Publication-based Doctoral Thesis

Correlation Based Modeling of the Archeomagnetic Field

by

Maximilian Arthus Schanner

Univ.-Diss.
zur Erlangung des akademischen Grades
“doctor rerum naturalium”
(Dr. rer. nat.)

in der Wissenschaftsdiziplin

Angewandte Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Mathematik
der Universität Potsdam

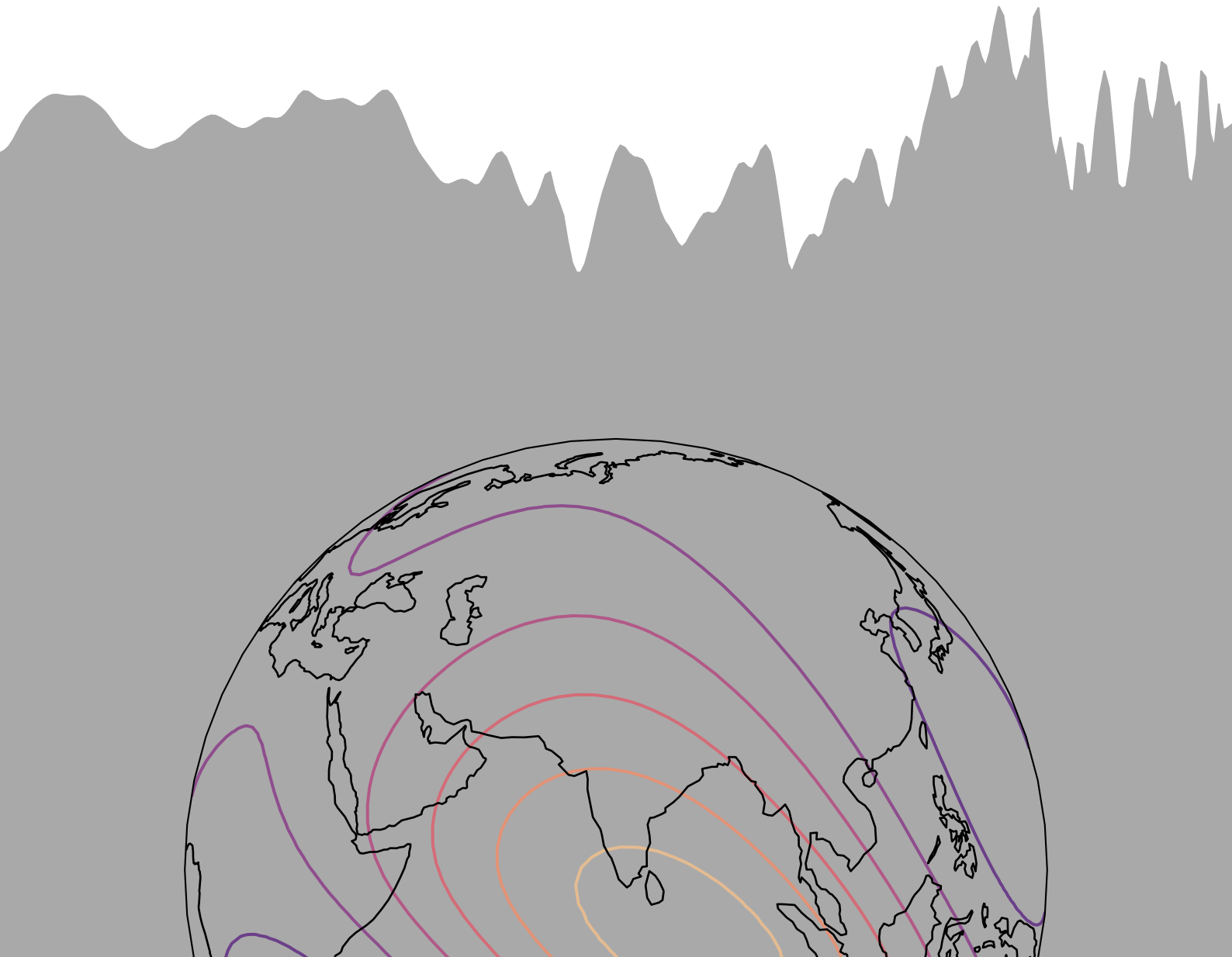
Disputation: Potsdam, den 31.05.2022

Betreuer*innen:
Prof. Dr. Matthias Holschneider
Dr. Monika Korte
Gutachter*innen:
Prof. Dr. Matthias Holschneider
Dr. Monika Korte
Dr. Andreas Nilsson

Published online on the Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-55587
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-555875

Maximilian Arthus Schanner

Correlation Based Modeling of the Archeomagnetic Field



Summary

The geomagnetic main field is vital for live on Earth, as it shields our habitat against the solar wind and cosmic rays. It is generated by the geodynamo in the Earth's outer core and has a rich dynamic on various timescales. Global models of the field are used to study the interaction of the field and incoming charged particles, but also to infer core dynamics and to feed numerical simulations of the geodynamo. Modern satellite missions, such as the SWARM or the CHAMP mission, support high resolution reconstructions of the global field. From the 19^{th} century on, a global network of magnetic observatories has been established. It is growing ever since and global models can be constructed from the data it provides. Geomagnetic field models that extend further back in time rely on indirect observations of the field, i.e. thermoremanent records such as burnt clay or volcanic rocks and sediment records from lakes and seas. These indirect records come with (partially very large) uncertainties, introduced by the complex measurement methods and the dating procedure.

Focusing on thermoremanent records only, the aim of this thesis is the development of a new modeling strategy for the global geomagnetic field during the Holocene, which takes the uncertainties into account and produces realistic estimates of the reliability of the model. This aim is approached by first considering snapshot models, in order to address the irregular spatial distribution of the records and the non-linear relation of the indirect observations to the field itself. In a Bayesian setting, a modeling algorithm based on Gaussian process regression is developed and applied to binned data. The modeling algorithm is then extended to the temporal domain and expanded to incorporate dating uncertainties. Finally, the algorithm is sequentialized to deal with numerical challenges arising from the size of the Holocene dataset.

The central result of this thesis, including all of the aspects mentioned, is a new global geomagnetic field model. It covers the whole Holocene, back until 12000 BCE, and we call it ArchKalmag14k. When considering the uncertainties that are produced together with the model, it is evident that before 6000 BCE the thermoremanent database is not sufficient to support global models. For times more recent, ArchKalmag14k can be used to analyze features of the field under consideration of posterior uncertainties. The algorithm for generating ArchKalmag14k can be applied to different datasets and is provided to the community as an open source python package.

Zusammenfassung

Das geomagnetische Hauptfeld ist essenziell für das Leben auf der Erde, da es unseren Lebensraum gegen den Sonnenwind und kosmische Strahlung abschirmt. Es wird vom Geodynamo im Erdkern erzeugt und zeigt eine komplexe Dynamik auf unterschiedlichen Zeitskalen. Globale Modelle des Magnetfelds werden zur Studie der Wechselwirkung von einströmenden geladenen Teilchen genutzt, aber auch um Kerndynamiken zu untersuchen und um sie in numerische Simulationen des Geodynamos einzuspeisen. Moderne Satellitenmissionen, wie SWARM und CHAMP, stützen hochauflösende Rekonstruktionen des globalen Felds. Seit dem 19. Jahrhundert wird ein globales Netzwerk von magnetischen Observatorien aufgebaut. Es wächst stetig und globale Modelle können aus den Daten, die es liefert, konstruiert werden. Geomagnetische Feldmodelle, die weiter in der Zeit zurückreichen, basieren auf indirekten Beobachtungen des Felds, d.h. auf thermoremanenten Daten, wie gebrannten Tonen oder vulkanischen Gesteinen, und auf Sedimentdaten aus Seen und Meeren. Diese indirekten Beobachtungen werden mit (teilweise sehr hohen) Unsicherheiten geliefert, die aus den komplexen Datierungs- und Messmethoden resultieren.

Ziel dieser Arbeit ist die Entwicklung einer neuen Modellierungsmethode für das globale geomagnetische Feld während des Holozäns, welche die Unsicherheiten berücksichtigt und realistische Schätzungen für die Verlässlichkeit des Modells liefert. Dabei werden lediglich thermoremanente Daten betrachtet. Diesem Ziel wird sich zunächst genähert, indem ein Schnappschuss-Modell konstruiert wird, um die unregelmäßige räumliche Verteilung der Daten und die nichtlineare Beziehung zwischen Daten und Magnetfeld zu untersuchen. In einem Bayesianischen Rahmen wird ein auf Gaussprozessen basierender Algorithmus entwickelt und zunächst auf diskretisierte Daten angewendet. Dieser Algorithmus wird dann um eine zeitabhängige Komponente ergänzt und erweitert, um Datierungsfehler zu berücksichtigen. Zuletzt wird der Algorithmus sequenzialisiert, um mit numerischen Herausforderungen umzugehen, die aufgrund der Größe des Holozän-Datensatzes bestehen.

Das zentrale Ergebnis dieser Arbeit, welches alle genannten Aspekte beinhaltet, ist ein neues globales geomagnetisches Feldmodell. Es deckt das gesamte Holozän ab, bis ins Jahr 12000 BCE, und wir nennen es ArchKalmag14k. Bei Betrachtung der Unsicherheiten, die gemeinsam mit dem Modell ermittelt werden, wird deutlich, dass die thermoremanente Datenbasis nicht ausreicht, um globale Modelle vor dem Jahr 6000 BCE zu stützen. Für jüngere Zeiträume kann ArchKalmag14k genutzt werden, um Merkmale des Erdmagnetfelds unter Berücksichtigung der a posteriori Unsicherheiten zu analysieren. Der Algorithmus, mit dem ArchKalmag14k erzeugt wurde, kann auf weitere Datensätze angewendet werden und wird als quelloffenes python-Paket zur Verfügung gestellt.

Acknowledgements

Working on this thesis and the publications it contains during a global pandemic was a peculiar experience. Luckily, the pandemic had a relatively low impact on my work. However, working mostly in home office it was at times hard to grasp that research is a team effort and greatly relies on communication and collaboration.

I would like to thank my supervisors Prof. Dr. Matthias Holschneider and Dr. Monika Korte for generously sharing advice, insights and knowledge in countless virtual meetings. I appreciate that when I needed it, they reassured me that things would work out. Next, I wish to thank my close colleague Stefan Mauerberger for the great time we had in our shared office at the beginning of this project. Unfortunately, we could not continue this fruitful collaboration for the whole length of the project. I further wish to thank all of my colleagues at GFZ, Section 2.3, and at the Institute of Applied Mathematics at the University of Potsdam, for creating a productive and pleasant working atmosphere. Special thanks go to Dr. Sanja Panovska, Dr. Hannes Matuschek, Dr. Julien Baerenzung and Dr. Ahmed Nasser Mahgoub for valuable scientific discussions. Also to Dr. Leonie Pick and Dr. Juan Rodríguez-Zuluaga for helping me get settled and for much appreciated general advice. And to Mareike Lausberg and Alex Jordan for administrative and technical assistance.

Thank you Felix, Lennart, Lukas, Martin and Robert, for helping me through the challenging parts of this process and you, Moritz for your unique approach to things. Thank you Inga for your views on academia and on life in general. Katrin, you always have my back and I am grateful that you are a part of my life. Without you, my dear housemates, I would not have made it through the pandemic, you bring me joy everyday. And you, Jule, you're the best.

Finally, I want to thank my parents and my brother for their unconditional support on all my endeavors and for always being there.

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant 388291411.

Contents

Overview

The work presented in this thesis is closely tied to DFG project 388291411. The aim of the project was the development of a novel correlation based modeling strategy, applicable to paleomagnetic data from the Holocene. Related research tasks include

- the assessment of reliable uncertainty estimates, that represent both the modeling process and uncertainties in the data, especially due to the dating methods
- the estimation of the temporal model resolution
- the extraction and study of model features under consideration of the estimated model uncertainties

These tasks are mainly motivated by studies of geomagnetism and a brief overview of global geomagnetic modeling is given in Chapter 1. To fill the more technical work in the rest of this thesis with some life, Chapter 2 describes the measurement process for paleomagnetic data of archaeological and volcanic origin. The modeling procedure is developed within the rich mathematical framework of Gaussian process regression. Chapter 3 provides some insight into more mathematical aspects that are missing in Part II. The foundation in this regard is presented in Holschneider et al. (2016). The publications included in Part II extend and apply the fundamental concepts of Holschneider et al. (2016) to paleomagnetism. Chapter 4 approaches the general model by first considering snapshots and focusing on the linearization of the paleomagnetic observation functionals and the estimation of model parameters and uncertainties from the magnetic measurements. Chapter 5 builds on this by introducing temporal correlations and considering uncertainties from the dating process. In Chapter 6 a new global geomagnetic model for the Holocene is proposed, that is constructed by sequentializing the developed method in order to make it suitable for larger datasets. Part III completes this thesis and includes overall conclusions in Chapter 7 as well as an outlook to following work in Chapter 8.

Part I

Introduction

1 Modeling the global geomagnetic field

This section mainly follows Chapter 1 of Merrill, McElhinny, and McFadden (1996).

Magnetism has fascinated humans for thousands of years. Already in the first century CE, the Chinese used magnetic compasses. About a thousand years later, the compass arrived in Europe and was extensively studied. Independently of the Chinese, who discovered magnetic declination around eight hundred years earlier, the magnetic directions where discovered in Europe in the 16th century. Resting on almost three hundred years of magnetic studies, William Gilbert proposed in his 1600 CE publication *De Magnete*, that the Earth itself is a giant magnet.

Another two hundred years later, Carl Friedrich Gauss was the first to represent the global geomagnetic field mathematically, in form of spherical harmonics (SH). He used sheets of magnetic declination, inclination and intensity, most of which were prepared during ship cruises. Although the inversion methods changed, the SH representation is still the most common form for global geomagnetic field models today.

With the dawn of archaeo- and paleomagnetism in the beginning of the 20th century, findings about the dynamics of the global geomagnetic field where applied to geological questions and among other things contributed to the theory of continental drift and the nature of the Earth's core. The latter is still a major application of modern paleomagnetic models, although nowadays complex geodynamo simulations are considered, and applications also include the study of shielding against cosmic rays, isotope production rates, paleomagnetic dating and investigations of the relation of the paleomagnetic field and paleoclimate.

Modern global geomagnetic field models are built from databases of paleomagnetic observations, most commonly via regularized least squares inversion. Even though sophisticated techniques for accessing information about paleomagnetic directions and intensities from specimen have been developed (see also chapter 2), observations still come in a wide range of quality and often together with large uncertainties, both of the magnetic elements themselves and of the corresponding ages. To represent these uncertainties in models built from least squares inversion, bootstrapping or sampling techniques are employed. In order to represent the uncertainties more accurately, recent models rely on statistical techniques. This chapter will introduce both approaches.

1.1 Brief magnetic field theory

Assuming the origin of the Earth's magnetic field to be in a constrained region of its interior and the mantle to be insulating, the geomagnetic field can be represented as the gradient of a scalar field outside of a sphere of radius R confining the source region.

$$\mathbf{B}(\mathbf{x}, t) = -\boldsymbol{\nabla}\Phi(\mathbf{x}, t) \tag{1.1}$$

This is a direct consequence of Maxwell's equations and Stoke's theorem. The geomagnetic potential fulfills the Laplace equation

$$\Delta\Phi(\mathbf{x}, t) = 0 \tag{1.2}$$

and therefore can be represented in terms of spherical harmonics (SH):

$$\Phi(\mathbf{x}, t) = R \sum_{\ell=1}^{L} \left(\frac{R}{|\mathbf{x}|}\right)^{\ell+1} \sum_{-\ell \le m \le \ell} g_\ell^m(t)\, Y_\ell^m(\hat{\mathbf{x}})\,. \tag{1.3}$$

$\hat{\mathbf{x}}$ is the unit vector $\mathbf{x}/|\mathbf{x}|$ 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. Often the time dependence of the Gauss coefficients $g_\ell^m(t)$ is expressed in form of a B-spline model (Bloxham and Jackson, 1992)

$$g_\ell^m(t) = \sum_{n=1}^{N} g_{\ell,n}^m M_n(t)\,, \tag{1.4}$$

with cubic spline basis functions $M_n(t)$ defined at N knot points t_n. This way, a global geomagnetic field model is specified by the set of $P = N \cdot L \cdot (L+2)$ coefficients $\left\{g_{\ell,n}^m\right\}$. Combining the temporal and spacial basis functions into one function $q_{\ell,n}^m$, the geomagnetic field can be expressed as

$$\mathbf{B}(\mathbf{x}, t) = -\nabla\Phi = \sum_{n,\ell,m} -\nabla q_{\ell,n}^m(\mathbf{x}, t) g_{\ell,n}^m = \mathbf{Q}(\mathbf{x}, t) \cdot \mathbf{m}\,, \tag{1.5}$$

with the matrix $\mathbf{Q}$ relating the model coefficients $\mathbf{m} = \left(g_{\ell,n}^m\right)$ to the magnetic field $\mathbf{B}$ at locations and times $(\mathbf{x}, t)$. For models differing from the B-spline approach, such as the one proposed by Hellio and Gillet (2018), the $q_{\ell,n}^m$ will be different.

Paleomagnetic observations of the magnetic field are given by a functional $H[\mathbf{B}]$, where H is either declination, inclination or intensity (see also equation 4.44).

1.2 Regularized least squares inversion

A prominent way of inferring magnetic field model coefficients $\hat{\mathbf{m}}$ from observations is via regularized least squares inversion (e.g. Jackson, Jonkers, and Walker, 2000). From observations $o(\mathbf{x}, t)$, the coefficients are determined by

$$\hat{\mathbf{m}} = \underset{\mathbf{m}\in\mathbb{R}^P}{\arg\min}\left[\left(o(\mathbf{x},t) - H\left[\mathbf{Q}(\mathbf{x},t)\cdot\mathbf{m}\right]\right)^\top \Sigma_o^{-1}\left(o(\mathbf{x},t) - H\left[\mathbf{Q}(\mathbf{x},t)\cdot\mathbf{m}\right]\right) + \mathbf{m}^\top \mathbf{C}_m^{-1}\mathbf{m}\right]. \quad (1.6)$$

The smoothing matrix $\mathbf{C}_m$ is calculated from two restrictions (Bloxham and Jackson, 1992). The solution is sought to minimize the Ohmic heat dissipation, which gives a spacial norm:

$$\mathbf{m}^\top \mathbf{S}^{-1}\mathbf{m} = \frac{4\pi}{t_e - t_s}\int_{t_s}^{t_e} \mathcal{F}(\mathbf{B})\,dt \quad (1.7)$$

with

$$\mathcal{F}(\mathbf{B}) = \sum_{\ell=1}^{L} \frac{(\ell+1)(2\ell+1)(2\ell+3)}{\ell}\left(\frac{R_{Earth}}{R}\right)^{2\ell+4} \sum_{-\ell\le m\le\ell} \left(g_\ell^m\right)^2$$

The temporal norm gives a smoothness condition for the radial component of the magnetic field B_r:

$$\mathbf{m}^\top \mathbf{T}^{-1}\mathbf{m} = \frac{1}{t_e - t_s}\int_{t_s}^{t_e}\oint_{CMB}\left(\partial_t^2 B_r\right)^2 d\Omega\,dt \quad (1.8)$$

$\mathbf{C}_m$ is then constructed by weighing the two regularizing matrices

$$\mathbf{C}_m = \lambda_S\mathbf{S} + \lambda_T\mathbf{T}. \quad (1.9)$$

One motivation for applying regularization to the least squares scheme is the convergence of the solution to (1.6): The solution is insensitive to the cutoff degrees L and N, if they are chosen appropriately large (Bloxham and Jackson, 1992). The minimizing solution $\hat{\mathbf{m}}$ is then either determined iteratively or via linearizing the observation functional H (c.f. equations (4.48)-(4.50)).

Accurate assessment of uncertainty in the solution has been discussed already in early papers on the regularized least-squares method (Bloxham and Jackson, 1992). However, later works did not pursue analytical approaches and relied on ensemble techniques instead (e.g. Korte, Donadini, and Constable, 2009).

1.3 Statistical models

In order to accurately assess model uncertainties and their characteristics, stochastic inversion techniques for magnetic field data are studied (e.g. Gubbins and Bloxham, 1985; Hellio and Gillet, 2018; Nilsson and Suttie, 2021). Instead of a set of coefficients, probabilistic models are given in form of distributions (or their parameters), most prominently via Bayesian posteriors

$$p(\mathbf{m}|o) \propto p(o|\mathbf{m}) \cdot p(\mathbf{m}) \,.$$

Some methods, for example the ones presented in Chapters 4 and 5, model the magnetic field directly

$$p(\mathbf{B}|o) \propto p(o|\mathbf{B}) \cdot p(\mathbf{B}) \,.$$

Non-linearity of the observation functionals, the amount of data and the complexity of the model itself often lead to the complication, that the distribution is not accessible in closed form. Recent models rely on Markov Chain Monte-Carlo (MCMC) sampling or on analytical approximations, in order to construct a proxy distribution, which is then reported as an equivalent to the solution to (1.6). This thesis is concerned with developing a Gaussian process based model and providing the analytical approximations necessary to access the posterior.

2 Paleomagnetic measurements

This chapter mainly follows Chapter 9 of Tauxe et al. (2018).

Acquiring high quality paleomagnetic data is a challenge on its own (e.g. Tauxe et al., 2018, Section 9.1). To illustrate some of the obstacles and shine light on possible sources of uncertainty in the data, we briefly introduce paleomagnetic measurement methods in this chapter.

Paleomagnetic observations come from a variety of sources, that can roughly be separated into specimen with thermoremanent magnetization, such as volcanic rocks or lava flows and archaeological artifacts like burnt clay, and sediments from marine or lacustrine drill cores. In all cases, a challenge in taking samples lies in the precise determination and preservation of the orientation of the sample. Furthermore, in addition to the paleomagnetic main field that is of interest, samples often carry additional fields. Elaborate techniques exist, in order to remove these remanent fields from the specimen and isolate the main field of interest.

2.1 Directional measurements

Directional information is inferred from specimen by sequentially demagnetizing the probe and measuring the magnetic vector between the demagnetization steps, either via a spinner magnetometer or via a superconducting quantum interference device (SQUID). Via orthogonal projections or (more prominently) via principle component analysis, the strongest component of remanent magnetization is inferred. This is assumed to be the direction of the paleomagnetic field direction. Often multiple specimen are taken from the same site and Fisher-statistics are applied to several directions, in order to determine declination and inclination from the location parameter. The concentration parameter is reported in order to quantify the uncertainty of the analysis.

2.2 Intensity measurements

Inferring the intensity of the paleomagnetic field from specimen is more complicated than the directions, as the field imprinted on the specimen points in the same direction, but the information about the intensity is only indirectly preserved. While other methods exists (Tauxe and Yamazaki, 2015), we focus here on the commonly used method introduced by Thellier and Thellier (1959) and later refined by others (e.g. Nagata, Arai, and Momose, 1963; Coe, 1967). The idea is that the remament magnetization is proportional to the intensity of the paleomagnetic field, and that an induced magnetization will share the same proportionality to the inducing field. The measurement procedure is then as follows

1. The specimen is heated to temperature T and left to cool down in zero field.
2. The magnetization M_N is measured. In this step, the specimen only contains a (decreasing) part of the remanent magnetization with intensity F_{paleo}.
3. The specimen is heated to temperature T again and left to cool down in a reference field of intensity F_{ref}.
4. The magnetization M_N is measured. In this step, the specimen contains the remanent magnetization as well as magnetization due to the reference field.
5. Increase T and repeat until the remanent magnetization left in the probe is zero (resp. the Curie-temperature is reached).

From the two measurements in each cycle of heating, the ratio of remanent to imprinted magnetization can be measured. The key assumption that the mechanism in which the reference field is imprinted is similar to the one that magnetized the specimen in the first place leads to a linear relationship between the two and the constant of proportionality relates to the ratio of paleomagnetic and reference field F_{paleo}/F_{ref} (Nagata, Arai, and Momose, 1963). This procedure is also repeated for multiple specimen from the same site and the sample set mean and standard deviation are reported.

2.3 Dating

A final relevant information for paleomagnetic data is the age of the specimen. To obtain this information, either radiometric or archaeological methods are employed. In radiometric methods, decaying elements (often ^{14}C or ^{40}K) are studied. By analyzing the decay the age can either be estimated directly or the amount of radioactive material

can be compared to a reference model (in case of ^{14}C, this is a model for the carbon distribution in the atmosphere) and the age estimated from the comparison. Archaeological methods rely on secondary knowledge about the specimen, e.g. about the site on which the specimen was acquired. The age is then estimated from archaeological knowledge about the site. Both methods rely on the precision on reference knowledge and can lead to large uncertainties in the age determination. In case of radiometric dating, modern techniques exist to estimate the distribution of the age, while in archaeological dating often only upper and lower bounds for the age are available.

3 Mathematical framework

The publications in Part II are addressed at the geomagnetism community and therefore do not focus on the mathematical details of the material that is presented in this thesis. To work out some of these, and especially to highlight differences to existing modeling methods, this chapter concentrates on the framework into which the developed method is embedded.

3.1 Gaussian processes

Gaussian process

Definition 3.1 A continuous stochastic process $f(\mathrm{x}), \mathrm{x} \in \mathrm{X}$ is called **Gaussian process (GP)**, if and only if for every finite set $\mathrm{x}_1, \ldots \mathrm{x}_k$ in X, the vector

$$\mathbf{f} = \begin{pmatrix} \mathrm{f}_1 \\ \vdots \\ \mathrm{f}_k \end{pmatrix} = \begin{pmatrix} f(\mathrm{x}_1) \\ \vdots \\ f(\mathrm{x}_k) \end{pmatrix}$$

is a multivariate Gaussian random variable.

In Chapter 1, we outlined different approaches to modeling the Earth's magnetic field (EMF). The modeling method we adopt and extend in Part II is based on Gaussian processes and falls into the class of Bayesian statistical models (c.f. Section 1.3). A key part is thus the definition of a prior distribution and a likelihood function, which together determine the posterior distribution that represents the model. We construct the prior as a Gaussian process (GP), while the non-linear likelihood is fixed by the observation functional. A GP prior is defined on a space of functions. We will consider the functional analytic implications of this in the next section. If the likelihood is Gaussian, the posterior is a GP as well. If not, the posterior is often not accessible directly and approximation techniques have to be employed.

A GP can be considered the infinite dimensional generalization of a multivariate normal distribution. Similar to the finite dimensional case, a GP is fully characterized by a mean and a covariance function (e.g. Rasmussen and Williams, 2006). In the presented studies, the specific choice of these functions is motivated by existing (outside) knowledge about the EMF (details can be found Chapter 4, as well as Sections 5.2.1 and Sections 6.2.3). In order to illustrate some properties of GPs, we consider a one-dimensional example in the following.

3.1.1 Gaussian process regression in one dimension

Consider as a prior the GP f with mean function $\bar{\mathrm{f}}$ and covariance function K

$$f \sim \mathcal{GP}\left(\bar{\mathrm{f}},\, \mathrm{K}\right) \,. \tag{3.10}$$

Given linear observations

$$\mathrm{o}(\mathrm{y}) = f(\mathrm{y}) + \varepsilon \,, \tag{3.11}$$

with $\varepsilon \sim \mathcal{N}\left(0,\, \Sigma\right)$, mean and covariance of the posterior GP are given by (e.g. Rasmussen and Williams, 2006, Equations (2.23) and (2.34))

$$\mathbb{E}\left[f(\mathrm{x})|\, \mathrm{o}\right] = \bar{\mathrm{f}}(\mathrm{x}) + \mathrm{K}(\mathrm{x},\mathrm{y})\left(\mathrm{K}(\mathrm{y},\mathrm{y}) + \Sigma\right)^{-1}\left(\mathrm{o} - \bar{\mathrm{f}}(\mathrm{y})\right) \tag{3.12}$$

$$\mathrm{Cov}\left[f(\mathrm{x}), f(\mathrm{x}')|\, \mathrm{o}\right] = \mathrm{K}(\mathrm{x},\mathrm{x}') - \mathrm{K}(\mathrm{x},\mathrm{y})\left(\mathrm{K}(\mathrm{y},\mathrm{y}) + \Sigma\right)^{-1}\mathrm{K}(\mathrm{y},\mathrm{x}') \tag{3.13}$$

Evidently, the posterior covariance does not depend on the observations o but only on the locations of observation y and the error covariance Σ. Figure 3.1 illustrates one dimensional regression with a squared exponential kernel. In the vicinity of the data, the posterior variance is reduced wrt. the prior (left side of the right panel). Away from the data, the prior is reproduced (right side of the right panel).

3.2 Reproducing kernel Hilbert spaces

In the previous section we mentioned that a Gaussian process prior is defined on a space of functions. In this section we investigate this space. The right toolbox is provided by the theory of reproducing kernel Hilbert spaces (e.g. Berlinet and Thomas-Agnan, 2004).

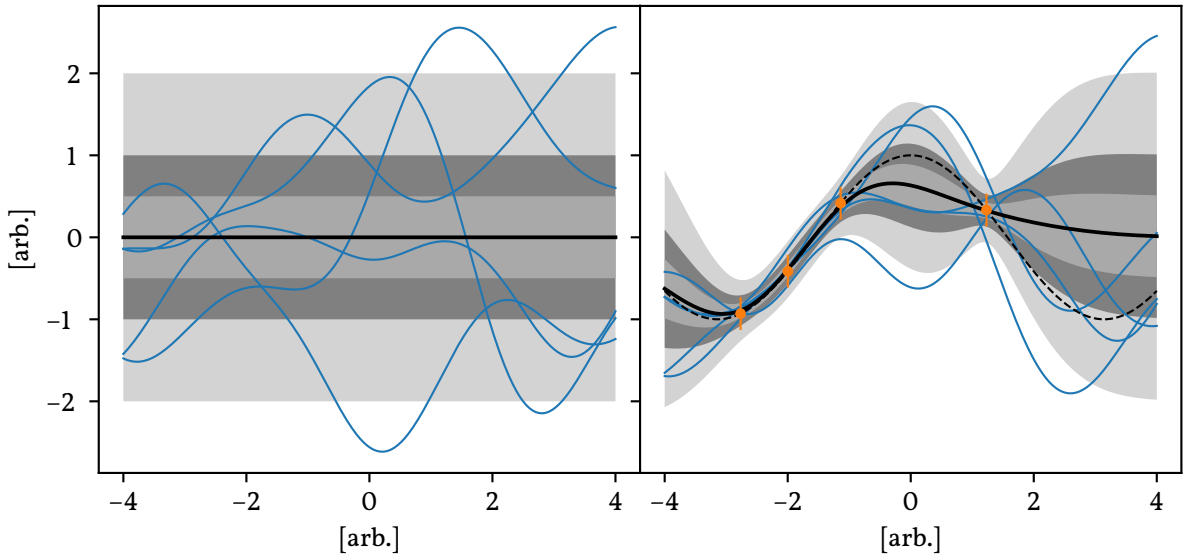


Figure 3.1: One dimensional GP regression with a squared exponential kernel. The left panel shows the prior mean as a thick black line, the shaded areas cover 0.5, 1 and 2 standard deviations respectively. Blue lines show samples drawn from the prior. The right panel shows artificial data in orange, together with the generating function as a black dashed line. The solid black line and shaded area are as on the left, blue lines show samples from the posterior process.

Reproducing kernel Hilbert space

Definition 3.2 Let $\mathcal{H}$ be a real Hilbert space of real valued functions on some set X. The **evaluation functional** over $\mathcal{H}$ is a linear functional

$$\begin{aligned} L_x : \mathcal{H} &\to \mathbb{R} \\ f &\mapsto f(x) \end{aligned}$$

$\mathcal{H}$ is called a **reproducing kernel Hilbert space (RKHS)**, if for all $x \in X$, L_x is continuous (or equivalently, bounded).

From Riesz' Lemma it follows, that if $\mathcal{H}$ is an RKHS,

$$\forall x \, \exists K_x \in \mathcal{H} : \quad f(x) = L_x(f) = \langle K_x, f \rangle_{\mathcal{H}} \quad \forall f \in \mathcal{H} \, .$$

K_x is said to have the *reproducing property*.

Reproducing kernel

Definition 3.3 The function K

$$\begin{aligned} K : X \times X &\to \mathbb{R} \\ K(x, y) &\mapsto L_x(K_y) = \langle K_x, K_y \rangle_{\mathcal{H}} \end{aligned}$$

is called **reproducing kernel of** $\mathcal{H}$.

The Moore-Aronszjain theorem (e.g. Theorem 3 in Berlinet and Thomas-Agnan, 2004) states, that every symmetric, positive definite kernel on a set X has an associated, unique RKHS for which it is the reproducing kernel. Every Gaussian process is uniquely defined by a mean function $\bar{f}$ and a positive definite, symmetric covariance function K. Thus every GP has a RKHS associated to it, namely the RKHS with reproducing kernel K. In fact, this RKHS is isometrically isomorphic to the Hilbert space spanned by the random variables $f(t)$ with $\mathrm{Cov}\left[f(t), f(s)\right] = K(t, s)$ (Wahba, 1990, Section 1.4). The random variables $f(t)$ can be one-to-one related to the kernel functions $K(\cdot, t)$. From this it is clear that the RKHS is spanned by the kernel functions and thus every element in it is a (possibly infinite) linear combination of kernel functions (Wahba, 1990, Section 1.4). Therefore the elements in the RKHS share the smoothness properties of the reproducing kernel. The imprinting of kernel smoothness onto the functions in the RKHS can also be understood from looking at the scalar product in the RKHS, if the kernel is known and fulfills some mild conditions (see also Berlinet and Thomas-Agnan, 2004, Section 3.3). Let $\mathcal{H}$ be a RKHS with kernel K and $\int K(x, x)\, d\mu(x) < \infty$ for some finite measure μ. Then all elements in $\mathcal{H}$ are in $L^2(\mu)$. From Mercer's theorem (e.g. Rasmussen and Williams, 2006, Theorem 4.2) it follows, that K can be expanded into a sequence of orthonormal eigenfunctions φ_i with corresponding eigenvalues λ_i

$$K(x, y) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(x) \varphi_i(y)\,. \tag{3.14}$$

Further, K has an associated integral operator T

$$T : \mathcal{H} \to L^2(\mu) \qquad T f(x) = \int K(x, y) f(y)\, d\mu(y) \tag{3.15}$$

with square root (e.g Berlinet and Thomas-Agnan, 2004, Section 3.3)

$$T^{1/2} : \mathcal{H} \to L^2(\mu) \qquad T^{1/2} f = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \int \varphi_i(y) f(y)\, d\mu(y)\, \varphi_i\,. \tag{3.16}$$

With $f_i = \int f(x)\varphi_i(x)\,d\mu(x)$ we have (Wahba, 1990, Lemma 1.1.1): f is an element of $\mathcal{H}$ if and only if

$$\sum_{i=1}^{\infty} \frac{f_i^2}{\lambda_i} < \infty \tag{3.17}$$

On the other hand, we have

$$\sum_{i=1}^{\infty} \frac{f_i^2}{\lambda_i} = \sum_{i=1}^{\infty} \left(\frac{\int f(x)\varphi_i(x)\, d\mu(x)}{\sqrt{\lambda_i}} \right)^2 = \left\| T^{-1/2} f \right\|_{L^2}^2, \tag{3.18}$$

where $T^{-1/2}$ is the pseudo inverse of $T^{1/2}$. From (3.18), the smoothness condition can be formulated as: every function in $\mathcal{H}$ is square integrable over μ against the pseudo inverse of the kernel's root. Similar to (3.18), the scalar product in the RKHS $\mathcal{H}$ can be related to the scalar product in $L^2(\mu)$ via

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{\infty} \frac{f_i g_i}{\lambda_i} = < T^{-1/2} f, T^{-1/2} g >_{L^2} . \tag{3.19}$$

Figure 3.2 shows elements from three different RKHS to illustrate different smoothness properties. In GP regression the covariance function of the prior defines the properties of possible models and the posterior mean lies in the RKHS, as can be seen from Equation (3.12) (see also the representer theorem, e.g. Rasmussen and Williams, 2006, Equation (2.27)).

3.2.1 RKHS and regularized least squares inversion

Many existing models of the EMF are constructed via regularized least squares inversion (see also Section 1.2). In the context of RKHS, the regularization conditions (1.7) and (1.8) define a norm, that clearly satisfies the parallelogram law and thus can be used to define an inner product via the polarization identity. (1.6) can then be alternatively formulated as a minimization problem over the associated RKHS:

$$\hat{\mathbf{m}}(t) = \underset{\mathbf{m}(t)\in\mathcal{H}}{\arg\min} \left[\left(o(\mathbf{x}, t) - H\left[\mathbf{Q}(\mathbf{x}) \cdot \mathbf{m}(t)\right] \right)^{\top} \Sigma_o^{-1} \left(o(\mathbf{x}, t) - H\left[\mathbf{Q}(\mathbf{x}) \cdot \mathbf{m}(t)\right] \right) + \|\mathbf{m}\|_{\mathcal{H}}^2 \right] \tag{3.20}$$

Since the regularization conditions penalize derivatives, the solution consists of piecewise polynomials (Murphy, 2012, Section 15.4.6). For computational reasons, Bloxham and Jackson (1992) resort to a finite basis of B-splines to approximate this function. This finite basis also spans an RKHS, where the kernel can straight-forwardly be constructed

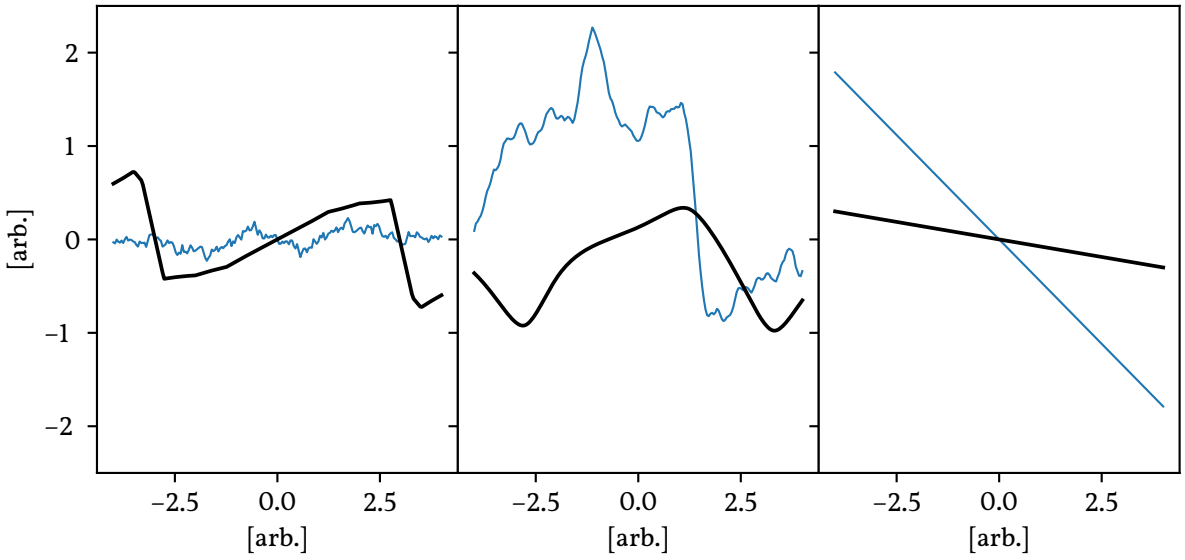


Figure 3.2: Elements (black) of different RKHS and samples (blue) from the associated zero mean GP. The elements are given by linear combinations of the kernel functions and thus share the kernels smoothness properties.

left Ornstein-Uhlenbeck kernel
$K(x, x') = \sigma^2/(2\theta)\big(\exp\big(-\theta|x - x'|\big) - \exp\big(-\theta|x + x'|\big)\big)$

center Matérn-3/2 kernel
$K(x, x') = (1 + \sqrt{3}|x - x'|/\ell)\exp\Big(-\sqrt{3}|x - x'|/\ell\Big)$

right scalar-product kernel
$K(x, x') = x \cdot x'$

from a combination of orthonormal basis functions. The inversion scheme proposed in Part II differs from regularized least squares in geomagnetism. However, there are some parallels. Instead of Equations (1.7) and (1.8), the regularization in GP regression is given by the choice of a particular kernel, which then defines a norm via the associated scalar product. The posterior mean, given by Equation (3.12), is then the solution to the equivalent of Equation (3.20) (Kanagawa et al., 2018, Section 3.3).

3.3 Model representation and uncertainty estimates

Representer Theorem

Theorem 3.1 Any solution to (the equivalent of) Equation (3.20) admits a representation of the form

$$\mathbb{E}\big[f(x)|\, o\big] = \bar{f}(x) + \sum_{k=1}^{n} \alpha_k K(x, y_k)\,. \tag{3.21}$$

Models in geomagnetism are traditionally given by B-spline approximations to the solution of Equation (1.6) and are parametrized by a fixed number of coefficients (see Section 1.1). In contrast, the true solution is an element of the corresponding RKHS and is parametrized by one coefficient per observation, as evident from the representer theorem (and Equation (3.12)). The result of GP regression is the posterior *distribution*. The result of the least squares inversion can be identified with the maximum a posteriori probability estimator (MAP) of the posterior (a function as well). This view further relates both approaches. A central reason for employing statistical modeling techniques is the quantification of uncertainties. With the posterior distribution, a notion of model uncertainties is naturally at hand and with the posterior being a GP, the posterior covariance is directly accessible (Equation (3.13)). Reporting the model in terms of a distribution may be interpreted as stating that all realizations of the GP are possible descriptions of reality. Which one is the *ground truth* can not be inferred, only statements about model probabilities are possible. With a GP, the mean (3.13) has the highest posterior density and is often considered when features and evolution of the model are discussed.

The posterior mean is an element of the RKHS associated to the GP, while every realization almost surely is not (Kanagawa et al., 2018, Section 4). Samples from the GP are less smooth than elements of the RKHS, as evident from Figure 3.2. This can be intuitively understood from the sample generation procedure: The Cholesky decomposition of the kernel matrix is used, which heuristically can be seen as the *root* of the kernel (GP sampling algorithms are prominent in the literature, e.g. Rasmussen and Williams, 2006, Appendix A.3). See Kanagawa et al. (2018) for further analysis of the Gaussian process sample space. The posterior standard deviation, that is often reported along with the mean to capture uncertainties, gives a measure of the spread of samples and can be used to estimate the regime in which possible realities most probably lie.

3.4 Estimating Gaussian process kernel parameters

In a linear setting, the predictive performance of GPs depends exclusively on the suitability of the kernel (Murphy, 2012, Section 15.2.3). The general form of the kernels we employ in Part II is motivated by (external) knowledge about the EMF. However, the exact form still depends on a number of parameters. These parameters influence the kernel, and therefore its suitability for modeling the archeomagnetic field. Choosing a set of parameters will also fix the RKHS of possible solutions. One way of fixing these parameters is to use existing models from other datasets (e.g. Hellio and Gillet, 2018; Nilsson

and Suttie, 2021). A different way is to estimate the kernel parameters from the data via maximum marginal likelihood estimation (e.g. Rasmussen and Williams, 2006, Section 5.4). We follow this approach in Part II and illustrate the procedure for a one dimensional example in the following.

Consider the posterior of a GP f with kernel K_θ, that depends on a (set of) parameters θ:

$$p(f|o;\theta) = \frac{p(o|f;\theta)\cdot p(f;\theta)}{p(o|\theta)} \tag{3.22}$$

Equation (3.22) gives the posterior distribution for f, given observations o and kernel parameters θ. Instead, we are interested in a measure for how appropriate a set of parameters is for the data o, without focusing on the distribution of f itself. This is provided by the denominator $p(o|\theta)$, where the functions f have been *marginalized* (integrated out). For this reason $p(o|\theta)$ is called the *marginal likelihood*. In the Bayesian hierarchy, the marginal likelihood is one level above the posterior distribution. It discriminates between different sets of parameters that in term define different kernels (and thus different RKHSs). Due to the higher level, the kernel parameters are also referred to as *hyperparameters*. For the linear GP from Section 3.1.1, the logarithm of the marginal likelihood is given by (e.g. Rasmussen and Williams, 2006, Equation (2.30))

$$\ln p(o|\theta) = -\frac{1}{2}\left(o - \bar{f}(y)\right)^\top \left(K_\theta(y,y) + \Sigma\right)^{-1}\left(o - \bar{f}(y)\right) - \frac{1}{2}\ln\left|K_\theta(y,y) + \Sigma\right| + \text{const.}\,. \tag{3.23}$$

The first term describes the goodness of fit to the data while the second term (involving the determinant of the kernel) is a measure of model complexity. The maximum marginal likelihood estimate balances the two terms. This is illustrated for the length scale estimation of a squared exponential kernel in Figure 3.3.

Depending on the form of the kernels K_θ, the parameter(s) θ that maximize (3.23) may be available analytically or have to be determined via numerical methods. Since every evaluation of $\ln p(o|\theta)$ involves inverting the kernel matrix, the latter will become computationally impossible when the number of records grows large. This is especially true if multiple parameters are optimized and the marginal likelihood contains many local optima.

3.5 Linear approximation for non-Gaussian likelihoods

In the previous sections we illustrated GP regression on the example of one dimensional linear data. In the natural sciences however, observations are often non-linear and multi dimensional. The one dimensional regression straight forwardly generalizes to vector data, however, if non-linear observations are addressed, the likelihood will no longer be

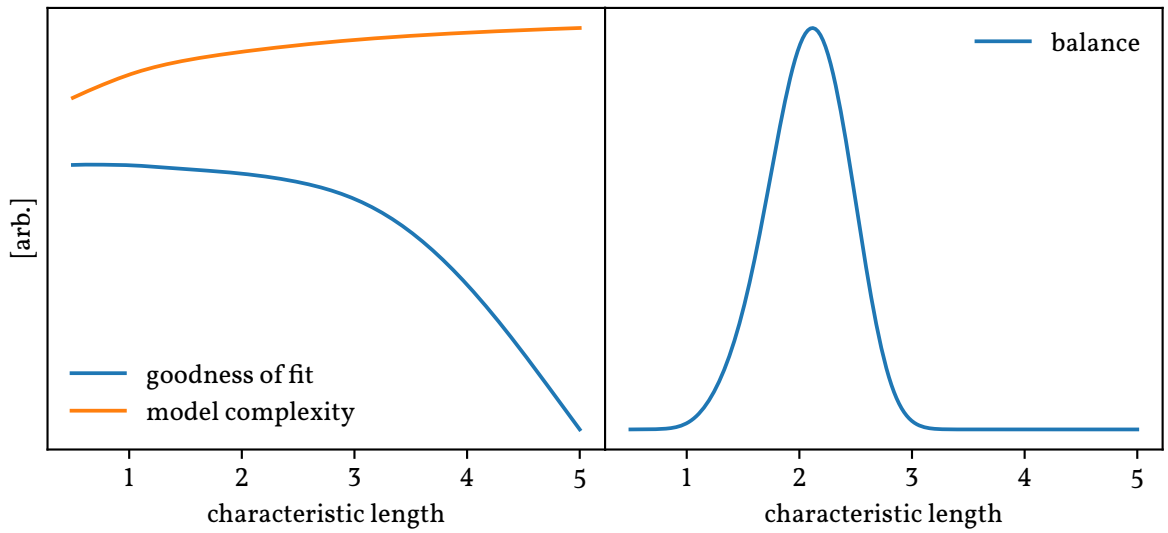


Figure 3.3: Estimation of the characteristic length scale for a squared exponential kernel in one dimension. The left panel shows the two terms in the log marginal likelihood individually. As the maximum of the marginal likelihood is sought, low values are penalized. The goodness of fit increases with smaller length, while model complexity grows. With bigger length scales, the model complexity decreases while the goodness of fit declines. The right panel shows the marginal likelihood, which balances both terms. The artificial data was generated from a kernel with characteristic length of 2.3.

Gaussian and the simple equations (3.12) and (3.13) break down. Sophisticated techniques exist to approximate the posterior distribution in this case. On the probabilistic side, sampling strategies such as MCMC are employed (e.g. Titsias, Rattray, and Lawrence, 2011; Hellio et al., 2014; Nilsson and Suttie, 2021). Deterministic approaches mostly rely on finding a Gaussian approximation to the likelihood analytically (e.g. Opper and Archambeau, 2009). More advanced strategies combine multiple Gaussian approximations (e.g. Rue, Martino, and Chopin, 2009; Minka, 2013). In Part II we implement a linear approximation to the non-linear observation functionals. This has the advantage of low computational overhead and good scaling properties (in terms of observations). The downside is that the linearized distribution depends on the point of expansion (see also Figure 4.4 below) and may strongly deviate from the actual one, especially if the variance is large. The main reason why we resort to a linearization based approach is computational cost. As argued in the previous section, finding the optimal set of hyperparameters based on marginal likelihood maximization requires performing many inversions and therefore renders more complex approaches like MCMC unfeasible.

3.6 Numerical aspects of Gaussian process regression

The main numerical workload in GP regression comes from inverting the covariance, which requires $\mathcal{O}(n^3)$ operations. For stability reasons, one often resorts to calculating the Cholesky decomposition first and then inverting the matrix by solving a triangular linear system (e.g. Rasmussen and Williams, 2006, Algorithm 2.1). Besides computational costs, the memory required to store the covariance matrix is $\mathcal{O}(n^2)$. Both cost in computational power and memory are the reason that standard GP regression does not scale well for large datasets. To tackle this, various approaches exist. These include sparse approximations (e.g. Smola and Bartlett, 2000) and the implementation of highly parallelized, approximate Cholesky factorizations (Gardner et al., 2018). In Chapter 6, we implement a Kalman-filter (Kalman, 1960) and smoother to tackle the memory demand. The downside is that the posterior covariance is not fully available, as cross correlations are not calculated explicitly. Therefore, posterior samples are not easily available. They can be generated from the forward model by requiring the ensemble to share the same statistics as the smoothed model (Baerenzung and Holschneider, in preparation).

Part II

Publications

This chapter is a transcript of Mauerberger et al. (2020)

Correlation based snapshot models of the archeomagnetic field

S. Mauerberger[1] M. Schanner[2] M. Korte[2] M. Holschneider[1]

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

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

Submitted to Geophysical Journal International on February 4, 2020; accepted on July 4, 2020.

Abstract

For the time stationary global geomagnetic field, a new modelling concept is presented. A Bayesian non-parametric approach provides realistic location dependent uncertainty estimates. Modelling related variabilities are dealt with systematically by making little subjective a priori assumptions. Rather than parameterizing the model by Gauss coefficients, a functional analytic approach is applied. The geomagnetic potential is assumed a Gaussian process to describe a distribution over functions. A priori correlations are given by an explicit kernel function with non-informative dipole contribution. A refined modelling strategy is proposed that accommodates non-linearities of archeomagnetic observables: First, a rough field estimate is obtained considering only sites that provide full field vector records. Subsequently, this estimate supports the linearisation that incorporates the remaining incomplete records. The comparison of results for the archeomagnetic field over the past 1000 years is in general agreement with previous models while improved model uncertainty estimates are provided.

4.1 Introduction

Global geomagnetic field reconstructions of the past millennia are useful to investigate the geodynamo process or the complex interaction of the field with solar wind particles and cosmic rays, and they find application in archeomagnetic and paleomagnetic dating. Reconstructions are typically built from volcanic and archeomagnetic samples collected at the Earth's surface providing records of the ancient EMF. Unfortunately, on a global scale records are clustered, unevenly distributed towards the Western Eurasian region and corrupted by various uncertaintes. This considerably complicates the reconstruction of the ancient EMF.

Dating back to 1985, Gubbins and Bloxham were amongst the first to propose a Bayesian inference for modelling the EMF, already discussing non-linear observables and model uncertainties. Their parametrized implementation of a truncated SH representation with norm optimization has become a widely used modelling scheme. The majority of historical and archeomagnetic field models published over the past years essentially rely upon the same inverse strategy. Early models such as Jackson, Jonkers, and Walker (2000), Constable, Johnson, and Lund (2000), and Korte and Constable (2003) provide estimates without quantifying uncertainties. More recent attempts – e.g. Korte, Donadini, and Constable (2009), Licht et al. (2013), Hellio and Gillet (2018), and Senftleben (2019) – describe variabilities by deriving ensembles of equivalent solutions. Roughly speaking, those models differ in two aspects: On the one hand, the error handling, data selection and outlier detection have been refined over the years (Licht et al., 2013). On the other hand, different strategies are used to incorporate a priori knowledge. Early models are typically starting off from an axial dipole and are regularized by a physically motivated norm. Since regularized field models are known to underestimate uncertainties at small length scales (Gillet, 2019), more elaborate modelling concepts are under investigation. Recent attempts deduce a priory information including temporal dynamics from the statistics of satellite era models (e.g. Hellio and Gillet, 2018) or from geodynamo simulations (e.g. Sanchez et al., 2016). Existing models, however, have in common that uncertainties related to modelling, in particular due to model parameters and the uneven data distribution, are not dealt with systematically.

The present paper introduces an advanced concept to model snapshots of the EMF. This work should be considered as a fist step towards a new inverse strategy in which the notion of modelling related uncertainties is well defined. Therefore, we adapt the correlation based inversion developed by Holschneider et al. (2016) that is known from modelling observatory and satellite data. Several modifications are required to adjust the concept to archeo- and paleomagnetic data.

We pursue a fully Bayesian approach that determines the EMF's posterior distribution which simultaneously encodes the most probable field model and its uncertainties. To obtain the posterior distribution we use a functional analytic approach where inference takes place directly in the space of functions. Observables and quantities of interest are expressed in terms of functionals that act on the geomagnetic potential. Rather than using a model that is parametrized by a finite SH basis, the geomagnetic potential is assumed a GP. The GP in use is non-parametric in the sense that it is a distribution over functions and is specified by a two-point covariance function.

From a parametric point of view, GPs have been used for a long time in modelling the EMF (e.g. Bouligand et al., 2005; Khokhlov, Hulot, and Bouligand, 2006), known under the term Giant Gaussian Process (GGP). That term was coined by Constable and Parker (1988) who proposed a GP based model focusing on the estimation of model parameters. Our approach may be seen as the functional analytic extension of the GGP model.

Our a priori distribution of the geomagnetic potential is characterized by its mean power spectral behaviour, which is represented by an explicit correlation function that takes all SH degrees into account. If a SH truncation was desired, transdimensional modelling (Livermore et al., 2018) may be applied to also infer the cut-off degree. Using an explicit kernel function – not truncated at a certain SH degree – circumvents that problem. This does not necessarily mean that our approach reaches a higher resolution at a global scale but, improves treatment of the uneven data coverage and allows the exploitation of the records to their fullest. In addition, a low SH degree truncation may lead to spurious oscillations and ringings if the data include pronounced local anomalies. We try to be the least subjective and specify the a priori field model using uninformative distributions, when possible. Our a priori model depends only on a single parameter that controls the a priori power spectral behaviour.

In the case of satellite and observatory data, EMF full vector components are observed directly and observables are linearly related to the geomagnetic potential. Thus the posterior distribution for the GP is explicit and may be computed using ordinary linear algebra. Archeomagnetic data, however, call for a refined modelling strategy that takes the non-linearity of declination, inclination and intensity into account. The majority of sites only have incomplete vector information so that a direct linearisation of each record is not possible. Therefore, we propose a two step Bayesian update system: First, a rough field estimate is obtained considering only sites that provide complete field vector records. Subsequently, this estimate supports the linearisation that incorporates the remaining incomplete records.

To demonstrate the potential of our modelling strategy we present a case study using archeomagnetic and volcanic data of the past 1000 years. Joint maps of best prediction and point-wise uncertainty are presented, which allow an improved interpretation of the spatial field structure. Although our modelling is not based upon a SH basis, we predict Gauss coefficients and quantify their uncertainties. In addition, we calculate the posterior mean of the spatial power spectrum and estimate error bounds. Finally we present the posterior probability density function (PDF) for the dipole strength and the location of the geomagnetic north pole. The latter results are obtained for records of the past millennium, coarsely sorted into 100 year bins and arranged into a discrete time-series.

The structure of the document is as follows: Section 4.2 gives an overview of the modelling theory, which closely follows Holschneider et al. (2016). First, we introduce the general construction of our dipole and non-dipole priors and correlations kernels (Section 4.2.1). We describe the link between observations and model, and how to obtain the posterior distribution from linear observations (Section 4.2.2), and then discuss the need for linearisation of archeomagnetic observables (Section 4.2.3). The treatment of data uncertainties is laid out in Section 4.2.4. The following two Sections, 4.3 and 4.4, give details about the necessary adjustments to model archeo- and paleomagnetic data. Section 4.3 focuses on the two step strategy used to handle the non-linearities, formulated as a Bayesian update system and includes synthetic tests. Section 4.4 covers the translation of a priori uncertainties in the model parameters to the posterior. Finally, Section 4.5 provides the case study. We close the document by drawing conclusions and showing future perspectives in Section 4.6.

4.2 Modelling Concept

In this section we review the non-parametric and correlation-based modelling strategy (Holschneider et al., 2016) that underlies our Bayesian approach for modelling time-stationary snapshots of the EMF. We lay out our field model together with our a priori assumptions and establish the nomenclature that we adopt throughout this paper. We point out difficulties arising when working with non-linear observables such as archeo- and paleomagnetic records and discuss the general treatment of data uncertainties.

4.2.1 Magnetic Field Model

When modelling the geomagnetic core field on archeo- to paleomagnetic time scales, further contributions to the geomagnetic field – the crust, ionosphere and magnetosphere – are neglected since archeo- and paleomagnetic measurement errors are assumed to be significantly larger than the respective field contributions (Constable and Korte, 2015).

Close to the surface – far from magnetic sources – the EMF, $\mathbf{B}$, can be approximated by the gradient of a scalar potential satisfying Laplace's equation

$$\mathbf{B} = -\nabla\Phi \quad , \qquad \nabla^2\Phi = 0 \, , \tag{4.24}$$

where Φ is referred to as the geomagnetic potential (Backus, Parker, and Constable, 1996, Chp. 4). In contrast to the magnetic field, the potential is not directly observable. For an internal source and with respect to some reference sphere of radius R, the potential Φ at location $\mathbf{x}$, $|\mathbf{x}| > \mathrm{R}$, can be expanded in SHs

$$\Phi(\mathbf{x}) = \mathrm{R} \sum_{\ell} \left(\frac{\mathrm{R}}{|\mathbf{x}|}\right)^{\ell+1} \sum_{-\ell \le \mathrm{m} \le \ell} \mathrm{g}_\ell^\mathrm{m} \, \mathrm{Y}_\ell^\mathrm{m}(\hat{\mathbf{x}}) \tag{4.25}$$

where $\mathrm{Y}_\ell^\mathrm{m}$ refers to the real valued and Schmidt semi-normalized SH of degree l and order m with related Gauss coefficient $\mathrm{g}_\ell^\mathrm{m}$. The dependence of $\mathrm{g}_\ell^\mathrm{m}$ on a reference radius R is not explicitly typed.

We use a spherical coordinate system with $\mathrm{B_N}$ pointing to geographic north, $\mathrm{B_E}$ to the east and $\mathrm{B_Z}$ vertically downward. The components of the magnetic field vector $\mathbf{B}$ at a location of radius r, co-latitude θ and longitude φ, are

$$\mathrm{B_N} = -\frac{1}{\mathrm{r}} \frac{\partial \Phi}{\partial \theta} \, , \qquad \mathrm{B_E} = \frac{1}{\mathrm{r} \sin(\theta)} \frac{\partial \Phi}{\partial \varphi} \, , \qquad \mathrm{B_Z} = -\frac{\partial \Phi}{\partial \mathrm{r}} \, . \tag{4.26}$$

The ellipticity of the Earth is neglected and geocentric coordinates are treated as if they were geodetic.

Motivated by magnetic field theory, we express our lack of knowledge by making a priori assumptions about all Gauss coefficients. Considering Holocene time-scales, the dipole part of the field does not have the same statistical distribution as the non-dipole part (Constable and Parker, 1988). Thus, our model of choice is dipole dominated with additional random field contributions.

The dipole part is specified by the Gauss coefficients of degree $\ell = 1$. A priori the coefficients are assumed normal

$$\mathbf{g}_1 \sim \mathcal{N}(\bar{\mathbf{g}}_1, \, \Sigma_1) \, , \tag{4.27}$$

with mean vector $\bar{\mathbf{g}}_1$ and covariance matrix Σ_1, nine model parameters to be determined (three for the mean and six for the covariance). The subscript is hinting at the SH degree and is going to be generalized. The dipole potential is a GP with mean and covariance function

$$\mathbb{E}\left[\Phi_{\mathrm{DP}}\right] = \mathbf{Y}_1^\top \mathbf{g}_1 \,, \qquad \mathbb{V}\left[\Phi_{\mathrm{DP}}\right] = \mathbf{Y}_1^\top \Sigma_1 \mathbf{Y}_1 \tag{4.28}$$

where $\mathbf{Y}_1$ refers to the SH basis of degree $\ell = 1$ i.e.

$$\mathbf{Y}_1(\mathbf{x})^{\mathrm{T}} = \mathrm{R}\left(Y_1^0(\mathbf{x}), Y_1^1(\mathbf{x}), Y_1^{-1}(\mathbf{x})\right)\left(\frac{\mathrm{R}}{|\mathbf{x}|}\right)^2 \,. \tag{4.29}$$

Section 4.4.1 deals with choosing a priori dipole parameters. The correlation pattern for independent and identically distributed (IID) Dipole coefficients is shown in the left panel of Figure 4.1.

Our model of the non-dipole part is similar to the one proposed by Constable and Parker (1988) but within a Bayesian setting. The non-dipole potential is assumed a GP of zero mean. Rather than truncating the SH basis at a certain degree, a covariance function of closed form is used. Therefore, we adopt the kernel construction method from Holschneider et al. (2016, Sec. 4). With respect to a reference sphere, the potential is characterized by its mean power spectral behaviour (see Sec. 4.5.6). That reference sphere may be seen as a virtual source region and the reference radius has no particular physical significance (Constable and Parker, 1988). Except for the dipole, at the reference radius Gauss coefficients are assumed IID normal random variables. The according Legendre type kernel is of the form

$$\mathrm{K_L}(\mathbf{x}, \mathbf{y}) = \lambda^2 \mathrm{R}^2 \sum_{\ell=2}^{\infty} \left(\frac{\mathrm{R}^2}{|\mathbf{x}||\mathbf{y}|}\right)^{\ell+1} \sum_{-l \le m \le l} Y_\ell^m(\hat{\mathbf{x}}) Y_\ell^m(\hat{\mathbf{y}}) \tag{4.30}$$

where we introduced λ, a scaling factor that controls the amount of the non-dipole contribution and its dimension. Dipole and non-dipole parts are assumed statistically independent. According to Holschneider et al. (2016, Eq. 54), $\mathrm{K_L}$ can be expressed in closed form. We choose this kernel as it is computationally simple and depends only on two parameters, R and λ. The right panel of Figure 4.1 depicts an a priori correlation pattern for the non-dipole potential.

In principle an SH decomposition is possible but computationally limited. To cover highly localized modelling errors, the covariance has to map the characteristic length scales present in the data. Small scale correlations among records are a valuable source of information that should enter the model. In case of an expansion, the highest SH de-

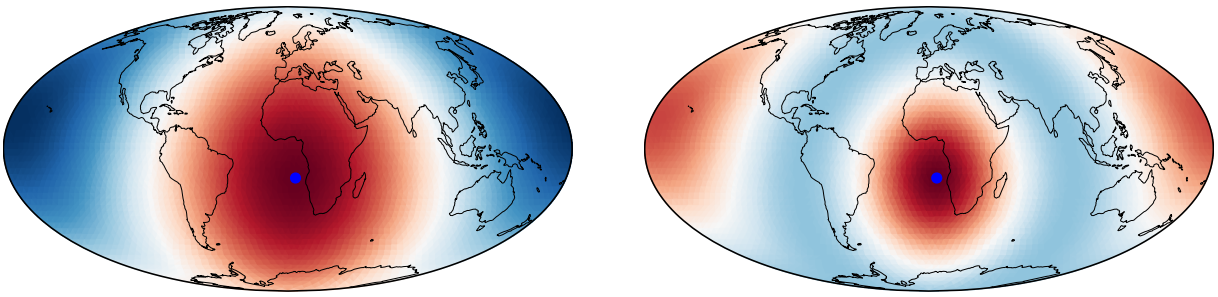


Figure 4.1: Visualization of dipole (left) and non-dipole (right) correlation structures at the Earth's surface. The reference location is indicated by the blue dot. The dipole pattern corresponds to IID dipole coefficients. The apparent correlation of antipodal points stems from the fact that this kernel describes the fluctuations around the removed dipole, which is dominated by the quadrupole contributions. Overall cross correlations result from superimposing dipole and non-dipole parts. Correlations are normalized and a diverging colormap is used, ranging from –1 (blue) to 1 (red).

gree must be chosen such that the smallest spatial wavelength coincides with distances between sites. If some sites were clustered with distances of about 100 km, an expansion up to degree $\ell \leq 175$ would be needed. Since high degree expansions are demanding, using an explicit kernel function is computationally beneficial and makes it feasible to adopt a global point of view while preserving the accuracy of local length scales.

The reference radius R controls the predominant slope of the a priori power spectrum. The smaller R, the smoother the a priori field at the surface. The scaling factor λ causes a shift along the axis of ordinates. Similar to Constable and Parker (1988), the power spectrum is used to tune the reference radius. R is chosen such that the prior mean power spectrum roughly conforms with the International Geomagnetic Reference Field (IGRF) (Thébault et al., 2015) time average from 1900 to 2020. Figure 4.2 depicts the alignment, carried out by visual inspection. Throughout we are going to use the fixed value $R = 2800$ km. In contrast, as λ is highly uncertain, it receives special treatment in Section 4.4.2. A reference radius below the core-mantle boundary (CMB) implies a non-erratic covariance structure at the CMB. A virtual source region within the outer core may seem unconventional but is becoming more popular. To give an example, Sanchez et al. (2016) are using complex correlation patterns at the CMB obtained from dynamo simulations as prior information.

The EMF is modelled as the negative gradient of the potential (Eq. 4.24). Differentiation is a linear operation and thus the field model is a GP as well. The a priori mean reads

$$\bar{\mathbf{B}}(\mathbf{x}) = -\nabla \mathbf{Y}_1^\top(\mathbf{x})\bar{\mathbf{g}}_1 = -\sum_{-1\leq m\leq 1} \nabla Y_1^m(\mathbf{x})\bar{g}_1^m \tag{4.31}$$

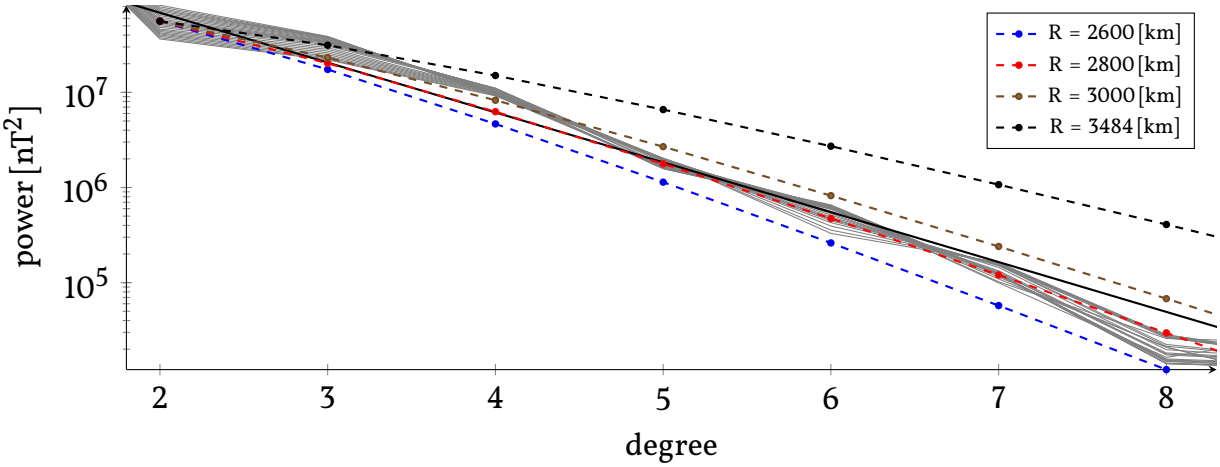


Figure 4.2: The mean power spectral behaviour chosen a priori is indicated by red dots. As a reference, the power spectra of the IGRF from 1900 to 2020 are drawn in grey. For comparison, the solid black line refers to the mean spectrum suggested by Constable and Parker (1988).

and the correlation kernel is composed of dipole and non-dipole covariance functions

$$K_{\mathbf{B}}(\mathbf{x},\mathbf{y}) = K_{\mathbf{B},DP} + K_{\mathbf{B},ND} = \nabla \mathbf{Y}_1^\top(\mathbf{x})\Sigma_1 \mathbf{Y}_1(\mathbf{y})\nabla^\dagger + \nabla K_L(\mathbf{x},\mathbf{y})\nabla^\dagger \tag{4.32}$$

where the right-hand gradient acts on the left at the second argument.

4.2.2 Inference

In contrast to existing archeomagnetic field models we pursue a functional analytic approach and perform a regression directly in the space of functions. The corner stone of our modelling strategy is formed by a GP regression. GP regression is also known in the field of geostatistics as Kriging (see Rasmussen and Williams, 2006, Sec. 2.2 and 2.8). For linear **B**-field observations this strategy has already been adopted by Holschneider et al. (2016). We briefly recall the overall modelling concept before we introduce non-linear observations.

Suppose we observe the components B_N, B_E and B_Z of the EMF. Measurements are denoted by

$$o = \left\{ \mathbf{o}_i = \left(B_N(\mathbf{z}_i), B_E(\mathbf{z}_i), B_Z(\mathbf{z}_i)\right)^\top \right\}_{i=1,\dots,n} , \tag{4.33}$$

recorded at locations $\mathbf{z}_i$. Observations are corrupted by additive noise $\mathbf{E}_i$ and the data model reads

$$O = \{\mathbf{B}(\mathbf{z}_i) + \mathbf{E}_i\}_{i=1,\dots,n} . \tag{4.34}$$

The error model is assumed normal of zero mean and covariance Σ_E. Recorded values o are assumed a realization of O. The data's a priori mean vector and covariance matrix are

$$\bar{O} = \left\{\bar{\mathbf{B}}(\mathbf{z}_i)\right\}_{i=1,\dots,n} \tag{4.35}$$

$$\Sigma_O = \left\{K_{\mathbf{B}}(\mathbf{z}_i, \mathbf{z}_j)\right\}_{i,j=1,\dots,n} + \Sigma_E \tag{4.36}$$

with typically diagonal error-covariance matrix Σ_E.

To obtain information about the EMF, we need to compute **B**'s posterior distribution. If we assume the **B**-field and measurement errors are independent, the cross covariance matrix follows to read

$$\Sigma_{\mathbf{B}(\mathbf{x})O} = \mathrm{Cov}\left[\mathbf{B}(\mathbf{x}), O\right] = \left\{K_{\mathbf{B}}(\mathbf{x}, \mathbf{z}_i)\right\}_{i=1,\dots,n} \tag{4.37}$$

for any design point **x** outside the reference sphere. Since O and **B** are jointly Gaussian, the posterior distribution is normal as well. It is fully determined by the conditional mean and conditional covariance

$$\mathbb{E}\left[\mathbf{B}(\mathbf{x})|o\right] = \bar{\mathbf{B}}(\mathbf{x}) \qquad + \Sigma_{\mathbf{B}O}\Sigma_O^{-1}\left(o - \bar{O}\right) \tag{4.38}$$

$$\mathrm{Cov}\left[\mathbf{B}(\mathbf{x}), \mathbf{B}(\mathbf{y})|o\right] = K_{\mathbf{B}}(\mathbf{x}, \mathbf{y}) - \Sigma_{\mathbf{B}O}\Sigma_O^{-1}\Sigma_{\mathbf{B}O}^{\top}\,. \tag{4.39}$$

Gauss coefficients are modelled analogously. Magnetic potential and Gauss coefficients are related through

$$g_\ell^m = \frac{2\ell+1}{4\pi R}\iint Y_\ell^m(\mathbf{x})\Phi(\mathbf{x})\,\mathrm{d}^2\mathbf{x}\,, \tag{4.40}$$

where integration is carried out over the sphere of radius R. The collection of Gauss coefficients up to SH degree ℓ is denoted by $\mathbf{g}_\ell$. Accordingly, Σ_ℓ refers to the prior covariance matrix of Gauss coefficients up to degree ℓ. Except for the dipole and according to Eq. 4.30, the a priori covariance Σ_ℓ is diagonal. At the reference radius the a priori variance is λ^2. The cross covariance matrix between $\mathbf{g}_\ell$ and the observations reads

$$\Sigma_{\ell O} = \mathrm{Cov}\left[\mathbf{g}_\ell, O\right] = \left\{-\Sigma_\ell \nabla \mathbf{Y}_\ell(\mathbf{z}_i)\right\}_{i=1,\dots,n} \tag{4.41}$$

and $\mathbf{Y}_\ell$ refers to the SH basis up to degree ℓ. Since Gauss coefficients and potential are linearly related, the posterior distribution is normal as well and, again, fully determined by the conditional mean and covariance

$$\mathbb{E}\left[\mathbf{g}_\ell|\mathbf{o}\right] = \bar{\mathbf{g}}_\ell + \Sigma_{\ell\mathrm{O}}\Sigma_{\mathrm{O}}^{-1}\left(\mathbf{o} - \bar{\mathrm{O}}\right) \tag{4.42}$$

$$\mathbb{V}\left[\mathbf{g}_\ell|\mathbf{o}\right] = \Sigma_\ell - \Sigma_{\ell\mathrm{O}}\Sigma_{\mathrm{O}}^{-1}\Sigma_{\ell\mathrm{O}}^\top \,. \tag{4.43}$$

It is worth mentioning that Bouligand et al. (2005) used geodynamo simulations to conclude that there are significant cross-correlations among the Gauss coefficients. Our statistical model places no restrictions on the posterior cross-correlations although the a priori assumptions are based on IID Gauss coefficients.

These are the formulae we are going to build our modelling strategy upon. However, they require an extension since archeomagnetic records and the magnetic potential are non-linearly related. Before elaborating our approach to this problem, let us first recall the observational functionals in question.

4.2.3 Observational functionals & Linearisation

Archeomagnetic data is not provided in the form of Cartesian field vector components. The quantities that are determined in laboratory experiments are the two angles, declination (D) and inclination (I), and intensity (F), acquired at locations $\mathbf{z}$. These quantities and the vector components are non-linearly related. The corresponding functionals read

$$\mathbf{H} : \mathbf{B} \to \begin{pmatrix} \mathrm{D} \\ \mathrm{I} \\ \mathrm{F} \end{pmatrix} = \begin{pmatrix} \arctan\left(\frac{\mathrm{B_E}}{\mathrm{B_N}}\right) \\ \arctan\left(\frac{\mathrm{B_Z}}{\mathrm{F_H}}\right) \\ \sqrt{\mathrm{B_N^2} + \mathrm{B_E^2} + \mathrm{B_Z^2}} \end{pmatrix} \tag{4.44}$$

where the horizontal intensity

$$\mathrm{F_H} = \sqrt{\mathrm{B_N^2} + \mathrm{B_E^2}} \tag{4.45}$$

is introduced as an auxiliary variable (Backus, Parker, and Constable, 1996, Eq. 1.2.1 – 1.2.4). The components of $\mathbf{H}$ are referred to as observation functionals and are denoted by $\mathrm{H_i}[\mathbf{B}]$ for $\mathrm{i} = \mathrm{D}, \mathrm{I}, \mathrm{F}$. The inverse map to magnetic field vector components reads

$$\mathbf{H}^{-1} : \begin{pmatrix} \mathrm{D} \\ \mathrm{I} \\ \mathrm{F} \end{pmatrix} \to \mathbf{B} = \mathrm{F} \begin{pmatrix} \cos(\mathrm{I})\cos(\mathrm{D}) \\ \cos(\mathrm{I})\sin(\mathrm{D}) \\ \sin(\mathrm{I}) \end{pmatrix} \,. \tag{4.46}$$

The inverse map is only unique if all three observables are at hand, i.e. three vector components relate uniquely to three observables and vice versa.

As already pointed out in the previous Section, the vital prerequisite for the modelling strategy is joint normality of observations and EMF. Certainly, the functionals D, I and F do not preserve **B**'s normality, nor are the measurement errors Gaussian.

To adopt the modelling concept by Holschneider et al. (2016) we approximate D, I and F by a 1^{st} order Taylor expansion

$$H_i[\mathbf{B}] \approx H_i[\tilde{\mathbf{B}}] + \nabla H_i[\tilde{\mathbf{B}}]^\top \left(\mathbf{B} - \tilde{\mathbf{B}}\right) , \tag{4.47}$$

where $\tilde{\mathbf{B}}$ refers to an arbitrary point of expansion (POE). This is the non-parametric counterpart compared with the approach presented by Gubbins and Bloxham (1985, Eq. 10). The functionals approximating D, I and F arise to

$$D \approx \tilde{D} + \frac{1}{\tilde{F}_H^2} \begin{bmatrix} -\tilde{B}_E \\ \tilde{B}_N \\ 0 \end{bmatrix}^\top \mathbf{B} , \tag{4.48}$$

$$I \approx \tilde{I} + \frac{1}{\tilde{F}_H} \left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{\tilde{B}_Z}{\tilde{F}} \frac{\tilde{\mathbf{B}}}{\tilde{F}} \right)^\top \mathbf{B} , \tag{4.49}$$

$$F \approx \frac{\tilde{\mathbf{B}}^\top}{\tilde{F}} \mathbf{B} \tag{4.50}$$

where $\tilde{D}$, $\tilde{I}$ and $\tilde{F}$ are referring to the 0^{th} order terms. From those equations it is obvious that performing an expansion about zero is not going to work. Existing models (e.g. Licht et al., 2013; Hellio and Gillet, 2018) typically use an axial dipole as initial POE. Since a Taylor expansion performs better the less **B** deviates from the POE, we propose not to use an axial dipole. Section 4.3 is dedicated to which point to linearise about.

Since **B** is assumed a GP and due to linearity, the approximating observational functionals are normally distributed. However, to actually achieve joint normality amongst observations and **B**-field, measurement errors need a Gaussian proxy as well.

4.2.4 Measurement Errors

Although the observational functionals were linearised, the data model is still not normal since measurement errors are not necessarily Gaussian. It is common to characterize the uncertainty of archeomagnetic directions (D and I) using the von Mises-Fisher (vMF) distribution and the variability of intensity using the normal distribution (Love and Constable, 2003). As long as intensity records report the standard error, linearising F provides a normal proxy model which we are going to use for inference. However, we lack a Gaussian proxy for directional errors.

The commonly used approach is to calculate directional errors individually. The vMF distribution is parametrized by a concentration factor and a location parameter. The larger the value of the concentration factor, the more the distribution tends towards concentrating around the location parameter. Provided a large concentration factor and that the location parameter is not pointing towards high latitudes, the marginal probability densities for D and I are approximately Gaussian. Proxy errors for declination and inclination are assumed independent, of zero mean and standard deviation

$$\sigma_I = \frac{57.3^\circ}{140}\alpha_{95} \qquad \text{and} \qquad \sigma_D = \frac{1}{\cos o_I}\sigma_I \tag{4.51}$$

(Suttie and Nilsson, 2019). Typically directional records report the 95% confidence cone α_{95} of the vMF distribution. This is a pragmatic approach that does not necessarily reflect the Gaussian moment matching proxy, since correlations are dropped and – in general – the mean of a vMF distribution does not coincide with the location parameter. In case of isolated declinations this approach does not work and such records are rejected. Another drawback of a Gaussian proxy error model is the intolerance against outliers. In the context of optimization theory there exist more robust approaches e.g. Farquharson and Oldenburg (1998), Walker and Jackson (2008), and Hellio et al. (2014). For the vast majority of non-Gaussian likelihoods, however, no explicit solution to the Bayesian inverse problem is known.

Two philosophies have been used in previous work when modelling archeo- and paleomagnetic data. Either very strict selection criteria are applied and these often contain tests that had not been applied in data published a few decades ago. As the absence of the test does not necessarily mean a result would not have passed the test, the other philosophy is to include as much data as possible without applying very strict criteria, aiming to increase the signal to noise ratio. However, it is likely that the reported measurement errors in several cases underestimate the true data uncertainties, which might contain systematic biases if corrections for, e.g., cooling rate or anisotropy have not been performed.

For practical reasons, we adopt the second philosophy. To compensate possibly non-conforming error estimates we introduce a scaling parameter ε. Then, the individual proxy uncertainties are given by

$$E_i \to \varepsilon E_i \sim \mathcal{N}\left(0,\ \varepsilon^2 \sigma_i^2\right) \tag{4.52}$$

and ε is regarded as a model parameter. Although we do not know the specific value for ε, its order of magnitude is assumed to be one.

In addition, we introduce a residual term $\mathbf{P}$ that compensates modelling related errors and accounts for observational biases. Amongst others, effects that our model does not include are temporal correlations, dating errors and crustal field anomalies. Therefore, we assume $\mathbf{P} \sim \mathcal{N}(0,\ \mathbf{I})$, i.e. uncorrelated standard normal at every pair of distinct sites. Our final data model becomes

$$o_i = H_i[\mathbf{B}_i + \rho \mathbf{P}_i] + \varepsilon E_i \tag{4.53}$$

and the magnitude of the residual is controlled by ρ, another not yet known model parameter. The residual term can be thought of as an error term that describes the simplification of the underlying physics statistically. In other words, those real world contributions that are not covered by our model are treated as if they were random errors.

Because we focus on time stationary snapshots of the EMF, dating uncertainties are displaced into the residual term. Nonetheless, these errors are of importance if the temporal behaviour of the EMF is reconstructed. There already exist several approaches to accommodate dating errors e.g. Jack-knife (Korte, Donadini, and Constable, 2009; Licht et al., 2013), MCMC sampling (Hellio et al., 2014) or transdimensional modelling (Livermore et al., 2018).

To proceed with the modelling concept outlined in Section 4.2.2, we combine the linearisation and the Gaussian proxy for the directional errors. However, a POE for the linearisation is still missing. The subsequent section covers this problem and also discusses the concrete incorporation of the error term and the residual. As the final ingredient of our modelling concept, the treatment of model parameters is subject of Section 4.4.

4.3 Bayesian Update System

The need for linearisation described in Section 4.2.3 requires a suitable POE. The linearisation as a Taylor expansion performs better the less the POE deviates from the *truth*. Archeo- and paleomagnetic directional and intensity data are determined from different laboratory experiments, and the majority of records report either one or two field components (incomplete records). The complete vector information (D, I, F) is only available in rare cases. Noting that it is easier to determine a POE from full vector records, we introduce a Bayesian update system to treat complete and incomplete records successively.

The posterior distribution is computed by a two step strategy only considering a subset of observations at a time. Records are partitioned into two disjoint groups $o_{\mathcal{I}}$ and $o_{\mathcal{C}}$ where subscripts are referring to incomplete and complete measurements. Making use of the conditional probability rule – i.e. $p(X|Y)p(Y) = p(X, Y)$ – and according to Bayes' law the posterior **B**-field factorizes

$$p(\mathbf{B}|o) = p(\mathbf{B}|o_{\mathcal{C}}, o_{\mathcal{I}}) = \frac{p(o_{\mathcal{I}}|\mathbf{B}, o_{\mathcal{C}})}{p(o_{\mathcal{I}}|o_{\mathcal{C}})} p(\mathbf{B}|o_{\mathcal{C}}) \,, \tag{4.54}$$

i.e. the posterior EMF based on the complete observations $o_{\mathcal{C}}$ serves as prior for the Bayesian posterior based on $o_{\mathcal{I}}$. Not to incorporate the data all at once appears to be a promising strategy due to strong magnetic field correlations.

The complexity of the developed algorithm is growing through this and the following section. Figure 4.3 provides a schematic illustration so as not to loose the overview. The two step strategy is shown in the top panel of Figure 4.3.

4.3.1 Complete records

In the initial step only complete records are taken into account. Triplets of declination, inclination and intensity are forming the set of complete records

$$o_{\mathcal{C}} = \left\{ \mathbf{o}_i = (D(\mathbf{z}_i), I(\mathbf{z}_i), F(\mathbf{z}_i))^\top \right\}_{i=1,\dots,n_{\mathcal{C}}} \,. \tag{4.55}$$

In order to apply the linearisation, a POE is missing. The special case of knowing all three components allows the calculation of the inverse map (Eq. 4.46)

$$\tilde{\mathbf{B}}_{\mathcal{C}} = \left\{ \mathbf{H}^{-1}[\mathbf{o}_i] \right\}_{i=1,\dots,n_{\mathcal{C}}} \tag{4.56}$$

which will serve as the point to linearise about. That point is reasonable as long as the prior **B**-field variance is large in comparison with the measurement errors. If measurement errors were not negligible w.r.t. the a priori distribution, then the POE is no longer known with confidence and we have to propagate uncertainties. To ensure this, Section 4.4 is devoted to choosing an uninformative a priori field. A Gaussian approximation in this way is also known as Laplace's method (Murphy, 2012, Sec. 8.4.1).

In order to apply the modelling scheme introduced in Section 4.2.2 we use the Gaussian proxy error model. The diagonal error covariance matrix is denoted by

$$\Sigma_{\mathrm{E},\mathcal{C}} = \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_{\mathrm{n}_\mathcal{C}}^2\right) , \tag{4.57}$$

where σ_i refers to individual standard errors w.r.t. D, I and F. The linearised observation functionals translate between (D, I, F) and **B**. To keep equations concise, the dipole basis and the Jacobi matrices are collected in big matrices

$$\acute{\mathrm{Y}}_{1,\mathcal{C}} = \left\{\nabla \mathbf{Y}_1(\mathbf{z}_\mathrm{i})\right\}_{\mathrm{i}=1,\ldots,\mathrm{n}_\mathcal{C}} \qquad (3\mathrm{n}_\mathcal{C} \times 3) \tag{4.58}$$

$$\acute{\mathrm{H}}_\mathcal{C} = \left\{\delta_{\mathrm{ij}} \nabla \mathbf{H}[\tilde{\mathbf{B}}_\mathrm{i}]\right\}_{\mathrm{i,j}=1,\ldots,\mathrm{n}_\mathcal{C}} \qquad (3\mathrm{n}_\mathcal{C} \times 3\mathrm{n}_\mathcal{C}) \tag{4.59}$$

where δ_{ij} refers to the Kronecker delta and $\acute{\mathrm{H}}_\mathcal{C}$ is 3×3 block-diagonal. For $\mathrm{o}_\mathcal{C}$, the approximative prior mean vector is given by

$$\bar{\mathrm{O}}_\mathcal{C} \approx \mathrm{o}_\mathcal{C} + \acute{\mathrm{H}}_\mathcal{C}^\top \left(-\acute{\mathrm{Y}}_{1,\mathcal{C}} \bar{\mathbf{g}}_1 - \tilde{\mathbf{B}}_\mathcal{C}\right) \tag{4.60}$$

where $\mathbf{B}_\mathcal{C}$ means evaluated at all the locations of observation. Due to assumed independence of error model and EMF, the linearised covariance matrix for complete records reads

$$\Sigma_\mathcal{C} = \mathbb{V}\left[\mathrm{O}_\mathcal{C}\right] \approx \acute{\mathrm{H}}_\mathcal{C}^\top \left(\mathbb{V}\left[\mathbf{B}_\mathcal{C}\right] + \rho^2 \mathbf{1}\right) \acute{\mathrm{H}}_\mathcal{C} + \varepsilon^2 \Sigma_{\mathrm{E},\mathcal{C}} , \tag{4.61}$$

where $\mathbb{V}\left[\mathbf{B}_\mathcal{C}\right]$ is constructed from the kernel (Eq. 4.32) at any two locations of observation. Due to bi-linearity of the covariance $\acute{\mathrm{H}}_\mathcal{C}$ is factored out and all constant terms are stripped off. For arbitrary design points, the linearised cross covariance amongst EMF and measurements is given by

$$\Sigma_{\mathbf{B}\mathcal{C}} = \mathrm{Cov}\left[\mathbf{B}, \mathrm{O}_\mathcal{C}\right] \approx \mathrm{Cov}\left[\mathbf{B}, \mathbf{B}_\mathcal{C}\right] \acute{\mathrm{H}}_\mathcal{C} . \tag{4.62}$$

According to Equations 4.38 and 4.39, a Gaussian proxy of **B**'s posterior distribution is determined through the conditional mean $\mathbb{E}\left[\mathbf{B}|\mathrm{o}_\mathcal{C}\right]$ and conditional co-variance $\mathbb{V}\left[\mathbf{B}|\mathrm{o}_\mathcal{C}\right]$. This first step of our modelling scheme is illustrated in the top left panel of Figure 4.3.

Analogous to Eq. 4.41, the linearised cross covariance amongst Gauss coefficients and observations is given by

$$\Sigma_{\ell\mathcal{C}} = \mathrm{Cov}\left[\mathbf{g}_\ell, \mathrm{O}_\mathcal{C}\right] \approx \mathrm{Cov}\left[\mathbf{g}_\ell, \mathbf{B}_\mathcal{C}\right] \acute{\mathrm{H}}_\mathcal{C} = -\Sigma_\ell \acute{\mathrm{Y}}_{\ell,\mathcal{C}} \acute{\mathrm{H}}_\mathcal{C} \ . \tag{4.63}$$

where $\acute{\mathrm{Y}}_{\ell,\mathcal{C}}$ extends Eq. 4.58 up to SH degree ℓ. The approximations of the conditional mean $\mathbb{E}\left[\mathbf{g}_\ell|\mathrm{o}_\mathcal{C}\right]$ and conditional co-variance $\mathbb{V}\left[\mathbf{g}_\ell|\mathrm{o}_\mathcal{C}\right]$ are given through Equations 4.42 and 4.43.

Although the subset of complete records is comparatively small, as a first *guess*, we anticipate a reconstruction of the EMF's dominating features due to strong magnetic field correlations.

4.3.2 Incomplete records

Inference with the incomplete records only implicitly depends on the a priori mean and covariance function through the first step. The distribution of the EMF posterior to $\mathrm{o}_\mathcal{C}$ may be understood as the prior to incorporate the remaining measurements $\mathrm{o}_\mathcal{I}$. To carry out the linearisation, the mean conditional on $\mathrm{o}_\mathcal{C}$ will serve as POE

$$\tilde{\mathbf{B}}_{|\mathcal{C}} = \mathbb{E}\left[\mathbf{B}|\mathrm{o}_\mathcal{C}\right] \ . \tag{4.64}$$

From a theoretical point of view this is a function that is going to be evaluated by the observational functionals. We expect those points to be well suited, as in a Gaussian model the mean is the most likely solution. However, this is an arbitrary choice and does not necessary imply any optimality.

Because the EMF's proxy posterior to $\mathrm{o}_\mathcal{C}$ is normal we are going to use the same modelling concept as we have already done (Sec. 4.2.2). To facilitate the second step it is necessary not only to predict on design points but also on *auxiliary* quantities in the first step.

Incomplete records are treated individually even though a certain location may report more than one observable. The approximative mean is given by

$$\bar{\mathrm{O}}_{\mathcal{I}|\mathcal{C}} \approx \left\{ \mathrm{H}_\mathrm{i}[\tilde{\mathbf{B}}_{|\mathcal{C}}] \right\}_{\mathrm{i}=1,\ldots,\mathrm{n}_\mathcal{I}} \tag{4.65}$$

and we use the subscript i to indicate both the location $\mathbf{z}_\mathrm{i}$ and the type of record in H_i, either D, I or F. If we collect all gradients within one big matrix

$$\acute{\mathrm{H}}_{\mathcal{I}|\mathcal{C}} = \left\{ \delta_{\mathrm{ij}} \nabla \mathrm{H}_\mathrm{i}[\tilde{\mathbf{B}}_{|\mathcal{C}}] \right\}_{\mathrm{ij}=1,\ldots,\mathrm{n}_\mathcal{I}} \ , \tag{4.66}$$

then the covariance matrix for incomplete records reads

$$\Sigma_{\mathcal{I}|\mathcal{C}} \approx \acute{\mathrm{H}}_{\mathcal{I}|\mathcal{C}}^{\top} \left(\mathbb{V}\left[\mathbf{B}_{\mathcal{I}}|\mathrm{o}_{\mathcal{C}}\right] + \rho^2 \mathbf{1} \right) \acute{\mathrm{H}}_{\mathcal{I}|\mathcal{C}} + \varepsilon^2 \Sigma_{\mathrm{E},\mathcal{I}} \,. \tag{4.67}$$

The error covariance matrix is analogous to Eq. 4.57. The auxiliary quantities we have to carry along are conditional mean vector and covariance matrix at points of observation.

To model the posterior **B**-field we again have to calculate linearised cross correlations amongst $\mathbf{B}|\mathrm{o}_{\mathcal{C}}$ and $\mathrm{O}_{\mathcal{I}}$. If we store the matrix $\mathrm{Cov}\left[\mathbf{B}, \mathbf{B}_{\mathcal{I}}|\mathrm{o}_{\mathcal{C}}\right]$ within the first step, linearised cross correlations amongst design points and incomplete records are given by

$$\Sigma_{\mathbf{B}\mathcal{I}|\mathcal{C}} = \mathrm{Cov}\left[\mathbf{B}, \mathrm{O}_{\mathcal{I}}|\mathrm{o}_{\mathcal{C}}\right] \approx \mathrm{Cov}\left[\mathbf{B}, \mathbf{B}_{\mathcal{I}}|\mathrm{o}_{\mathcal{C}}\right] \acute{\mathrm{H}}_{\mathcal{I}} \,. \tag{4.68}$$

The Gaussian proxy for the EMF's posterior distribution is again determined through Equations 4.38 and 4.39. Conditional mean and co-variance read

$$\mathbb{E}\left[\mathbf{B}|\mathrm{o}\right] = \mathbb{E}\left[\mathbf{B}|\mathrm{o}_{\mathcal{C}}\right] + \Sigma_{\mathbf{B}\mathcal{I}|\mathcal{C}} \Sigma_{\mathcal{I}|\mathcal{C}}^{-1} (\mathrm{o}_{\mathcal{I}} - \bar{\mathrm{O}}_{\mathcal{I}|\mathcal{C}}) \tag{4.69}$$

$$\mathbb{V}\left[\mathbf{B}|\mathrm{o}\right] = \mathbb{V}\left[\mathbf{B}|\mathrm{o}_{\mathcal{C}}\right] - \Sigma_{\mathbf{B}\mathcal{I}|\mathcal{C}} \Sigma_{\mathcal{I}|\mathcal{C}}^{-1} \Sigma_{\mathbf{B}\mathcal{I}|\mathcal{C}}^{\top} \,. \tag{4.70}$$

The top panel of Figure 4.3 illustrates how the posterior EMF is built within two steps.

Gauss coefficients are estimated analogously. If we store the matrix $\mathrm{Cov}\left[\mathbf{g}_{\ell}, \mathbf{B}_{\mathcal{I}}|\mathcal{C}\right]$ while performing the first step, then, the linearised cross correlations are given by

$$\Sigma_{\ell\mathcal{I}|\mathcal{C}} = \mathrm{Cov}\left[\mathbf{g}_{\ell}, \mathrm{O}_{\mathcal{I}}|\mathrm{o}_{\mathcal{C}}\right] \approx \mathrm{Cov}\left[\mathbf{g}_{\ell}, \mathbf{B}_{\mathcal{I}}\middle|\mathrm{o}_{\mathcal{C}}\right] \acute{\mathrm{H}}_{\mathcal{I}|\mathcal{C}} \,. \tag{4.71}$$

Approximations of conditional mean and co-variance translate according to Equations 4.42 and 4.43.

Unfortunately, we cannot yet directly apply that algorithm since we do not know specific values for the model parameters $\bar{\mathbf{g}}_1$, Σ_1, λ, ρ and ε. Before we deal with abandoning these parameters (see Sec. 4.4) we carry out tests with synthetic data to validate the proposed framework.

4.3.3 Tests

The current IGRF coefficients are used as a realistic reference field. Synthetic data is generated from the reference field and is corrupted by artificial noise. A gamma distribution is used to corrupt intensity records whereas directional data is randomly drawn from a vMF distribution. To check for robustness, the measurement errors are chosen on purpose not to coincide with the proposed Gaussian proxy error model. For the tests we use several datasets that differ in the error level, the complete/incomplete ratio and the spa-

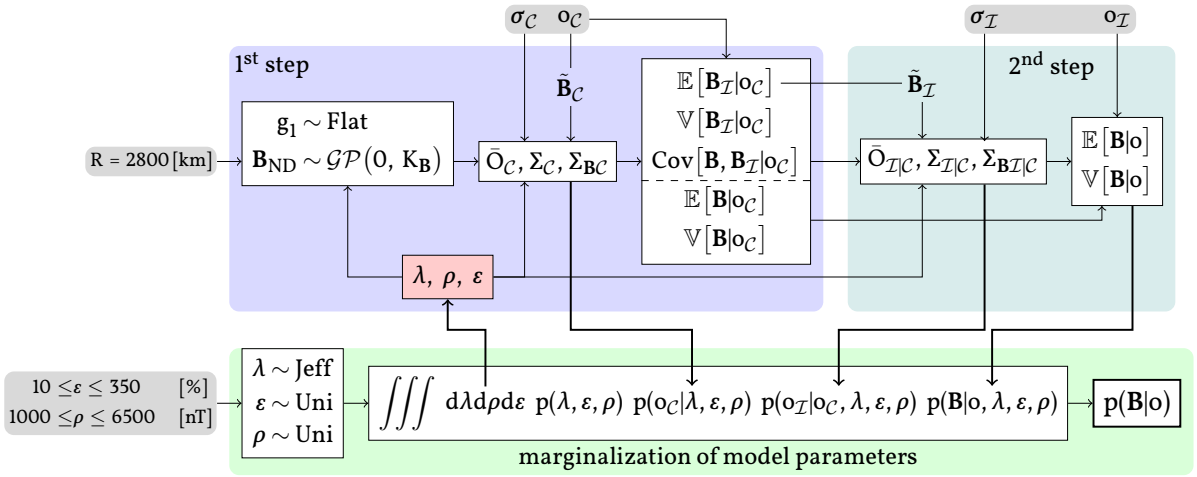


Figure 4.3: Illustration of how the two-step strategy and a marginalization intertwine to compute the posterior compound distribution. Invariants – such as observations – are shaded in grey. The upper part refers to the update system where model parameters are highlighted in red. Top left panel shows the initial step (Sec. 4.3.1) whereas top right illustrates the update (Sec. 4.3.2). The marginalization is highlighted in green (Sec. 4.4.2). Arrows indicate how information is passed.

tial distribution. Since the reference model is known, we use the mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{n}\sum_{j}^{n} \left|\left| \mathbf{B}(\mathbf{z}_j) - \mathbb{E}\left[\mathbf{B}(\mathbf{z}_j)|o\right] \right|\right|_1 \tag{4.72}$$

as a test characteristic. Reconstruction and reference field are compared at the Earth's surface, sampled at a rate that accounts for the length scales present in the reference field. Our modelling strategy is able to recover the reference field for all datasets considered.

Furthermore the influence of the POE on the linearisation is examined. We compare the proposed strategy to linearisation about an axial dipole of $g_1^0 = -23\,\mu$T. For all datasets under consideration, the MAE of our modelling strategy falls below linearisation about the axial dipole. Especially if the dataset mimics reality, the proposed strategy performs better.

To make sure that $g_1^0 = -23\,\mu$T is not a coincidence, we also vary the axial dipole strength. For the synthetic data that mimic reality, Figure 4.4 compares the MAE of the proposed strategy with linearising about a range of axial dipoles. Even the axial dipole featuring the smallest MAE has a value that is still above the proposed POE built from the complete data. Throughout the synthetic datasets considered, the proposed strategy outperforms the linearisation about the *best* suited axial dipole. A detailed description and the full test results are available together with the source code by Schanner and Mauerberger (2019).

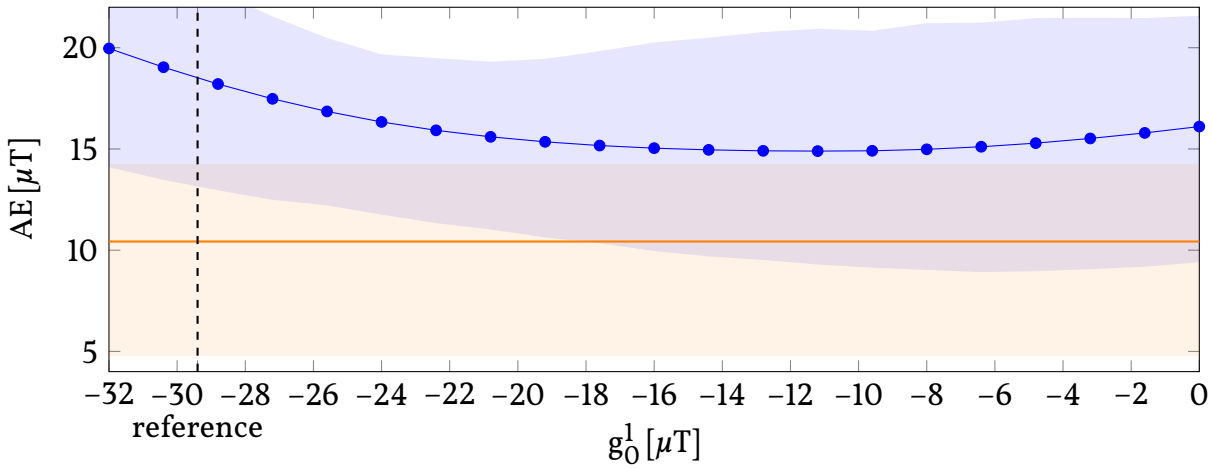


Figure 4.4: Linearisation about an axial dipole compared to the strategy proposed. The horizontal red line indicates the MAE of our approach whereas the blue dots refer to the MAE by varying the axial dipole strength. The shaded areas extend from the 25% to the 75% quantile to illustrates the spread of absolute errors. The dashed vertical line refers to the axial dipole of the reference field.

4.4 Model Parameters

Before applying the outlined modelling strategy we need to address several model parameters. As the posterior EMF must not depend on a certain choice of model parameters, the a priori mean $\bar{\mathbf{g}}_1$ and covariance Σ_1 of the dipole and the scaling factors λ, ρ and ε are considered so-called *nuisance* parameters, i.e. parameters that are not of primary interest. To abandon those quantities from the EMF's posterior distribution we are going to use two different techniques: Section 4.4.1 analytically eliminates $\bar{\mathbf{g}}_1$ and Σ_1 by exploring the limit of a flat dipole prior. Section 4.4.2 addresses a hierarchical Bayesian approach to marginalize the scaling factors λ, ρ and ε.

4.4.1 Uninformative Dipole

To make as few assumptions as possible we aim to explore the limit of an uninformative prior dipole. This approach is beneficial in two different ways: 1) It prevents us from accidentally choosing an overly confident prior. 2) We abandon all of the nine dipole-parameters. We expect the data to be strong enough to estimate $\mathbf{g}_1$ with confidence.

Loosely speaking, a flat prior may be understood as the limit of a Gaussian with variance sent to infinity. Since an unbounded covariance is not well defined the standard approach is to express formulae w.r.t. the precision and explore the limiting case of vanishing dipole precision i.e. $\Sigma_1^{-1} \to 0$. In the following, we closely follow Rasmussen and Williams (2006, Sec. 2.7), additionally taking our update system and the linearisation into account.

Even though the prior dipole precision is sent to zero, the posterior distribution remains normal. The second step in our update system depends only implicitly on the a priori assumptions made, thus, only the first step needs modifications. This means that the overall modelling concept persists, although we have to modify the conditional mean and covariance (Eqs. 4.38 and 4.39).

First of all, we have to partition the covariance matrix into dipole and non-dipole contributions

$$\Sigma_{\mathcal{C}} = \Sigma_{\mathcal{C},\mathrm{DP}} + \Sigma_{\mathcal{C},\mathrm{ND}} = \acute{\mathrm{H}}_{\mathcal{C}} \acute{\mathrm{Y}}_{1,\mathcal{C}} \Sigma_1 \acute{\mathrm{Y}}_{1,\mathcal{C}}^{\top} \acute{\mathrm{H}}_{\mathcal{C}}^{\top} + \Sigma_{\mathcal{C},\mathrm{ND}} \tag{4.73}$$

where $\Sigma_{\mathcal{C},\mathrm{ND}}$ is constructed from $\mathrm{K}_{\mathbf{B},\mathrm{ND}}$ also containing measurement errors and residuals. To keep equations concise, it is beneficial to predict the dipole coefficients first. Making use of the matrix inversion lemma Rasmussen and Williams, 2006, A.9, conditional mean and inverse of the covariance result in

$$\mathbb{V}\left[\mathbf{g}_1|\mathrm{o}_{\mathcal{C}}\right]^{-1} = \Sigma_1^{-1} + \acute{\mathrm{Y}}_{1,\mathcal{C}}^{\top} \acute{\mathrm{H}}_{\mathcal{C}}^{\top} \Sigma_{\mathcal{C},\mathrm{ND}}^{-1} \acute{\mathrm{H}}_{\mathcal{C}} \acute{\mathrm{Y}}_{1,\mathcal{C}} \tag{4.74}$$

$$\mathbb{E}\left[\mathbf{g}_1|\mathrm{o}_{\mathcal{C}}\right] = \mathbb{V}\left[\mathbf{g}_1|\mathrm{o}_{\mathcal{C}}\right]^{-1} \left(\Sigma_1^{-1}\bar{\mathbf{g}}_1 - \acute{\mathrm{Y}}_{1,\mathcal{C}}^{\top} \acute{\mathrm{H}}_{\mathcal{C}}^{\top} \Sigma_{\mathcal{C},\mathrm{ND}}^{-1} \acute{\mathrm{H}}_{\mathcal{C}} \tilde{\mathrm{B}}_{\mathcal{C}}\right) \tag{4.75}$$

and the data $\mathrm{o}_{\mathcal{C}}$ enters through the POE (see Eq. 4.60). Considering the limit of the uninformative dipole yields

$$\Sigma_{1|\mathcal{C}}^{-1} := \lim_{\Sigma_1^{-1}\to 0} \mathbb{V}\left[\mathbf{g}_1|\mathrm{o}_{\mathcal{C}}\right]^{-1} = \acute{\mathrm{Y}}_{1,\mathcal{C}}^{\top} \acute{\mathrm{H}}_{\mathcal{C}}^{\top} \Sigma_{\mathcal{C},\mathrm{ND}}^{-1} \acute{\mathrm{H}}_{\mathcal{C}} \acute{\mathrm{Y}}_{1,\mathcal{C}} \tag{4.76}$$

$$\bar{\mathbf{g}}_{1|\mathcal{C}} := \lim_{\Sigma_1^{-1}\to 0} \mathbb{E}\left[\mathbf{g}_1|\mathrm{o}_{\mathcal{C}}\right] = -\Sigma_{1|\mathcal{C}}^{-1} \acute{\mathrm{Y}}_{1,\mathcal{C}}^{\top} \acute{\mathrm{H}}_{\mathcal{C}}^{\top} \Sigma_{\mathcal{C},\mathrm{ND}}^{-1} \acute{\mathrm{H}}_{\mathcal{C}} \tilde{\mathrm{B}}_{\mathcal{C}} \,. \tag{4.77}$$

Interestingly, $\mathbf{g}_1|\mathrm{o}_{\mathcal{C}}$ does not depend on $\bar{\mathbf{g}}_1$, rendering the dipole's prior mean irrelevant.

Predicting Gauss coefficients of higher SH degree is straightforward since a priori dipole and non-dipole contributions are assumed independent. For $\ell > 1$ cross correlations are of no concern as they do not depend on $\mathbf{g}_1$. Analogous to Equation 4.63, conditional mean and covariance result in

$$\mathbb{E}\left[\mathbf{g}_{2:\ell}|\mathrm{o}_{\mathcal{C}}\right] = \quad - \Sigma_{2:\ell} \acute{\mathrm{Y}}_{2:1,\mathcal{C}}^{\top} \acute{\mathrm{H}}_{\mathcal{C}}^{\top} \Omega_{\mathcal{C}} \acute{\mathrm{H}}_{\mathcal{C}} \tilde{\mathrm{B}}_{\mathcal{C}} \tag{4.78}$$

$$\mathbb{V}\left[\mathbf{g}_{2:\ell}|\mathrm{o}_{\mathcal{C}}\right] = \Sigma_{2:\ell} - \Sigma_{2:\ell} \acute{\mathrm{Y}}_{2:1,\mathcal{C}}^{\top} \acute{\mathrm{H}}_{\mathcal{C}}^{\top} \Omega_{\mathcal{C}} \acute{\mathrm{H}}_{\mathcal{C}} \acute{\mathrm{Y}}_{2:1,\mathcal{C}} \Sigma_{2:\ell} \tag{4.79}$$

where $\Omega_{\mathcal{C}}$ refers to the limiting precision matrix

$$\Omega_{\mathcal{C}} = \Sigma_{\mathcal{C},\mathrm{ND}}^{-1} - \Sigma_{\mathcal{C},\mathrm{ND}}^{-1} \acute{\mathrm{H}}_{\mathcal{C}} \acute{\mathrm{Y}}_{1,\mathcal{C}} \Sigma_{1|\mathcal{C}} \acute{\mathrm{Y}}_{1,\mathcal{C}}^{\top} \acute{\mathrm{H}}_{\mathcal{C}}^{\top} \Sigma_{\mathcal{C},\mathrm{ND}}^{-1} \tag{4.80}$$

and we again made use of the matrix inversion lemma. To perform the second step, the whole posterior covariance matrix is necessary. However, the posterior cross covariance amongst dipole and non-dipole coefficients is missing. The difficulty is to find the limiting case. According to Equation 4.43 we have

$$\mathrm{Cov}\left[\mathbf{g}_1, \mathbf{g}_{2:\ell} | \mathrm{o}_\mathcal{C}\right] = -\mathrm{Cov}\left[\mathbf{g}_1, \mathrm{O}_\mathcal{C}\right] \Sigma_\mathcal{C}^{-1} \mathrm{Cov}\left[\mathrm{O}_\mathrm{C}, \mathbf{g}_{2:\ell}\right] \tag{4.81}$$

since a priori dipole and non-dipole are assumed independent. Plugging in the linearisation we end up with

$$\mathrm{Cov}\left[\mathbf{g}_1, \mathbf{g}_{2:\ell} | \mathrm{o}_\mathcal{C}\right] = -\Sigma_{1|\mathcal{C}} \acute{\mathrm{Y}}_{1,\mathcal{C}} \acute{\mathrm{H}}_\mathcal{C} \Sigma_{\mathcal{C},\mathrm{ND}}^{-1} \acute{\mathrm{H}}_\mathcal{C}^\top \acute{\mathrm{Y}}_{2:1,\mathcal{C}}^\top \Sigma_{2:\ell} \,, \tag{4.82}$$

which no longer depends on Σ_1. To see this, use the matrix inversion lemma to express the precision matrix, factor in the left hand side and expand by $\Sigma_{1|\mathcal{C}}^{-1} \Sigma_{1|\mathcal{C}}$. Using Equation 4.74 and rearranging terms yields Equation 4.82.

To predict on the EMF, we divide the cross covariance into dipole and non-dipole contributions. Since a priori the dipole term is assumed independent of the non-dipole contribution we have

$$\Sigma_{\mathbf{B}\mathcal{C}} = \Sigma_{\mathbf{B}\mathcal{C},\mathrm{DP}} + \Sigma_{\mathbf{B}\mathcal{C},\mathrm{ND}} = \acute{\mathrm{Y}}_1^\top \Sigma_1 \acute{\mathrm{Y}}_{1,\mathcal{C}} \acute{\mathrm{H}}_\mathcal{C} + \Sigma_{\mathbf{B}\mathcal{C},\mathrm{ND}} \tag{4.83}$$

and $\Sigma_{\mathbf{B}\mathcal{C},\mathrm{ND}}$ is constructed from $\mathrm{K}_{\mathbf{B},\mathrm{ND}}$. Using the same strategy as we did with the Gauss coefficients, the posterior mean and covariance arise to

$$\mathbb{E}\left[\mathbf{B} | \mathrm{o}_\mathcal{C}\right] = -\acute{\mathrm{Y}}_1^\top \bar{\mathrm{g}}_{1|\mathcal{C}} + \Sigma_{\mathbf{B}\mathcal{C},\mathrm{ND}} \Sigma_{\mathcal{C},\mathrm{ND}}^{-1} \left(\tilde{\mathrm{B}}_\mathcal{C} - \acute{\mathrm{Y}}_{1,\mathcal{C}}^\top \bar{\mathrm{g}}_{1|\mathcal{C}}\right) \tag{4.84}$$

$$\mathbb{V}\left[\mathbf{B} | \mathrm{o}_\mathcal{C}\right] = \acute{\mathrm{Y}}_1^\top \Sigma_{1|\mathcal{C}} \acute{\mathrm{Y}}_1 + \mathrm{K}_{\mathbf{B},\mathrm{ND}} - \Sigma_{\mathbf{B}\mathcal{C},\mathrm{ND}} \Omega_\mathcal{C} \Sigma_{\mathbf{B}\mathcal{C},\mathrm{ND}}^\top \,. \tag{4.85}$$

From those equations we can obtain all quantities needed to proceed with the second step incorporating incomplete records.

The importance in the result is that conditional mean and covariance no longer depend on the choice of the a priori dipole, since we assumed zero precision.

4.4.2 Compound Distribution

Although we are not particularly interested in reconstructing the probability distribution of $\vartheta = (\lambda, \rho, \varepsilon)$, their variabilities must be taken into account. The final result of the proceeding is the EMF's compound distribution

$$\mathrm{p}(\mathbf{B}|\mathrm{o}) = \int \mathrm{p}(\mathbf{B}|\mathrm{o}, \vartheta)\, \mathrm{p}(\vartheta|\mathrm{o}) \mathrm{d}\vartheta \tag{4.86}$$

which results from marginalizing the scaling factors. Figure 4.3 illustrates the interaction of the two-step strategy and the compound distribution, with the marginalization being depicted in the bottom part. This approach makes it possible to escape from Gaussianity. The integral will not be tractable analytically and must be evaluated by numerical methods.

Let us inspect the two PDFs we want to integrate over. For a certain choice of parameters, $p(\mathbf{B}|o, \vartheta)$ is calculated according to our two step strategy. To keep track of the scaling factor's posterior $p(\vartheta|o)$ we add a hierarchical stage. Applying Bayes' law, the posterior density is given by

$$p(\vartheta|o) \propto p(o|\vartheta)p(\vartheta) \tag{4.87}$$

where we neglected the normalizing constant. To suitably normalize we have to carry out another quadrature since $p(o) = \int p(o|\vartheta)p(\vartheta)d\vartheta$ is unknown. Because of the flat dipole prior, calculating $p(o|\vartheta)$ needs special attention. Not to distract from our endeavour, this is the focus of Section 4.4.3.

A priori, all three parameters are assumed statistically independent. Both, residual and error level are considered uniformly distributed. As we are roughly aware of magnitudes, the chosen range is well-spaced (weakly informative). On the contrary, λ is a scale parameter bearing across orders of magnitude. The according uninformative prior – representing the state of no prior information – is Jeffrey's prior

$$p(\lambda) \propto \frac{1}{\lambda}\,[\mathrm{nT}] \tag{4.88}$$

i.e. values ten times larger are just as likely as values ten times smaller (Murphy, 2012, Sec. 5.4.2). However, there is a subtlety arising for the compound prior PDFs. The hierarchical approach has an impact because the a priori $\mathbf{B}$-field depends on λ. The scale invariance is passed on rendering the compound prior distributions improper, i.e. the density can not be normalized. As an example, the compound PDF for the a priori non-dipole Gauss coefficients reads

$$p(g_\ell^m) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\lambda^2} \exp\left\{-\frac{1}{2}\frac{(g_\ell^m)^2}{\lambda^2}\right\} d\lambda \propto \frac{1}{|g_\ell^m|}\,. \tag{4.89}$$

That density remains centred at zero whereas the variance does not exist. Nevertheless, the compound posterior PDFs are normalizable and well defined. Marginalizing λ results in an extremely weak a priori assumption.

Although the compound distribution is not normal, calculating higher-order moments is one of its potentials. Posterior mean and covariance are given by

$$\mathbb{E}\left[\mathbf{B}|\mathrm{o}\right] = \int \mathrm{p}(\vartheta|\mathrm{o}) \quad \mathbb{E}\left[\mathbf{B}|\mathrm{o},\vartheta\right] \mathrm{d}\vartheta \tag{4.90}$$

$$\begin{aligned}\mathbb{V}\left[\mathbf{B}|\mathrm{o}\right] = &\int \mathrm{p}(\vartheta|\mathrm{o})\left(\ \mathbb{E}\left[\mathbf{B}|\mathrm{o},\vartheta\right]^2 + \mathbb{V}\left[\mathbf{B}|\mathrm{o},\vartheta\right]\ \right)\mathrm{d}\vartheta - \\ &- \mathbb{E}\left[\mathbf{B}|\mathrm{o}\right]^2\end{aligned} \tag{4.91}$$

and the mean and co-variance we are integrating over are explicitly given. For modelling Gauss coefficients the above formulae translate analogously.

An actual implementation requires to perform three numeric integrations, in total. Since numeric quadrature in three dimensions is feasible, both the fully Bayesian posterior density and also the Gaussian moment matching proxy are right at hand.

4.4.3 Marginal Likelihood

We postponed the calculation of the *marginal likelihood* until here. The terms *marginal* and *likelihood* refer to the marginalization over the EMF as a function in ϑ. To actually discretise and integrate the compound distribution, $\mathrm{p}(\mathrm{o}|\vartheta)$ is still missing. We again separate into complete and incomplete records. The marginal likelihood factorizes

$$\mathrm{p}(\mathrm{o}|\vartheta) = \mathrm{p}(\mathrm{o}_\mathcal{I}|\mathrm{o}_\mathcal{C},\vartheta)\mathrm{p}(\mathrm{o}_\mathcal{C}|\vartheta) \tag{4.92}$$

where we used the conditional probability rule. The major benefit is that we already know the quantities needed to evaluate the PDFs on the right hand side of Eq. 4.92 through our two step strategy. Figure 4.3 illustrates where Eq. 4.92 enters the modelling scheme and how quantities are passed between marginalization and two-step strategy.

For the sake of simplicity we analyse incomplete records, first. For a certain choice of parameters, $\mathbf{B}|\mathrm{o}_\mathcal{C},\vartheta$ and $\mathrm{O}_\mathcal{I}|\vartheta$ are jointly normal due to linearisation and error approximation. By taking advantage of the joint normality, we can directly observe the prior predictive distribution of the incomplete records. The PDF follows to read

$$\mathrm{p}(\mathrm{o}_\mathcal{I}|\mathrm{o}_\mathcal{C},\vartheta) = \frac{\exp\left\{-\frac{1}{2}(\mathrm{o}_\mathcal{I}-\bar{\mathrm{O}}_{\mathcal{I}|\mathcal{C}})^\top\Sigma_{\mathcal{I}|\mathcal{C}}^{-1}(\mathrm{o}_\mathcal{I}-\bar{\mathrm{O}}_{\mathcal{I}|\mathcal{C}})\right\}}{\sqrt{(2\pi)^{\mathrm{n}_\mathcal{I}}|\Sigma_{\mathcal{I}|\mathcal{C}}|}} \tag{4.93}$$

and the mean and covariance are given by Eqs. 4.65 and 4.67, implicitly depending on λ, ρ and ε through the first step in our update system. Regarded as a function in ϑ, the marginal likelihood is certainly not normal.

Because of the flat dipole prior, the marginal likelihood for complete records needs special attention. We proceed analogously to the incomplete data, however, having to bear the limiting case

$$p(o_{\mathcal{C}}|\vartheta) = \lim_{\Sigma_1^{-1}\to 0} \frac{\exp\left\{-\frac{1}{2}(o_{\mathcal{C}} - \bar{O}_{\mathcal{C}})^\top \Sigma_{\mathcal{C}}^{-1}(o_{\mathcal{C}} - \bar{O}_{\mathcal{C}})\right\}}{\sqrt{(2\pi)^{n_{\mathcal{C}}}|\Sigma_{\mathcal{C}}|}} . \tag{4.94}$$

We closely follow Rasmussen and Williams (2006, Sec. 2.7) and again split into dipole and non-dipole contributions. As already mentioned in Section 4.4.1, in the limit $\bar{g}_1$ is irrelevant and we set $\bar{O}_{\mathcal{C}} = o_{\mathcal{C}} - \acute{H}_{\mathcal{C}}\tilde{B}_{\mathcal{C}}$ (see Eq. 4.60). The limiting precision matrix is of no concern and given by Equation 4.80. The big concern, however, is the determinant as the dipole variance tends to infinity. According to Rasmussen and Williams (2006, Eq. 2.45), the marginal likelihood for the complete records results in

$$p(o_{\mathcal{C}}|\vartheta) = \frac{\exp\left\{-\frac{1}{2}\tilde{B}_{\mathcal{C}}^\top \acute{H}_{\mathcal{C}}^\top \Omega_{\mathcal{C}} \acute{H}_{\mathcal{C}} \tilde{B}_{\mathcal{C}}\right\}}{\sqrt{(2\pi)^{n_{\mathcal{C}}-3}|\Sigma_{\mathcal{C},\mathrm{ND}}||\Sigma_{1|\mathcal{C}}^{-1}|}} \tag{4.95}$$

and we already computed all relevant quantities in the first step of our update system. Although not explicitly indicated, the non-dipole covariance, the limiting precision and the dipole covariance depend on ϑ.

4.5 Application

As a practical proof of concept we apply the suggested method to the archeomagnetic and volcanic data offered by the GEOMAGIA50.v3 database (Brown et al., 2015). We used all directional and intensity records between 753 AD and 1950 AD that were included in GEOMAGIA50 version 3.3 in November 2019. To simplify matters, the Earth is assumed a sphere of radius $R_E = 6\,371.2$ km and coordinates are treated as if they were spherical. We estimate the committed error to be less than $\frac{1}{2}\,\mu$T which is small compared to modelling uncertainties.

We use the individual, originally reported error estimates. If uncertainties are not available (ca. 8.4 % of the data), we assign $\alpha_{95} = 4.5^\circ$ as directional errors and $\sigma_F = 8.25\,\mu$T as intensity errors Licht et al., 2013, Sec. 2.2. Single unpaired declination records are not used.

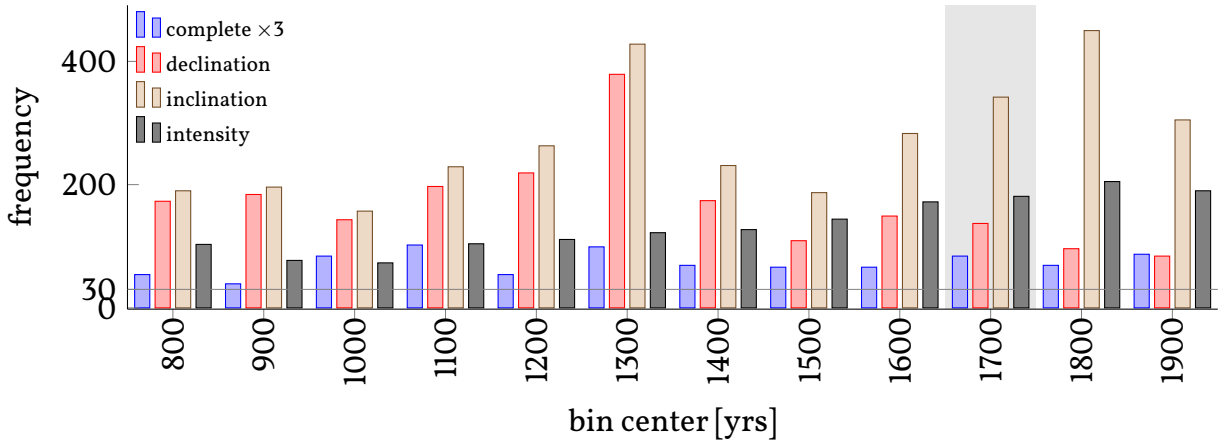


Figure 4.5: Number of data per 100 years bin (D, I and F counts are without complete records). For comparability each complete record is counted as three data points. The bin that we use as example in the discussion is shaded in grey. The horizontal line refers to the criterion of how the window length is chosen.

As our model does not yet account for time dependence we group the data into disjoint bins of 100 years. The decisive factor for the window length is the number of complete records per bin. Our two step strategy and the linearisation are the basic rationale behind. The number of complete records has to determine a reasonable POE for incorporating the incomplete records. Let us assume that at the surface the EMF is dipole dominated (about 90%) for the timespan under consideration. As a rule of thumb, a Taylor expansion performs reasonably well when deviating less than 10% from the *truth*. Having a parametric view back in mind, the field's dominating features may be described by only nine parameters ($\mathbf{g}_1$ and Σ_1). With a minimum of ten complete records per bin we anticipate a coarse field estimate suited as POE. This consideration leads to a window length of 100 years. Figure 4.5 shows the temporal data distribution.

To demonstrate the potential of our modelling strategy in recovering the stationary field we use the interval $[1650, 1750]$ as an example. In total, this bin summarizes 744 observations acquired at 480 sites. Figure 4.6 depicts the highly irregular data coverage which is dominated by the northern hemisphere.

We are going to compare our findings to three previous, continuous magnetic field models. The models ARCH10k.1 (Constable, Korte, and Panovska, 2016), arhimag1k (Senftleben, 2019) and COV-ARCH (Hellio and Gillet, 2018) are considered eligible competitors as they stem from a similar data basis, although arhimag1k additionally includes direct historical observations. All three models report Gauss coefficients up to SH degree $\ell = 10$. ARCH10k.1 and arhimag1k are continuous in time, but do not report modelling errors. In

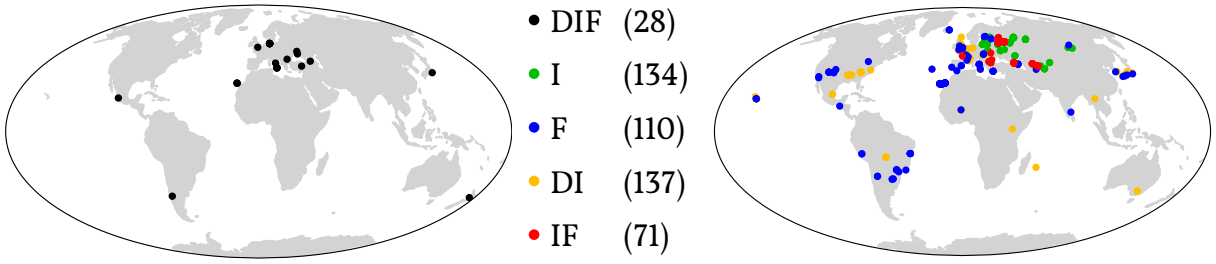


Figure 4.6: Illustration of the very irregular data coverage for the 1700 epoch. The left panel shows complete records whereas the right panel illustrates combinations of D, I and F measurements.

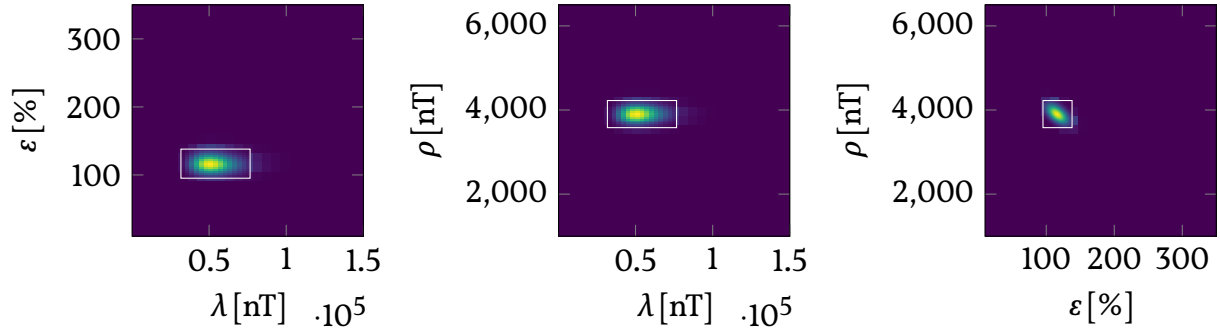


Figure 4.7: Discretisation of the parameter space for the epoch 1700. The accurately sampled marginal densities for the integration phase (white framed box) are drawn atop of the coarse exploratory grid.

contrast, COV-ARCH provides an ensemble of 50 realizations but uses the same coarse time steps as we do. To compare with COV-ARCH's ensemble we calculate sample mean and sample variance. Histograms are computed according to *Scott's rule of thumb* (Scott, 1979).

4.5.1 Numeric Integration

To evaluate the compound distribution $p(\mathbf{B}|o)$ we discretise the integral in Equation 4.86. The posterior PDF is approximated by a simple Riemann sum

$$p(\mathbf{B}|o) \approx \sum_i p(\mathbf{B}|o, \vartheta_i)\, p(\vartheta_i|o) \Delta\vartheta_i \,. \tag{4.96}$$

Although not highly accurate, we favour using a Riemann sum because of its simplicity. Since by construction the mixture components $p(\mathbf{B}|o, \vartheta_i)$ are normally distributed, the r.h.s. of Equation 4.96 gives a (finite) Gaussian mixture distribution (Murphy, 2012, Sec. 11.2.1), for which moments are readily available by translating Equations 4.90 and 4.91 accordingly.

To actually discretise the parameter space, we use a regular and equally spaced grid. Thus we have the posterior

$$p(\mathbf{B}|o) \approx \sum_{i,j,k=0}^{N} p(\mathbf{B}|o, \lambda_i, \varepsilon_j, \rho_k)\, p(\lambda_i, \varepsilon_j, \rho_k|o) \cdot \Delta\lambda\Delta\varepsilon\Delta\rho\,, \tag{4.97}$$

where N^3 is the number of grid points for the Riemann sum, $\Delta\lambda$ is the step width, λ_0 and λ_N specify the bounds of the interval we integrate over and $\lambda_i = \lambda_0 + i\Delta\lambda$ (Similar for ε and ρ). As mentioned above, we carry out a second quadrature to normalize the parameter posterior $p(\lambda_i, \varepsilon_j, \rho_k|o)$. This also serves the purpose of having a well defined Gaussian mixture density. We normalize it with respect to the Riemann sum, i.e. such that

$$\sum_{i,j,k=0}^{N} p(\lambda_i, \varepsilon_j, \rho_k|o)\Delta\lambda\Delta\varepsilon\Delta\rho = 1\,. \tag{4.98}$$

Under these considerations Equation 4.97 can be seen as the discrete version of Equation 4.86, where all terms are replaced by their respective discrete equivalent.

We consider the approximation (Eq. 4.96) reasonable, since we found the probability mass of the parameter posterior $p(\vartheta|o)$ to be unimodal and localized in a finite region. Thus to calculate the Gaussian mixture proxy (Eq. 4.97) for the compound distribution, we perform two steps:

Exploration of the parameter space We need to identify the region in which the parameter posterior is localized. Hence we span a coarse grid over all values we believe to be (physically) reasonable and calculate the posterior $p(\vartheta|o)$ on this grid. We choose $N = 25$ grid points along each of the three dimensions and choose the bounds as

$$\begin{aligned} \lambda_0 &= 100\ \text{nT} & \lambda_N &= 150000\ \text{nT} \\ \varepsilon_0 &= 10\ \% & \varepsilon_N &= 350\ \% \\ \rho_0 &= 1000\ \text{nT} & \rho_N &= 6500\ \text{nT}\,. \end{aligned} \tag{4.99}$$

Although in principle a wider extent of the grid may be “physically reasonable”, by trial and error we found these bounds to be sufficient (i.e. outside of the bounds the probability mass was approximately zero for all considered cases). Finally, from the posterior we calculate the marginal distribution for each parameter λ, ε, ρ via another Riemann

sum. We then calculate each (empirical) mean and standard deviation from these coarse marginals. The region where the probability mass is located is then covered by the cuboid centred at the empirical mean with edge-lengths given by two empirical standard deviations.

Calculation The actual numerical integration only takes place within the cuboid that is derived in the exploration step. Inside that cuboid a refined grid is spanned with N = 15 nodes per dimension. By calculating not only the posterior on this new grid, but running for each grid point an inversion for the EMF, the Gauss coefficients and other quantities of interest, we can calculate a proxy for the full compound posterior for each of these quantities, using Equation 4.97. This proxy, which is a Gaussian mixture distribution, is the final result of our modelling strategy.

For all epochs the parameter posterior is of good-nature only featuring a single mode. For epoch 1700, Figure 4.7 depicts the posterior parameter PDF. Presented are all combinations of 2D marginal parameter posterior at the coarse and refined grid. The white rectangle refers to the edges of the refined grid that is used for numeric integration. The maximum of the error level ε is slightly above 100% at a rather small spread. Taking the approximations into account, a shift toward higher values seems reasonable. The residual term ρ has its maximum at about 4 μT and is also relatively sharp. It is interesting to note that the residual is of the same order of magnitude as the equatorial dipole. This need not only be due to unexplained sources, it can also be an effect of the linearisation or not taking time dependencies into account. The scaling parameter λ features the widest distribution. In comparison with the IGRF, the magnitude and range appear reasonable. However, the posterior PDF of the model parameters is not sharp enough to justify a point estimate. Therefore we integrate out model parameters and thus the uncertainties from the parameter posteriors are translated into the posterior of the quantities of interest. This way the posterior variance, which is easily available due to the Gaussian mixture structure, does not only reflect uncertainties arising from the data, but also reproduces the model uncertainties.

4.5.2 Vector field predictions

The discretised versions of $\mathbf{B}$'s posterior mean and variance translate according to Equations 4.90 and 4.91 and are given by

$$\mathbb{E}\left[\mathbf{B}|o\right] = \sum p(\vartheta_i|o)\mathbb{E}\left[\mathbf{B}|o,\vartheta_i\right]\Delta\vartheta \tag{4.100}$$

$$\mathbb{V}\left[\mathbf{B}|o\right] = \sum p(\vartheta_i|o)\left(\mathbb{E}\left[\mathbf{B}|o,\vartheta_i\right]^2 + \mathbb{V}\left[\mathbf{B}|o,\vartheta_i\right]\right)\Delta\vartheta - \mathbb{E}\left[\mathbf{B}|o\right]^2 . \tag{4.101}$$

The pointwise posterior standard deviation serves for realistic location dependent uncertainty estimates. Again, **B**-field predictions are non-parametric and do not depend on Gauss coefficients.

The top row of Figure 4.8 depicts both posterior mean and standard deviation of the down component at the Earth's surface. The field is evaluated at 2000 design points, equally distributed over the sphere (Deserno, 2004). We are able to quantify what previous studies are suggesting: The EMF is reconstructed with confidence within areas of dense data coverage, e.g. in Europe. Structures of large parts of the southern hemisphere, however, remain vague. We see quite similar patterns across all epochs under investigation.

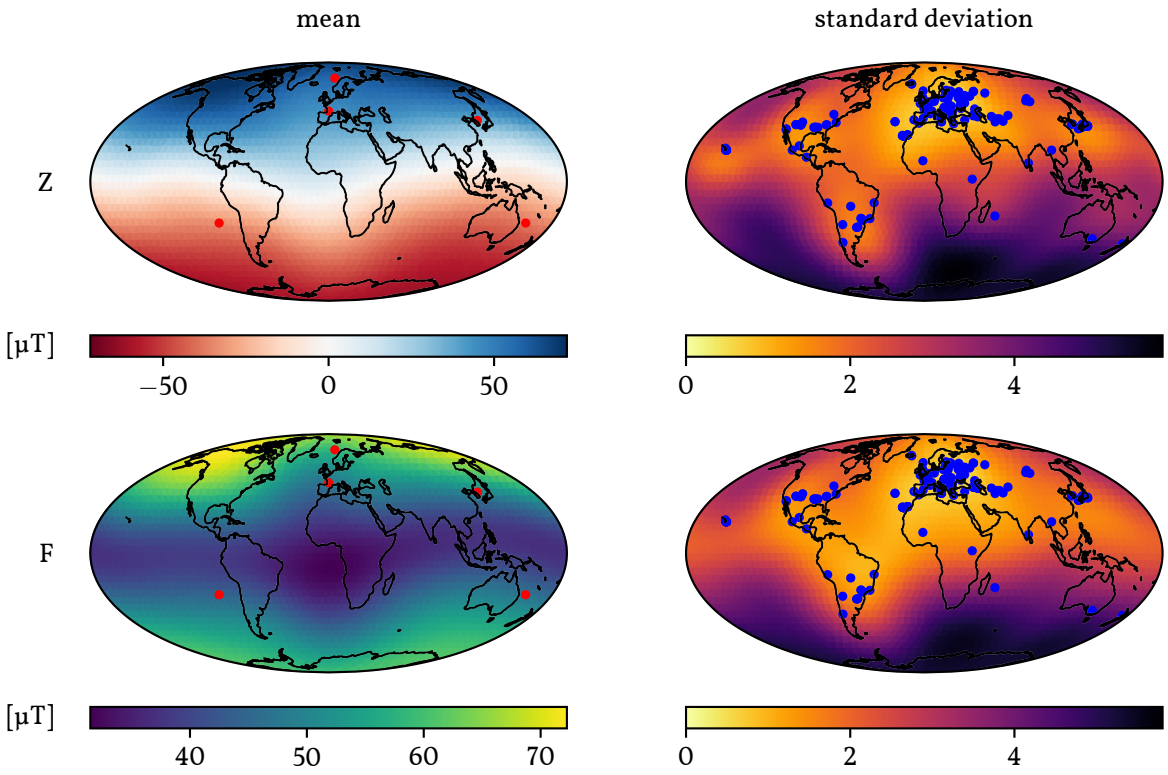


Figure 4.8: Posterior mean (left) and standard deviation (right) of the EMF at the Earth's surface for the 1700 epoch. The top row shows the down component B_Z. The locations at which we perform normality tests are indicated in red (see Tab. 4.1). The bottom row shows the field intensity F. Blue dots refer to the underlying data sites.

4.5.3 Declination, Inclination and Intensity

Because the EMF shares a non-linear relation with declination, inclination and intensity we do not see the possibility of analytically deriving properties such as the posterior mean for D, I and F. For a moderate number of design points we can obtain the distribution by sampling strategies. However, trying to draw from a high dimensional normal mixture distribution appears absurd. For the 2000 design points we are interested in, storing all the covariance matrices of the entire parameter grid is hardly possible, let alone drawing samples from the resulting mixture.

The idea behind uncovering the posterior mean and variance of D, I and F is once more a linearisation. To do so the same strategy as pointed out in Section 4.2.3 is used. The linearisation of D, I and F are given by Equations 4.48, 4.49 and 4.50. We build upon the EMF estimate from the previous section and the posterior mean $\mathbb{E}\left[\mathbf{B}|\mathbf{o}\right]$ serves as POE. Because of the linearisation approximations for mean and (co)-variance are explicit.

To provide an example we consider the intensity, only. Utilizing Equation 4.50, the approximate posterior mean and variance are given by

$$\mathbb{E}\left[\mathrm{F}|\mathrm{o}\right] \approx \| \, \mathbb{E}\left[\mathbf{B}|\mathbf{o}\right] \, \| \tag{4.102}$$

$$\mathbb{V}\left[\mathrm{F}|\mathrm{o}\right] \approx \frac{\mathbb{E}\left[\mathbf{B}|\mathbf{o}\right]^{\top} \mathbb{V}\left[\mathbf{B}|\mathbf{o}\right] \mathbb{E}\left[\mathbf{B}|\mathbf{o}\right]}{\| \, \mathbb{E}\left[\mathbf{B}|\mathbf{o}\right] \, \|^2} \, . \tag{4.103}$$

We pursue a non-parametric approach but the concept is similar to that of Hellio et al. (2014, Appendix A). For the 1700 epoch, the bottom row of Figure 4.8 depicts mean and standard deviation of the intensity at the Earth's surface. Although not directly observed by a single record, the evolving area of weak field known as the South Atlantic Anomaly (e.g., Mandea et al., 2007; Hartmann and Pacca, 2009) is certainly significant within one standard deviation. Looking at the other epochs, the westward drift is also visible. This feature is constrained here by only a few points around the actual anomaly, which shows that our modelling approach is capable of uncovering features that are known from models with stronger data basis. This ability stems from long ranging spatial correlations of the kernel and is controlled by the reference radius R (see Figure 4.1).

When comparing epoch 800 with figure 11(a) in Hellio and Gillet (2018) two things stand out. The magnitudes are similar but a low intensity patch is found in the Pacific rather than the Atlantic. In the standard deviation we also see a rather sharp transition between north and south. Since this is a proof of concept a detailed comparison or even an interpretation would be premature.

Since our posterior distributions are certainly not normal it is questionable if the standard deviation is adequate to describe uncertainties. If the posterior is highly skewed or even multi modal, the standard deviation would not be well suited to quantify modelling errors. One would preferably use percentiles which, however, cannot be derived analytically. As long as the number of design points is moderate, one brute-force method is to calculate percentiles by sampling. This works well for percentiles that are not far from the bulk of the probability mass.

A simplistic approach to obtain samples from a mixture distribution is the following algorithm. First, by chance the k^{th} random variable is selected from the mixing distribution $p(\vartheta_k|o)$ (categorical). Then the value of the selected random variable $p(\mathbf{B}|o, \vartheta_k)$ is realized (multivariate normal). Repeat until the desired amount of samples is achieved.

Although we have to radically reduce the number of design points we can gain insight into how strong the proxy standard deviation and percentiles deviate. We compare 16- and 84-percentiles to mean $\pm$ standard deviation via

$$\Delta_{16} = 1 - \frac{\mu - \alpha_{16}}{\sigma} \qquad \text{and} \qquad \Delta_{84} = 1 - \frac{\alpha_{84} - \mu}{\sigma} \tag{4.104}$$

where α refers to percentiles, μ and σ indicate mean and standard deviation. The interpretation of Δ is as follows: If Δ is larger (smaller) than zero, the standard deviation overestimates (underestimates) the uncertainty given by the percentile, by $|\Delta|$ standard deviations. Note, that to keep this interpretation consistent for both Δ_{16} and Δ_{84} the sign of the second term in (4.104) changes.

At 5 randomly selected locations we apply this check to the down component B_Z and to the three commonly used archeomagnetic observables D, I and F. As can be seen in Table 4.1, at all locations the deviation is well below 10%. Thus we believe the proxy standard deviation is qualified to describe uncertainties. Furthermore, the proxy standard deviation obtained by linearisation is computationally feasible and easy to visualize and interpret.

4.5.4 Predictions at the core-mantle boundary

Until here predictions were carried out at the Earth's surface. It is straightforward to predict the EMF at arbitrarily chosen design points outside of the reference sphere of radius R. Figure 4.9 presents mean and standard deviation of the down component B_Z at the CMB. We again use 2000 equidistributed design points but at radius $R_{CMB} = 3480$ km.

	Lat	-26°	45°	71°	39°	-26°
	Lon	-88°	0°	9°	131°	159°
D	Δ_{16}	2.5	0.7	0.3	-0.2	1.5
	Δ_{84}	-0.3	2.9	3.9	1.8	0.4
I	Δ_{16}	3.6	-0.3	-0.4	-1.5	3.8
	Δ_{84}	-3.0	1.0	3.0	6.7	-2.3
F	Δ_{16}	4.6	-2.4	0.7	1.5	7.1
	Δ_{84}	-1.7	0.5	-0.1	-0.3	0.6
Z	Δ_{16}	2.9	-2.6	-1.1	-0.4	1.2
	Δ_{84}	1.6	0.1	0.6	0.6	1.9

Table 4.1: Estimation of the quality of the proxy standard deviation as a measure of error, compared to the 16-/84-percentiles. If Δ is larger (smaller) than 0, the standard deviation overestimates (underestimates) the error. Values are given in percent of one standard deviation.

One observes that uncertainties are greater than at the Earth's surface. Roughly speaking, the standard deviation is about 40% compared to the mean, while at the Earth's surface relative errors only amount up to 10% (see right column of Figure 4.8). The modelling error strongly depends on how far design points are from the reference sphere.

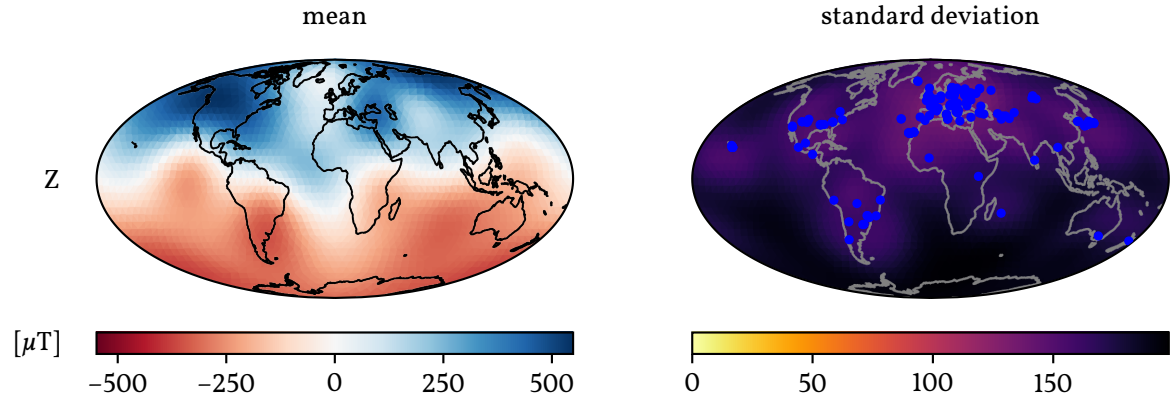


Figure 4.9: Posterior mean (left) and standard deviation (right) of the EMF down component B_Z at the CMB for the 1700 epoch. Blue dots refer to the sites where records are acquired.

This behaviour is best understood thinking in terms of Gauss coefficients. In Equation 4.25 the term that is raised to the power of $\ell+1$ causes this effect. Since design points lie outside the reference sphere, the ratio $\frac{R}{|x|}$ is smaller than one. Thus, Gauss coefficients are more *penalized* the larger the SH degree, and the ratio $\frac{R}{|x|}$ determines the rate of de-

cline. In turn, the closer to the reference radius the more impact higher SH degrees have. Related to smaller structures, the higher the SH degree the less certain we are. When scaling down – e.g. to the CMB – higher SH degrees increase and so do the respective uncertainties. This idea translates to our non-parametric model.

Compared to what Hellio and Gillet (2018, figure 8c) found, our reconstruction of epoch 1400 looks very similar. Because of the large variabilities, a detailed comparison does not make much sense.

4.5.5 Gauss Coefficients

We want to stress again that our model is inherently non-parametric. The fundamental quantity of the inference is the geomagnetic potential. Nevertheless, our approach allows to infer Gauss coefficients since geomagnetic potential and Gauss coefficients are linearly related (see Eq. 4.40). The procedure is similar to inferring the EMF. The discretised versions of the compound mean (Eq. 4.90) and (co)-variance (Eq. 4.91) for the Gauss coefficients g_ℓ^m read

$$\mathbb{E}\left[g_\ell^m|o\right] = \sum p(\vartheta_i|o)\mathbb{E}\left[g_\ell^m|o,\vartheta_i\right]\Delta\vartheta \tag{4.105}$$

$$\mathbb{V}\left[g_\ell^m|o\right] = \sum p(\vartheta_i|o)\left(\mathbb{E}\left[g_\ell^m|o,\vartheta_i\right]^2 + \mathbb{V}\left[g_\ell^m|o,\vartheta_i\right]\right)\Delta\vartheta - \mathbb{E}\left[g_\ell^m|o\right]^2 . \tag{4.106}$$

In principle we can predict up to arbitrary SH degree. However, storing the component's mean and covariances for the whole parameter grid becomes memory intense. If we restrict our selves to a moderate SH degree – e.g. $\ell \leq 10$ – we are able to calculate the full mixture PDF, sample from the posterior and calculate percentiles.

Although not explicitly indicated, Gauss coefficients depend on the reference radius R. Given the g_ℓ^m at the reference radius, we can scale them to a radius $\tilde{R} > R$ by

$$\tilde{g}_\ell^m = g_\ell^m \left(\frac{R}{\tilde{R}}\right)^{\ell+2} . \tag{4.107}$$

Mean and co-variances translate accordingly.

For comparison, we scale Gauss coefficients with respect to the Earth's surface. The mixture PDFs for the dipole coefficients are shown in Figure 4.10. As can be seen, the mixture is quite close to the moment matching normal proxy. Across all epochs the dipole coefficients are close to being normally distributed, except for the bin $[850, 950]$, for which the mixture distribution deviates slightly from the normal proxy. However, the higher the SH degree the less normal the posterior is and the more the scale invariant prior dominates (see Eq. 4.89). We further see that for the epoch under consideration, the histogram

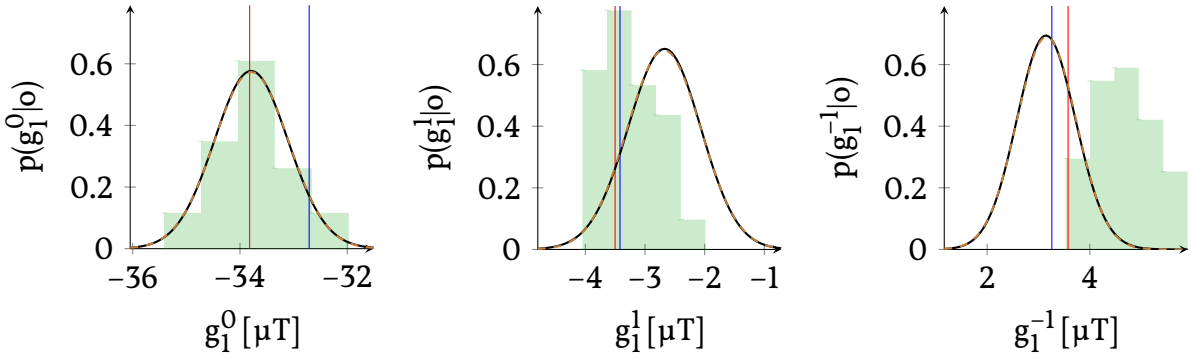


Figure 4.10: Distribution of the dipole coefficients at the Earth's surface for epoch 1700. The resulting mixture PDF is indicated by the solid black lines whereas the, nearly identical, dashed lines refers to the moment matching Gaussian proxies. The green histograms illustrates the COV-ARCH ensemble coefficients. The blue vertical lines refer to arhimag1k and the red ones to ARCH10k.1 .

built from the COV-ARCH ensemble is in good agreement with the mixture and the mean aligns with the ARCH10k.1 estimate. It is not too much of a surprise that the arhimag1k prediction deviates since it additionally incorporates historical records, which are not included in our dataset.

Although well suited to compare with existing models, care has to be taken when interpreting Gauss coefficients. A finite set of Gauss coefficients does not represent the full information contained in our non-parametric modelling approach. Areas of dense data coverage may feature a resolution than can not be captured by an expansion e.g. until degree 10. However, this is less relevant in the context of geomagentic core field modelling, where SH degrees larger than around 14 are dominated by lithospheric field signals, which cannot be resolved by a sparse data distribution. Figure 4.11 compares the predicted Gauss coefficients to the selected reference models until SH degree 5, which is considered the approximate global resolution of the spherical harmonics based models (Korte, Donadini, and Constable, 2009; Licht et al., 2013; Sanchez et al., 2016; Constable, Korte, and Panovska, 2016; Hellio and Gillet, 2018). With few exceptions, our findings are on a par with existing models within one standard deviation.

4.5.6 Spatial Power Spectrum

It is hard to digest all the information contained in a collection of Gauss coefficients. Therefore, it has become common to consider the geomagnetic power spectrum Backus, Parker, and Constable, 1996, Sec. 4.4.2, that reflects the contributions of different spatial wavelengths in the SHs. For degree ℓ the corresponding wavelength is $\lambda_\ell \approx \frac{4\pi R}{2\ell+1}$ Langel

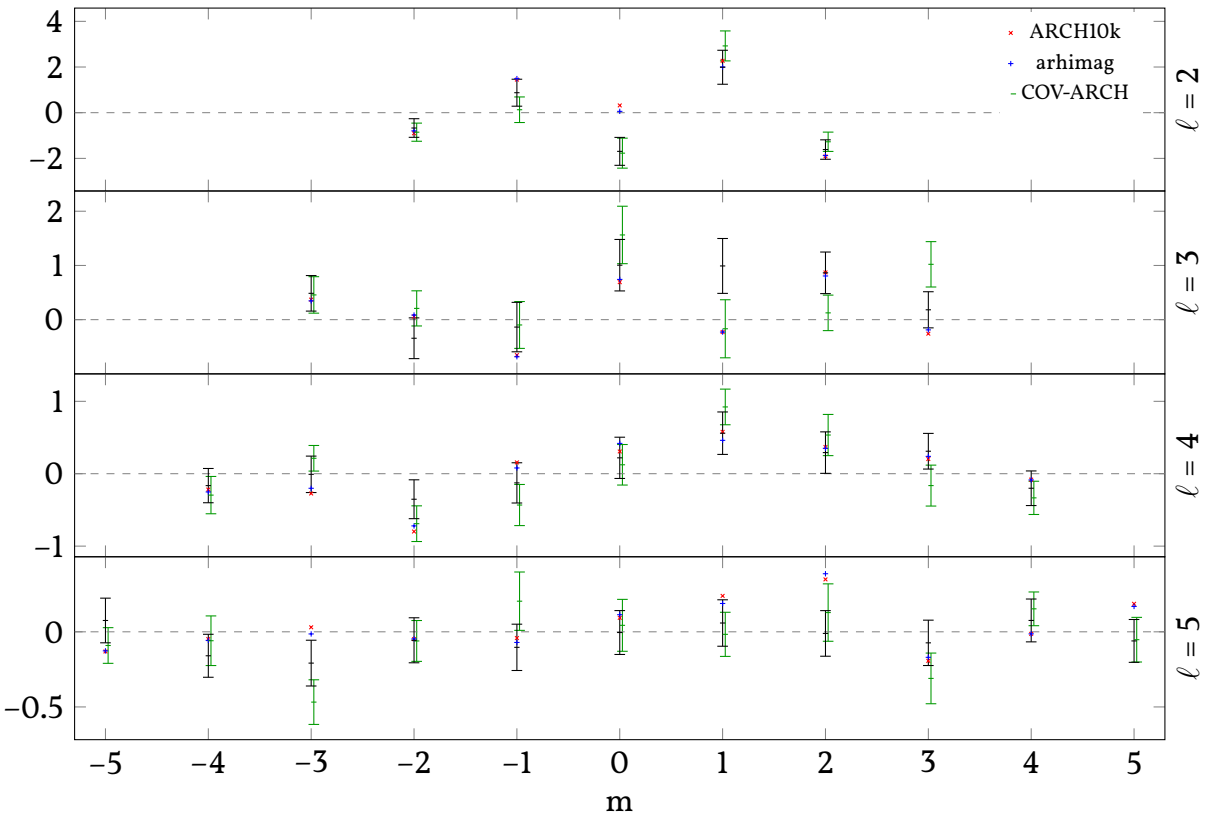


Figure 4.11: Comparison of Gauss coefficients for SH degrees 2 to 5 at Earth's surface for epoch 1700. Results from this study are shown in black. Error bars indicate one standard deviation.

and Hinze, 1998, Sec 4.3.5. The EMF is apportioned such that

$$\mathbf{B} = \sum_{\ell} \mathbf{B}_{\ell} = -\nabla \sum_{\ell} \Phi_{\ell} \tag{4.108}$$

where

$$\Phi_{\ell}(\mathbf{x}) = R \left(\frac{R}{|\mathbf{x}|} \right)^{\ell+1} \sum_{-\ell \le m \le \ell} g_{\ell}^{m} \, Y_{\ell}^{m}(\hat{\mathbf{x}}) \, . \tag{4.109}$$

The components Φ_{ℓ} are certainly orthogonal, since SHs form an orthogonal system. A characteristic that describes $\mathbf{B}_{\ell}$'s variations is the so-called *average square value*. Since $\mathbf{B}$ is divergence free, the net total flux through a closed surface is zero. Thus, the *average* over the sphere vanishes

$$\langle \mathbf{B}_{\ell} \rangle_R = 0 \, , \tag{4.110}$$

where the angle brackets are an abbreviation for the surface integral (see Eq. 4.40). The definition of the average can be extended to the (centred) average square value

$$\left\langle (\mathbf{B}_{\ell} - \langle \mathbf{B}_{\ell} \rangle_R)^2 \right\rangle_R = \left\langle \mathbf{B}_{\ell}^2 \right\rangle_R \, . \tag{4.111}$$

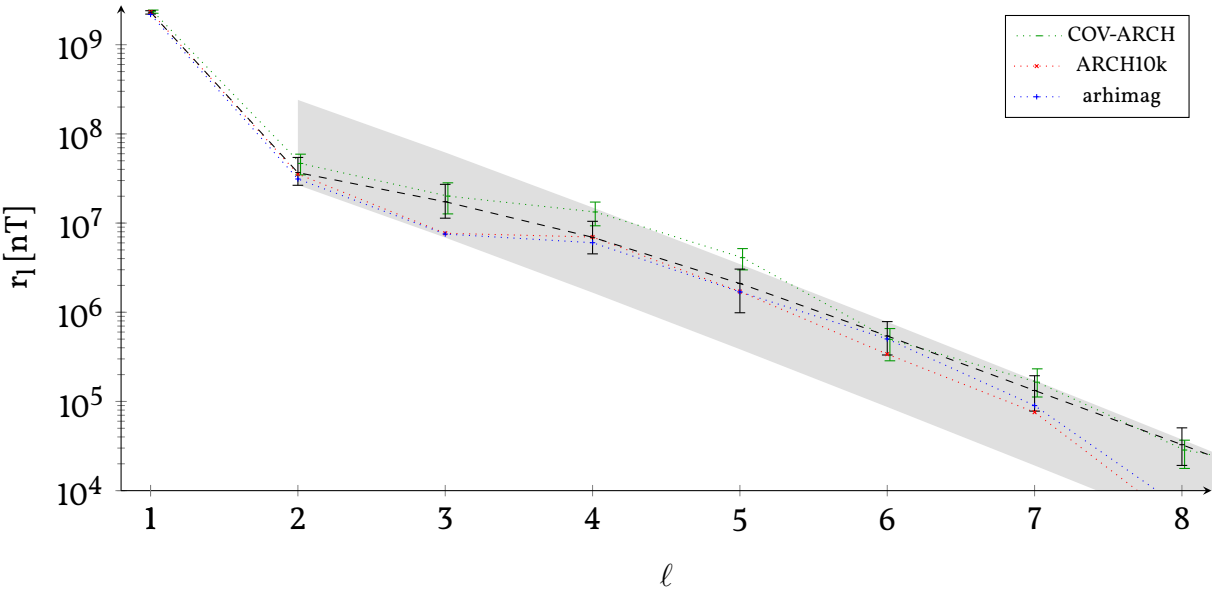


Figure 4.12: Power spectrum at the Earth' surface for epoch 1700. The mean of the posterior power spectrum is indicated by the dashed line. The top and bottom of the error bars refer to 16 and 84 percentiles, respectively. Except for a constant offset, the grey shaded area indicates the slope of the a priori power spectrum.

Due to Parseval's Theorem Backus, Parker, and Constable, 1996, Eq. 4.4.21, the average square can be expressed in terms of Gauss coefficients. Regarded as a function in ℓ, the quantity

$$r_\ell := \left\langle \mathbf{B}_\ell^2 \right\rangle_R = (\ell+1) \sum_{-\ell \le m \le \ell} (g_\ell^m)^2 \tag{4.112}$$

is called the *geomagnetic power spectrum* (Lowes, 1974). Again, Gauss coefficients g_ℓ^m depend on the reference radius and so does r_ℓ.

Within the setting of statistical inversions, calculating the power spectrum requires special attention since uncertainties have an appreciable effect. Squaring and summing normal distributed Gauss coefficients gives a random variable (RV) that is distributed as the sum of weighted non-central χ^2 RVs. Unfortunately, for the PDF of a linear combination of non-central chi-square RVs no closed, analytic expression is known (Bausch, 2013). Nonetheless, using the algebraic formula for the variance $\mathbb{E}[XY] = \mathrm{Cov}[X,Y] + \mathbb{E}[X]\,\mathbb{E}[Y]$ we obtain an expression for the expectation of the power spectrum

$$\mathbb{E}[r_\ell | o] = (\ell+1) \sum_m \left(\mathbb{E}[g_\ell^m | o]^2 + \mathbb{V}[g_\ell^m | o] \right) \tag{4.113}$$

and it is obvious that variances play an important role. In other words, the larger the uncertainties the bigger the impact on the spectrum's mean.

Although the second moment may be accessible, the standard deviation is not suited to quantify errors. Roughly speaking, if standard deviations of the g_ℓ^ms dominate over the mean, the PDF of r_ℓ is highly skewed with wide tails. In turn, a moment matching Gaussian proxy would violate the positivity constraint of r_ℓ. For an increasing SH degree this is certainly the case. Percentiles are better suited to estimate the error level, but impossible to access analytically. We again calculate the distribution and percentiles empirically by brute-force sampling Gauss coefficients from the posterior. In Section 4.5.3 the sampling strategy is described.

For epoch 1700 the resulting power spectrum is shown in Figure 4.12. Within the error margins our findings and the COV-ARCH model are in good agreement. Differences arise comparing with ARCH10k.1 and arhimag1k. Both models report less power at degree $\ell = 3$ and feature a rapid loss for degrees $\ell \geq 8$. While the degree 3 deviation might be due to differences in the underlying data basis, the latter likely is caused by the influence of the global regularization in the spherical harmonic models. Nonetheless, for $\ell \leq 8$ ARCH10k.1 and arhimag1k are potential realizations from what we find.

4.5.7 Dipole Moment

To a first approximation, the EMF is dipolar. This is, its shape is similar to that of a hypothetical bar magnet placed at the centre of the Earth. With respect to a Cartesian coordinate frame, the corresponding vector dipole moment is given by

$$\mathbf{m} = \frac{4\pi R_E^3}{\mu_0}(g_1^1\hat{\mathbf{x}} + g_1^{-1}\hat{\mathbf{y}} + g_1^0\hat{\mathbf{z}})\,, \tag{4.114}$$

where $\mu_0 \approx 4\pi \cdot 10^{-7}$ [Tm/A] refers to the permeability of free space (Backus, Parker, and Constable, 1996, Eq. 4.4.17). Vector components are indicated by subscripts x, y and z, e.g. $m_x \propto g_1^1$. Considering the discretisation, the vector dipole moment $\mathbf{m}$ is a Gaussian mixture borrowing its statistical properties from the dipole coefficients. The dipole's magnitude can be computed directly via

$$\tau = \|\mathbf{m}\| = \frac{4\pi R_E^3}{\mu_0}\sqrt{(g_1^1)^2 + (g_1^{-1})^2 + (g_1^0)^2} \tag{4.115}$$

and is proportional to the square root of the power r_1. We are interested in statistical properties of τ. Because of the square root, we are not able to derive an analytic expression for $\mathbb{E}\left[\tau|o\right]$. Nonetheless, if we knew $\mathbb{E}\left[\tau|o\right]$, the variance is right at hand

$$\mathbb{V}\left[\tau|o\right] = \mathbb{E}\left[\tau^2|o\right] - \mathbb{E}\left[\tau|o\right]^2\,, \tag{4.116}$$

since $\mathbb{E}\left[\tau^2\right] \propto \mathbb{E}\left[r_1\right]$, which is given by Eq. 4.113. To obtain a proxy of $\mathbb{E}\left[\tau|o\right]$ and to calculate the empiric distribution we again use sampling. An ordinary Gaussian kernel density estimate is used to smoothen the histogram Murphy, 2012, Sec. 14.7.2. The bandwidth is selected by *Scott's rule of thumb* which – due to its simplicity – strongly influences the estimate.

For the epoch of choice, Figure 4.13 compares the density estimate, the according histogram and the Gaussian proxy. The density estimate looks rather normal and is approximated well by the moment matching Gaussian. This is also the case for all other epochs in our study. At least for the epoch 1700 our findings agree with ARCH10k.1 and COV-ARCH. Presumably, due to its stronger data basis arhimag1k deviates. However, the agreement varies through epochs.

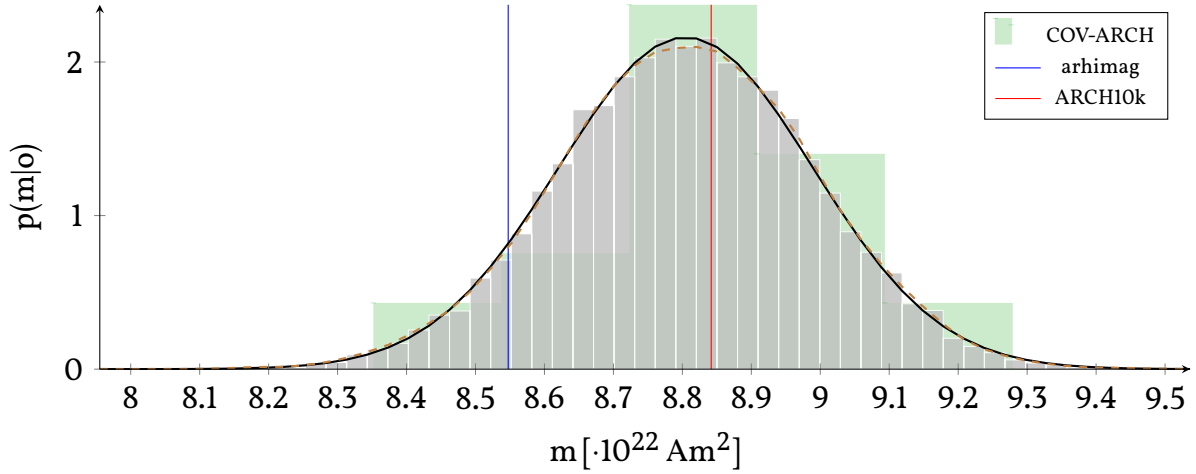


Figure 4.13: Distribution of the dipole moment for epoch 1700 based on 10000 drawn samples. The kernel density estimate is indicated by the red dashed line, that is nearly identical to the Gaussian proxy (black line).

As the dipole moment is known to change with time, Figure 4.14 displays the time series of all epochs under consideration. Compared to COV-ARCH, ARCH10k.1 and arhimag1k, we find a similar temporal evolution of the dipole moment. Presumably the most recent epoch deviates from the other models, since we do not include historical information in our model. Especially for the earlier epochs we see slightly higher intensities than reported by existing models. The strong deviation of epochs 1100 and 1500 may be caused by outliers, which we did not test for so far, but it is also possible that the variations are real and earlier models underestimate strong variations due to the treatment of dating uncertainties and temporal regularization. It is beyond the scope of our proof-of-concept model to resolve this question. For both epochs the deviation is caused by g_1^0. The parameter distributions $p(\vartheta|o)$ for these epochs do not show any noticeable problems. However, epoch 1500 features a rather weak data basis.

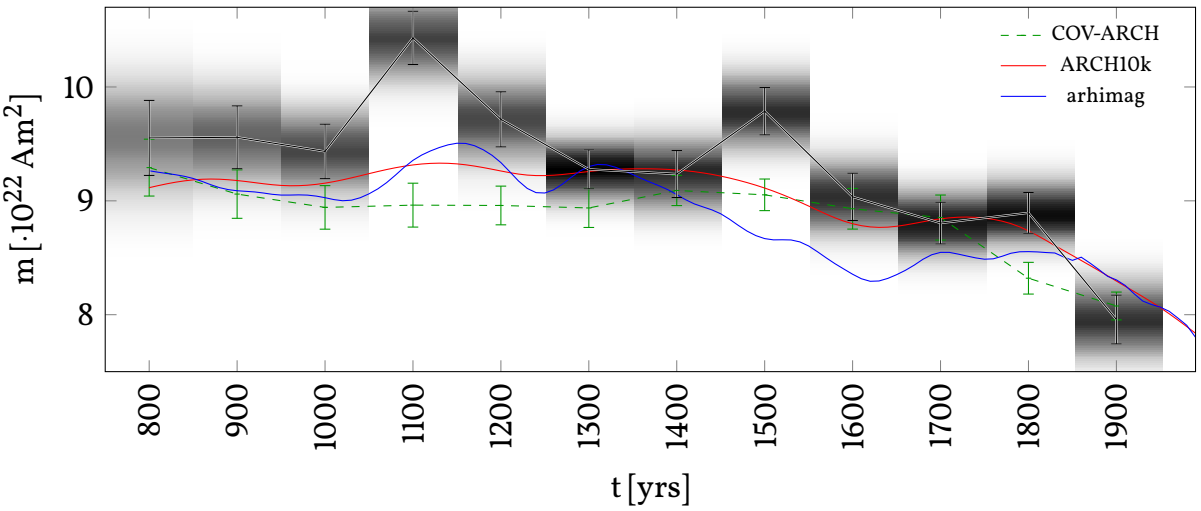


Figure 4.14: Temporal variation of the dipole moment magnitude based on 10000 drawn samples. The posterior means are connected by black lines and the error bars indicate one standard deviation. The gray shaded background refers to the kernel density estimate. Results from models COV-ARCH, ARCH10k and arhimag are shown for comparison.

We are further interested in the distribution of the dipole's north pole a.k.a. *geomagnetic north*. The geomagnetic north pole is given as the antipode of the projection of the vector dipole moment onto the sphere. In other words, the intersection of the axis of the hypothetical bar magnet with the Earth's surface. The location w.r.t. spherical coordinates is given by

$$\theta_m = \arccos\left(-\frac{m_z}{\tau}\right), \qquad \varphi_m = \arctan\left(\frac{-m_y}{-m_x}\right) \tag{4.117}$$

where θ refers to co-latitude and φ to longitude. We proceed with a similar approach as presented in Khokhlov, Hulot, and Bouligand, 2006 and translate the dipole's PDF that is interpreted w.r.t. a Cartesian reference frame into spherical coordinates. The PDF of the vector dipole moment transformed to spherical coordinates is given by

$$p\left(\begin{pmatrix}\tau\\ \theta_m\\ \varphi_m\end{pmatrix}\middle| o\right) = p\left(\begin{matrix}m_x = \sin\theta_m\cos\varphi_m\\ m_y = \sin\theta_m\sin\varphi_m\\ m_z = \cos\theta_m\end{matrix}\middle| o\right)\tau^2\sin\theta_m\,, \tag{4.118}$$

where we made use of the change of variables theorem (Murphy, 2012, Sec. 2.6). To obtain the distribution of the location we have to marginalize the magnitude from the mixture distribution

$$\mathrm{p}\left(\begin{pmatrix}\theta_\mathrm{m}\\ \varphi_\mathrm{m}\end{pmatrix}\middle|\mathrm{o}\right) = \sum_\mathrm{i} \mathrm{p}(\vartheta_\mathrm{i}|\mathrm{o}) \int_0^\infty \mathrm{p}\left(\begin{pmatrix}\tau\\ \theta_\mathrm{m}\\ \varphi_\mathrm{m}\end{pmatrix}\middle|\mathrm{o}, \vartheta_\mathrm{i}\right) \mathrm{d}\tau \,. \tag{4.119}$$

Since individual dipole coefficients are normally distributed, we can analytically solve the integral. To do so, factor out m, complete the square and marginalize via the standard Gaussian integral equations (Owen, 1980, Eqs. 10, 11 and 12). We deliberately skip the resulting expression as it is lengthy and of no particular interest.

However, to visualize the wander of geomagnetic north it is useful to numerically evaluate the resulting PDF. As can be seen in Figure 4.15 the general movement is similar to the models we are comparing with. Again the 1500 epoch deviates and is ahead of the comparison path, which may be caused by outliers as stated previously. For comparison we only show COV-ARCH's sample mean because a scatter plot of the complete ensemble leads to an overstuffed picture. The ensemble covariance features similar magnitudes compared to what we find.

4.6 Conclusions and Perspectives

The extensive theory of Sections 4.2, 4.3 and 4.4 build the foundation of a new modelling strategy for archeo- and paleomagnetic field models. The key advantage of this probabilistic approach is that realistic modelling uncertainty estimates are obtained, e.g., via the standard deviation. The a priori distributions we choose have pros and cons. Taking the least subjective choice is an advantage as we a priori do not specify any preferred direction and, thus, our method is even well suited for time periods featuring reversals. However, a shortcoming – in particular concerning the prior dipole – is that we could not visualize a comparison of prior and posterior uncertainties.

Besides the a priori covariance structure and the weakly informative parameter priors, our modelling strategy depends only on a single parameter, the reference radius R. Conceptually, it is no problem to also integrate out this remaining parameter. Only a technicality arises if we want to integrate out R, Gauss coefficients require scaling to a common radius – e.g. the Earth's surface – as they depend on R. The limiting factor is a four dimensional parameter space with excessive computational demands.

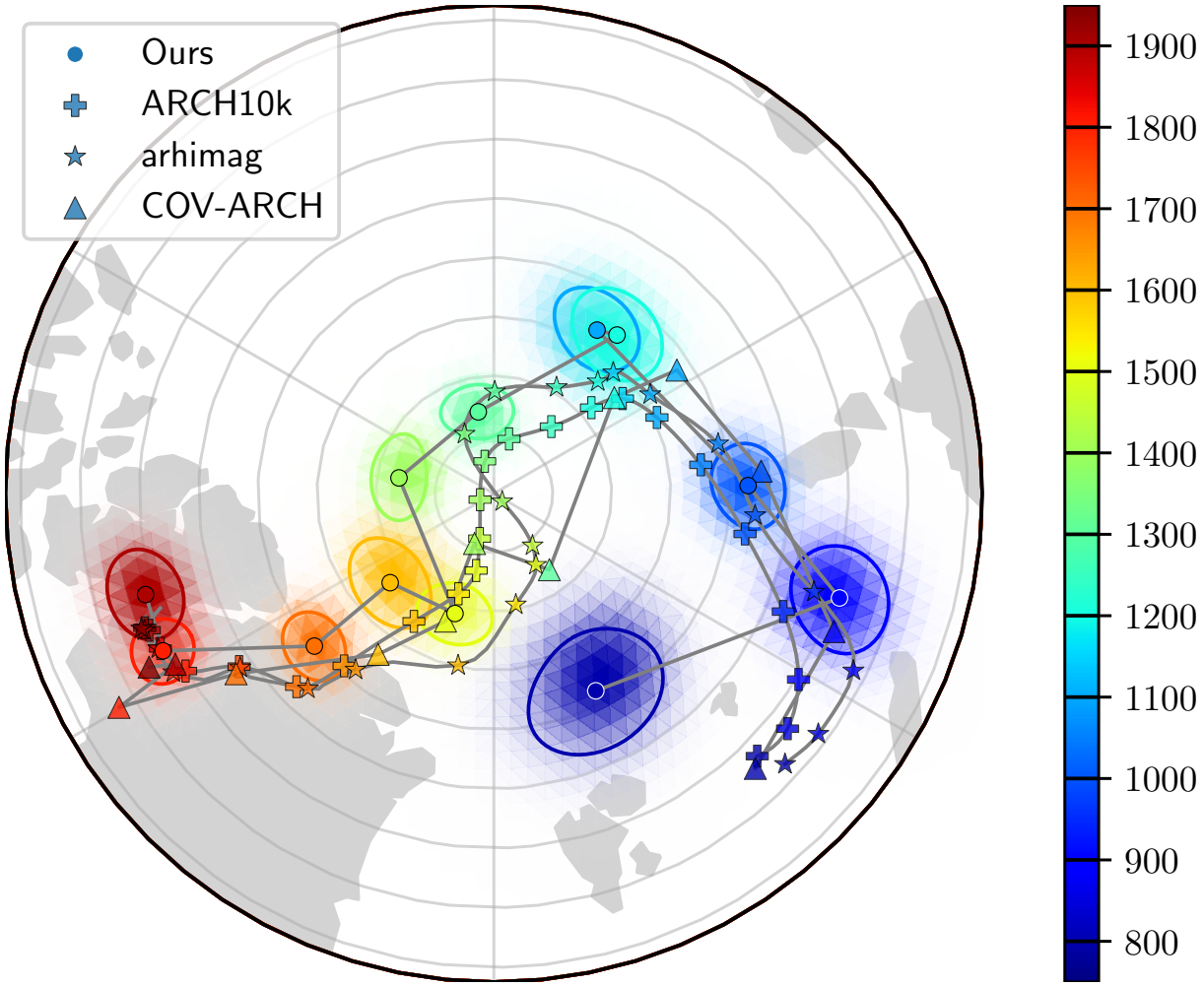


Figure 4.15: Wander of the geomagnetic north pole. For each epoch in the study, the probability distribution of the location is shown with the mean marked by dots and one sigma marked by an ellipse. The colour varies according to time. For comparison, wander paths for the reference models are shown, without uncertainties in order not to clutter the figure.

To save the effort of implementing the exhaustive theory, a ready-to-use software suite called CORBASS (Schanner and Mauerberger, 2019) was developed as part of the project. CORBASS is written in python and licensed under the GPLv3. A public GIT repository serves for development, maintenance and support. To facilitate first steps we provide usage examples in the form of a web based interactive environment (Jupyter notebooks), that further illustrate the modelling concept and the algorithm. To lighten system requirements we make use of the package and environment manager *conda*.

We carried out a case study to demonstrate the potential of our statistical model. All results presented in Section 4.5 are produced using CORBASS. Even though we use a rather small data set, computational costs are not negligible. At the time of writing, processing all epochs under consideration took about 30 hours on an ordinary workstation. The computational complexity is set by the number of observations and the parameter grid chosen. For a certain choice of model parameters, the computational complexity of GP regression is cubic w.r.t. number of observations. Performing a Riemann sum along one dimension scales quadratically according to the number of collocation points. Thus the complexity of the quadrature scheme for all three model parameters grows with the sixth power. Although numeric integration offers room for optimizations, the computational complexity cannot be lightened.

Although a proof of concept rather than a fully featured EMF model, our case study already supports the findings of existing studies. In comparison with models using traditional methods this is useful since it is another source that quantifies what was described qualitatively. As an example, early studies questioned the reconstruction of the EMF in the southern hemisphere from archeomagnetic and volcanic data only due to poor data coverage e.g. Korte, Donadini, and Constable (2009) and Constable and Korte (2015). That statement is quantified by our findings: The EMF's posterior standard deviation is small in areas of good data coverage, such as Europe, while uncertainties are large on the southern hemisphere (see top panel of Fig. 4.8). Even though our uninformative prior assumptions are significantly weaker, we find uncertainties similar to e.g. Licht et al. (2013), while noting that bootstrap ensemble methods tend to underestimate uncertainties in regions where there are no data to draw from. Surprisingly, our uncertainty estimates are on a par with more elaborate modelling concepts (e.g. Hellio and Gillet, 2018) whereas our approach does not yet account for the temporal evolution. Future work will show if this is a coincidence, and if taking temporal dynamics into account yields different results.

In a general context, caution is advised when performing any further processing with posterior Gauss coefficients. Whenever possible, use the posterior EMF instead. At densely covered regions a vast amount of Gauss coefficients are necessary to represent all the information that is contained in the posterior EMF at the surface. However, this is irrelevant when studying the core field which cannot be retrieved beyond degrees around 14 anyway, due to the distance from the source and dominance of the lithospheric field at higher degrees.

The spatial correlation structure we employ makes truncating the SH decomposition obsolete and the model resolves according to the availability of data. This raises the obvious question of how our global, non-parametric model compares to higher resolving studies. To investigate this question one would have to include other sources of data such as historical logs and observatory data for recent times. Although the modelling strategy inherently works with magnetic field components B_N, B_E and B_Z we deliberately left out recent observatory data to put attention on non-linear observables. If we focus on areas of small modelling errors, precise locations must not be neglected. Considering elevation and the coordinate conversion from geodetic to geocentric is straightforward. The difficulty with historical logs is that the direction of travel is affected by large inaccuracies (Jackson, Jonkers, and Walker, 2000; Jonkers, Jackson, and Murray, 2003). To account for location uncertainties, our data model requires an extension. Another difficulty is computational costs related to the large amount of historical and observatory data. Therefore one has to introduce a data selection and reduction process as the interest is in time spans of years but not days or less. Furthermore, our proxy Gaussian error model is intolerant of outliers. With only a few records that strongly deviate, the Gaussian error model causes a highly distorted reconstruction. Therefore it is important to perform outlier analysis and select data with care. The work by Khokhlov, Hulot, and Bouligand (2006) appears to be well suited to discriminate data that are incompatible with our modelling approach.

For the time increments of interest – $\Delta t \geq 1\,\mathrm{yr}$ – we know with confidence when observatory data and historical records were acquired. For volcanic and archeomagnetic records, the average dating uncertainties amount to about a hundred years Licht et al., 2013, Sec. 2.2, and they tend to increase further back in time. For the present study, we assume that our rather long bin width of 100 years balances temporal errors, although it results in poor temporal resolution. Arguably, our model is still overly confident as we did not consider dating errors. In order to apply our method more generally and to longer times the inclusion of sediment records has to be implemented. Sediment records are affected by large dating uncertainties which require a data model that also accounts

for temporal errors and preserves the stratification (e.g., Nilsson et al., 2014). Moreover, a strategy to deal with the scaling or relative intensity has to be developed and our two-step approach might become a challenge if the number of available full vector records with absolute intensity information is small.

Nonetheless, the presented snapshot model should be considered a first step in the direction of a time continuous correlation based Holocene magnetic field model, and, more generally, a new modelling method for the paleomagnetic field on various time-scales. We regard the time stationary binning an interim solution as it does not capture the dynamics of the EMF well. In our opinion a temporally continuous model also considering dating errors is needed to fulfil the needs of paleomagnetic field modelling. For the extension of our modelling concept we are currently working on an empirical continuous time correlation kernel, similar to Bouligand et al. (2016) and Hellio and Gillet (2018). Combining a temporal and a spatial kernel to a space-time kernel will make any binning obsolete. Within this space-time GP setting, existing techniques may be used to address dating uncertainties (McHutchon and Rasmussen, 2011).

We conclude by emphasizing again that this is initial work towards a new paleomagnetic field modelling strategy and an improved full Holocene model, and that it is destined to receive many improvements in the future. The open source modelling concept offers vast flexibility and a allows for a variety of refinements.

This chapter is a transcript of Schanner et al. (2021)

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

Submitted to JGR: Solid Earth on December 14, 2020; accepted June 18, 2021.

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.

5.1 Introduction

Existing models of the 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 (see e.g. Hartmann and Pacca, 2009; Jackson and 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 SH coefficients, truncated at a certain degree. Most models, such as Jackson, Jonkers, and Walker (2000), Korte, Donadini, and Constable (2009) and Senftleben (2019), implement spline interpolation in the temporal domain, while some alternative approaches exist (Constable and Parker, 1988; Bouligand et al., 2005; Hellio et al., 2014; Hellio and 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 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, Jonkers, and Walker, 2000; Constable, Johnson, and Lund, 2000; Korte and Constable, 2003) do not provide uncertainty estimates, more recent field models use ensemble techniques to quantify (modeling related) errors (Korte, Donadini, and Constable, 2009; Licht et al., 2013; Pavón-Carrasco et al., 2014; Hellio and 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 and Williams, 2006; McHutchon and 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 5.2 we discuss our prior assumptions, construct the correlation kernel and describe the full modeling algorithm. Section 5.3 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 5.4. The appendix provides further insight into the mathematical footing of the introduced methods.

5.1.1 Magnetic Field Theory

Assuming an insulating mantle, outside of the conducting core, the EMF $\mathbf{B}$ can be approximated by the gradient of the geomagnetic potential Φ (Backus, Parker, and Constable, 1996):

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

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 $\mathbf{x}$ outside of this sphere $|\mathbf{x}| > R$ the field can be represented using spherical harmonics (SH)

$$\Phi(\mathbf{x}) = R \sum_{\ell} \left(\frac{R}{|\mathbf{x}|}\right)^{\ell+1} \sum_{-\ell \le m \le \ell} g_\ell^m(t)\, Y_\ell^m(\hat{\mathbf{x}})\,. \tag{5.120}$$

$\hat{\mathbf{x}}$ is the unit vector $\mathbf{x}/|\mathbf{x}|$ 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 $\mathbf{x} = (\mathbf{x}, t)$ to distinguish space-time inputs from purely spatial inputs. Often the Gauss coefficients are expressed in form of a spline model (Bloxham and Jackson, 1992)

$$g_\ell^m(t) = \sum_n g_{\ell,n}^m M_n(t)\,, \tag{5.121}$$

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

$$B_N = \frac{1}{r}\frac{\partial \Phi}{\partial \theta}\,, \qquad B_E = -\frac{1}{r\sin(\theta)}\frac{\partial \Phi}{\partial \varphi}\,, \qquad B_Z = \frac{\partial \Phi}{\partial r}\,, \tag{5.122}$$

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

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:

$$\mathbf{H} : \mathbf{B} \to \begin{pmatrix} D \\ I \\ F \end{pmatrix} = \begin{pmatrix} \arctan\left(\frac{B_E}{B_N}\right) \\ \arctan\left(\frac{B_Z}{F_H}\right) \\ \sqrt{B_N^2 + B_E^2 + B_Z^2} \end{pmatrix} \tag{5.123}$$

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

5.1.2 Gaussian Process regression

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

$$\mathbf{B} \sim \mathcal{GP}\left(\bar{\mathbf{B}},\, K_{\mathbf{B}}\right). \tag{5.124}$$

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

$$\mathbb{E}\left[\mathbf{B}(\mathbf{x})|O\right] = \bar{\mathbf{B}}(\mathbf{x}) + K_{\mathbf{B}}(\mathbf{x},\mathbf{y})\left(K_{\mathbf{B}}(\mathbf{y},\mathbf{y}) + \Sigma_o\right)^{-1}\left(o(\mathbf{y}) - \bar{\mathbf{B}}(\mathbf{y})\right) \tag{5.125}$$

$$\mathrm{Cov}\left[\mathbf{B}(\mathbf{x}), \mathbf{B}(\mathbf{x}')|O\right] = K_{\mathbf{B}}(\mathbf{x},\mathbf{x}') - K_{\mathbf{B}}(\mathbf{x},\mathbf{y})\left(K_{\mathbf{B}}(\mathbf{y},\mathbf{y}) + \Sigma_o\right)^{-1} K_{\mathbf{B}}(\mathbf{y},\mathbf{x}')\,. \tag{5.126}$$

$\mathbf{x} = (\mathbf{x}, t)$ 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 and Gillet, 2018) or spatial correlations (Sanchez et al., 2016; Mauerberger et al., 2020).

5.2 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.

5.2.1 A priori process

Translating the uninformative dipole prior from MSKH20 to a time-dynamic realm presents a challenge, as temporal correlations vanish in the limit of zero a priori precision and thus 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:

$$\bar{\mathbf{B}}(\mathbf{x}) = -\nabla\left(\gamma_1^0 \cdot Y_1^0(\hat{\mathbf{x}}) \cdot \frac{R^3}{|\mathbf{x}|^2}\right) \tag{5.127}$$

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 $K_{\Phi,S,ND}$ 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 5.5.1 we lay out 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)

$$K_L(\mathbf{x}, \mathbf{x}') = \frac{1}{\sqrt{1 - 2b + a^2}}, \tag{5.128}$$

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

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

$$K_{\Phi,T}(t, t') = K_{SQE}(t, t') = \exp\left(-\frac{(t - t')^2}{\tau^2}\right) \tag{5.129}$$

τ is the a priori correlation time, that gives an idea about the timescale on which the dynamics of the process happen. Note that the posterior curve may be smoother or more detailed than this scale, depending on the data. 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 (Constable and Parker, 1988). Each part is coupled to its own temporal correlation kernel and thus has its own correlation time:

$$\begin{aligned} K_\Phi(\mathbf{x}, \mathbf{x}') =& \alpha_{DP}^2 K_{DP,S}(\mathbf{x}, \mathbf{x}') \cdot \exp\left(-\frac{(t - t')^2}{\tau_{DP}^2}\right) \\ &+ \alpha_{ND}^2 K_{ND,S}(\mathbf{x}, \mathbf{x}') \cdot \exp\left(-\frac{(t - t')^2}{\tau_{ND}^2}\right) \end{aligned} \tag{5.130}$$

The index S stands for spatial and $\alpha_\bullet$ are the a priori variances of the dipole and non-dipole part. They can be interpreted as the expected standard deviation in the dipole (g_1^m) and non-dipole coefficients. See 5.5.1 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

$$K_{\mathbf{B}}(\mathbf{x}, \mathbf{x}') = \nabla_{\mathbf{x}} \nabla_{\mathbf{x}'} K_\Phi(\mathbf{x}, \mathbf{x}'). \tag{5.131}$$

5.2.2 Linearization

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

$$o(\mathbf{x}) = \mathbf{H}\Big(\mathbf{B}(\mathbf{x}, t)\Big) + \mathbf{E}. \tag{5.132}$$

Clearly, the relationship to the field vector is non-linear (Eq. 5.123). 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, Rasmussen, and Ghahramani, 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

$$D \approx \tilde{D} + \frac{1}{\tilde{F}_H^2} \begin{bmatrix} -\tilde{B}_E \\ \tilde{B}_N \\ 0 \end{bmatrix}^\top \mathbf{B} \,, \tag{5.133}$$

$$I \approx \tilde{I} + \frac{1}{\tilde{F}_H} \left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{\tilde{B}_Z}{\tilde{F}} \frac{\tilde{\mathbf{B}}}{\tilde{F}} \right)^\top \mathbf{B} \,, \tag{5.134}$$

$$F \approx \frac{\tilde{\mathbf{B}}^\top}{\tilde{F}} \mathbf{B} \,. \tag{5.135}$$

Here, $\tilde{D}$, $\tilde{I}$, $\tilde{F}$ and $\tilde{\mathbf{B}}$ indicate the 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

$$o(\mathbf{x}) \approx \mathbf{H}_{\text{lin.}} \mathbf{B}(\mathbf{x}, t) + \mathbf{E} \,. \tag{5.136}$$

$\mathbf{H}_{\text{lin.}}$ refers to the linearized observation functionals Eqs. 5.133-5.135.

Previous works implemented an axial dipole as POE (e.g. Hellio and 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.

5.2.3 Measurement errors

In order to apply the GP regression formalism (Eq. 5.124), 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 and Nilsson, 2019)

$$\sigma_I = \frac{57.3^\circ}{140}\alpha_{95} \qquad \text{and} \qquad \sigma_D = \frac{1}{\cos o_I}\sigma_I \,. \tag{5.137}$$

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

$$o(\mathbf{x}) \approx \mathbf{H}_{\text{lin.}}\Big(\mathbf{B}(\mathbf{x}, t) + \rho\mathbf{P}\Big) + \varepsilon\mathbf{E}_{\text{prox.}} \,, \tag{5.138}$$

where $\mathbf{E}_{\text{prox.}}$ are the approximate errors, constructed from Eq. 5.137.

5.2.4 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, Donadini, and Constable, 2009; Hellio and Gillet, 2018; Senftleben, 2019). Hellio et al. (2014) used a normal error model for the dates, and applied MCMC methods to estimate the posterior distribution. We pursue a similar, hierarchical approach, but instead of MCMC methods, we perform analytic approximations. The dating uncertainties are translated to measurement errors, as presented below, by weighing them with the temporal derivatives 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. This covariance mediates the dating uncertainties to the measurement errors.

Summarizing the errors from the previous section as ε for readability, the data model is

$$o(\mathbf{x}) \approx \mathbf{H}_{\text{lin.}}\mathbf{B}(\mathbf{x}, t) + \varepsilon \,. \tag{5.139}$$

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

$$t_o = t + e_t \,, \qquad \text{where} \qquad e_t \sim \mathcal{N}(0, \sigma_t^2) \tag{5.140}$$

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

$$o(\mathbf{x}) = \mathbf{H}_{\text{lin.}} \mathbf{B}(\mathbf{x}, t_o - e_t) + \varepsilon \,. \tag{5.141}$$

This is known as the noisy input Gaussian process (NIGP) (McHutchon and Rasmussen, 2011). Due to the random variable e_t appearing at the inputs of the GP $\mathbf{B}$, this data model is non-Gaussian again. To tackle it, McHutchon and Rasmussen (2011) suggest once more a linearization. This gives

$$o(\mathbf{x}) \approx \mathbf{H}_{\text{lin.}} \left(\mathbf{B}(\mathbf{x}, t_o) - e_t^\top \partial_t \mathbf{B}(\mathbf{x}, t)|_{t_o} \right) + \varepsilon \,. \tag{5.142}$$

The first term is normal and the second term allows for easy construction of a moment matching proxy. With this modifications, Eq. 5.142 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

$$\Sigma_{tt'} \circ \partial_t \partial_{t'} K_{\mathbf{B}}(\mathbf{x}, \mathbf{x}')|_{t_o} \,. \tag{5.143}$$

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, $K_{\mathbf{B}}(\mathbf{x}, \mathbf{x}')$ 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 5.1 we present a synthetic, one dimensional example, to compare 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 5.1, the NIGP gives a reasonable proxy to the MCMC result at immensely reduced computational cost.

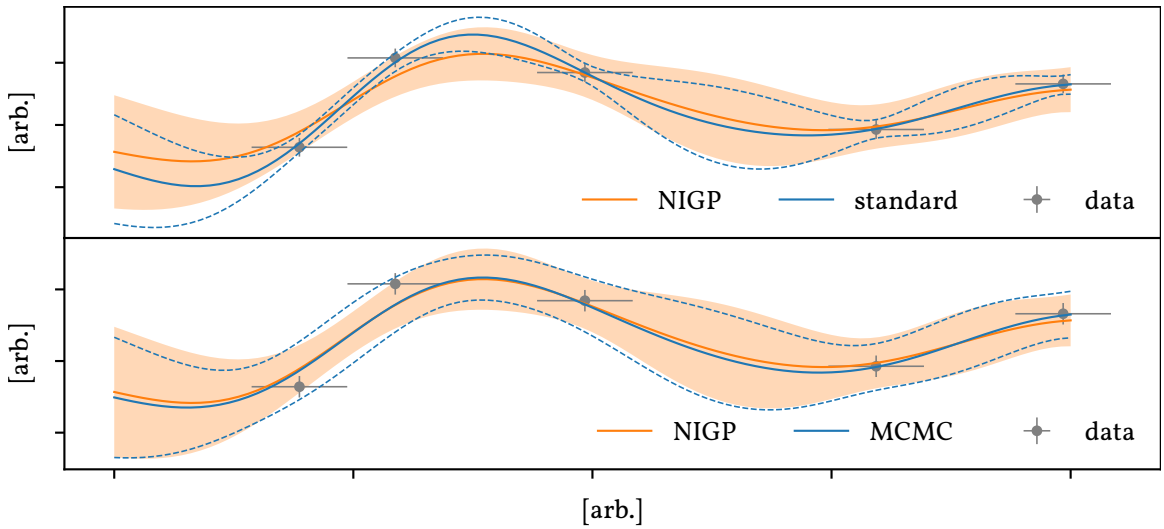


Figure 5.1: Different inversion strategies for a synthetic one dimensional model. The vertical uncertainty is deliberately chosen smaller than the horizontal one, to highlight the effect of the different strategies. The shaded area and dashed lines present one standard deviation. In the top panel, the standard GP inversion, which can not incorporate input uncertainty, is compared to the noisy input Gaussian process (NIGP) approach. 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. In the bottom panel, an MCMC based approach is compared to the NIGP. 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.

5.2.5 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 = 2800km, which gives a slope similar to the IGRF power spectra (Thébault et al., 2015). Note, that R gives the radius of a sphere of "virtual" sources and has no direct physical meaning. The other parameters are marginalized, so that the outcome of the modeling procedure is a compound distribution

$$\mathrm{p}(\mathbf{B}|\mathrm{o}) = \int \mathrm{p}(\mathbf{B}|\mathrm{o}, \eta) \cdot \mathrm{p}(\eta|\mathrm{o})\, \mathrm{d}\eta. \tag{5.144}$$

Here η is the set of hyperparameters $\{\gamma_1^0, \alpha_{DP}, \alpha_{ND}, \tau_{DP}, \tau_{ND}, \varepsilon, \rho\}$ and $p(\eta|o)$ is the marginal posterior. See 5.5.2 and MSKH20 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.

5.3 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 5.5.2. However, two points are to be discussed here. One is the explicit second derivative of the temporal kernel, appearing in Eq. 5.143. In this study the correlation kernels for dipole and non-dipole contributions are each considered tensor product kernels, so that the time derivative only affects the temporal part. The full derivative of each contribution is then given by the product of the respective spatial part and the second order derivative of the SQE-kernel. For the SQE-kernel, the derivative is straightforward to calculate and reads

$$\partial_t \partial_{t'} K_{SQE}(t, t') = \left(\frac{2}{\tau^2} - \frac{4}{\tau^4}(t - t')^2 \right) \cdot K_{SQE}(t, t') \,. \tag{5.145}$$

When using a more realistic kernel, especially one with different correlation times for different degrees ℓ, the kernel can no longer be constructed as a tensor product, and 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).

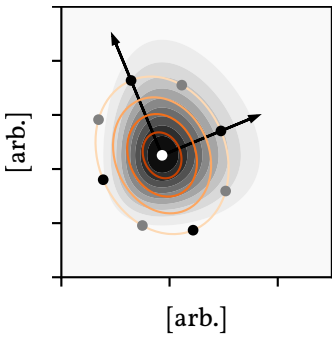


Figure 5.2: Illustration of the CCD integration in two dimensions. Integration over the actual distribution (grey in the background) is performed by calculating the MAP (white) and constructing a proxy multivariate Gaussian (orange), using the inverse Hessian at the MAP. The dots indicate the integration 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, Martino, and Chopin (2009, Sec. 6.5) and Sanchez and Sanchez (2005).

The second point is the marginalization integral in Eq. 5.144. 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 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, Martino, and Chopin (2009, Sec. 6.5). The idea is to center the integration around the maximum a posteriori probability estimator (MAP) of the marginal posterior. Collocation points are added according to central composite design (CCD) (Sanchez and Sanchez, 2005), in order to capture the bulk of the uncertainties in the hyperparameters (see Figure 5.2). In seven dimensions, the integration is approximated by a sum over 79 collocation points. To find the MAP of the marginal posterior, we use the LIPO-TR global optimization algorithm (King, 2009; King, 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

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

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

$$\mathrm{p}(g_1^0) \propto \frac{1}{g_1^0}, \qquad \mathrm{p}(\alpha_\bullet) \propto \frac{1}{\alpha_\bullet}, \qquad \mathrm{p}(\tau_\bullet) \propto \frac{1}{\tau_\bullet} \qquad \text{and} \qquad \mathrm{p}(\rho) \propto \frac{1}{\rho}, \tag{5.146}$$

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. 5.144, the compound distribution is approximated by a Gaussian mixture. Details can again be found in 5.5.2.

5.3.1 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 (cf. Fig. 5.5). At these locations we generate data from the ARCH10k.1 model (Constable, Korte, and Panovska, 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. Additionally, to simulate effects of other magnetic field sources and from the measurement process, we add random contributions with a constant standard deviation of 2.5 μT to the generated data.

The resulting MAP for the hyperparameters is

$$\begin{aligned}
\hat{g}_1^0 &= -396.641\,\mu\mathrm{T} & \hat{\alpha}_{\mathrm{DP}} &= 19.314\,\mu\mathrm{T} & \hat{\tau}_{\mathrm{DP}} &= 318.454\ \text{yrs.} \\
\hat{\varepsilon} &= 101.656\,\% & \hat{\alpha}_{\mathrm{ND}} &= 40.4634\,\mu\mathrm{T} & \hat{\tau}_{\mathrm{ND}} &= 339.471\ \text{yrs.} \\
\hat{\rho} &= 2.59371\,\mu\mathrm{T}
\end{aligned}$$

Note that $\hat{g}_1^0$, $\hat{\alpha}_{\mathrm{D}}$ and $\hat{\alpha}_{\mathrm{ND}}$ are given w.r.t. the reference radius, i.e. at the Earth's surface the axial dipole and both variances are much smaller. To transform g_1^0 and α_{D} to the Earth's surface, they have to be multiplied by a factor of $(\mathrm{R}/\mathrm{R}_{\mathrm{Earth}})^3 \approx 0.08$. Scaling the non-dipole variance α_{ND} is not possible, due to the degree dependent scaling and the closed form kernel. The error level scalings are as expected. As the reported error levels have been used to corrupt the data, they are accurate ($\varepsilon \approx 100\%$). The residual term reflects the random contributions that have been added.

In Figure 5.3 we compare the dipole of the reference model to our findings from running the model on the synthetic data. One can see that only the long-term behavior of the reference model can be recovered. Nonetheless, we believe that the results are meaningful and related to the data distribution. To illustrate this, we present a simplified exam-

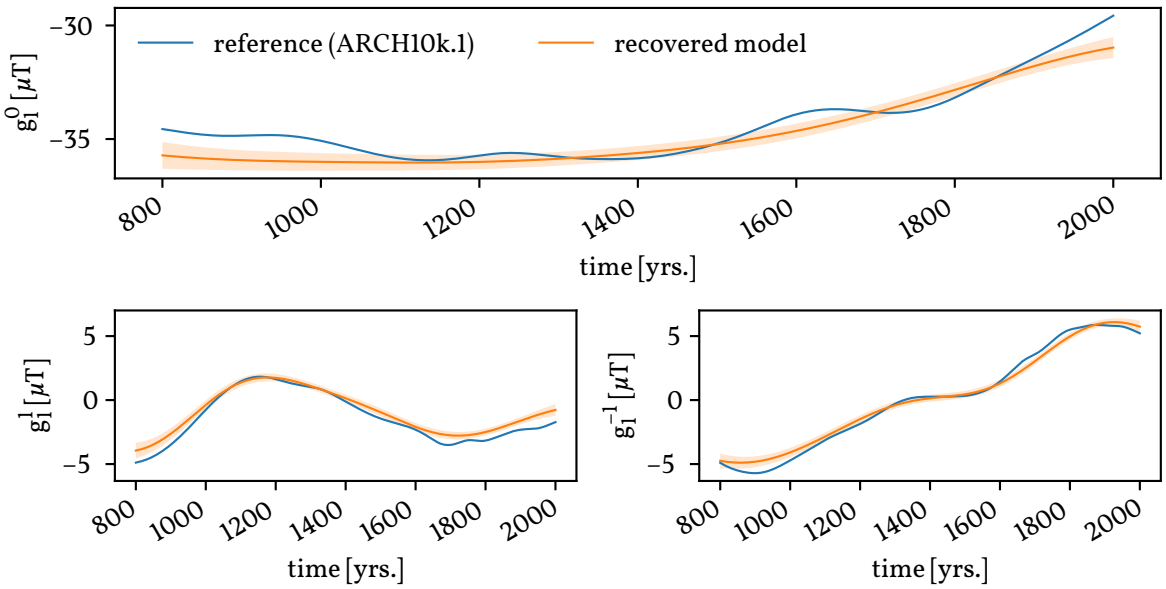


Figure 5.3: Comparison of the reference model to the one recovered from synthetic test data. The shaded area represents one standard deviation. The long-term behavior could be recovered, while data is too sparsely distributed and uncertainties are too large to recover the short timescale features. See the text for additional discussion.

ple. Consider a one-dimensional time series that we want to model by a Gaussian process. We generate data from some reference model (here we chose the axial dipole of the ARCH10k.1 model) and corrupt it with artificial noise. For simplicity and to emphasize the effect, we neglect dating uncertainties. We synthesize two datasets, one with many observations and small uncertainties and one with few observations and large errors. The "data poor" situation was designed to mimic the actual situation, based on findings in MSKH20: Every one hundred years an observation was generated and the error level chosen to agree with the one found in MSKH20.

Then for both datasets, a separate correlation time is estimated by maximizing the marginal posterior. In the situation with many observations, the proposed modeling algorithm gives a correlation time corresponding to the short term variability of the reference model and the reference model is recovered well and detailed (Figure 5.4, left panel). In the situation with few observations however, one can only recover the long-term variability of the reference model (Figure 5.4, right panel). The actual situation is more complicated, but we believe that the large correlation times and smooth curves come from a similar mechanism. While locally, the data distribution is much denser than in the "data poor" case, global information (e.g. about the axial dipole) is sparse and unreliable, as can be seen from the results of MSKH20. We performed a similar experiment for synthetic field data and found similar results, i.e. only the long-term information can be recovered.

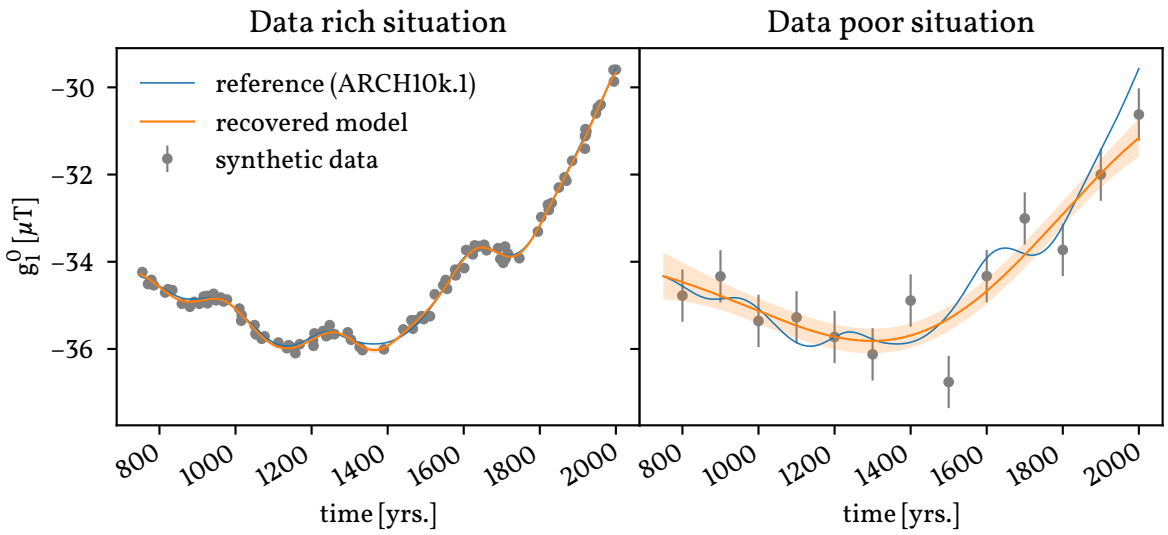


Figure 5.4: Synthetic one-dimensional example illustrating how the lack of data influences the information one can recover from the reference model. The shaded area represents one standard deviation. In the data rich situation on the left, the reference model can be recovered quite well and posterior uncertainties are so small, they are barely visible. In the data poor situation on the right only the long-term behavior can be recovered.

5.3.2 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 5.5 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 $\alpha_{95} = 4.5^\circ$ as directional errors and $\sigma_F = 8.25\,\mu$T 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 5.8.

We compare our findings to two existing magnetic field models as well as to the results of MSKH20. The models are ARCH10k.1 (Constable, Korte, and Panovska, 2016) and COV-ARCH (Hellio and Gillet, 2018). They are considered reasonably comparable, as they are based on similar data compilations. Both models report Gauss coefficients up to SH degree $\ell = 10$, while the actual spatial resolution is determined by regulariza-

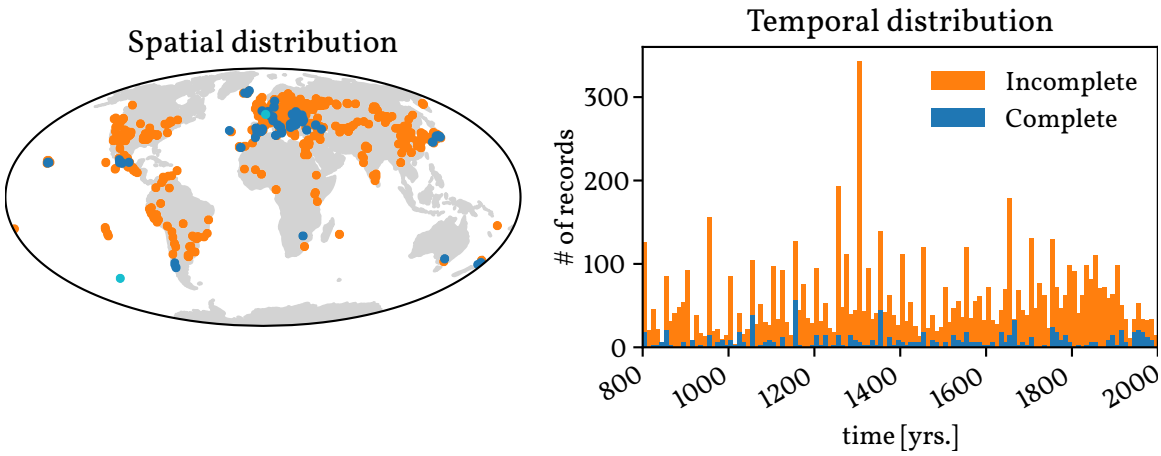


Figure 5.5: Spatial and temporal distribution of the data. We separate the data only into complete and incomplete without indicating declination, inclination and intensity separately, to avoid overloading the plots. The turquoise dots indicate the locations at which we present the time series (Paris, Fig. 5.8 and pacific, Fig. 5.9).

tion 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[1] 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 5.3.1):

$$\hat{\gamma}_1^0 = -425.242\,\mu\text{T} \qquad \hat{\alpha}_{\text{DP}} = 13.6831\,\mu\text{T} \qquad \hat{\tau}_{\text{DP}} = 348.555\,\text{yrs.}$$
$$\hat{\varepsilon} = 135.781\,\% \qquad \hat{\alpha}_{\text{ND}} = 39.4199\,\mu\text{T} \qquad \hat{\tau}_{\text{ND}} = 293.025\,\text{yrs.}$$
$$\hat{\rho} = 3.82749\,\mu\text{T}$$

[1] https://geodyn.univ-grenoble-alpes.fr/

Again, $\hat{\gamma}_1^0$, $\hat{\alpha}_\mathrm{D}$ and $\hat{\alpha}_\mathrm{ND}$ are given w.r.t. the reference radius and are lower at the Earth's surface. At the Earth's surface, $\hat{\gamma}_1^0 \approx -36.09\ \mu$T. 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 and Mauerberger, 2020).

Field predictions

Predicting on the EMF's vector components is straightforward and given by wrapping equations 5.125 and 5.126, 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 5.6 shows mean and standard deviation for the down component of the EMF for the epoch 1700. Similarly, predictions at the CMB can be performed, by translating the design points accordingly. The results for the down component are shown for the epoch 1700 in Figure 5.7. 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 5.6 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 5.7) 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.

Similar to the map, predictions at a specific location can be obtained for various times and time series of the observables can be constructed. 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 5.8 and 5.9 show time series at two distinct locations. Figure 5.8 presents time series for Paris, together with comparison models and data, while the series in Figure 5.9 are for a location in the Pacific, where no data is present in the surroundings. The data in Figure 5.8 stem from a surrounding of 250 km.

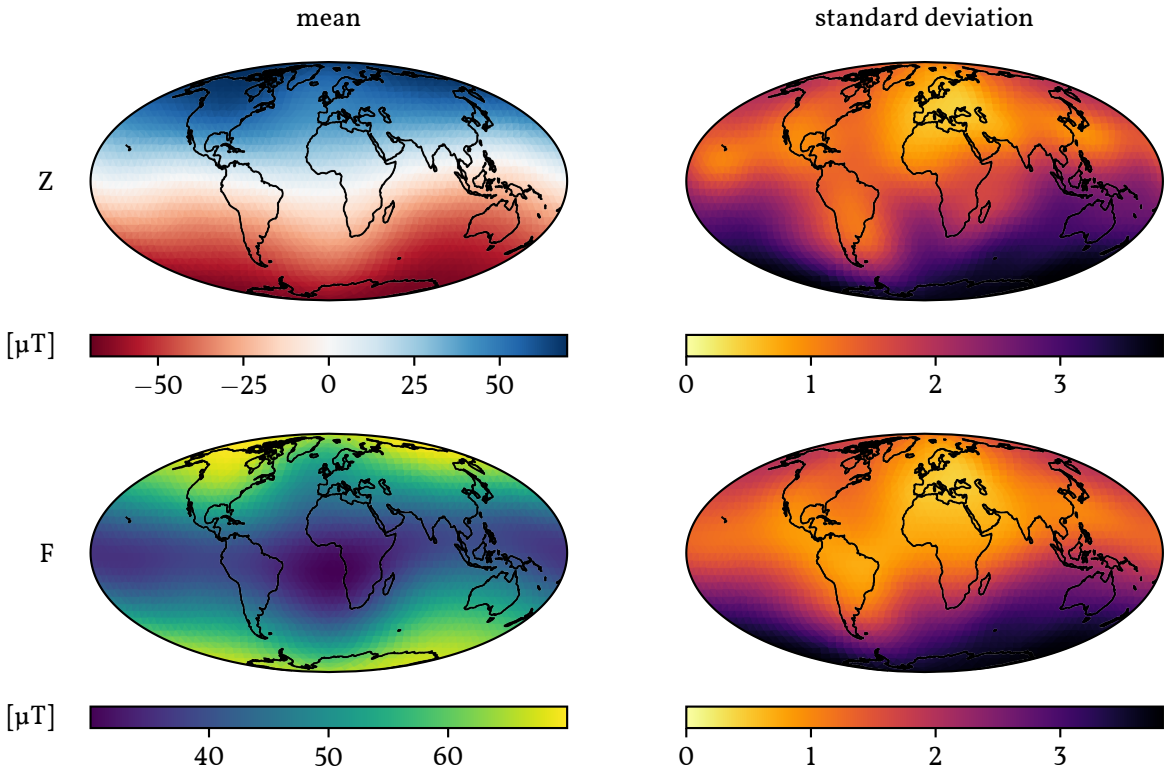


Figure 5.6: Maps of the EMFs down component (top) and intensity (bottom), together with standard deviations for the epoch 1700.

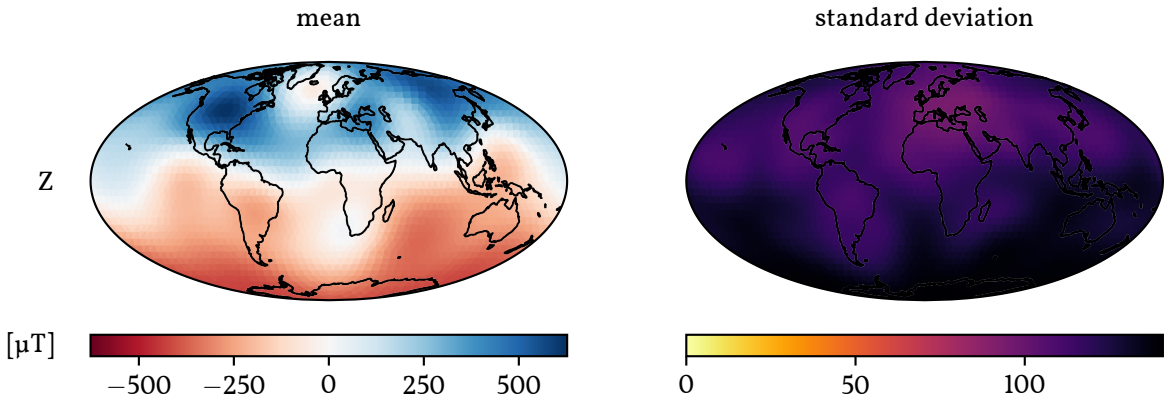


Figure 5.7: Map of the down component of the EMF at the CMB, together with standard deviation for the epoch 1700.

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

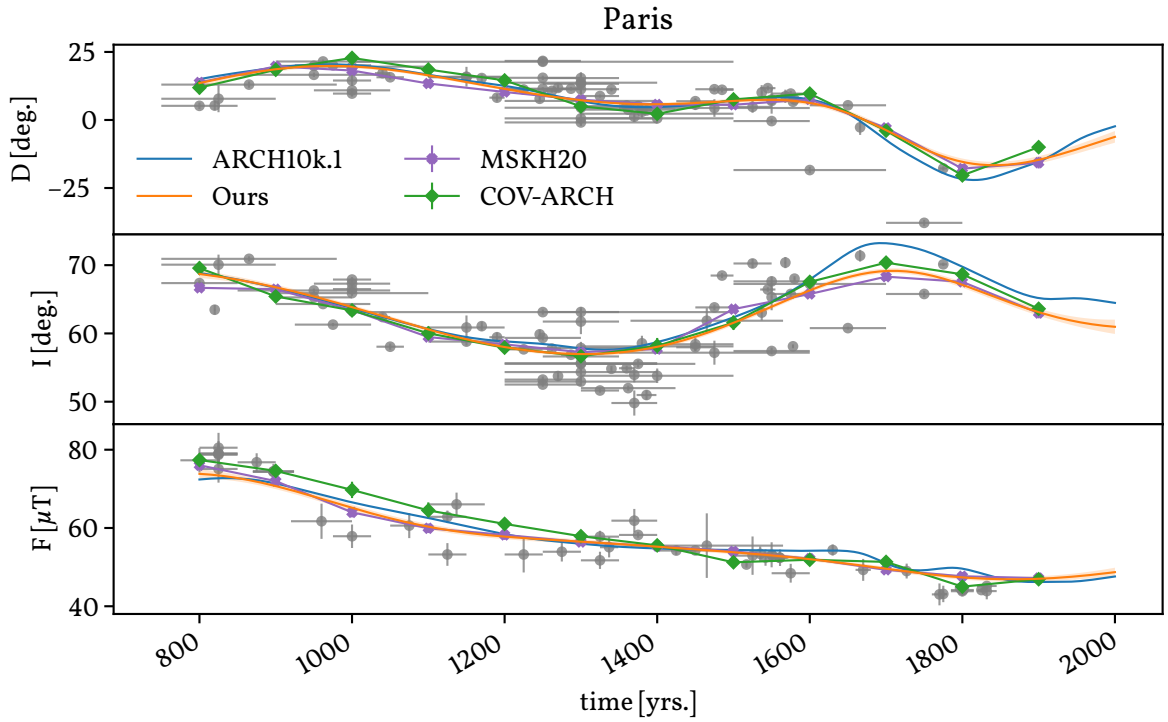


Figure 5.8: Data (grey dots) and model time series of the observables at Paris. Horizontal and vertical grey bars indicate the one sigma temporal and field component data uncertainties, respectively. The shaded area and vertical bars on the models also represent one sigma. For the MSKH20 snapshot model and COV-ARCH, which is reported in 100 year intervals, dots for the epochs are linearly connected by coloured lines.

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 5.10 present the moment matching Gaussian proxy to the actual mixture. The dipole coefficients from

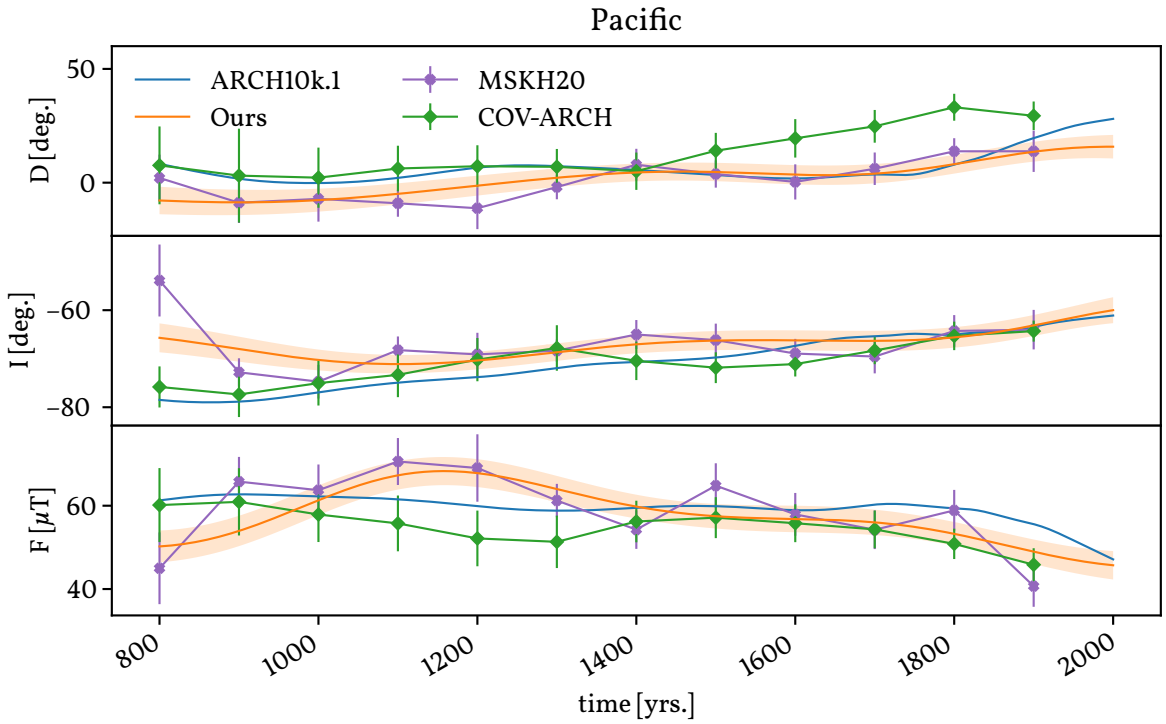


Figure 5.9: Model time series of the observables for a location in the Pacific (−48.18°, −128.03°, see also Fig. 5.5) as in Fig. 5.8. The shaded area and vertical bars on the models represent one sigma. Uncertainties are higher than for Paris (Figure 5.8) due to the lack of data in the surrounding.

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.

Spectra

Power spectra are considered to condense the information contained in the Gauss coefficients (e.g. Backus, Parker, and Constable, 1996). Using sampling techniques, mean and percentiles of the geomagnetic power spectrum distribution are available (for further details consider MSKH20, Section 5.6). Figure 5.11 shows the geomagnetic power spectrum for the epoch 1700, together with 16- and 84-percentiles. Within the reported uncertainties, the 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 $\ell = 4 \ldots 7$ when compared to MSKH20, which

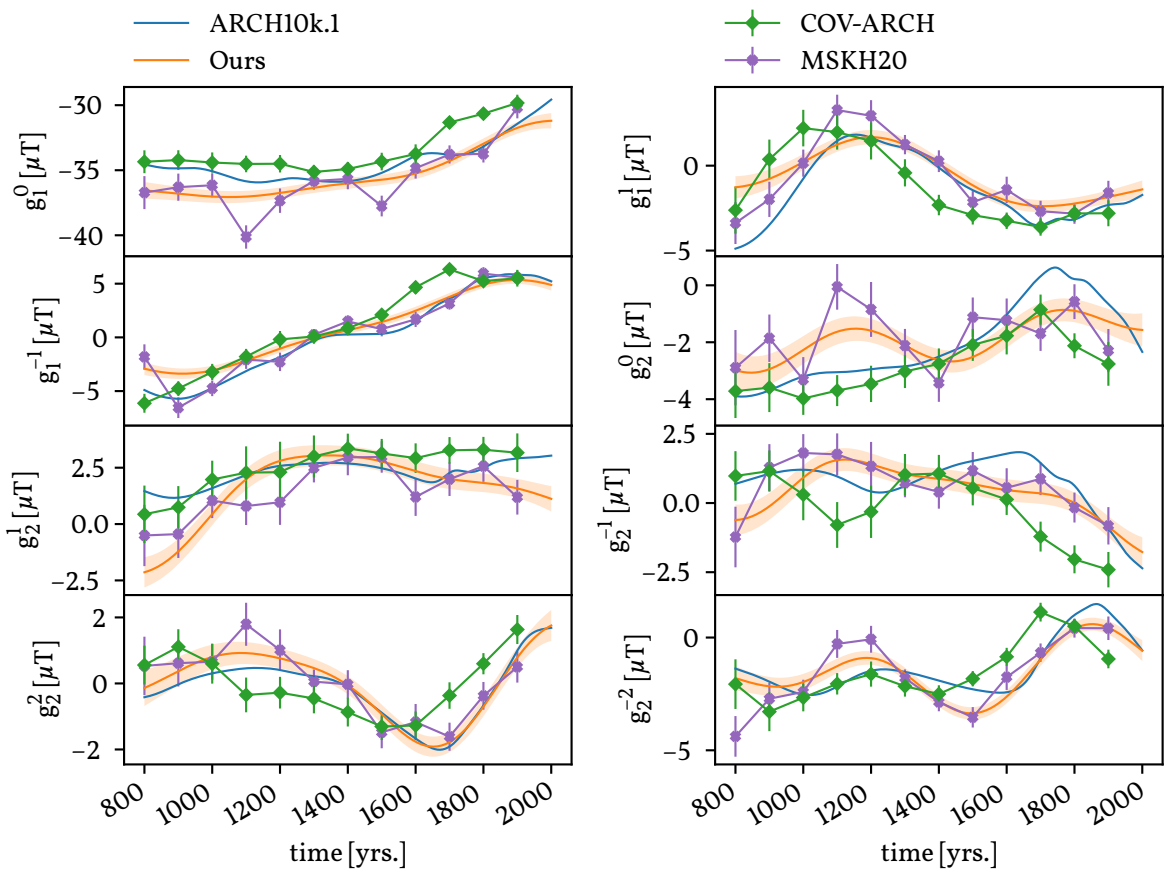


Figure 5.10: Gauss coefficient time series for degrees $\ell = 1$ and $\ell = 2$, together with comparable models. One sigma uncertainties are given by the shaded area for our model and by vertical bars where available. MSKH20 and COV-ARCH, which are available in 100 year steps are shown by coloured dots linked linearly by solid lines.

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:

$$\begin{aligned}\mathbb{E}\left[\dot{g}(t)|O\right] =& \partial_t \mathbb{E}\left[g(t)|O\right] \\ =& \partial_t \bar{g}(t) + \partial_t K_{g,\mathbf{B}}(t,\mathbf{y})\left(K_{\mathbf{B}}(\mathbf{y},\mathbf{y}) + \Sigma_o\right)^{-1}\left(o(\mathbf{y}) - \bar{\mathbf{B}}(\mathbf{y})\right)\end{aligned} \tag{5.147}$$

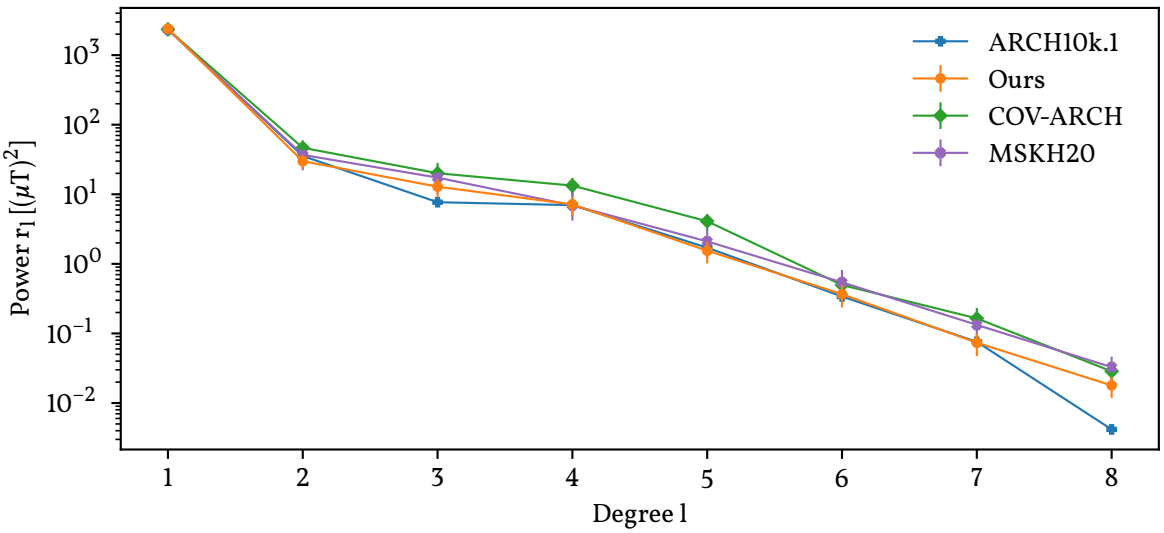


Figure 5.11: Comparison of the spatial power spectra at the Earth's surface for the epoch 1700 AD. The errorbars represent 16- and 84-percentiles. Thus the error covers an area of 68%, which corresponds to one standard deviation in the Gaussian case.

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 derivative of the SQE kernel:

$$\partial_t K_{SQE}(t, t') = -\frac{2}{\tau^2}(t - t')K_{SQE}(t, t') \tag{5.148}$$

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 5.12. Note, that we choose an earlier epoch for the secular variation spectrum, as ARCH10k.1 is constrained to gufm1 (Jackson, Jonkers, and Walker, 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.

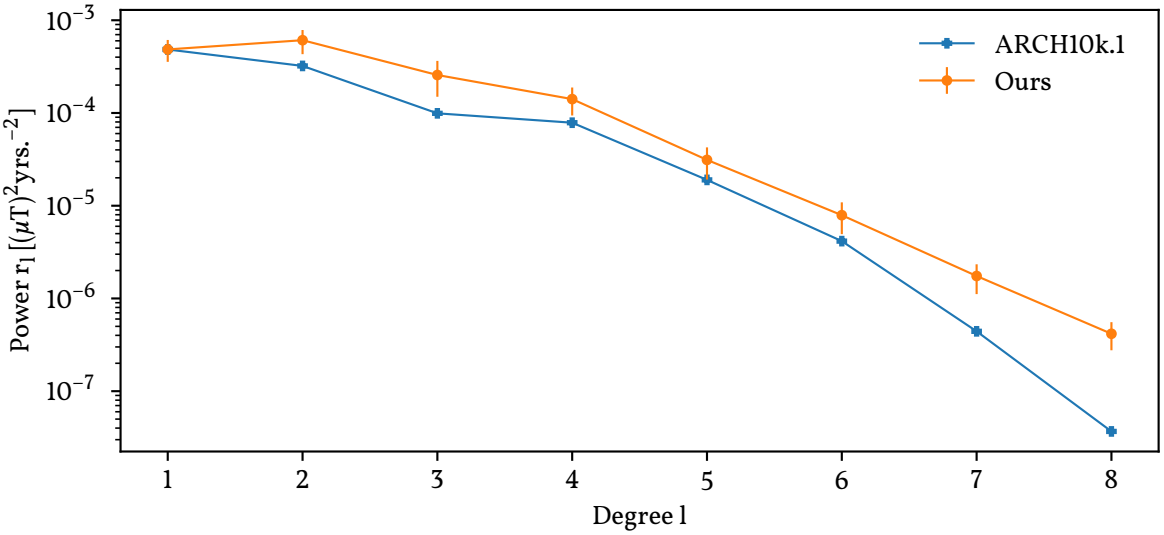


Figure 5.12: Secular variation spectra at the Earth's surface for 1400 AD. The errorbars represent 16- and 84-percentiles. Thus the error covers an area of 68%, which corresponds to one standard deviation in the Gaussian case.

Dipole

Finally we present the dynamics of the EMFs dipole. Figure 5.13 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 5.14 shows the movement of the geomagnetic north pole. The mean curve of our model (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 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 more recent times, the path lies further to the west. Deviance for the most recent epochs is again caused by a lack of data.

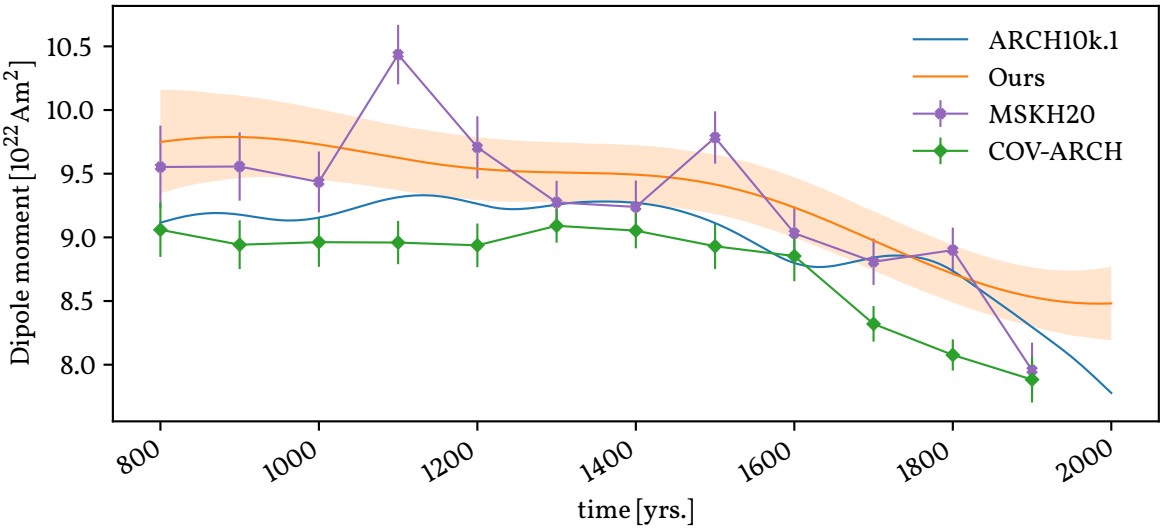


Figure 5.13: Time series of the dipole moment magnitude, together with comparison models. One sigma uncertainties are given by the shaded area for our model and by vertical bars where available. The MSKH20 snapshot model and COV-ARCH, which are available in 100 year steps, are shown by colored dots linked linearly by solid lines.

5.4 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 5.2, all necessary modifications are discussed, together with a novel approach to include dating uncertainties. In contrast to previous works (Hellio et al., 2014; Hellio and Gillet, 2018; Korte, Donadini, and Constable, 2009; Licht et al., 2013; Hellio and Gillet, 2018; Senftleben, 2019), using a NIGP (McHutchon and 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 vanish in the limit of zero a priori precision and thus 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 (Sanchez and Sanchez, 2005) based integration, as proposed by Rue, Martino, and Chopin (2009).

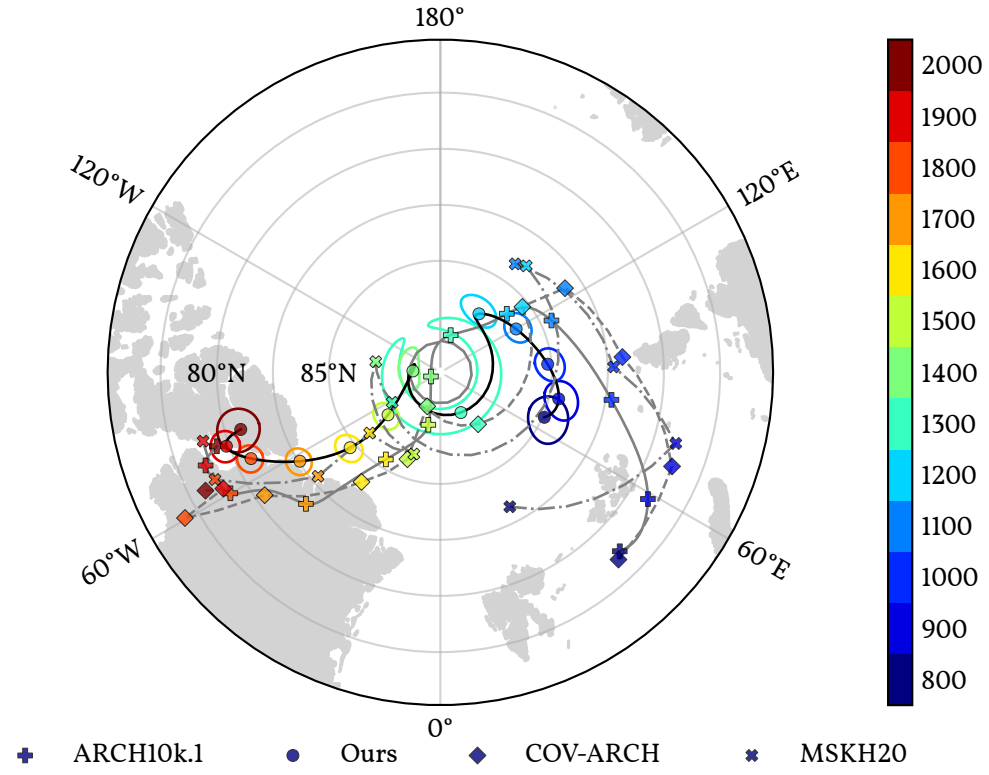


Figure 5.14: Movement of the geomagnetic north pole. Every one hundred years mean and one-sigma ellipses of the snapshot-distributions are shown. Color varies according to time, with markers every one hundred years. The black curve gives the mean pole track from our model. The grey curves give the same for the other models: Solid grey is ARCH10k.1, dashed grey is COV-ARCH and dash-dotted grey is MSKH20.

The major challenge in implementing this strategy consists of finding the MAP of the marginal posterior. Running the LIPO-TR algorithm (King, 2009; King, 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 ker-

nel 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 smoothness imposed by the SQE-kernel, which suppresses the influence of single records. The results of the case study show smoother curves than the models we compare to, in particular for the axial dipole contribution. While partly this can be attributed to the SQE-kernel, the proposed method is able to resolve faster variations locally, if the data supports these variations. Due to the closed form, such variations can be reflected by higher order coefficients and do not introduce global variations.

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 and Mauerberger, 2020). Together with extensive documentation, the software source code provides further insight into the modeling algorithm. It is available at `https://sec23.git-pages.gfz-potsdam.de/korte/corbam/`.

5.5 Appendix

5.5.1 Constructing the spatial covariance kernel

The Lagrange kernel (Eq. 5.128) is constructed from the Gauss coefficient correlations as follows:

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

$$\mathrm{Cov}\left[\Phi(\mathbf{x}), \Phi(\mathbf{x}')\right] = R^2 \sum_{\ell,m} \sum_{\ell',m'} \left(\frac{R^2}{|\mathbf{x}||\mathbf{x}'|}\right)^{\ell+1} Y_\ell^m(\hat{\mathbf{x}}) Y_{\ell'}^{m'}(\hat{\mathbf{x}}') \mathrm{Cov}\left[g_\ell^m, g_{\ell'}^{m'}\right] . \tag{5.149}$$

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

$$\mathrm{Cov}\left[g_\ell^m, g_{\ell'}^{m'}\right] = \delta_{\ell,\ell'}\delta_{m,m'}\,, \tag{5.150}$$

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

$$\mathrm{Cov}\left[\Phi(\mathbf{x}), \Phi(\mathbf{x}')\right] = R^2 \sum_{\ell=0}^{\infty} \left(\frac{R^2}{|\mathbf{x}||\mathbf{x}'|}\right)^{\ell+1} \sum_m Y_\ell^m(\hat{\mathbf{x}}) Y_\ell^m(\hat{\mathbf{x}}')\,. \tag{5.151}$$

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

$$K_L(\mathbf{x}, \mathbf{x}') = \frac{1}{\sqrt{1 - 2b + a^2}}\,, \tag{5.128}$$

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

$$K_{DP,S}(\mathbf{x}, \mathbf{x}') = \frac{b}{a^3}\,. \tag{5.152}$$

Thus the non-dipole kernel reads

$$K_{ND,S}(\mathbf{x}, \mathbf{x}') = K_L(\mathbf{x}, \mathbf{x}') - K_{DP,S}(\mathbf{x}, \mathbf{x}') - \frac{1}{a}\,, \tag{5.153}$$

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

5.5.2 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.

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 $\mathbf{B}$ 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:

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

$\nabla \mathbf{H}|_{\tilde{\mathbf{B}}(\mathbf{x})}$ refers to the gradient of the observation functionals (c.f. Eqs. 5.133-5.135), evaluated at the POE $\tilde{\mathbf{B}}$ at locations $\mathbf{x}$.

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

$$\bar{\mathbf{H}}(\mathbf{x}) = \mathbf{H}\left(\tilde{\mathbf{B}}(\mathbf{x})\right) + \nabla \mathbf{H}|_{\tilde{\mathbf{B}}(\mathbf{x})}\left(\tilde{\mathbf{B}}(\mathbf{x}) - \bar{\mathbf{B}}(\mathbf{x})\right) \tag{5.154}$$

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

$$\Sigma_{\mathbf{yy}} = \mathrm{K}_{\mathbf{B}}(\mathbf{y}, \mathbf{y}) \tag{5.155}$$

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

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

$$\bar{\mathbf{B}}_{|\mathrm{o}}(\mathbf{y}) = \mathbb{E}\left[\mathbf{B}(\mathbf{y})|\mathrm{o}\right] \tag{5.156}$$

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

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

$$\Sigma_{\mathbf{yy}|\mathrm{o}} = \mathrm{Cov}\left[\mathbf{B}(\mathbf{y}), \mathbf{B}(\mathbf{y})|\mathrm{o}\right] \tag{5.157}$$

$\Sigma_{\mathbf{yo}}$ 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

$$\Sigma_{\mathbf{yo}} = \mathrm{K}_{\mathbf{B}}(\mathbf{y}, \mathbf{x}_{\mathrm{o}}) \nabla \mathbf{H}|_{\tilde{\mathbf{B}}(\mathbf{x}_{\mathrm{o}})} \,. \tag{5.158}$$

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

Σ_{oo} refers to the covariance amongst observations:

$$\Sigma_{\mathrm{oo}} = \nabla \mathbf{H}|^{\top}_{\tilde{\mathbf{B}}(\mathbf{x}_{\mathrm{o}})}\left(\mathrm{K}_{\mathbf{B}}(\mathbf{x}_{\mathrm{o}}, \mathbf{x}_{\mathrm{o}}) + \Sigma_{\mathrm{T}} + \rho \Sigma_{\mathrm{p}}\right) \nabla \mathbf{H}|_{\tilde{\mathbf{B}}(\mathbf{x}_{\mathrm{o}})} + \varepsilon \Sigma_{\mathrm{e}} \tag{5.159}$$

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 5.2.3. Σ_T is the correction term for dating uncertainties, see Section 5.2.4:

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

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:

$$o = \{c, i\} \tag{5.160}$$

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

$$\mathbb{E}\left[\mathbf{B}(\mathbf{y})|c\right] = \bar{\mathbf{B}}(\mathbf{y}) + \Sigma_{\mathbf{yc}} \cdot \Sigma_{cc}^{-1} \cdot \left(c - \bar{\mathbf{H}}(\mathbf{x}_c)\right) \tag{5.161}$$

$$\mathrm{Cov}\left[\mathbf{B}(\mathbf{y}), \mathbf{B}(\mathbf{y})|c\right] = \Sigma_{\mathbf{yy}} - \Sigma_{\mathbf{yc}} \cdot \Sigma_{cc}^{-1} \cdot \Sigma_{\mathbf{yc}}^{\top} \tag{5.162}$$

Posterior correlations for the incomplete records are given by

$$\Sigma_{\mathbf{yi}|c} = \Sigma_{\mathbf{yi}} - \Sigma_{\mathbf{yc}} \cdot \Sigma_{cc}^{-1} \cdot \Sigma_{ic}^{\top} \tag{5.163}$$

$$\Sigma_{ii|c} = \Sigma_{ii} - \Sigma_{ic} \cdot \Sigma_{cc}^{-1} \cdot \Sigma_{ic}^{\top} \tag{5.164}$$

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

$$\tilde{\mathbf{B}}(\mathbf{x}_c) = \mathbf{H}^{-1}(c) \tag{5.165}$$

with

$$\mathbf{H}^{-1} : \begin{pmatrix} D \\ I \\ F \end{pmatrix} \rightarrow \mathbf{B} = F \begin{pmatrix} \cos(I)\cos(D) \\ \cos(I)\sin(D) \\ \sin(I) \end{pmatrix} . \tag{5.166}$$

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

$$\tilde{\mathbf{B}}(\mathbf{x}_i) = \bar{\mathbf{B}}_{|c}(\mathbf{x}_i) \tag{5.167}$$

The remaining records are incorporated and give a posterior with

$$\mathbb{E}\left[\mathbf{B}(\mathbf{y})|\mathrm{c},\mathrm{i}\right] = \bar{\mathbf{B}}_{|\mathrm{c}}(\mathbf{y}) + \Sigma_{\mathbf{yi}|\mathrm{c}} \cdot \Sigma_{\mathrm{ii}|\mathrm{c}}^{-1} \cdot \left(\mathrm{i} - \bar{\mathbf{H}}_{|\mathrm{c}}(\mathbf{x}_\mathrm{i})\right) \tag{5.168}$$

$$\mathrm{Cov}\left[\mathbf{B}(\mathbf{y}),\mathbf{B}(\mathbf{y})|\mathrm{c},\mathrm{i}\right] = \Sigma_{\mathbf{yy}|\mathrm{c}} - \Sigma_{\mathbf{yi}|\mathrm{c}} \cdot \Sigma_{\mathrm{ii}|\mathrm{c}}^{-1} \cdot \Sigma_{\mathbf{yi}|\mathrm{c}}^{\top} \tag{5.169}$$

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.

Marginal posterior

The density

$$\mathrm{p}(\mathrm{o}|\eta) = \int \mathrm{p}(\mathrm{o}|\mathbf{B})\mathrm{p}(\mathbf{B}|\eta)\ \mathrm{d}\mathbf{B} \tag{5.170}$$

is called *marginal likelihood* where the term marginal refers to the integration over the EMF. $\mathrm{p}(\mathrm{o}|\eta)$ is a function in the hyperparameters η, given the data o. For a certain choice of hyperparameters, $\mathrm{p}(\mathrm{o}|\eta)$ 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, $\eta = \{\gamma_1^0, \alpha_{\mathrm{DP}}, \alpha_{\mathrm{ND}}, \tau_{\mathrm{DP}}, \tau_{\mathrm{ND}}, \varepsilon, \rho\}$, see also Section 5.2.5. Building the compound distribution for the EMF (Eq. 5.144) 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

$$\mathrm{p}(\mathrm{o} = \{\mathrm{c},\mathrm{i}\}|\eta) = \mathrm{p}(\mathrm{i}|\eta,\mathrm{c})\mathrm{p}(\mathrm{c}|\eta)\,. \tag{5.171}$$

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

$$\mathrm{p}(\mathrm{c}|\eta) = \frac{\exp\left(-\frac{1}{2}\left(\mathrm{c} - \bar{\mathbf{H}}(\mathbf{x}_\mathrm{c})\right)^{\mathrm{T}}\Sigma_{\mathrm{cc}}^{-1}\left(\mathrm{c} - \bar{\mathbf{H}}(\mathbf{x}_\mathrm{c})\right)\right)}{\sqrt{(2\pi)^{3n_\mathrm{c}}|\Sigma_{\mathrm{cc}}|}} \tag{5.172}$$

$$\mathrm{p}(\mathrm{i}|\mathrm{c},\eta) = \frac{\exp\left(-\frac{1}{2}\left(\mathrm{i} - \bar{\mathbf{H}}_{|\mathrm{c}}(\mathbf{x}_\mathrm{i})\right)^{\mathrm{T}}\Sigma_{\mathrm{ii}|\mathrm{c}}^{-1}\left(\mathrm{i} - \bar{\mathbf{H}}_{|\mathrm{c}}(\mathbf{x}_\mathrm{i})\right)\right)}{\sqrt{(2\pi)^{n_\mathrm{i}}|\Sigma_{\mathrm{ii}|\mathrm{c}}|}}\,. \tag{5.173}$$

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:

$$\mathrm{p}(\eta|\mathrm{o}) \propto \mathrm{p}(\mathrm{o}|\eta)\mathrm{p}(\eta) \tag{5.174}$$

For numerical reasons, one often uses the log marginal posterior

$$\ln \mathrm{p}(\eta|\mathrm{o}) = \ln \mathrm{p}(\mathrm{o}|\eta) + \ln \mathrm{p}(\eta) + \text{const.}\,. \tag{5.175}$$

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, Martino, and Chopin (2009) and Sanchez and Sanchez (2005). With weights Δ_η, the integral for the compound distribution Eq. 5.144 is approximated by a sum

$$\int \mathrm{p}(\mathbf{B}|\mathrm{o},\eta) \cdot \mathrm{p}(\eta|\mathrm{o})\,\mathrm{d}\eta \approx \sum_{\eta\in\mathcal{S}_\eta} \mathrm{p}(\mathbf{B}|\mathrm{o},\eta) \cdot \mathrm{p}(\eta|\mathrm{o})\,\Delta\eta\,. \tag{5.176}$$

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.

This chapter is a transcript of Schanner, Korte, and Holschneider (2021)

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

Submitted to JGR: Solid Earth on September 6, 2021.

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 reintroduced 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 mantle 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 data is not sufficient to support global models.

6.1 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. Constable and Korte, 2015). This includes studying the evolution of field features, such as dipole decay, the South Atlantic Anomaly (SAA) and flux patches (Hartmann and Pacca, 2009; Jackson and Finlay, 2015). In the past, several techniques for constructing global field models have been developed and employed. Truncated SH in the spatial domain combined with spline interpolation in time are widely used (Jackson, Jonkers, and Walker, 2000; Korte, Donadini, and Constable, 2009). In the eighties, 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 and Gillet, 2018; Nilsson and Suttie, 2021).

Even though Bloxham and Jackson (1992) already discussed the accurate assessment of uncertainties, most following studies did not pursue the suggested analytical approaches, and either did not report uncertainties at all (Jackson, Jonkers, and Walker, 2000; Constable, Johnson, and Lund, 2000; Korte and Constable, 2003) or relied on ensemble techniques to quantify (modeling related) errors (Korte, Donadini, and Constable, 2009; Licht et al., 2013; Pavón-Carrasco et al., 2014; Hellio and 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 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; 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 (Schanner et al., 2021, in the following referred to as SMKH21) and propose a new global geomagnetic field model for the Holocene.

Global geomagnetic field models on archaeological scales are inferred from two classes of data: Data from materials with thermoremanent magnetisation, such as volcanic rocks, bricks or burnt clay fragments from archaeologic sites, and data from marine or lacustrine sediments with embedded magnetic particles. As in earlier studies, we focus on the former class and loosely refer to it as archeomagnetic data in this paper. 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 MCMC based strategy, we rely on a deterministic inversion based on Kalman-filtering.

The rest of this article is structured as follows: In Section 6.2 we discuss prior assumptions, showcase the modeling method and introduce the dataset. Section 6.3 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 6.4. We conclude in Section 6.5 by reconsidering possible extensions and shortcomings of the method, as well as an outlook to future work.

6.2 Method and Data

6.2.1 Gaussian process based modeling

In the eighties Constable and Parker (1988) proposed using GPs to model the 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 $\bar{\mathbf{B}}$ and a covariance function $\mathrm{K}_{\mathbf{B}}$

$$\mathbf{B} \sim \mathcal{GP}\left(\bar{\mathbf{B}},\, \mathrm{K}_{\mathbf{B}}\right). \tag{6.177}$$

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. 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 and 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.

6.2.2 Data model

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

$$o(\mathbf{x}, t) = \mathbf{H}(\mathbf{B}(\mathbf{x}, t)) + \mathbf{E}\,, \tag{6.178}$$

where the observation functional $\mathbf{H} = (D, I, F)$ contains the usual expressions for declination, inclination and intensity and $\mathbf{E}$ 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 (6.178):

$$D \approx \tilde{D} + \frac{1}{\tilde{F}_H^2} \begin{bmatrix} -\tilde{B}_E \\ \tilde{B}_N \\ 0 \end{bmatrix}^\top \mathbf{B}\,, \tag{6.179}$$

$$I \approx \tilde{I} + \frac{1}{\tilde{F}_H} \left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{\tilde{B}_Z}{\tilde{F}} \frac{\tilde{\mathbf{B}}}{\tilde{F}} \right)^\top \mathbf{B}\,, \tag{6.180}$$

$$F \approx \frac{\tilde{\mathbf{B}}^\top}{\tilde{F}} \mathbf{B}\,. \tag{6.181}$$

$\tilde{D}, \tilde{I}, \tilde{F}$ and $\tilde{\mathbf{B}}$ indicate the POE and we summarize the linearized expressions as $\mathbf{H}_{\text{lin.}}$. The observation errors $\mathbf{E}$ 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 and Nilsson, 2019)

$$\sigma_I = \frac{57.3^\circ}{140} \alpha_{95} \qquad \text{and} \qquad \sigma_D = \frac{1}{\cos o_I} \sigma_I\,. \tag{6.182}$$

We label these approximate errors $\mathbf{E}_{\text{prox.}}$. 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, even though dating errors for archaeological artifacts may have a non-Gaussian distribution. This error in the inputs is handled by another linearization, as proposed by McHutchon and Rasmussen (2011, the 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

$$\Sigma_{tt'} \circ \partial_t \partial_{t'} K_{\mathbf{B}}(\mathbf{x}, t; \mathbf{x}', t')|_{t_o} . \tag{6.183}$$

$\Sigma_{tt'}$ is the covariance matrix of the dating errors and $\circ$ is the Hadamard product (see also Schanner et al., 2021, Sec. 2.4). The effect of the NIGP model is thus the inclusion of dating errors as contributions to the data covariance, similar to measurement errors. To realize this, dating errors are weighed by the second order time derivative of the kernel. The idea is related, but not equal, to the approach of estimating the contribution of dating uncertainties by using the secular variation (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 $\boldsymbol{\xi}\mathbf{P}$ is added to cover any effects that are not modeled, like crustal field or ellipticity of the Earth (see also Mauerberger et al., 2020; Schanner et al., 2021). This way, the data model reads

$$o(\mathbf{x}, t) \approx \mathbf{H}_{\text{lin.}} \left(\mathbf{B}(\mathbf{x}, t_o) - e_t^\top \partial_t \mathbf{B}(\mathbf{x}, t)|_{t_o} + \boldsymbol{\xi}\mathbf{P} \right) + \mathbf{E}_{\text{prox.}} . \tag{6.184}$$

6.2.3 A priori process

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

$$\Phi(\mathbf{x}, t) = R \sum_{\ell} \left(\frac{R}{|\mathbf{x}|} \right)^{\ell+1} \sum_{-\ell \le m \le \ell} g_\ell^m(t)\, Y_\ell^m(\hat{\mathbf{x}}) . \tag{6.185}$$

$\hat{\mathbf{x}}$ is the unit vector $\mathbf{x}/|\mathbf{x}|$ 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

$$\mathbf{B} = -\nabla \Phi , \tag{6.186}$$

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 mean function for the axial dipole is assumed constant, with value γ_1^0. We assume all coefficients to be uncorrelated and identically distributed at a reference radius R = 2800 km (within the Earth's core). This is the "virtual" source region where the spectrum is flat and the field has no direct physical meaning. The magnetic field given by this assumption is only a valid representation of the actual field above the 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 and Gillet, 2018; Baerenzung et al., 2020; Ropp et al., 2020; Nilsson and Suttie, 2021). This way, the temporal correlation of each coefficient is given by

$$\rho_\ell(t - t') = \left(1 + \frac{|t - t'|}{\tau_\ell}\right) \exp\left(-\frac{|t - t'|}{\tau_\ell}\right). \tag{6.187}$$

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

$$\tau_\ell = \frac{\tau_{\mathrm{ND}}}{\ell}. \tag{6.188}$$

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

6.2.4 Sequentialization

In previous studies (Mauerberger et al., 2020; 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 $\mathbf{z} = (g_\ell^m, \dot{g}_\ell^m)$. The prediction equations from step i to i + 1 are given by

$$\mathbb{E}\left[\mathbf{z}_{i+1|i}\right] = \mathbf{F}_i \mathbb{E}\left[\mathbf{z}_i\right] \tag{6.189}$$

$$\mathrm{Cov}\left[\mathbf{z}_{i+1|i}, \mathbf{z}_{i+1|i}\right] = \mathbf{F}_i \mathrm{Cov}\left[\mathbf{z}_i, \mathbf{z}_i\right] \mathbf{F}_i^\top + \tilde{\Sigma}, \tag{6.190}$$

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 $\tilde{\Sigma} = \Sigma - \mathbf{F}\Sigma\mathbf{F}^\top$ with the a priori correlations Σ. 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, Tung, and Striebel (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.

6.2.5 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 and Gillet, 2018; Nilsson and 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

$$\mathcal{L}_{\text{fwd.}} = \sum_{i=1}^{N} \left[-\ln|\Sigma_{o,i}| - \frac{1}{2}\left(o_i - H_{\text{lin.}}\mathbf{B}(\mathbf{z}_i)\right)^\top \Sigma_{o,i}^{-1}\left(o_i - H_{\text{lin.}}\mathbf{B}(\mathbf{z}_i)\right) \right] \tag{6.191}$$

with observations in the i-th step o_i and their observation covariance $\Sigma_{o,i}$. N refers to the number of steps in the Kalman filter. 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; King, 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:

$$-100\,\mu\mathrm{T} \leq \gamma_1^0 \leq -10\,\mu\mathrm{T} \qquad 1\,\mu\mathrm{T} \leq \alpha_\bullet \leq 1000\,\mu\mathrm{T}$$
$$10\text{ yrs.} \leq \tau_\bullet \leq 2000\text{ yrs.} \qquad 0.01\,\mu\mathrm{T} \leq \xi \leq 5\,\mu\mathrm{T}$$

where $\bullet$ stands for DP and ND. Note that γ_1^0 and the $\alpha_\bullet$ are given at the reference radius R = 2800km.

6.2.6 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. Part of the records from Mexico contain wrong age and dating uncertainty estimates (Mahgoub, pers. comm.). Some of them are too old by several thousand years and have been removed, while for others updated ^{14}C ages have been published for the lava flows that they originate from. These updates have been used (see Table 6.4) and will be included in GEOMAGIA in the future. To identify other records that deviate from the rest, we use a Naive Bayes classifier (e.g. Berrar, 2018). 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 $((100^\circ)^2$ for declination, $(50^\circ)^2$ 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.

In the geomagnetic community, it is common to use L1 or Huber norms and employ reweighing techniques to address outliers (Walker and Jackson, 2000; Olsen, 2002). Hellio et al. (2014) and Nilsson and Suttie (2021) implement longer tailed error distributions and include all records in their models. In contrast to these approaches, we resort to outlier rejection for two practical reasons: First, the precise influence of the error distribution on the model is a question that is yet to be addressed by the community. Records which are accepted by the naive Bayes classifier are likely to be described well by a Gaussian error model, and by rejecting the rest we postpone a detailed analysis. Second, by

considering a Gaussian error model only, the inversion is feasible analytically and can be addressed by numerical linear algebra, without resorting to sampling techniques. We acknowledge that the error model is a point that could be addressed more thoroughly in the future.

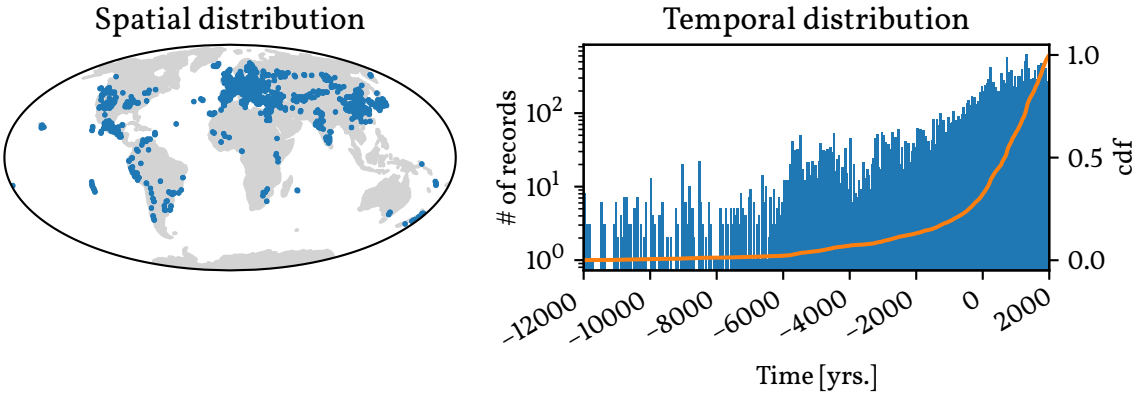


Figure 6.1: Spatial and temporal distribution of the data. Every declination, inclination and intensity is counted as one record and represented by one dot. Note the logarithmic scale (left) on the histogram. To emphasize the inhomogeneity in the temporal distribution, the normalized cumulative sum of the data is shown in orange (right scale).

6.3 Results

6.3.1 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 6.2.6. 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 6.1 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 variations that are present in the data can be resolved with the larger a priori correlation time and shorter variations can not be recovered. 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 6.2 shows generated and inferred axial dipole and quadrupole. Especially in earlier times, the quadrupole is better resolved than the axial dipole. This may be due to the data distribution, as indicated by Figure 6.3. We there show the relative covariance gain (the diagonal of the last term on the right hand side of Eq. (7) in Schanner et al. (2021), relative to the prior covariance) per coefficient for a stationary inversion. This quantity only depends on the data distribution. With the actual distribution, the gain is stronger in the quadrupole. With a symmetrized dataset, large scale degrees are resolved more evenly. A similar behavior is also visible in the power spectra of the test inversion, which are provided with the supplementary material (Figure 6.13). In general, Figure 6.2 shows a promising agreement, although some variation in the dipole, prominently between 10000 and 8000 BCE, is not present 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 (Fig. 6.12).

Table 6.1: Hyperparameters that have been used to generate synthetic data for the validation ("fixed") and the ones inferred using the proposed method. 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. ξ is the scaling factor of the residual term. Note that γ_1^0 and $\alpha_\bullet$ are given at the reference radius.

Model	$\gamma_1^0\,[\mu\mathrm{T}]$	$\alpha_{\mathrm{DP}}\,[\mu\mathrm{T}]$	τ_{DP} [yrs.]	$\alpha_{\mathrm{ND}}\,[\mu\mathrm{T}]$	τ_{ND} [yrs.]	$\xi\,[\mu\mathrm{T}]$
Fixed	-412.3	13.8	250	39.4	393	-
Inferred	-408.55	9.87	302.48	30.70	724.76	0.01

6.3.2 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 ArchKalmag14k, 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 6.2. We compare ArchKalmag14k to the models ARCH10k.1 (Constable, Korte, and Panovska, 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 6.2 gives 281 sets of 440 main field and 440 secular variation coefficients together with the respective covariances, one set every 50 years. A comparison of the model coefficients to the prior is given with the supplementary material (Figure 6.14). Figure 6.4 shows the dipole and axial quadrupole and

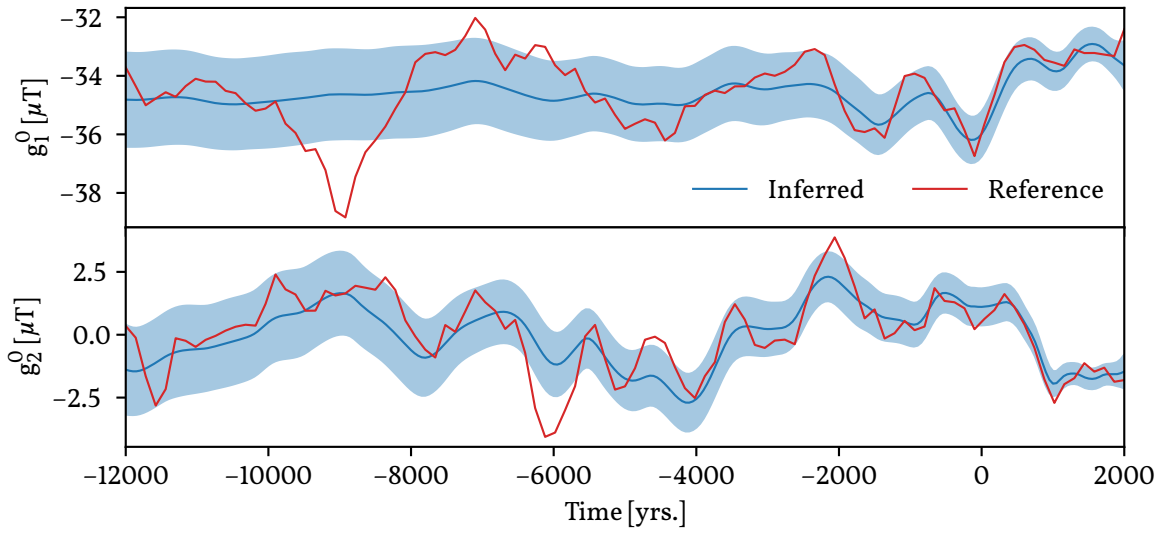


Figure 6.2: Axial dipole (top) and quadrupole (bottom) of the synthetic model, together with the corresponding inferred ones from the proposed inversion. The inferred (blue) and reference curves (red) agree within the pointwise 95%-regions shown in light blue. Some variations, most prominently in the axial dipole between 10000 and 8000 BCE, can not be resolved.

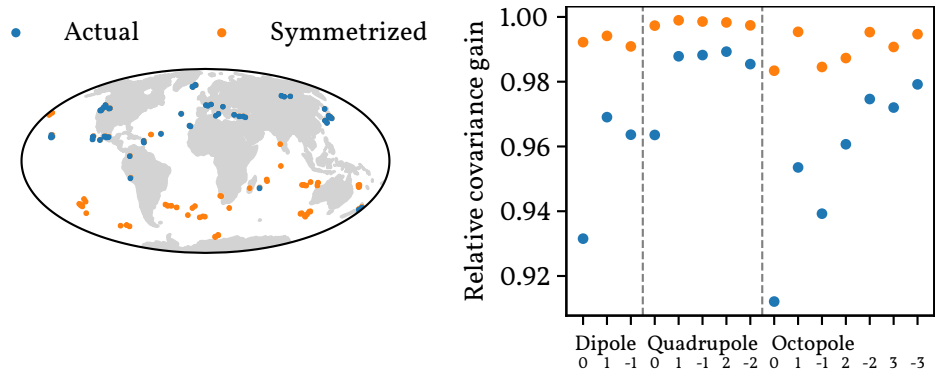


Figure 6.3: Relative covariance gain per degree for a stationary inversion. The blue dots refer to the actual dataset, while orange ones refer to a symmetrized version. The bottom row of the horizontal axis refers to the coefficient order.

Table 6.2: Prior hyperparameters for ArchKalmag14k. Note that γ_1^0 and $\alpha_\bullet$ are given at the reference radius. At the Earth's surface, $\gamma_1^0 \approx -36.19\,\mu$T.

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

octopole coefficients together with 95%-uncertainties and comparison models. The proposed model ArchKalmag14k shows less variation in the dipole degrees than comparable models, especially 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 6.5 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 (i.e. a quantity non-linearly derived from the posterior GP), 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 6.6 and 6.7 show local curves for Paris and Hawaii respectively. Data from a surrounding of 250km around the respective location is included with the prediction. Inclination and intensity are translated along the corresponding axial dipoles (Merrill, McElhinny, and McFadden, 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. Two additional local predictions, for the Indian ocean and New Zealand, are provided with the supplementary material (Figures 6.15 and 6.16).

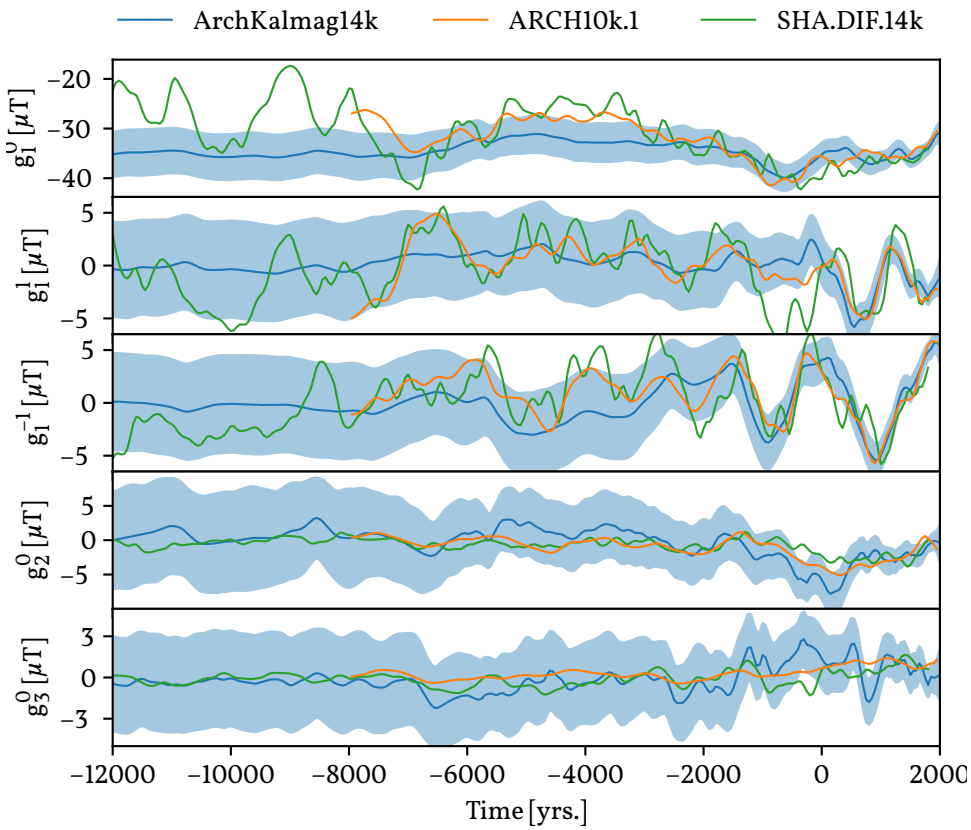


Figure 6.4: Gauss coefficients of the dipole and the axial quadru- and octopole. ArchKalmag14k is shown in blue. The shaded area covers the pointwise 95%-region. ARCH10k.1 is shown in orange and SHA.DIF.14k in green.

We investigate the misfit of the model in table 6.3. The sum of residuals squared, divided by the variance is a χ^2-distributed random variable:

$$\mathcal{T} = \sum_{i=1}^{N} \frac{\left(o_i(\mathbf{x}, t) - H_i[\mathbf{B}(\mathbf{x}, t)]\right)^2}{\sigma_{o,i}^2 + \sigma_{t,i}^2 + \sigma_{B,i}^2} \sim \chi_N^2 \tag{6.192}$$

$\sigma_{o,i}$ is the observation error, $\sigma_{t,i}$ the weighed dating uncertainty as given by Eq. (6.183) and $\sigma_{B,i}$ is the posterior standard deviation. N refers to the number of records. The normalized misfit can be calculated from $\mathcal{T}$ as

$$\mathcal{M} = \sqrt{\frac{\mathcal{T}}{N}}\,. \tag{6.193}$$

It is evident, that $\mathcal{T}$ lies below the 95% confidence interval of the corresponding χ^2 distribution in almost all cases. Also $\mathcal{M}$ is below 1 for all subsets. Both results might indicate, that the model is overfitting the data, but when looking at the relatively big MAE, this does not seem to be the case. Instead, we think the reason for the low misfit and $\mathcal{T}$

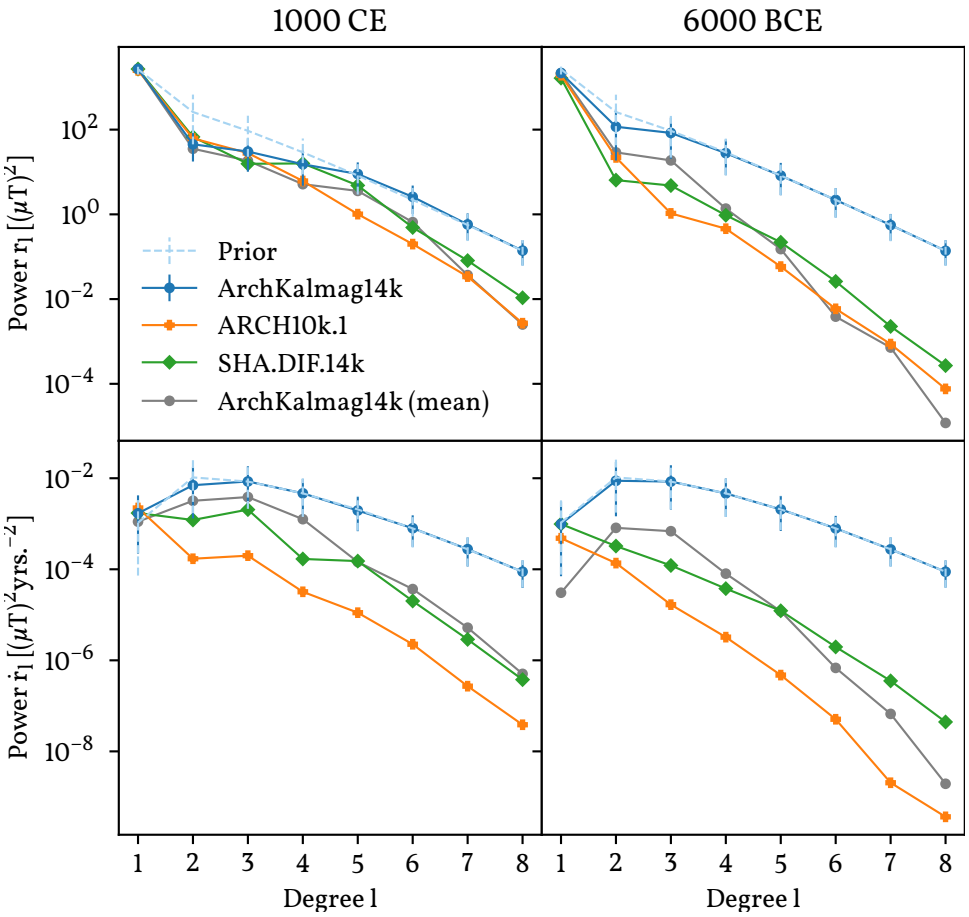


Figure 6.5: Geomagnetic main field (top) and secular variation spectra (bottom) at Earth's surface for two selected epochs. The random variable power spectrum for ArchKalmag14k is shown in blue. The errorbars report 2.5- and 97.5-percentiles, covering 95%. For comparison, the spectra of the mean model 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.

values lies in the large contributions from the dating uncertainties. These may be overestimated, as we use the reported uncertainties as standard deviations for normal error distributions. If in fact they come from a different distribution, for example a uniform one, this procedure gives errors that are too large. The large errors also result in a low impact of such records on the model. The large MAE may be caused by those records as well. As discussed above, a more thorough treatment of the error model (and with this also the dating uncertainties) in general may be necessary to address this.

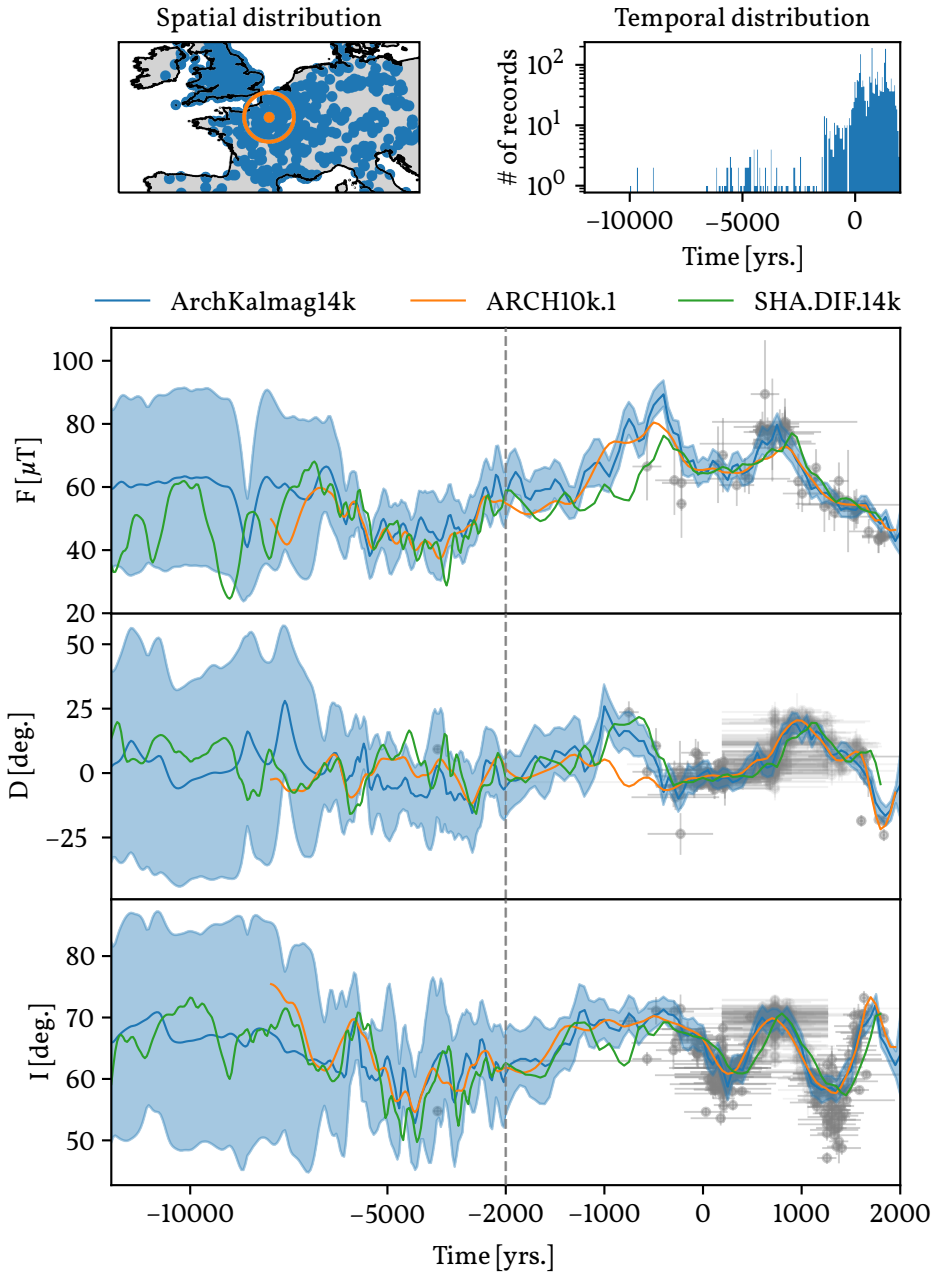


Figure 6.6: Local predictions of intensity F, declination D and inclination I for Paris. Note the different timescales in the left and right parts of the bottom panels! ArchKalmag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. In the top row, the spatial and temporal distribution of the surrounding are shown. Data in the orange ellipse (250km radius) are translated (Merrill, McElhinny, and McFadden, 1996) to the location of prediction (orange dot) and shown as gray dots. Horizontal and vertical gray bars indicate the two sigma temporal and field component data uncertainties, respectively. The temporal distribution (top right) includes all data visible in the top left plot.

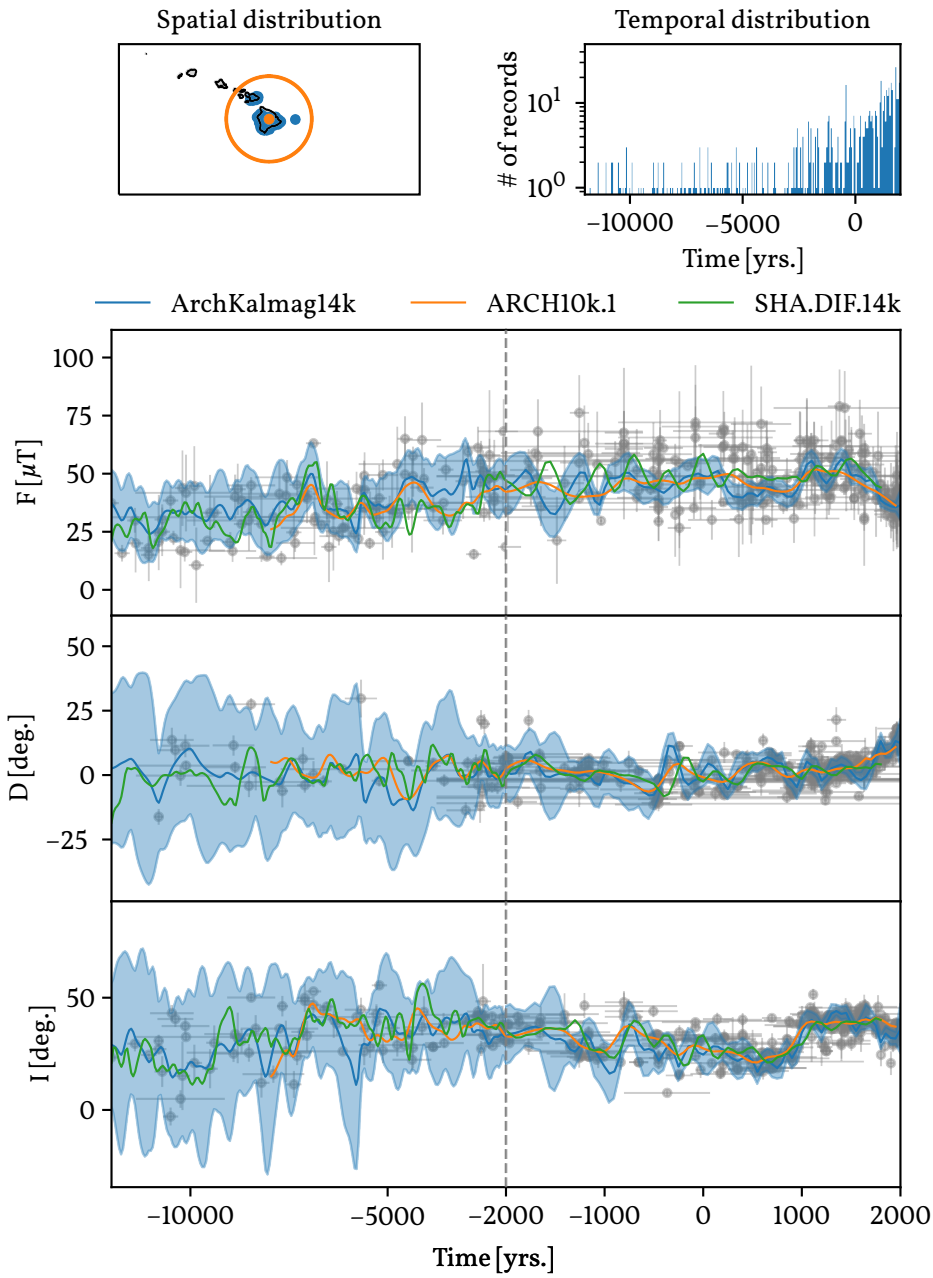


Figure 6.7: Local predictions of intensity F, declination D and inclination I for Hawaii. Note the different timescales in the left and right parts of the bottom panels! ArchKalmag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. In the top row, the spatial and temporal distribution of the surrounding are shown. Data in the orange ellipse (250km radius) are translated (Merrill, McElhinny, and McFadden, 1996) to the location of prediction (orange dot) and shown as gray dots. Horizontal and vertical gray bars indicate the two sigma temporal and field component data uncertainties, respectively. The temporal distribution (top right) includes all data visible in the top left plot.

Table 6.3: Data misfit for several data subsets. N refers to the number of records in the data subset. $\mathcal{T}$ gives the sum of the residuals squared, divided by the variances. This is a χ^2-distributed random variable and we give the 95%-intervals of the corresponding distribution as CI_{χ^2}. $\mathcal{M}$ is the normalized misfit, i.e. $\sqrt{\mathcal{T}/\mathrm{N}}$. MAE refers to the mean absolute error. The early subset consists of all records before -6000 BCE, the middle subset covers -6000 BCE to 0 CE and the recent subset includes all records after 0 CE. A map of the localized subsets is given with the supplementary material (Fig. 6.17).

Name	Type	N	$\mathcal{T}$	CI_{χ^2}	$\mathcal{M}$	MAE
Early	D	119	19.0	[90.7, 151.1]	0.40	9.9°
	I	132	13.6	[102.1, 165.7]	0.32	7.1°
	F	155	39.6	[122.4, 191.4]	0.51	8.7 μT
Middle	D	1307	489.0	[1208.7, 1409.1]	0.61	7.3°
	I	1372	461.2	[1271.2, 1476.6]	0.58	4.1°
	F	3010	1984.7	[2859.8, 3164.0]	0.81	7.1 μT
Recent	D	4208	2494.5	[4030.1, 4389.7]	0.77	5.1°
	I	5552	5346.2	[5347.4, 5760.4]	0.98	3.4°
	F	2958	1829.7	[2809.2, 3110.6]	0.79	6.0 μT
Europe	D	3434	1802.3	[3273.5, 3598.3]	0.72	5.8°
	I	3952	2916.1	[3779.7, 4128.1]	0.86	3.0°
	F	2546	1987.3	[2408.0, 2687.7]	0.88	6.3 μT
North America	D	943	515.0	[859.8, 1030.0]	0.74	5.4°
	I	945	703.2	[861.7, 1032.1]	0.86	4.3°
	F	488	370.4	[428.7, 551.1]	0.87	8.0 μT
East Asia	D	349	214.1	[299.1, 402.6]	0.78	5.6°
	I	412	365.9	[357.7, 470.1]	0.94	4.2°
	F	601	331.0	[535.0, 670.8]	0.74	7.2 μT
South Pacific	D	0	-	-	-	-
	I	0	-	-	-	-
	F	28	11.8	[15.3, 44.5]	0.65	5.2 μT

6.3.3 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 6.8 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. Interestingly we observe a higher dipole moment than the comparison models for the interval 6000 to 2000 BCE and also from 12000 to 8000 BCE.

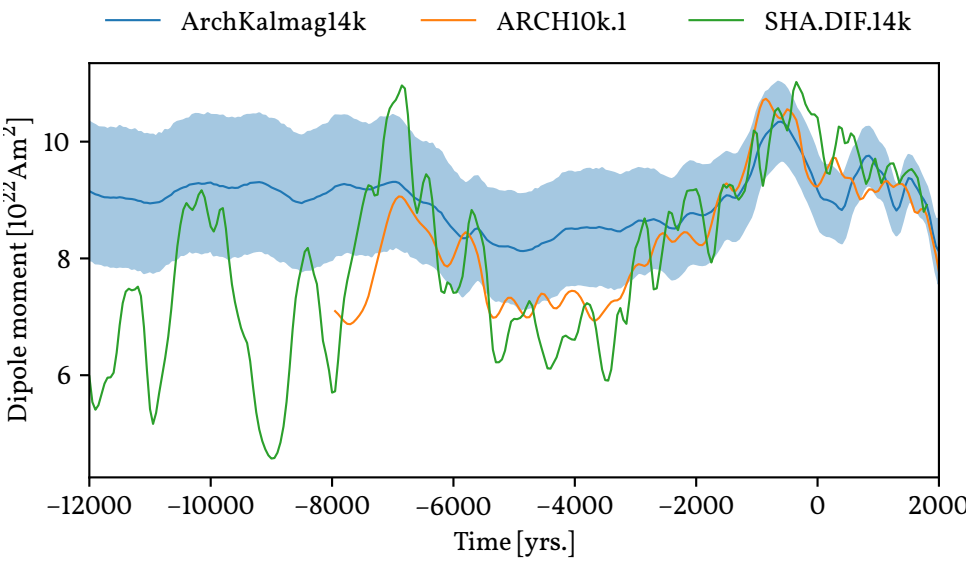


Figure 6.8: Dipole moment of the geomagnetic field. ArchKalmag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. Mean and standard deviation of ArchKalmag14k are inferred from sampling. This sampling is the reason for the small scale noise in the blue curve and area.

Figure 6.9 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; 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. 6.182). 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 6.9, bottom row) increases towards earlier times, as is expected from the thinning data distribution.

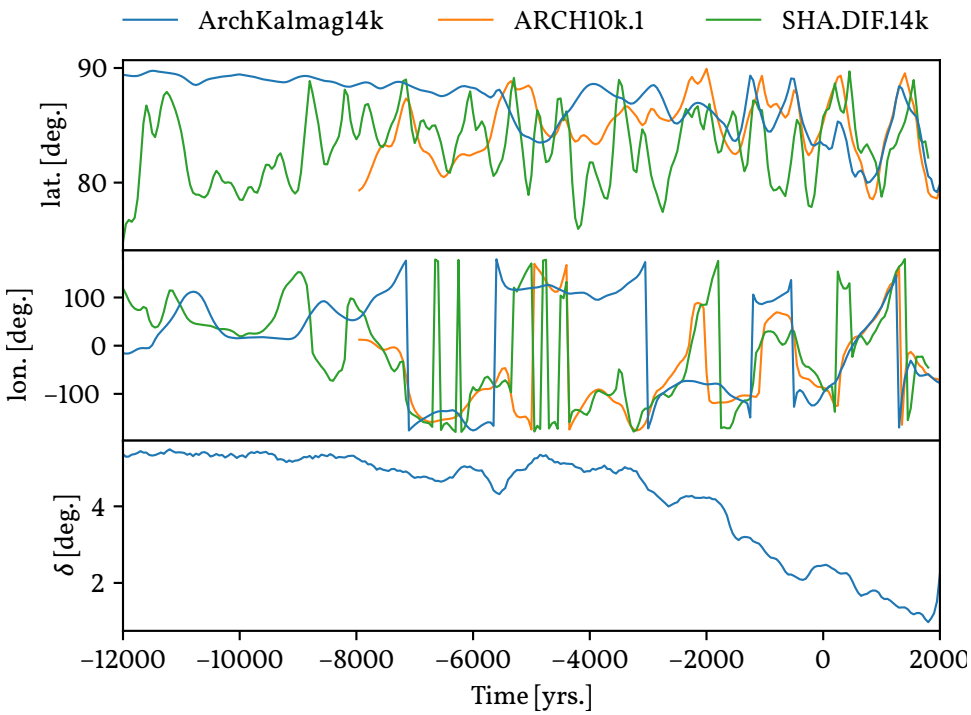


Figure 6.9: Latitude (top) and longitude (middle) of the geomagnetic dipole axis. ArchKalmag14k is shown in blue. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. The bottom plot shows the angular standard deviation δ (Butler, 2004) for ArchKalmag14k, which is inferred from sampling. This sampling is the reason for the small scale noise in earlier times.

6.3.4 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 and 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 model is not resolved reliably 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 (bottom right in Figure 6.11). 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. 6.10, b)). This contrasts the findings of Campuzano et al.

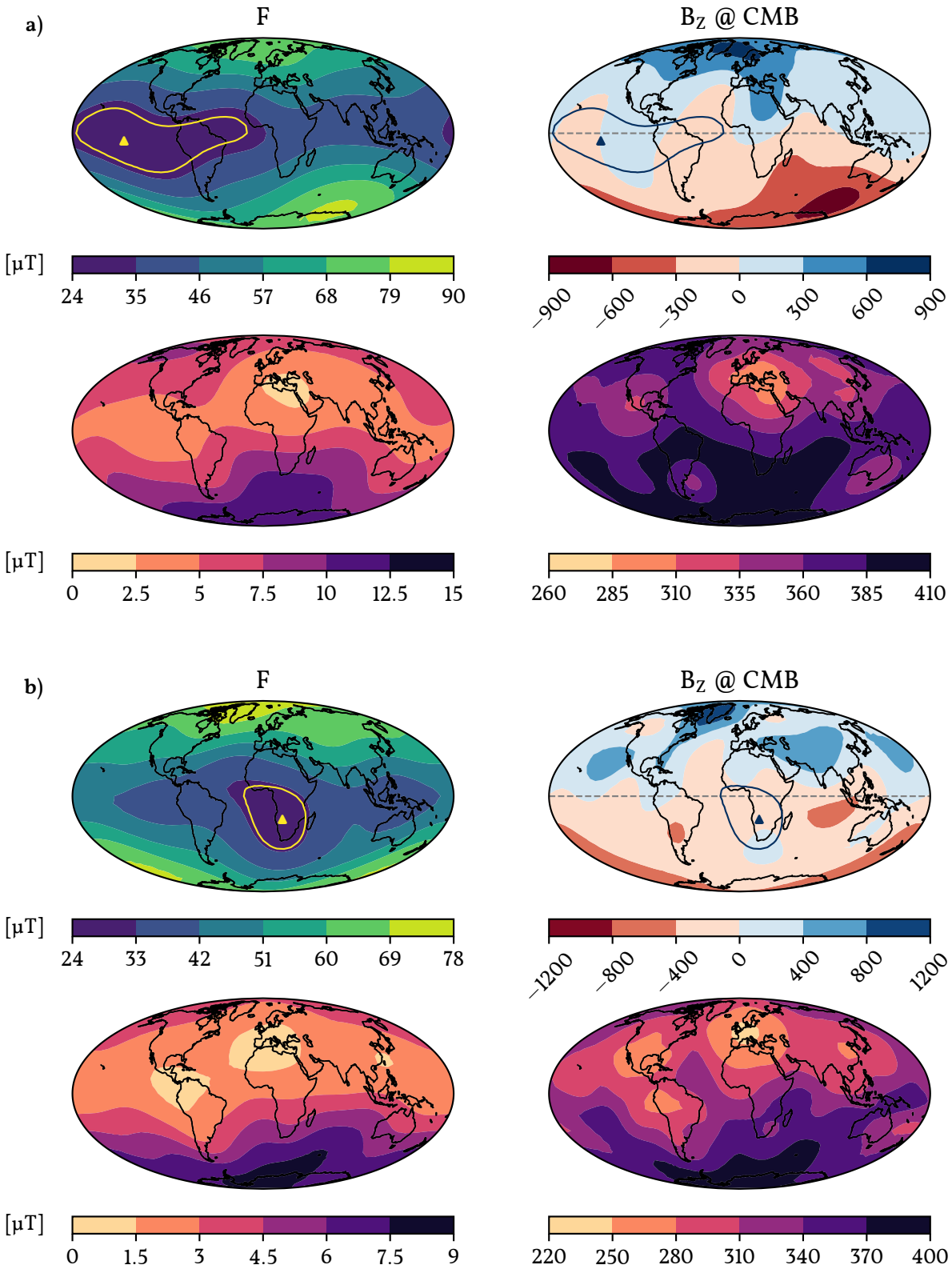


Figure 6.10: The South Atlantic Anomaly (SAA). The top rows show the field intensity at the Earth's surface and the magnetic field radial component (downwards). The bottom rows show the respective standard deviation. a) is for the year 1500 BCE and b) for 1300 CE. The yellow triangles indicate the location 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.

(2019), where the low intensity region emerges approximately 100 years earlier close to Madagascar, although an earlier emergence is within the uncertainties of our model. 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 (bottom left in Figure 6.11). 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 reliable hints on a low intensity field region in the Indian ocean are present around 3000 BCE, with the region drifting eastward (top left in Figure 6.11) and a second low intensity region appearing over the Northern part of South America at 2600 BCE (top right in Figure 6.11). 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. 6.10, 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.

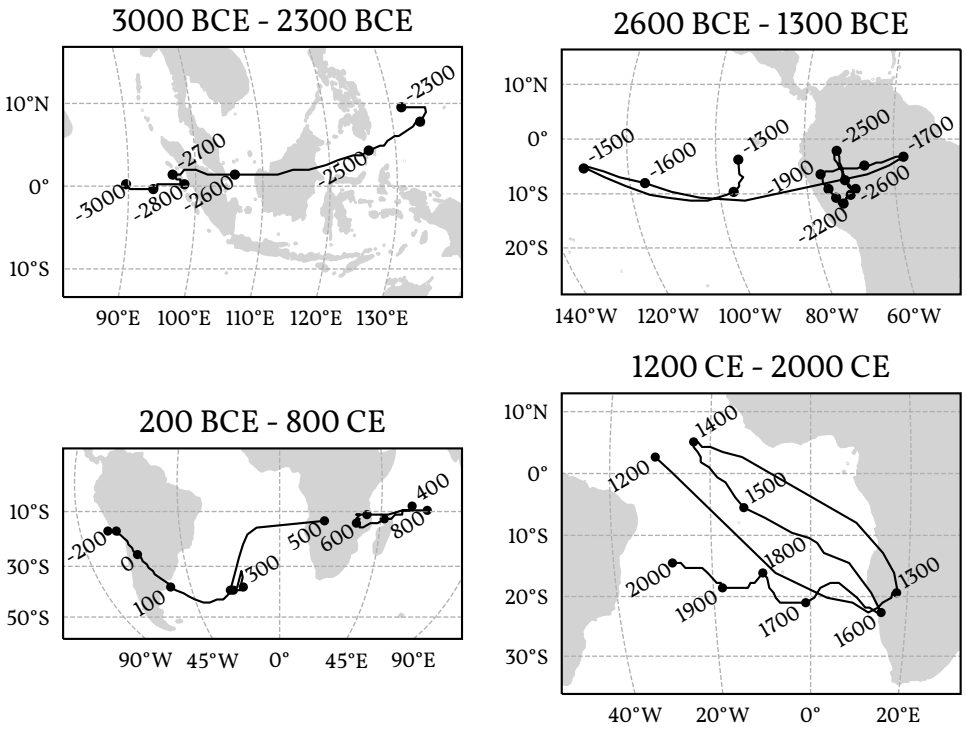


Figure 6.11: Paths of the minima of low intensity field regions described in the text. Dots are drawn every hundred years. The locations of the minima have been obtained using grid search, which leads to several discrete jumps in the curves.

6.4 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 (Figure S3)). If the data cannot resolve the global scales, the prior is reproduced, which is evident from local curves with no data coverage (Fig. 6.6) and the analysis of the dipole itself (Figs. 6.8 and 6.9). For times earlier than 6000 BCE, the axial dipole varies only slightly around the prior mean value of approx. –36.19 μT (Fig. 6.4, top row; see also Figure 6.14 in the supplementary material). Nevertheless, local variations are resolved, if supported by the data (Fig. 6.7, 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. 6.5, 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 and Gillet, 2018; Nilsson and 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 (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 6.4 and Figure 6.8. 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 adaptation of the modeling method.

6.5 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; 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 (Schanner, Mauerberger, and Korte, 2020), so that feature analysis is straight-forward. Together with the software, we provide a Jupyter notebook, that illustrates how to reproduce ArchKalmag-14k. The model is also available via a website: https://ionocovar.agnld.uni-potsdam.de/Kalmag/Archeo/

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 and Suttie, 2021) shows, this requires significant modifications due to aspects of the sedimentation process and the respective statistical implications.

6.6 Supplementary material

Introduction

This supplementary material provides validation plots for additional coefficients in Figure 6.12 and spectra of the synthetic inversion in Figure 6.13. A comparison of the model coefficients with the prior is given in Figure 6.14 and local field predictions at two additional locations in Figures 6.15 and 6.16. Figure 6.17 shows the local data subsets used for misfit evaluation. Table 6.4 contains a list of changes made to the GEOMAGIA v.3.4 dataset (Brown et al., 2015). A separately available Movie shows the evolution of the geomagnetic field intensity at the Earth's surface and of the radial component (downwards) at the core-mantle boundary, together with respective uncertainties.

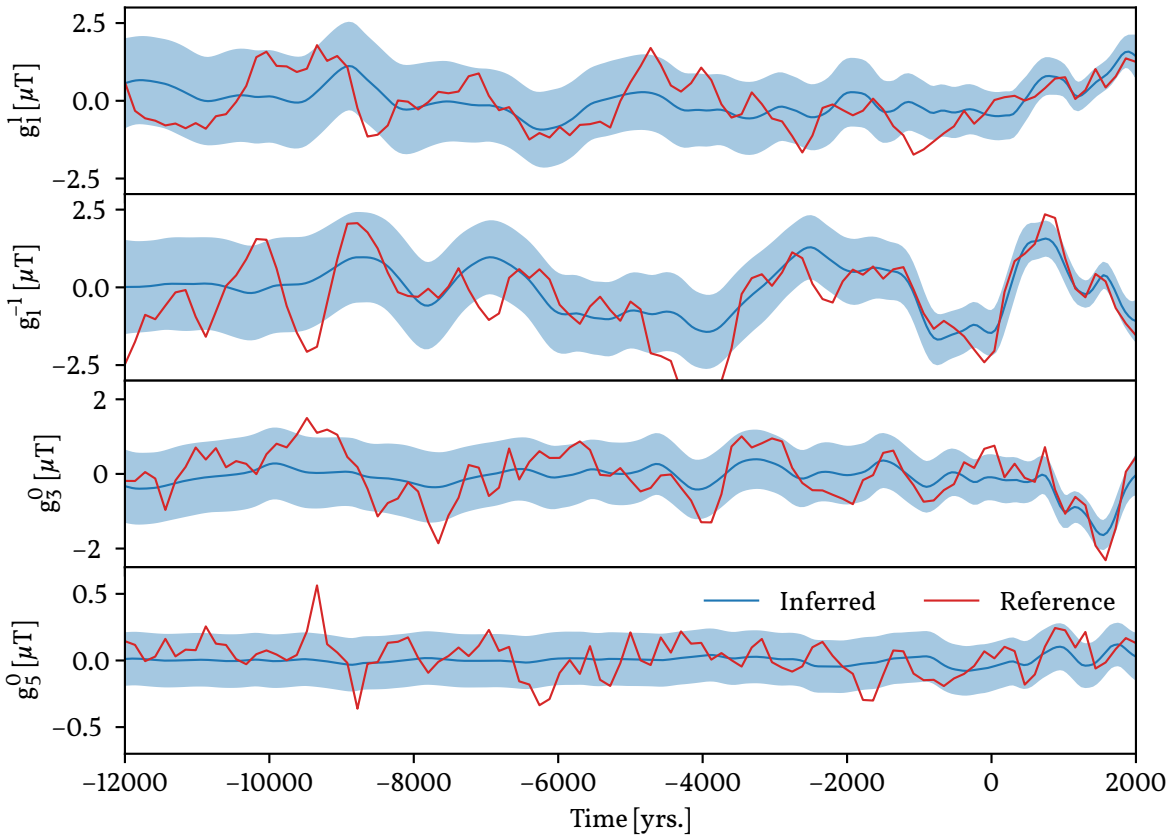


Figure 6.12: Additional dipole and higher order coefficients of the synthetic model, together with the corresponding inferred ones from the proposed inversion. The inferred (blue) and reference curves (red) agree within the pointwise 95%-regions shown in light blue.

Table 6.4: 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

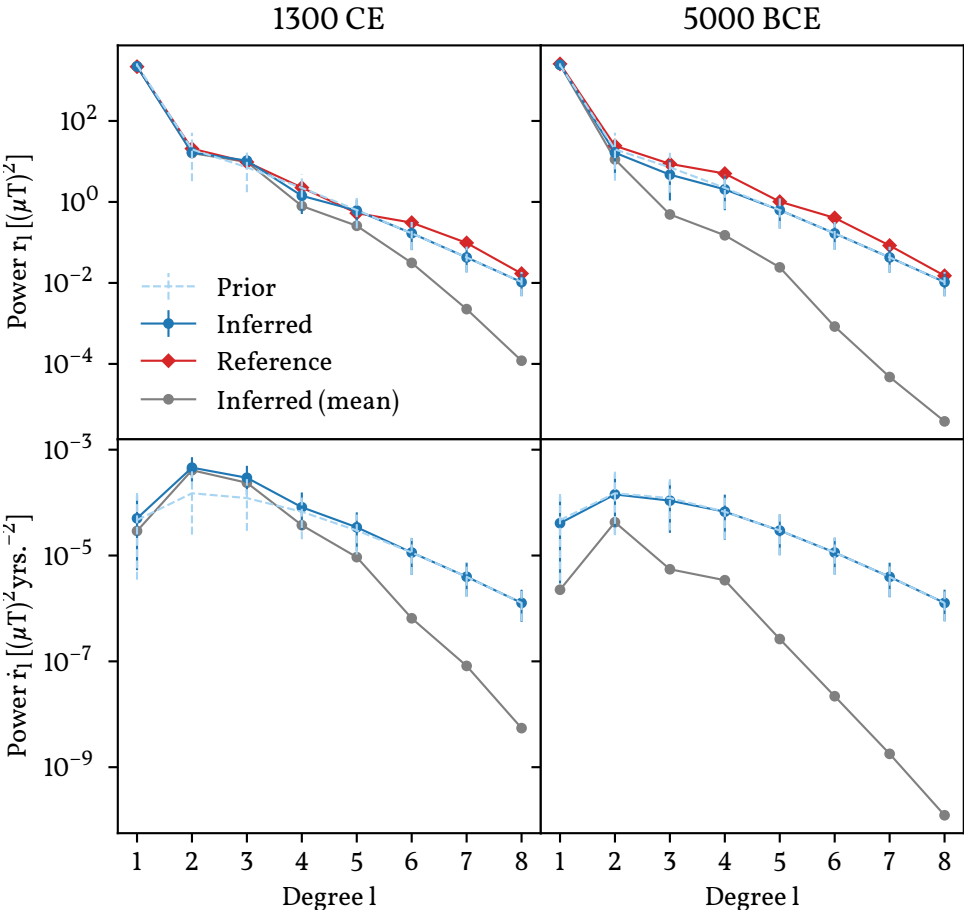


Figure 6.13: Geomagnetic main field (top) and secular variation spectra (bottom) at Earth's surface for two selected epochs of the synthetic test case. The random variable power spectrum is shown in blue. The errorbars report 2.5- and 97.5-percentiles, covering 95%. For comparison, the spectra of the mean model are shown in grey. The prior spectrum is shown as a light blue dashed line. The main field spectrum of the reference model is shown in red, the secular variation spectrum of the reference model is not available.

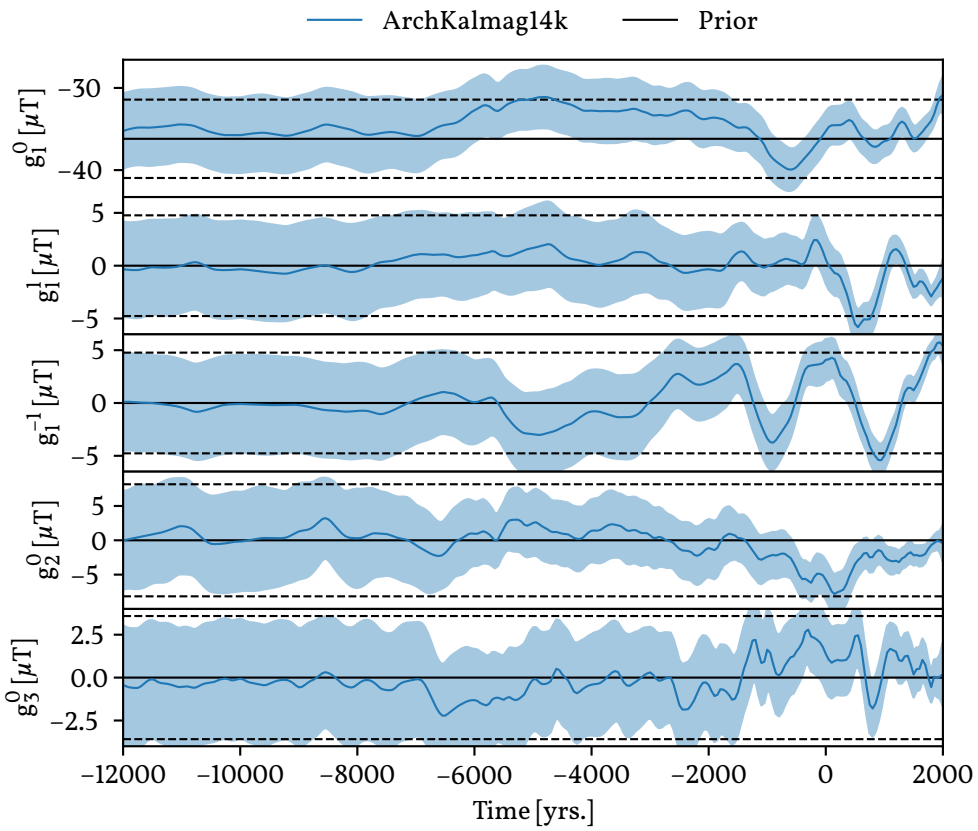


Figure 6.14: ArchKalmag14k model coefficients together with the prior. The shaded area and dashed lines indicate the pointwise 95%-regions.

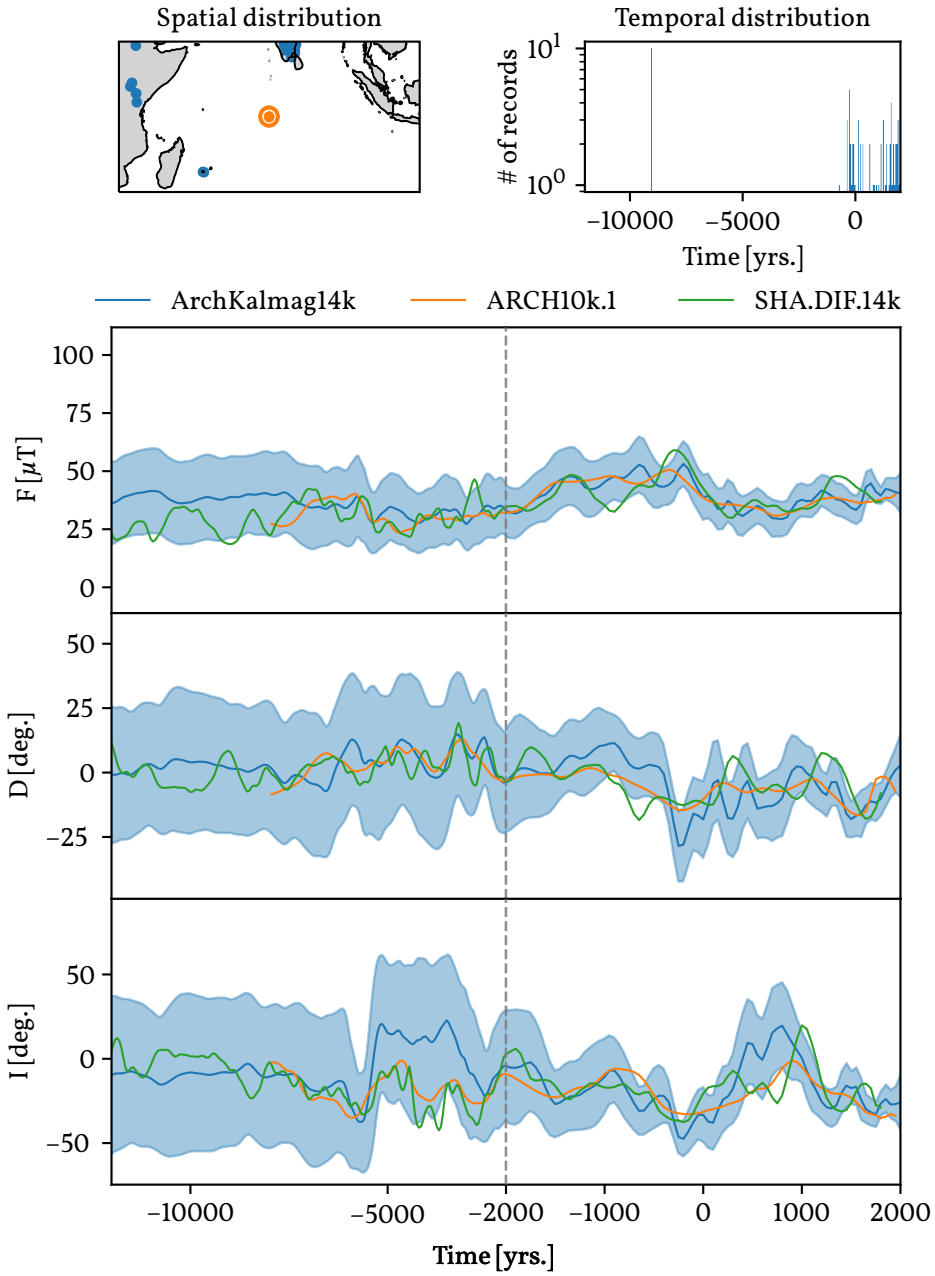


Figure 6.15: Local predictions of intensity F, declination D and inclination I for the Indian Ocean. Note the different timescales in the left and right parts of the bottom panels! ArchKalmag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. In the top row, the spatial and temporal distribution of the surrounding are shown. In contrast to the other local predictions (Fig. 6.6, 6.7 and 6.16) no data is present in the lower plots, as no data is available in the region. Horizontal and vertical gray bars indicate the one sigma temporal and field component data uncertainties, respectively. The temporal distribution (top right) includes all data visible in the top left plot.

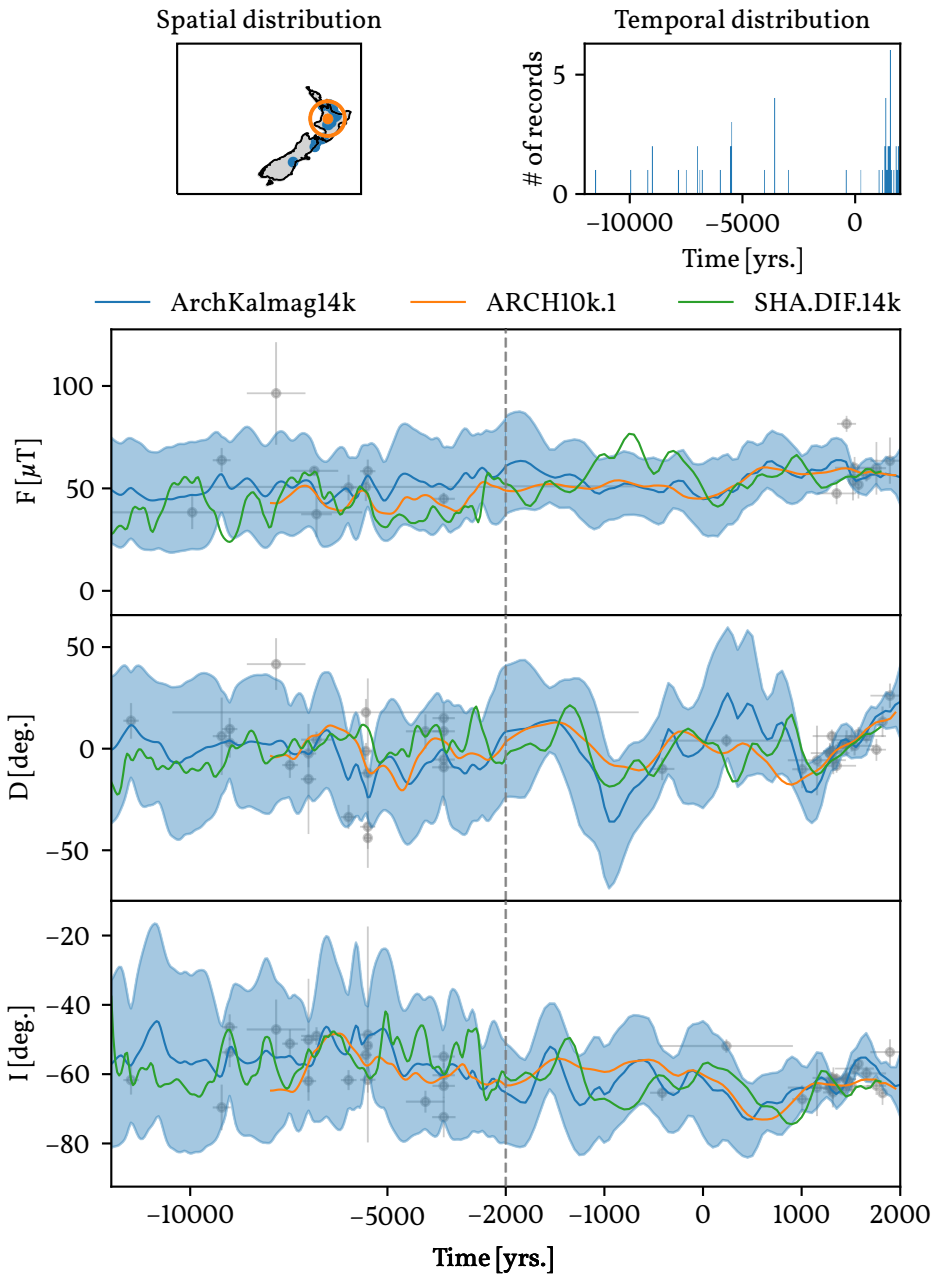


Figure 6.16: Local predictions of intensity F, declination D and inclination I for New Zealand. Note the different timescales in the left and right parts of the bottom panels! ArchKalmag14k is shown in blue. The shaded area covers 95%. ARCH10k.1 is shown in orange and SHA.DIF.14k in green. In the top row, the spatial and temporal distribution of the surrounding are shown. Data in the orange ellipse (250km radius) are translated (Merrill, McElhinny, and McFadden, 1996) 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 (top right) includes all data visible in the top left plot.

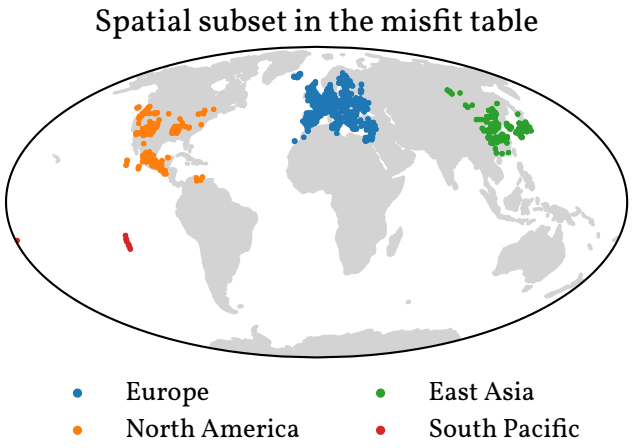


Figure 6.17: Spatial subsets considered for the misfit in Table 6.3.

Part III

Discussion

7 Conclusions and scientific contribution

The culminating result of this thesis is the global geomagnetic field model ArchKalmag-14k, presented in Chapter 6. Resting on the work of the other two publications, included in Chapters 4 and 5, ArchKalmag14k provides new insights on global geomagnetic modeling. A central conclusion is that the current archaeo- and volcanic database before 6000 BCE is not sufficient to support global models. For times more recent, ArchKalmag14k allows feature analysis under consideration of uncertainties arising from modeling, magnetic measurements and dating procedure. Together with the mean model, which has the highest posterior density and is therefore considered for feature analysis, the provided standard deviation allows to discriminate between reliable features and features that are not supported well by data. I carried out an exemplary analysis for the South Atlantic Anomaly (SAA) in Section 6.3.4. Together with the model itself comes a novel modeling method. Due to its nature, the resolution of the model is governed by the data distribution and varies over time. This flexibility allows exploiting the data to its fullest and does not introduce variations on the global scale, if they are only supported locally. On a more general level, in Chapter 6 I adapted the sequentialized modeling technique, proposed by Baerenzung et al. (2020), to the paleomagnetic setting. The work of Chapters 4 and 5 is applied in the correction step of the Kalman-filter (Kalman, 1960), that is utilized for sequentialization.

Initial work, presented in Chapter 4, was conducted in close collaboration with Stefan Mauerberger. The focus of this work was the modification of the kernel based modeling technique for the global geomagnetic field, presented in Holschneider et al. (2016), to make it applicable to paleomagnetic data. Care was taken to be as objective as possible about the prior parameters and our contribution in this regard is the implementation of a non-informative prior for the axial dipole. A major difficulty in comparison to the work of Holschneider et al. (2016) is the non-linear relation of paleomagnetic observations to the magnetic field vector. To tackle this, we linearize the observation functional. This is a common approach (e.g. Hellio et al., 2014), but requires fixing a point to linearize about. In this aspect our work presented in Chapter 4 contains another novelty. The data where full vector information is available is separated from the rest. From this reduced dataset, a first model is built and then serves as a linearization point for the remaining observations in a Bayesian update system (Section 4.3). Figure 4.4 indicates that

this approach to linearization performs better than linearization about any axial dipole, even the axial dipole used for generating the synthetic test dataset. Due to the sequentialized approach used for constructing ArchKalmag14k, neither the non-informative prior nor the Bayesian update system were directly applied in full time-dynamic modeling. However, the idea of updating the linearization point is also found in the adaption of the Kalman-filter, where in every correction step the current model serves as point of expansion.

The static approach of Chapter 4 already provided first results on the assessment of realistic uncertainties, especially due to the marginalization of model parameters by means of a Riemann sum. However, an important source of uncertainty is the dating of magnetic specimen that only comes into play when time dynamic models are considered. I extended the static approach to the temporal domain conceptually by an unphysical squared exponential kernel in Chapter 5. This comes with several difficulties. The non-informative prior is not applicable in this setting, as temporal correlations of the dipole are sent to zero and can not be recovered in the posterior. Instead, a constant axial dipole is included as an additional prior parameter. Together with the temporal correlation times, the parameter space becomes too large to apply the Riemann sum approach. To tackle the high dimensional integration, I implemented recently proposed numerical integration strategies based on central composite design (Sanchez and Sanchez, 2005; Rue, Martino, and Chopin, 2009). Beyond the translation of concepts from Chapter 4, my key contribution in Chapter 5 is the consideration of dating errors by means of the noisy input Gaussian process approach (McHutchon and Rasmussen, 2011). This approach strictly differs from existing studies, as it does not rely on sampling techniques.

In recent years, several statistical approaches to geomagnetic field modeling on different scales have been proposed (e.g. Hellio et al., 2014; Hellio and Gillet, 2018; Nilsson and Suttie, 2021). All of these rely on sampling, either directly because the posterior is approximated via MCMC or indirectly to incorporate dating uncertainties via bootstrapping methods. In this regard, the technique we developed throughout this thesis is different. Sampling is only employed for posterior quantities that do non-linearly relate to the field vector (for example the power spectrum). Instead of sampling, an approximation to the posterior distribution is constructed analytically, by two linearizations and the sequentialization of the inversion. The linearization approach embeds the proposed formalism in the rich theory of RKHS, as the proxy posterior defines a subspace of the RKHS associated to the prior. The prior proposed by Nilsson and Suttie (2021) is almost identical to the one used for ArchKalmag14k, however their approach is aimed towards

sediment records, a class of data excluded completely here. Another central difference to existing models is the determination of model parameters. Throughout Part II, parameters are estimated from the data and marginalized if possible. Most existing strategies rely on estimating the model parameters from models built on satellite data.

Besides the applied mathematics and geoscientific aspects of the presented formalism, a third fragment of the work lies in the software implementation of the algorithms. For every chapter in Part II there is a repository, through which the proposed methods are available to the public. Together with the model ArchKalmag14k I developed a python package, called paleokalmag, that allows easy reproduction of the results of Chapter 6, as well as application of the formalism to other paleomagnetic datasets. A side product of the work in this thesis is the python package pymagglobal (Schanner, Mauerberger, and Korte, 2020), that provides access to existing global field models to the public and replaces some Fortran scripts that were circulating in the geomagnetism community. It is possible to access ArchKalmag14k via pymagglobal as well, so that a unified interface for global field models is provided.

8 Outlook

The presented work can be extended in several directions. The most obvious one is the inclusion of a larger dataset, which is provided by marine and lacustrine drill cores and the corresponding sediment records. Sediment records go further back in time and may fill the gap in data before 6000 BCE, therefore providing access to field models dating farther back in time. However, sediment records come with several additional challenges, that require further extension and adaption of the current method (see also Nilsson and Suttie, 2021). First, only relative variations of paleomagnetic intensity and declination are available from sediments. This requires external calibration, which possibly can be addressed by another two step strategy, where a first model is constructed from archaeological and volcanic data, such as ArchKalmag14k. Second, data along a single core can not be considered independent. Representing this in the modeling method requires significant adaption, especially when it comes to representing the dating uncertainties. Third, sediment records inherently present a smoothed version of magnetic field variations. The amount of smoothing depends on the sedimentation rate and experimental methods and varies from core to core.

Another dataset is given in form of historical data, especially from ship logs and land surveys from the sixteenth to nineteenth century. This data may be fed into the algorithm directly in principle, however the database is too large to be handled with the proposed strategy. If an effective and accurate reduction scheme is developed, the application of the method in Chapter 6 is straight forward.

A different direction considers the approximation of the non-linear likelihood. Either higher order approximations or more sophisticated algorithms may be considered. Especially the expectation propagation algorithm proposed by Minka (2013) appears promising in this regard. The problem of high costs when estimating hyperparameters remains, but it may be possible to estimate these from a simplified, linearized procedure and then use a more sophisticated approximation for the actual model building.

A third direction is the embedding of the proposed strategy into modern GP frameworks, e.g. the one by Gardner et al. (2018). This would allow the access of the full posterior covariance, in other words the direct application of the algorithm of Chapter 5 to large datasets. This is possible due to an efficient, parallelized implementation of approximate Cholesky factorization, which allows the inversion to be performed on GPUs.

The main difficulty here is of technical nature. Existing libraries (gpytorch (Gardner et al., 2018) and KeOps (Charlier et al., 2021)) currently either do not support the update strategy required for effective linearization or the multi-dimensional covariance structure necessary for the modeling of vector data.

Declaration of authorship

I, Maximilian Arthus Schanner, hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text.

July 13, 2022

Maximilian Arthus Schanner

Acronyms

CCD central composite design. 78, 90

CMB core-mantle boundary. 29, 53, 54, 55, 83, 84, 103, 117, 118, 119, 121

EMF Earth's magnetic field. 11, 12, 15, 17, 24, 25, 26, 27, 29, 30, 31, 33, 35, 36, 37, 38, 39, 41, 43, 45, 47, 50, 51, 52, 53, 54, 55, 57, 59, 64, 65, 66, 68, 69, 70, 71, 83, 84, 87, 89, 96, 101, 103, 121

GGP Giant Gaussian Process. 25

GP Gaussian process. 11, 12, 13, 14, 15, 16, 17, 18, 20, 25, 28, 29, 30, 33, 64, 68, 69, 70, 71, 73, 74, 75, 76, 100, 101, 102, 103, 104, 105, 110, 135

IGRF International Geomagnetic Reference Field. 29, 30, 39, 50

IID independent and identically distributed. 28, 29

MAE mean absolute error. 40, 41, 111, 112, 115

MAP maximum a posteriori probability estimator. 17, 78, 79, 82, 91, 93, 97

MCMC Markov Chain Monte-Carlo. 6, 19, 74, 75, 76, 101, 132

NIGP noisy input Gaussian process. 75, 76, 90, 91, 92, 102, 103

PDF probability density function. 26, 44, 45, 48, 50, 55, 56, 58, 59, 61, 62

POE point of expansion. 33, 35, 36, 37, 38, 40, 42, 47, 52, 73, 94, 95, 96, 102, 105

RKHS reproducing kernel Hilbert space. 13, 14, 15, 17, 18, 132

RV random variable. 58

SAA South Atlantic Anomaly. 100, 117, 118, 119, 121, 131

SH spherical harmonics. 3, 4, 24, 25, 26, 27, 28, 31, 32, 38, 42, 47, 54, 55, 56, 57, 59, 65, 68, 69, 71, 81, 92, 100, 103

SQE squared exponential. 68, 72, 75, 77, 88, 91, 92

vMF von Mises-Fisher. 34, 39

Bibliography

Alldredge, L. R. (1984). "Harmonics Required in Main Field and Secular Variation Models". In: *Journal of geomagnetism and geoelectricity* 36.2, pp. 63–72.

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

Baerenzung, J. and M. Holschneider (in preparation). "Kalmag: A high spatio temporal model of the geomagnetic field". In: *Earth Planets Space*.

Baerenzung, J. et al. (2020). "The Kalmag model as a candidate for IGRF-13." In: *Earth Planets Space* 72, p. 163.

Bausch, J. (2013). "On the efficient calculation of a linear combination of chi-square random variables with an application in counting string vacua". In: *Journal of Physics A: Mathematical and Theoretical* 46.50, p. 505202.

Berlinet, A. and C. Thomas-Agnan (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, Boston, MA.

Berrar, D. (2018). "Bayes' theorem and naive Bayes classifier". In: *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics; Elsevier Science Publisher: Amsterdam, The Netherlands*, pp. 403–412.

Bloxham, J. and A. Jackson (1992). "Time-dependent mapping of the magnetic field at the core-mantle boundary". In: *Journal of Geophysical Research: Solid Earth* 97.B13, pp. 19537–19563.

Bouligand, C. et al. (2005). "Statistical paleomagnetic field modeling and dynamo numerical simulation". In: *Geophysical Journal International* 161, pp. 603–626.

Bouligand, C. et al. (2016). "Frequency spectrum of the geomagnetic field harmonic coefficients from dynamo simulations". In: *Geophysical Journal International* 207, pp. 1142–1157.

Brown, M. C. et al. (2015). "GEOMAGIA50.v3: 2. A new paleomagnetic database for lake and marine sediments". In: *Earth, Planets and Space* 67.1, p. 70.

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

Campuzano, S. et al. (2019). "Emergence and evolution of the South Atlantic Anomaly revealed by the new paleomagnetic reconstruction SHAWQ2k". In: *Earth and Planetary Science Letters* 512, pp. 17–26.

Charlier, B. et al. (2021). "Kernel Operations on the GPU, with Autodiff, without Memory Overflows". In: *Journal of Machine Learning Research* 22.74, pp. 1–6.

Coe, R. S. (1967). "The Determination of Paleo-Intensities of the Earth's Magnetic Field with Emphasis on Mechanisms which Could Cause Non-ideal Behavior in Thellier's Method". In: *Journal of geomagnetism and geoelectricity* 19.3, pp. 157–179.

Constable, C. and M. Korte (2015). "5.09 - Centennial- to Millennial-Scale Geomagnetic Field Variations". In: *Treatise on Geophysics (Second Edition)*. Ed. by G. Schubert. Second Edition. Oxford: Elsevier, pp. 309–341.

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

Constable, C. G. and R. L. Parker (1988). "Statistics of the geomagnetic secular variation for the past 5 m.y." In: *Journal of Geophysical Research: Solid Earth* 93.B10, pp. 11569–11581.

Constable, C., M. Korte, and S. Panovska (2016). "Persistent high paleosecular variation activity in southern hemisphere for at least 10 000 years". In: *Earth and Planetary Science Letters* 453, pp. 78 –86.

Deserno, M. (2004). "How to generate equidistributed points on the surface of a sphere". In: *If Polymerforshung (Ed.)*, p. 99.

Farquharson, C. G. and D. W. Oldenburg (1998). "Non-linear inversion using general measures of data misfit and model structure". In: *Geophysical Journal International* 134.1, pp. 213–227.

Gardner, J. R. et al. (2018). "GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration". In: *CoRR* abs/1809.11165.

Gillet, N. et al. (2013). "Stochastic modeling of the Earth's magnetic field: Inversion for covariances over the observatory era". In: *Geochemistry, Geophysics, Geosystems* 14.4, pp. 766–786.

Gillet, N. (2019). "Spatial And Temporal Changes Of The Geomagnetic Field: Insights From Forward And Inverse Core Field Models". In: *Geomagnetism, aeronomy and space weather: a journey from the Earth's core to the sun*.

Gubbins, D. and J. Bloxham (1985). "Geomagnetic field analysis – III. Magnetic fields on the core-mantle boundary". In: *Geophysical Journal of the Royal Astronomical Society* 80.3, pp. 695–713.

Hartmann, G. A. and I. G. Pacca (2009). "Time evolution of the South Atlantic Magnetic Anomaly". en. In: *Anais da Academia Brasileira de Ciências* 81, pp. 243 –255.

Hellio, G. and N. Gillet (2018). "Time-correlation-based regression of the geomagnetic field from archeological and sediment records". In: *Geophysical Journal International* 214.3, pp. 1585–1607.

Hellio, G. et al. (2014). "Stochastic modelling of regional archaeomagnetic series". In: *Geophysical Journal International* 199, pp. 931–943.

Holschneider, M. et al. (2016). "Correlation-based modeling and separation of geomagnetic field components". In: *Journal of Geophysical Research: Solid Earth* 121.5, pp. 3142–3160.

Jackson, A., A. Jonkers, and M. Walker (2000). "Four centuries of geomagnetic secular variation from historical records". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 358.1768, pp. 957–990.

Jackson, A. and C. Finlay (2015). "Geomagnetic Secular Variation and Its Applications to the Core". English. In: *Treatise on Geophysics*. Ed. by G. Schubert. 2nd. Vol. 5. United Kingdom: Elsevier, pp. 137–184.

Jonkers, A. R. T., A. Jackson, and A. Murray (2003). "Four centuries of geomagnetic data from historical records". In: *Reviews of Geophysics* 41.2.

Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems". In: *Transactions of the ASME–Journal of Basic Engineering* 82.Series D, pp. 35–45.

Kanagawa, M. et al. (2018). *Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences*.

Khokhlov, A., G. Hulot, and C. Bouligand (2006). "Testing statistical palaeomagnetic field models against directional data affected by measurement errors". In: *Geophysical Journal International* 167.2, pp. 635–648.

King, D. E. (2009). "Dlib-ml: A Machine Learning Toolkit". In: *Journal of Machine Learning Research* 10, pp. 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. and C. G. Constable (2003). "Continuous global geomagnetic field models for the past 3000 years". In: *Phys. Earth Planet. Interiors* 140, pp. 73–89.

Korte, M., F. Donadini, and C. Constable (2009). "Geomagnetic field for 0-3ka: 2. A new series of time-varying global models". In: *Geochem. Geophys. Geosys.* 10, Q06008.

Korte, M. et al. (2005). "Continuous geomagnetic field models for the past 7 millennia: 1. A new global data compilation". In: *Geochemistry, Geophysics, Geosystems* 6.2.

Langel, R. and W. Hinze (1998). *The Magnetic Field of the Earth's Lithosphere: The Satellite Perspective*. Cambridge University Press.

Licht, A. et al. (2013). "Ensembles of low degree archeomagnetic field models for the past three millennia". In: *Physics of the Earth and Planetary Interiors* 224, pp. 38 –67.

Livermore, P. W. et al. (2018). "Transdimensional inference of archeomagnetic intensity change". In: *Geophysical Journal International* 215.3, pp. 2008–2034.

Love, J. J. and C. G. Constable (2003). "Gaussian statistics for palaeomagnetic vectors". In: *Geophysical Journal International* 152.3, pp. 515–565.

Lowes, F. J. (1974). "Spatial power spectrum of the main geomagnetic field, and extrapolation to the core". In: *Geophysical Journal International* 36.3, pp. 717–730.

Mandea, M. et al. (2007). "The magnetic field changing over the Southern African continent - a unique behaviour". In: *S. Afr. J. Geol.* 110, pp. 193–202.

Mauerberger, S et al. (2020). "Correlation based snapshot models of the archeomagnetic field". In: *Geophysical Journal International*. ggaa336.

McHutchon, A. and C. E. Rasmussen (2011). "Gaussian Process Training with Input Noise". In: *Advances in Neural Information Processing Systems 24*. Ed. by J. Shawe-Taylor et al. Curran Associates, Inc., pp. 1341–1349.

Merrill, R. T., M. W. McElhinny, and P. L. McFadden (1996). *The Magnetic Field of the Earth: Paleo-magnetism, the Core, and the Deep Mantle*. Academic Press, San Diego.

Minka, T. P. (2013). "Expectation Propagation for approximate Bayesian inference". In: *CoRR* abs/1301.2294.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press.

Nagata, T., Y. Arai, and K. Momose (1963). "Secular Variation of the Geomagnetic Total Force during the Last 5000 Years". In: *Journal of Geophysical Research (1896-1977)* 68.18, pp. 5277–5281.

Nilsson, A. and N. Suttie (2021). "Probabilistic approach to geomagnetic field modelling of data with age uncertainties and post-depositional magnetisations". In: *Physics of the Earth and Planetary Interiors* 317, p. 106737.

Nilsson, A. et al. (2014). "Reconstructing Holocene geomagnetic field variation: new methods, models and implications". In: *Geophysical Journal International* 198.1, pp. 229–248.

Olsen, N. (2002). "A model of the geomagnetic field and its secular variation for epoch 2000 estimated from Ørsted data". In: *Geophysical Journal International* 149.2, pp. 454–462.

Opper, M. and C. Archambeau (2009). "The Variational Gaussian Approximation Revisited". In: *Neural Computation* 21.3, pp. 786–792.

Owen, D. B. (1980). "A table of normal integrals". In: *Communications in Statistics - Simulation and Computation* 9.4, pp. 389–419.

Pavón-Carrasco, F. J. et al. (2014). "A geomagnetic field model for the Holocene based on archaeomagnetic and lava flow data". In: *Earth and Planetary Science Letters* 388, pp. 98–109.

Piper, J. D. A. (1989). "Paleomagnetism". In: *Geomagnetism*. Ed. by J. A. Jacobs. Vol. 3. Academic Press, pp. 31–61.

Rasmussen, C. and C. Williams (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA.

Rauch, H. E., F. Tung, and C. T. Striebel (1965). "Maximum likelihood estimates of linear dynamic systems". In: *AIAA Journal* 3.8, pp. 1445–1450.

Ropp, G. et al. (2020). "Sequential modelling of the Earth's core magnetic field." In: *Earth Planets Space* 72, p. 153.

Rue, H., S. Martino, and N. Chopin (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 319–392.

Sanchez, S. et al. (2016). "Modelling the archaeomagnetic field under spatial constraints from dynamo simulations: a resolution analysis". In: *Geophysical Journal International* 207.2, pp. 983–1002.

Sanchez, S. M. and P. J. Sanchez (2005). "Very Large Fractional Factorial and Central Composite Designs". In: *ACM Trans. Model. Comput. Simul.* 15.4, pp. 362–377.

Schanner, M., M. Korte, and M. Holschneider (2021). "ArchKalMag14k: A Kalman-filter based global geomagnetic model for the Holocene". In: Submitted to Journal of Geophysical Research: Solid Earth.

Schanner, M. et al. (2021). "Correlation Based Time Evolution of the Archeomagnetic Field". In: *Journal of Geophysical Research: Solid Earth* 126.7, e2020JB021548.

Schanner, M. A. and S. Mauerberger (2020). *CORBAM: CORrelation Based Archeomagnetic Modeling*. Potsdam: GFZ Data Services.

Schanner, M. A., S. Mauerberger, and M. Korte (2020). *`pymagglobal` - Python interface for global geomagnetic field models*. Potsdam: GFZ Data Services.

Schanner, M. and S. Mauerberger (2019). *CORBASS: CORrelation Based Archeomagnetic SnapShot model V.1.0*.

Scott, D. W. (1979). "On optimal and data-based histograms". In: *Biometrika* 66.3, pp. 605–610.

Senftleben, R. (2019). "Earth's magnetic field over the last 1 000 years". PhD thesis. University of Potsdam.

Smola, A. J. and P. Bartlett (2000). "Sparse Greedy Gaussian Process Regression". In: *Proceedings of the 13th International Conference on Neural Information Processing Systems*. NIPS'00. Denver, CO: MIT Press, pp. 598–604.

Snelson, E., C. E. Rasmussen, and Z. Ghahramani (2003). "Warped Gaussian Processes". In: *IN ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NIPS*. MIT Press, p. 2003.

Suttie, N. and A. Nilsson (2019). "Archaeomagnetic data: The propagation of an error". In: *Physics of the Earth and Planetary Interiors* 289, pp. 73 –74.

Tauxe, L. and T. Yamazaki (2015). "5.13 - Paleointensities". In: *Treatise on Geophysics (Second Edition)*. Ed. by G. Schubert. Second Edition. Oxford: Elsevier, pp. 461–509.

Tauxe, L. et al. (2018). *Essentials of Paleomagnetism*. Vol. 5th Web Edition. UC Press.

Terra-Nova, F. et al. (2017). "Relating the South Atlantic Anomaly and geomagnetic flux patches". In: *Physics of the Earth and Planetary Interiors* 266, pp. 39–53.

Thébault, E. et al. (2015). "International Geomagnetic Reference Field: the 12th generation". In: *Earth, Planets and Space* 67.1, p. 79.

Thellier, E. and O. Thellier (Jan. 1959). "Sur l'intensité du champ magnétique terrestre dans le passé historique et géologique". In: *Annales de Geophysique* 15, p. 285.

Titsias, M. K., M. Rattray, and N. D. Lawrence (2011). "Markov chain Monte Carlo algorithms for Gaussian processes". In: *Bayesian Time Series Models*. Ed. by D. Barber, A. T. Cemgil, and S. Chiappa. Cambridge Universtiy Press.

Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.

Walker, M. and A. Jackson (2008). "Robust modeling of the Earth's magnetic field". In: *Geophysical Journal International* 143, pp. 799 –808.

Walker, M. R. and A. Jackson (2000). "Robust modelling of the Earth's magnetic field". In: *Geophysical Journal International* 143.3, pp. 799–808.