ArchKalMag14k: A Kalman-filter based global geomagnetic model for the Holocene

Maximilian Arthus Schanner[1], Monika Korte[1], and Matthias Holschneider[2]

[1]GFZ German Research Center for Geosciences
[2]University of Potsdam

November 22, 2022

Abstract

We propose a global geomagnetic field model for the last fourteen thousand years, based on thermoremanent records. We call the model ArchKalMag14k. ArchKalMag14k is constructed by modifying recently proposed algorithms, based on space-time correlations. Due to the amount of data and complexity of the model, the full Bayesian posterior is numerically intractable. To tackle this, we sequentialize the inversion by implementing a Kalman-filter with a fixed time step. Every step consists of a prediction, based on a degree dependent temporal covariance, and a correction via Gaussian process regression. Dating errors are treated via a noisy input formulation. Cross-correlations are re-introduced by a smoothing algorithm and model parameters are inferred from the data. Due to the specific statistical nature of the proposed algorithms, the model comes with space and time dependent uncertainty estimates. The new model ArchKalMag14k shows less variation in the large scale degrees than comparable models. Local predictions represent the underlying data and agree with comparable models, if the location is sampled well. Uncertainties are bigger for earlier times and in regions of sparse data coverage. We also use ArchKalMag14k to analyze the appearance and evolution of the South Atlantic anomaly together with reverse flux patches at the core mantel boundary, considering the model uncertainties. While we find good agreement with earlier models for recent times, our model suggests a different evolution of intensity minima prior to 1650 CE. In general, our results suggest that prior to 6000 BCE the database is not strong enough to support global models.

ArchKalMag14k: A Kalman-filter based global geomagnetic model for the Holocene

M. Schanner[1], M. Korte[1], M. Holschneider[2]

[1]German Research Centre for Geosciences GFZ, Section 2.3, Potsdam, Germany
[2]Applied Mathematics, University of Potsdam, Potsdam, Germany

Key Points:

- We propose a new global geomagnetic field model for the Holocene based on thermoremanent records.
- Existing algorithms based on space-time correlation are modified by sequentialization via a Kalman-filter and smoothing.
- The results suggest that prior to 6000 BCE the database is not strong enough to support global models.

Corresponding author: M. Schanner, `arthus@gfz-potsdam.de`

Abstract

We propose a global geomagnetic field model for the last fourteen thousand years, based on thermoremanent records. We call the model ArchKalMag14k. ArchKalMag14k is constructed by modifying recently proposed algorithms, based on space-time correlations. Due to the amount of data and complexity of the model, the full Bayesian posterior is numerically intractable. To tackle this, we sequentialize the inversion by implementing a Kalman-filter with a fixed time step. Every step consists of a prediction, based on a degree dependent temporal covariance, and a correction via Gaussian process regression. Dating errors are treated via a noisy input formulation. Cross-correlations are re-introduced by a smoothing algorithm and model parameters are inferred from the data. Due to the specific statistical nature of the proposed algorithms, the model comes with space and time dependent uncertainty estimates.

The new model ArchKalMag14k shows less variation in the large scale degrees than comparable models. Local predictions represent the underlying data and agree with comparable models, if the location is sampled well. Uncertainties are bigger for earlier times and in regions of sparse data coverage. We also use ArchKalMag14k to analyze the appearance and evolution of the South Atlantic anomaly together with reverse flux patches at the core mantel boundary, considering the model uncertainties. While we find good agreement with earlier models for recent times, our model suggests a different evolution of intensity minima prior to 1650 CE. In general, our results suggest that prior to 6000 BCE the database is not strong enough to support global models.

Plain Language Summary

We use data of archaeological and volcanic origin from the last fourteen thousand years to construct a global geomagnetic field model. We call the model ArchKalMag14k. The database is uneven in space, with significantly more records in the Northern hemisphere and multiple clusters. Further, the number of available records decreases in time with a distinct drop 6000 BCE. Previous studies introduced a modeling method that was adapted to this inhomogeneities, but could not be applied to the whole database for computational reasons. To tackle this, we modify the method and implement an approach which handles only a number of records at a time. Relations between the individual steps are re-introduced later in the algorithm. Uncertainties in the data and in their ages contribute to estimating reasonable model uncertainties. The model parameters are inferred from the data.

ArchKalMag14k shows less variation on a global scale than comparable models. On a local scale, predictions represent the underlying data and agree with comparable models, if the location is covered well by data. Uncertainties are bigger for times and regions of sparse data coverage. The results suggest that prior to 6000 BCE the database is not strong enough to support global models.

1 Introduction

Global field reconstructions of the past are a key tool for understanding the dynamics of the Earth's magnetic field and the underlying processes in the Earth's core (e.g. C. Constable & Korte, 2015). This includes studying the evolution of field features, such as dipole decay, the South Atlantic Anomaly (SAA) and flux patches (Hartmann & Pacca, 2009; Jackson & Finlay, 2015). In the past, several techniques for constructing global field models have been developed and employed. Truncated spherical harmonics (SH) in the spatial domain combined with spline interpolation in time are widely used (Jackson et al., 2000; Korte et al., 2009; Senftleben, 2019). In the eighties, C. G. Constable and Parker (1988) first proposed using Gaussian processes to model the field dynamics, but until recently, the technique had not been applied to global field modeling. Only in the last years, statistical methods implementing this approach have been suggested (Hellio & Gillet, 2018; Nilsson & Suttie, 2021).

While early models (Jackson et al., 2000; C. G. Constable et al., 2000; Korte & Constable, 2003) do not provide uncertainty estimates, more recent field models use ensemble techniques to quantify (modeling related) errors (Korte et al., 2009; Licht et al., 2013; Pavón-Carrasco et al., 2014; Hellio & Gillet, 2018; Senftleben, 2019). In contrast, Nilsson and Suttie (2021) (and earlier Hellio et al. (2014) for local field models) used a Bayesian formulation of the proposed Gaussian process (GP) approach, to estimate uncertainties based on the posterior distribution.

Holschneider et al. (2016) extended the GP approach to the spatial domain, to also reflect uncertainties resulting from the data distribution, and in two recent studies this method was adapted to paleomagnetic records (Mauerberger et al., 2020; M. Schanner et al., 2021). The major challenge with the modeling strategies proposed there is related to the inversion of large scale matrices, and the methods were found computationally unfeasible for the number of records available for the Holocene. In the area of modeling the recent field, this challenge was overcome by applying sequentialization by means of a Kalman-filter (Kalman, 1960) to the inversion problem (Baerenzung et al., 2020; Ropp et al., 2020). This way, models from a way higher number of satellite observations have been constructed, while retaining the strategies proposed by Holschneider et al. (2016). In this study we apply sequentialization to the earlier developed strategy (M. Schanner et al., 2021, in the following referred to as SMKH21) and propose a new global geomagnetic field model for the Holocene.

Usually, global geomagnetic field models are inferred from two classes of data: Data from materials with thermoremanent magnetisation, such as volcanic rocks, bricks or burnt clay fragments from archeologic sites, and data from marine or lacustrine sediments with embedded magnetic particles. In this paper we focus on the former class and loosely refer to it as archeomagnetic data. The extension to sediments poses several additional challenges, some of which are addressed and discussed by Nilsson and Suttie (2021). The a priori model that results from the sequentialization of SMKH21 is similar to the one proposed by Nilsson and Suttie (2021). Besides a focus on a different and smaller dataset, the main difference lies in the inversion procedure: While Nilsson and Suttie (2021) employ a probabilistic Markov Chain Monte-Carlo (MCMC) based strategy, we rely on a deterministic inversion based on Kalman-filtering.

The rest of this article is structured as follows: In Section 2 we discuss prior assumptions, showcase the modeling method and introduce the dataset. Section 3 contains a brief validation section, using synthetic data, but mainly focuses on the description of features of the new model, which are discussed in section 4. We conclude in Section 5 by reconsidering possible extensions and shortcomings of the method, as well as an outlook to future work.

2 Method and Data

2.1 Gaussian process based modeling

In the eighties C. G. Constable and Parker (1988) proposed using GPs to model the Earth's magnetic field (EMF). The technique was later applied by Gillet et al. (2013) and extended by Holschneider et al. (2016). A GP is a stochastic process that is uniquely characterized by a mean function $\bar{\boldsymbol{B}}$ and a covariance function $K_{\boldsymbol{B}}$

$$\boldsymbol{B} \sim \mathcal{GP}\big(\bar{\boldsymbol{B}},\ K_{\boldsymbol{B}}\big)\ . \tag{1}$$

Gaussian process based modeling is a Bayesian approach, where a GP is used as a prior and an update is given by some normal likelihood, relating observations to the magnetic field. The posterior is then a GP as well, so that the model is also uniquely characterized by a mean function and a covariance function (Rasmussen & Williams, 2006). The main difficulty in applying this technique to paleomagnetic records lies in constructing the normal likelihood, as archeomagnetic observations are non-linearly related to the magnetic field.

2.2 Data model

To apply GP based modeling, one has to construct a normal likelihood, relating observations to the magnetic field. In paleomagnetism, the observations are the field directions (declination D and inclination I) and intensity F. At locations $\boldsymbol{x}$ and times t, the data model can then be formulated as

$$o(\mathbf{x}) = \mathbf{H}(\boldsymbol{B}(\boldsymbol{x}, t)) + \boldsymbol{E} \ , \tag{2}$$

where the observation functional $\mathbf{H} = (D, I, F)$ contains the usual expressions for declination, inclination and intensity and $\boldsymbol{E}$ are the observation errors. This data model is not Gaussian, as $\mathbf{H}$ is non-linear. We linearize the observation functional, to construct a normal proxy for the data model (2):

$$D \approx \tilde{D} + \frac{1}{\tilde{F}_H^2} \begin{bmatrix} -\tilde{B}_E \\ \tilde{B}_N \\ 0 \end{bmatrix}^\top \boldsymbol{B} \ , \tag{3}$$

$$I \approx \tilde{I} + \frac{1}{\tilde{F}_H} \left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{\tilde{B}_Z}{\tilde{F}} \frac{\tilde{\boldsymbol{B}}}{\tilde{F}} \right)^\top \boldsymbol{B} \ , \tag{4}$$

$$F \approx \frac{\tilde{\boldsymbol{B}}^\top}{\tilde{F}} \boldsymbol{B} \ . \tag{5}$$

$\tilde{D}, \tilde{I}, \tilde{F}$ and $\tilde{\boldsymbol{B}}$ indicate the point of expansion (POE) and we summarize the linearized expressions as $\mathbf{H}_{\text{lin.}}$. The observation errors $\boldsymbol{E}$ are also non-Gaussian, as the directional errors are given by a Fisher-von Mises distribution. We approximate this two dimensional distribution with 95% confidence cone (α_{95}) by two centered normal distributions with standard deviations (Piper, 1989; Suttie & Nilsson, 2019)

$$\sigma_I = \frac{57.3^\circ}{140} \alpha_{95} \qquad \text{and} \qquad \sigma_D = \frac{1}{\cos o_I} \sigma_I \ . \tag{6}$$

We label these approximate errors $\boldsymbol{E}_{\text{prox.}}$. Next, we consider dating uncertainties as suggested in SMKH21. The precise times t at which the archeomagnetic specimen received their magnetization are unknown. Instead, a corrupted date $t_o = t + e_t$ is reported, and we consider e_t to be a centered normal error. This error in the inputs is handled by another linearization, as proposed by McHutchon and Rasmussen (2011, the noisy input Gaussian process (NIGP)). As the errors are centered, the a priori mean is not affected by this procedure. However, via linearization the dating uncertainties are translated into observation uncertainties, and the covariance gets an additional term

$$\Sigma_{tt'} \circ \partial_t \partial_{t'} K_{\boldsymbol{B}}(\mathbf{x}, \mathbf{x}')|_{t_o} \ . \tag{7}$$

Here $\Sigma_{tt'}$ is the dating error covariance matrix and $\circ$ is the Hadamard product, i.e. element wise multiplication along the t direction. To this end, $K_{\boldsymbol{B}}(\mathbf{x}, \mathbf{x}')$ is considered as a matrix consisting of 3×3 blocks. The effect of the NIGP model is thus the inclusion of dating errors as contributions to the data covariance, similar to measurement errors. The translation is realized by weighing the dating uncertainties by the second order time derivative of the kernel. This is related, but not equal, to the idea of using the secular variation to estimate the contribution of dating uncertainties (see e.g. Korte et al., 2005). Due to the GP structure of the proposed model, the covariance structure for the secular variation is available a priori. Finally, a residual term is added to cover any effects that are not modeled, like crustal field or ellipticity of the Earth. This way, the data model reads

$$o(\mathbf{x}) \approx \mathbf{H}_{\text{lin.}}\Big(\boldsymbol{B}(\boldsymbol{x}, t_o) - e_t^\top \partial_t \boldsymbol{B}(\boldsymbol{x}, t)|_{t_o} + \rho \boldsymbol{P}\Big) + \boldsymbol{E}_{\text{prox.}} \ . \tag{8}$$

2.3 A priori process

We consider the common SH expansion of the geomagnetic potential Φ, which is valid outside of the Earth's conducting core, assuming an insulating mantle:

$$\Phi(\mathbf{x}) = R \sum_\ell \left(\tfrac{R}{|\boldsymbol{x}|}\right)^{\ell+1} \sum_{-\ell \le m \le \ell} g_\ell^m(t)\, Y_\ell^m(\hat{\boldsymbol{x}}) \ . \tag{9}$$

$\hat{\boldsymbol{x}}$ is the unit vector $\boldsymbol{x}/|\boldsymbol{x}|$ and Y_ℓ^m refers to the real valued and Schmidt semi-normalized SH of degree ℓ and order m with related Gauss coefficient g_ℓ^m. From this, the Earth's magnetic field is given as the gradient

$$\boldsymbol{B} = -\nabla\Phi \ , \tag{10}$$

and mean and covariance function of the EMF can be derived from assumptions about correlations of the Gauss coefficients. A priori we assume all Gauss coefficients except for the axial dipole to be of zero mean. The axial dipole is assumed constant, with value γ_1^0. We assume all coefficients to be uncorrelated at a reference radius $R = 2800$ km within the Earth's core. This is the "virtual" source region where the field is uncorrelated, with no direct physical meaning. The magnetic field given by this assumption is only a valid representation of the actual field above the core-mantle boundary (CMB). Inside of the core it can be seen as an artificial connection of the physical field at the CMB to the virtual sources inside of the core. We assume two different a priori variances, one for the dipole coefficients α_{DP} and one for all higher degrees α_{ND}. For each coefficient we assume a temporal correlation in the form of an AR(2)-process, as proposed by Gillet et al. (2013) and employed also by others (Hellio & Gillet, 2018; Baerenzung et al., 2020; Ropp et al., 2020; Nilsson & Suttie, 2021). This way, the temporal correlation of each coefficient is given by

$$\rho_\ell(t-t') = \left(1 + \frac{|t-t'|}{\tau_\ell}\right) \exp\left(-\frac{|t-t'|}{\tau_\ell}\right) \ . \tag{11}$$

Similar to Baerenzung et al. (2020), we assume one correlation time τ_{DP} for the dipole and a relation for all higher degrees $\ell \ge 2$

$$\tau_\ell = \frac{\tau_{\text{ND}}}{\ell} \ . \tag{12}$$

The posterior may be smoother or more detailed than these scales, depending on the data.

2.4 Sequentialization

In previous studies (Mauerberger et al., 2020; M. Schanner et al., 2021) we aimed at performing standard GP regression in the introduced setting. However, as determining the hyperparameters of the model requires this regression to be performed many times, this proved to be computationally unfeasible. To overcome this, we perform a sequentialized inversion, in form of a Kalman filter (Kalman, 1960; Baerenzung et al., 2020). Starting at an initial time, the Kalman filter consists of a series of steps, each consisting of a prediction based on the current model and a correction, which updates the model if data is available. In contrast to the previous study SMKH21, this requires us to define a cutoff degree $\ell_{\max}$, so that the model can be characterized by a finite vector of coefficients and their derivatives $\boldsymbol{z} = (g_\ell^m, \dot{g}_\ell^m)$. The prediction equations from step i to $i+1$ are given by

$$\mathbb{E}\big[z_{i+1|i}\big] = \mathbf{F}_i \mathbb{E}[z_i] \tag{13}$$

$$\operatorname{Cov}\big[z_{i+1|i}, z_{i+1|i}\big] = \mathbf{F}_i \operatorname{Cov}[z_i, z_i]\, \mathbf{F}_i^\top + \tilde{\boldsymbol{\Sigma}} \ , \tag{14}$$

where

$$\mathrm{F}_i(\ell, \Delta t = t_{i+1} - t_i) = \begin{pmatrix} 1 + |\Delta t|/\tau_\ell & \Delta t \\ -\Delta t/\tau_\ell^2 & 1 - |\Delta t|/\tau_\ell \end{pmatrix} \exp\left(-\frac{|\Delta t|}{\tau_\ell}\right)$$

is the forward operator of the AR(2)-process and $\tilde{\mathbf{\Sigma}} = \mathbf{\Sigma} - \mathbf{F}\mathbf{\Sigma}\mathbf{F}^\top$ with the a priori correlations $\mathbf{\Sigma}$. The correction step consists of a Bayesian GP inversion, as described in detail in SMKH21. The linearization is performed around the current model, beginning with the prior. We run the Kalman filter "backwards", i.e. from modern times to the past, as the data distribution is sparser towards earlier years. We expect the bigger amount of data in the beginning of the filtering to constrain the model and improve the POE for earlier times. We choose a cutoff degree of $\ell_{\max} = 20$ and a step size of $\Delta t = 10$ years. Both choices are believed to allow for a way higher resolution than present in the data, so that every dynamic present in the data can be captured by the model. After running the Kalman-filter we run a smoothing algorithm, following the formulation of Rauch et al. (1965) (see also Baerenzung et al. (2020)). This way, cross correlations that are not present in the Kalman-filter are re-introduced to the posterior.

We store a set of coefficients every 50 years, so that the output of a sequentialized inversion consists of 281 sets of 440 main field coefficients, 440 secular variation coefficients and the respective covariances.

2.5 Hyperparameters

The a priori model depends on several parameters, that have to be inferred before the actual inversion can be performed. One approach (e.g. Hellio & Gillet, 2018; Nilsson & Suttie, 2021) is to infer these parameters from outside knowledge, for example from models based on observatory and satellite data. We followed this approach in selecting the reference radius R, which effectively controls the slope of the a priori spectrum, by comparison to the IGRF models. For the other parameters we suggest a more self-consistent strategy and estimate them based on a maximum likelihood procedure. This strategy did not work for the reference radius, most likely because the sparse data in earlier years do not constrain it well enough.

Consider the forward log-marginal likelihood

$$\mathcal{L}_{\text{fwd.}} = \sum_{i=1}^{n} \left[-\ln|\Sigma_{o,i}| - \frac{1}{2}(o_i - \mathrm{H}_{\text{lin.}}\boldsymbol{B}(\boldsymbol{z}_i))^\top \Sigma_{o,i}^{-1} (o_i - \mathrm{H}_{\text{lin.}}\boldsymbol{B}(\boldsymbol{z}_i)) \right] \tag{15}$$

with observations o and observation covariance Σ_o. The forward likelihood depends on the hyperparameters and is considered a measure for how good a choice of hyperparameters describes the data. We maximize this expression using LIPO-TR (King, 2009, 2017) and use the maximum estimator for the parameters in the inference. The search region is specified by lower and upper bounds for the hyperparameters, these are as follows:

$$\begin{aligned} -100\ \mu\mathrm{T} \le \gamma_1^0 \le -10\ \mu\mathrm{T} \qquad & 1\ \mu\mathrm{T} \le \alpha_\bullet \le 1000\ \mu\mathrm{T} \\ 10\ \text{yrs.} \le \tau_\bullet \le 2000\ \text{yrs.} \qquad & 0.01\ \mu\mathrm{T} \le \rho \le 5\ \mu\mathrm{T} \end{aligned}$$

where $\bullet$ stands for DP and ND.

2.6 Dataset

The dataset is a slight variation of all records from the archaeological and volcanic database from GEOMAGIA v3.4 (Brown et al., 2015) with ages between 12000 BCE and 2000 CE. Some of the records from Mexico contain wrong age and dating uncertainty estimates (Mahgoub, pers. comm.), so they have been altered or removed, if no better estimate was available. To identify records that deviate from the rest, we use a Naive Bayes classifier. This procedure is integrated into the Kalman-filter as follows:

When a step $i+1$ contains new data, we evaluate the probability of every record to either come from a normal distribution with standard deviation of the size of the reported error or from a flat distribution of larger variance ($(100°)^2$ for declination, $(50°)^2$ for inclination and $(100\mu\mathrm{T})^2$ for intensities). Records that are more likely to stem from the flat distribution are considered outliers. In comparison to the standard approach of rejecting all data that deviates by a specific amount from the model, this procedure is more flexible and allows larger deviations, especially if the current model reports high uncertainties. By this procedure 276 records are identified and removed from the dataset. The final dataset contains 18735 records from 11637 locations. It consists of 5611 declinations, 7028 inclinations and 6096 intensities.

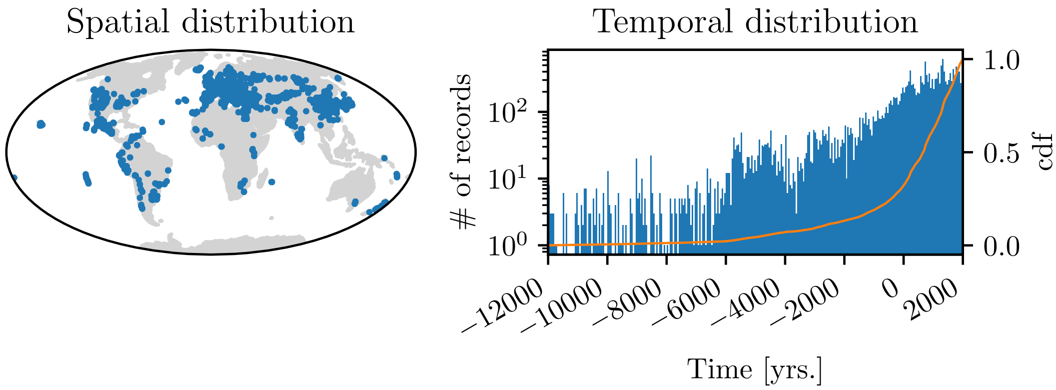

Figure 1. Spatial and temporal distribution of the data. Every declination, inclination and intensity is counted as one record and represented by one dot. Note the logarithmic scale (left) on the histogram. To emphasize the inhomogeneity in the temporal distribution, the normalized cumulative sum of the data is shown in orange (right scale).

3 Results

3.1 Validation

In order to validate the proposed modeling method, we performed a test inversion on synthetic data. We therefore set up a model with fixed hyperparameters and sampled coefficients from the prior distribution, which serve as reference. From these coefficients we generated data at the same input locations and times as the ones in the dataset described in section 2.6. The data was then corrupted by artificial noise from a Gamma distribution for the intensity and a von Mises-Fisher distribution for the directions and by normal noise in the ages. The error levels reported in the database were used. Table 1 shows the fixed hyperparameters and the inferred ones. Apart from one parameter they agree reasonably well. The deviance in the non-dipole correlation time is likely due to the data distribution. We believe that the inferred a priori correlation time is sufficient to resolve the variations that are present in the data. No additional contributions (white noise) were added to the synthetic dataset and the algorithm chooses the lowest possible value for the residual scaling accordingly.

Figure 2 shows generated and inferred axial dipole and quadrupole. Again, a promising agreement is observed, although some variation in the dipole, prominently between 10000 and 8000 BCE, is not resolved in the inferred model. This already hints at the data not containing enough information to recover global features during early times. Further figures from the validation process, showing the other dipole and some higher order coefficients, are available with the supplementary material.

Table 1. Hyperparameters that have been used to generate synthetic data for the validation ("fixed") and the ones inferred using the proposed method.[2]

Model	γ_1^0 [μT]	α_{DP} [μT]	τ_{DP} [yrs.]	α_{ND} [μT]	τ_{ND} [yrs.]	ρ [μT]
Fixed	-412.3	13.8	250	39.4	393	-
Inferred	-408.55	9.87	302.48	30.70	724.76	0.01

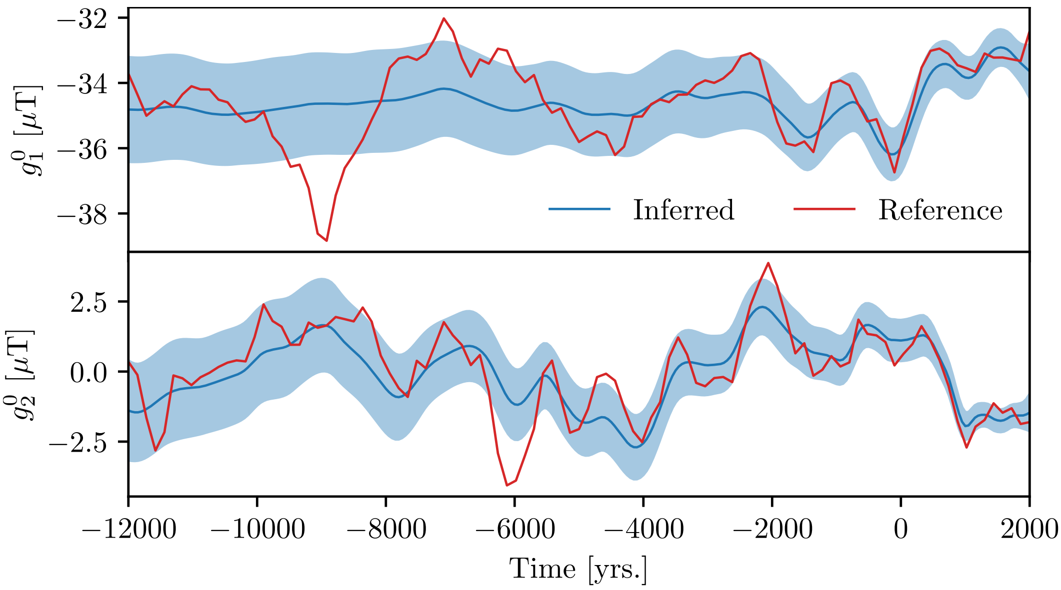

Figure 2. Axial dipole (top) and quadrupole (bottom) of the synthetic model, together with the corresponding inferred ones from the proposed inversion. The inferred (blue) and reference curves (red) agree within the one-sigma region shown in light blue. Some variations, most prominently in the axial dipole between 10000 and 8000 BCE, can not be resolved.

3.2 ArchKalMag14k

In the following we propose and describe a new global geomagnetic field model, based on archeomagnetic records. It covers the last 14000 years and we call it ArchKalMag-14k, as it is based on methods similar to the KalMag model by Baerenzung et al. (2020). The hyperparameters that maximize the marginal likelihood and define the prior used for constructing the model are given in Table 2. We compare ArchKalMag14k to the models ARCH10k.1 (C. Constable et al., 2016) and SHA.DIF.14k (Pavón-Carrasco et al., 2014), as both rest on a similar database and cover a similar timespan.

Running the inversion as described in Section 2 gives 281 sets of 440 main field and 440 secular variation coefficients together with the respective covariances, one set every 50 years. Figure 3 shows the dipole and axial quadrupole and octopole coefficients together with 95%-uncertainties and comparison models. The proposed model ArchKal-Mag14k shows less variation in the dipole degrees than comparable models, especially

[2] γ_1^0 is the constant a priori axial dipole, α_{DP} and α_{ND} give the a priori scaling of the dipole and non-dipole covariance kernel respectively. τ_{DP} and τ_{ND} give the corresponding a priori correlation times. ρ is the scaling factor of the residual term. Note that γ_1^0 and $\alpha_\bullet$ are given at the reference radius.

Table 2. Prior hyperparameters for ArchKalMag14k. Note that γ_1^0 and $\alpha_\bullet$ are given at the reference radius. At the Earth's surface, $\gamma_1^0 \approx -36.19$ μT.

γ_1^0 [μT]	α_{DP} [μT]	τ_{DP} [yrs.]	α_{ND} [μT]	τ_{ND} [yrs.]	ρ [μT]
-426.33	28.66	183.22	111.63	316.00	3.35

during earlier times when data is sparse. More variation is present in the quadrupole and octopole, with variation decreasing towards earlier times.

This behavior is also reflected in the power spectra. Figure 4 shows the spatial (top row) and secular variation (bottom row) spectra for two selected epochs, one with dense (1000 CE) and one with sparse (6000 BCE) data coverage. The blue lines show the power spectrum as a random variable, together with the corresponding prior as a light blue dashed line. These curves represent the non-linear transformations of the prior and posterior distribution. We also plot the power spectrum of the mean model (grey lines), i.e. the power spectrum directly inferred from the mean coefficients. The random variable gives higher values than the mean and comparison models, as it also includes the variance of the coefficients. The random variable can be compared to the prior, to determine the model resolution, while the power spectrum of the mean is better suited for comparison to existing models. For the recent epoch, the spectrum lies between the one for ARCH10k.1 (orange) and SHA.DIF.14k (green). For the earlier epoch, more power is present in degrees 2 and 3 and a more rapid decrease in power is observed for the higher degrees, than in the comparison models. For the secular variation the prior is reproduced from degree 3 on at both epochs. For the earlier epoch, the dipole secular variation power is also close to the prior. The mean model shows less secular variation in the dipole than the comparison models, with more power in degrees 2 to 4. For the recent epoch, more variation is observed in the higher degrees with a more rapid decrease in power for the earlier epoch, similar to the spatial spectrum.

Figures 5 and 6 show local curves for Paris and Hawaii respectively. Data from a surrounding of 250km is translated to the location of prediction. Inclination and intensity are translated along the corresponding axial dipoles (Merrill et al., 1996). Declinations are taken as reported. The two locations were chosen because they have very different data coverage: Paris is covered well during recent times with a decrease in data from 1000 BCE on and virtually no data for epochs earlier than 6000 BCE. This is reflected in the prediction curves, which show less variation and increasing uncertainties for times with low data coverage. Hawaii is not as densely covered during recent times, but due to the volcanic area, records are available over the whole timespan of the model. Consequently, the predictions show variations during earlier times and the reported uncertainties are smaller. The comparison models agree within the reported 95%-intervals for both locations. For Paris, the SHA.DIF.14k model shows more variation during times earlier than 5000 BCE and most prominently from 12000 to 8000 BCE. For Hawaii, all models show a similar amount of variation, with SHA.DIF.14k varying slightly more and ARCH10k.1 slightly less, especially in the intensity.

3.3 Dipole moment and location

During the Holocene, the geomagnetic field is dipole dominated. Therefore it is of special interest to infer the dynamics of the dipole. Figure 7 shows the evolution of the dipole moment. To access the dipole moment mean and standard deviation, sampling techniques are employed. The proposed model ArchKalMag14k shows significantly less variation in the dipole moment than comparable models. We observe some rapid variations from 1000 BCE to today, but for earlier times no rapid variations are found. In-

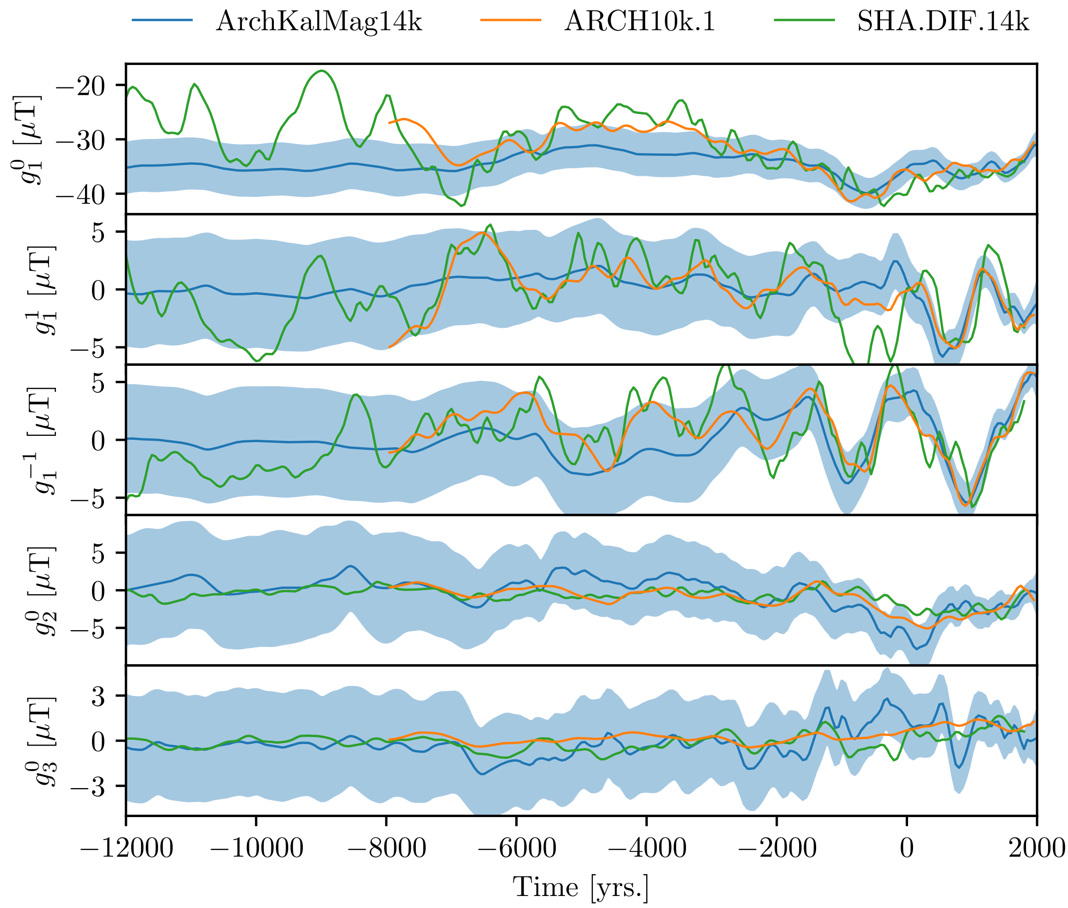

Figure 3. Gauss coefficients of the dipole and the axial quadru- and octopole. ArchKalMag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green.

terestingly we observe a higher dipole moment than the comparison models for the interval 6000 to 2000 BCE and also from 12000 to 8000 BCE.

Figure 8 shows the latitude and longitude of the dipole location, together with the angular standard deviation (Butler, 2004). The latter is inferred via sampling. In earlier studies (Mauerberger et al., 2020; M. Schanner et al., 2021) we analyzed the statistics of the dipole axis coordinates directly. Here we analyze the projection of the dipole onto the sphere instead. The corresponding distribution is approximated by a von Mises-Fisher distribution and we report the latitude and longitude of its location parameter, instead of the mean of the marginal distributions. The advantage of performing statistics on the sphere instead of considering the marginal distribution is that there is no critical point (resp. meridian). The disadvantage is that the distribution is not available in closed form and that uncertainties can not easily be translated to latitude and longitude, as approximations become unreliable when close to the pole (singularity in Eq. 6). Similar to the dipole moment, the proposed model shows less variation during earlier times. The dipole latitude shows a trend opposite to the SHA.DIF.14k model for the interval 12000 to 6000 BCE, with the geomagnetic pole being very close to the geographic one in the beginning and a decrease in latitude towards recent times, in contrast to an increase present in the SHA.DIF.14k model. The angular standard deviation (Figure 8,

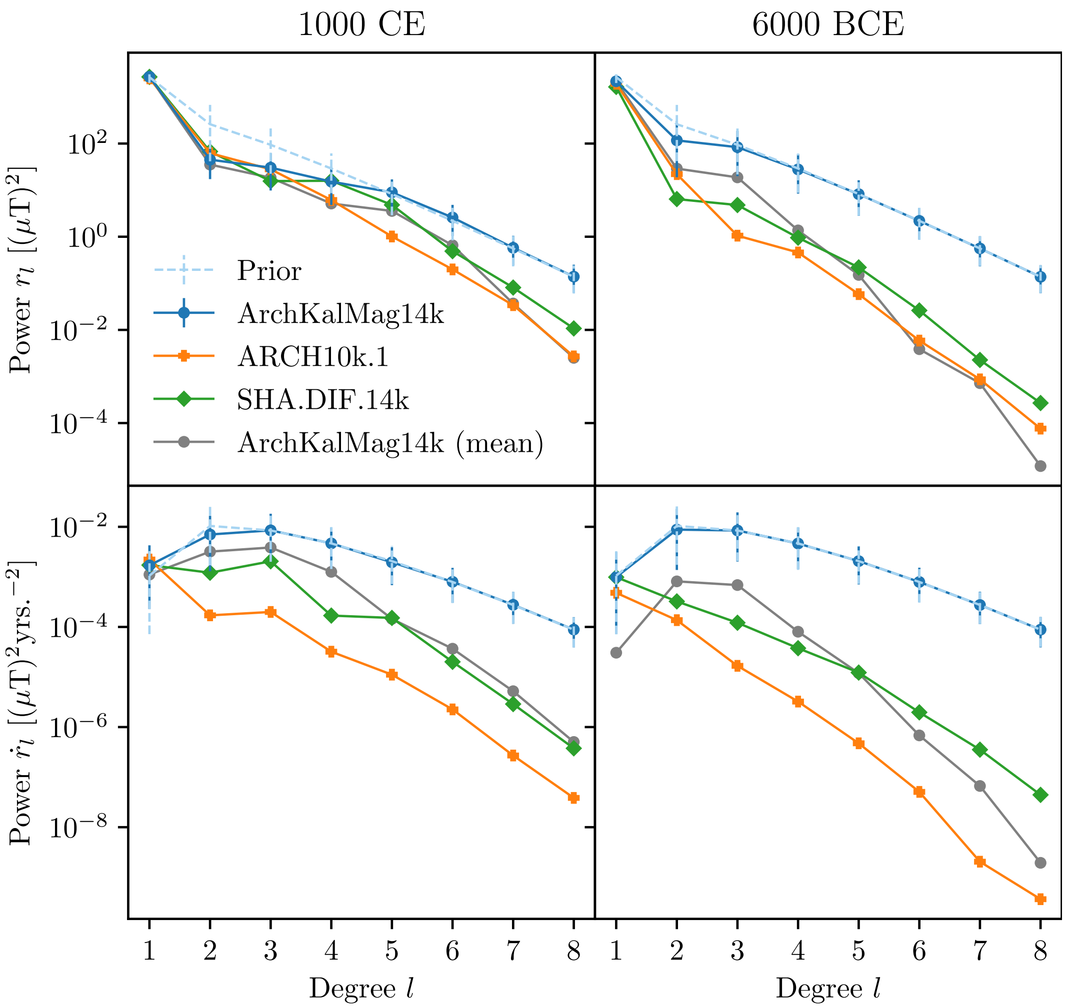

Figure 4. Geomagnetic main field (top) and secular variation spectra (bottom) at Earth's surface for two selected epochs. The random variable power spectrum for ArchKalMag14k is shown in blue. The errorbars report 2.5- and 97.5-percentiles, covering 95%. For comparison, the spectra of the mean model are shown in grey. The prior spectrum is shown as a light blue dashed line. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. See the text for additional discussion.

bottom row) increases towards earlier times, as is expected from the thinning data distribution.

3.4 South Atlantic anomaly

To conclude the results, we present investigations of the South Atlantic Anomaly (SAA). The SAA is a region of low field intensity, that has been linked to reverse flux patches at the CMB during recent times (e.g. Terra-Nova et al., 2017). We compare the appearance and evolution of the SAA as predicted by ArchKalMag14k to other studies (Hartmann & Pacca, 2009; Campuzano et al., 2019). We do not follow the kernel-based approach of Terra-Nova et al. (2017), but investigate maps of the magnetic fields radial component at the CMB. In general, due to the projection into the Earth's interior, uncertainties at the CMB are so large that reverse flux in the mean is not resolved reliably

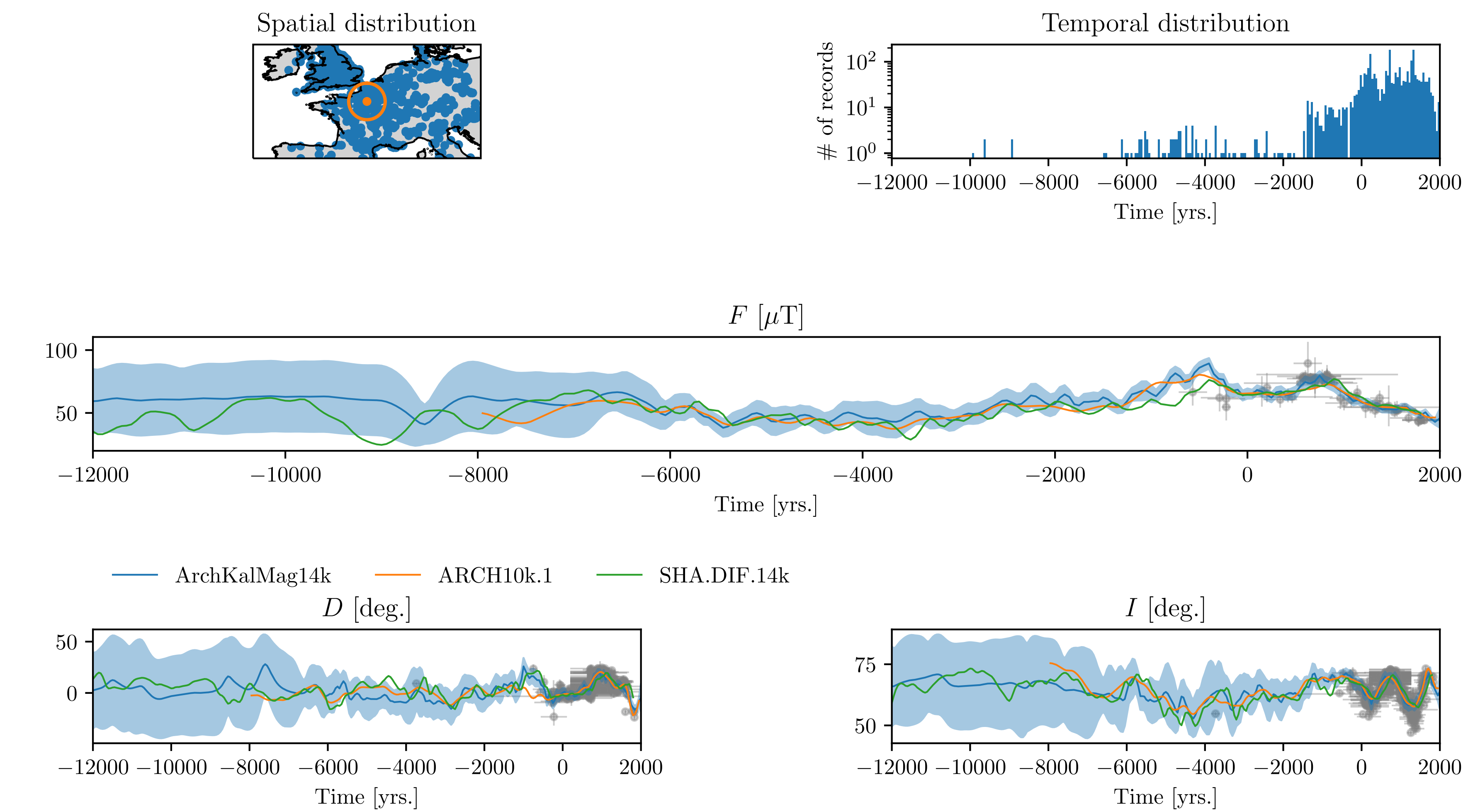

Figure 5. Local predictions of intensity F, declination D and inclination I for Paris. ArchKalMag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. In the top row, the spatial and temporal distribution of the surrounding are shown. Data in the orange ellipse (250km radius) are translated to the location of prediction (orange dot) and shown as gray dots. Horizontal and vertical gray bars indicate the one sigma temporal and field component data uncertainties, respectively. The temporal distribution includes all data visible in the left plot.

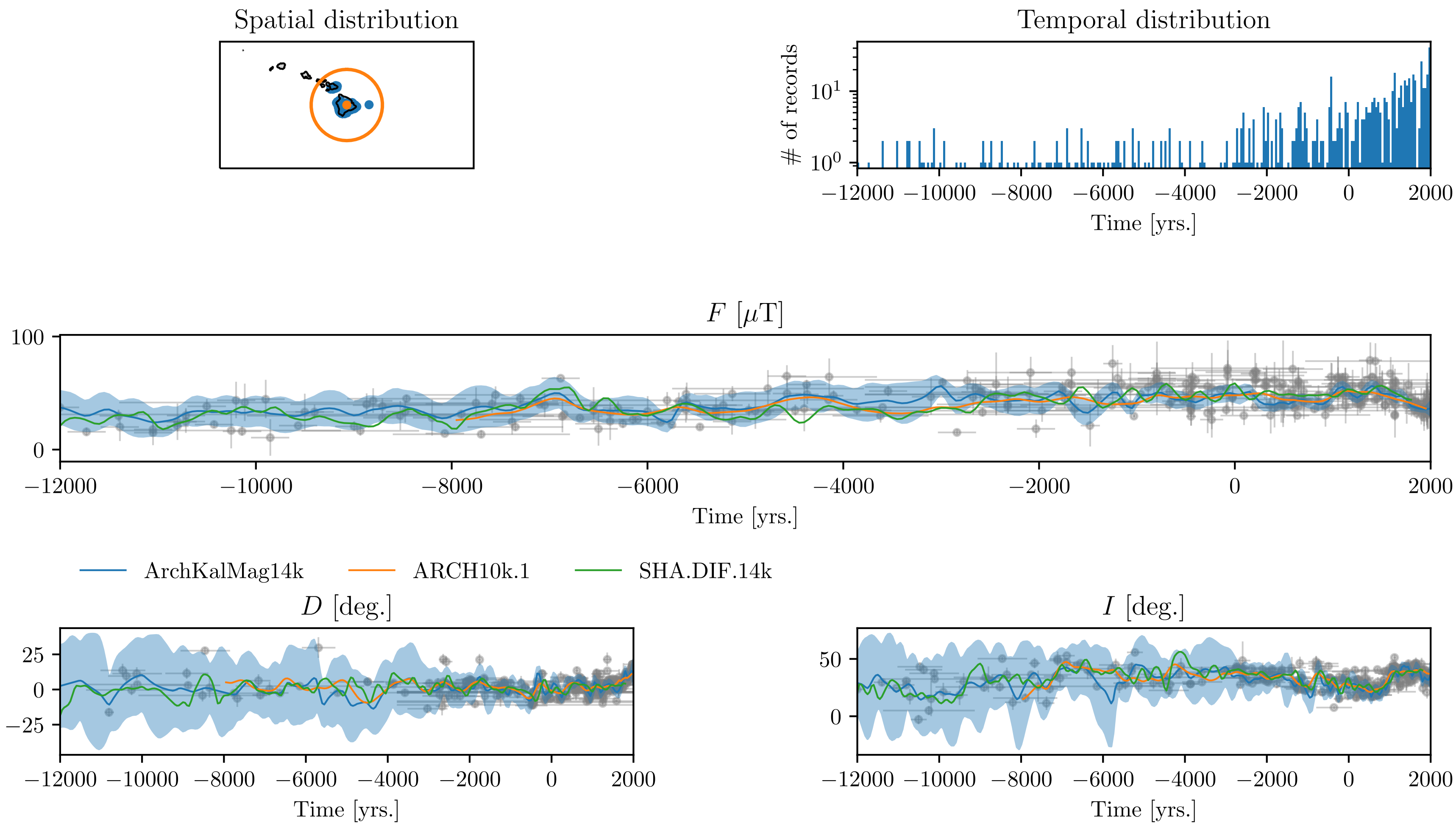

Figure 6. Local predictions of intensity F, declination D and inclination I for Hawaii. ArchKalMag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. In the top row, the spatial and temporal distribution of the surrounding are shown. Data in the orange ellipse (250km radius) are translated to the location of prediction (orange dot) and shown as gray dots. Horizontal and vertical gray bars indicate the one sigma temporal and field component data uncertainties, respectively. The temporal distribution includes all data visible in the left plot.

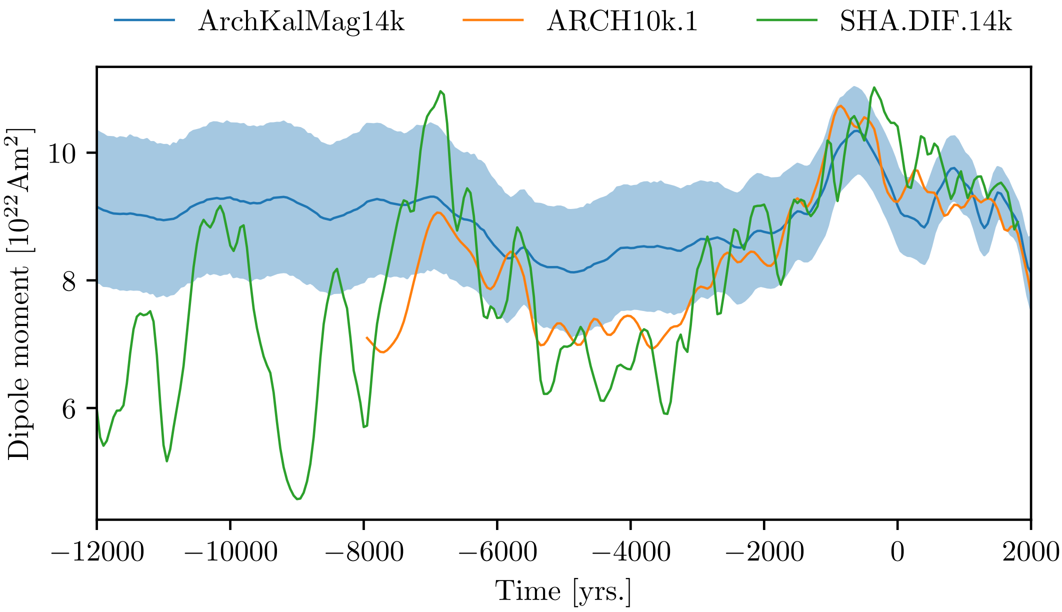

Figure 7. Dipole moment of the geomagnetic field. ArchKalMag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. Mean and standard deviation of ArchKalMag14k are inferred from sampling. This sampling is the reason for the small scale noise in the blue curve and area.

and more data and future work are required to confirm these findings. We consider the projections qualitatively nevertheless.

We find a region of field intensity lower than 32 μT emerging close to the tip of Brazil at 1200 CE. Reverse flux is present to the north and a patch of reverse flux is located directly south of the region. Together with this patch, the region of low intensity rapidly moves south-eastward to the coast of today's Namibia, where it is located in 1300 CE (Fig. 9, b)). This contrasts the findings of Campuzano et al. (2019), where the low intensity region emerges approximately 100 years earlier close to Madagascar. The SAA then extends to the West and slightly to the East, with the center drifting westward until 1500 CE, back to the origin of the region. From there it moves East and constricts at the coast of today's Namibia, almost disappearing at 1650 CE. This dynamic is also not present in SHA.WQ.2k by Campuzano et al. (2019), where the SAA persists at the coast of Namibia and does not decrease in size. The described evolution precedes the dynamics found by Hartmann and Pacca (2009). The subsequent westward drift of the low intensity region generally agrees with their findings and the findings of Campuzano et al. (2019) within the uncertainties.

Further, we find a low field intensity region emerging in 250 BCE west of today's Peru. It drifts south-eastward and in 500 CE merges with a second low field intensity region that emerges around 400 CE North-East of Madagascar. Both anomalies are accompanied by reverse flux in the Southern hemisphere. The joint low intensity region continues to drift eastward and shrinks, persisting until 900 CE. Campuzano et al. (2019) find a low intensity field region emerging at the coast of Namibia at 175 CE. In their findings the earlier anomaly is static and grows until 500 CE. It then shrinks and disappears at 700 CE, earlier than in our findings.

Low intensity regions around the equator are present from the beginning of the model timespan on, but uncertainties are too large to reliably interpret their appearance. First

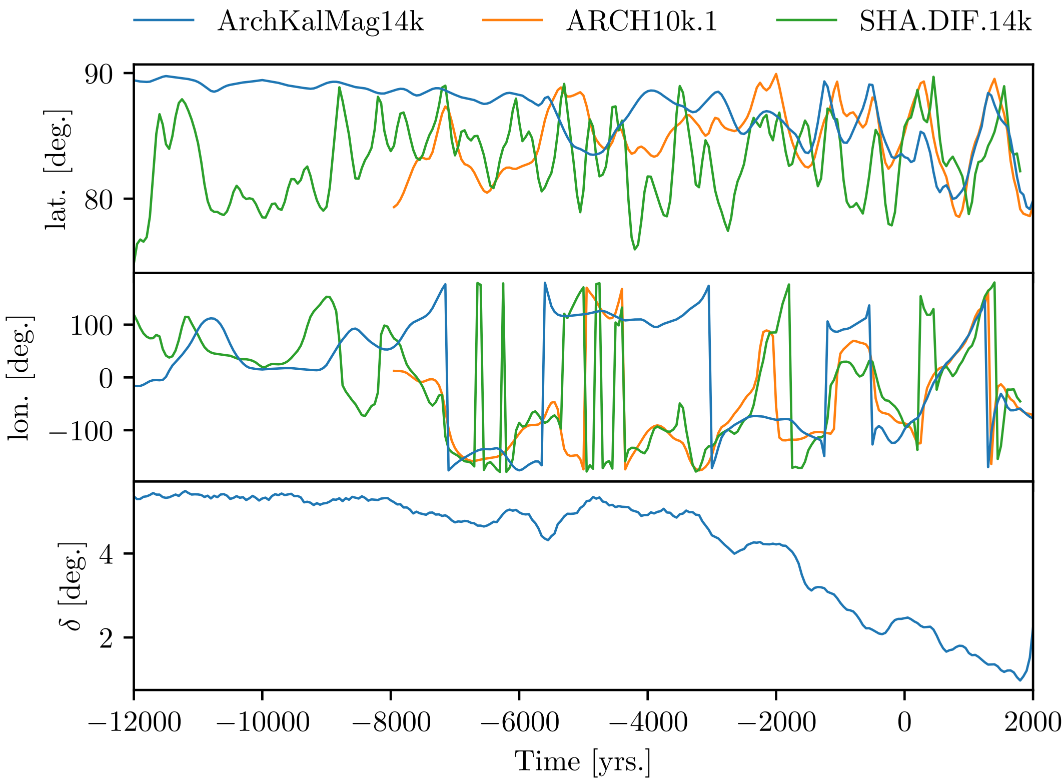

Figure 8. Latitude (top) and longitude (middle) of the geomagnetic dipole axis. ArchKalMag14k is shown in blue. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. The bottom plot shows the angular standard deviation δ (Butler, 2004) for ArchKalMag14k, which is inferred from sampling. This sampling is the reason for the small scale noise in earlier times.

reliable hints on a low intensity field region in the Indian ocean are present around 3000 BCE, with the region drifting eastward and a second low intensity region appearing over the Northern part of South America at 2600 BCE. The anomaly in the Indian ocean disappears at 2200 BCE. The one above South America is accompanied by pronounced reverse flux, although during these epochs uncertainties at the CMB are even higher than during recent times and caution has to be taken when interpreting the results. The anomaly persists over South America, extends until 1500 BCE (Fig. 9, a)) and vanishes in 1200 BCE.

Overall the model shows low field intensity anomalies, accompanied by reverse flux, emerging and vanishing regularly, with a cycle in the order of 1000 years. An animation of the field at the Earth's surface and the CMB can be found with the supplementary material.

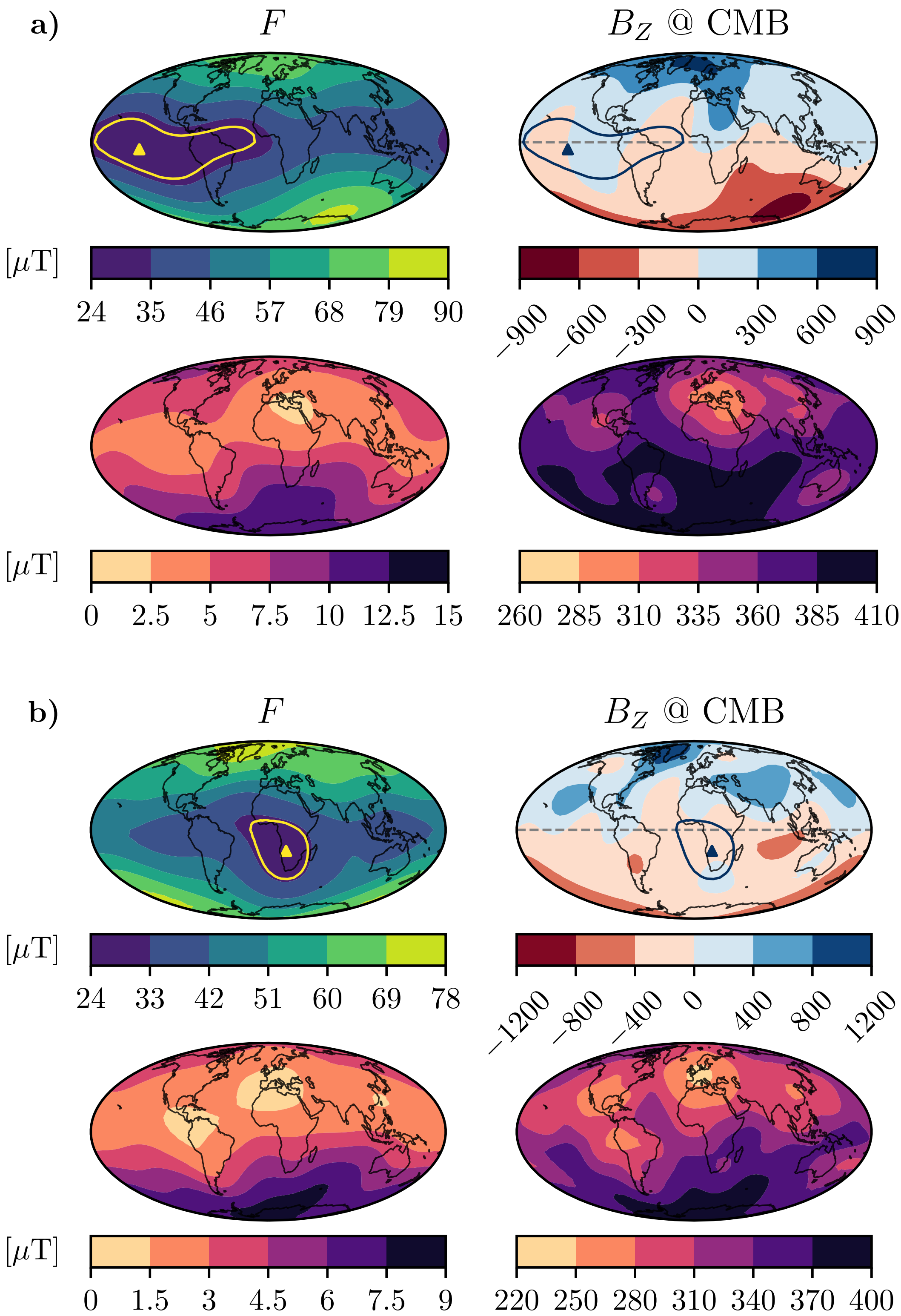

Figure 9. The South Atlantic Anomaly (SAA). The top rows show the field intensity at the Earth's surface and the magnetic field radial component (downwards). The bottom rows show the respective standard deviation. a) is for the year 1500 BCE and b) for 1300 CE. The yellow triangles indicate the location of lowest field intensity. The yellow contour line corresponds to a field value of 32 μT. For reference, both location of lowest intensity and contour are also shown in the CMB plots in blue.

4 Discussion

In the preceding section we proposed the new global geomagnetic field model ArchKalMag14k and presented its features. The local predictions give a reasonable representation of the underlying archeomagnetic data and agree with comparison models within the uncertainties. If no data is present, local curves show significantly less variation than the compared models. Low order, global scale degrees are only resolved if a sufficient amount of data is present. In this case, local predictions for remote locations also show rapid variations and uncertainties are relatively small (see the local predictions for the Indian ocean in the supplementary material). If the data cannot resolve the global scales, the prior is reproduced, which is evident from local curves with no data coverage (Fig. 5) and the analysis of the dipole itself (Figs. 7 and 8). For times earlier than 6000 BCE, the axial dipole varies only slightly around the prior mean value of approx. $-36.19\ \mu$T (Fig. 3, top row). Nevertheless, local variations are resolved, if supported by the data (Fig. 6, especially the dip in declination at 11000 BCE). Spatial power spectra provide insight on the resolution of the model on global scales. From a comparison of the spectra to the respective prior it is evident, that for recent times information up to degree 6 is obtained, while for the earlier times the prior is reproduced already at degree 3 (Fig. 4, top row). An investigation of low intensity field regions reproduces the emergence and evolution of the South Atlantic Anomaly (SAA) in recent times (from 1600 CE on), while the preceding dynamics differ from other studies (Campuzano et al., 2019). Low intensity field regions can be resolved from 3000 BCE on. Although uncertainties at the CMB are large, hints for reverse flux patches associated with these field anomalies are found. A detailed evaluation relating these patches to the anomalies, e.g. based on kernels (Terra-Nova et al., 2017) remains to be done and more data are needed to reduce the uncertainties.

In contrast to other recently proposed Bayesian models (Hellio & Gillet, 2018; Nilsson & Suttie, 2021), most prior parameters of ArchKalMag14k are inferred from the data via maximization of the log marginal likelihood. As the marginal likelihood drops off quickly around the maximum, we did not perform an integration as proposed in the last study (M. Schanner et al., 2021). The a priori assumption of a constant axial dipole may lead to an underestimation of uncertainties in the dipole degrees, moment and location, as the prior mean is constrained well by data from recent times and variations during earlier times are considered around this fixed, constant value. Using only part of the recent records to create a dataset that is more homogeneous in time may improve this, but leads to other complications as hyperparameters become less constrained and harder to determine, when fewer records are available. Artificially increasing the a priori dipole variance leads to more variation around the constant mean during earlier times, but also to higher posterior uncertainties and the model we propose lies well within these. Two scenarios are reasonable, to explain the absence of variations during earlier times in our model. Either the statistical properties (and thus the underlying processes) of the EMF changed during the Holocene, some time around 3000 BCE. This is supported by a visual inspection of the top row in Figure 3 and Figure 7. Or the data do not contain enough information to recover the global dynamics of the field, which is supported by the findings of the validation section. Additional data, e.g. from sediments may help recovering the actual field dynamics, but require significant adaption of the modeling method.

5 Conclusions

This study proposes a new global geomagnetic model for the Holocene, called ArchKalMag14k. We modified the algorithms suggested in earlier works (Mauerberger et al., 2020; M. Schanner et al., 2021) to be applicable to the archeomagnetic database. The inversion is sequentialized by means of a Kalman-filter (Kalman, 1960; Baerenzung et al., 2020). The resulting model consists of sets of Gauss coefficients, secular variations and covariances, stored every 50 years. The model can be reproduced by code that is publicly available (`https://sec23.git-pages.gfz-potsdam.de/korte/paleokalmag/`) or is provided upon request. ArchKalMag14k can be imported by pymagglobal (M. A. Schanner et al., 2020), so that feature analysis is straight-forward.

The central result of this study is that for times earlier than 6000 BCE the current database of thermoremanent records alone does not contain enough information to construct global models. For times earlier than 6000 BCE, ArchKalMag14k reproduces the prior on a global scale and only local variations are resolved. Existing models may further overconfidently report variations during times later than 6000 BCE, as local variations that are resolved by higher degrees in ArchKalMag14k result in variations of the large scale dipole in existing models.

The next step is to extend and adapt the modeling framework to incorporate sediment records. As the recent study by (Nilsson & Suttie, 2021) shows, this requires significant modifications due to aspects of the sedimentation process and the respective statistical implications.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant 388291411. M. Schanner performed theoretical and conceptual work, with support from both co-authors. The manuscript was assembled by M. Schanner, with support from all co-authors. Software development and data processing was conducted by M. Schanner. The work and findings were supervised by M. Korte and M. Holschneider.

Special thanks go to A. N. Mahgoub for sharing his experience and knowledge on records from Mexico, to J. Baerenzung for sharing insights on the Kalman filter algorithm and to S. Mauerberger for valuable discussions on different aspects of this study.

The dataset used in this study is a slight variation of all records from the archaeological and volcanic database from GEOMAGIA v3.4 (Brown et al., 2015) with ages between 12000 BCE and 2000 CE. Some of the records from Mexico contain wrong ages (Mahgoub, pers. comm.), so they have been altered or removed, if no better estimate was available. A list of altered records is available with the supplementary material. All results were produced using a python implementation of the discussed algorithm, which is publicly available at `https://sec23.git-pages.gfz-potsdam.de/korte/paleokalmag/`.

References

Baerenzung, J., Holschneider, M., Wicht, J., Lesur, V., & Sanchez, S. (2020). The kalmag model as a candidate for IGRF-13. *Earth Planets Space*, *72*, 163. Retrieved from `https://earth-planets-space.springeropen.com/articles/10.1186/s40623-020-01295-y` doi: https://doi.org/10.1186/s40623-020-01295-y

Brown, M. C., Donadini, F., Nilsson, A., Panovska, S., Frank, U., Korhonen, K., ... Constable, C. G. (2015). Geomagia50.v3: 2. a new paleomagnetic database for lake and marine sediments. *Earth, Planets and Space*, *67*(1), 70. Retrieved from `https://doi.org/10.1186/s40623-015-0233-z` doi: 10.1186/s40623-015-0233-z

Butler, R. F. (2004). *Paleomagnetism.* Blackwell Scientific Publications. (Electronic edition)

Campuzano, S., Gómez-Paccard, M., Pavón-Carrasco, F., & Osete, M. (2019). Emergence and evolution of the south atlantic anomaly revealed by the new paleomagnetic reconstruction shawq2k. *Earth and Planetary Science Letters*, *512*, 17-26. Retrieved from `https://www.sciencedirect.com/`

science/article/pii/S0012821X19300822 doi: https://doi.org/10.1016/j.epsl.2019.01.050

Constable, C., & Korte, M. (2015). 5.09 - centennial- to millennial-scale geomagnetic field variations. In G. Schubert (Ed.), *Treatise on geophysics (second edition)* (Second Edition ed., p. 309-341). Oxford: Elsevier. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780444538024001032 doi: https://doi.org/10.1016/B978-0-444-53802-4.00103-2

Constable, C., Korte, M., & Panovska, S. (2016). Persistent high paleosecular variation activity in southern hemisphere for at least 10 000 years. *Earth and Planetary Science Letters*, *453*, 78 - 86. doi: 10.1016/j.epsl.2016.08.015

Constable, C. G., Johnson, C. L., & Lund, S. P. (2000). Global geomagnetic field models for the past 3000 years: transient or permanent flux lobes? *Phil. Trans. R. Soc. Lond. A*, *358*, 991-1008.

Constable, C. G., & Parker, R. L. (1988). Statistics of the geomagnetic secular variation for the past 5 m.y. *Journal of Geophysical Research: Solid Earth*, *93*(B10), 11569-11581. doi: 10.1029/JB093iB10p11569

Gillet, N., Jault, D., Finlay, C. C., & Olsen, N. (2013). Stochastic modeling of the Earth's magnetic field: Inversion for covariances over the observatory era. *Geochemistry, Geophysics, Geosystems*, *14*(4), 766-786. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/ggge.20041 doi: 10.1002/ggge.20041

Hartmann, G. A., & Pacca, I. G. (2009). Time evolution of the south atlantic magnetic anomaly. *Anais da Academia Brasileira de Ciências*, *81*, 243 - 255. Retrieved from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0001-37652009000200010&nrm=iso doi: 10.1590/S0001-37652009000200010

Hellio, G., & Gillet, N. (2018). Time-correlation-based regression of the geomagnetic field from archeological and sediment records. *Geophysical Journal International*, *214*(3), 1585-1607. doi: 10.1093/gji/ggy214

Hellio, G., Gillet, N., Bouligand, C., & Jault, D. (2014). Stochastic modelling of regional archaeomagnetic series. *Geophysical Journal International*, *199*, 931-943. doi: 10.1093/gji/ggu303

Holschneider, M., Lesur, V., Mauerberger, S., & Baerenzung, J. (2016). Correlation-based modeling and separation of geomagnetic field components. *Journal of Geophysical Research: Solid Earth*, *121*(5), 3142–3160. doi: 10.1002/2015JB012629

Jackson, A., & Finlay, C. (2015). Geomagnetic secular variation and its applications to the core. In G. Schubert (Ed.), *Treatise on geophysics* (2nd ed., Vol. 5, pp. 137–184). United Kingdom: Elsevier. doi: 10.1016/B978-0-444-53802-4.00099-3

Jackson, A., Jonkers, A., & Walker, M. (2000). Four centuries of geomagnetic secular variation from historical records. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *358*(1768), 957–990. doi: 10.1098/rsta.2000.0569

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, *82*(Series D), 35–45.

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, *10*, 1755-1758.

King, D. E. (2017). *A global optimization algorithm worth using.* http://blog.dlib.net/2017/12/a-global-optimization-algorithm-worth.html. (Accessed: 2020-07-07)

Korte, M., & Constable, C. G. (2003). Continuous global geomagnetic field models for the past 3000 years. *Phys. Earth Planet. Interiors*, *140*, 73-89.

Korte, M., Donadini, F., & Constable, C. (2009). Geomagnetic field for 0-3ka: 2.

a new series of time-varying global models. *Geochem. Geophys. Geosys.*, *10, Q06008*, doi:10.1029/2008GC002297.

Korte, M., Genevey, A., Constable, C. G., Frank, U., & Schnepp, E. (2005). Continuous geomagnetic field models for the past 7 millennia: 1. a new global data compilation. *Geochemistry, Geophysics, Geosystems*, *6*(2). doi: https://doi.org/10.1029/2004GC000800

Licht, A., Hulot, G., Gallet, Y., & Thébault, E. (2013). Ensembles of low degree archeomagnetic field models for the past three millennia. *Physics of the Earth and Planetary Interiors*, *224*, 38 - 67. doi: 10.1016/j.pepi.2013.08.007

Mauerberger, S., Schanner, M., Korte, M., & Holschneider, M. (2020). Correlation based snapshot models of the archeomagnetic field. *Geophysical Journal International*. Retrieved from https://doi.org/10.1093/gji/ggaa336 (ggaa336) doi: 10.1093/gji/ggaa336

McHutchon, A., & Rasmussen, C. E. (2011). Gaussian process training with input noise. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 24* (pp. 1341–1349). Curran Associates, Inc.

Merrill, R. T., McElhinny, M. W., & McFadden, P. L. (1996). *The magnetic field of the earth: Paleo-magnetism, the core, and the deep mantle.* Academic Press, San Diego.

Nilsson, A., & Suttie, N. (2021). Probabilistic approach to geomagnetic field modelling of data with age uncertainties and post-depositional magnetisations. *Physics of the Earth and Planetary Interiors*, *317*, 106737. Retrieved from https://www.sciencedirect.com/science/article/pii/S0031920121000959 doi: https://doi.org/10.1016/j.pepi.2021.106737

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 and Planetary Science Letters*, *388*, 98-109. Retrieved from https://www.sciencedirect.com/science/article/pii/S0012821X13006869 doi: https://doi.org/10.1016/j.epsl.2013.11.046

Piper, J. D. A. (1989). Paleomagnetism. In J. A. Jacobs (Ed.), *Geomagnetism* (Vol. 3, p. 31-61). Academic Press.

Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning.* MIT Press, Cambridge, MA.

Rauch, H. E., Tung, F., & Striebel, C. T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, *3*(8), 1445-1450. doi: 10.2514/3.3166

Ropp, G., Lesur, V., Baerenzung, J., & Holschneider, M. (2020). Sequential modelling of the Earth's core magnetic field. *Earth Planets Space*, *72*, 153. Retrieved from https://earth-planets-space.springeropen.com/articles/10.1186/s40623-020-01230-1 doi: https://doi.org/10.1186/s40623-020-01230-1

Schanner, M., Mauerberger, S., Korte, M., & Holschneider, M. (2021). Correlation based time evolution of the archeomagnetic field. *Journal of Geophysical Research: Solid Earth*, *126*(7), e2020JB021548. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JB021548 doi: https://doi.org/10.1029/2020JB021548

Schanner, M. A., Mauerberger, S., & Korte, M. (2020). *pymagglobal - python interface for global geomagnetic field models.* Potsdam: GFZ Data Services. doi: https://doi.org/10.5880/GFZ.2.3.2020.005

Senftleben, R. (2019). *Earth's magnetic field over the last 1 000 years* (Unpublished doctoral dissertation). University of Potsdam.

Suttie, N., & Nilsson, A. (2019). Archaeomagnetic data: The propagation of an error. *Physics of the Earth and Planetary Interiors*, *289*, 73 - 74. doi: 10.1016/j.pepi.2019.02.008

Terra-Nova, F., Amit, H., Hartmann, G. A., Trindade, R. I., & Pinheiro, K. J.

(2017). Relating the south atlantic anomaly and geomagnetic flux patches. *Physics of the Earth and Planetary Interiors*, *266*, 39-53. Retrieved from https://www.sciencedirect.com/science/article/pii/S0031920116302205 doi: https://doi.org/10.1016/j.pepi.2017.03.002

Supporting Information for "ArchKalMag14k: A Kalman-filter based global geomagnetic model for the Holocene"

M. Schanner[1], M. Korte[1], M. Holschneider[2]

[1]German Research Centre for Geosciences GFZ, Section 2.3, Potsdam, Germany

[2]Applied Mathematics, University of Potsdam, Potsdam, Germany

Contents of this file

Additional Supporting Information (Files uploaded separately)

1. Captions for Movie S1

Introduction

This supplementary material provides validation plots for additional coefficients in Figure S1, a comparison of the model coefficients with the prior in Figure S2 and local field predictions at two additional locations in Figures S3 and S4. Table S1 contains a list of changes made to the GEOMAGIA v.3.4 dataset (Brown et al., 2015). A separately available Movie S1 shows the evolution of the geomagnetic field intensity at the Earth's surface and of the radial component (downwards) at the core mantel boundary, together with respective uncertainties.

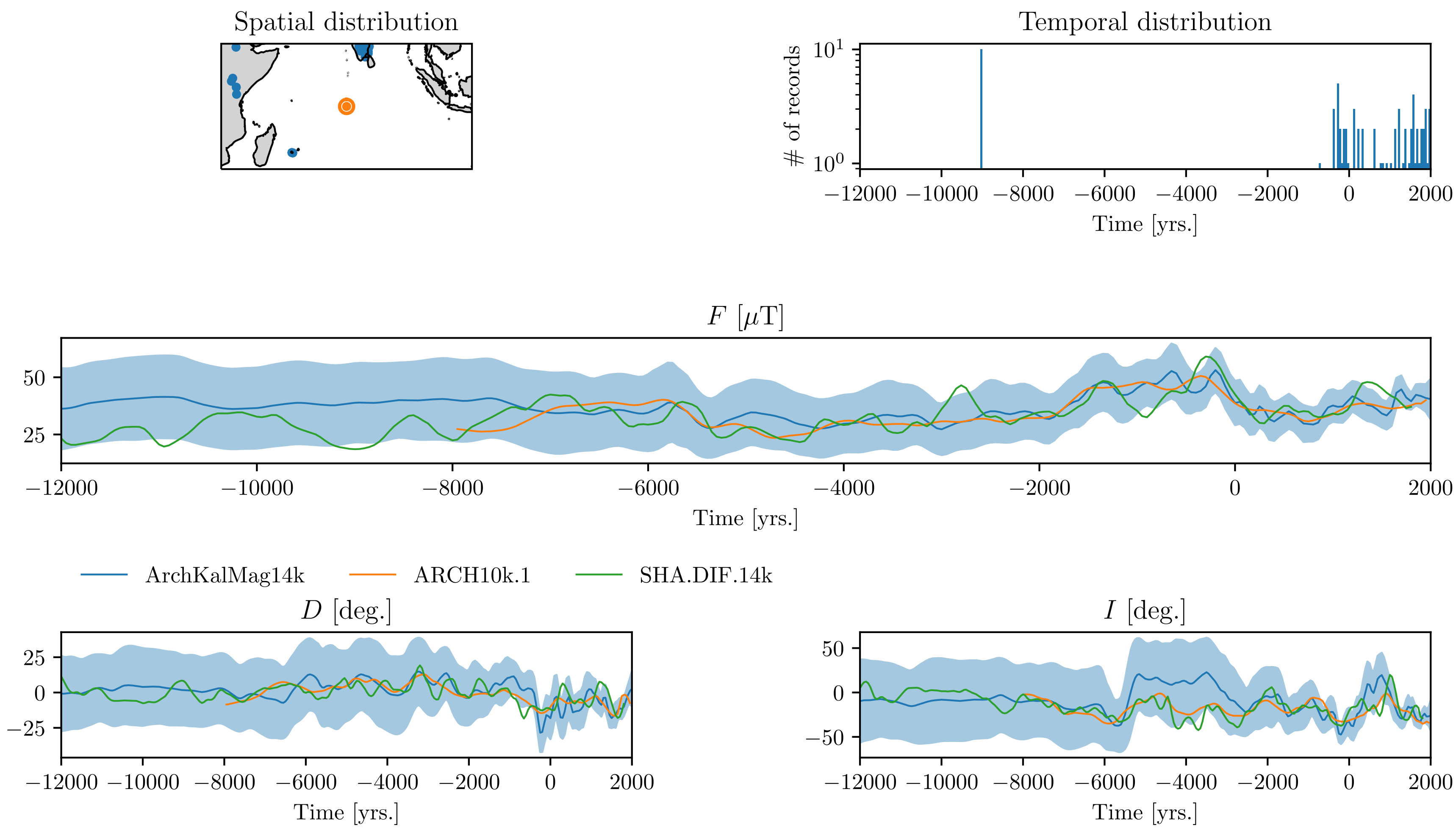

Figure S3. Local predictions of intensity F, declination D and inclination I for the Indian Ocean. ArchKalMag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. In the top row, the spatial and temporal distribution of the surrounding are shown. Data in the orange ellipse (250km radius) are translated to the location of prediction (orange dot) and shown as gray dots. Horizontal and vertical gray bars indicate the one sigma temporal and field component data uncertainties, respectively. The temporal distribution includes all data visible in the left plot.

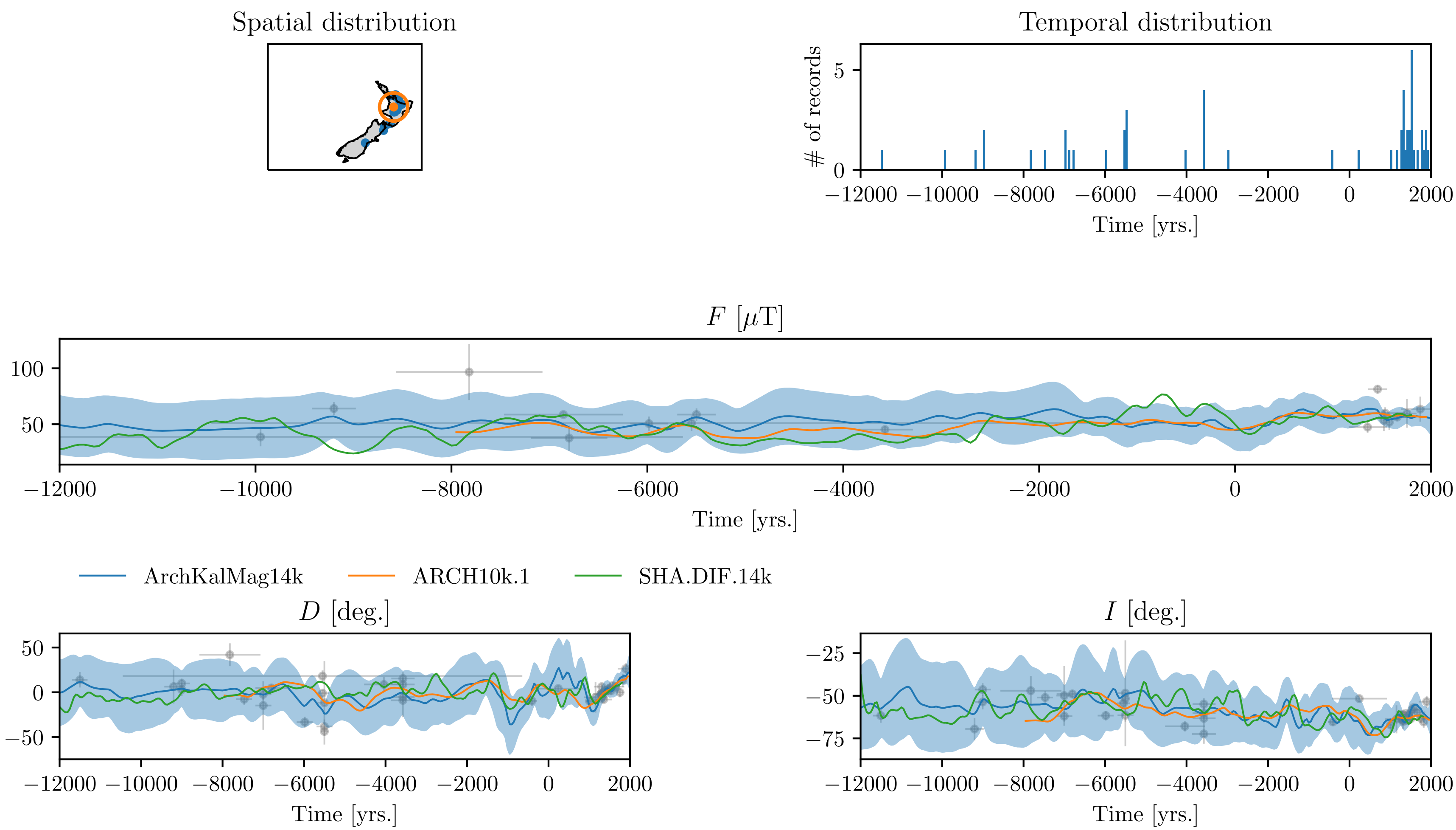

Figure S4. Local predictions of intensity F, declination D and inclination I for New Zealand. ArchKalMag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. In the top row, the spatial and temporal distribution of the surrounding are shown. Data in the orange ellipse (250km radius) are translated to the location of prediction (orange dot) and shown as gray dots. Horizontal and vertical gray bars indicate the one sigma temporal and field component data uncertainties, respectively. The temporal distribution includes all data visible in the left plot.

Movie S1. Evolution of the geomagnetic field intensity at the Earth's surface (left) and of the radial component (downwards, right) at the core mantel boundary, together with respective uncertainties. The time interval of 50 years corresponds to the full resolution of ArchKalMag14k. Note, that the scales change during the movie. The yellow triangle indicates the location of lowest field intensity. The yellow contour line corresponds to a field value of 32 μT. For reference, both location of lowest intensity and contour are also shown in the CMB plots in blue.

References

Brown, M. C., Donadini, F., Nilsson, A., Panovska, S., Frank, U., Korhonen, K., ... Constable, C. G. (2015). Geomagia50.v3: 2. a new paleomagnetic database for lake and marine sediments. *Earth, Planets and Space*, *67*(1), 70. Retrieved from `https://doi.org/10.1186/s40623-015-0233-z` doi: 10.1186/s40623-015-0233-z

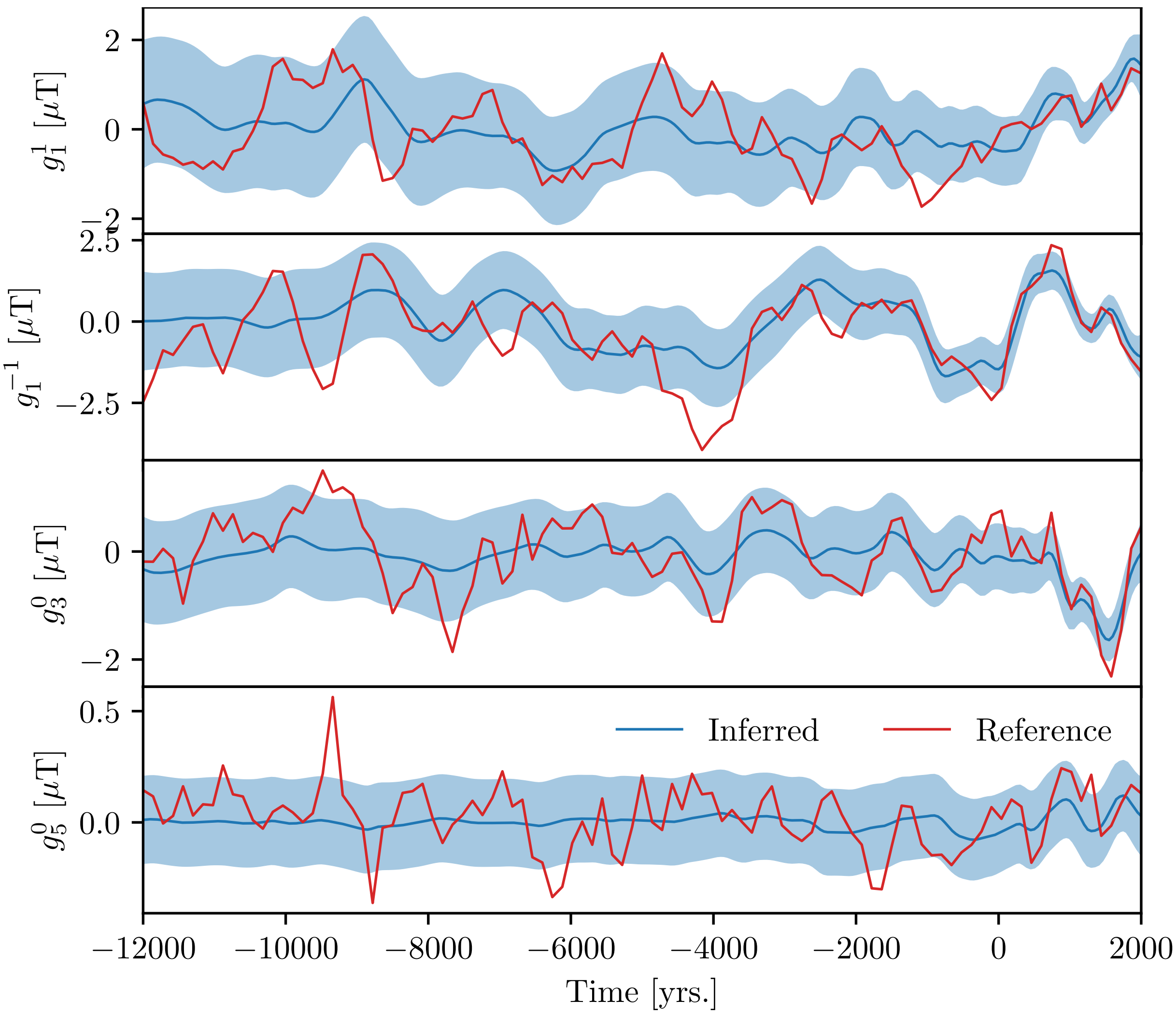

Figure S1. Additional dipole and higher order coefficients of the synthetic model, together with the corresponding inferred ones from the proposed inversion. The inferred (blue) and reference curves (red) agree within the 95%-region shown in light blue.

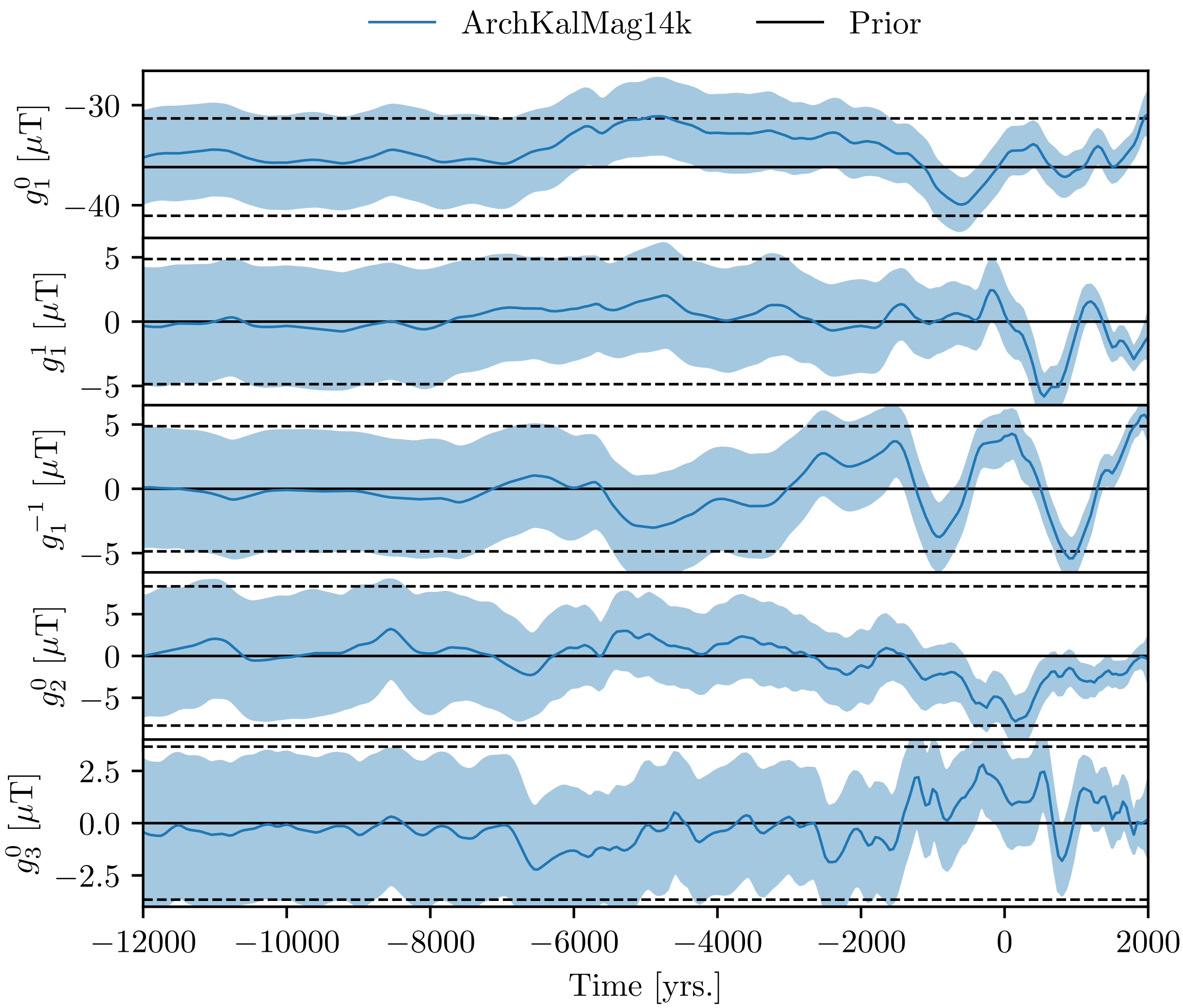

Figure S2. ArchKalMag14k model coefficients together with the prior. The shaded area and dashed lines cover 95%.

Table S1. Updates to the GEOMAGIA dataset (Brown et al., 2015) used to assemble the database for ArchKalMag14k. GEOMAGIA provides a unique ID for every record, that we use to identify the records from Mexico that we changed, as they have wrong age and dating uncertainty estimates (Mahgoub, pers. comm.). Records with IDs 11237, 2773, 6891 and 13149 have been removed from the dataset as no updated information is available.

UID	Updated age [yrs.]	Updated standard deviation [yrs.]
13153	-7550	422
2768	-8523	800
2769	-7450	270
11967	-10000	338
6893	-10000	338
11966	-5707	184
2770	1250	5
6892	1250	5
13086	8	62
13118	8	62
11992	1545	94