. Most models claim to resolve field structure up to spherical harmonic degree 4 to 5, but no finer details and higher degrees. This sets the spatial limit of our study and we will discuss only the field morphology related to the first 5 spherical harmonic degrees, i.e. the first 35 G coefficients.

3. Evaluation of archeo- and paleomagnetic field model

Each of the models that are presented in the previous section show different spatial and temporal characteristics and may differ substantially from an other. These differences are due to different data sets (i.e. different versions of *GEOMAGIA*) used in the model derivation, whether or not sediment data are used, and different modeling strategies. For instance, there are different policies to reject data that are identified as outliers. However, apart from these differences, models are supposed to show coherent features that are more or less equally pronounced in each of them. We attempt to characterize these features in order to provide a digestive description of the archeo- and paleomagnetic field and to derive values of structural criteria to allow comparisons to numerical dynamo simulations.

Similarly, as in the process of deriving the International Geomagnetic Reference Field model (IGRF) (Alken et al., 2021a; Alken et al., 2021b), we have to deal with a set of field models that are equally realistic representations of the geomagnetic field. These representations are equal in the sense that their model priors can not be proven to be false, and therefore, can not be weighted differently with respect to each other.

Table 1

List of evaluated archeomagnetic field models and their setup. Abbreviations stand for **A**rcheomagnetic artifacts, **L**ava flows, **S**ediments and **H**istorical data. The model tags are used in the text and graphics.

name	tag	max. SHD	temporal basis	data types
Models based on stochastic inversion (SI)				
CALS3k.4b	cals3k4b	10	cubic B-splines	A & L & S
CALS10k.2	cals10k	10	cubic B-splines	A & L & S
HFM.OL1.A1	hfm	10	cubic B-splines	A & L & S
SHA.DIF.14 k	shadif14k	10	cubic B-splines	A & L
SHAWQ2k	shawq2k	10	cubic B-splines	A & L
LSMOD.2	lsmod2	10	cubic B-splines	A & L & S
GGFSS70k	GGFSS70	6	cubic B-splines	A & L & S
GGF100k	GGF100k	10	cubic B-splines	A & L & S
Models based on ensemble and Bayesian approaches (BI)				
ArchKalmag14k.r	aklm	20	Gaussian process	A & L
pfm9k.2	pfm9k2	5	Gaussian process	A & L & S
COV-ARCH	covarch	10	Gaussian process	A & L
COV-LAKE	covlake	10	Gaussian process	A & L & S
BIGMUDI4k	bigmudi4k	8	Gaussian process	H & A & L
A_FM, ASD_FM, ASDI_FM		5	not specified	A & L & S

There are at least two ways to summarize the field morphology of the models listed in Table 1 at each time step; either by computing the statistical mean and median of their Gauss coefficients, or by a spatial average of the models on a spherical surface (spherical averaging). The latter method requires to compute field components on a global grid for each model, and to invert the grid for a new set of Gauss coefficients that represents the spatial average of these models at a given epoch. This procedure was proposed for the evaluation of previous IGRF candidates (Finlay et al., 2010; Thébault et al., 2015). To apply this procedure, we compute vector-triplets of the magnetic field components (B_θ, B_ϕ, B_r) at each point of a uniform grid over Earth's surface from each model every 50 years for archeo- and historical field models, and every 100 years for Holocene and Pleistocene models. The number of grid points is set to 4050, 45 points in latitudinal direction and 90 in longitudinal direction, respectively. We seek the model parameters that represent equally all models, i.e. the magnetic field morphology that is commonly manifested by all models. These model parameters are estimated by an iterative least squares fitting scheme to the grided data. For a linear inverse problem (least squares fit) the model vector m that contains the Gauss coefficients is found at the minimum of an objective function

$$\Theta(m) = (\mathrm{y} - \mathbf{A}\mathrm{m})^{\mathrm{T}}\mathbf{C}_{\mathbf{w}}(\mathrm{y} - \mathbf{A}\mathrm{m}), \tag{2}$$

where y is the data vector containing the grided magnetic data, **A** is the spherical harmonic design matrix, and $\mathbf{C}_{\mathrm{W}}$ the data weight matrix. The solution is iteratively given by

$$\mathrm{m}_{\mathrm{i}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{C}_{\mathbf{w},\mathrm{i}}\mathbf{A}\right)\Big/\left(\mathbf{A}^{\mathrm{T}}\mathbf{C}_{\mathbf{w},\mathrm{i}}\mathrm{y}\right). \tag{3}$$

The iteration concerns only the $\mathbf{C}_{\mathbf{w},\mathbf{i}}$ matrix and is updated every iteration step *i*. The weight for each datum is estimated numerically as the residual between the magnetic field components B_k of the grided data and $\widehat{B}_{\mathrm{k}}$ derived from the actual model m_{i} at each point *p*:

$$\epsilon_{\mathrm{k,p}} = B_{\mathrm{k,p}} - \widehat{B}_{\mathrm{k,p}}. \tag{4}$$

The weight matrix $\mathbf{C}_{\mathbf{w}}$ is updated at every iteration step i and contains weights for the magnetic field components B_k at the position *p* of the grid, namely

$$w_{\mathrm{k,p}} = 1\big/\epsilon_{\mathrm{k,p}}. \tag{5}$$

The minimal weight is set to $w_{\mathrm{k,p}} = 0.01$ as a precaution to avoid divisions by zeros. Each iteration starts from an initial model $\mathrm{m}_{\mathrm{i}=0}$ where all Gauss coefficients are set to 1.0 and the iteration converges after three steps.

We also apply a Huber-weighting scheme for the spherical averaging as it is applied in the derivation of the IGRF-13 (Alken et al., 2021a). The weights are derived as

$$w_{\mathrm{k,p}} = \begin{pmatrix} 1 & \text{if} \quad \| \epsilon_{\mathrm{k,p}}/\sigma_{\mathrm{it}} \| \quad \leq c \\ c/ \| \epsilon_{\mathrm{k,p}}/\sigma_{\mathrm{it}} \| & \text{otherwise,} \end{pmatrix} \tag{6}$$

where the Huber constant $c = 1.345$ and σ_{it} the standard deviation of the residuals at one grid point $\epsilon_{\mathrm{k,p}}$ at each iteration step. σ_{it} is approximated over n grid point as

$$\sigma_{\mathrm{it}} \approx \left(1\Big/n\sum_{i}^{n}\left|B_{\mathrm{k,p}} - \widehat{B}_{\mathrm{k,p}}\right|\right)\Big/0.6745 \tag{7}$$

which is the mean absolute deviation that is less sensitive to large outliers compared to the standard deviation (Thébault et al., 2015). The Huber formalism assigns lower weights to values that are more than 1 standard deviation away from the model m_i of the actual iteration step. We will refer to these spherical averaging models as M_{RW} (mean residual-weighted) and M_{HW} (mean Huber-weighted), respectively.

Another and simpler way is to summarize the morphology of these models by computing their statistical mean and median from the Gauss coefficients directly at the same time steps as for the spherical averaging. In detail, this computation is performed by taking the SI models and the ensemble means of the BI models at a given time equally without considering their model uncertainties. We refer to these models as statistical mean and statistical median models, respectively.

For both methods, the spherical averaging and the statistical averaging, we obtain 293 mean models, 186 for the Pleistocene, 70 for the Holocene and 37 for Archaeological and Historical period, respectively. These models represent individual time slices.

3.1. Pleistocene period

The temporal variability of the field coefficients during the upper Pleistocene are shown in Fig. 1. Here, we focus on the period from 47,000 to 28,000 BCE, as for this period three models are available and therefore may give a more robust mean field model, than for the rest of the period which is covered by two models. We derive the average models from the lsmod2, GGFSS70 and the GGF100k models. The period from 47,000 to 28,000 BCE also includes the Laschamp excursion from 39,000 to 36,000 BCE, and a weak field phase during which axial dipole and non-dipole power at the CMB are comparable leading to more than one excursion between 34,000 and 28,000 BCE (Korte et al., 2019). One of these excursion is known as Mono-Lake excursion that centered around 31,000 BCE.

We find no significant differences between the Gauss coefficients of

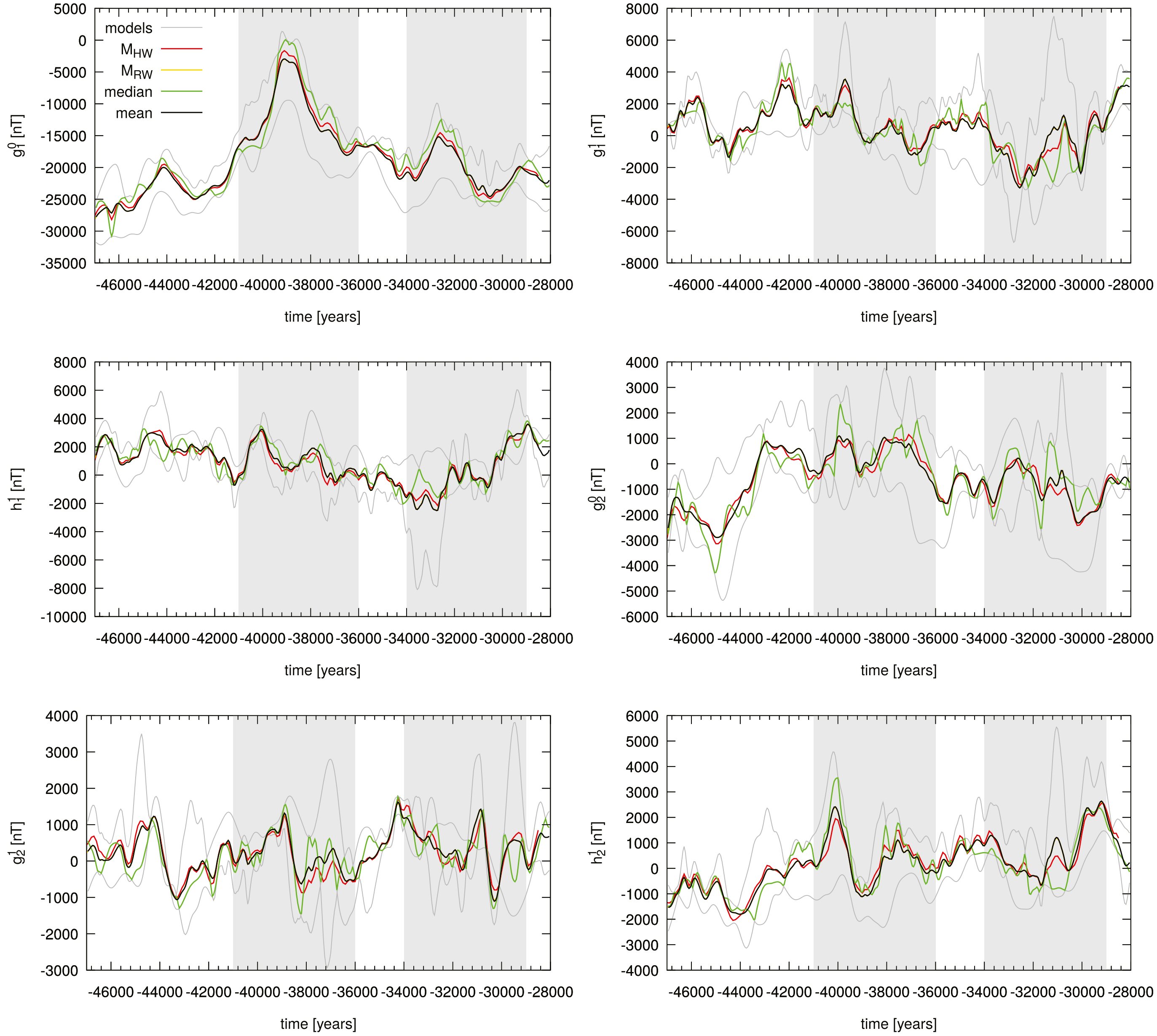


Fig. 1. Temporal evolution of the first 6 Gauss-coefficients over the Pleistocene period that centers around the Laschamp excursion at 40000 BCE. Colored curves are derived from the models given in Table 1, their statistical mean in black and the statistical median in green, respectively. The spherical averages are shown in yellow (M_{RW}) and red (M_{HW}). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the statistical mean models and those coefficients derived by the spherical averaging procedure that uses a residual-weighted iterative inversion. Absolute differences are smaller than 1 nT. Larger differences occur between the statistical mean and the spherical averaging that is using a Huber-weighted iterative inversion. Differences are about a few 100 nT. The largest deviation from the statistical mean is carried by the statistical median model. We display maps of the mean and median models during the Laschamp and Mono-Lake excursions. Differences in spatial features appear to be minor, see Appendix Fig. B.1.

3.2. *Holocene period*

The derivation of the average models during the Holocene period before 7000 BCE is based on three models (hfm, shadif14k, cals10k), thereafter and before 3000 BCE it includes also the pfm9k2, and since 3100 BCE it includes the aklm which can be considered as realistic since then (Schanner 2024, priv. Comm.).

Fig. 2 displays the temporal variability of the large scale magnetic field Gauss coefficients in the Holocene. Higher temporal variability and the large amplitudes are seen for g_1^1 and h_1^1 which are derived from the SHA.DIF.14 k during most of the period. These features of the SHA. DIF.14 k may likely result from the disregard of sediment data as model inputs (Panovska et al., 2015).

3.3. *Archaeological and historical period*

Prior to 0 CE, the averaged models of the archaeological and historical period (1100 BCE - 1800 CE) are computed from 9 models, thereafter from 13 models including the shawq2k, ASD_FM and A_FM. The two later are considered every 200 years, as their temporal sampling is 40 years and therefore matches the other models epochs every 200 years. Fig. 3 shows the temporal evolution of the first six Gauss-coefficients of all models of the archaeo- and historical period. We note the large spread of the model curves, particularly related to the BI-

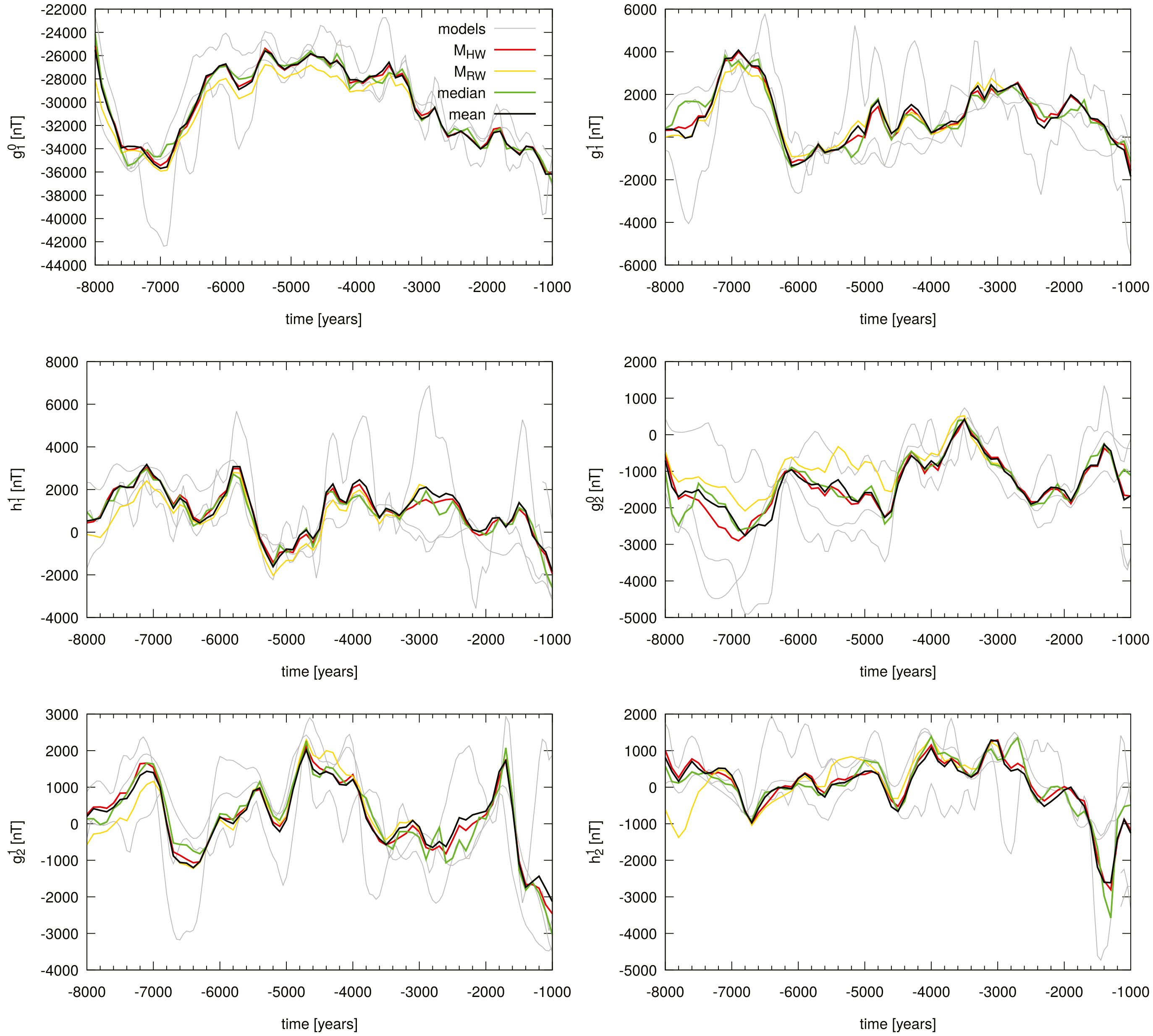


Fig. 2. Temporal evolution of the first 6 Gauss-coefficients over the Holocene period from 8000 BCE to 1000 BCE. Colored curves are derived from the models given in Table 1, their statistical mean in black and the statistical median in green, respectively. The spherical averages are shown in yellow (M_{RW}) and red (M_{HW}). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

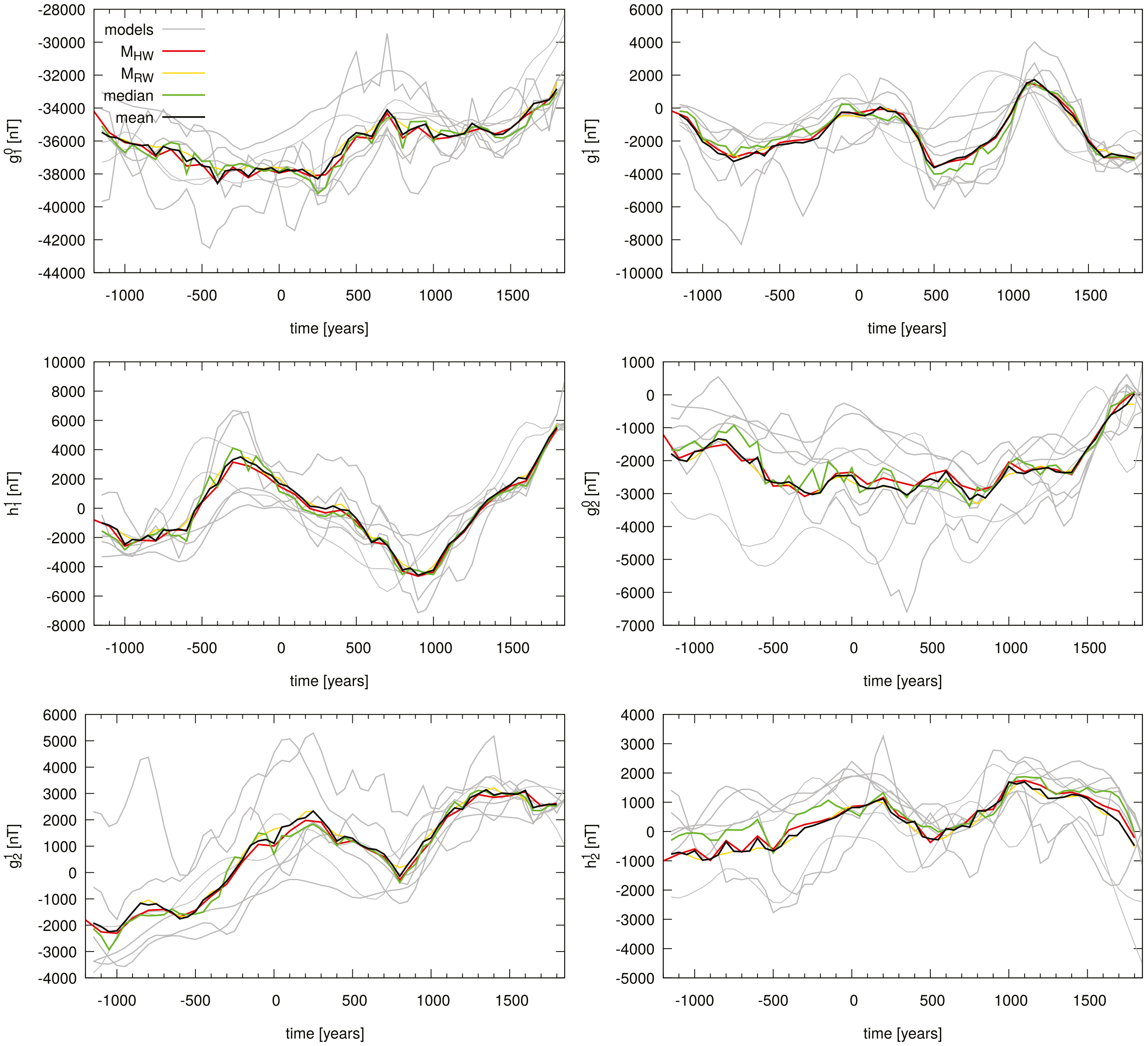


Fig. 3. Temporal evolution of the first 6 Gauss-coefficients over the last 3000 years. Gray-shaded curves are derived from the models given in Table 1, their statistical mean in black and the statistical median in green, respectively. The spherical averages are shown in yellow (M_{RW}) and red (M_{HW}).

models. However, model curves seemingly start to behave similar around 800 to 500 BCE and tend to converge after ~700 CE. Additional Appendix Figs. A.1 & A.2 compare the curves for SI and BI models during the archeo and historical period. They indicate that curves of the SI-models tend to gather more closely (Appendix Fig. A.1) than those of the BI-models (Appendix Fig. A.1). The reason for the large spread is not fully understood, but it is certainly related to the applied modeling technique, the individual treatment of age uncertainties and possibly different data sets (different versions of *GEOMAGIA* database, that evolved over time).

We observe that the statistical mean and the M_{RW} almost match, slightly larger deviations are shown by the M_{HW}, and the median shows the largest offsets from the other curves.

Finally, we decide to use the mean model to be representative for the archaeo- and paleomagnetic field, mainly because results of spherical averaging are close to the statistical mean.

4. Differences of field models

Before we discuss common features of the models, we highlight their differences in the description of the field. The applied measures are particularly designed to highlight model differences. First, we analyze spectral differences between the mean model and the individual models, by using

$$R_\ell = \left(\frac{a}{c}\right)^{2\ell+4} (\ell+1) \sum_{m=0}^{\ell} \left(\delta g_\ell^m\right)^2 + \left(\delta h_\ell^m\right)^2, \tag{8}$$

where ℓ, m are the spherical harmonic degree and order of the Gauss coefficients g_ℓ^m, h_ℓ^m, $\delta g_\ell^m = \hat{g}_\ell^m - g_{i,\ell}^m$ and likewise δh_ℓ^m the differences between the mean and individual models, and a, c Earth's and outer core radii, respectively.

While (8) indicates differences between models, it cannot identify features of models to be different. To identify such features, we derive the azimuthal spectral differences between individual models and the mean model. The azimuthal spectrum is arranged by azimuthal modes that are defined by the azimuthal ratio az $= m/\ell$ which ranges from 0 to 1 for the g_ℓ^m and from -1 to 0 for the h_ℓ^m

$$R(az) = \left(\frac{a}{c}\right)^{2\ell+4}(\ell+1)\sum_{az}(\delta g(az))^2 + (\delta h(az))^2. \tag{9}$$

Models with $\ell_{\max} = 5$ have 21 modes, where the center mode az $= 0/\ell$ is made up from the spectral differences of the 5 zonal coefficients and the mode $\ell = m$ is derived from the 10 sectorial coefficients (5 for g and h each).

Fig. 4 shows the spectral differences and the azimuthal spectral differences for four different epochs whose maps are discussed below. The spectral differences between the mean model and individual models are compared to the spectral differences between the mean model and a model of randomly shuffled Gauss coefficients (random model) shown as

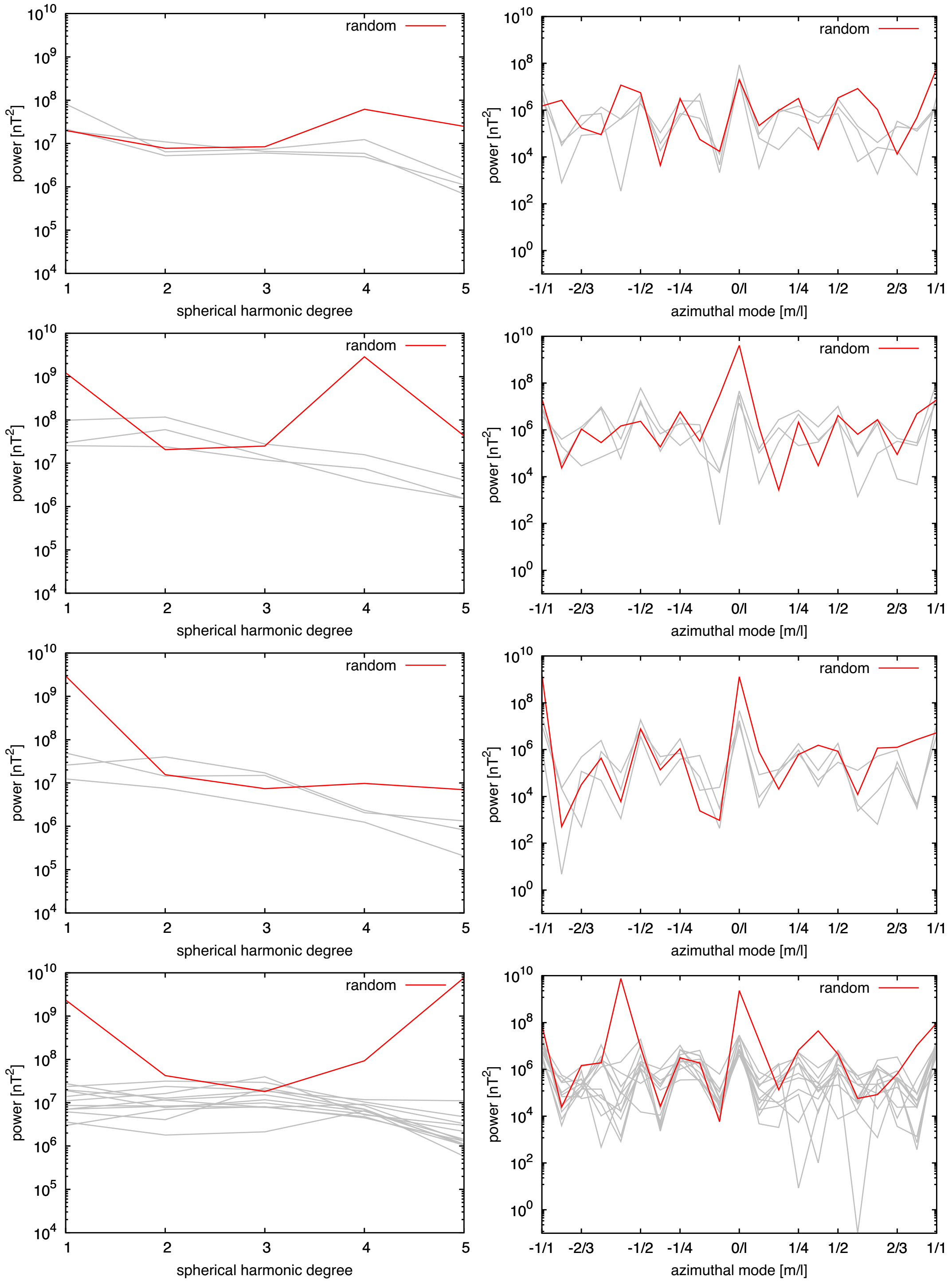


Fig. 4. Spectral differences (left) and azimuthal spectral differences (right) derived for the epochs −39,000, −31,000, −6200 and 1000 CE, respectively. Red curves are derived using a randomly shuffled magnetic field model, gray lines represent spectra derived from the differences of the mean and individual models. The ticks of the azimuthal spectral differences are arranged by the azimuthal modes, their distances do not vary linearly. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

red curve in the plots. For most epochs the spectral difference with the random is larger than those of the individual models. Only models of the Mono-Lake excursion show for some spherical harmonic degrees larger spectral differences as the random model. The azimuthal spectral differences are more heterogeneous, spectral differences for most of the azimuthal modes are well below the curve of the azimuthal spectral differences between mean and a random model, which is particularly true for zonal and sectorial terms of the individual models. This indicates a better agreement between mean and individual models in zonal and sectorial terms, than between a random model and individual models.

5. Common features of the averaged field

In this section we intend to provide further justification to derive a mean model from the set of archaeo- and paleomagnetic field models and evaluate the resemblance between models. First, we derive the degree correlation function (Arkani-Hamed et al., 1988; Toksöz et al., 1969) to quantify the similarity of the models during the different periods. The degree correlation, $r_i(\ell)$, between two models is defined as

$$r_i(\ell) = \frac{\sum_{m=0}^{\ell}\left(\widehat{g}_{\ell}^{m} g_{i,\ell}^{m} + \widehat{h}_{\ell}^{m} h_{i,\ell}^{m}\right)}{\sqrt{\left(\sum_{m=0}^{\ell}\left[\left(\widehat{g}_{\ell}^{m}\right)^2 + \left(\widehat{h}_{\ell}^{m}\right)^2\right]\right)\left(\sum_{m=0}^{\ell}\left[\left(g_{i,\ell}^{m}\right)^2 + \left(h_{i,\ell}^{m}\right)^2\right]\right)}}, \qquad (10)$$

where $\widehat{g}_{\ell}^{m}, \widehat{h}_{\ell}^{m}$ and $g_{i,\ell}^{m}, h_{i,\ell}^{m}$ are the Gauss coefficients of the mean and the individual model i, and ℓ, m their spherical harmonic degree and order, respectively. We derive $r(\ell)$, a function of the spherical harmonic degree between the mean model and individual models listed in Table 1.

In the text, we discuss the degree correlation of the three periods by computing an average over the degree correlation of the individual models which is also temporarily averaged. This is simply because the vast number of plots can not be accommodated in the main text. However, we note that the degree correlations of all epochs are evaluated.

Fig. 5 shows the averaged degree correlation for the three periods. The plots are arranged in the same manner, where the blue regions represent the averaged degree correlation between the mean model and the others, averaged over the respective period, i.e. Pleistocene, Holocene and Archaeological period. We intend to quantify the expected range of the degree correlation derived from random sets of Gauss coefficients. The random shuffling of Gauss coefficients eliminates any spatial correlation between them. Therefore, the separation of gray and blue areas reflects spatial correlation between the mean model and the individual field models. This, in turn, indicates common features between the mean model and the individual field models. The gray regions represent averaged degree correlation derived between the mean model and 30,000 field models with randomly shuffled Gauss coefficients. In all three plots of Fig. 5 gray and blue regions largely separate from each other, and therefore suggest common features of the different models to be captured by the mean field model. We note that the averaged degree correlation of the three periods is ~ 0.7 for spherical harmonic degrees > 1, and that most of the averaged degree correlations are above the 95 % level of significance, which also confidentially indicates similarity between the mean model and individual models.

A qualitative assessment of model similarities may be gained through drawing maps of different field models and comparing those to maps of the mean field model at given epochs. All models are truncated at spherical harmonic degree and order five. The reasons for the truncation were already provided in the previous section and simply relates to the limited spatial distribution of archeo- and paleomagnetic data which may not allow a finer spatial resolution of the geomagnetic field. In the following, Figs. 6 to 9 present maps of the individual models, the mean model, the Huber-weighted model and a map displaying the averaged Huber weights of the radial component, respectively.

During the Laschamp excursion at 39000 BCE, see Fig. 6, the radial field morphology largely deviates from being dipole dominated. The GGF100k and the lsmod2 models agree in the magnetic field amplitude, whereas the field features derived from the GGFSS70 model are less intense than those of the two others. However, the morphology appears similar with intense flux patches of alternating polarity occur around the geographic equator and also in polar regions. Further, the magnetic equator is heavily bended in maps of all three models. We note that only maps of the GGF100k model show an intense positive magnetic flux patch under West-Antarctica, whereas the two other models show negative flux in this region. Maps of the mean model and M_{HW} largely agree in the field morphology and only minor differences in the amplitude of some flux features. The subplot that is showing the averaged Huber weights of the radial component indicates that most disagreements between the individual models are found in polar regions, where weights are around 0.5 which is a significant down-weighting of some of the models. Agreement is found in a band around the geographic equator and the western Pacific, where models equally describe the magnetic field morphology. Here, Huber weights range between 0.9 and 1.0.

For the Mono-Lake excursion, see Fig. 7, we note that the GGFSS70 and GGF100k models show a higher resemblance than with the lsmod2. Maps of the mean field and M_{HW} largely match each other. The high resemblance of all models is also seen in the map of the averaged Huber weights, only relative small patches of down-weighting are located in the west Pacific and West Antarctica. The Huber-weights map may further indicate that the Huber-weights seem to be mainly determined by the two models which agree well (GGFSS70 and GGF100k). The dominant patches of the lsmod2 under the Caribbean and the mid Atlantic receive different Huber-weights. While the intense field in the Caribbean region more or less concur with the other two models, the mid Atlantic patch is down-weighted (or considered less robust) in the derivation of the Huber model. Similarly, this can be seen for the smaller Huber-weights in the West-Antarctic region, where the models disagree.

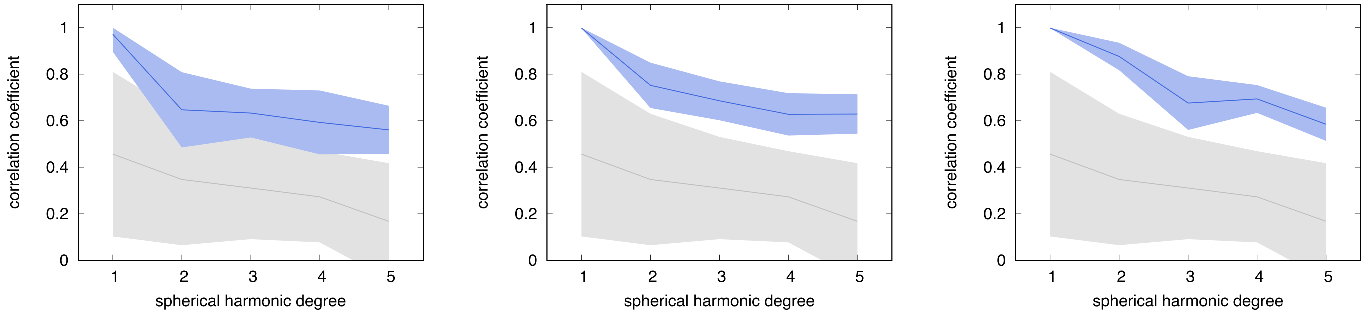


Fig. 5. The averaged degree correlations for the three different periods. From left: Pleistocene, Holocene and archaeological period. The blue areas represent the averaged degree correlation and its temporal variance. Gray areas represent the degree correlation between randomly shuffled field models and the mean model. (See the text for further description and discussion.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

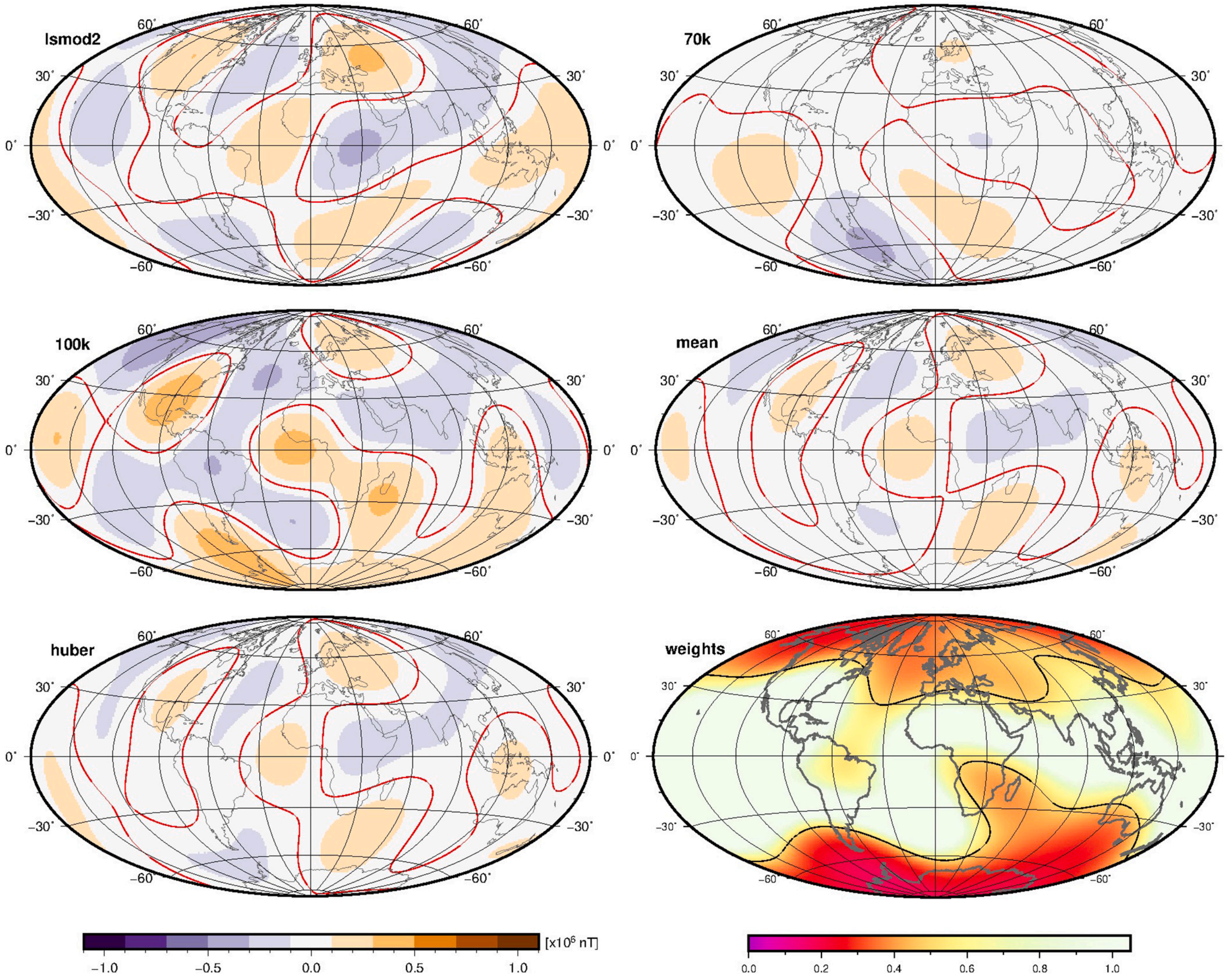


Fig. 6. The radial magnetic field at the core surface derived from three different models, as indicated in the left upper corner of each map including the mean model. The maps are derived for the center of the Laschamp excursion at 39000 BCE and share the same colour-bar that is attached to the map of the Huber-weighted model. The magnetic equator is displayed as a red curve. An additional map is showing the mean Huber weights of the B_r component, there the black curve marks weights = 0.5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

By comparing maps of the two excursions, we note that the radial magnetic field of the Mono-Lake excursion show a stronger dipolarity than maps of the Laschamp excursion, where the field dipolarity is disturbed by large-sized reversed flux patches in polar regions. This is not the case for the Mono-Lake excursion, where normal flux patches dominate their respective hemisphere and notch with their counterparts along the geographical equator, which leads to a highly bended magnetic equator.

Fig. 8 shows maps of the radial magnetic field at the core surface for the epoch 6200 BCE. As already seen from the curves of the individual Gauss coefficients (see Fig. 2), models can largely deviate from each other, sometimes by a few thousands nT. However, at this epoch Gauss coefficients are very similar, magnetic field maps show only minor deviations from each other. Most noticeable is a reverse flux patch in the eastern Pacific region seen in the pfm9k2 and aklm, and two strong normal flux patches in the north polar region seen in hfm and cals10k. In the mean field model and also the M_{HW} these spatial features are less pronounced, so that these models are largely dipole dominated. Note that aklm is not considered in the derivation of the mean model and the M_{HW}. The distribution of Huber weights show a larger dissension around Antarctica, north polar region and North America.

Figs. 9 and 10 show the radial component of the magnetic field at the core surface around 1000 CE. The maps show a wide range of spatial complexity and differ in some aspects. Some features are only seen in a few models like the presence of an intensified field patch east of Brazil, or the ditch of the magnetic equator in the Indian ocean. Maps derived from the Covarch, Covlake and the BigMudi4k models show rich and different spatial details, e.g. the curvature of magnetic equator and a number of normal and reversed flux patches in the northern polar region as well as in the equatorial region. Apart from these differences most models indicate the existence of a high normal flux patch west of North America. Less clear is the extend of the normal flux patch under Siberia that appears in all models, but at slightly different locations and with different shape. Patches in the Southern hemisphere at high latitude are less evident, which is possibly related to the sparser data distribution, but some of the models suggest the existence of large reverse flux under the Indian Ocean, i.e. AKLM, BigMudi4k and the SHAWQ2k. This anomaly may give rise to a South Indian Ocean Anomaly at Earth's surface, a possible ancestor of the South Atlantic Anomaly. The mean field model reduces the amplitude of a few features, but seems to accommodate a number of them. The mean field model and the M_{HW} largely agree with each other and only differ in details, like for the

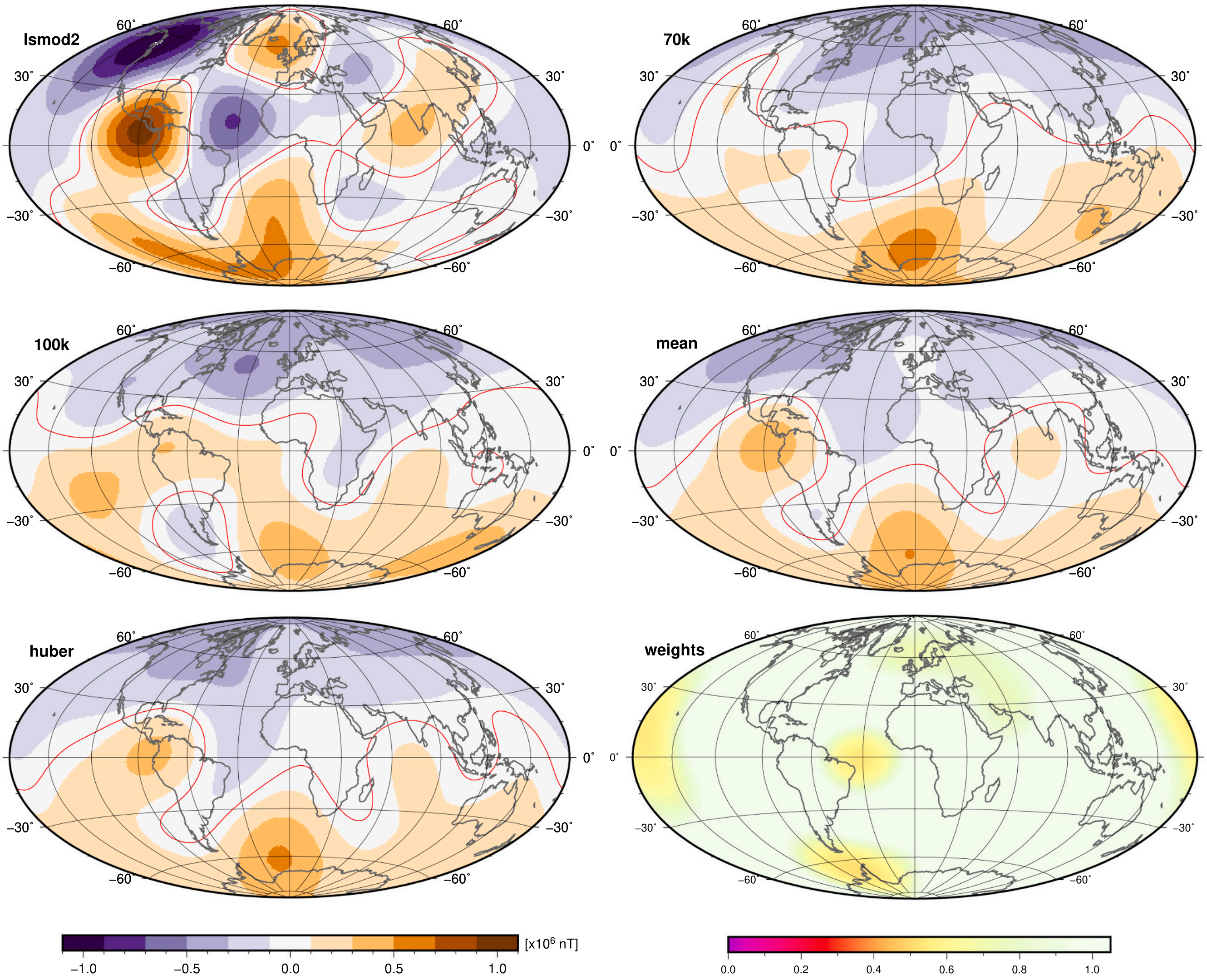


Fig. 7. Maps of the radial magnetic field at the core surface for the center of the Mono-Lake excursion at 31000 BCE and the according map of the Huber weights.

enhanced curvature of the magnetic equator under the Indian Ocean. The Huber weights suggest that models differ in some patches in the equatorial regions.

6. Criteria of the geomagnetic field morphology

Now, we turn to quantities that are used to assess the Earth likeliness of numerical dynamo simulations. These criteria measure the complexity of the geomagnetic field morphology and were introduced by (Christensen et al., 2010). Most of these criteria are calculated based on the power spectrum of the spherical harmonics coefficients (Lowes, 1974; Mauersberger, 1956):

$$R_\ell = \left(\frac{a}{c}\right)^{2\ell+4}(\ell+1)\sum_{m=0}^{\ell}\left(g_\ell^m\right)^2+\left(h_\ell^m\right)^2, \tag{11}$$

where ℓ, m are the spherical harmonic degree and order of the Gauss coefficients g_ℓ^m, h_ℓ^m, and a, c Earth's radius and the outer core radius, respectively. Similar to (8).

6.1. Axial dipole dominance

Earth's magnetic field being predominantly dipolar, we start by evaluating the level of axial dipolarity estimated by the ratio between the power spectrum of the axial dipole over the sum of power spectrum of all other spherical harmonics. Following (Christensen et al., 2010) who derived values of the Axial dipole dominance: AD/NAD as

$$\mathrm{AD}\Big/\mathrm{NAD} = \frac{P_{1,0}}{P_{1,1}+\sum_{\ell=2}^{\ell_{max}}(a/c)^{2\ell-2}\sum_{\ell,m}P_{\ell,m}}, \tag{12}$$

where the terms

$$\begin{aligned} P_{1,0} &= 2\left(g_1^0\right)^2 \\ P_{1,1} &= 2\left(\left(g_1^1\right)^2+\left(h_1^1\right)^2\right) \\ P_{\ell m} &= (\ell+1)\left(\left(g_\ell^m\right)^2+\left(h_\ell^m\right)^2\right) \end{aligned} \tag{13}$$

are divided by a common factor $(a/c)^6$ to ease the formulation. We, further, evaluate the axial dipole dominance by estimating the dipole latitude θ

$$\theta = \tan^{-1}\left(\frac{g_1^0}{\sqrt{\left(g_1^1\right)^2+\left(h_1^1\right)^2}}\right), \tag{14}$$

where g_1^0 is the axial, and g_1^1, h_1^1 are the equatorial dipole terms, respectively.

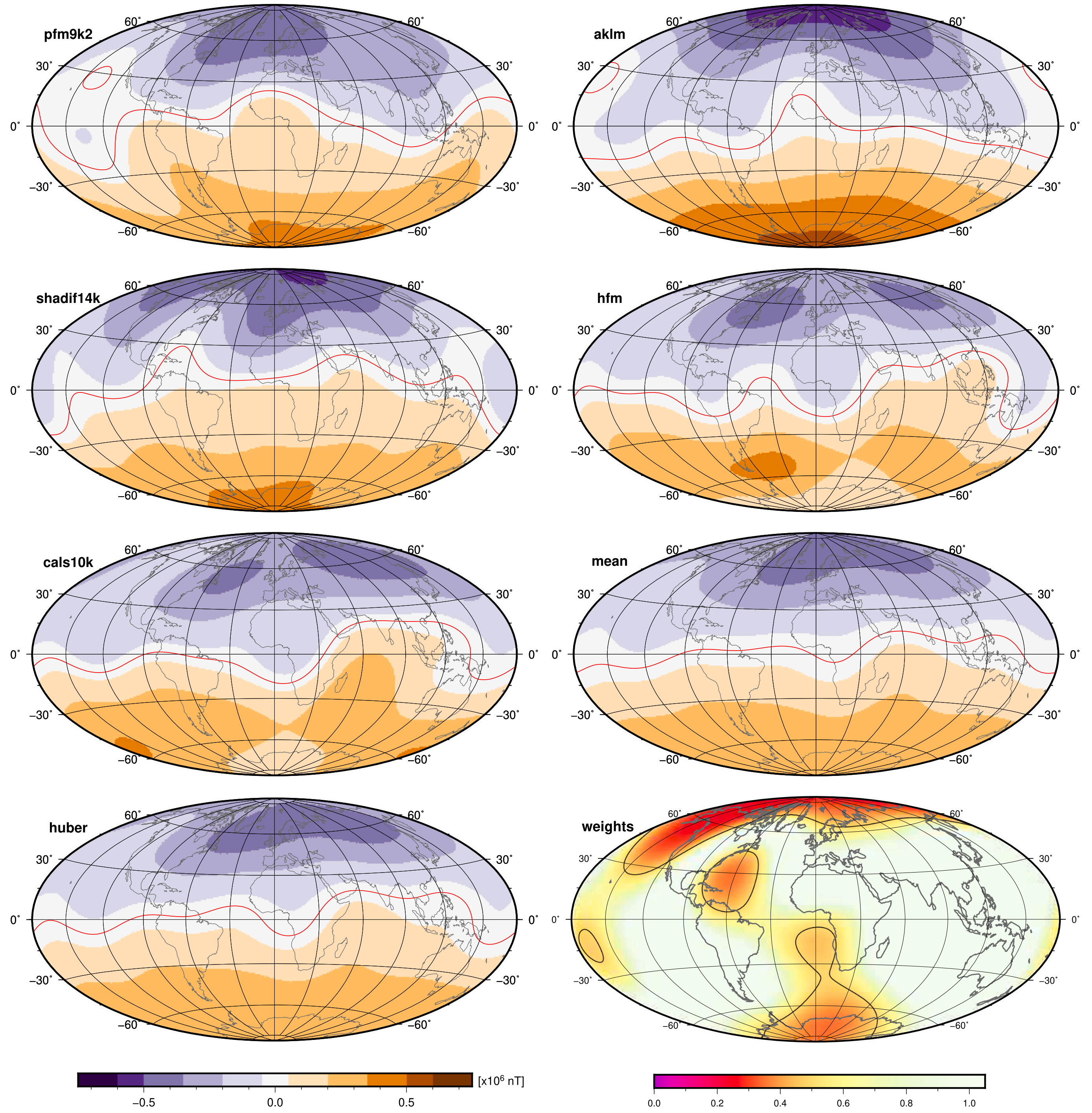


Fig. 8. Maps of the radial magnetic field at the core surface at 6200 BCE and the according map of the Huber weights.

(Panovska et al., 2019) investigated the impact of truncation on the Earth-like criteria, see their Figs. 22–24, table 3 and Figs. S2-S3. They showed that for archeo- and paleomagnetic field models the impact of truncating on $\ell_{max} = 5 - 10$ is weak. This is not the case for magnetic field models of the historical and satellite era, where the values of AD/NAD change between 1.54 and 0.85 for maximum degree truncation 5–13, respectively.

6.2. Field symmetries

A similar expression as (12) is used to quantify different symmetry properties of the magnetic field. The symmetry of the field with respect to the geographic equator is given by the O/E ratio. The O/E ratio relates the power in equatorial antisymmetric non-dipole components, O (coefficients with $\ell + m$ odd) to the power in equatorial symmetric non-dipole components, E (coefficients with $\ell + m$ even).

The zonality of the field is similarly evaluated by the ratio of zonal ($m = 0$) and non-zonal ($m \neq 0$) field power (Z/NZ). If the value is larger than unity, then this indicates a large axis-symmetry of the field, whereas small values suggest a deviation from zonality and from an axis-symmetric field morphology.

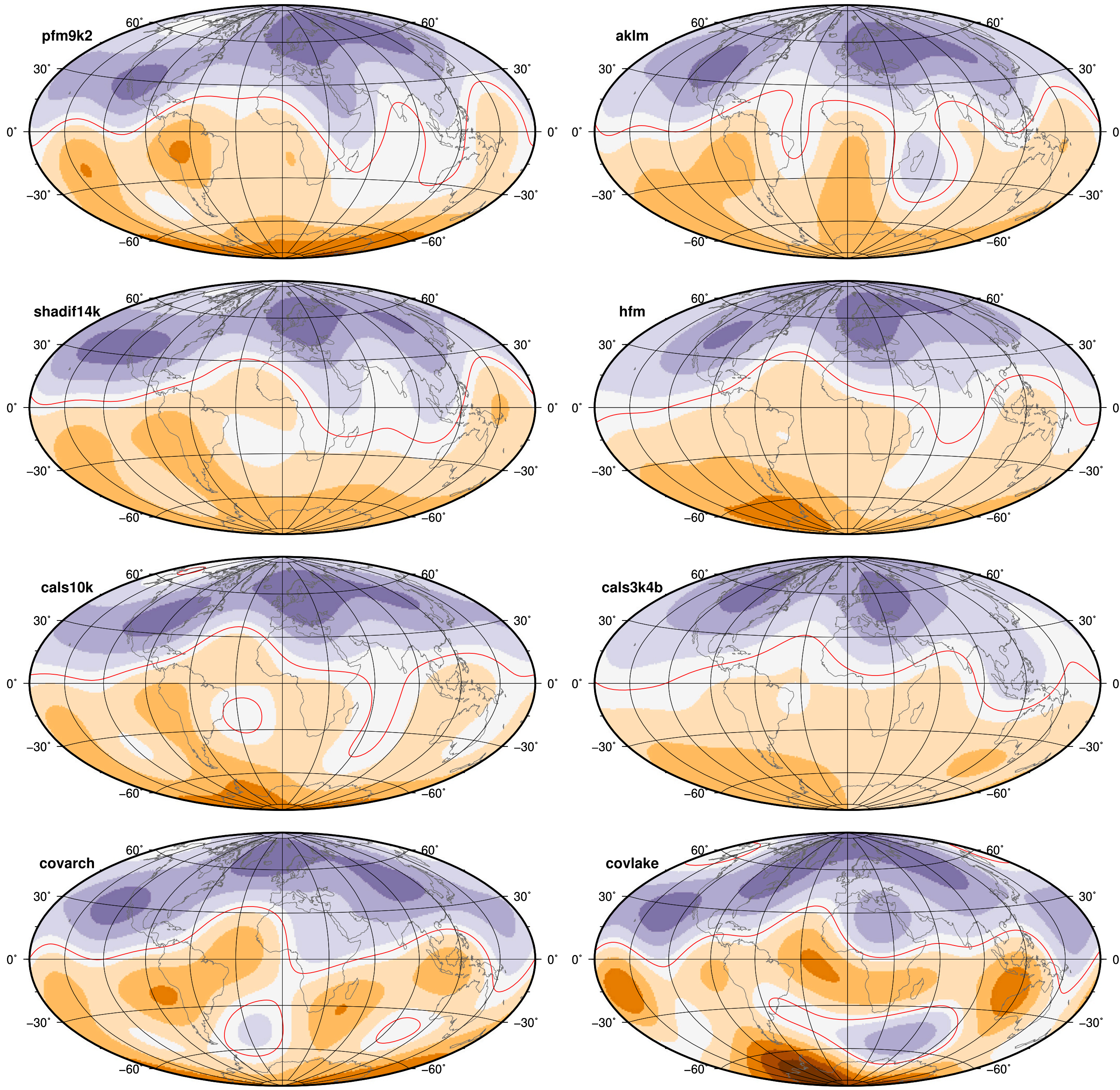


Fig. 9. Maps of the radial magnetic field at the core surface at 1000 CE. Colors are referred to the colour-bar in the next Figure. To be continued next Figure.

6.3. Flux concentration

(Christensen et al., 2010) suggested a fourth criterion to compare results of numerical dynamo simulations to the geomagnetic field, which is the flux concentration:

$$\mathrm{FCF} = \frac{\langle B_r^4 \rangle - \langle B_r^2 \rangle^2}{\langle B_r^2 \rangle}. \qquad (15)$$

B_r is the radial component of the magnetic field at the core surface and are the spherical averages over the core surface.

Table 2 summarizes all criteria during archeo- and paleomagnetic periods derived from the statistical mean model and the Huber-weighted model. It also lists mean values derived from the COV-OBS.x1 model truncated at $\ell_{max} = 5$.

Figs. 11, 12 and 13 show the dipole dominance and field symmetries for the three different periods, i.e. the upper Pleistocene, the Holocene and the archaeological and historical period, respectively. These graphics are derived from the statistical mean model.

The three periods show different ranges of the field dipolarity. During the upper Pleistocene it was the lowest and the magnetic field showed excursional behavior. During the Laschamp and Mono-Lake excursion the weak dipolarity is adjoined by a lower dipole latitude. The smallest dipole latitude (largest dipole tilt) is $\sim 69^\circ$ (21°) and occurred during the Laschamp excursion. It is also the smallest dipole latitude (largest dipole tilt) that occurred in the last 100 kyrs. The strongest dipolarity is displayed during the Holocene with a maximum around 6200 BCE. At that epoch the dipole field dominates the rest of the field by a factor of $\sim$20. Maps of the radial magnetic field at this epoch are shown in Fig. 8.

The archaeological period shows an intermediate dipolarity, that is

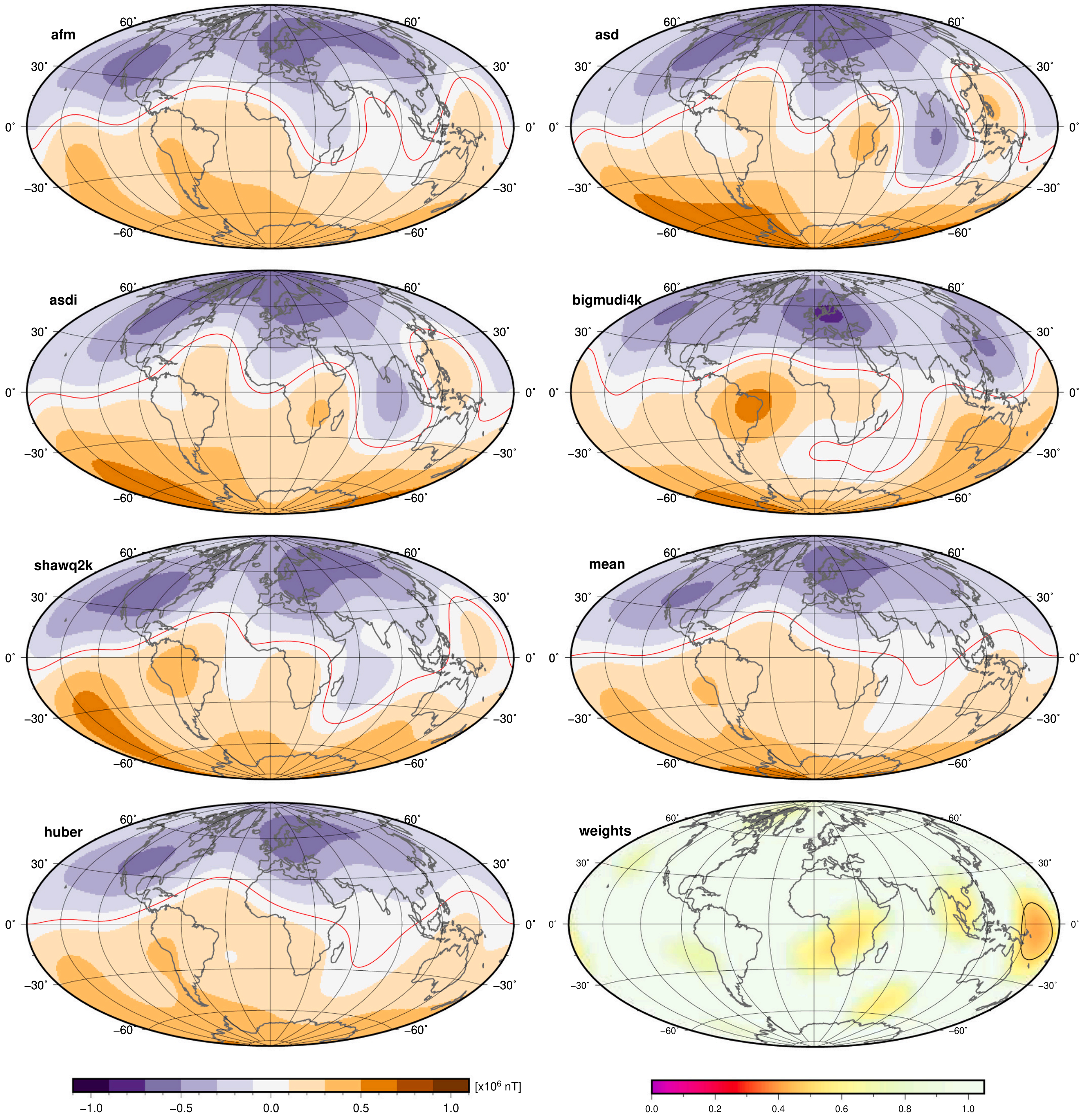


Fig. 10. Continued from Fig. 9: Maps of the radial magnetic field at the core surface at 1000 CE and the according map of the Huber weights.

less pronounced than in the Holocene, but stronger than during the Pleistocene. Also the dipole latitude gets to lower values in the archaeological and historical period. The symmetry properties of the magnetic field are very similar for the different periods. We note that the zonality is rather low and that the Z/NZ rarely gets close to 1.0, as it mostly varies around ~ 0.25. Therefore, the field never gets to a state where it is zonal, i.e. rotational symmetric. Non-zonality dominates the field morphology, also during the Laschamp and Mono-Lake excursions. Z/NZ drops to a minimum at the center of the Laschamp excursion, at 39000 BCE. The O/E value of the Pleistocene is slightly lower than for all other periods. We note occasional high O/E values during the excursion phases that indicate a strong equator-asymmetry in this period. The magnetic flux-concentration measured by the FCF (Christensen et al., 2010) shows its largest value during the Pleistocene. However, (Terra-Nova et al., 2024) find a tendency of this criteria to vary with the datatype, depending on whether sediments, lava flows and archaeoligical artifacts are used in the model derivation, which limits implications based on this criterion.

7. Discussion and conclusion

Our analysis showed that magnetic field models over the last decades of millennia have a wide range of temporal variability and of spatial content, see Figs. 1 to 4. While the temporal variability of the models

Table 2
Ranges and mean values of the structural criteria derived from the statistical mean and Huber-weighted field models.

	AD/NAD	*O/E*	*Z/NZ*	FCF
Statistical mean model				
range	0.05–25.9	0.2–3.5	0.01–1.0	0.7–2.8
Pleistocene	2.76	0.81	0.26	1.61
Holocene	10.03	0.95	0.29	1.12
Archeo	5.88	0.78	0.31	1.10
Huber-weighted model				
range	0.02–20.6	0.2–3.5	0.01–0.8	0.7–3.2
Pleistocene	2.45	0.74	0.27	1.64
Holocene	9.17	1.01	0.27	1.16
Archeo	5.77	0.79	0.30	1.12
COV-OBS	1.67	0.89	0.31	1.16

may represent the true variability of the magnetic field over the time span, their different spatial content represents to some account different modeling techniques and data sets. Therefore, the spatial complexity of the magnetic field morphology is uncertain. We attempt to reduce this uncertainty by truncating each model to spherical harmonic degree and order $\ell_{max} = 5$ and then to average these models for common epochs. The truncation to $\ell_{max} = 5$ can be justified by the limited spatial distribution of archeo- and paleomagnetic data. A truncation at $\ell_{max} = 5$ is supported by a companion study (Terra-Nova et al., 2024) who finds that the O/E and Z/NZ criteria derived from the COV-OBS.x2 (Huder et al., 2020) do not show a clear monotonic dependence with spherical harmonic degree, but rather an uneven dependence from degrees 5–6 onward.

To overcome spatial uncertainties we suggest to average field models for common epochs. Averaging a set of geomagnetic field models can sustain and emphasize field features that are commonly present in all models. Degree correlations (Fig. 5) showed a significant correlation between the mean field model and individual models for all spherical harmonic degrees. These correlations become more pronounced from the upper Pleistocene to the Archaeological period, which may be related to the growing number of available field models. We have studied two different ways of deriving an averaged field model: (i) the statistical mean model and (ii) the spherical averaging where models are projected onto spherical grids at Earth's surface that are stacked and inverted for a representative set of Gauss coefficients. Two schemes of deriving these spherical averaging are compared, and we decided to use results of the Huber-weighted inversion, as this scheme is also applied to derive IGRF-13 models (Alken et al., 2021a; Alken et al., 2021b). Results of the Huber-weighted inversion slightly differ from the statistical mean model, simply because the statistical mean model is an unweighted average, where each model is considered equally in the derivation of the mean model. The unweighted averaging might be problematic in case of an individual model that largely deviates from the rest of the models. Huber-weighting reduces the effect of outliers. However, marking models as outliers consequently requires to favor one modeling strategy upon another, which may not be surely justified, as we find that archeo- and paleomagnetic field models are also sensitive to the modeling strategy and their specific treatment of data uncertainties, see Appendix Figs. A.1 and A.2. This sensitivity can be illustrated as follows. The Bayesian methods (BI models) assume the temporal relations between models of adjacent epochs to fit a linear Gaussian process [e.g. 6, 48,

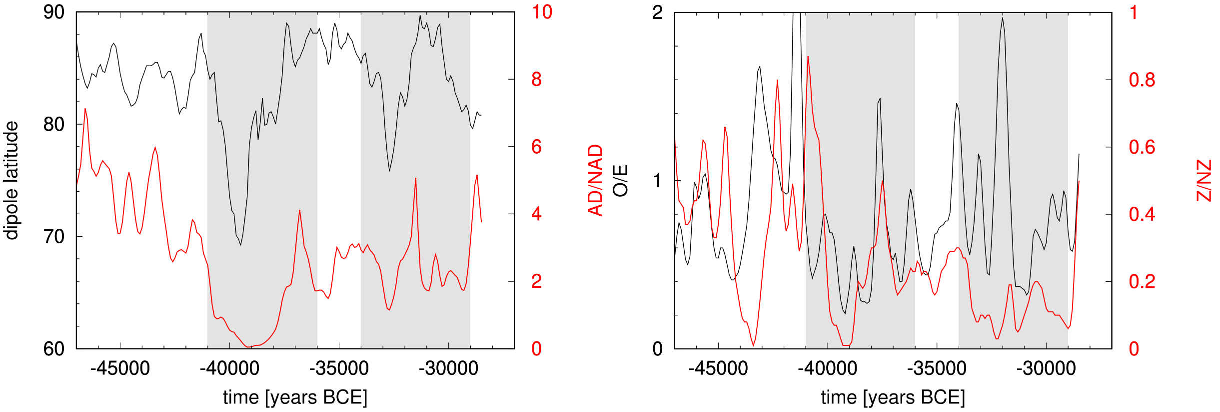


Fig. 11. Structural criteria derived from a mean model for the period from 47,000 to 28,000 BCE. The left panel shows criteria that quantify the dipole dominance, where the right panels shows the symmetry properties of the magnetic field. The gray regions mark the periods of the Laschamp and the Mono-Lake excursions.

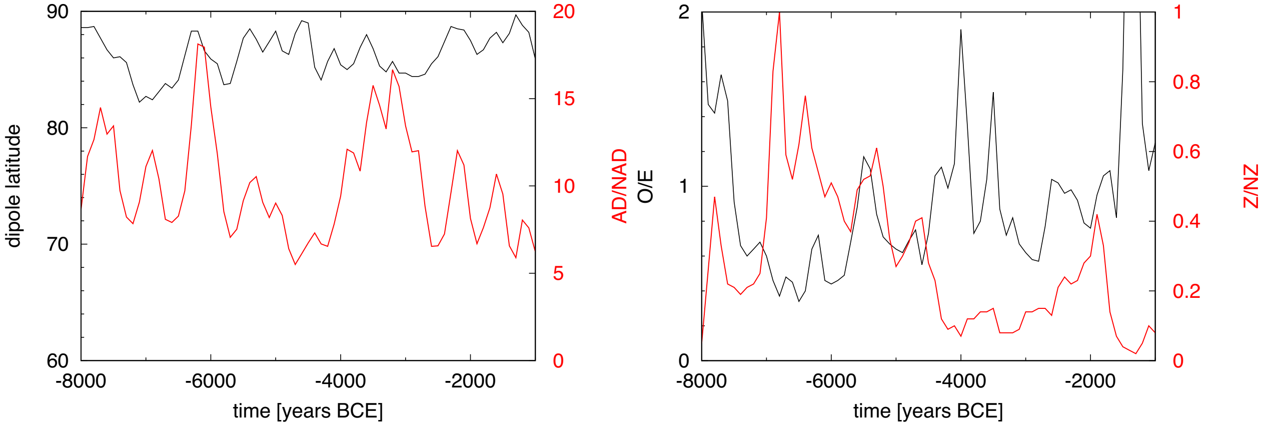


Fig. 12. Structural criteria derived from a mean model for the Holocene period from 8000 to 0 BCE. Same arrangement as the previous Figure.

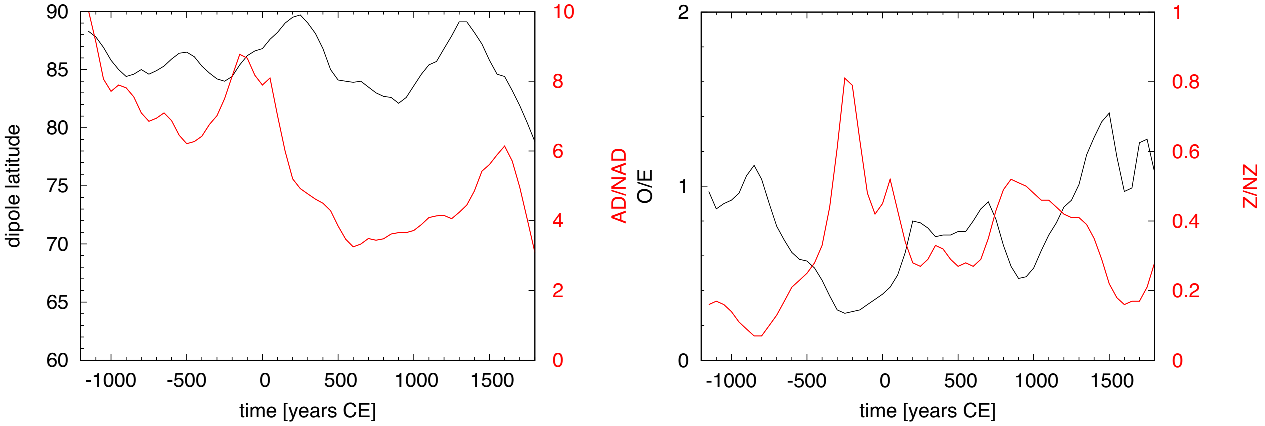


Fig. 13. Structural criteria derived from a mean model for the archeo and historical period from 0 to 1800 CE. Same arrangement as the previous Figure.

55]. Other models (SI models) are derived using cubic B-splines for a continuous representation in time [e.g. 20, 54]. It can be expected that BI models and SI models show different temporal behaviors. Indeed, spline based models that apply trade-off curves to obtain optimal damping parameters provide smoother and more dipolar models (Constable et al., 2016; Panovska et al., 2015). We may also expect that timing uncertainties have a different impact to the modeling strategies. Models, where adjacent epochs with different data uncertainties are connected linearly may be temporally rougher, than models that connect adjacent epochs using a spline interpolation. For the latter models, the data uncertainty at an epoch will be determined not only by the uncertainties of the direct neighboring epoch, but also from a wider vicinity, which is due to the definition of cubic B-splines to account for the uncertainties of the 4 neighboring epochs. This will directly translate into the spatial representation of magnetic field features and the derived structural criteria. In an ideal situation with equally distributed data and well-approximated uncertainties, the modeling strategies will result in very similar models (Alken et al., 2021b). This may be also seen from Appendix Figs. A.1 and A.2, where models based on different methods tend to agree towards periods with a better data distribution, i.e. 0 CE - 1900 CE. For earlier archaeological periods with poorer data distributions, the sensitivity to the modeling strategy and their specific treatment of data uncertainties dominates the sensitivity to the data distribution, as models that are using almost the same data set, but using different modeling techniques largely deviate. This can be seen in Fig. 14 where large regions have low Huber weights (disagreement between BI and SI models), in comparison to the Huber weight maps of Fig. 10 with large regions of high Huber weights (agreement between BI and SI models). However, the extent to which this may alter the derivation of the structural criteria remains to be estimated. Perhaps, future developments to derive consensus paleomagnetic field models from a wide set of models may include model uncertainties of individual field models, or some kind of scheme that rates the individual modeling strategy.

As an additional results the Huber-weighted inversion provides maps of the averaged Huber weights to qualify the agreement between models. However, these maps may not allow an interpretation in terms of model reliability and model robustness, as we do not have access to the error assumptions of the individual models. In the future such kind of information may be included, when deriving an averaged model from a set of models including their paleomagnetic data uncertainties. The access to inversion diagnostics (resolution matrix, etc.) and the individual trade-off behavior of the models may allow to provide error margins for the derivations of structural criteria and a better way to highlight the limitations of current archeo- and paleomagnetic field models.

The computation of a representative model also eases the comparison and discussion with results of numerical dynamo simulations, as we do not have to discuss field models individually. It allows to draw maps of the radial magnetic field at the core surface and to analyze these maps

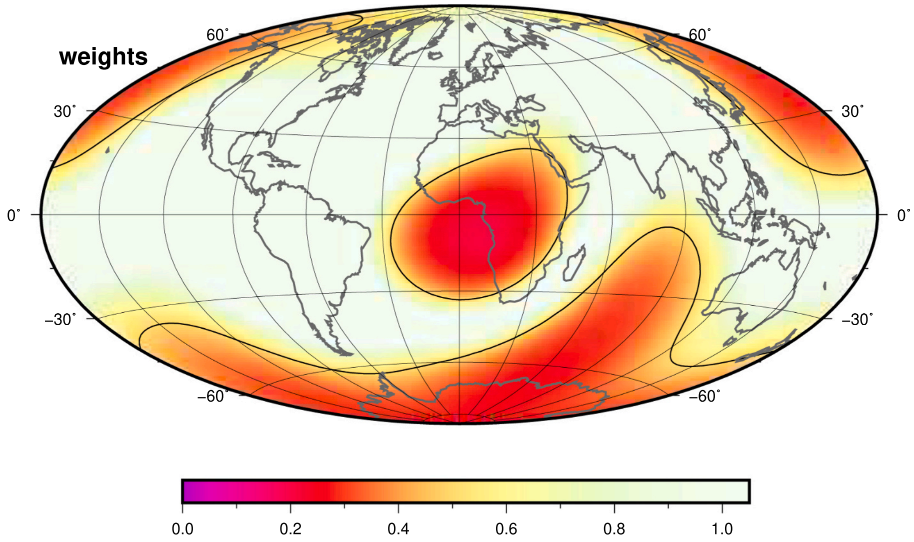


Fig. 14. Huber weights for 500 BCE.

and to discuss field features that are common in all models. Moreover, an averaged model provides statistically robust values for structural criteria of the geomagnetic field which can be compared with those resulting from dynamo simulations. Table 2 summarizes the structural criteria derived from the statistical mean model and Huber-weighted spherical averages. We find that the temporal variability of the structural criteria indicates larger ranges of these criteria as previously derived from historical field models (Christensen et al., 2010). The magnetic field dipolarity derived from the mean model ranges from 0.05 to $\sim$ 20 over the studied period. The dipolarity of the mean model is noticeably larger than the dipolarity of individual models. To what amount this is a result of the averaging remains to be studied further. We note, that such effect would emphasize large scale field structures and reduces the strength of small scale features. However, the large range of the dipolarity variation, and even half of it, may constitute the limited validity of temporal average of this quantity to evaluate the Earth-likeness of numerical dynamo simulations, see Table 2. On the contrary, this enormous range suggests a larger set of simulations to be considered as Earth-like. Even though the mean values of the symmetry properties (the ratios O/E and Z/NZ, see section 6.2) are comparable to the respective values derived from COV-OBS.x1, they vary considerably and are not always at the same scale than values derived for the current field. This may also widen the set of Earth-like dynamo simulations.

We note that the dipole strength was higher in the Holocene and at the beginning of the Archaeological period than it is for the current field and as it was on average in the Pleistocene. (Nilsson et al., 2024) concluded from production rates of cosmogenic radionuclides that there is a systematic underestimation of g_2^0 during these periods. The bias is likely due to an uneven data distribution between northern and southern hemispheres and due to inaccurate assumptions about the error distribution in the model derivation. This is very similar to a problem encountered in the derivation of large-scale global magnetic field models for the planet Mercury from hemispherical unevenly distributed satellite data (Thébault et al., 2018; Wardinski et al., 2021; Wardinski et al., 2019). These models are largely determined in the southern hemisphere by the model prior and led to a high cross-correlation between g_1^0 and g_2^0 and therefore limited resolution of g_2^0, which may also be the case for the archeo- and paleomagnetic field models. However, if the large variability of the dipole strength, to some extent, is robust against data biases, then this would suggest a variation of the field reversibility, i.e. the ability of the field to reverse. Clearly, the strong field dipolarity during the Holocene made it difficult for the field to reverse. Whereas, during the Pleistocene a rather low dipolarity concurred with transitional and excursionary field behavior. A similar dipole weakness of the current field led to speculations upon a nearing field reversal and/or excursion, but further ingredients are needed for a field reversal to occur (Brown et al., 2018). Therefore, a variation of the field reversibility further suggest that either the geodynamo varies from a non-reversing to a reversing regime, or that an external forcing flicks the geodynamo between these regimes in a continuous fashion on a time scale certainly longer than a few thousands years. The latter effect might be captured by quantities that are describing the influence of the lower mantle and the heterogeneous heat flux at the CMB onto the geodynamo (Terra-Nova et al., 2024).

So far, studies were performed to quantify the observable features over 10 Myr of the temporal averaged field (Meduri et al., 2021; Sprain et al., 2019) or to measure the morphological complexity of the geomagnetic field (Christensen et al., 2010; Terra-Nova et al., 2024) to provide criteria for comparisons with results of numerical dynamo simulations. Yet, the geomagnetic temporal variability on millennial time scale has been analyzed only marginally, and no specific criterion to quantify the temporal complexity or the dynamical behavior of the magnetic field and dynamo simulations has been devised. One reason for this is the challenging character of implementing such criterion, considering the temporal uncertainties related to the dating of paleomagnetic field samples. One way to account for these uncertainties was proposed by (Sprain et al., 2019), who derived statistical fluctuations of paleomagnetic key features from a time-averaged field over 10 Myr normalized by the dipole moment variance. This provides a set of criteria to compare with dynamo simulations. Another attempt was made by (Constable et al., 2016) to analyze the morphology of the paleosecular variation (PSV) over the last 10,000 years. They found that the hemispherical asymmetry of strong PSV in the atlantic hemisphere and weak PSV in the pacific region have persisted at least the last 10,000 years and possibly even longer (Panovska et al., 2018). (Dumberry and Bloxham, 2006; Dumberry and Finlay, 2007b; Wardinski and Korte, 2008) studied the long-term variation of the outer core dynamics and found that the motion in the liquid outer core undergoes different regimes of zonal flow direction and that this is reflected in the azimuthal drift of radial magnetic field features. These, possibly, regular changes in the core flow direction could be related to the convectional overturn in the outer core, which is a key feature of the dynamo action. Dynamical similarities related to the convectional overturn and the field drift may provide temporal criteria to compare the geomagnetic field and results of numerical dynamo simulations.

This contribution does not supersede any previous field model of the magnetic field for the Archaeological, Holocene and Pleistocene periods. Here, we only studied the feasibility of summarizing different magnetic field models that are obtained by using different modeling strategies. The derivation of archeo- and paleomagnetic reference field models should be addressed by an eventual community effort.

CRediT authorship contribution statement

Ingo Wardinski: Writing – review & editing, Writing – original draft, Software, Resources, Methodology, Investigation, Funding acquisition, Conceptualization. **Filipe Terra-Nova:** Writing – review & editing, Writing – original draft, Software, Investigation, Conceptualization. **Erwan Thébault:** Writing – review & editing, Software, Resources, Investigation.

Declaration of competing interest

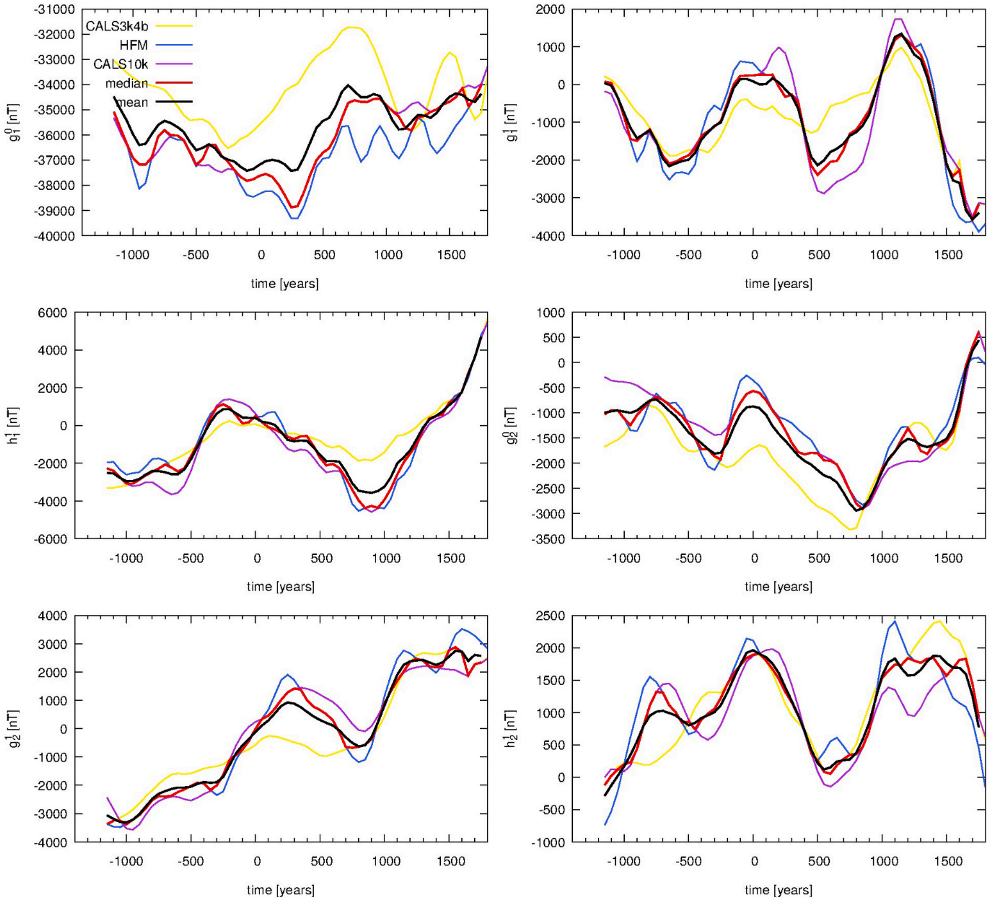
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

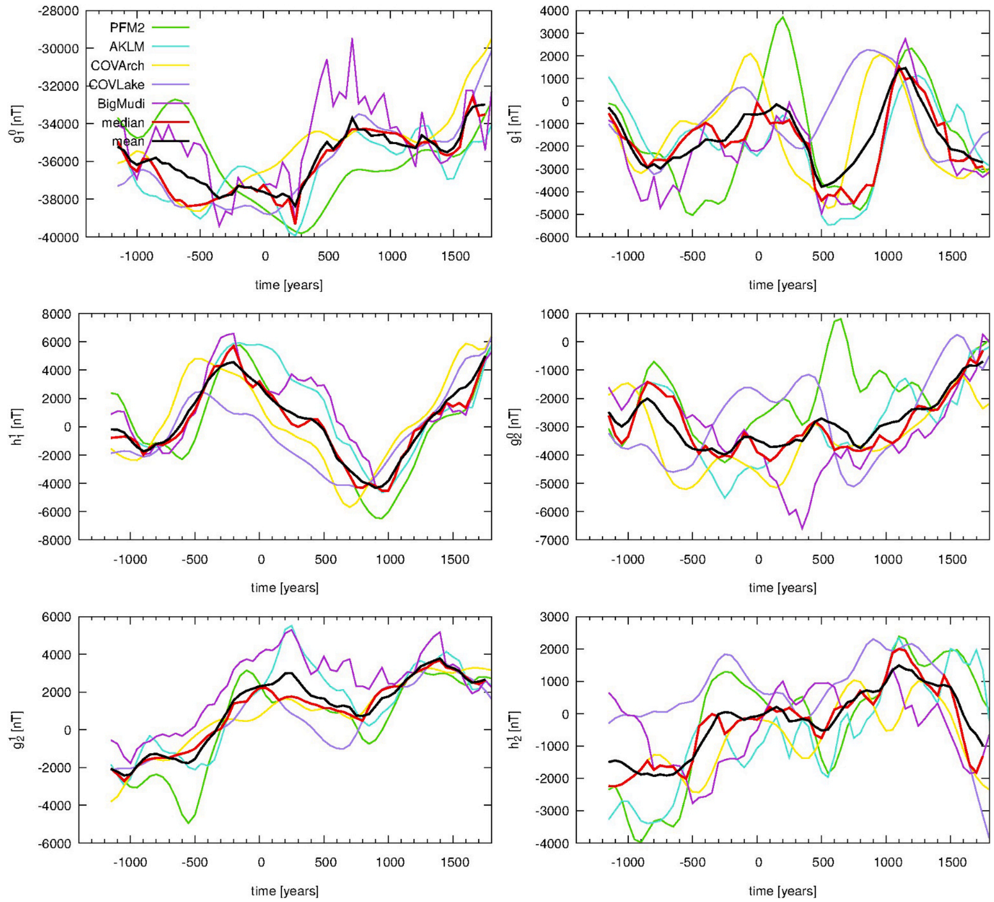
We are very grateful for the comments and criticisms of two reviewers which improved the manuscript. We acknowledge the financial support from the French Agence Nationale de Recherche, project DYRECOMB (grant ANR-22-CE49-0016–01).

Appendix A. Archeo- and Paleomagnetic field models

The Appendix Figs. A.1 & A.2 compare the temporal evolution of the SI-Models and BI-models respectively.



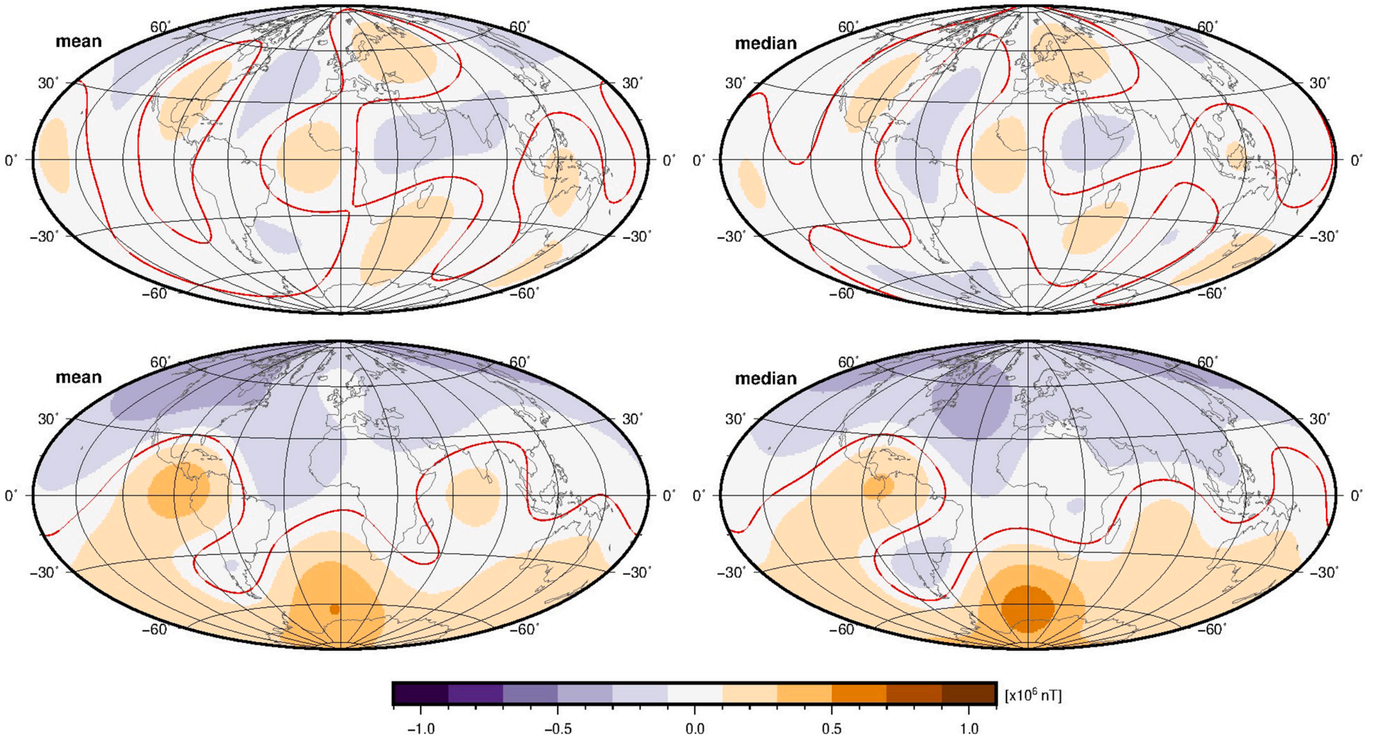
Appendix Fig. A.1. Temporal evolution of the first 6 Gauss-coefficients over the last 3000 years. Colored curves are derived from the Stochastic inversion models (SI-models) given in Table 1, their mean in red and the median in black, respectively.



Appendix Fig. A.2. Temporal evolution of the first 6 Gauss-coefficients over the last 3000 years. Colored curves are derived from the Bayesian inference models (BI-models), their mean in red and the median in black, respectively.

Comparing the results of this procedure, see Appendix Figs. A.1 & A.2, indicates that curves of the SI-models tend to gather more closely (Appendix Fig. A.1) than those of the BI-models (Appendix Fig. A.1). The reason for the large spread is not fully understood, but it is certainly related to the applied modeling technique, the individual treatment of age uncertainties and possibly are due to different data sets (different versions of *GEOMAGIA* database, that evolved over time).

Appendix B. Differences of mean and median field models during the excursions



Appendix Fig. B.1. Maps of the radial magnetic field at the core surface at 39000 BCE (top) and 31,000 BCE (bottom) derived from the statistical mean (left) and median (right) models.

As we have noted in section 3.1 there are differences between mean and median models during the Laschamp (39,000 BCE) and Mono-Lake (31,000 BCE) excursions. These differences show up in the times series plot of the Gauss coefficients, Fig. 1, though they do not look to be particularly large. Appendix Fig. B.1 shows maps of the radial magnetic field derived from the mean model and the median model at the center of the Laschamp and Mono-Lake excursion, respectively. Most noticeably are differences in the curvature of the magnetic equator during the Laschamp excursions (top panel of Appendix Fig. B.1). But the morphology of radial field, with reverse flux on either hemisphere, is similar. Less clearer are differences during the Mono-Lake excursions (bottom panel of Appendix Fig. B.1).

Data availability

The different types of models are publicly available and hosted in the Github repository at https://github.com/filipecros/Eval_Paleo_Archeo. Shell scripts to reproduce the figures will also be made available. The scripts utilize awk aho 1988 and GMT wessel 2013. Aho, A.V., Kernighan, B.W., Weinberger, P.J., 1988. The AWK programming language. Series in Computer Science. Addison-Wesley. Wessel, P., Smith, W. H.F., Scharroo, R., Luis, J., Wobbe, F., Nov. 2013. Generic Mapping Tools: Improved Version Released. EOS Transactions 94(45), 409-410.

References

Alken, P., Thébault, E., Beggan, C.D., Amit, H., Aubert, J., Baerenzung, J., Bondar, T.N., Brown, W.J., Califf, S., Chambodut, A., Chulliat, A., Cox, G.A., Finlay, C.C., Fournier, A., Gillet, N., Grayver, A., Hammer, M.D., Holschneider, M., Huder, L., Hulot, G., Jager, T., Kloss, C., Korte, M., Kuang, W., Kuvshinov, A., Langlais, B., Léger, J.M., Lesur, V., Livermore, P.W., Lowes, F.J., Macmillan, S., Magnes, W., Mandea, M., Marsal, S., Matzka, J., Metman, M.C., Minami, T., Morschhauser, A., Mound, J.E., Nair, M., Nakano, S., Olsen, N., Pavón-Carrasco, F.J., Petrov, V.G., Ropp, G., Rother, M., Sabaka, T.J., Sanchez, S., Saturnino, D., Schnepf, N.R., Shen, X., Stolle, C., Tangborn, A., Tøffner-Clausen, L., Toh, H., Torta, J.M., Varner, J., Vervelidou, F., Vigneron, P., Wardinski, I., Wicht, J., Woods, A., Yang, Y., Zeren, Z., Zhou, B., 2021a. International geomagnetic reference field: the thirteenth generation. Earth Planets Space 73 (1), 49. Dec.

Alken, P., Thébault, E., Beggan, C.D., Aubert, J., Baerenzung, J., Brown, W.J., Califf, S., Chulliat, A., Cox, G.A., Finlay, C.C., Fournier, A., Gillet, N., Hammer, M.D., Holschneider, M., Hulot, G., Korte, M., Lesur, V., Livermore, P.W., Lowes, F.J., Macmillan, S., Nair, M., Olsen, N., Ropp, G., Rother, M., Schnepf, N.R., Stolle, C., Toh, H., Vervelidou, F., Vigneron, P., Wardinski, I., 2021b. Evaluation of candidate models for the 13th generation international geomagnetic reference field. Earth Planets Space 73 (1), 48. Dec.

Amit, H., Choblet, G., 2009. Mantle-driven geodynamo features - effects of post-perovskite phase transition. Earth Planets Space 61, 1255–1268. Nov.

Amit, H., Christensen, U.R., Langlais, B., 2011. The influence of degree-1 mantle heterogeneity on the past dynamo of Mars. Phys. Earth Planet. Inter. 189, 63–79.

Arkani-Hamed, J., Zhao, S.K., Strangway, D.W., 1988. Geophysical interpretation of the magnetic anomalies of China derived from Magsat data. Geophys. J. Int. 95, 347–359.

Arneitz, P., Egli, R., Leonhardt, R., Fabian, K., 2019. A Bayesian iterative geomagnetic model with universal data input: self-consistent spherical harmonic evolution for the geomagnetic field over the last 4000 years. Phys. Earth Planet. Inter. 290, 57–75.

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 (1–4), 337–370. Aug.

Aubert, J., Finlay, C.C., Fournier, A., 2013. Bottom-up control of geomagnetic secular variation by the Earth's inner core. Nature 502 (7470), 219–223.

Backus, G.E., 1988a. Bayesian inference in geomagnetism. Geophys. J. R. Astron. Soc. 92, 125–142.

Backus, G.E., 1988b. Comparing hard and soft prior bounds in geophysical inverse problems. Geophys. J. R. Astron. Soc. 94, 249–261.

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 Space 61, 9–22. Jan.

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 (7572), 245–248. Oct.

Bonhommet, N., Babkine, J., 1967. Sur la presence d'aimentation inverse dans la Chaine des Puys. Seances de l'Acad. Sci. Ser. B 264, 92–94.

Brown, M.C., Donadini, F., Korte, M., Nilsson, A., Korhonen, K., Lodge, A., Lengyel, S.N., Constable, C.G., 2015a. GEOMAGIA50.v3: 1. General structure and modifications to the archeological and volcanic database. Earth Planets Space 67, 83.

Brown, M.C., Donadini, F., Nilsson, A., Panovska, S., Frank, U., Korhonen, K., Schuberth, M., Korte, M., Constable, C.G., 2015b. GEOMAGIA50.v3: 2. A new paleomagnetic database for lake and marine sediments. Earth Planets Space 67, 70.

Brown, M., Korte, M., Holme, R., Wardinski, I., Gunnarson, S., 2018. Earth's magnetic field is (probably) not reversing. Proc. Natl. Acad. Sci. 115 (20), 5111–5116.

Campuzano, S.A., Gómez-Paccard, M., Pavón-Carrasco, F.J., Osete, M.L., 2019. Emergence and evolution of the South Atlantic anomaly revealed by the new paleomagnetic reconstruction SHAWQ2k. Earth Planet. Sci. Lett. 512, 17–26. Apr.

Christensen, U.R., Wicht, J., 2007. Numerical dynamo simulations. In: Olson, P. (Ed.), Treatise on Geophysics, vol. 8. Elsevier Sci, New York, pp. 245–282.

Christensen, U.R., Aubert, J., Hulot, G., 2010. Conditions for Earth-like geodynamo models. Earth Planet. Sci. Lett. 296, 487–496.

Constable, C., Korte, M., Panovska, S., 2016. Persistent high paleosecular variation activity in southern hemisphere for at least 10 000 years. Earth Planet. Sci. Lett. 453, 78–86. Nov.

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. Geochem. Geophys. Geosyst. 19 (5), 1533–1558. May.

Denham, C.R., Cox, A., 1971. Evidence that the Laschamp polarity event did not occur 13 300-30 400 years ago. Earth Planet. Sci. Lett. 13 (1), 181–190. Dec.

Dumberry, M., Bloxham, J., 2006. Azimuthal flows in the Earth's core and changes in length of day at millennial timescales. Geophys. J. Int. 165, 32–46.

Dumberry, M., Finlay, C.C., 2007a. Eastward and westward drift of the Earth's magnetic field for the last three millennia. Earth Planet. Sci. Lett. 254, 146–157.

Dumberry, M., Finlay, C.C., 2007b. Eastward and westward drift of the Earth's magnetic field for the last three millennia. Earth Planet. Sci. Lett. 254, 146–157.

Efron, B., 2013. Bayesian inference and the parametric bootstrap arXiv e-prints, arXiv: 1301.2936. Jan.

Finlay, C.C., Maus, S., Beggan, C.D., Hamoudi, M., Lowes, F.J., Olsen, N., Thébault, E., 2010. Evaluation of candidate geomagnetic field models for IGRF-11. Earth Planets Space 62 (10), 787–804. Oct.

Gubbins, D., 1983. Geomagnetic field analysis – I. Stochastic inversion. Geophys. J. R. Astron. Soc. 73, 641–652.

Gubbins, D., Bloxham, J., 1985. Geomagnetic field analysis – III. Magnetic fields on the core-mantle boundary. Geophys. J. R. Astron. Soc. 80, 695–713.

Hellio, G., Gillet, N., 2018. Time-correlation-based regression of the geomagnetic field from archeological and sediment records. Geophys. J. Int. 214 (3), 1585–1607.

Huder, L., Gillet, N., Finlay, C.C., Hammer, M.D., Tchoungui, H., 2020. Cov-obs.x2: 180 years of geomagnetic field evolution from ground-based and satellite observations. Earth Planets Space 72, 160.

Jackson, D.D., 1979. The use of a priori data to resolve nonuniqueness in linear inversion. Geophys. J. R. Astron. Soc. 57, 137–157.

Jackson, A., Jonkers, A.R.T., Walker, M.R., 2000. Four centuries of geomagnetic secular variation from historical records. Philos. Trans. R. Soc. Lond. A358, 957–990.

Jonkers, A.R.T., Jackson, A., Murray, A., 2003. Four centuries of geomagnetic data from historical records. Rev. Geophys. 41. https://doi.org/10.1029/2002RG000115.

Korte, M., Constable, C.G., 2011. Improving geomagnetic field reconstructions for 0-3 ka. Phys. Earth Planet. Inter. 188 (3–4), 247–259.

Korte, M., Holme, R., 2010. On the persistence of geomagnetic flux lobes in global field models. Phys. Earth Planet. Inter. 182, 179–186.

Korte, M., Donadini, F., Constable, C.G., Jun., 2009. Geomagnetic field for 0-3 ka: 2. A new series of time-varying global models. Geochem. Geophys. Geosyst. 10 (6). Q06008.

Korte, M., Brown, M.C., Panovska, S., Wardinski, I., 2019. Robust characteristics of the Laschamp and Mono Lake geomagnetic excursions: results from global field models. Front. Earth Sci. 7, 86. Apr.

Lanci, L., Kissel, C., Leonhardt, R., Laj, C., 2008. Morphology of the Iceland Basin Excursion from a spherical harmonics analysis and an iterative Bayesian inversion procedure of sedimentary records. Phys. Earth Planet. Inter. 169 (1–4), 131–139. Aug.

Leonhardt, R., Fabian, K., 2007. Paleomagnetic reconstruction of the global geomagnetic field evolution during the Matuyama/Brunhes transition: iterative Bayesian inversion and independent verification. Earth Planet. Sci. Lett. 253 (1–2), 172–195. Jan.

Licht, A., Hulot, G., Gallet, Y., Thebault, E., 2013. Ensembles of low degree archeomagnetic field models for the past three millennia. Phys. Earth Planet. Inter. 224, 38–67.

Liddicoat, J.C., Coe, R.S., 1979. Mono Lake geomagnetic excursion. J. Geophys. Res. 84 (B1), 261–271. Jan.

Lowes, F., 1974. Spatial power spectrum of the main magnetic field. Geophys. J. R. Astron. Soc. 36, 717–730.

Mauersberger, P., 1956. Das Mittel der Energiedichte des geomagnetischen Hauptfeldes an der Erdoberfläche und seine säkulare Änderung. Gerlands Beitr. Geophys. 65, 207–215.

Meduri, D.G., Biggin, A.J., Davies, C.J., Bono, R.K., Sprain, C.J., Wicht, J., 2021. Numerical dynamo simulations reproduce paleomagnetic field behavior. Geophys. Res. Lett. 48 (5), e90544. Mar.

Nilsson, A., Suttie, N., 2021. Probabilistic approach to geomagnetic field modelling of data with age uncertainties and post-depositional magnetisations. Phys. Earth Planet. Inter. 317, 106737. Aug.

Nilsson, A., Holme, R., Korte, M., Suttie, N., Hill, M., 2014. Reconstructing Holocene geomagnetic field variation: new methods, models and implications. Geophys. J. Int. 198, 229–248.

Nilsson, A., Nguyen, L., Panovska, S., Herbst, M., Zheng, M., Suttie, N., Muscheler, R., 2024. Holocene solar activity inferred from global and hemispherical cosmic-ray proxy records. Nat. Geosci. 317, 1–6. Jun.

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 (1–2), 66–79. May.

Panovska, S., Korte, M., Finlay, C.C., Constable, C.G., 2015. Limitations in paleomagnetic data and modelling techniques and their impact on Holocene geomagnetic field models. Geophys. J. Int. 202 (1), 402–418. Jul.

Panovska, S., Constable, C.G., Brown, M.C., 2018. Global and regional assessments of paleosecular variation activity over the past 100 ka. Geochem. Geophys. Geosyst. 19 (5), 1559–1580. May.

Panovska, S., Korte, M., Constable, C.G., 2019. One hundred thousand years of geomagnetic field evolution. Rev. Geophys. 57 (4), 1289–1337. Dec.

Panovska, S., Korte, M., Liu, J., Nowaczyk, N., 2021. Global evolution and dynamics of the geomagnetic field in the 15-70 kyr period based on selected paleomagnetic sediment records. J. Geophys. Res. 126 (12), e2021JB022681. Dec.

Pavón-Carrasco, F.J., Osete, M.L., Torta, J.M., De Santis, A., 2014. A geomagnetic field model for the Holocene based on archaeomagnetic and lava flow data. Earth Planet. Sci. Lett. 388, 98–109. Feb.

Schanner, M., Korte, M., Holschneider, M., Feb., 2022. ArchKalmag14k: a Kalman-filter based global geomagnetic model for the Holocene. J. Geophys. Res. 127 (2), e2021JB023166.

Sprain, C.J., Biggin, A.J., Davies, C.J., Bono, R.K., Meduri, D.G., 2019. An assessment of long duration geodynamo simulations using new paleomagnetic modeling criteria (Q_{PM}). Earth Planet. Sci. Lett. 526, 115758. Nov.

Terra-Nova, F., Hartmann, G., Trindade, R., 2013. Evolution of non-dipole features of the geomagnetic field from archeomagnetic models. In: AGU SPRING MEETING, CANCUN, MEX.

Terra-Nova, F., Amit, H., Hartmann, G.A., Trindade, R.I.F., 2015. The time dependence of reversed archeomagnetic flux patches. J. Geophys. Res. 120 (2), 691–704.

Terra-Nova, F., Wardinski, I., Panovska, S., Korte, M., 2024. Novel geomagnetic field morphological criteria: from modern to Pleistocene eras. ısubmitted to JGR. Solid Earth.

Thébault, E., Finlay, C.C., Alken, P., Beggan, C.D., Canet, E., Chulliat, A., Langlais, B., Lesur, V., Lowes, F.J., Manoj, C., Rother, M., Schachtschneider, R., 2015. Evaluation of candidate geomagnetic field models for IGRF-12. Earth Planets Space 67, 112. Jul.

Thébault, E., Langlais, B., Oliveira, J.S., Amit, H., Leclercq, L., 2018. A time-averaged regional model of the Hermean magnetic field. Phys. Earth Planet. Inter. 276, 93–105. Mar.

Toksöz, M.N., Arkani-Hamed, J., Knight, C.A., 1969. Geophysical data and long-wave heterogeneities of the Earth's mantle. J. Geophys. Res. 74, 3751–3770.

Wardinski, I., Korte, M., 2008. The evolution of the core-surface flow over the last seven thousands years. J. Geophys. Res. 113, 05101.

Wardinski, I., Thébault, E., 2019. Modelling Internal and External Geomagnetic Fields Using Satellite Data. Special Publications of the International Union of Geodesy and Geophysics. Cambridge University Press, pp. 84–97.

Wardinski, I., Langlais, B., Thébault, E., 2019. Correlated time-varying magnetic fields and the core size of mercury. J. Geophys. Res. Planets 124 (8), 2178–2197.

Wardinski, I., Amit, H., Langlais, B., Thébault, E., 2021. The internal structure of mercury's core inferred from magnetic observations. J. Geophys. Res. 126 (12), e06792. Dec.