Correlation based time evolution of the archeomagnetic field

M. Schanner[1], S. Mauerberger[2], 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:

- Extension of a previous study on spatial correlation based archeomagnetic modeling to the temporal domain
- Dating uncertainties are incorporated by the noisy input Gaussian process formalism

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

Abstract

In a previous study, a new snapshot modeling concept for the archeomagnetic field was introduced (Mauerberger et al., 2020). By assuming a Gaussian process for the geomagnetic potential, a correlation based algorithm was presented, which incorporates a closed form spatial correlation function. This work extends the suggested modeling strategy to the temporal domain. A space-time correlation kernel is constructed from the tensor product of the closed form spatial correlation kernel with a squared exponential kernel in time. Dating uncertainties are incorporated into the modeling concept using a noisy input Gaussian process. All but one modeling hyperparameters are marginalized, to reduce their influence on the outcome and to translate their variability to the posterior variance. The resulting distribution incorporates uncertainties related to dating, measurement and modeling process. Results from application to archeomagnetic data show less variation in the dipole than comparable models, but are in general agreement with previous findings.

1 Introduction

Existing models of the Earth's magnetic field (EMF) for the past millennia show a variety of time-dependent features: The evolution of the South Atlantic Anomaly, the observed dipole decay in recent centuries and the movement of flux patches all take place on timescales of several hundred years (Hartmann & Pacca, 2009; Jackson & Finlay, 2015). To accurately describe and study these features, time resolved models are necessary. Usually these 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. Existing models differ in the approach to global modeling, but are usually constructed using inversion for spherical harmonics (SH) coefficients, truncated at a certain degree. Most models, such as Jackson et al. (2000), Korte et al. (2009) and Senftleben (2019), implement spline interpolation in the temporal domain, while some alternative approaches exist (C. G. Constable & Parker, 1988; Bouligand et al., 2005; Hellio et al., 2014; Hellio & Gillet, 2018).

By regularizing the SH model in both space and time, global features can be extracted from the sparse and clustered database. However, this way also local variations might be wrongly explained by the dynamics of the large scale coefficients. Regions of good data coverage may thus imprint a fast, well supported local dynamic into the dynamic of global coefficients such as the di- and quadrupole. To avoid such variations in the low degrees, and to exploit the data to its fullest, we suggest a Bayesian modeling approach based on Gaussian processes (GPs), both in space and time. With this already in mind, we implemented a closed form covariance function for the spatial domain in a previous study (Mauerberger et al., 2020, hereafter referred to as MSKH20). The present work extends this study to the temporal domain. We again employ the closed form correlation kernel, introduced by Holschneider et al. (2016), and extend it to a space-time kernel using a squared exponential (SQE) kernel. Knowing that such a kernel is unphysical, we abstain from suggesting a new geomagnetic field model. Instead, the aim of this work is to show the potential of the proposed modeling approach and to lay out the modeling strategy and its implications in detail.

The application of a spatio-temporal GP in a Bayesian framework includes the natural availability of well quantified uncertainties via the posterior standard deviation. 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; Hellio & Gillet, 2018; Senftleben, 2019). Within the space-time correlation framework that we suggest,

uncertainties arising from the uneven data distribution, from inaccurate dating and from the modeling itself can be accounted for in a well defined, statistically sound manner (Rasmussen & Williams, 2006; McHutchon & Rasmussen, 2011). The inversion scheme is embedded in a functional analytic frame of non-parametric modeling. The result is a distribution over functions, in this case in both space and time. This distribution is characterized by a mean function, which gives the most likely field model, and a two-point covariance function, describing the variability of the field.

This article is structured as follows: The rest of this section covers some basic introduction into magnetic field theory and GP inversion. We use those paragraphs mainly to introduce our notation. In Section 2 we discuss our prior assumptions, construct the correlation kernel and describe the full modeling algorithm. Section 3 contains a brief validation section, using synthetic data, as well as a case study to showcase the capabilities of our method. We conclude with a discussion in Section 4. The appendix provides further insight into the mathematical footing of the introduced methods.

1.1 Magnetic Field Theory

Outside of the conducting core, the EMF $\boldsymbol{B}$ can be approximated by the gradient of the geomagnetic potential Φ (Backus et al., 1996):

$$\boldsymbol{B} = -\nabla\Phi$$

Φ is a scalar potential, satisfying Laplace's equation $\nabla^2\Phi = 0$. Assuming the sources of the potential lie at some reference sphere with radius R, at locations $\boldsymbol{x}$ outside of this sphere $|\boldsymbol{x}| > R$ the field can be represented using spherical harmonics (SH)

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

$\hat{\boldsymbol{x}}$ 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. Similar to MSKH20, we do not consider the Earth's ellipticity. The dependence of g_ℓ^m on a reference radius R is not explicitly expressed. The time dependence of the field is typically encoded in the Gauss coefficients $g_\ell^m(t)$. We use upright letters $\mathbf{x} = (\boldsymbol{x}, t)$ to distinguish space-time inputs from purely spatial inputs. Often the Gauss coefficients are expressed in form of a spline model (Bloxham & Jackson, 1992)

$$g_\ell^m(t) = \sum_n g_{\ell,n}^m M_n(t) \ , \tag{2}$$

where $M_n(t)$ are typically cubic B-spline basis functions and the model is defined by the set of Gauss coefficients $\{g_{\ell,n}^m\}$ at knot times t_n.

The field vector components are given in a spherical coordinate system, consisting of north B_N, east B_E and down B_Z components

$$B_N = -\frac{1}{r}\frac{\partial\Phi}{\partial\theta} \ , \qquad B_E = \frac{1}{r\sin(\theta)}\frac{\partial\Phi}{\partial\phi} \ , \qquad B_Z = -\frac{\partial\Phi}{\partial r} \ , \tag{3}$$

where θ, ϕ and r are colatitude, longitude and radius of a field location $\boldsymbol{x}$.

Paleomagnetic records of the EMF are provided as declination D, inclination I and intensity F, which relate to the field vector in a non-linear fashion:

$$\mathbf{H} : \boldsymbol{B} \to \begin{pmatrix} D \\ I \\ F \end{pmatrix} = \begin{pmatrix} \arctan\left(\frac{B_E}{B_N}\right) \\ \arctan\left(\frac{B_Z}{F_H}\right) \\ \sqrt{B_N^2 + B_E^2 + B_Z^2} \end{pmatrix} \tag{4}$$

The horizontal intensity $F_H = \sqrt{B_N^2 + B_E^2}$ is an auxiliary quantity. $\mathbf{H}$ is called *observation functional*.

1.2 Gaussian Process regression

In the eighties C. G. Constable and Parker (1988) already proposed using GPs to model the EMF. A GP is a stochastic process, characterized by a mean function $\bar{\boldsymbol{B}}$ and a covariance function $K_{\boldsymbol{B}}$

$$\boldsymbol{B} \sim \mathcal{GP}\big(\bar{\boldsymbol{B}},\ K_{\boldsymbol{B}}\big)\ . \tag{5}$$

Given observations $o(\mathbf{y})$ of $\boldsymbol{B}$ at locations and times $\mathbf{y} = (\boldsymbol{y}, s)$ with Gaussian measurement errors, characterized by a covariance matrix Σ_o, the posterior of $\boldsymbol{B}$ is again a GP. Its (conditional) mean and covariance functions read (Rasmussen & Williams, 2006)

$$\mathbb{E}[\boldsymbol{B}(\mathbf{x})|O] = \bar{\boldsymbol{B}}(\mathbf{x}) + K_{\boldsymbol{B}}(\mathbf{x},\mathbf{y})\big(K_{\boldsymbol{B}}(\mathbf{y},\mathbf{y}) + \Sigma_o\big)^{-1}\big(o(\mathbf{y}) - \bar{\boldsymbol{B}}(\mathbf{y})\big) \tag{6}$$

$$\mathrm{Cov}[\boldsymbol{B}(\mathbf{x}), \boldsymbol{B}(\mathbf{x}')|O] = K_{\boldsymbol{B}}(\mathbf{x},\mathbf{x}') - K_{\boldsymbol{B}}(\mathbf{x},\mathbf{y})\big(K_{\boldsymbol{B}}(\mathbf{y},\mathbf{y}) + \Sigma_o\big)^{-1}K_{\boldsymbol{B}}(\mathbf{y},\mathbf{x}')\ . \tag{7}$$

$\mathbf{x} = (\boldsymbol{x}, t)$ refers to location and time of interest. Note, that herein already a difference to previous GP based models is visible: The covariance function is defined both in space and time. While in principle the truncated spherical harmonics and the B-spline basis may also be used to construct a covariance function, the language and formalism of GP regression have so far only been applied to either temporal correlations (Gillet et al., 2013; Hellio & Gillet, 2018) or spatial correlations (S. Sanchez et al., 2016; Mauerberger et al., 2020).

2 Modeling concept

We propose a fully Bayesian modeling concept, embedded in a functional analytic setting. Therefore, GP based techniques are employed. One key ingredient to GP regression is the a priori covariance function, also called the (correlation) kernel. In this section we formulate the covariance function we employ, based on our a priori assumptions. Additionally, we formulate the paleomagnetic data model and discuss approximations that are necessary to apply the GP regression scheme.

2.1 A priori process

Translating the uninformative dipole prior from MSKH20 to a time-dynamic realm presents a challenge, as temporal correlations are send to zero together with the a priori precision and cannot easily be recovered in the posterior. Instead, the a priori mean function of the EMF is assumed to be constant in time, with only axial dipole contribution:

$$\bar{\boldsymbol{B}}(\mathbf{x}) = -\nabla\Big(\gamma_1^0 \cdot Y_1^0(\hat{\boldsymbol{x}}) \cdot \frac{R^3}{|\boldsymbol{x}|^2}\Big) \tag{8}$$

The strength γ_1^0 of the a priori dipole is a free parameter, that will later be marginalized.

We suggest building the space-time covariance using a sum of tensor products. The proposed closed form covariance function for the spatial correlations of the non-dipole part $K_{\Phi,\mathrm{S,ND}}$ includes all SH degrees. Holschneider et al. (2016) describe how to construct this kernel, and in MSKH20 we describe in detail how to adapt it for paleomagnetic applications. In Appendix A we wrap up this procedure. The strategy is to translate an idea about correlations amongst Gauss coefficients to the potential, using the SH representation. The field covariance function then follows from the gradient. Our main a priori assumption for the covariance is that at some reference sphere, close to where the core field is generated, the geomagnetic Gauss coefficients are uncorrelated. Assuming a flat spectrum at this reference sphere's radius R, it is possible to derive a closed form for the potential covariance function. This closed form is called Legendre kernel and reads (Holschneider et al., 2016, Eq. 53)

$$K_{\mathrm{L}}(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{\sqrt{1 - 2b + a^2}}\ , \tag{9}$$

where $b = \boldsymbol{x} \cdot \boldsymbol{x}'/R^2$ and $a = |\boldsymbol{x}||\boldsymbol{x}'|/R^2$.

Temporal correlations are incorporated by a tensor product of this spatial kernel with a squared exponential (SQE) kernel:

$$K_{\Phi,\mathrm{T}}(t,t') = K_{\mathrm{SQE}}(t,t') = \exp\left(-\frac{(t-t')^2}{\tau^2}\right) \tag{10}$$

τ is the correlation time. Similar to MSKH20, we split the kernel into dipole and non-dipole part, as the statistical properties of the dipole are known to differ from the higher field degrees (C. G. Constable & Parker, 1988). Each part is coupled to its own temporal correlation kernel and thus has its own correlation time:

$$\begin{aligned} K_\Phi(\mathbf{x},\mathbf{x}') =& \alpha_{\mathrm{DP}}^2 K_{\mathrm{DP,S}}(\boldsymbol{x},\boldsymbol{x}') \cdot \exp\left(-\frac{(t-t')^2}{\tau_{\mathrm{DP}}^2}\right) \\ &+ \alpha_{\mathrm{ND}}^2 K_{\mathrm{ND,S}}(\boldsymbol{x},\boldsymbol{x}') \cdot \exp\left(-\frac{(t-t')^2}{\tau_{\mathrm{ND}}^2}\right) \end{aligned} \tag{11}$$

The index S stands for spatial and $\alpha_\bullet$ are the a priori variances of the dipole and non-dipole part. See Appendix A for the explicit forms and further details. The kernel implements a single, constant correlation time τ_{ND} for all degrees $\ell \geq 2$. We are aware, that previous work indicates different behavior (Bouligand et al., 2016). However, implementing the SQE kernel as suggested is straightforward and sufficient for the conceptual work we present here. The field's covariance function reads

$$K_{\boldsymbol{B}}(\mathbf{x},\mathbf{x}') = \nabla_{\boldsymbol{x}} \nabla_{\boldsymbol{x}'} K_\Phi(\mathbf{x},\mathbf{x}') \ . \tag{12}$$

2.2 Linearization

Paleomagnetic observations are reported as declination, inclination and intensity. With measurement errors $\boldsymbol{E}$, the data model reads

$$o(\mathbf{x}) = \mathbf{H}\Big(\boldsymbol{B}(\boldsymbol{x},t)\Big) + \boldsymbol{E} \ . \tag{13}$$

Clearly, the relationship to the field vector is non-linear (Eq. 4). Handling non-linear transformations in the framework of GP regression is technically demanding and often analytically impossible, as the transformed random variables are no longer Gaussian. While more sophisticated methods exist (e.g. Snelson et al., 2003), the standard approach is to linearize the observation functional by means of a Taylor approximation of 1st order. For declination, inclination and intensity, the approximate, linear functionals read

$$D \approx \tilde{D} + \frac{1}{\tilde{F}_H^2} \begin{bmatrix} -\tilde{B}_E \\ \tilde{B}_N \\ 0 \end{bmatrix}^\top \boldsymbol{B} \ , \tag{14}$$

$$I \approx \tilde{I} + \frac{1}{\tilde{F}_H} \left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{\tilde{B}_Z}{\tilde{F}} \frac{\tilde{\boldsymbol{B}}}{\tilde{F}} \right)^\top \boldsymbol{B} \ , \tag{15}$$

$$F \approx \frac{\tilde{\boldsymbol{B}}^\top}{\tilde{F}} \boldsymbol{B} \ . \tag{16}$$

Here, $\tilde{D}, \tilde{I}, \tilde{F}$ and $\tilde{\boldsymbol{B}}$ indicate the point of expansion (POE). We implement this approximate transformation, to have a linear relation between the observations and the modeled quantity. Linear transformations preserve normality and thus the standard GP formalism is applicable. The proxy data model reads

$$o(\mathbf{x}) \approx \mathbf{H}_{\mathrm{lin.}} \boldsymbol{B}(\boldsymbol{x},t) + \boldsymbol{E} \ . \tag{17}$$

$\mathbf{H}_{\mathrm{lin.}}$ refers to the linearized observation functionals Eqs. 14-16.

Previous works implemented an axial dipole as POE (e.g. Hellio & Gillet, 2018). MSKH20 shows that the performance of inversion for the POE can be enhanced if we separate the data into two disjoint groups. One group consists of records with full vector information (complete) and the other of records with at least one component missing (incomplete). In a first step, only complete records are considered and the posterior distribution for these records is calculated. This first step posterior then serves as the prior and POE for the second step, where the remaining, incomplete records are treated.

2.3 Measurement errors

In order to apply the GP regression formalism (Eq. 5), the full data model has to be Gaussian. Therefore, linearizing the observation functional as described in the previous section is not sufficient, but a normal proxy error model has to be constructed as well. Intensity records often provide the error as standard deviations of a normal distribution, and thus linearizing the observation functional is sufficient for the intensities. Records of the archeomagnetic directions (declination and inclination) on the other hand are reported together with the 95% confidence cone (α_{95}) of a von Mises-Fisher distribution. Thus for the archeomagnetic directions, we construct a Gaussian proxy, using (Suttie & Nilsson, 2019)

$$\sigma_I = \frac{57.3^\circ}{140}\alpha_{95} \qquad \text{and} \qquad \sigma_D = \frac{1}{\cos o_I}\sigma_I \ . \tag{18}$$

Additionally, similar to MSKH20, we implement a scaling factor ϵ to compensate possible false error estimates, and a residual term $\boldsymbol{P}$ with scaling factor ρ, to address modeling related errors (e.g. observational bias due to crustal field contributions). This way, the data model reads

$$o(\mathbf{x}) \approx \mathbf{H}_{\text{lin.}}\Big(\boldsymbol{B}(\boldsymbol{x},t) + \rho\boldsymbol{P}\Big) + \epsilon\boldsymbol{E}_{\text{prox.}} \ , \tag{19}$$

where $\boldsymbol{E}_{\text{prox.}}$ are the approximate errors, constructed from Eq. 18.

2.4 Dating uncertainties

To a large amount, archeomagnetic specimen are dated using either radiocarbon dating or archeologic age estimation. Both methods suffer from uncertainties, as the former depends on carbon models of the atmosphere and the latter on contextual knowledge. To incorporate these uncertainties, and to represent them in the resulting models, previous studies mostly relied on sampling strategies (e.g. Korte et al., 2009; Hellio & Gillet, 2018; Senftleben, 2019). Hellio et al. (2014) used a normal error model for the dates, and applied Markov Chain Monte-Carlo (MCMC) methods to estimate the posterior distribution. We pursue a similar, hierarchical approach, but instead of MCMC methods, we perform analytic approximations. Summarizing the errors from the previous section as $\boldsymbol{\varepsilon}$ for readability, the data model is

$$o(\mathbf{x}) \approx \mathbf{H}_{\text{lin.}}\boldsymbol{B}(\boldsymbol{x},t) + \boldsymbol{\varepsilon} \ . \tag{20}$$

However, one does not know the precise time t at which the specimen received its magnetization, but a corrupted date

$$t_o = t + e_t \ , \qquad \text{where} \qquad e_t \sim \mathcal{N}(0,\sigma_t^2) \tag{21}$$

is a normal error. Plugging this into the data model gives

$$o(\mathbf{x}) = \mathbf{H}_{\text{lin.}}\boldsymbol{B}(\boldsymbol{x},t_o - e_t) + \boldsymbol{\varepsilon} \ . \tag{22}$$

This is known as the noisy input Gaussian process (NIGP) (McHutchon & Rasmussen, 2011). Due to the random variable e_t appearing at the inputs of the GP $\boldsymbol{B}$, this data

model is non-Gaussian again. To tackle it, McHutchon and Rasmussen (2011) suggest once more a linearization. This gives

$$o(\mathbf{x}) \approx \mathbf{H}_{\text{lin.}}\Big(\boldsymbol{B}(\boldsymbol{x}, t_o) - e_t^\top \partial_t \boldsymbol{B}(\boldsymbol{x}, t)|_{t_o}\Big) + \boldsymbol{\varepsilon} \ . \tag{23}$$

The first term is normal and the second term allows for easy construction of a moment matching proxy. With this modifications, Eq. 23 can be used for GP regression in the usual way. Since the error e_t is centered, the a priori mean is not affected by the dating uncertainties. However, the covariance gets an additional term

$$\Sigma_{tt'} \circ \partial_t \partial_{t'} K_{\boldsymbol{B}}(\mathbf{x}, \mathbf{x}')|_{t_o} \ . \tag{24}$$

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.

In Figure 1 we present a comparison of the proposed NIGP strategy to the standard GP inversion and inference via MCMC. Data was generated from a one dimensional SQE kernel and assigned large input uncertainties and small errors, to mimic the situation of large dating uncertainties. The standard GP regression shows the typical constrictions at the input points, while the NIGP shows a larger standard deviation, especially at the input points. We believe that an MCMC approach gives a better estimate of the actual posterior, though in a realistic setting this is computationally unfeasible. However, as can be seen from the bottom panel of Figure 1, the NIGP gives a reasonable proxy to the MCMC result at immensely reduced computational cost.

2.5 Hyperparameters

The model we constructed throughout this section consists of several parameters, which are a priori unknown. Most obvious are the a priori dipole strength γ_1^0, the variances α_{DP} and α_{ND} and the correlation times τ_{DP} and τ_{ND}. Additionally, there are the two scalings, ϵ and ρ, for the measurement errors and the residual, respectively. The least obvious is the kernel's reference radius R. R basically controls the slope of the prior power spectrum. As suggested in our MSKH20 (Fig. 2), we take $R = 2800$km, which gives a slope similar to the IGRF power spectra (Thébault et al., 2015). The other parameters are marginalized, so that the outcome of the modeling procedure is a compound distribution

$$p(\boldsymbol{B}|o) = \int p(\boldsymbol{B}|o, \eta) \cdot p(\eta|o) \ \mathrm{d}\eta. \tag{25}$$

Here η is the set of hyperparameters $\{\gamma_1^0, \alpha_{\text{DP}}, \alpha_{\text{ND}}, \tau_{\text{DP}}, \tau_{\text{ND}}, \epsilon, \rho\}$ and $p(\eta|o)$ is the marginal posterior. See Appendix B for further details.

The compound distribution is no longer Gaussian and includes modeling uncertainties, resulting from the a priori lack of knowledge about the hyperparameters, but does not depend on these parameters. This distribution is the central result of the suggested modeling strategy. It includes a most probable field model, the mean of the distribution, as well as uncertainty estimates, resulting both from modeling and measurement process. Similar distributions can be obtained for other quantities of interest, such as the Gauss coefficients. To actually evaluate these expressions, numerical approximations have to be employed, as described in the following section. The integral is approximated by a sum, which results in a Gaussian mixture distribution. Moments for this mixture are easily obtained.

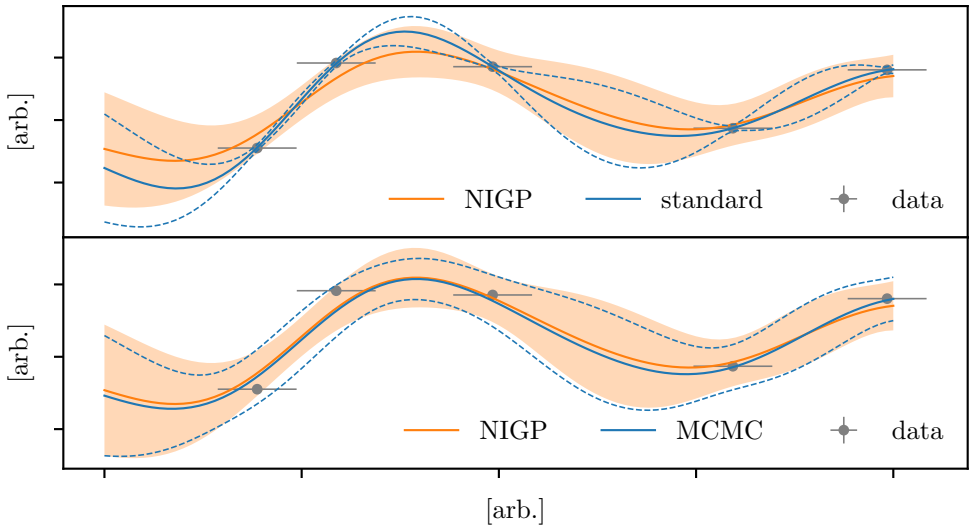

Figure 1. Comparison of the noisy input Gaussian process (NIGP) approach to standard GP inversion (top) and inference using MCMC (bottom) for artificial data in arbitrary units. The NIGP approach translates uncertainties in the inputs to uncertainties in the posterior, in stark contrast to the standard GP regression, which shows the typical constrictions at the inputs. From the bottom panel one can see, that the NIGP gives a reasonable proxy to the MCMC posterior, which we believe to be a good estimate for the actual posterior. For this small dataset, the runtime for the MCMC was 43s, in distinction to 0.3ms for the NIGP. This factor of roughly 10^4 makes using MCMC infeasible for the later steps of our proposed modeling procedure.

3 Application

In this section we demonstrate the potential of the suggested modeling scheme, by first applying it to synthetic test data and finally conducting a case study based on actual archeomagnetic records. The major task to this end is the implementation of the covariance matrices from the kernel, the linearization and the two step strategy. As this process is described in detail in MSKH20, we outsource it to Appendix B. However, two points are to be discussed here. One is the explicit second derivative of the temporal kernel, appearing in Eq. 24. In this study the correlation kernels for dipole and non-dipole contributions are each considered tensor product kernels such that the temporal and spatial parts can be separated. Thus the time derivative only affects the SQE-part. For this simple kernel, the derivative is straightforward to calculate and reads

$$\partial_t \partial_{t'} K_{\mathrm{SQE}}(t,t') = \left(\frac{2}{\tau^2} - \frac{4}{\tau^4}(t-t')^2 \right) \cdot K_{\mathrm{SQE}}(t,t') \ . \tag{26}$$

When using a more realistic kernel, especially one with different correlation times for different degrees ℓ, calculating this derivatives analytically may pose a challenge, so that numerical methods have to be employed. This is one reason why in this conceptual study we chose the SQE kernel over a more realistic one e.g. the one proposed by Gillet et al. (2013).

The second point is the marginalization integral in Eq. 25. As the proposed model contains seven parameters and as the data is incorporated all at once, instead of in bins as in MSKH20, the brute-force parameter space exploration and integration suggested in MSKH20 are now computationally unfeasible. This is due to the cost for fixed point

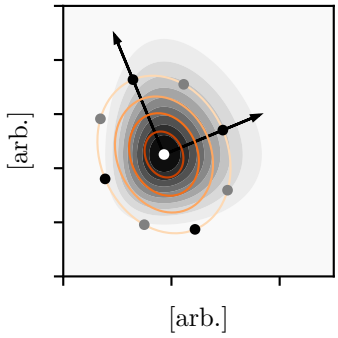

Figure 2. Illustration of the CCD integration in two dimensions. Integration over the actual distribution (grey in the background) is performed by calculating the MAP (white) and constructing a proxy multivariate Gaussian (orange), using the inverse Hessian at the MAP. The dots indicate the integration points. The black dots are called *star points* and are placed along the transformed main axis (the arrows). The grey points are added to the CCD to better capture the covariance structure. In higher dimensions they are not symmetric. See also Rue et al. (2009, Sec. 6.5) and S. M. Sanchez and Sanchez (2005).

integration growing exponentially with the number of dimensions and the high cost of matrix inversion ($\mathcal{O}(n_{\text{Data}}^3)$).

Instead, we perform numerical integration similar to the strategy suggested by Rue et al. (2009, Sec. 6.5). The idea is to center the integration around the maximum a posteriori probability estimator (MAP). Collocation points are added according to central composite design (CCD) (S. M. Sanchez & Sanchez, 2005), in order to capture the bulk of the uncertainties in the hyperparameters (see Figure 2). In seven dimensions, the integration is approximated by a sum over 79 collocation points. To find the MAP, we use the LIPO-TR global optimization algorithm (King, 2009, 2017). The parameters are assigned box priors, as is required by most global optimization algorithms. We choose as upper and lower bonds for the hyperparameters

$$\begin{aligned} -10000\ \mu\text{T} \leq \gamma_1^0 \leq -1\ \mu\text{T} \qquad & 1\ \mu\text{T} \leq \alpha_\bullet \leq 10000\ \mu\text{T} \\ 10\ \% \leq \epsilon \leq 350\ \% \qquad & 10\ \text{yrs.} \leq \tau_\bullet \leq 10000\ \text{yrs.} \\ 0.2\ \mu\text{T} \leq \rho \leq 10\ \mu\text{T} & \end{aligned}$$

Within the box, all parameters except for ϵ are additionally assigned Jeffrey's priors, for scale independence inside of said box.

$$p(\gamma_1^0) \propto \frac{1}{\gamma_1^0}\ , \qquad p(\alpha_\bullet) \propto \frac{1}{\alpha_\bullet}\ , \qquad p(\tau_\bullet) \propto \frac{1}{\tau_\bullet} \qquad \text{and} \qquad p(\rho) \propto \frac{1}{\rho}, \tag{27}$$

where $\bullet$ stands for DP and ND. The poles induced by Jeffery's priors do not cause trouble, as the box constraints are far enough from zero. By numerically approximating the integral in Eq. 25, the compound distribution is approximated by a Gaussian mixture. Details can again be found in Appendix B.

3.1 Synthetic Tests

To validate the proposed algorithm, we test it on synthetic data. As inputs we choose dates and locations from the archeomagnetic and volcanic data in GEOMAGIA v3.3 (Brown

et al., 2015) for the interval from 800 to 2000. At these locations we generate data from the ARCH10k.1 model (C. Constable et al., 2016) and corrupt it by artificial noise. For the directions we use a von Mises-Fisher distribution. Intensities are corrupted by gamma distributed noise and the dates by normal noise. The error levels are taken from GEOMAGIA as well. The resulting MAP for the hyperparameters is

$$\hat{\gamma}_1^0 = -33.6673\ \mu\text{T} \qquad \hat{\alpha}_{\text{DP}} = 1.63939\ \mu\text{T} \qquad \hat{\tau}_{\text{DP}} = 318.454\ \text{yrs.}$$
$$\hat{\epsilon} = 101.656\ \% \qquad \hat{\alpha}_{\text{ND}} = 40.4634\ \mu\text{T} \qquad \hat{\tau}_{\text{ND}} = 339.471\ \text{yrs.}$$
$$\hat{\rho} = 2.59371\ \mu\text{T}$$

Note that α_{ND} is given w.r.t. the reference radius, i.e. the variance at the Earth's surface is much smaller. The error level scalings are as expected, as the reported errors have been used to corrupt the data ($\epsilon \approx 100\%$) and a random contribution of 2.5 μT has been added. The correlation times appear rather large. Nonetheless, we believe that the results are meaningful, as can be understood from looking at a one-dimensional example. Assume for simplicity, that there are direct observations of the axial dipole g_1^0. In reality these may be reconstructed from records, but in this example we consider synthetic ones generated from some reference model. In a situation with many observations with

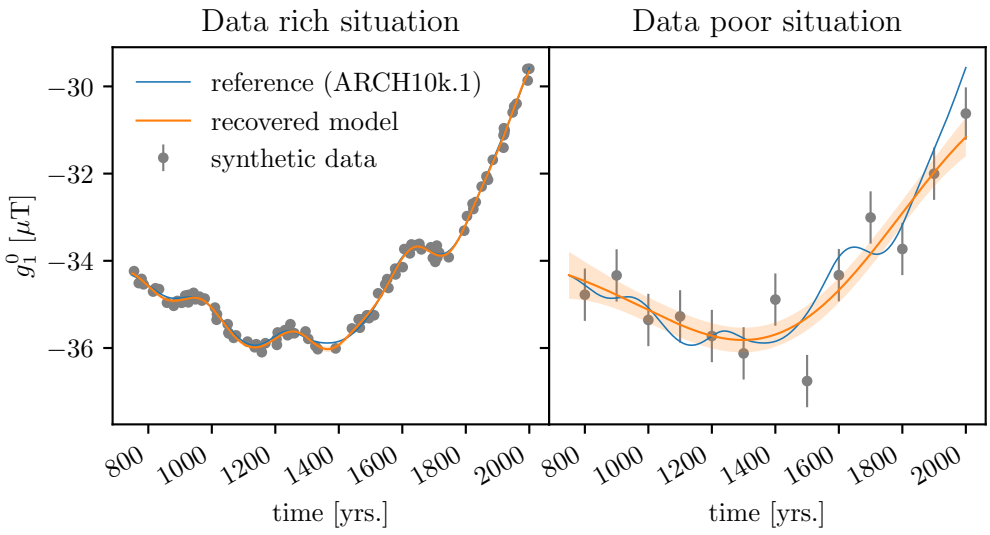

Figure 3. Synthetic one-dimensional example illustrating how the lack of data influences the information one can recover from the reference model. In the data rich situation on the left, the reference model can be recovered quite well, while in the data poor situation on the right only the long-term behavior can be recovered.

small uncertainties, the proposed modeling algorithm gives a correlation time corresponding to the short term variability of the reference model and the reference is recovered well and detailed (Figure 3, left panel). In a situation with few observations with large errors however, one can only recover the long-term variability of the reference model (Figure 3, right panel). The "data poor" situation was designed in a way that is close to the actual situation: Every one hundred years a dipole with errors similar to the one we observed in MSKH20 was generated. In the actual modeling procedure, there is no "first step" of predicting on the axial dipole, but we believe that the large correlation times come from a similar mechanism. To emphasize the effect, in this toy example neither cross-correlations nor dating uncertainties have been considered. We performed a similar ex-

periment for synthetic field data and found similar results, i.e. only the long-term information can be recovered.

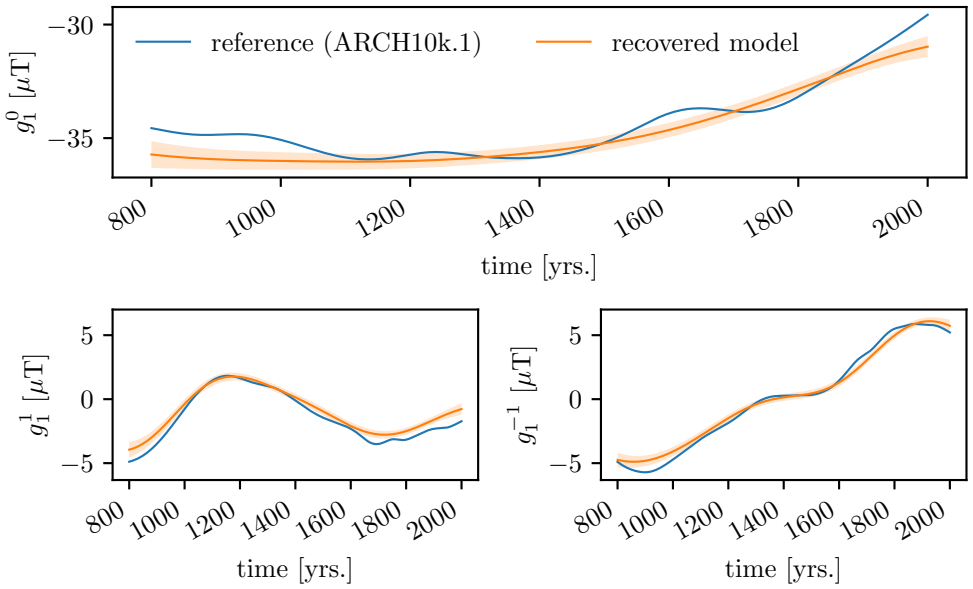

Figure 4. Comparison of the reference model to the one recovered from synthetic test data. The long-term behavior could be recovered, while data is too sparsely distributed and uncertainties are too large to recover the short timescale features. See the text for additional discussion.

In Figure 4 we compare the dipole of the reference model to our findings for the three dimensional synthetic data described at the beginning of this subsection. One can see that again only the long-term behavior of the reference model can be recovered from the data.

3.2 Case study

Here we present the results from applying the proposed modeling strategy to actual archeomagnetic data, taken from GEOMAGIA v3.3 (Brown et al., 2015). The data covers the interval $[800, 2000]$ AD and consists of 7801 records, of which 3.9 percent are complete vector triples. Figure 5 shows the spatial and temporal distribution of the data. Similar to MSKH20, we do not consider the Earth's ellipticity. We use the originally reported error estimates and assign $\alpha_{95} = 4.5°$ as directional errors and $\sigma_F = 8.25$ μT as intensity errors to the 8.4 percent of data, where no error is reported. As described above, the dating uncertainties are considered as standard deviations of independent normal distributions. When different values for upper and lower temporal error are reported, we use the bigger value. The 0.7 percent of data for which no dating uncertainty is reported are assigned a standard deviation $\sigma_t = 100$ yrs. Note that for the recent times fewer records are available from the archeomagnetic dataset. This results in bigger uncertainties towards recent times, as can be seen for example in the inclination series in Figure 8.

We compare our findings to two existing magnetic field models as well as to the results of MSKH20. The models are ARCH10k.1 (C. Constable et al., 2016) and COV-ARCH (Hellio & Gillet, 2018). They are considered reasonably comparable, as they are based on similar data compilations. Both models report Gauss coefficients up to SH de-

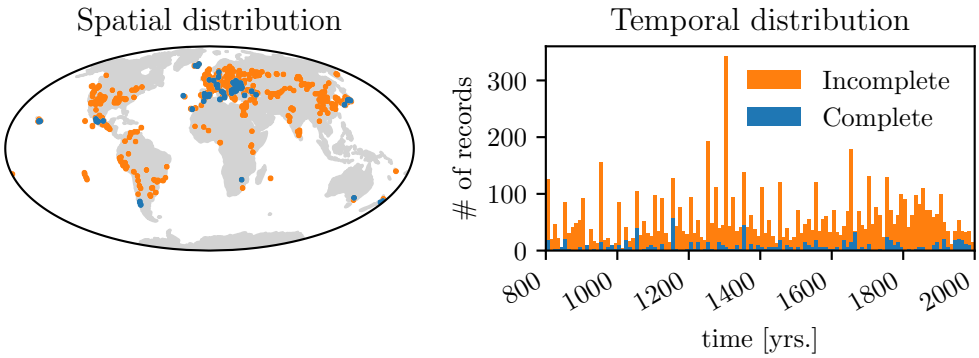

Figure 5. Spatial and temporal distribution of the data. We separate the data only into complete and incomplete without indicating declination, inclination and intensity separately, to avoid overloading the plots.

gree $\ell = 10$, while the actual spatial resolution is determined by regularization in ARCH10k.1 and by cross-covariances based on prior assumptions in COV-ARCH and lies around $\ell = 5$ in both cases. ARCH10k.1 does not report uncertainties, while COV-ARCH provides an ensemble of 50 models from which uncertainties are constructed by calculating sample mean and sample standard deviation. In contrast to the results presented by Hellio and Gillet (2018), the publicly available model is not time continuous but reports coarsely binned coefficients for every hundred years in the interval.

We want to stress again, that the presented results stem from a conceptual design. Especially the common temporal correlation time for all degrees $\ell \geq 2$ should be reconsidered, when building an actual model from the proposed strategy. The hyperparameter-MAP for the actual data is quite similar to the one in the synthetic data test (see section 3.1):

$$\hat{\gamma}_1^0 = -36.095\ \mu\text{T} \qquad \hat{\alpha}_{\text{DP}} = 1.16144\ \mu\text{T} \qquad \hat{\tau}_{\text{DP}} = 348.555\ \text{yrs.}$$
$$\hat{\epsilon} = 135.781\ \% \qquad \hat{\alpha}_{\text{ND}} = 39.4199\ \mu\text{T} \qquad \hat{\tau}_{\text{ND}} = 293.025\ \text{yrs.}$$
$$\hat{\rho} = 3.82749\ \mu\text{T}$$

Again, $\hat{\alpha}_{\text{ND}}$ is reported w.r.t. the reference radius and is way lower at the Earth's surface. All values are in a reasonable order of magnitude. Surprisingly the residual scaling did not decrease in comparison to MSKH20. The proximity of both correlation times may be explained by the dominance of the quadrupole over the larger degrees. For further insight, we provide profiled distributions together with the modeling software (Schanner & Mauerberger, 2020).

3.2.1 Field predictions

Predicting on the EMF's vector components is straightforward and given by wrapping equations 6 and 7, into the marginalization procedure described in the appendix. However, to get a reasonable spatial resolution too many design points have to be included to store the covariance matrices for all integration points, which is necessary to calculate the mixture distribution. Instead, similar to MSKH20, we resort to calculating the moment matching Gaussian proxy. The top row of Figure 6 shows mean and standard deviation for the down component of the EMF for the epoch 1700. Similarly, predictions at the core-mantle boundary (CMB) can be performed, by translating the design points accordingly. The results for the down component are shown for the epoch

1700 in Figure 7. Inferring the observables (declination, inclination and intensity) is hindered by the non-linear relation to the field. Utilizing again a linearization, mean and covariance can be constructed, similar to MSKH20 and Hellio et al. (2014, appendix A). The bottom row of Figure 6 shows a prediction of the EMF's intensity for the epoch 1700.

Compared to MSKH20 for snapshots in time, the new results show a slightly lower strength of the down component and lower field intensity. Moreover, the new results have a reduced standard deviation, which can be attributed to additional constraints due to the temporal correlations. The overall field structure is similar, showing features such as the South Atlantic Anomaly. The reconstruction at the CMB (Figure 7) reveals a region of lower field strength at the southern tip of Africa, which is at the limit of significance, but was not present in the snapshot model.

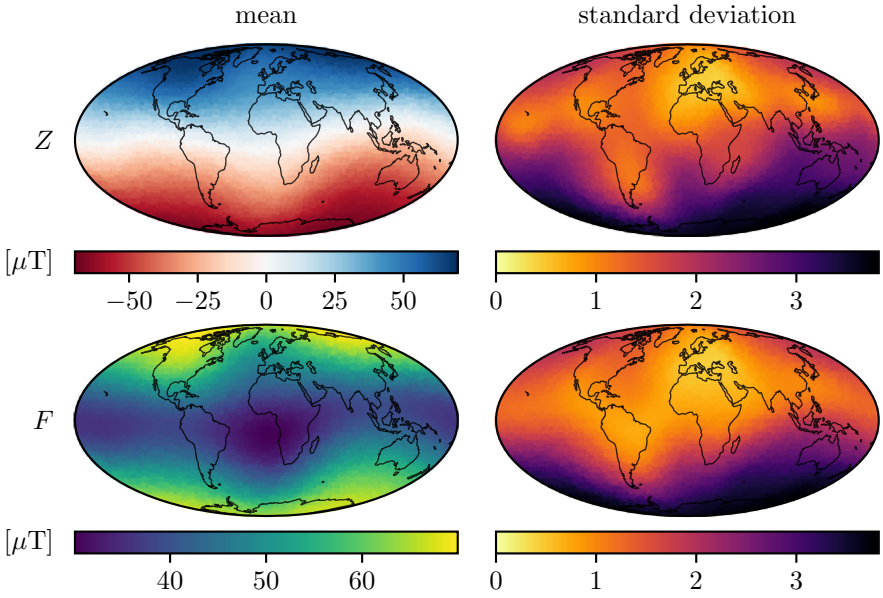

Figure 6. Maps of the EMFs down component (top) and intensity (bottom), together with standard deviations for the epoch 1700.

Similar to the map, predictions at a specific location can be obtained for various times and time series of the observables can be constructed. This is in contrast to the typical B-spline interpolation, as temporal design points may be chosen arbitrarily. Moment matching proxies have to be employed, as the number of parameters to infer is again too big when considering a reasonable temporal resolution. Figures 8 and 9 show time series at two distinct locations. Figure 8 presents time series for Paris, together with comparison models and data, while the series in Figure 9 are for a location in the Pacific, where no data is present in the surroundings. The data in Figure 8 stem from a surrounding of 250 km.

Inclinations and intensities are translated to the coordinates of Paris along the corresponding axial dipoles (Merrill et al., 1996). Declinations are taken as reported. For Paris, the different models mostly agree, with larger deviations towards the recent epoch,

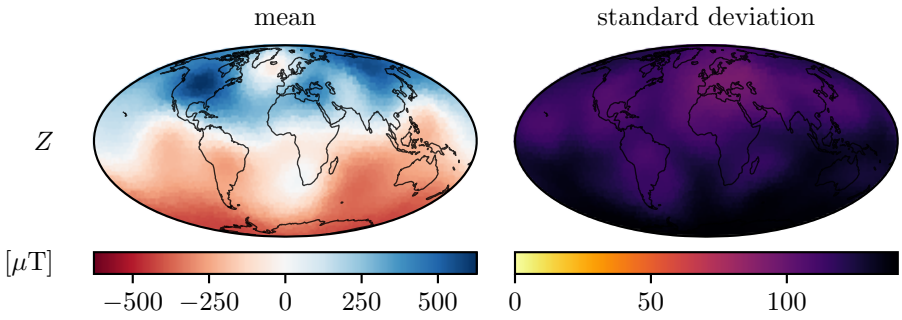

Figure 7. Map of the down component of the EMF at the CMB, together with standard deviation for the epoch 1700.

where the database is thinning out. Our inferred series shows less variation than comparable models, which probably can be attributed to our somewhat unrealistic temporal kernel with equal correlation time for all spatial wavelengths from ℓ=2 on. As expected, the uncertainties are bigger for the location in the Pacific, and the comparing model series show larger deviations.

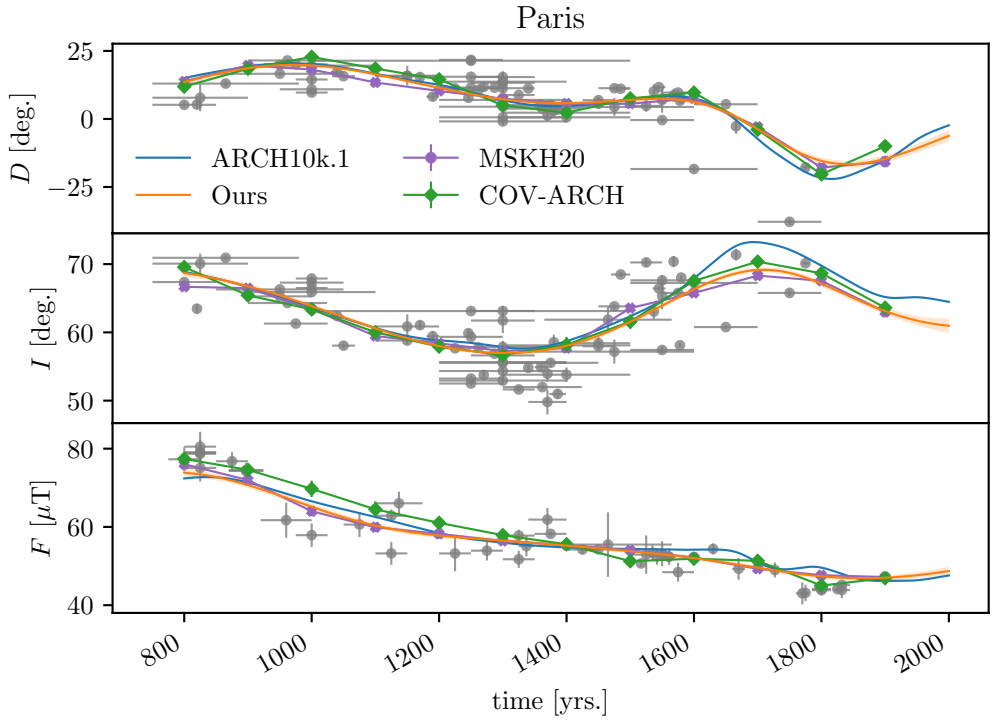

Figure 8. Data (grey dots) and model time series of the observables at Paris. Horizontal and vertical grey bars indicate the one sigma temporal and field component data uncertainties, respectively. For the MSKH20 snapshot model and COV-ARCH, which is reported in 100 year intervals, dots for the epochs are linearly connected by coloured lines.

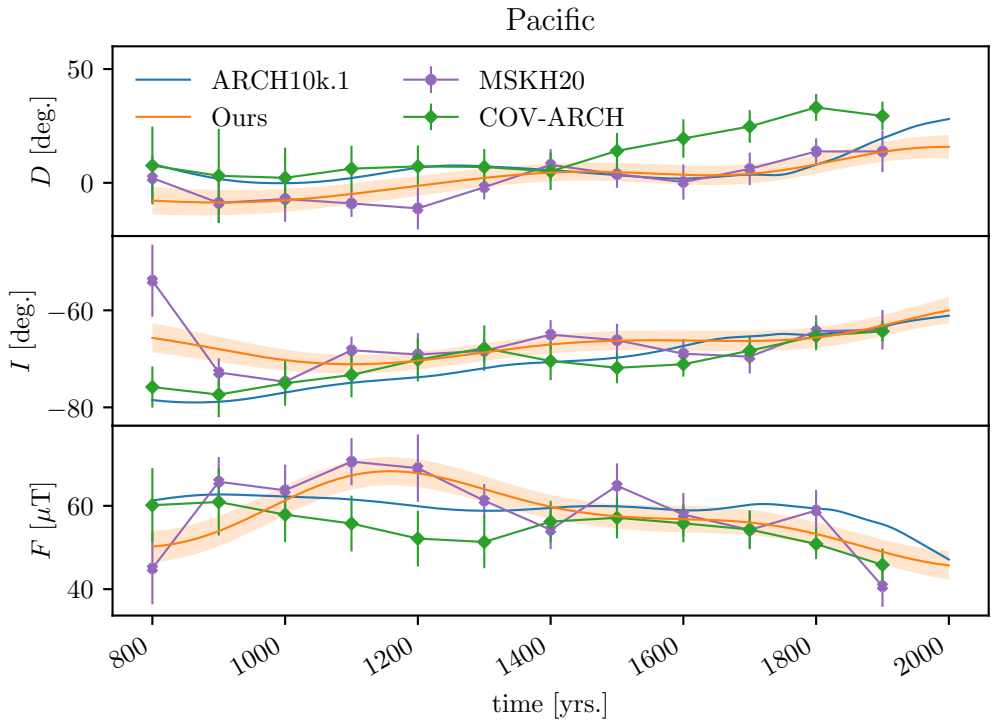

Figure 9. Model time series of the observables for a location in the Pacific (−48.18°, −128.03°) as in Fig. 8. Uncertainties are higher than for Paris (Figure 8) due to the lack of data in the surrounding.

3.2.2 Gauss coefficients

Although the proposed model is inherently non-parametric in both space and time, predictions on Gauss coefficients can be performed. As they are linearly related to the field, the procedure is straightforward (see Eq. 21 in Holschneider et al. (2016)). When predicting coefficients for a specific epoch, the full mixture distribution is accessible. When predicting on coefficient time series however, the number of parameters one has to include in the prediction to get a reasonable temporal resolution is too memory intense to store the covariance matrices of every individual collocation point. This is similar to predictions of the field in MSKH20. Therefore, the time series shown in Figure 10 present the moment matching Gaussian proxy to the actual mixture. The dipole coefficients from the new modeling strategy show a similar dynamic as those of the other models, with the "outliers" from MSKH20 disappearing. The series of the quadrupole coefficients show different behavior, with an interval of lower axial quadrupole strength around 1200 AD. Differences to ARCH10k.1 and COV-ARCH might partly be due to some differences in the underlying data compilation.

3.2.3 Spectra

Power spectra are considered to condense the information contained in the Gauss coefficients (e.g. Backus et al., 1996). Using sampling techniques, mean and percentiles of the geomagnetic power spectrum distribution are available (for further details consider MSKH20, Section 5.6). Figure 11 shows the geomagnetic power spectrum for the epoch 1700, together with 16- and 84-percentiles. Within the reported uncertainties, the

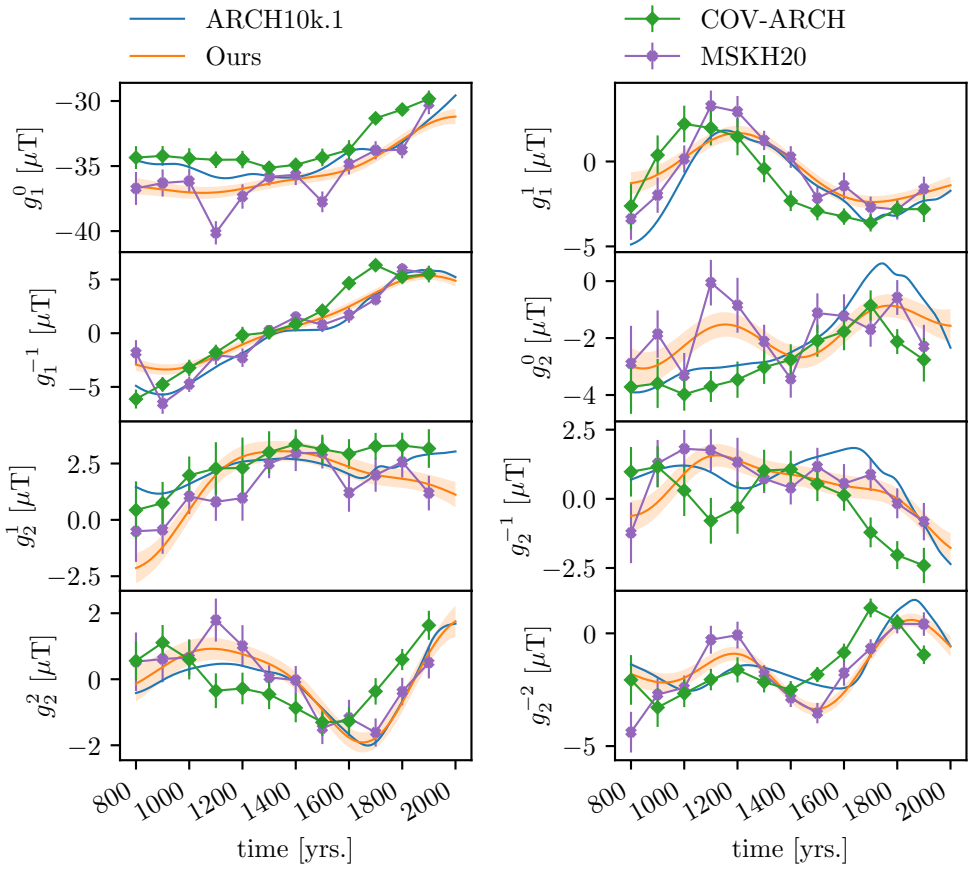

Figure 10. Gauss coefficient time series for degrees $\ell = 1$ and $\ell = 2$, together with comparable models. One sigma uncertainties are shown for our model in orange. MSKH20 and COV-ARCH, which are available in 100 year steps are shown by coloured dots linked linearly by solid lines.

power in dipole and quadrupole agrees to the comparison models. For $\ell = 3$ the reconstruction reports less power than COV-ARCH and MSKH20. Noteworthy is the faster power decay for degrees $\ell = 4 \dots 7$ when compared to MSKH20, which also indicates larger deviations from the prior. This may be due to temporal correlations increasing the information or the long correlation time damping small scale structures. Implementing a separate correlation time for each coefficient may provide further insight.

Similar to the geomagnetic power spectrum, the spectrum of the secular variation can be calculated (Alldredge, 1984). Therefore one has to predict on the derivatives of Gauss coefficients, also called the secular variations. As the derivative is a linear operator, this is straightforward. To explain the basic concept, consider the simplified example of direct observations of the EMF:

$$\begin{aligned}\mathbb{E}[\dot{g}(t)|O] =& \partial_t \mathbb{E}[g(t)|O] \\ =& \partial_t \bar{g}(t) + \partial_t K_{g,\boldsymbol{B}}(t,\mathbf{y})\big(K_{\boldsymbol{B}}(\mathbf{y},\mathbf{y}) + \Sigma_o\big)^{-1}\big(o(\mathbf{y}) - \bar{\boldsymbol{B}}(\mathbf{y})\big)\end{aligned} \tag{28}$$

Since the a priori mean in our model is constant, the first term vanishes. Thus, to predict on the secular variations one only has to calculate the derivative of the correlation between coefficients and observations, which in the suggested model reduces to the first

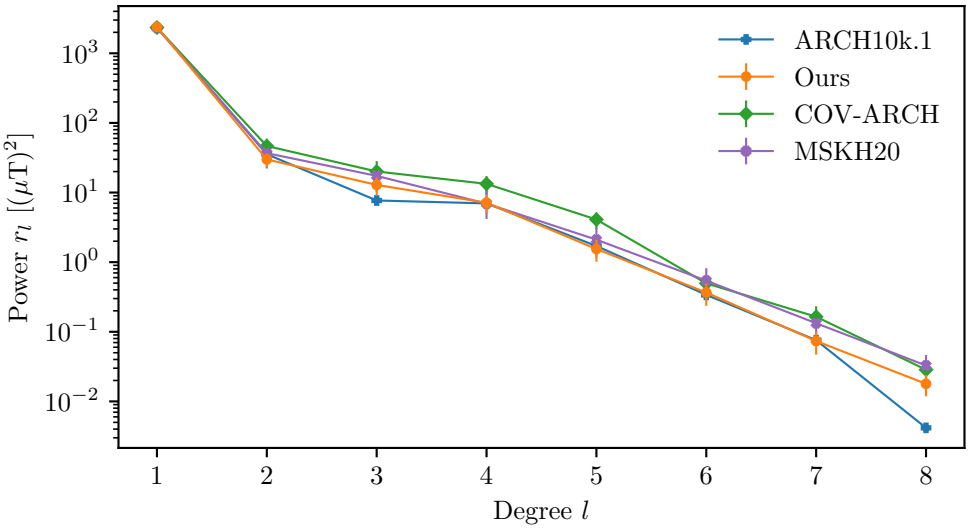

Figure 11. Comparison of the spatial power spectra for the epoch 1700 AD.

derivative of the SQE kernel:

$$\partial_t K_{\mathrm{SQE}}(t,t') = -\frac{2}{\tau^2}(t-t')K_{\mathrm{SQE}}(t,t') \tag{29}$$

All other quantities are known from inferring the Gauss coefficients themselves. The covariance of the secular variation translates analogously. Mean and percentiles of the secular variation spectrum are then again available via sampling.

As MSKH20 consists of snapshot models, the respective secular variation spectrum is not accessible. Similarly, as the publicly available version of COV-ARCH only reports values every 100 years, no secular variation can be calculated and the spectrum is missing in Figure 12. Note, that we choose an earlier epoch for the secular variation spectrum, as ARCH10k.1 is constrained to gufm1 (Jackson et al., 2000) for the recent times and therefore shows higher than average secular variation for these centuries. The secular variation spectra for the two models are fairly similar, with a very good agreement for the dipole and slightly higher values for the higher degrees in our new model.

3.2.4 Dipole

Finally we present the dynamics of the EMFs dipole. Figure 13 shows time series of the dipole moment magnitude. The magnitude is higher than the ones reported by comparison models, while the "outliers" from MSKH20 are not present in the new results. From 1840 on ARCH10k.1 is constrained by the gufm1 model, which in turn is constrained by a large amount of direct observations and can be considered to represent the dipole moment quite reliably from that time on. The deviation of our model from ARCH10k.1 during the last century is likely caused by a lack of archeomagnetic data for these epochs.

Figure 14 shows the movement of the geomagnetic north pole. The mean curve (black line) is calculated via sampling. For a given epoch the full density is available analytically (c.f. Mauerberger et al., 2020, Eq. 96). To not overload the plot, we only show mean and one-sigma ellipses for every century. The stereographic projection is responsible for the crescent-shaped distortion of the ellipses. The rapid movement of the dipole for earlier epochs suggested by other models is not found by our new reconstruction and for

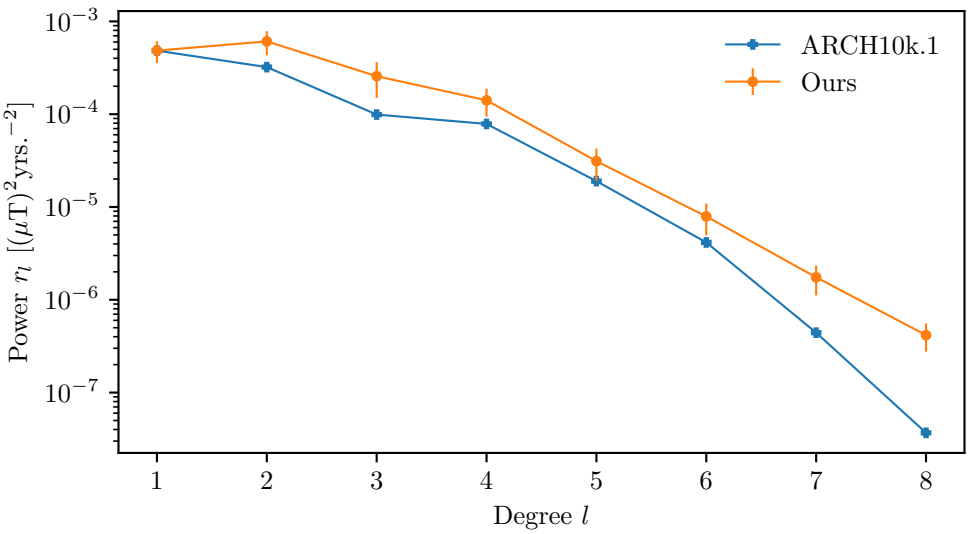

Figure 12. Secular variation spectrum for 1400 AD.

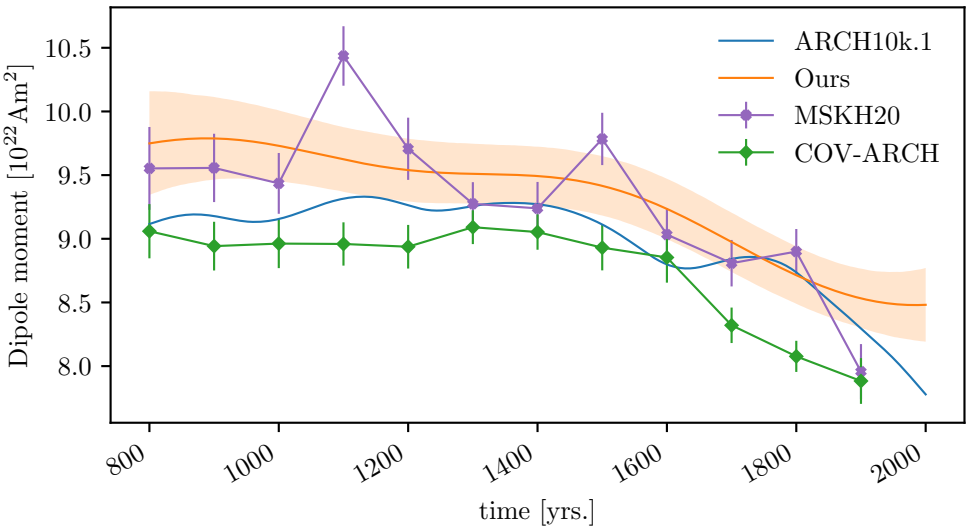

Figure 13. Time series of the dipole moment magnitude, together with comparison models. One sigma uncertainties are shown for our model in orange. The MSKH20 snapshot model and COV-ARCH, which are available in 100 year steps, are shown by colored dots linked linearly by solid lines.

more recent times, the path lies further to the west. Deviance for the most recent epochs is again caused by a lack of data.

4 Conclusions

The presented work extends the Bayesian strategy for correlation-based modelling of the archeomagnetic field introduced in MSKH20 to the temporal domain. In Section

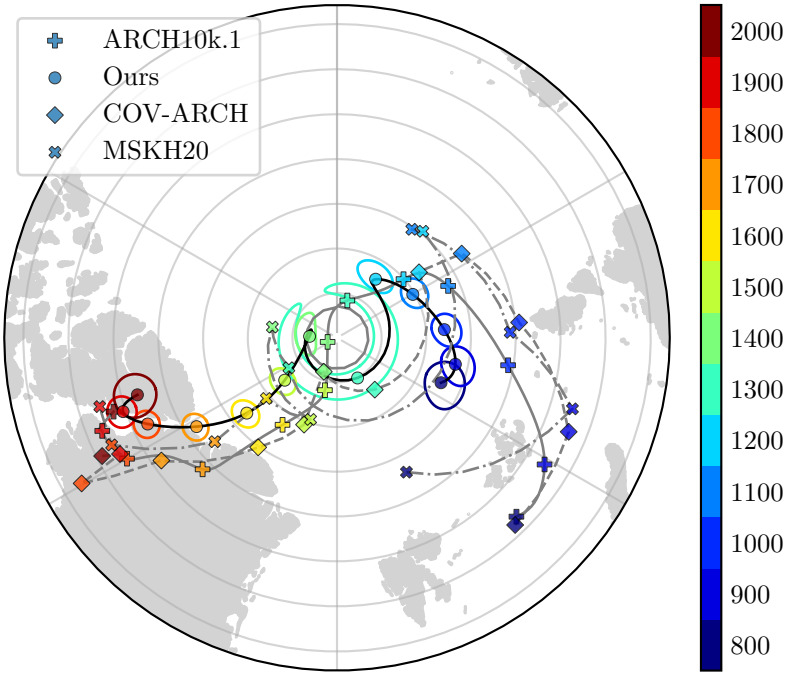

Figure 14. Movement of the geomagnetic north pole. Every one hundred years mean and one-sigma ellipses of the snapshot-distributions are shown. Color varies according to time, with markers every one hundred years.

2, all necessary modifications are discussed, together with a novel approach to include dating uncertainties. In contrast to previous works (Hellio et al., 2014; Hellio & Gillet, 2018; Korte et al., 2009; Licht et al., 2013; Hellio & Gillet, 2018; Senftleben, 2019), using a NIGP (McHutchon & Rasmussen, 2011) to incorporate dating uncertainties does not rely on sampling techniques. The a priori model is again constructed with the aim of being as objective as possible. The uninformative dipole prior from MSKH20 cannot easily be translated to the time-dynamic realm, as temporal correlations are send to zero together with the a priori precision and cannot easily be recovered in the posterior. Instead, we assume a priori a constant axial dipole with the dipole strength being a free parameter. Together with all but one other model parameters, the dipole strength is marginalized so that the model does not depend on the specific value. This marginalization presents another challenge, as numerical integration in a seven dimensional space has to be performed. The Riemann sum approach from MSKH20 is not applicable, due to the curse of dimensionality (i.e. unfeasible computation time). As a practicable alternative to the brute force integration we implement a CCD (S. M. Sanchez & Sanchez, 2005) based integration, as proposed by Rue et al. (2009). The major challenge in implementing this strategy consists of finding the MAP of the marginal posterior. Running the LIPO-TR algorithm (King, 2009, 2017) for the validation and case study datasets took around 25 hours each on a regular workstation. Once the MAP is found, the set of integration points consists of 79 hyperparameter combinations and the marginalization takes between half an hour and five hours, depending on the quantity one predicts on, as some quantities

require sampling, which is more computationally demanding. With the marginalization performed, the model depends only on the a priori choice of the Gauss coefficients covariance structure at the reference radius, the value of the reference radius and the temporal covariance structure. For the conceptual work presented, we chose an unphysical SQE kernel. This has the advantage of being easy to implement, but the main point why we use this kernel instead of a more reasonable one is the necessity to calculate temporal derivatives to implement the NIGP. For the SQE kernel this is straightforward. We have shown by means of a synthetic test and a case study on real data from 800 AD to 2000 AD that even with the simplified kernel the results compare well with previous models. Notably the "outliers" for the years 1100 and 1300, present in MSKH20, do not appear in the present work. This may be explained by the new model considering temporal errors, and thus covering a false binning, or by the long correlation time suppressing the influence of single records.

Implementing a more realistic kernel, such as the one proposed by Gillet et al. (2013), will be the direction of future work. Together with a Bayesian framework for data selection this will allow the construction and proposition of a new correlation based field model. In MSKH20, the expansion of the database by records from ship logs was discussed. Incorporating uncertainties arising from imprecise locations may be performed by the use of the proposed NIGP. Instead of temporal derivatives, the spatial gradients of the kernel are used to translate the input uncertainties to contributions to the measurement errors. The challenges to scale relative intensities and preserve stratification (Nilsson et al., 2014) persist, so that sediment records require a different approach than the application of the NIGP proposed here.

We again developed a python framework to save the effort of implementing the proposed algorithm (Schanner & Mauerberger, 2020). Together with extensive documentation, the software source code provides further insight into the modeling algorithm. It is available at `https://sec23.gitext-pages.gfz-potsdam.de/korte/corbam/`.

Acronyms

CCD central composite design
CMB core-mantle boundary

EMF Earth's magnetic field

GP Gaussian process

MAP maximum a posteriori probability estimator
MCMC Markov Chain Monte-Carlo

NIGP noisy input Gaussian process

POE point of expansion

SH spherical harmonics
SQE squared exponential

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 S. Mauerberger and M. Holschneider. The manuscript was assembled

by M. Schanner, with support from all co-authors. Software development and data processing was conducted by M. Schanner, with contributions from S. Mauerberger. M. Korte took care of data selection as well as interpretation and embedding of the case study. The work and findings were supervised by M. Korte and M. Holschneider.

Special thanks go to H. Matuschek for providing FieldTools (Matuschek & Mauerberger, 2019), assistance and support and to S. Panovska for helpful discussions on the secular variation spectrum.

There is no new data involved in this publication. The data used in the case study is available via GEOMAGIA v3.3 (Brown et al., 2015). All results were produced using a python implementation of the discussed algorithm, which is publicly available (Schanner & Mauerberger, 2020).

Appendix A Constructing the spatial covariance kernel

The Lagrange kernel (Eq. 9) is constructed from the Gauss coefficient correlations as follows:

Consider the covariance of the magnetic potential in SH decomposition. Then, for a pontential of internal origin

$$\mathrm{Cov}[\Phi(\boldsymbol{x}),\Phi(\boldsymbol{x}')] = R^2 \sum_{\ell,m}\sum_{\ell',m'} \left(\frac{R^2}{|\boldsymbol{x}||\boldsymbol{x}'|}\right)^{\ell+1} Y_\ell^m(\hat{\boldsymbol{x}}) Y_{\ell'}^{m'}(\hat{\boldsymbol{x}}') \mathrm{Cov}\left[g_\ell^m, g_{\ell'}^{m'}\right] \ . \tag{A1}$$

Assuming that at some reference sphere the Gauss coefficients are uncorrelated with a flat spectrum, i.e.

$$\mathrm{Cov}\left[g_\ell^m, g_{\ell'}^{m'}\right] = \delta_{\ell,\ell'}\delta_{m,m'} \ , \tag{A2}$$

where δ_{ij} refers to the Kronecker delta. This gives

$$\mathrm{Cov}[\Phi(\boldsymbol{x}),\Phi(\boldsymbol{x}')] = R^2 \sum_{\ell=0}^{\infty} \left(\frac{R^2}{|\boldsymbol{x}||\boldsymbol{x}'|}\right)^{\ell+1} \sum_m Y_\ell^m(\hat{\boldsymbol{x}}) Y_\ell^m(\hat{\boldsymbol{x}}') \ . \tag{A3}$$

Following Holschneider et al. (2016), evaluating the sums gives the kernel

$$K_{\mathrm{L}}(\boldsymbol{x},\boldsymbol{x}') = \frac{1}{\sqrt{1-2b+a^2}} \ , \tag{9}$$

where $b = \boldsymbol{x}\cdot\boldsymbol{x}'/R^2$ and $a = |\boldsymbol{x}||\boldsymbol{x}'|/R^2$. The dipole kernel can be extracted by setting $\ell = 1$. This yields

$$K_{\mathrm{DP,S}}(\boldsymbol{x},\boldsymbol{x}') = \frac{b}{a^3} \ . \tag{A4}$$

Thus the non-dipole kernel reads

$$K_{\mathrm{ND,S}}(\boldsymbol{x},\boldsymbol{x}') = K_{\mathrm{L}}(\boldsymbol{x},\boldsymbol{x}') - K_{\mathrm{DP,S}}(\boldsymbol{x},\boldsymbol{x}') - \frac{1}{a} \ , \tag{A5}$$

where the last term excludes the monopole ($\ell = 0$).

Appendix B Detailed modeling algorithm

The modeling algorithm consists of two stages. The first one deals with finding the MAP of the hyperparameters. The MAP then serves as a center point for marginalizing the hyperparameters in the second step. We begin this section by laying out the inversion process. From quantities that are calculated along the way, the marginal posterior can be constructed. Using both procedures, the full algorithm can presented in a compact way.

B1 Inversion

The inversion closely follows the modeling concept described in MSKH20. To provide insight into the mathematical background of the inversion, we lay out the full inversion process for the field $\boldsymbol{B}$ at locations of interest $\mathbf{y}$. Inverting for other quantities, such as the Gauss coefficients or the field's intensity, is straightforward.

To keep the equations concise, we use the following notation for matrices:

$\bar{\boldsymbol{B}}(\mathbf{x})$ is the a priori (mean) field at locations $\mathbf{x}$.

$\nabla\mathbf{H}|_{\tilde{\boldsymbol{B}}(\mathbf{x})}$ refers to the gradient of the observation functionals (c.f. Eqs. 14-16), evaluated at the POE $\tilde{\boldsymbol{B}}$ at locations $\mathbf{x}$.

$\bar{\boldsymbol{H}}(\mathbf{x})$ refers to the linearized, transformed prior field at locations $\mathbf{x}$

$$\bar{\boldsymbol{H}}(\mathbf{x}) = \mathbf{H}\big(\tilde{\boldsymbol{B}}(\mathbf{x})\big) + \nabla\mathbf{H}|_{\tilde{\boldsymbol{B}}(\mathbf{x})}\Big(\tilde{\boldsymbol{B}}(\mathbf{x}) - \bar{\boldsymbol{B}}(\mathbf{x})\Big) \tag{B1}$$

The transformed prior field serves as a mean proxy to observations at locations $\mathbf{x}$.

$\Sigma_{\mathbf{yy}}$ refers to the a priori covariance of the field at locations $\mathbf{y}$. This is

$$\Sigma_{\mathbf{yy}} = K_{\boldsymbol{B}}(\mathbf{y}, \mathbf{y}) \tag{B2}$$

This is a matrix, composed of 3×3 blocks, containing correlations at each point.

$\bar{\boldsymbol{B}}_{|o}(\mathbf{y})$ is short hand for the posterior mean of the field at locations $\mathbf{y}$, given observations o, i.e.

$$\bar{\boldsymbol{B}}_{|o}(\mathbf{y}) = \mathbb{E}[\boldsymbol{B}(\mathbf{y})|o] \tag{B3}$$

$\bar{\boldsymbol{H}}_{|o}(\mathbf{x})$ is similar to $\bar{\boldsymbol{H}}(\mathbf{x})$ and refers to the linearized transformed mean at locations $\mathbf{x}$, posterior to observations o.

$\Sigma_{\mathbf{yy}|o}$ is short hand for the posterior covariance of the field, i.e.

$$\Sigma_{\mathbf{yy}|o} = \mathrm{Cov}[\boldsymbol{B}(\mathbf{y}), \boldsymbol{B}(\mathbf{y})|o] \tag{B4}$$

$\Sigma_{\mathbf{y}o}$ refers to cross-covariance between the field at points of interest $\mathbf{y}$ and observations of the field o. As these are linearized, the matrices involve a gradient

$$\Sigma_{\mathbf{y}o} = K_{\boldsymbol{B}}(\mathbf{y}, \mathbf{x}_o)\nabla\mathbf{H}|_{\tilde{\boldsymbol{B}}(\mathbf{x}_o)} \ . \tag{B5}$$

$\mathbf{x}_o$ are the locations of the observations o. The dot product is taken pointwise, i.e. for every observation $\mathbf{x}_o$.

Σ_{oo} refers to the covariance amongst observations:

$$\Sigma_{oo} = \nabla\mathbf{H}|^{\top}_{\tilde{\boldsymbol{B}}(\mathbf{x}_o)}\Big(K_{\boldsymbol{B}}(\mathbf{x}_o, \mathbf{x}_o) + \Sigma_T + \rho\Sigma_p\Big)\nabla\mathbf{H}|_{\tilde{\boldsymbol{B}}(\mathbf{x}_o)} + \epsilon\Sigma_e \tag{B6}$$

The residual Σ_p is an identity matrix of the dimension of number of observations and Σ_e is the (typically diagonal) matrix of approximate measurement errors, see Section 2.3. Σ_T is the correction term for dating uncertainties, see Section 2.4:

$$\Sigma_T = \Sigma_{tt'} \circ \partial_t\partial_{t'}K_{\boldsymbol{B}}(t, t')|_{t_o}$$

As the linearization is tackled by means of a two-step strategy, at first the data is partitioned into complete c and incomplete i records:

$$o = \{c, i\} \tag{B7}$$

The first step in the two-step strategy only deals with complete records. However, as the posterior mean from the first step serves as the point of expansion (POE) in the second

step, predictions at locations of incomplete records $\mathbf{x}_i$ have to be included as well. The posterior mean and covariance from the first step read

$$\mathbb{E}[\boldsymbol{B}(\mathbf{y})|c] = \bar{\boldsymbol{B}}(\mathbf{y}) + \Sigma_{\mathbf{y}c} \cdot \Sigma_{cc}^{-1} \cdot \Big(c - \bar{\boldsymbol{H}}(\mathbf{x}_c)\Big) \tag{B8}$$

$$\mathrm{Cov}[\boldsymbol{B}(\mathbf{y}), \boldsymbol{B}(\mathbf{y})|c] = \Sigma_{\mathbf{yy}} - \Sigma_{\mathbf{y}c} \cdot \Sigma_{cc}^{-1} \cdot \Sigma_{\mathbf{y}c}^{\top} \tag{B9}$$

Posterior correlations for the incomplete records are given by

$$\Sigma_{\mathbf{y}i|c} = \Sigma_{\mathbf{y}i} - \Sigma_{\mathbf{y}c} \cdot \Sigma_{cc}^{-1} \cdot \Sigma_{ic}^{\top} \tag{B10}$$

$$\Sigma_{ii|c} = \Sigma_{ii} - \Sigma_{ic} \cdot \Sigma_{cc}^{-1} \cdot \Sigma_{ic}^{\top} \tag{B11}$$

To calculate the relevant linearized quantities, the POE is calculated as the inverse observation functional, i.e. in the first step

$$\tilde{\boldsymbol{B}}(\mathbf{x}_c) = \mathbf{H}^{-1}(c) \tag{B12}$$

with

$$\mathbf{H}^{-1} : \begin{pmatrix} D \\ I \\ F \end{pmatrix} \rightarrow \boldsymbol{B} = F \begin{pmatrix} \cos(I)\cos(D) \\ \cos(I)\sin(D) \\ \sin(I) \end{pmatrix} . \tag{B13}$$

In the second step, the posterior mean of the first step is used as POE:

$$\tilde{\boldsymbol{B}}(\mathbf{x}_i) = \bar{\boldsymbol{B}}_{|c}(\mathbf{x}_i) \tag{B14}$$

The remaining records are incorporated and give a posterior with

$$\mathbb{E}[\boldsymbol{B}(\mathbf{y})|c, i] = \bar{\boldsymbol{B}}_{|c}(\mathbf{y}) + \Sigma_{\mathbf{y}i|c} \cdot \Sigma_{ii|c}^{-1} \cdot \Big(i - \bar{\boldsymbol{H}}_{|c}(\mathbf{x}_i)\Big) \tag{B15}$$

$$\mathrm{Cov}[\boldsymbol{B}(\mathbf{y}), \boldsymbol{B}(\mathbf{y})|c, i] = \Sigma_{\mathbf{yy}|c} - \Sigma_{\mathbf{y}i|c} \cdot \Sigma_{ii|c}^{-1} \cdot \Sigma_{\mathbf{y}i|c}^{\top} \tag{B16}$$

During the illustrated procedure, the hyperparameters are assumed to be known. As this is a priori not the case, we next illustrate how to marginalize them.

B2 Marginal posterior

The density

$$p(o|\eta) = \int p(o|\boldsymbol{B}) p(\boldsymbol{B}|\eta) \, \mathrm{d}\boldsymbol{B} \tag{B17}$$

is called *marginal likelihood* where the term marginal refers to the integration over the EMF. $p(o|\eta)$ is a function in the hyperparameters η, given the data o. For a certain choice of hyperparameters, $p(o|\eta)$ describes how likely the observations are. In our case, η consists of the a priori dipole strength and variances, the correlation times and the error and residual scalings, $\eta = \{\gamma_1^0, \alpha_{\mathrm{DP}}, \alpha_{\mathrm{ND}}, \tau_{\mathrm{DP}}, \tau_{\mathrm{ND}}, \epsilon, \rho\}$, see also Section 2.5. Building the compound distribution for the EMF (Eq. 25) requires calculating the marginal likelihood. Fortunately, this can be done using expressions from the previous section. In the same language as before, the marginal likelihood is also given by a two-step formula

$$p(o = \{c, i\}|\eta) = p(i|\eta, c) p(c|\eta) \ . \tag{B18}$$

The conditional in $p(i|\eta, c)$ refers again to the POE in calculating the gradients for linearization. With the matrix notation from before, one has (Rasmussen & Williams, 2006)

$$p(c|\eta) = \frac{\exp\Big(-\frac{1}{2}\big(c - \bar{\boldsymbol{H}}(\mathbf{x}_c)\big)^T \Sigma_{cc}^{-1} \big(c - \bar{\boldsymbol{H}}(\mathbf{x}_c)\big)\Big)}{\sqrt{(2\pi)^{3n_c}|\Sigma_{cc}|}} \tag{B19}$$

$$p(i|c, \eta) = \frac{\exp\Big(-\frac{1}{2}\big(i - \bar{\boldsymbol{H}}_{|c}(\mathbf{x}_i)\big)^T \Sigma_{ii|c}^{-1} \big(i - \bar{\boldsymbol{H}}_{|c}(\mathbf{x}_i)\big)\Big)}{\sqrt{(2\pi)^{n_i}|\Sigma_{ii|c}|}} \ . \tag{B20}$$

Here n_c and n_i refer to the numbers of complete and incomplete records respectively. Multiplying the marginal likelihood with a prior over the hyperparameters gives the *marginal posterior*, up to a normalization constant:

$$p(\eta|o) \propto p(o|\eta)p(\eta) \tag{B21}$$

For numerical reasons, one often uses the log marginal posterior

$$\ln p(\eta|o) = \ln p(o|\eta) + \ln p(\eta) + \text{const.} \; . \tag{B22}$$

B3 Exploration and integration

With the details outlined in the previous sections, we can now describe the main two stages of the modeling algorithm:

Exploration The first stage consists of finding the maximum a posteriori probability estimator (MAP) $\hat{\eta}$ of the marginal posterior. Therefore, the log marginal posterior is optimized using global optimization techniques.

Integration With the MAP as center, a set of integration points $\mathcal{S}_\eta$ is constructed as described by Rue et al. (2009) and S. M. Sanchez and Sanchez (2005). With weights Δ_η, the integral for the compound distribution Eq. 25 is approximated by a sum

$$\int p(\boldsymbol{B}|o,\eta) \cdot p(\eta|o) \; \mathrm{d}\eta \approx \sum_{\eta \in \mathcal{S}_\eta} p(\boldsymbol{B}|o,\eta) \cdot p(\eta|o) \; \Delta\eta \; . \tag{B23}$$

This way the compound distribution is approximated by a Gaussian mixture. Similar expressions exist for the compound distributions of all quantities of interest, such as Gauss coefficients or observables like the linearized intensity F.

References

Alldredge, L. R. (1984). Harmonics required in main field and secular variation models. *Journal of geomagnetism and geoelectricity*, *36*(2), 63-72. doi: 10.5636/jgg.36.63

Backus, G., Parker, R., & Constable, C. (1996). *Foundations of geomagnetism*. Cambridge University Press.

Bloxham, J., & Jackson, A. (1992). Time-dependent mapping of the magnetic field at the core-mantle boundary. *Journal of Geophysical Research: Solid Earth*, *97*(B13), 19537-19563. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/92JB01591` doi: 10.1029/92JB01591

Bouligand, C., Gillet, N., Jault, D., Schaeffer, N., Fournier, A., & Aubert, J. (2016, 11). Frequency spectrum of the geomagnetic field harmonic coefficients from dynamo simulations. *Geophysical Journal International*, *207*, 1142-1157. doi: 10.1093/gji/ggw326

Bouligand, C., Hulot, G., Khokhlov, A., & Glatzmaier, G. (2005, 06). Statistical paleomagnetic field modeling and dynamo numerical simulation. *Geophysical Journal International*, *161*, 603-626. doi: 10.1111/j.1365-246X.2005.02613.x

Brown, M. C., Donadini, F., Nilsson, A., Panovska, S., Frank, U., Korhonen, K., ... Constable, C. G. (2015, May 12). 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

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, 06). 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, 08). 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

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.

Licht, A., Hulot, G., Gallet, Y., & Thbault, 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

Matuschek, H., & Mauerberger, S. (2019). *Toolbox for manipulating vector fields on the sphere.* GFZ Data Services. Retrieved from `http://doi.org/10.5880/fidgeo.2019.033` doi: 10.5880/fidgeo.2019.033

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 in-

put 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., Holme, R., Korte, M., Suttie, N., & Hill, M. (2014, 05). Reconstructing Holocene geomagnetic field variation: new methods, models and implications. *Geophysical Journal International*, *198*(1), 229-248. doi: 10.1093/gji/ggu120

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

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319-392. Retrieved from `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00700.x` doi: 10.1111/j.1467-9868.2008.00700.x

Sanchez, S., Fournier, A., Aubert, J., Cosme, E., & Gallet, Y. (2016, 08). Modelling the archaeomagnetic field under spatial constraints from dynamo simulations: a resolution analysis. *Geophysical Journal International*, *207*(2), 983-1002. doi: 10.1093/gji/ggw316

Sanchez, S. M., & Sanchez, P. J. (2005, October). Very large fractional factorial and central composite designs. *ACM Trans. Model. Comput. Simul.*, *15*(4), 362377. Retrieved from `https://doi.org/10.1145/1113316.1113320` doi: 10.1145/1113316.1113320

Schanner, M. A., & Mauerberger, S. (2020). *CORBAM: CORrelation Based Archeomagnetic Modeling.* Potsdam: GFZ Data Services. Retrieved from `https://doi.org/10.5880/GFZ.2.3.006`

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

Snelson, E., Rasmussen, C. E., & Ghahramani, Z. (2003). Warped Gaussian processes. In *In advances in neural information processing systems (nips* (p. 2003). MIT Press.

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

Thébault, E., Finlay, C. C., Beggan, C. D., Alken, P., Aubert, J., Barrois, O., ... Zvereva, T. (2015, May 27). International geomagnetic reference field: the 12th generation. *Earth, Planets and Space*, *67*(1), 79. doi: 10.1186/s40623-015-0228-9